CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*

Bernard R. Brooks,[†] **Robert E. Bruccoleri,**[‡] **Barry D. Olafson,**[§] **David J. States,**[¶] **S. Swaminathan, and Martin Karplus**

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
Received 6 July 1982; accepted 14 October 1982

CHARMM (*C*hemistry at *HAR*vard *M*acromolecular *M*echanics) is a highly flexible computer program which uses empirical energy functions to model macromolecular systems. The program can read or model build structures, energy minimize them by first- or second-derivative techniques, perform a normal mode or molecular dynamics simulation, and analyze the structural, equilibrium, and dynamic properties determined in these calculations. The operations that CHARMM can perform are described, and some implementation details are given. A set of parameters for the empirical energy function and a sample run are included.

I. INTRODUCTION

In recent years, the usefulness of empirical energy functions for investigating the physical properties of a wide variety of molecules has been demonstrated.[1-8] Studies range from those concerned with the structures, energies, and vibrational frequencies of small molecules,[9-14] through those dealing with Monte Carlo and molecular dynamics simulations of pure liquids and solutions,[15-23] to the analysis of the conformational energies and fluctuations of large molecules, such as proteins and nucleic acids in vacuum, in solution, and in a crystal environment.[24-33] It is desirable to have available a general and flexible computer program which can efficiently handle all aspects of such computations with empirical energy functions for the various systems of interest. In this article we describe the design of such a program (CHARMM) and give details concerning its implementation. It is possible with the present version of the program to model build the struc-

ture of interest, energy minimize that structure by first- and second-derivative techniques, perform a normal mode or molecular dynamics simulation, and analyze the structural, equilibrium, and dynamic properties determined in these calculations. Currently, CHARMM can treat isolated molecules, molecules in solutions, and molecules in crystalline solids. The information for proteins, nucleic acids, prosthetic groups (e.g., heme groups), and substrates are available, and the extension to other molecules is straightforward. A wide range of analysis is possible, including static structure and energy analysis, structure and energy comparisons, time series, correlation functions and statistical properties of molecular dynamic trajectories, and interfaces to computer graphics programs.

In developing the program, we have concentrated on three aspects of its design that lead to an efficient calculational system. First, the program should complete a specific task in as short a time as possible; second, we have attempted to simplify the input necessary to obtain any desired results; and third, we have tried to make it easy to understand and modify the program. It is these aspects of the program which have concerned us most and which have led to some original aspects in its implementation.

The organization of the program is presented in Figure 1. Table I gives a brief description of each of the boxes in the figure. The support routines indicated in the figure provide the programming

Brooks et al.

CHARMM Command Parser

I/O	Data Structures	Coordinate Manipulation	Mechanics & Energetics	Analysis Facility
Parameters	Topology File	Internal Coordinates	Minimization	Tables Statistics
Coordinates	Structure File	Build	Dynamics	Static Properties Correlations
Data Structures	Hbond List	Orient	Normal Modes	Dynamic Properties Spectral Analysis
Results	Nonbond List	Transformation		Comparison Graphics Interfaces
	Symmetric Images			Homologies

Support Routines

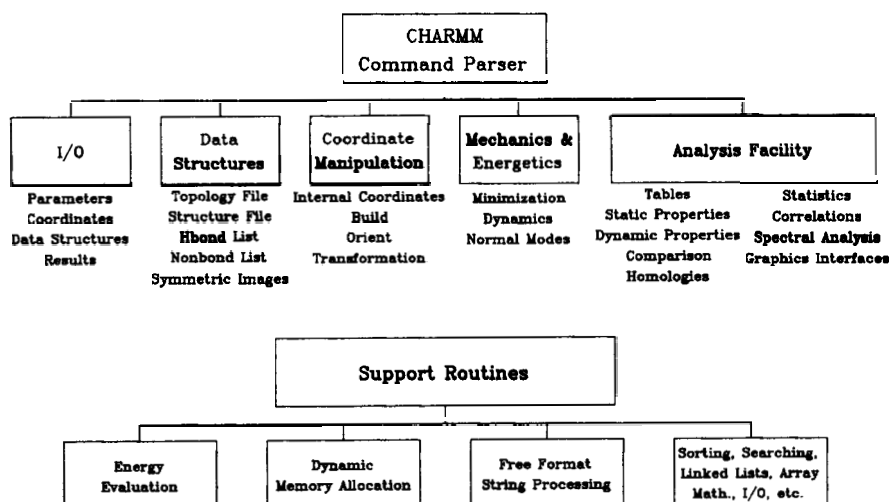| Energy Evaluation	Dynamic Memory Allocation	Free Format String Processing	Sorting, Searching, Linked Lists, Array Math., I/O, etc.

Figure 1. Organization of CHARMM. The top half of the figure shows the major groupings of commands under CHARMM. The bottom half outlines the support subroutines which provide the programming environment for the implementation of CHARMM. Table I gives a verbal description of the boxes in the figure.

Table I. An overview of CHARMM (see Fig. 1).

CHARMM Program Sections

Input/Output	Parameters	Force constants and equilibrium geometries for computing the energy
	Coordinates	The program reads and write coordinates in binary or arbitrarily ordered card formats
	Data Structures	I/O facilities are available for all of the major data structures in CHARMM
	Results	Printing of results during the course of the computation
Data Structures	Topology File	The topology file contains structural and bonding information about each monomer unit which is used for constructing the system.
	Structure File	The Structure File (PSF) is the main repository of structural information on the system. It is built using entries from the topology file.
	Hbond List	A list of favorable hydrogen bond pairs
	Nonbond List	A list of energetically significant nonbonded interaction pairs
	Symmetric Images	Symmetry related images such as dimers or crystal contacts are specially treated to conserve and exploit their symmetry
Coordinate Manipulation	Internal Coordinates	Internal coordinate representations are available for analysis and comparison of structures
	Build	Facilities to construct coordinates using internal coordinate information
	Orient	Routines to orient molecules with respect to their principal axis system or with respect to other molecules
	Transformation	General purpose routines to translate, rotate, and transform from nonorthogonal coordinate systems
Mechanics & Energetics	Minimization	Energy minimization to relax and cool structures
	Dynamics	The integration of Newton's equations of motion with Gear or Verlet algorithms and the option of internal coordinate constraints (SHAKE)

	Normal Modes	Diagonalization of the second derivative matrix and analysis of the normal modes
Analysis Facility	Tables	Tables are a general purpose data structure used to store and manipulate analysis information
	Static Properties	Many properties of the system at a fixed point in time may be entered into a table.
	Dynamic Properties	A table can store averages, fluctuations, and other properties computed from a molecular dynamics simulation.
	Comparison	Structural and energetic comparison of systems
	Homology	Homologies in the residue sequence of two related systems can be computed for use in a comparison.
	Statistics	Applicable to all table data
	Correlations	Generation of auto and cross correlation functions for the analysis of trajectories
	Spectral Analysis	Frequency analysis of dynamical properties
	Graphics Interfaces	Interface routines and programs to display results on an Evans & Sutherland picture system and generate hardcopy plots.

