FEDER/2: program for static and dynamic conformational energy analysis of macro-molecules in dihedral angle space

Hiroshi Wako a, Shigeru Endo b, Kuniaki Nagayama b, Nobuhiro Gō c

a School of Social Sciences, Waseda University, Nishi-Waseda, Shinjuku-ku, Tokyo 169-50, Japan
b Protein Array Project, ERATO, JRDC, 5-9-1, Tokodai, Tsukuba 300-26, Japan
c Department of Chemistry, Faculty of Sciences, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

Received 17 October 1994; revised 16 January 1995

Abstract

The computer program, FEDER/2, has been developed to carry out static and dynamic conformational energy analysis of macromolecules by treating dihedral angles as independent variables. The original program, FEDER (H. Wako and N. Gō, J. Comput. Chem. 8 (1987) 625), developed to rapidly calculate first and second derivatives of a conformational energy function with respect to dihedral angles for a single one protein molecule, has been extended by generalizing the tree representation and by revising the program. The tree topology of a molecular system, which is essential to this program, is defined in terms of rigid ‘unit’ and rotatable ‘bond’. Then, algorithms and formulae based on the tree topology are developed to calculate the first and second derivatives. In this revised version of the program we have constructed a library of units, from which a set of data required for a given specific system of molecules is generated. By separating such parts of the program that take care of specifics of various molecular systems into a form of data in the library of units, the main part of the program to calculate the first and second derivatives has become general enough to be applicable to wider types of molecules.

Keywords: Conformational analysis; Energy calculations; Macromolecules

1. Introduction

Computer programs developed for static and dynamic conformational energy analysis of biopolymer systems can be classified into two groups depending on independent variables used. In one group including CHARMM [1] and AMBER [2], atomic Cartesian coordinates in a molecule are used as independent variables. In the other group including ECEPP [3,4], FEDER [5], FANTOM [6], and DADAS90 [7], only dihedral angles are treated as independent variables by keeping bond lengths and bond angles fixed. Computer programs of the former type have a merit that they are simpler, whereas those of the latter type have the merit that the space-time load on the computer is much less owing to the use of a smaller number of independent variables, which is about one eighth of that of the former type, when applied for proteins [5].

The key to the programs of the latter type is the discovery of an algorithm for the fast calculation of the first and second derivatives of the conformational energy function with respect to dihedral angles [8,9]. Extending this algorithm the computer program FEDER (Fast ECEPP with DERivatives) was developed so as to adapt it to a supercomputer with vector
processors [5] (ECEPP is a program developed for polypeptide by the group of Scheraga [3,4]). The rapidly calculable first and second derivatives are used for rapidly converging energy minimization with the Newton–Raphson method [5] and for conformational dynamics studies. For use in normal mode analysis explicit expressions are derived [10] for the coefficient matrix that converts small changes in dihedral angles to small atomic displacements, and for the coefficient matrix in the expression of the kinetic energy of internal motions in terms of first order time derivatives of variable dihedral angles. Normal mode analyses have been performed on various protein molecules [11–20] with dihedral angles as independent variables. The effect of fixing bond lengths and bond angles in the normal mode analysis has been compared with the analysis treating Cartesian coordinates as independent variables [21].

Normal mode analysis is based on the assumption of the harmonicity of the energy surface. For studying nonlinear aspects of conformational dynamics, a Monte Carlo simulation method with an anisotropic step size was invented, in which variables corresponding to the normal modes are used as independent collective variables [22]. By this method a more efficient sampling of conformations than the method of molecular dynamics can be achieved [22,23].

Lately, progress has been made in the development of the program and the algorithm. Firstly, it has been shown that the algorithm, which was originally developed for a system consisting of one molecule, can be expanded to a system of two or more molecules with minor modifications [24,25]. Dynamic properties of deoxymyoglobin have been studied by the conformational normal mode analysis with this extended algorithm [26,27]. Secondly, it has been realized that the calculations with dihedral angles as independent variables are useful and powerful in determination of solution conformations of proteins based on NMR distance analysis [6,7,28,29]. DADAS90 is an improved version of DADAS [28] containing a special version of FEDER for that purpose [7]. Thirdly, hydration energy has been taken into account through a hydration-shell model proposed by Ooi et al. [30]. Differences between the simulations with and without hydration energy are examined [31].

It is to incorporate these improvements that the second version of FEDER, termed FEDER/2, has been developed. In this paper characteristic points of FEDER/2 are described. At first, general and explicit definitions are given for a tree representation of a system consisting of any number of molecules, because such a representation is essential for the algorithm. Some properties of the tree representation useful for programming are derived from them. The algorithm and formulæ given in the previous papers (Refs. [5,8,9,25]) are summarized in terms of these definitions and properties.

Secondly, we have constructed a library of units, in which necessary information about various types of units is stored. The chemical structure of a given molecule represented by a sequence of unit names is converted to a set of data required in the FEDER/2 computation (referred to as MOL data) by a supplementary program MKMOL. The MOL data includes properties to represent a system in a tree structure, and energy parameters assigned to atoms and rotatable bonds. Once the MOL data specific to a given molecule is generated, the main part of the program for static and dynamic conformational energy analysis works in a way that is independent of the type of molecule. This is in contrast to the original FEDER, which involved some treatments that were specific to proteins. Therefore, FEDER/2 can, in principle, treat any molecule with a tree topology, once the MOL data for the molecule can be constructed from the unit library. To treat new types of molecules, we only have to add the information about relevant units to the unit library.

In this paper we consider a system comprising one or more molecules, each of which can be represented by a tree in the sense of graph theory. Independent variables of a system are (a) dihedral angles about rotatable covalent bonds and (b) translational and orientational variables between molecules. Bond lengths and bond angles of the molecules are kept fixed in computation. In the next section we define a tree topology of a single molecule and its properties are studied. In Section 3 the formulation and algorithm to calculate the first and second derivatives of the conformational energy and/or objective functions are described according to the tree topology. In Section 4 it is shown that the formulation and algorithm can be extended to a system consisting of two or more molecules with minor modifications. Such a
system can be also represented by a tree topology. In Section 5 we describe a library of units, with which the part of the program specific to a molecule to be calculated can be separated from the main part. Consequently, the main part of the program becomes applicable to wider types of molecules.

Fig. 1. Example of a system, which consists of three molecules. Circles and squares represent real and pseudo atoms, respectively, and figures in them are serial atom numbers. Three molecular chains comprise atoms 1 to 168, 174 to 588, and 594 to 846, respectively. Atoms 169 to 173 and 589 to 593 are pseudo ones corresponding to translational and rotational variables. Enclosing curves and arrows represent units and bonds, respectively. Three molecular chains comprises units 0 to 46, 52 to 177, and 183 to 267, and bonds 1 to 46, 53 to 177, and 184 to 267, respectively (serial unit and bond numbers are written in italics and romans, respectively). Units 47 to 51 and 178 to 182 and bonds 47 to 52 and 178 to 183 correspond to translational and rotational variables. There is an interchain bond between the first and second molecules, which is treated by regarding the two segments of atoms 158–162–163–164 and 584–583–582–578 as coincident with each other; i.e., pseudo atoms 163, 164, 583 and 584 are regarded as the same atoms with real atoms 582, 578, 162 and 158, respectively, and the artificial potential (A6) in Appendix A is imposed on these pairs of the pseudo and real atoms.
Table 1
Glossary