Support Routines

Energy Evaluation	The CHARMM energy function is available as a subroutine called in many different parts of the program for tasks including total energy evaluation, first and second derivative evaluation, and term by term analysis of the potential energy.
Dynamic Memory Allocation	Dynamic memory allocation routines written and called in FORTRAN are used in many parts of the program to maximize the flexibility of the program and minimize demands on the host system.
Free Format String Processing	A free format string manipulation library used for input and command parsing and for adaptable formats for outputs.
Sorting, Searching, Linked Lists, Array Math., I/O, etc.	A variety of subroutines exists for sorting and searching arrays, manipulating linked lists, performing arithmetic on arrays and vectors, performing common input/output support, and other useful miscellaneous functions.

environment for the implementation of the commands in CHARMM.

In this article, we describe the energy expression in Sec. II and the generation of data structures required to compute the energy in Sec. III. This order is followed because the energy expression provides the rationale for the generation process. The generation of unknown coordinates and the manipulation of conformations is described in Sec. IV. In Sec. V, we outline the energy manipulations required for energy minimization, normal mode analysis, and dynamics. The analysis facility is discussed in Sec. VI. In Sec. VII, we treat specific programming aspects of CHARMM that made it possible to implement the desired design features of CHARMM. Section VIII presents the discussion and conclusions. Appendix 1 gives a set of parameters for the explicit hydrogen representation of molecular structure. Appendix 2 gives a sample CHARMM run on a small molecule which demonstrates many of the program features. A table of execution times for a few representative tasks is given in Appendix 3.

A number of programs with some similar features have been described.[34-44] By not comparing these programs with the one presented here, we do not mean to imply that our implementation is preferable, but rather to limit the discussion to the present program and to demonstrate how its design features have made it a useful research tool. The closest in conception and formulation to the present program is the program AMBER,[44] which is based, as is CHARMM, on the macromolecular program developed at Harvard by Bruce Gelin and Martin Karplus.[36,45]

II. THE EMPIRICAL ENERGY FUNCTION

In this section, we describe the empirical energy function that is used in CHARMM.

A. Atomic Representation

The fundamental unit used in CHARMM is the atom. An atom is considered to be a charged point mass with no directional properties and without internal degrees of freedom. Masses are expressed in atomic mass units. Each atom of a real system could correspond to a single atom in CHARMM, but for large systems some (or all) hydrogens are combined with neighboring heavy atoms to which they are bound. This combining of atoms is re-

ferred to as the "extended-atom representation." For the peptide group and amino acid side chains, it has been shown[36,46] that the extended-atom dihedral angle potentials can essentially reproduce those obtained from an all-atom representation. Some advantages and disadvantages of the extended-atom approximation are presented.

Advantages:

(1) The most important advantage of extended atoms is that they significantly reduce the size of most problems, since roughly half of the atoms in biological or other organic macromolecules are hydrogens. This results in fewer nonbonded interactions and internal degrees of freedom.
(2) Motions involving heavy atoms are separated from hydrogen stretching motions. The large gap in the IR spectra between these types of motions implies that removing one type should have a small effect on the other.[47]
(3) Larger dynamics integration step sizes can be used by not including hydrogens since the small mass of hydrogen atoms requires a smaller time step for accurate integration.
(4) Hydrogen positions are usually not available from x-ray crystallography and must be generated from the heavy atom coordinates.

Disadvantages:

(1) It is difficult to represent accurately hydrogen bonding, since the position of a hydrogen atom has a large effect on hydrogen bond strength. When only heavy atom positions are included, a crude approximation to the hydrogen bond must be introduced.
(2) Dipole and quadrupole moments are lost combining hydrogens into extended atoms; e.g., the loss of a dipole in a N—H bond can adversely affect its interaction with other nearby groups.
(3) There is a loss of steric effects to hydrogens, since an extended atom is always spherical.
(4) Hydrogen coordinates are necessary for some types of analysis (e.g., proton and ^{13}C-NMR phenomena).

The extended atom representation has been shown to provide a satisfactory representation of the internal vibrations and bulk properties of small molecules including simple peptides.[23,36,46,48,49] In the past, when computational considerations posed severe limitations, the extended-atom rep-

resentation was used for all groups involving hydrogens. More recently, with improvements in efficiency and increased computing power, only aliphatic hydrogens, which are not significantly charged and cannot participate in hydrogen bonds, continue to be represented as extended atoms. In this way, the most important disadvantages of the extended-atom representation are removed [(1) and (2) above]. For small molecules, all atoms can easily be included explicity. Consequently, all three options are supported by CHARMM (all hydrogens, explicit hydrogen-bonding hydrogens only, and extended atoms only). See Appendix 1 for a list of atom types.

B. Functional Forms

The empirical energy function is made up of a sum of many terms. Some of these have optional forms or parameters, and others are optional altogether. The CHARMM energy function is based on separable internal coordinate and pairwise nonbond interaction terms. The total energy is expressed in the form

$$E = E_b + E_\theta + E_\phi + E_\omega + E_{vdW}$$
$$+ E_{el} + E_{hb} + E_{c_r} + E_{c_\phi} \quad (1)$$

where the formula for each of these terms is presented below. For each type of term, the relevant definitions and discussion follow in the text.

(1) Internal energy terms

Bond potential:

$$E_b = \sum k_b (r - r_0)^2 \quad (2)$$

Bond angle potential:

$$E_\theta = \sum k_\theta (\theta - \theta_0)^2 \quad (3)$$

Dihedral angle (torsion) potential:

$$E_\phi = \sum |k_\phi| - k_\phi \cos(n\phi), \quad \text{where } n = 1,2,3,4,6 \quad (4)$$

Improper torsions:

$$E_\omega = \sum k_\omega (\omega - \omega_0)^2 \quad (5)$$

The first two terms account for bond and angle deformations, which in most cases at ordinary temperatures and in the absence of chemical reactions are sufficiently small for the harmonic approximation to apply. The torsion energy term is a four-atom term based on the dihedral angle about an axis defined by the middle pair of atoms. For this term, the energy constant may be negative
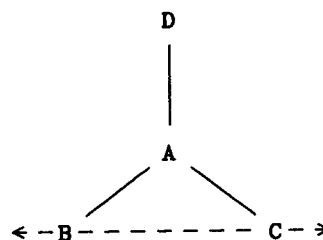

Figure 2. Definition of an improper torsion. The improper torsion angle ω, is defined as the torsion angle ϕ_{ABCD} or equivalently the angle between the planes ABC and BCD.

(indicating a maximum at the cis conformation), and there may be several contributions with different k_ϕ and different periodicities for a given set of four atoms. The improper torsion term has been designed both to maintain chirality about a tetrahedral extended heavy atom (e.g., an α carbon without an explicit hydrogen), and to maintain planarity about certain planar atoms (such as a carbonyl carbon) with a quadratic distortion potential; without such a term, out-of-plane potentials tend to be quartic. In addition, this term provides a better force field near the minimum-energy geometry, which is important for dynamics and vibrational analysis. Figure 2 depicts an improper torsion where the axis used to compute the dihedral angle is along B-C.

Force constants (k_x) and geometric constants $(r_0, \theta_0, n, \omega_0)$ are selected from the parameter table (which are read into CHARMM) based on the atom types of the atoms involved. For bonds and angles, the atoms' types must match. For the four-atom terms, the ability (but not the necessity) to use a "wild-card" specification (X) that can correspond to any atom type greatly reduces the potential size of the parameter lists. Appendix 1 gives a list of parameters used for proteins. Parameters for nucleic acids are available in the literature.[30]

Force constants have been obtained by fitting to vibrational data in some cases[4,50,51] and from the literature for others.[10,30,36] For the most part, geometric constants have been derived from crystallographic data.

First and second derivatives of the torsional potential with respect to Cartesian displacements are found by multiple use of the chain rule. The energy is never differentiated with respect to ϕ because of singularities when angles become planar (which is rather common). Instead, the functional form is differentiated with respect to $\cos(\phi)$. This has the added advantage that trigonometric function evaluations are not needed when evalu-

ating the energy. To facilitate this procedure, the cosine term of eq. (4) is replaced by

$$\cos(2\phi) = 2\cos^2(\phi) - 1 \tag{6}$$

$$\cos(3\phi) = 4\cos^3(\phi) - 3\cos(\phi) \tag{7}$$

$$\cos(4\phi) = 8\cos^4(\phi) - 8\cos^2(\phi) + 1 \tag{8}$$

$$\cos(6\phi) = 32\cos^6(\phi) - 48\cos^4(\phi) + 18\cos^2(\phi) - 1 \tag{9}$$

For the same reason, the derivatives of planar improper torsions are expanded in a Taylor series when ω is small. The actual form that is used is

$$E = k\omega^2 \tag{10}$$

$$\frac{\partial E}{\partial \cos(\omega)} = -2k\left(1 + \frac{\omega^2}{6}\right) \tag{11}$$

$$\frac{\partial^2 E}{[\partial \cos(\omega)]^2} = \frac{2k}{3}\left(1 + \frac{\omega^2}{6}\right) \tag{12}$$

When ω is large (6° or more), a straightforward approach is used.

(2) Nonbonded interactions

Van der Waals interactions:

$$E_{\text{vdW}} = \sum_{\text{excl}(i,j)=1} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}\right) \text{sw}(r_{ij}^2, r_{\text{on}}^2, r_{\text{off}}^2) \tag{13}*$$

Electrostatic potential (only one of the below):

(a) Constant dielectric:

$$E_{\text{el}} = \sum_{\text{excl}(i,j)=1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{14}$$

(b) Distance-dependent dielectric (linear):

$$E_{\text{el}} = \sum_{\text{excl}(i,j)=1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}^2} \text{sw}(r_{ij}^2, r_{\text{on}}^2, r_{\text{off}}^2) \tag{15}$$

(c) Shifted dielectric:

$$E_{\text{el}} = \sum_{\text{excl}(i,j)=1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\left(1 - \frac{2r_{ij}^2}{r_{\text{cut}}^2} + \frac{r_{ij}^4}{r_{\text{cut}}^4}\right) \tag{16}$$

(d) Electrostatics by groups: E_{el} from Eq.

---
* sw is a switching function defined by

$\text{sw}(x, x_{\text{on}}, x_{\text{off}}) = 1$ when $x \leqslant x_{\text{on}}$

$\text{sw}(x, x_{\text{on}}, x_{\text{off}}) = \dfrac{(x_{\text{off}} - x)^2 (x_{\text{off}} + 2x - 3x_{\text{on}})}{(x_{\text{off}} - x_{\text{on}})^3}$

            when $x_{\text{on}} < x \leqslant x_{\text{off}}$

$\text{sw}(x, x_{\text{on}}, x_{\text{off}}) = 0$ when $x > x_{\text{off}}$

$\text{excl}(i,j)$ = 0 if atoms are connected by angles or bonds or $i \geq j$

            = 1 otherwise (unless explicitly specified to be zero)

(15) above, except replace $\text{sw}(r_{ij}^2, r_{\text{on}}^2, r_{\text{off}}^2)$ by $\text{sw}(r_{IJ}^2, r_{\text{on}}^2, r_{\text{off}}^2)$, where r_{IJ} is the distance between the center of geometry of the groups containing atoms i and j.

(e) Extended electrostatics:

$$E_{\text{el}} = E_{\text{el}_{\text{dir}}} + E_{\text{el}_{\text{ext}}} \tag{17}$$

$$E_{\text{el}_{\text{dir}}} = \sum_{\substack{r_{ij} < \text{cutnb} \\ \text{excl}(i,j)=1}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \text{ (direct term)} \tag{18}$$

$$E_{\text{el}_{\text{ext}}} = \sum_{\text{atoms}} q_i \left[\frac{1}{2}\text{pot}_i + (\mathbf{r}_i - \mathbf{r}_{i_0}) \cdot \mathbf{F}_i\right] \tag{19}$$

(extended term)

where pot_i and $\mathbf{F}_i$ are the approximate potential and field at atom i due to atoms that are not present in the nonbond pair list [eq. (18)], and $\mathbf{r}_{i_0}$ are the coordinates at the last update of the nonbonded pair list.

$$\text{pot}_i = \text{pot}_{i_{\text{atoms}}} + \text{pot}_{i_{\text{groups}}} \tag{20}$$

$$\text{pot}_{i_{\text{atoms}}} = \sum_{\substack{r_{ij} \geq \text{cutnb} \\ C(I,J)=1}} \frac{q_j}{4\pi\epsilon_0 r_{ij}} \tag{21}$$

$$\text{pot}_{i_{\text{groups}}} = \sum_{C(I,J)=0} \frac{q_J}{4\pi\epsilon_0 r_{iJ}} + \frac{\boldsymbol{\mu}_J \cdot \mathbf{r}_{iJ}}{4\pi\epsilon_0 r_{iJ}^3} + \frac{\mathbf{r}_{iJ} \cdot \mathbf{Q}_J \cdot \mathbf{r}_{iJ}}{4\pi\epsilon_0 r_{iJ}^5} \tag{22}$$

$$\mathbf{F}_i = \mathbf{F}_{i_{\text{atoms}}} + \mathbf{F}_{i_{\text{groups}}} \tag{23}$$

$$\mathbf{F}_{i_{\text{atoms}}} = \sum_{\substack{r_{ij} \geq \text{cutnb} \\ C(I,J)=1}} \frac{-q_j \mathbf{r}_{ij}}{4\pi\epsilon_0 r_{ij}^3} \tag{24}$$

$$\mathbf{F}_{i_{\text{groups}}} = \sum_{C(I,J)=0} \frac{-q_J \mathbf{r}_{iJ}}{4\pi\epsilon_0 r_{iJ}^3} + \frac{\boldsymbol{\mu}_J}{4\pi\epsilon_0 r_{iJ}^3}$$
$$+ \frac{-3\mathbf{r}_{iJ}(\boldsymbol{\mu}_J \cdot \mathbf{r}_{iJ})}{4\pi\epsilon_0 r_{iJ}^5} + \frac{\mathbf{Q}_J \cdot \mathbf{r}_{iJ}}{4\pi\epsilon_0 r_{iJ}^5} + \frac{\mathbf{r}_{iJ} \cdot \mathbf{Q}_J}{4\pi\epsilon_0 r_{iJ}^5}$$
$$+ \frac{-5\mathbf{r}_{iJ}(\mathbf{r}_{iJ} \cdot \mathbf{Q}_J \cdot \mathbf{r}_{iJ})}{4\pi\epsilon_0 r_{iJ}^7} \tag{25}$$

Here $\boldsymbol{\mu}_J$ is the electric dipole moment of group J, and $\mathbf{Q}_J$ is the electric quadrupole moment. $C(I,J) = 1$ indicates that the groups containing atoms i and j are in close contact; $C(I,J) = 0$ if they are not or if $I \geq J$. Groups are defined as being in contact if the closest approach between the minimum-sized rectangular boxes containing each of the groups is less than the nonbond cutoff distance (cutnb). The sums over atoms in the extended terms include the atoms within groups which are in contact but are outside of the nonbond cutoff

distance. The sums over groups $[C(I,J) = 0]$ are sums over the remaining groups which are not in contact.