- $A_{\alpha}$: accessible surface area of atom $\alpha$ (Eq. (A.7)).
- $A_{\alpha\beta}$, $B_{\alpha\beta}$: parameters for Lennard–Jones type non-bonded energy functions (Eq. (A.2)).
- $a$, $b$: serial unit or bond numbers.
- BRANCH($b$): tail unit number of bond $c$ that has the smallest bond number satisfying $MDESCN(b) = MDESCN(c)$.
- $c(a)$: child bond number of bond $a$.
- $c_{\alpha\beta}$: defined by Eq. (B.5).
- $D$: dielectric constant (Eq. (A.3)).
- $d_{\alpha\beta}$: defined by Eq. (B.13).
- DEPTH($a$): depth of unit $a$ (Eq. (1)).
- $E$: sum of the conformational energy and objective functions (Eq. (A.13)).
- $E_c$: conformational energy function (Eq. (A.1)).
- $E_{el}$: electrostatic energy (Eq. (A.3)).
- $E_{hyd}$: hydration energy (Eq. (A.7)).
- $E_{LJ}$: Lennard–Jones type non-bonded energy (Eq. (A.2)).
- $E_1$: objective function for lower limit constraints (Eq. (A.10)).
- $E_{ob}$: objective function for dihedral angle constraints (Eq. (A.11)).
- $E$: objective function for upper limit constraints (Eq. (A.9)).
- $E_{X}$: objective function for a regularization of X-ray data (Eq. (A.12)).
- $e$: factor to convert the dimension of the objective function to that of energy (Eq. (A.13)).
- $e_a$: unit vector lying on bond $a$.
- $F_a$: defined by Eq. (5).
- $f_a$: defined by Eq. (7).
- $J_{a\beta}$: defined by Eq. (B.4).
- $g_{\alpha}$: hydration free energy per unit ASA for atom $\alpha$ (Eq. (A.7)).
- $h_{\mu\nu}$: defined by Eq. (D.3).
- $k(a, b)$: distance between bonds $a$ and $b$ along the chain (Eq. (2)).
- $L_{\alpha\beta}$: defined by Eq. (B.6).
- $l_{ab}$: defined by Eq. (11).
- $l_{\alpha\beta}$: lower limit distance between atoms, $\alpha$ and $\beta$ (Eq. (A.10)).
- $M_a$: a set of atoms in a subtree $t_a$.
- $\overline{M}_a$: a set of atoms complementary to $M_a$.
- $MDESCN(b)$: maximum bond number of descendants of bond $b$.
- $n_{a}$: multiplicity in a torsional potential for bond $a$ (Eq. (A.4)).
- $p$: factor to determine weights on potential and objective functions (Eq. (A.13)).
- $p(a)$: parent bond number of bond $a$.
- $Q_{\alpha}$: partial charge of atom $\alpha$ (Eq. (A.3)).
- $q_i$: representative variables of $\theta_i$, $\tau_i$ and $\sigma_i$ (Section 4).
- $R_{ab}$: defined by Eq. (10).
- $r_a$: position vector of a certain point properly chosen on bond $a$.
- $r^F$: position vector of an atom with respect to $F$-system.
- $r^S$: position vector of an atom with respect to $S$-system.
- $r_{\alpha\beta}$: distance between atoms, $\alpha$ and $\beta$.
- $S_{ab}$: $-1$ if bond $a$ is a descendant of bond $b$, and $1$ otherwise (Eq. (9)).
- $T$: rotation matrix.
- $t_a$: subtree with a joint bond $a$.
- $t_{\mu\nu}$: element of matrix $T$ (Eq. (18)).
- $U_a$: barrier height of rotation about bond $a$ (Eq. (A.4)).
- $u_{\alpha\beta}$: upper limit distance between atoms, $\alpha$ and $\beta$ (Eq. (A.9)).
Table 1 (continued)

<table>
<thead>
<tr>
<th>$W_{a\beta}$</th>
<th>defined by Eqs. (B.6) and (B.7).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{a,\alpha\beta}$</td>
<td>weighting factor for a lower limit distance (Eq. (A.10)).</td>
</tr>
<tr>
<td>$w_{l,a,\beta}$</td>
<td>weighting factors for loop closure (Eqs. (A.5) and (A.6)).</td>
</tr>
<tr>
<td>$w_{lp,a,\beta}$</td>
<td>weighting factor for an upper limit distance (Eq. (A.9)).</td>
</tr>
<tr>
<td>$w_{3,a,\beta}$</td>
<td>weighting factor in an objective function for regularization (Eq. (A.12)).</td>
</tr>
<tr>
<td>$w_{0,a}$</td>
<td>weighting factor for a dihedral angle constraint (Eq. (A.11)).</td>
</tr>
<tr>
<td>$X_{a\alpha\beta}$</td>
<td>defined by Eqs. (B.6) and (B.8).</td>
</tr>
<tr>
<td>$Y_{a\alpha\beta}$</td>
<td>defined by Eqs. (B.6) and (B.9).</td>
</tr>
<tr>
<td>$Z_{a\alpha\beta}$</td>
<td>unit vector pointing from atom $\alpha$ to atom $\beta$ (Eq. (B.3)).</td>
</tr>
<tr>
<td>$\alpha$, $\beta$</td>
<td>defined by Eqs. (B.6) and (B.10).</td>
</tr>
<tr>
<td>$\gamma_1$, $\gamma_2$, $\gamma_3$</td>
<td>serial atom numbers.</td>
</tr>
<tr>
<td>$\delta_{i\alpha}$</td>
<td>unit vectors defined by Eq. (D.1).</td>
</tr>
<tr>
<td>$e_{i\alpha\beta}$</td>
<td>defined by Eq. (B.11).</td>
</tr>
<tr>
<td>$e_{x,y,z}$</td>
<td>defined by Eq. (B.12).</td>
</tr>
<tr>
<td>$\xi_{i,\alpha\beta}$</td>
<td>defined by Eq. (D.2).</td>
</tr>
<tr>
<td>$\eta_{i,\alpha\beta}$</td>
<td>factor for 1-4 interactions in non-bonded energy function (Eq. (A.2)).</td>
</tr>
<tr>
<td>$\theta_{i,\alpha\beta}$</td>
<td>dihedral angle about bond $\alpha$.</td>
</tr>
<tr>
<td>$\pi_{i,b_1, b_j}$</td>
<td>path between units, $b_1$ and $b_j$.</td>
</tr>
<tr>
<td>$\pi_{i}(b_1 \rightarrow b_j)$</td>
<td>one-direction path from unit $b_1$ to unit $b_j$.</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>vector used in a multiple-chain system, which has the same direction as the direction axis of rotation and the length equal to the magnitude of rotation. $\sigma = (\sigma_1, \sigma_2, \sigma_3)$.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>position vector of the origin of the second system with respect to the first system. $\tau = (\tau_1, \tau_2, \tau_3)$.</td>
</tr>
<tr>
<td>$\kappa$, $\mu$, $\nu$, $\xi$</td>
<td>($x$, $y$, $z$)-coordinates or three components of $\sigma$.</td>
</tr>
<tr>
<td>$\Phi_{i}(r_{a\beta})$</td>
<td>conformational energy and/or objective function given as a function of interatomic distances, $r_{a\beta}$.</td>
</tr>
<tr>
<td>$\psi_{i}(\theta_{a})$</td>
<td>conformational energy and/or objective function given as a function of dihedral angles, $\theta_{a}$.</td>
</tr>
<tr>
<td>$\psi_{i}$</td>
<td>defined in Table 2.</td>
</tr>
</tbody>
</table>

2. Tree structure representation of a single molecule

At first we consider a single molecule, in which bond lengths and bond angles are kept fixed and only dihedral angles about rotatable bonds are treated as independent variables. A molecule is assumed to have a tree topology comprising units and bonds. Units and bonds are defined as follows. (The example given in Fig. 1 may be helpful for readers to understand the following definitions and properties. The variables and parameters to appear in this paper are summarized in Table 1.)

**Definition 1. (unit and bond)** A **bond** is a rotatable covalent bond in a molecule. The molecule is divided into units by the bonds. In other words a **unit** is defined as a set of atoms separated by the bonds. Every atom must belong to one of the units.

The term **bond** is used only for a rotatable bond in this study, since there is no need for taking into account explicitly unrotatable covalent bonds in the formulation and algorithm described below. Unrotatable bonds must be included in units. A bond is thus synonymous with an independent variable. Interactions between atoms in the same unit need not be considered, since mutual distances between them never change by any rotation about any bond.

In Fig. 1 units 0, 1 and 2, for example, consist of atoms 1 to 3, 4 to 5, and 6 to 9, respectively. Rotatable bonds such as those connecting atoms 3-4, 4-6, 4-13, etc. (i.e., connecting units 0-1, 1-2, 1-5, etc.) are referred to as bonds in this study. For unrotatable bonds such as those connecting atoms 4-5 in unit 1, 6-7 and 6-8 in unit 2, etc., connectivity of atoms does not have much meaning in the calculation except that it is necessary for classification of interactions between atoms into 1-4 and 1-5 interactions.
types [3]. We can also introduce a pseudo-atom such as atom 9 in unit 2, whose distances from the other atoms in the same unit are fixed.

**Definition 2. (path and tree structure)** A path is a sequence of bonds, \(b_1, b_2, \ldots, b_n\), connecting units, \(a_1, a_2, \ldots, a_n, a_{n+1}\), respectively, as \(a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_n \rightarrow a_{n+1}\). The path between \(b_1\) and \(b_n\) is denoted by \(\pi(b_1, b_n)\). The units \(a_1\) and \(a_{n+1}\) are called end units of the path \(\pi(b_1, b_n)\). A length of a path is defined as the number of units on the path exclusive of the end units. As a special case, a single bond has a path of length 0. If path is unique between any pair of bonds, the molecule is said to have a tree topological structure, or simply a tree structure.

**Definition 3. (directed tree structure and directed bonds)** In a tree structure any one of the units can be defined as a root unit. Between the root unit and any one unit in the molecule, there is a unique path as required in Definition 2. We now introduce direction of the path such that it is directed from the root to the unit of the other end. Directions of bonds are also defined to be the same as the direction of the path involving them.