These terms represent the nonbonded interactions used in CHARMM. For a typical protein, the nonbonded interactions represent the bulk of the energy evaluation time. For large systems, it becomes necessary to avoid computing explicitly all pair interactions [an $O(n^2)$ task]. This is accomplished for the van der Waals term by switching off the potential at some distance (usually 7.5 Å) and not computing any interaction beyond this distance. When selecting the nonbonded list, the cutoff for this selection is usually 0.5 Å larger to allow for some displacement of atoms without introducing impulsive forces when the nonbonded list is subsequently updated.

Two distinctly different approaches are included for treating the electrostatic interactions (see above). One option introduces an approximate solvent screening term in the dielectric constant by setting the dielectric equal to r.[24,36,52] The electrostatic potential is then an even power of r and can be evaluated rapidly. Because of the need to limit the number of pair interactions and to avoid discontinuities in the forces (to conserve energy during dynamics), several schemes are used. A switching function is available, but the energy can still be significant at the cutoff distance which may result in artificially large forces at long range. This problem can be overcome in a number of ways. A simple approach referred to as the "shifted" potential, modifies the radial function so that the energy and forces go to zero at some cutoff distance. One drawback to the function presented in eq. (16) is that it has a discontinuity in the second derivative at the cutoff distance. At short distances, the forces are almost the same as found with the constant dielectric, but the overall magnitude of the energy of interaction is reduced by a constant. Another choice is to switch off the potential energy based on groups, rather than individual atoms. If two groups are neutral with nonzero dipoles, their long range interaction is proportional to $1/\epsilon(r)r^3$, which is much smaller than the individual atom interactions $[1/\epsilon(r)r]$.

The second approach (extended electrostatics) avoids the need for cutoffs through the use of the multipole approximation to permit the efficient evaluation of all of the electrostatic interactions in the system. This approach is particularly designed for simulations where the solvent is being included explicitly and the details of solvation and solvent polarization are of interest; therefore, a

constant dielectric is used. This is done by treating short-range interactions in the usual way, but expressing long-range group–group interaction in terms of multipoles; the multipole expansion is truncated after quadrupole moments. Since the separations involved are large compared with the size of the groups, the overall electric field at each atom is computed and updated only periodically.

(3) Hydrogen bonding

Involving atoms AA, A, H, D (acceptor antecedent, acceptor, hydrogen, and donor heavy atom):

$$E_{hb} = \sum \left(\frac{A'}{r_{AD}^i} - \frac{B'}{r_{AD}^j} \right) \cos^m(\theta_{A-H-D})$$
$$\times \cos^n(\theta_{AA-A-H}) \times sw(r_{AD}^2, r_{h_{on}}^2, r_{h_{off}}^2)$$
$$\times sw[\cos^2(\theta_{A-H-D}), \cos^2(\theta h_{on}), \cos^2(\theta h_{off})] \quad (26)$$

where i and j are positive integers, $m = (0,2,4)$ and $n = (0,2)$. The van der Waals term between the hydrogen atom and the acceptor atom can be reduced.

The selection of hydrogen bonds is based on the A–D distance and the A–H–D angle. The total term is zeroed if θ_{A-H-D} is less than 90° or if n is greater than 0 and θ_{AA-A-H} is less than 90°. The exponents i and j are determined when the parameters are read, and the exponents n and m are determined in the topology file when donor and acceptor atoms are listed. The exponent n is determined by the type of acceptor, and the exponent m is determined by the donor type. When donors are treated as extended atoms, only the radial terms are used. In addition to the explicit hydrogen bond potential (eq. 26), the electrostatic interactions between the atoms involved contribute.

In earlier work, the i and j exponents were set to 12 and 10, respectively, and only the A–H–D angular term was included in the potential (i.e., $n = 0$). Preliminary *ab initio* calculations of the formamide dimer indicate that the hydrogen-bond energy minimum is considerably broader than that of the 12–10 function and that the hydrogen bond energy depends on both the AA–A–H and A–H–D angles.[53] The hydrogen bond energy function is being reformulated in the light of these findings.

(4) Water–water interactions

To describe water–water interaction, CHARMM

uses either the ST2 potential,[15] or a combination of above terms (electrostatic, van der Waals, and hydrogen bonding). The use of the ST2 potential is preferable both in terms of computational efficiency and accuracy in describing water–water interations. At present, second derivatives are not available with the use of this option. Other water potentials[17,54] could be introduced without difficulty.

(5) Constraints

Atom harmonics:

$$E_{c_r} = \sum K_i (\mathbf{r}_i - \mathbf{r}_{i_0})^2,$$
$$K_i \text{ can be mass weighted} \quad (27)$$

Rigid:

$$\frac{\partial E}{\partial \mathbf{r}} \equiv 0 \text{ and } \delta\mathbf{r} = 0 \text{ for all operations} \quad (28)$$

Dihedral constraints:

$$E_{c_\phi} = \sum K_i (\phi_i - \phi_{i_0})^2 \quad (29)$$

Rigid distance constraints (SHAKE):

$$\delta_{\mathbf{r}_{ij}} = 0 \quad (30)$$

For some purposes, it is useful to restrict the changes that occur in a structure. For this reason, there is the option of including different types of constraints in the energy when manipulating the structure through minimization or dynamics. One choice is to maintain rigidly the position of certain atoms and to delete the energy terms involving only these atoms. This is an effective way to study a small part of a very large system with relative ease and increased efficiency. Atom harmonic constraints are used primarily to avoid large displacements of atoms when minimizing (often when bad contacts are present), while still allowing the structure to relax. Dihedral constraints are used to maintain certain local conformation or when a series of different conformations needs to be examined in making potential energy maps (see Appendix 2 for an example of this). The use of SHAKE[55] is restricted to dynamics and is discussed in Sec. V B.

(6) User energy

CHARMM allows for the addition of an arbitrary function calculated by a user-supplied routine which has access to all structural and positional information.

III. GENERATION

One method of generating an energy expression is to specify all of the bonds, angles, dihedral angles, etc., as input. For very small molecules, this is a task that does not require much effort, but for large systems, another method must be used. In most macromolecules, there is a large amount of redundancy as similar subgroups can appear many times, and there is a finite number of these subgroups. In CHARMM, these subgroups are referred to as residues, and the definitions of the energy terms associated with each type of residue constitutes the (residue) topology file (RTF). The actual storage of energy terms for a particular system is represented in the (protein) structure file (PSF). The generation process describes the method used to construct the PSF from a specified sequence of residues and the RTF. The terms residue and protein are employed in an extended sense, as indicated below, but remind the user of the system for which CHARMM was developed originally.