**Definition 4. (tail and head units, parent and child units, and entering and outgoing bonds)** If a bond connects units \(a\) and \(b\) in the direction of \(a \rightarrow b\), we call \(a\) and \(b\) tail and head units of this bond, respectively. The units \(a\) and \(b\) are also called a parent of unit \(b\) and a child of unit \(a\), respectively. We call this bond an entering bond of unit \(b\) and an outgoing bond of unit \(a\).

Every unit except the root unit has exactly one parent unit. Every unit including the root unit can have any number of child units, including zero unit (see Fig. 1).

Since, as defined below, the bonds entering the units \(a\) and \(b\) are also called \(a\) and \(b\), respectively, we assume the same relationships between the bonds \(a\) and \(b\); i.e., bond \(a\) is a parent of bond \(b\) and bond \(b\) is a child of bond \(a\). In Fig. 1 directed bonds are represented by arrows. For example, units 1 and 5 are tail and head units of bond 5, respectively, and unit 5 is called a child of unit 1 and unit 1 a parent of unit 5.

**Definition 5. (one-direction path)** If every bond in a path has the same direction such as \(a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_n \rightarrow a_{n+1}\), the path is called a one-direction path and denoted by \(\pi(b_1 \rightarrow b_n)\).

In Fig. 1 the path \(\pi(1, 7)\), for example, is a sequence of bonds 1, 5, 6 and 7, and a one-direction path. The path \(\pi(3, 7)\) is a sequence of bonds 3, 2, 5, 6 and 7, but not a one-direction path. According to Definition 3 any path starting from the root unit is of one-direction.

**Definition 6. (numbering of bonds and units)** The numbering of units are defined by a preorder traversal of a tree [32]; i.e., the order to visit each unit of the tree in the following manner defines the serial number of each unit:

(a) Visit the root unit and assign a unit number \(a = 0\) to it.

(b) If the unit has children not visited yet, visit one of them and assign a unit number \(a \leftarrow a + 1\); otherwise, return to the parent unit. This procedure is repeated recursively until all units are visited.

(c) To a bond the number of its head unit is assigned.

The numbering of units and bonds can be made unique, if the order of outgoing bonds to visit is given for units with two or more outgoing bonds in advance.

**Definition 7. (ancestor, descendant, subtree and joint bond)** If there is a one-direction path from bond \(b\) to bond \(a\), then bond \(b\) is called an ancestor of bond \(a\), and bond \(a\) a descendant of bond \(b\). Each bond is regarded as an ancestor and a descendant of itself. A set of bonds involving bond \(b\) and all its descendants constructs a subtree, which is denoted by \(t_b\). The bond \(b\) is called a joint bond of the subtree.

In Fig. 1 bond 1, for example, is an ancestor of bonds 1 to 46, and bonds 1 to 46 are descendants of bond 1. Bonds 2 to 4 construct a subtree, the joint bond of which is bond 2.

**Definition 8. (branching bond)** If a path between bonds \(a\) and \(b\) is not of one-direction, there is exactly one common ancestor bond of \(a\) and \(b\).
whose head unit is on the path \( \pi(a, b) \). We call such a bond a branching bond of the path.

In Fig. 1 the path \( \pi(4, 7) \), for example, is not a one-direction path. Bond 1 is a branching bond of this path, because bond 1 is a common ancestor of both bonds 4 and 7 and its head unit (i.e., unit 1) is on the path.

In the above numbering system all bond numbers in a subtree with a joint bond \( b \) are greater than \( b \). In other words, making use of the maximum bond number of the descendants of \( b \), \( \text{MDESCN}(b) \), we can say that bond \( a \) is on a subtree \( t_b \) if and only if \( b \leq a \leq \text{MDESCN}(b) \). Owing to this property we can easily determine whether or not a given bond \( a \) is a descendant of \( b \). For example, \( \text{MDESCN}(1) = 46 \) and \( \text{MDESCN}(2) = 4 \) in Fig. 1. Once these values are given, we can know that bond 3 is a descendant of both bonds 1 and 2, whereas bond 6 is a descendant of bond 1 but not of bond 2, without examining Fig. 1.

In the previous papers [5,8,9] we called the subtree with the joint bond \( b \), \( t_b \), a moving side with respect to bond \( b \) and the remaining part of the tree was called a fixed side. A kinship of a pair of bonds, \( a \) and \( b \), is essential to the calculation of second derivatives with respect to dihedral angles as independent variables.

In the recurrent equations to rapidly calculate the second derivatives a distance between two bonds is used (see Appendix C). In the following, the distance of a given unit from the root unit is specifically called depth, and it is shown that the distance between any pair of bonds can be calculated from the depths of the units.

**Definition 9. (depth of a unit)** The depth of a unit \( a \) in a tree is the length of the path from the root unit to \( a \) and denoted by \( \text{DEPTH}(a) \).

The depth of a unit \( a \) is calculated easily by the following recurrent relation;

\[
\text{DEPTH}(a) = \begin{cases} 
0 & \text{if } a \text{ is the root unit,} \\
\text{DEPTH}(p(a)) + 1 & \text{otherwise,}
\end{cases}
\]

where \( p(a) \) denotes the parent unit of \( a \).

**Definition 10. (distance between bonds)** A distance between bonds \( a \) and \( b \), \( k(a, b) \), is defined as the number of units on the path, \( \pi(a, b) \).

Assume that \( a \) is greater than \( b \) (the case \( a = b \) is trivial; \( k(a, a) = 0 \)). If \( a \) is a descendant of \( b \), \( k(a, b) = \text{DEPTH}(a) - \text{DEPTH}(b) \). If \( a \) is not a descendant of \( b \), \( k(a, b) \) is not trivial. It is necessary to find the branching bond of the path \( \pi(a, b) \). A list, \( \text{BRANCH}(b) \), is introduced for this purpose. Assume that \( c \) is the smallest bond number satisfying \( \text{MDESCN}(b) = \text{MDESCN}(c) \). Then, \( \text{BRANCH}(b) \) is defined as the serial number of the tail unit of the bond \( c \) (note that \( c \leq b \)). Trace \( \text{BRANCH} \) to \( c_1 = \text{BRANCH}(b), c_2 = \text{BRANCH}(c_1), \ldots, c_j = \text{BRANCH}(c_{j-1}) \) and find \( c_j \) such that \( \text{MDESCN}(c_{j-1}) < \text{MDESCN}(a) \leq \text{MDESCN}(c_j) \) (where \( c_0 = b \)). Since \( c_j < b < a \leq \text{MDESCN}(a) \leq \text{MDESCN}(c_j) \), the bond \( c_j \) is a common ancestor of the bonds \( a \) and \( b \) and the unit \( c_j \) is on the path \( \pi(a, b) \), i.e., \( c_j \) is the branching bond of the path \( \pi(a, b) \). Eventually, using \( k(a, c_j) = \text{DEPTH}(a) - \text{DEPTH}(c_j) \) and \( k(b, c_j) = \text{DEPTH}(b) - \text{DEPTH}(c_j) \), we obtain the distance between \( a \) and \( b \) as

\[
k(a, b) = \text{DEPTH}(a) + \text{DEPTH}(b) - 2 \text{DEPTH}(c_j) - 1.
\]

In Fig. 1, for example, \( \text{BRANCH}(a) = 1 \) for \( a = 2, 3 \) and 4, \( \text{BRANCH}(a) = 0 \) for \( a = 1, 5, 6, 7, 12, \ldots \), and \( \text{BRANCH}(a) = 41 \) for \( a = 42, 43 \) and 44. For bonds 3 and 7 the branching bond of the path \( \pi(3, 7) \) is 1, and \( \text{DEPTH}(3) = 3, \text{DEPTH}(7) = 4 \) and \( \text{DEPTH}(1) = 1 \). Therefore, \( k(3, 7) = 4 \), which is equal to the number of units between bonds 3 and 7 (i.e., units 2, 1, 5 and 6).

To sum up, a molecule is represented as a directed tree comprising units and directed bonds. Once the root unit is chosen, the directions of bonds are determined uniquely, and the numbering of units and bonds are carried out by a preorder traversal from the root unit. Kinship between bonds and between units is important. Some properties, such as \( S_{ab} \) in Eq. (9) to appear later and \( k(a, b) \) above, used in the calculations of the first and second
derivatives can be obtained immediately from the three one-dimensional arrays, MDESCN, DEPTH, and BRANCH, defined above.

3. Calculations of first and second derivatives

Conformational energy and objective functions used in molecular mechanics calculations usually consist of two different types of terms (Explicit function forms used in FEDER/2 are given in Appendix A);

\[ E(\theta_1, \theta_2, \ldots, \theta_n) = \sum_a \Psi_a(\theta_a) + \sum_{a} \sum_{\beta} \Phi_{a\beta}(r_{a\beta}). \]  

(3)

The terms in the first summation are explicit function of the dihedral angles \( \theta_a \) \((a = 1, 2, \ldots, n)\) so that the first and second derivatives of these terms with respect to the dihedral angles can be easily calculated. Torsional potential around rotatable bonds in a conformational energy function and an objective function for torsional angle constraints derived from NMR J-coupling parameters are of this type.

The terms in the second summation are given as functions of interatomic distances \( r_{a\beta} \) (\( \alpha \) and \( \beta \) are used to specify atoms hereinafter). Lennard-Jones type non-bonded and Coulomb type electrostatic energy functions in a conformational energy function and objective functions for distance constraints in NMR distance analysis belong to this category. Since an interatomic distance is a function of dihedral angles, the terms in the second summation are implicit function of dihedral angles. Calculation of the first and second derivatives of these terms with respect to dihedral angles is not trivial. Rapid computational algorithm for the derivatives was developed by Gō and his coworkers \([5,8,9]\). The algorithm is briefly reviewed below with the terminology introduced in the previous section.

The first derivatives of the conformational energy function and objective function with respect to the \( \alpha \)th dihedral angle \( \theta_a \) is given by the following expressions;

\[ \frac{\partial E}{\partial \theta_a} = - (e_a, e_a \times r_a) F_a, \]  

(4)

\[ F_a = \sum_{\alpha} \sum_{\beta} f_{a\beta}, \]  

(5)

where \( e_a \) is a unit vector lying on bond \( a \), whose direction is the same as that of the bond, and \( r_a \) is a position vector of a certain point properly chosen on the bond \( a \). \( F_a \) and \( f_{a\beta} \) are six-dimensional column vectors. The explicit form of \( f_{a\beta} \) is given in Eq. (B.4). The summations in Eq. (5) are taken over all interactable atom pairs, \( \alpha \in M_a \) and \( \beta \in \overline{M}_a \), where \( M_a \) is a set of atoms in the subtree \( t_a \) with the joint bond \( a \) and \( \overline{M}_a \) a set of atoms complementary to \( M_a \).

Eq. (5) can be effectively calculated by the following recurrent relations;

\[ F_a = f_a^0 + \sum_{j = c(a)} f_j, \]  

(6)

where

\[ f_a^0 = \sum_b f_{a,b}^*, \]  

(7)

\[ f_{a,b}^* = \sum_{\alpha \in a} \sum_{\beta \in b} f_{a\beta}. \]  

(8)

The summations in Eqs. (6), (7) and (8) are taken, respectively, over the child bonds of bond \( a \), \( c(a) \), over all units \( b \) except \( a \), and over the atoms \( \alpha \) in unit \( a \) and \( \beta \) in unit \( b \).

The second derivative with respect to the \( \alpha \)th and \( \beta \)th dihedral angles, \( \theta_a \) and \( \theta_b \) \((b \leq \alpha)\), is given by the following expressions \([5,9]\);

\[ \frac{\partial^2 E}{\partial \theta^2_a} = - S_{ab} (e_a, e_a \times r_a) R_{ab} \left( e_b, e_b \times r_b \right), \]  

(9)

\[ R_{ab} = \sum_{\alpha} \sum_{\beta} L_{\alpha\beta}, \]  

(10)

where \( S_{ab} = -1 \) if bond \( a \) is a descendant of bond \( b \), and \( S_{ab} = 1 \) otherwise. \( S_{ab} \), i.e., the kinship of the pair of bonds \( a \) and \( b \), can be easily determined from MDESCN(b) as shown in the previous section. \( R_{ab} \) and \( L_{\alpha\beta} \) are 6 \( \times \) 6 matrices. \( L_{\alpha\beta} \) is defined by Eq. (B6). The summations in Eq. (10) are taken over atoms \( \alpha \in M_a \) and \( \beta \in \overline{M}_b \), if \( a \) is a descendant of \( b \); otherwise, they are taken over atoms \( \alpha \in M_a \) and \( \beta \in \overline{M}_b \).

For a given pair of bonds \( a \) and \( b \), we define \( I_{a^*b^*} \) as

\[ I_{a^*b^*} = \sum_{\alpha \in a^*} \sum_{\beta \in b^*} L_{\alpha\beta}. \]  

(11)

Here summations extend over atoms \( \alpha \) and \( \beta \) in the end units \( a^* \) and \( b^* \) of a path \( \pi(a, b) \). Making use
of \( I_{a^* b^*} \), we can now write down a recurrent equation for \( R_{ab} \):

\[
R_{ab} = I_{a^* b^*} + \sum_{a'} R_{a^* b'} + \sum_{b'} R_{a b'} - \sum_{a'} \sum_{b'} R_{a' b'}
\]  

(12)

Summations are taken over atoms in the units, \( a' \) and \( b' \), which are attached to the end units, \( a^* \) and \( b^* \), respectively.

The following two points are important for saving storage space during computation. Firstly, although \( R_{b a} \) is not symmetric with respect to \( a \) and \( b \), calculations of \( R_{a b} \) only for \( b \leq a \) are sufficient owing to the following relation [9];

\[
R_{a b} = R_{b a}^t,
\]  

(13)

where \( t \) means transpose. In actual fact, even if we assume \( a \geq b \), it may occur that \( b' > a \) and/or \( b' > a' \) in Eq. (12). Therefore, if \( j > i \) for \( R_{ij} \) in Eq. (12), it is assumed that \( R_{ji} \) is used instead of \( R_{ij} \).

Secondly, the following relations for distances between bonds included in the recurrent Eq. (12);

\[
k(a', b') = k(a, b') = k(a, b) + 1,
\]  

(14)

\[
k(a', b') = k(a, b) + 2.
\]  

(15)

mean that we need to prepare storage space of \( R \)'s only for bond pairs with distances, \( k(a, b), k(a, b) + 1 \) and \( k(a, b) + 2 \) at the same time [5]. Computational procedures to calculate the first and second derivatives based on the above recurrent relations are summarized in Appendix C.

4. Extension to systems consisting of two or more molecules

We consider a system consisting of two molecules here (the original formulation is given in Refs. [24, 25]). Extension to a system of three or more molecules is trivial. A local coordinate system is defined in each unit. We call the local coordinate systems of the root units of the first and second molecules \( F \)- and \( S \)-systems, respectively. If the relative position and orientation of \( S \)-system to \( F \)-system are given, coordinates of all atoms in the two molecules can be expressed in the \( F \)-system. Making use of the position vector of the origin of \( S \)-system with respect to \( F \)-system, \( \tau = (\tau_1, \tau_2, \tau_3) \), and the rotation matrix, i.e., relative orientation of \( S \)-system to \( F \)-system, \( T = \{t_{\mu \nu}\} \ (\mu, \nu = 1, 2, 3) \), the coordinate vector of an atom with respect to \( S \)-system, \( r^S \), is related to the coordinate vector of the same atom with respect to \( F \)-system, \( r^F \), by

\[
r^F = \tau + Tr^S.
\]  

(16)

\( T \) can be expressed by the use of rotation of a certain angle \( \sigma \) about a certain directed axis. We define a vector as \( \sigma = (\sigma_1, \sigma_2, \sigma_3) \) such that it has the same direction as the direction axis of rotation and the length equal to the magnitude of rotation;

\[
\sigma = \left( \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \right)^{1/2}.
\]  

(17)

In terms of the three components of the vector \( \sigma \), elements of the rotation matrix \( T \) are given by

\[
T_{\mu \nu} = \sigma^{-2} (1 - \cos \sigma) \delta_{\mu \nu} + \sigma^{-1} \sin \sigma \sum_{\xi} e_{\mu \xi} \delta_{\xi \nu} \cos \sigma,
\]  

(18)

(\( \mu, \nu = 1, 2, 3 \)). \( e_{\mu \xi} \) and \( \delta_{\mu \nu} \) are given in Eqs. (B11) and (B12), respectively.

The range of variation of the three components, \( \sigma_1, \sigma_2, \) and \( \sigma_3 \) is given by

\[
0 \leq \pi
\]  

(19)

In order to maintain a one-to-one correspondence between a relative orientation of the two systems and the set of \( \sigma = (\sigma_1, \sigma_2, \sigma_3) \), only the following cases are chosen when \( \sigma = \pi \);

(i) \( \sigma_1 > 0 \),

(ii) \( \sigma_1 = 0 \) and \( \sigma_2 > 0 \),

(iii) \( \sigma_1 = \sigma_2 = 0 \) and \( \sigma_3 = \pi \).