A. The Structure File

The most fundamental data structure used in CHARMM is the (protein) structure file (PSF), in that very little can be done until the PSF is generated or read from an existing file. With the addition of parameters and coordinates, the PSF contains all of the information necessary to generate nonbonded and hydrogen bond lists and evaluate the energy. A brief description of what is contained in the PSF is presented.

Residues and segments—how atoms are divided into residues and segments along with names and identifiers, etc.

Atom information—atom names, types, charges, masses, rigid constraints.

Bond, angle, dihedral, and improper dihedral lists—atoms involved in energy terms.

Donor and acceptor lists—lists used to construct hydrogen bonds.

Nonbonded exclusion list—atom pairs connected by close interactions (usually 1–2 and 1–3 interactions), so that no van der Waals or electrostatic term is desired.

Electrostatic groupings—how small numbers of atoms (six or fewer) are grouped for the purpose of computing long-range electrostatic or distance cutoff terms.

CHARMM builds a segment of the PSF each time

the generate command is invoked by stringing together the information of monomer units (referred to as residues) stored in a (residue) topology file (RTF). Some examples of typical segments are a single macromolecular chain, a collection of water molecules, a prosthetic group, and a collection of other small molecules.

Once the PSF, parameters, and coordinates are assembled, all the information is available for calculating the first four terms of the empirical energy function. However, the hydrogen bond and nonbonded lists must still be calculated from the PSF and coordinates.

B. The Topology File

The (residue) topology file (RTF) contains local information about atoms, bond, angles, etc., for each possible type of monomer unit (residue) that can be used in building a particular type of macromolecule.

The concept of residue as used in CHARMM is fundamental from an organizational point of view. It may be an amino acid residue as in a protein, but in the DNA topology files for example, a single residue corresponds to one nucleotide unit along a single strand of DNA (i.e., a phosphate group, ribose, and base).

The protein topology files contain residue information for each of the 20 amino acids (the titratable acids are present in multiple forms), terminal groups, heme, water, and assorted other small molecules such as ethanol. The DNA topology files contain one residue for each of the four different bases, terminal groups, and water. In addition to actual residues, topology files also contain information required to patch structures where special bonds and other features are needed. Examples of such patching operations are given below.

C. Sample Residue

To illustrate the information stored in the topology file, a sample residue from a protein RTF is presented followed by a brief description of its elements. The input below is read free field (spacing does not matter).

```
BOND N    CA          CA   C          C    +N         C    O
BOND N    H           CA   CB
THET -C   N    CA          N    CA   C          CA   C    +N
THET CA   C    O           O    C    +N         -C   N    H
THET H    N    CA          N    CA   CB         C    CA   CB
DIHE -C   N    CA   C          N    CA   C    +N
DIHE CA   C    +N   +CA
IMPH N    -C   CA   H          C    CA   +N   O
IMPH CA   N    C    CB
DONO H    N
ACCE O    C
IC  -C   CA   *N   H    0.0000   0.00   180.00   0.00   0.0000
IC  -C   N    CA   C    0.0000   0.00   180.00   0.00   0.0000
IC  N    CA   C    +N   0.0000   0.00   180.00   0.00   0.0000
IC  +N   CA   *C   O    0.0000   0.00   180.00   0.00   0.0000
IC  CA   C    +N   +CA  0.0000   0.00   180.00   0.00   0.0000
IC  N    C    *CA  CB   0.0000   0.00   120.00   0.00   0.0000
PATC FIRS NTER LAST CTER
```

The first line specifies that a new residue is started, its name is ALA (alanine), and its total charge is zero. Next, the atoms are defined. For each atom, a name, chemical type (see Appendix 1), and a charge is given. The interspersed keyword GROU specifies that a new electrostatic group follows. The groups are usually neutral, except for charged residues and special cases, and long-range interactions between groups can be determined by a multipole expansion, as described above. Atoms connected by bonds are then specified in pairs, angles by threes, and for dihedral angles (torsions) and improper torsions they are specified in groups of four. The use of a − or + before any atom indicates that this atom is to be found if possible in the previous or subsequent residues, respectively; if this atom cannot be found (which would be expected at chain termini), then the associated energy term is dropped and a warning issued. Following this is the declaration for donors and acceptors (this residue has one of each). The entries labeled IC (internal coordinates) contain all of the information required to build the coordinates of the residue from bond lengths, angles, and dihedrals (see Sec. IV A). A * preceding the third atom name indicates that this is an improper dihedral angle internal coordinate (see Sec. IV A). The zeroes in this table are filled with values that correspond to the minimum energy geometries found from the parameter file, except for the middle column (dihedral angle); any nonzero value is not replaced. The last line of this residue defines what patch to invoke if this residue happens to be the first or last residue in the segment. A different first or last patch (or none) may be specified in the generate command.

Once a segment of the PSF has been generated, the RTF is no longer needed until a subsequent generation is requested. Because of this, different segments of a single PSF can be built from different topology files. In this way, it is possible to include proteins and DNA in a PSF without the necessity of creating a single large RTF.

```
RESI ALA     0.00000
GROU
ATOM N    NH1    -0.35
ATOM H    H       0.25
ATOM CA   CH1E    0.10
GROU
ATOM CB   CH3E    0.00
GROU
ATOM C    C       0.45
ATOM O    O      -0.45
```

D. Patching of Structures

To treat the special features of specific macro-molecules (chain termination, disulfides, blocking groups, prosthetic groups, etc.), the program incorporates a "patching" mechanism. The RTF structure allows all such patching specifications to be stored as "patch residues," making them easily adaptable and their actions apparent.

There are two types of patch residues, those which are invoked to handle chain termination when generating a structure, and those which modify a structure after it has been built. The latter are invoked by the PATCH command from the CHARMM command parser. An example of each type is given here.

```
PRES CTER   -1.00000   ! PATCH FOR THE CARBOXY TERMINUS
GROU
ATOM C     C        0.14
ATOM OT1   OC      -0.57
ATOM OT2   OC      -0.57
ATOM O     O      999.00
BOND C     OT1          C     OT2
ANGL CA    C     OT1          CA    C     OT2          OT1   C     OT2
DIHE N     CA    C     OT2
IMPH C     CA    OT2   OT1
ACCE OT1   C
ACCE OT2   C
IC   N     CA    C     OT2   0.0   0.0   180.0   0.0   0.0
IC   OT2   CA    *C    O     0.0   0.0   180.0   0.0   0.0


PRES DISU 0.000   ! PATCH FOR DISULFIDES
ATOM 1CB   CH2E  0.19
ATOM 1SG   S    -0.19
ATOM 2SG   S    -0.19
ATOM 2CB   CH2E  0.19
BOND 1SG   2SG
ANGLE 1CB  1SG   2SG          1SG   2SG   2CB
DIHE 1CA   1CB   1SG   2SG          1CB   1SG   2SG   2CB          1SG   2SG   2CB   2CA
IC   1CA   1CB   1SG   2SG   0.0   0.0   180.0   0.0   0.0
IC   1CB   1SG   2SG   2CB   0.0   0.0   180.0   0.0   0.0
IC   1SG   2SG   2CB   2CA   0.0   0.0   180.0   0.0   0.0
```

The first PRES (patch residue) is the patch that is used for the carboxy terminus of a protein. The name and total charge are given its declaration command. Then follows the list of modified or added atoms. The atom named O has been deleted (this is indicated by its unacceptable charge). For the modified atoms, both the charge and chemical types may be changed. Following the atom specifications are the remaining terms required to define the patch which is appended to the last residue of the protein chain.