(20)

Thus, the independent variables to describe the system of two or more molecules are dihedral angles, \( \theta_\alpha \ (\alpha = 1, 2, \ldots, n) \); with proper numbering throughout all molecules, translational variables, \( \tau_\mu \ (\mu = 1, 2, 3) \), and orientational variables, \( \sigma_\mu \ (\mu = 1, 2, 3) \). The first derivatives of the conformational energy and objective functions with respect to these variables (represented by \( q_i \)) is given by:

\[
\frac{\partial E}{\partial q_i} = - (\phi_i, \psi_i) F_i
\]  

(21)

where \( \phi_i \) and \( \psi_i \) are given in Table 2.
The second derivatives with respect to $q_i$ is given by

$$
\frac{\partial^2 E}{\partial q_i \partial q_j} = - (\phi_i, \psi_i) R_{ij} \left( \begin{array}{c} \phi_i \\ \psi_i \end{array} \right)
$$

(22)

except for the cases that both $q_i$ and $q_j$ are one of $\sigma_\mu$ ($\mu = 1, 2, 3$) for the same molecule. For such exceptional cases the second derivative is given for $\mu \leq \nu$ by:

$$
\frac{\partial^2 E}{\partial \sigma_\mu \partial \sigma_\nu} = - \left( \xi_\mu, \xi_\mu \times \tau \right) R_{\mu\nu} \left( \begin{array}{c} \xi_\nu \\ \xi_\nu \times \tau \end{array} \right)
$$

$$
- \left( \frac{\partial \xi_\mu}{\partial \sigma_\nu}, \frac{\partial \xi_\mu}{\partial \sigma_\nu} \times \tau \right) F_\mu,
$$

(23)

where $\xi_\mu$ and $\partial \xi_\mu/\partial \sigma_\nu$ are given in Appendix D.

As seen above, for a system consisting of two or more molecules the first and second derivatives (Eqs. (21) and (22)) can be expressed in the same form as Eqs. (4) and (9) for a single molecule by extending the definition of $\phi_i$ and $\psi_i$ as given in Table 2, except for the cases that both variables $q_i$ and $q_j$ are one of $\sigma_\mu$ ($\mu = 1, 2, 3$) for the same molecule. Even for the latter cases, the only modification is the inclusion of the second term of Eq. (23).

5. Unit library and construction of a tree structure

We regard molecules as consisting of bonds and units. It is useful to prepare a library of units in which necessary information about each unit type is collected. If some combinations of unit types appear frequently (e.g., amino acid residues in protein molecules), it is better to treat them together in the library; we call a combination of two or more units a block (a single unit can be also conveniently regarded as a block; accordingly, there is no need to differentiate the unit and block from each other. Thus, we call a library of units and blocks simply a unit library).

The chemical structure of a molecule can be given in a sequence of block names as illustrated in Fig. 2. It should be noted that there are two types of blocks, i.e., a root block and a non-root one. A root block has no entering bond and can occur only at the root unit in the first molecule is connected with the root units of other molecules with six pseudo-bonds and five pseudo-units, respectively, which correspond to the translational and orientational variables. (The sixth pseudo-unit is regarded to coincide with the root units of other molecules.) It should be noted that, since the pseudo-bonds and pseudo-units are introduced for the sake of convenience to represent the system as a tree, they are not substances with bond lengths and real atoms. With this numbering and tree representation there is no need to change the recurrent Eqs. (6) and (12) for the first and second derivative calculations even for a system comprised of two or more molecules.

A tree representation of a system comprising three molecules is given in Fig. 1. The bonds 47 to 52 and 178 to 183 correspond to translational and rotational variables for chain 2 and chain 3, respectively, and the atoms 169 to 173 and 589 to 593 are pseudo ones. Since these bonds are not substances with bond lengths, it should be noted that no restrictions are imposed on the distances between units 0 and 52, and between 0 and 183. It should be also noted that there is no need to change the format of the MOL data, when we extend FEDER to a system consisting of two or more molecules.
beginning of the sequence. A non-root block has exactly one entering bond, while the number of outgoing bonds from the block can be zero, one, or more. If there are two or more outgoing bonds, it is necessary to assign the order in attaching child blocks. A bond with the smallest bond number is given priority for attaching a next block. Thus, the chemical structure of a molecule is determined uniquely from the sequence of block names, which is the same way to number bonds and units according to Definition 6.

In Fig. 3 an example of a unit library is shown for an aspartic acid residue, which consists of seven bonds (five inner, one entering, and one outgoing bonds). The block name ((ASP) in the first line is used to refer to the data. The numbers of atoms ((14), and the number of entering and inner bonds and the number of outgoing bonds (6 and 1, respectively) in the block are given in the second line.

For each unit and bond the following are given: the serial unit and bond number and the bond name, the serial numbers of four atoms that define the dihedral angle about the bond, the dihedral angle value which generates the set of atomic coordinates given below, energy parameters for torsional potential function for the bond, the serial number of its parent unit, the number of the child units (i.e., outgoing bonds) and their serial numbers, etc.

In each unit the smallest serial number of constituent atoms is given to the head atom of the entering bond. The third number in the four serial atomic number that define the dihedral angle corresponds to such a head atom. Owing to this rule we can know which atoms are included in a given unit. To take an example, bond 3 has a name, CHI1; the dihedral angle around this bond is defined by the atoms 1-3-5-9 (i.e., N-CA-CB-CG) and the torsional potential is 1.35 (1 + cos 3θ) kcal/mol; the parent bond is 2 and there is one child bond, which is 4. It can be known that unit 3 consists of four atoms, 5, 6, 7 and 8 (atom 8 is a pseudo atom), from the fact that the head atoms of bonds 3 and 4 are 5 and 9, respectively.

For each of the constituent atoms, the serial number and the atom name, the partial charge, the atomic coordinates generated with the above dihedral angle values, the serial number of its parent atom, the number of child atoms and their serial numbers, etc., should be included. The coordinates of the tail atom of the entering bond and the head atom(s) of the outgoing bond(s) are also required, although they are not member atoms of this block. For example, the atom 5, named CB, has a partial charge of \(-1.29\) \((= -0.100 \times (332.0/2)^{1/2})\); an actual partial charge is \(-0.100\) e.c.u.; 332.0 and 2 are the conversion factor to give the energy in units of kcal/mol and the effective dielectric constant, respectively [3] and is classified as type 6 in the calculation of the van der Waals energy; the parent atom is 3, and there are three child atoms, 6, 7, and 9.

The MOL data are generated by the program MKMOL using the library discussed above for the
system illustrated in Fig. 1. Parts of it are shown in Figs. 4a, b. The contents of the MOL data are very similar to those in the unit library, but it should be noticed that the three lists, DEPTH, MDESCN and BRANCH, defined in Section 2 are included. Information about the classification of atom pairs into 1–4 and 1–5 interactions and loop closing potential (for disulfide bond in this example) is also necessary.

Fig. 3. Example of a unit library. Data for aspartic acid are shown. a) block name, b) the number of atoms, c) the numbers of entering and inner bonds, and of outgoing bonds, d) serial numbers and names of bonds (or units), e) serial numbers of four atoms to define the dihedral angles about a bond, f) dihedral angle value, g) parameters for a torsional potential function, h) a serial unit number of a parent, i) the number of child units and their serial unit numbers, j) serial numbers and names of atoms, k) atom type, l) partial charge, m) x, y and z coordinates, n) a serial atom number of a parent, o) the number of child atoms and serial atom numbers of them.
Fig. 4. Example of the MOL data corresponding to the system shown in Fig. 1. (a) The beginning of the MOL data; a) number of bonds (variables), b) number of atoms, c) serial bond (or unit) number and bond name, d) serial atom numbers of both ends of a bond, e) serial bond number of a parent, f) DEPTH, g) MDESCN, h) BRANCH, i) parameters for a torsional potential function, j) dihedral angle value, k) number of child bonds, l) serial bond numbers of children. (b) The middle part of the MOL data; m) serial atom number and atom name, n) serial unit number to which the atom belongs, o) atom type, p) serial atom number for a real atom, q) partial charge, r) x, y, and z coordinates, s) number of child atoms, t) serial atom number of children (a minus sign indicates the relevant bond is rotatable), u) information for disulfide bonds, v) information for 1-4 and 1-5 interactions (not shown here).
<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HN1 CYS</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2.27</td>
<td>-0.37</td>
<td>-0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>HN2 CYS</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2.27</td>
<td>-0.37</td>
<td>-0.47</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>N   CYS</td>
<td>0</td>
<td>14</td>
<td>3</td>
<td>-4.59</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>CA  CYS</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>0.82</td>
<td>1.45</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>HA  CYS</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.26</td>
<td>1.76</td>
<td>-0.51</td>
<td>0.91</td>
</tr>
<tr>
<td>6</td>
<td>CB  CYS</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>-1.16</td>
<td>2.01</td>
<td>-0.78</td>
<td>-1.18</td>
</tr>
<tr>
<td>7</td>
<td>HB2 CYS</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>0.71</td>
<td>1.64</td>
<td>-1.80</td>
<td>-1.13</td>
</tr>
<tr>
<td>8</td>
<td>HB3 CYS</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0.71</td>
<td>1.64</td>
<td>-0.35</td>
<td>-2.10</td>
</tr>
<tr>
<td>9</td>
<td>QB  CYS</td>
<td>2</td>
<td>23</td>
<td>0</td>
<td>0.00</td>
<td>1.64</td>
<td>-1.08</td>
<td>-1.62</td>
</tr>
<tr>
<td>10</td>
<td>SG  CYS</td>
<td>3</td>
<td>21</td>
<td>9</td>
<td>0.13</td>
<td>3.84</td>
<td>-0.84</td>
<td>-1.26</td>
</tr>
<tr>
<td>11</td>
<td>SD  CYS</td>
<td>4</td>
<td>23</td>
<td>0</td>
<td>0.00</td>
<td>4.23</td>
<td>-1.95</td>
<td>-2.93</td>
</tr>
<tr>
<td>12</td>
<td>CE  CYS</td>
<td>4</td>
<td>23</td>
<td>0</td>
<td>0.00</td>
<td>6.05</td>
<td>-2.00</td>
<td>-3.01</td>
</tr>
<tr>
<td>13</td>
<td>C   CYS</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>5.80</td>
<td>1.93</td>
<td>1.45</td>
<td>0.00</td>
</tr>
<tr>
<td>14</td>
<td>O   CYS</td>
<td>5</td>
<td>17</td>
<td>11</td>
<td>-4.95</td>
<td>1.12</td>
<td>2.37</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>N   LEU</td>
<td>6</td>
<td>14</td>
<td>12</td>
<td>-4.59</td>
<td>3.24</td>
<td>1.60</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>HN  LEU</td>
<td>6</td>
<td>2</td>
<td>13</td>
<td>2.27</td>
<td>3.90</td>
<td>0.84</td>
<td>0.00</td>
</tr>
</tbody>
</table>