The second patch is invoked when a disulfide bond is added. Again, atom types and charges are modified, and new terms are added to the energy expression. For this kind of patch, all atom specifications must begin with a one-digit integer which is assigned to a particular residue in the PSF.

Because of the simplicity of storing patches in this manner, as opposed to hard wiring this information into the program, most patches are carried out with relative ease. Patches can be entered from the input stream to provide on-line editing of the PSF.

E. Generation of the Nonbonded List

To maximize the efficiency of the nonbonded calculation, a list is created which contains all pair interactions which are to be considered. Atom pairs are not included in the list if they are too far apart (beyond the long-range cutoffs), or if they are closely connected. In the latter case, the atom will be present in the excluded list [see eq. (13)].

The list generation approach was chosen over other alternatives (such as a pairwise search) for two reasons. During minimization or dynamics, the relative positions between atoms does not change radically between one step and the next. Also, a pairwise search through the list of atoms is relatively costly. Thus, the list is updated periodically during minimization or dynamics. If electrostatic groups are used, the list is stored in terms of group pairs. This leads to improved efficiency and a smaller list, as well as the ability to handle long-range group interactions.

A two-stage search is used to generate the nonbonded list. First, the minimum rectangular solid about each residue is computed so that a pairwise search through residues finds pairs that can potentially interact. Second, an atom search through these residue pairs is then used to construct the list.

When the electrostatic energy is evaluated without cutoffs, the long-range potential and field are evaluated and stored during nonbond updates (see Sec. II B). These terms are then used in computing long-range electrostatic interactions for every energy evaluation (until the next update).

F. Generation of the Hydrogen-Bond List

The list of hydrogen bonds is compiled using prespecified distance and angle cutoffs (see Sec. V B). During minimization and dynamics, the list can be kept fixed, or be updated periodically, or may be read from an external file. Another option selects only the "best" hydrogen bond for any donor, where best is based on an energy criterion.

In the case where the donor hydrogen is part of an extended atom representation (a method not recommended for future use), the hydrogen bond energy is a radial function only. Angular terms are, however, introduced in the selection of such hydrogen bonds using generated hydrogen positions as described in Sec. IV B. These hydrogen positions are temporary and are discarded after the hydrogen bond list has been created.

G. Symmetric Images

In CHARMM there is a generalized facility for reducing the problem size when symmetry is present and to be preserved.[56] The symmetry can be finite as in the case of a C_2 symmetric dimer such as hemoglobin, or the system can be infinite as is the case in crystal simulations or other systems with periodic boundary conditions. The translation rotation matrices are arbitrary which allows any crystal to be simulated. There is the obvious restriction that the rotation matrices be unitary (or antiunitary), and that each transformation must have an inverse.

At present, there are some restrictions on this feature in CHARMM. One limitation is that second derivatives are not available when this option is invoked. Another limitation is that only van der Waals, electrostatic, hydrogen bond, and ST2[15] water interactions are allowed between the primary set of atoms and image atoms (primary atoms being those actually represented in the PSF and image atoms, those which are obtained by applying a transformation).

IV. COORDINATE MANIPULATIONS

CHARMM has generalized facilities for the construction, manipulation, transformation, and interconversion of Cartesian, crystal, and internal coordinates as well as a specialized routine for the generation of hydrogen bonding proton coordinates. While crystallographic and simulation results are usually obtained and stored in Cartesian coordinates, it is frequently desirable to transform them either to reoriented Cartesian frames or to internal coordinates. All of the reorientations and transformations presented in this section can be applied to an arbitrary (and easily specified) set of atoms. Internal coordinates may be specified without regard to the bonding of atoms. This generality makes the routines useful in analysis as well as in setup and generation. This section presents the basic strategy used for these manipulations.

A. Coordinate Generation and Internal Coordinates (IC)

Coordinate generation is an important function. There are often situations where all coordinates are not available, or where some or all of a structure must be modified or built from internal

coordinate values. The internal coordinates (ICs) used to build a structure may be specified arbitrarily, taken from an existing structure, or chosen from the minimum energy values in the parameter tables. For example, specified ICs could be used to construct coordinates for a peptide backbone when only a phi–psi plot is available. The ability to take a Cartesian coordinate structure, fill and edit the internal coordinate (IC) table (i.e., specify what distances, angles, or dihedral values are to be modified), and then construct a new structure, is extremely useful. With this feature, entire sections of a protein can be moved relative to one another in an arbitrary but easily specified manner.

Given atoms of known position (A,B,C) and an unknown position (D), the values of the $C—D$ bond length, the B–C–D bond angle, and the A–B–C–D torsion angle are sufficient to place atom D. The IC table consists of entries specifying four atoms and five values (see Sec. III B for an example). There are two types of IC table entries corresponding to proper and improper torsions, the latter type occurring when atom C is central. This type is indicated by the presence of a * preceding this atom's name. For the proper torsion, the values correspond to R_{AB}, θ_{ABC}, ϕ_{ABCD}, θ_{BCD}, and R_{CD}, and for the improper type, the five values correspond to R_{AC}, θ_{ACB}, ϕ_{ABCD}, θ_{BCD}, and R_{CD}. From the IC table, the entire structure can be built when as few as three connected atom positions are known. The reason that fives values are used instead of the minimum of three is that with five, the structure can be built in any direction, regardless of which atoms in the structure have known positions.

The internal coordinate file is supported by several commands that allow the user to build or modify a coordinate set in terms of internal angles and distances. These commands include a complete editing facility (used to add or modify table entries), options that approximate unknown internal coordinate values based on the parameters, commands that build a Cartesian coordinate set from internal coordinates and the inverse, and I/O facilities.

B. Coordinate Modifications

In addition to the ability to freely interconvert between Cartesian and internal coordinates, there are a number of coordinate manipulation commands which have proved to be useful. For each of

these commands, any arbitrary subset of atoms may be employed. These commands can be summarized as follows:

Initialize —initialize selected coordinates (for subsequent building)

Copy —copy coordinates within CHARMM (CHARMM employs additional coordinate sets besides those used for evaluating the energy, for analysis, etc.)

Average —find a weighted average of two conformations

Translate —translate a selected set of atoms

Rotate —rotate a selected set of atoms

Orient —do a least-squares translation–rotation fit of two structures, or of one with a chosen axis (e.g., for a helix)

RMS —find the rms difference between two structures

The analysis capabilities of these operations are applicable for a wide range of problems. For example, the ligand asymmetry of O_2 or CO in heme binding can be observed by rotating to the principal axis system[57] of the heme atoms, translating the iron to the origin, and printing the Cartesian coordinates of the ligand. Since the heme will be parallel to the XY plane, deviation of the iron–ligand bond from the Z axis gives the asymmetry. The commands necessary to accomplish this set of transformations are

COORD ORIENT INCL ATOM SEG1 HEME *
COORD ORIENT INCL ATOM SEG1 HEME FE
PRINT COORDINATES INCL ATOM SEG1 02 *

Similar strategies can be employed to study helix packing, sheet distortion effects, and many other phenomena.