......

| 165| C   CYS | 45 | 7 | 138| 5.80 | 13.42 | 12.28 | -20.25 | 3 | -156 | 166 | -167 |
| 166| O   CYS | 45 | 17| 139| -4.95 | 14.37 | 13.05 | -20.34 | 1 | 165 |
| 167| OXT CYS | 46 | 18| 140| -4.90 | 12.49 | 12.19 | -21.14 | 2 | -165 | 168 |
| 168| HXT CYS | 46 | 4 | 141| 2.63 | 12.69 | 12.81 | -21.90 | 1 | 167 |
| 169| 47 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 170| 48 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 171| 49 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 172| 50 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 173| 51 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 174| N   ARG | 52 | 14| 144| -4.59 | 0.00 | 0.00 | 0.00 | 3 | 174 | 175 | -177 |
| 175| CA  ARG | 53 | 9 | 145| 0.82 | 1.45 | 0.00 | 0.00 | 4 | -176 | 178 | -179 | -200 |
| 176| HA  ARG | 53 | 1 | 146| 0.26 | 1.73 | -0.52 | 0.91 | 1 | 177 |

......

| 585| C   CYS | 176 | 7 | 493| 5.80 | 67.69 | 39.52 | 0.00 | 3 | -576 | 586 | -587 |
| 586| O   CYS | 176 | 17| 494| -4.95 | 68.09 | 38.36 | 0.00 | 1 | 585 |
| 587| OXT CYS | 177 | 18| 495| -4.90 | 68.41 | 40.59 | 0.00 | 2 | -585 | 588 |
| 588| HXT CYS | 177 | 4 | 496| 2.63 | 69.38 | 40.36 | 0.00 | 1 | 587 |
| 589| 179 | 23 | 0 | 0.00 | 0.00 | 17.00 | 0.00 | 0 |
| 590| 180 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 591| 181 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 592| 182 | 23 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| 593| 183 | 14| 149| -4.59 | 0.00 | 0.00 | 0.00 | 3 | 594 | 595 | -597 |
| 594| N   ARG | 184 | 9 | 500| 0.82 | 1.45 | 0.00 | 0.00 | 4 | -596 | 598 | -599 | -620 |
| 595| CA  ARG | 184 | 9 | 145| 0.82 | 1.45 | 0.00 | 0.00 | 4 | -596 | 598 | -599 | -620 |
| 596| HA  ARG | 184 | 1 | 501| 0.26 | 1.73 | -0.52 | 0.91 | 1 | 597 |

......

| 843| C   CYS | 266 | 7 | 713| 5.80 | 41.83 | 26.30 | 0.00 | 3 | -834 | 844 | -845 |
| 844| O   CYS | 266 | 17| 715| -4.90 | 42.31 | 25.17 | 0.00 | 1 | 843 |
| 845| OXT CYS | 267 | 18| 715| -4.90 | 42.46 | 27.42 | 0.00 | 2 | -843 | 846 |
| 846| HXT CYS | 267 | 4 | 716| 2.63 | 43.45 | 27.26 | 0.00 | 1 | 845 |
| 10| 841 | 100.0000 | 1 | 0 | 0.0000 |
| 11| 840 | 100.0000 | 1 | 0 | 0.0000 |
| 6 | 842 | 100.0000 | 1 | 0 | 0.0000 |
| 12| 836 | 100.0000 | 1 | 0 | 0.0000 |
| 162| 583 | 100.0000 | 1 | 0 | 0.0000 |
| 163| 582 | 100.0000 | 1 | 0 | 0.0000 |
| 158| 584 | 100.0000 | 1 | 0 | 0.0000 |
| 164| 578 | 100.0000 | 1 | 0 | 0.0000 |

......

Fig. 4. continued
in the MOL data. (The information about classification of atom pairs into 1–4 and 1–5 interactions are not shown in Fig. 4.)

6. Discussion

The program, FEDER/2, developed for static and dynamic conformational energy analysis of macromolecules has been described in this paper. In the program dihedral angles are treated as independent variables by keeping bond lengths and bond angles fixed. Although the program is more complicated, compared with other similar programs using atomic Cartesian coordinates as independent variables, the reduction in the number of variables is significant not only to efficiently simulate a large system consisting of macromolecules such as protein molecules, but also to make an analysis of such molecules more understandable. For example, in the determination of the solution conformation of a protein molecule from NMR distance analysis and modeling of a region with mutated amino acid residues, optimal conformations can be obtained more efficiently by reducing the conformational space to be searched. In dynamic conformational analyses such as normal mode analysis and Monte Carlo simulation, global or concerted motions of molecules are more intelligible with a smaller number of essential variables.

The tree representation of a system is fundamental to the program. On the basis of the definitions given in Section 2 a system consisting of one or more molecules can be described consistently. The system comprises units and bonds connecting them in a tree structure, and a set of data for molecules to be calculated can be generated from the unit library. By introducing the unit library the program becomes more general. Consequently, FEDER/2 can be applied to more extensive problems of protein conformations.

The study, in which several dihedral angles are fixed (it means that several units are unified into one larger unit), is an interesting approach in the analysis of the protein conformational dynamics. In modeling the mutated amino acid residues the protein may be divided into a large rigid part (a large unit comprised of many units) and small flexible parts [33]. It may be interesting to study a protein conformation by regarding one secondary structure segment (e.g., α-helix, β-strand and turn) as one unit.

In conclusion, the general tree representation of a system consisting of one or more molecules makes FEDER/2 a powerful and flexible software system in the dihedral angle space to study biopolymers.

Acknowledgements

We thank Dr. Steven Hayward for carefully reading the manuscript.

Appendix A

The program, FEDER/2, contains the following conformational energy functions and objective functions, which can be used in various combination specified by program users.

(a) Conformational energy function of molecules is a sum of several energy terms:

$$E_c = E_{lj} + E_{es} + E_{tor} + E_{lp} + E_{byd}.$$  \hspace{1cm} (A.1)

For the first four terms FEDER/2 uses the same function forms and parameter values as ECEPP [3,4]; an alternative function form is available for the forth term. The last term is newly introduced to take into account hydration energy.

The first term of Eq. (A1) represents the van der Waals energy for hydrogen bonding and nonhydrogen bonding atomic pairs:

$$E_{lj} = \begin{cases} 
\sum_{\alpha<\beta} \left( \frac{A_{\alpha\beta}}{r_{\alpha\beta}^{12}} - \frac{B_{\alpha\beta}}{r_{\alpha\beta}^{10}} \right) & \text{for hydrogen bonding atomic pairs} \\
\sum_{\alpha<\beta} \left( \frac{A_{\alpha\beta}}{r_{\alpha\beta}^{12}} - \frac{B_{\alpha\beta}}{r_{\alpha\beta}^{6}} \right) & \text{for non-hydrogen bonding atomic pairs}
\end{cases}$$  \hspace{1cm} (A.2)

where \( \eta \) is a factor whose magnitude is determined by consideration of the repulsive force constants for 1–4 interactions; \( \eta = 0.5 \) for 1–4 type interactions, and 1.0 for all other type interactions. The second
term sums the electrostatic energy over all non-bonded atomic pairs;

\[ E_{es} = \sum_{\alpha < \beta} \frac{Q_{\alpha} Q_{\beta}}{Dr_{\alpha \beta}}. \]  

(A.3)

In the third term the energy associated with the barriers to rotation about a rotatable bond, which is not completely accounted for by the above energy contributions, is taken into computation;

\[ E_{tor} = \sum_{\alpha} U_{\alpha}(1 \pm \cos n_{\alpha} \theta_{\alpha}). \]  

(A.4)

When a loop, in which at least one rotatable bond is included, occurs in a molecule, a special treatment is necessary, because a loop is unacceptable in the tree topology. In order to represent a molecule with a loop as a tree, one of the bonds in the loop must be cleaved. In compensation a proper potential is introduced to the conformational energy function to maintain the standard bond length and bond angles of the cleaved bond during computation. The two alternative treatments are considered in FEDER/2.