C. Hydrogen Bond Proton Coordinate Generation

In CHARMM, it is at times necessary to generate the coordinates of the proton in a hydrogen bond given only the heavy atom coordinates. Two frequently encountered occasions are the use of experimentally determined coordinate sets (where only heavy atoms can be resolved) and runs using all extended atom representations (where proton coordinates are needed for angular selection during the hydrogen-bond list update). When hy-

drogen atom positions are not known, an estimate of the best hydrogen positions is made.

There are two problems that must be solved in generating such proton coordinates. The first involves determining the hydrogen bonding pattern, while the second is placing the proton that is part of a given hydrogen bond. Inside macromolecules the pattern of hydrogen bonding is usually obvious, and placing the proton is relatively straightforward. On the molecular surface, and particularly in the bulk solvent, such is not the case. The difficulty comes from the exponential complexity of such systems of coupled bonds. Lacking a definitive method for finding the best configuration, we use an approximate method which attempts to generate a reasonable hydrogen bonding pattern in a short time. The algorithm that is employed takes the hydrogen bond donors (heavy atoms) in sequential order and for each of them, places a hydrogen bond with the best available acceptor. A second pass through the list of donors is used to place the remaining hydrogen bonds for donors (such as water) where more than one proton may hydrogen bond. In both cases, the placement of the new hydrogen bonds is constrained by requiring the preservation of the already existing hydrogen bonds. This results in hydrogen bond pattern which contains most of the strongest hydrogen bonds available to the system and places the correct number of protons on each site. The algorithm can generate pairs of protons which have strong van der Waals repulsions, particularly when bulk water is present, and it may be necessary to do dynamics on such a system to permit redistribution of the generated hydrogen bonds.[15]

Given the hydrogen bonding pattern, the problem of placing coordinates is straightforward. Coordinate generation is based on the orientation and chemical nature of the donor. For sp^2 hybridized donors, the bond length, angle, and planarity completely determine the coordinates of the protons. When the donor is an sp^3 hybridized atom, there is a dihedral angle degree of freedom which must be determined. The best hydrogen bond from the heavy atom donor is used to fix the dihedral angle so that one proton is in the optimal hydrogen bonding position. The remaining protons are place at the other rotamer positions. In the case of water, the first proton is directed toward the best acceptor, and the second proton is then oriented towards its acceptor, subject to the bond length and angle constraints from the placement of the first proton.

V. MECHANICS AND ENERGETICS

CHARMM supports three major operations involving the energy of the system. Minimization adjusts the coordinates of the system so as to lower the energy. Dynamics simulates the motion of the system and produces a trajectory; i.e., a collection of coordinates and velocities which describe the movement of the system through phase space. Normal mode analysis provides an orthonormal basis for the harmonic vibrations of a system about a particular configuration.

A. Minimization

Given a potential energy function, it is often desirable to find minimum-energy configurations of a system. In some cases, minimization is performed to relieve strain in conformations obtained experimentally or by the averaging of several structures. In other cases, finding a local or global energy minimum may be of prime interest, e.g., for determining the configuration of a peptide. For macromolecular systems, the number of local minima and the cost of the computations prevent exhaustive search of the energy surface, so it is frequently impossible to determine the global energy minimum; generally, a local minimum in the neighborhood of the x-ray structure, if available, is examined.

To provide flexibility in minimizing the energy of a range of systems for a variety of purposes, CHARMM supports five different, iterative minimization algorithms. They may be classified into a nonconvergent method (steepest descents), two convergent methods which make local harmonic approximations (conjugate gradients and ABNR), an explicit second derivative method (Newton–Raphson), and quenched dynamics. The convergent methods generally stop when they have achieved a minimum-energy configuration; the nonconvergent method never stops wandering on the energy surface.

The simplest minimization algorithm is steepest descents (SD).[2,36] In this procedure, a displacement opposite to the potential energy gradient is added to the coordinates at each step. The step size is increased if a lower energy results; otherwise, it is decreased. Although this method suffers from poor convergence, the positional shifts are gentle; it is therefore useful for small change such as removing bad contacts. However, atom harmonic constraints may also be used for this purpose.

A second method is the conjugate gradient (CG) technique which has better convergence characteristics.[58] The method makes use of the previous history of minimization steps as well as the current gradient to determine the next step. The method converges to the minimum energy in $O(N)$ steps for quadratic energy surfaces where N is the number of degrees of freedom. To achieve the same energy lowering for the systems studied, this method requires fewer evaluations of the energy and gradient than steepest descents minimization. Also, the method can produce larger coordinate shifts while progressing toward a minimum.

The third method is adopted-basis set Newton–Raphson (ABNR) minimization.[59] This algorithm stores more explicit information from the potential energy surface than does the conjugate gradients method, but avoids the $O(N^2)$ storage requirements of the full second derivative method described above. In this method, the positions and forces from each of the previous M steps are stored. The second derivative matrix is constructed by finite differences from the displacement and first derivative vectors. Newton–Raphson minimization is then applied in the subspace of the coordinates spanned by these displacement vectors. As in the explicit Newton–Raphson method described below, the second derivative matrix is diagonalized and the positive eigenvalues are inverted. Steepest descent steps are applied along the directions of zero or negative eigenvalues (again to avoid saddle points). New directions are incorporated into the displacements subspace by using a steepest descents step along the residual gradient vector calculated at each step. The behavior of the method is in many respects similar to that of conjugate gradients, but the number of energy evaluations is reduced by a factor of 2–3 by avoiding the linear interpolation phase of conjugate gradients. The storage requirements are linear in the number of atoms and the number M of previous steps that are being stored, so the method can be applied to very large systems. Since M is usually a small number (about 5), the computational requirements of the matrix diagonalization are negligble. This is the method of choice for most large systems.

The fourth method is multidimensional Newton–Raphson (NR) minimization applied to the full basis. The Newton–Raphson equations can be solved by a number of means, but the method adopted for CHARMM involves diagonalizing the second derivative matrix and then finding the

optimum step size along each eigenvector. When there are one or more negative eigenvalues, blind application of the Newton–Raphson algorithm will converge to a saddle point in the potential. To overcome this problem, a single additional energy and gradient determination is performed along the eigenvector displacement of small or negative eigenvalues. Using this additional data, an approximate cubic potential is fitted to the potential energy and the step size that minimizes this function is adopted. The advantages of this method are that the geometry cannot remain at a saddle point, as sometimes occurs with the standard procedure, and that the procedure converges rapidly when the potential is nearly quadratic (or cubic). The major disadvantage is that this procedure is difficult to apply to very large molecules because of excessive storage requirements, $O(N^2)$, and computation time per step, $O(N^3)$, for large molecules. The method is currently restricted to systems with about 200 atoms or less.