Assume that four atoms, \(a, b, c,\) and \(d\), define a dihedral angle around bond \(b-c\), which is cut off. In one treatment the bond length, \(b-c\), the bond angles, \(a-b-c\) and \(b-c-d\), and the dihedral angle, \(a-b-c-d\), are kept close to their standard values by introducing the pseudo-potential with respect to four distances between \(b\) and \(c\), \(a\) and \(c\), \(b\) and \(d\), and \(a\) and \(d\) [3,4].

\[ E_{lp} = \sum_{\alpha, \beta} w_{lp, \alpha \beta} \left( r_{\alpha \beta} - r_{0_{\alpha \beta}} \right)^2, \]  

(A.5)

where \(r_{\alpha \beta}\) and \(r_{0_{\alpha \beta}}\) are the distance between atoms, \(\alpha\) and \(\beta\), in the calculated conformation and the required (standard) distance, respectively, and \(w_{lp, \alpha \beta}\) is a constant chosen properly for each of the four atomic pairs.

In another treatment two independent chains, \(a--b-c--d^*\) and \(a^*-b^*-c--d\) are introduced with pseudo-atoms, \(a^*, b^*, c^*,\) and \(d^*\), which correspond to the real atoms, \(a, b, c,\) and \(d\), respectively. The bond lengths, \(b-c^*, c^*-d^*, a^*-b^*\) and \(b^*-c,\) and the bond angles, \(a--b-c^*, b-c^*-d^*, a^*-b^*-c,\) and \(b^*-c--d,\) are fixed to their standard values. Then the loop can be closed by the following potential functions;

\[ E_{lp} = \sum_{\alpha} w_{lp, \alpha} \left| r_{\alpha} - r_{\alpha^*} \right|^2, \]  

(A.6)

where \(r_{\alpha}\) and \(r_{\alpha^*}\) are position vectors of real and pseudo atoms, respectively, and \(w_{lp, \alpha}\) is a constant chosen properly. When we use this potential instead of Eq. (A5), the torsional potential about the \(b-c\) bond can be added on to \(E_{lp}\).

The last term in Eq. (A.1) is hydration energy according to the hydration-shell model;

\[ E_{hyd} = \sum_{\alpha} g_{\alpha} A_{\alpha} \]  

(A.7)

where \(A_{\alpha}\) is accessible surface area (ASA) of atom \(\alpha\) in a calculated conformation, and \(g_{\alpha}\) is a hydration free energy per unit ASA, which is given for seven chemical groups by Ooi et al. [30]. Since the derivatives of hydration energy function (i.e., derivatives of ASA) with respect to dihedral angles are not included in FEDER/2, this term can be taken into computation only in the Monte Carlo simulation [31].

(b) An objective function is used to search for optimal conformations satisfying given conditions. In this program it has to be a function of dihedral angles or distances between atoms. In FEDER/2 the objective function to determine solution conformations by using data from NMR distance analysis is incorporated according to Braun and Gö [28,29];

\[ E_{ob} = E_u + E_t + E_g, \]  

(A.8)

where

\[ E_u = \sum_{\alpha, \beta} w_{u, \alpha \beta} \left( r_{\alpha \beta}^2 - u_{\alpha \beta}^2 \right)^2, \]  

(A.9)

\[ E_t = \sum_{\alpha, \beta} w_{l, \alpha \beta} \left( r_{\alpha \beta}^2 - l_{\alpha \beta}^2 \right)^2, \]  

(A.10)

\[ E_g = \sum_{\alpha} w_{g, \alpha} \left[ 1 - \left( \frac{\theta_{\alpha} - \theta_{m}^e}{\theta_{w}} \right)^2 \right]^2, \]  

(A.11)

\(u_{\alpha \beta}\) and \(l_{\alpha \beta}\) are given upper and lower limits of the distance \(r_{\alpha \beta}\) between atoms, \(\alpha\) and \(\beta\), respectively. The limits of the dihedral angle, \(\theta_{\alpha}\), are given by \(\theta_{m}^m - \theta_{w}^m\) and \(\theta_{m}^m + \theta_{w}^m\). \(w_{u, \alpha \beta}, w_{l, \alpha \beta}\) and \(w_{g, \alpha}\) are weighting factors, which have properly assigned positive values, if \(r_{\alpha \beta} > u_{\alpha \beta}, r_{\alpha \beta} < l_{\alpha \beta}\) and \(\theta_{m}^m - \theta_{w}^m < \theta_{\alpha} < \theta_{m}^m + \theta_{w}^m\), respectively; otherwise, they vanish.
(c) A regularization of X-ray data must be performed before X-ray data of a protein molecule are used in FEDER/2, because FEDER/2 requires bond lengths and bond angles to adopt standard values. This procedure is a kind of optimization of an objective function. The objective function is defined as

\[ E_x = \sum_{\alpha < \beta} w_{x,\alpha\beta} \left( \frac{r_{x\beta}}{r_{\alpha\beta}} - 1 \right)^2, \]  

(A.12)

where \( r_{\alpha\beta} \) and \( r_{x\beta} \) are distances between atoms, \( \alpha \) and \( \beta \), in the calculated and X-ray conformations, respectively, and \( w_{x,\alpha\beta} \) is a weighting factor properly assigned.

(d) A combination of conformational energy function (A.1) and objective function (A.8) or (A.12) is a useful strategy to obtain reasonable conformations;

\[ E = pE_c + (1 - p)eE_{ob} \quad (0 \leq p \leq 1). \]  

(A.13)

where \( e \) is a factor to convert the dimension of the objective function to that of energy, and \( p \) and \( 1 - p \) are factors to determine the weights on the potential function and the objective function, respectively. Usually \( p \) is changed stepwise from 0 to 1.

(e) The variable target function method [28] is available in FEDER/2. As shown in Section 3 and Appendix C, the calculation of the first and second derivatives are performed recursively with respect to distances between bonds. Accordingly, it is easy to confine the computation to a set of selected pairs of atoms, depending on their distance along the chain (a distance between atoms along the chain is defined by the distance between the units they belong to; see Section 2). For example, by setting \( I_{a,b} = 0 \) for the bond pairs of \( k(a, b) > k_c \) in Eq. (11), it is possible to exclude the interactions of atoms more distant along the chain from each other than a given cutoff distance \( k_c \). The minimization performed by extending the cutoff distance gradually is a useful strategy to avoid being trapped in a local minimum of the function [7,28].

Appendix B

In this Appendix explicit expressions are summarized for the equations given in Section 3. We only consider the terms in which the conformational energy and objective functions of dihedral angles are given implicitly by functions of the interatomic distances;

\[ E(\theta_1, \theta_2, \ldots, \theta_a) = \sum_{\alpha < \beta} \Phi_{\alpha\beta}(r_{\alpha\beta}). \]  

(B.1)

(a) The basic quantity we are concerned with is the distance \( r_{\alpha\beta} = |r_a - r_\beta| \) between two atoms \( \alpha \) and \( \beta \), where \( r_a \) and \( r_\beta \) are position vectors of the atoms, respectively. The derivatives of \( r_{\alpha\beta} \) with respect to dihedral angle \( \theta_a \) is given as:

\[
\frac{\partial r_{\alpha\beta}}{\partial \theta_a} = \begin{cases} 
0 & \text{if } \alpha \in M_a \text{ and } \beta \in M_a \\
\gamma \cdot (e_a \times (r_\beta - r_a)) & \text{if } \alpha \in M_a \text{ and } \beta \in \overline{M_a}, \\
-\gamma \cdot (e_a \times (r_\beta - r_a)) & \text{if } \alpha \in \overline{M_a} \text{ and } \beta \in M_a \text{,} \\
0 & \text{if } \alpha \in \overline{M_a} \text{ and } \beta \in \overline{M_a}.
\end{cases}
\]  

(B.2)

where \( \cdot \) and \( \times \) denote the scalar and the cross products, respectively, and \( \gamma \) is the unit vector pointing from \( r_\beta \) to \( r_a \).

\[ y_{\alpha\beta} = (r_a - r_\beta)/r_{\alpha\beta}. \]  

(B.3)

\( e_a \) is a unit vector lying on bond \( a \), whose direction is the same as the direction of the bond, and \( r_a \) is a position vector of a certain point properly chosen on the bond \( a \).

(b) The six-dimensional column vector \( f_{\alpha\beta} \) is given as

\[ f_{\alpha\beta} = \begin{pmatrix} c_{\alpha\beta}(r_a \times r_\beta) \\ c_{\alpha\beta}(r_a - r_\beta) \end{pmatrix}, \]  

(B.4)

where

\[ c_{\alpha\beta} = \Phi'_{\alpha\beta}(r_{\alpha\beta}). \]  

(B.5)

(c) The \( 6 \times 6 \) matrix \( L_{\alpha\beta} \) can be written in a block matrix form,

\[ L_{\alpha\beta} = \begin{pmatrix} W_{\alpha\beta} & X_{\alpha\beta} \\ Y_{\alpha\beta} & Z_{\alpha\beta} \end{pmatrix}, \]  

(B.6)

where the \( 3 \times 3 \) matrices \( W_{\alpha\beta}, X_{\alpha\beta}, Y_{\alpha\beta}, \) and \( Z_{\alpha\beta} \) are given by:

\[ W_{\alpha\beta}^{\mu\nu} = \delta_{\mu\nu} \sum_\xi c_{\alpha\beta} r_\alpha^{\xi\mu} r_\beta^{\xi\nu} - c_{\alpha\beta} r_\alpha^{\nu\mu} r_\beta^{\nu\mu} \\
+ \delta_{\mu\nu} r_a^{\xi\mu}(r_a \times r_\beta)^\xi, \]  

(B.7)
\[ X_{\alpha\beta}^{\mu\nu} = \sum_\xi c_{\alpha\beta} \varepsilon_{\mu\nu\xi} r_\alpha^\xi + d_{\alpha\beta} (r_\alpha \times r_\beta)^\nu (r_\alpha - r_\beta)^\mu, \]  
\tag{B.8}

\[ Y_{\alpha\beta}^{\mu\nu} = - \sum_\xi c_{\alpha\beta} \varepsilon_{\mu\nu\xi} r_\beta^\xi + d_{\alpha\beta} (r_\alpha \times r_\beta)^\nu (r_\alpha - r_\beta)^\mu, \]  
\tag{B.9}

and

\[ Z_{\alpha\beta}^{\mu\nu} = \delta_{\mu\nu} c_{\alpha\beta} + d_{\alpha\beta} (r_\alpha - r_\beta)^\mu (r_\alpha - r_\beta)^\nu. \]  
\tag{B.10}

\( \delta_{\mu\nu} \) and \( \varepsilon_{\mu\nu\xi} \) are defined as

\[ \delta_{\mu\nu} = \begin{cases} 
1 & \text{if } \mu = \nu, \\
0 & \text{if } \mu \neq \nu,
\end{cases} \]  
\tag{B.11}

and

\[ \varepsilon_{\mu\nu\xi} = \begin{cases} 
1 & \text{if } (\mu, \nu, \xi) \text{ is even permutation of } (x, y, z), \\
-1 & \text{if } (\mu, \nu, \xi) \text{ is odd permutation of } (x, y, z), \\
0 & \text{otherwise.}
\end{cases} \]  
\tag{B.12}

\( d_{\alpha\beta} \) is given as

\[ d_{\alpha\beta} = \frac{c'_{\alpha\beta}}{r_{\alpha\beta}} = \frac{\Phi_{\alpha\beta}(r_{\alpha\beta})}{r_{\alpha\beta}^2} - \frac{\Phi'_{\alpha\beta}(r_{\alpha\beta})}{r_{\alpha\beta}^3}. \]  
\tag{B.13}

**Appendix C**

The computational procedures to calculate the first and second derivatives are described in this Appendix (the original discussion is given in Ref. [5]). For effective programming it is convenient to prepare the following lists:

(a) Lists of a parent bond number, the number of child bonds, and child bond numbers for every bond.

(b) A list of bond pairs which are arranged in descending order of distance \( k \) between bonds, a list of a starting position of the bond pairs of distance \( k \) in the bond pair list, and a list of a serial number of each bond pair in the bond pair list so as to retrieve it when a bond pair is given. The sign of \( S_{ab} \) for a bond pair, \( a \) and \( b \), in Eq. (9), can be memorized together in the bond pair list.

The lists (a) are necessary to obtain \( c(a) \) in Eq. (6), and \( a^*, b^* \) and \( a' \) and \( b' \) in Eq. (12). The lists (b) are used in the calculations of Eq. (12), because Eq. (12) has the following property about distances between bonds; i.e., if the distance between bonds \( a \) and \( b \) in the left-hand-side of Eq. (12) is \( k \), the distances between \( a' \) and \( b \) and between \( a \) and \( b' \) are \( k + 1 \) and the distance between \( a' \) and \( b' \) is \( k + 2 \) according to Eqs. (14) and (15). This means that we need to prepare storage space of \( R \)'s only for bond pairs with \( k, k + 1 \) and \( k + 2 \) at the same time.

Then, the computational procedure is as follows:

(i) Set \( k \) to the maximum value of \( k \).

(ii) For all pairs of bonds, \( a \) and \( b \), with a given value of \( k \), search the above list for \( (a^*, b^*) \) and \( (a', b') \) and then calculate \( f_{a,b} \) and \( I_{a,b} \) by Eqs. (8) and (11), respectively.

(iii) \( f_{a,b}^{*} \) is added to \( f_{a,b} \) (Eq. (7)).

(iv) \( R_{a,b} \) is calculated according to Eq. (12) and then store it. At the same time \( \partial^2 E / \partial \theta_a \partial \theta_b \) is calculated by Eq. (9).

(v) Set \( k = k - 1 \).

(vi) If \( k \geq 0 \), then go to (ii); otherwise go to (vii).

(vii) \( F_a \) is calculated according to Eq. (6).

(viii) Stop.

**Appendix D**

The equations which are necessary for a system consisting of two or more molecules (see Eq. (23) and Ref. [25]) are summarized in the following:

\[ \gamma_1 = (1, 0, 0) \]
\[ \gamma_2 = (0, 1, 0), \]  
\[ \gamma_3 = (0, 0, 1), \]  
\[ \xi = h_{\mu 1} \gamma_1 + h_{\mu 2} \gamma_2 + h_{\mu 3} \gamma_3, \]  
\tag{D.1}

\[ h_{\mu \nu} = \sigma^{-2} (1 - \sigma^{-1} \sin \sigma) \sigma_\mu \sigma_\nu \]
\[ + \sigma^{-2} (1 - \cos \sigma) \sum_\xi \varepsilon_{\mu\nu\xi} \sigma_\xi \]
\[ + \delta_{\mu\nu} \sigma^{-1} \sin \sigma, \]  
\tag{D.2}

\[ \frac{\partial \xi_{\mu}}{\partial \sigma_\xi} = \frac{\partial h_{\mu 1}}{\partial \sigma_\xi} \gamma_1 + \frac{\partial h_{\mu 2}}{\partial \sigma_\xi} \gamma_2 + \frac{\partial h_{\mu 3}}{\partial \sigma_\xi} \gamma_3, \]  
\tag{D.3}

\[ \frac{\partial h_{\mu \nu}}{\partial \sigma_\xi} = \frac{\partial h_{\mu 1}}{\partial \sigma_\xi} \frac{\partial \gamma_1}{\partial \sigma_\xi} + \frac{\partial h_{\mu 2}}{\partial \sigma_\xi} \frac{\partial \gamma_2}{\partial \sigma_\xi} + \frac{\partial h_{\mu 3}}{\partial \sigma_\xi} \frac{\partial \gamma_3}{\partial \sigma_\xi}. \]  
\tag{D.4}
\[ = -\sigma^{-4}(2 + \cos \sigma - 3\sigma^{-1} \sin \sigma)\sigma \sigma \sigma \sigma + \sigma^{-3}(\sin \sigma - 2\sigma^{-1}(1 - \cos \sigma)) + \sum_{k} e_{\mu \nu \kappa} \sigma \sigma \sigma \sigma + \sigma^{-2}(1 - \sigma^{-1} \sin \sigma)(\delta_{\mu \varepsilon} \sigma \sigma + \delta_{\nu \varepsilon} \sigma \sigma) + \sigma^{-2}(\cos \sigma - \sigma^{-1} \sin \sigma) \delta_{\mu \nu} \sigma \sigma + \sigma^{-2}(1 - \cos \sigma) e_{\mu \nu \varepsilon} \sigma \sigma \sigma \sigma. \]  

(D.5)

References