The final minimization method is quenched molecular dynamics.[60] Energy minimization is equivalent to cooling, since both remove energy from the internal modes of the system. In this method, the classical equations of motion are integrated (see below) for a system while kinetic energy is periodically removed by rescaling the velocities or by the application of a frictional damping term. The method is particularly useful when the sampling of large portions of phase space is desired in minimizing the system's energy, and it frequently converges to lower energies than either conjugate gradients or ABNR. It may be useful to follow a quenched molecular dynamics run with CG or ABNR minimization to obtain a converged structure. The random assignment of initial velocities can disrupt symmetries (such as planarity) that may be present in starting structures. Without this perturbation, symmetric systems often remain at a saddle point throughout steepest descents, conjugate gradient, or ABNR minimization. Quenched dynamics may drift significantly from the starting structure, so the method is not useful in studies of local minima or in making local structural comparisons between several minimized structures.

Table II gives a comparison of the various methods applied to two model systems, a dipeptide (Arg–Pro) with 27 atoms and a small protein (bovine pancreatic trypsin inhibitor[61]) with 580 atoms. Several criteria are available to judge the algorithms. The first is how much they lower the

Table II. Comparison of minimization techniques in CHARMM.[a]

Method[b]	Energy	Algorithm Cycles	Root Mean Square Gradient	Root Mean Square Shift	CPU Time[c]
Arg-Pro Dipeptide (Initial Energy is 23.90 kcal/mole)					
After 100 Energy Evaluations					
SD	3.81	100	0.4652	0.133	19.8 sec
CG	2.82	38	0.5954	0.225	18.0 sec
NR	-2.29	25	0.0004	0.862	12.4 min[d]
ABNR	2.09	98	0.2134	0.316	30.8 sec
QMD	4.64	100	2.8551	0.280	25.2 sec
After 500 Energy Evaluations					
SD	2.66	500	0.3694	0.218	59.8 sec
CG	0.37	189	0.0789	0.588	55.1 sec
ABNR	0.36	493	0.0380	0.580	118.8 sec
QMD	-1.46	500	4.3826	1.198	75.0 sec
PTI (Initial Energy is -167.59 kcal/mole)[e]					
After 100 Energy Evaluations					
SD	-540.40	100	1.4856	0.145	8.01 min
CG	-592.23	36	1.3356	0.305	7.48 min
ABNR	-633.27	99	0.5912	0.359	9.58 min
QMD	-584.95	100	3.0144	0.496	9.60 min
After 500 Energy Evaluations					
SD	-616.61	500	0.8179	0.293	37.46 min
CG	-693.97	186	0.8157	0.919	36.03 min
ABNR	-736.76	489	0.1980	0.933	45.12 min
QMD	-754.41	500	3.1267	1.386	41.01 min

[a] The computation of the energy and the details of the two structures minimized, an Arg–Pro dipeptide and PTI (bovine pancreatic trypsin inhibitor), are described in Appendix 3. Except as noted, none of these minimizations have converged although the CG and ABNR minimizations of the dipeptide and the ABNR minimization of PTI are approaching local minima.

[b] Steepest descents (SD), conjugate gradients (CG), Newton–Raphson (NR), adopted basis set Newton–Raphson (ABNR), quenched molecular dynamics (QMD).

[c] The CPU times listed were measured on a VAX 11/780 with a floating-point accelerator.

[d] Only 25 energy evaluations were performed. Convergence took 19 evaluations.

[e] A Newton–Raphson minimization on PTI is not feasible because diagonalizing the second derivative matrix requires about 1 CPU day.

potential energy using a given number of energy evaluations. The second is the reduction of forces, listed as the root mean square gradient of the energy evaluated over all the atoms. The third is the shift and indicates the motion of the atoms by the algorithms.

B. Dynamics

Dynamical simulations[15,62] are performed by classical mechanics and involve integration of Newton's equations of motion for a suitably prepared system; i.e., it is necessary to integrate simultaneously the equations

$$\frac{\partial^2 \mathbf{x}_i}{\partial t^2} = -\frac{\nabla E(\mathbf{x}_i)}{m_i} \qquad (31)$$

for all the atoms in the system.

CHARMM can solve these equations numerically

for all atoms in the molecule. It has two different methods available, a simple second-order predictor two-step method due to Verlet[63] and a fifth-order predictor–corrector multivalue method developed by Gear.[64] The essential approximation in both methods is that the acceleration varies slowly over the time step selected. The two numerical methods supported by CHARMM differ in their response to variations in the acceleration over a time step. The Gear method, being higher order, is more accurate than the Verlet method when the time step is short, but becomes unstable more quickly than the Verlet algorithm when the time step is increased. The smaller the time step, the better is the approximation, but a smaller time step requires a larger number of steps to simulate the system for a given period of time. In general, the time step must be significantly smaller (about 20-fold) than the period of the fastest local motion in the system, typically a bond stretch involving a hydrogen atom. Thus, for proteins with hydrogens, a time step of 5.0×10^{-4} ps is appropriate.

Since the numerical solution is obtained sequentially, there is a potential for roundoff errors to accumulate. To control such errors, these algorithms use double-precision variables for the various time derivatives of positions and for the accumulation of sums of many terms. Single precision arithmetic is used elsewhere for computational efficiency.

The accuracy of the numerical solution can be checked by several techniques. First, since the system has no external forces, the total energy should be conserved. Numerical errors will result in some fluctuations in the total energy. Since the heat capacity of the system is a measure of its response to external energy inputs, and since the heat capacity is proportional to fluctuations in kinetic energy,[65] the fluctuations in the total energy is compared to the fluctuations in kinetic energy to judge the quality of the numerical solution. A ratio of kinetic energy fluctuations to total energy fluctuations of about 100 is considered satisfactory.[55] Second, one can back integrate; i.e., run the system forward in time for some number of steps, reverse the velocities, and run it backwards the same number of steps to see how closely it returns to the starting point. This is a more rigorous check than those on the energy.

Since the period of the fastest local motion determines the maximum allowable time step, it is desirable to remove these motions, provided they do not interact with motions of interest. The fastest local motions are due to stretching

frequencies involving hydrogens. The SHAKE algorithm[55] which is capable of constraining the distance between any two atoms across time steps of the Verlet algorithm is used to remove these motions (and others optionally). SHAKE has been implemented only for the Verlet algorithm. The use of SHAKE on the bonds involving hydrogen permit a twofold increase in the step size (to 1.0×10^{-3} ps). The effect of using SHAKE on the results of the molecular dynamics simulation depends on the coupling of the constrained bond motions to the other motions in the system. This coupling is small due to the large separation in frequencies of these motions. Computations have shown that fixing all the bonds has a negligible effect on the results of a molecular dynamics simulation of bovine pancreatic trypsin inhibitor for properties involving times longer than 0.05 ps.[29]

The actual process of a molecular dynamics simulation is quite involved primarily because of the need to prepare the system in a suitable state; for most purposes, this is a system that corresponds to a microcanonical ensemble at a temperature T. There is no way of providing the numerical integration algorithms, a set of initial positions, and velocities which are equilibrated, so a preliminary molecular dynamics simulation is used to equilibrate the system in phase space at the desired temperature.

To begin the equilibration, a coordinate set is obtained from the x-ray structure (e.g., for a native protein) or, for small or idealized molecules, from the internal coordinates options described in section IV. These coordinates are then minimized using methods described earlier to eliminate any large forces due to repulsive van der Waals contacts or poor geometries. The minimizations are run either for a small number of cycles (e.g., 200 steepest descent cycles) or with constraints on the movement of all the atoms to prevent large positional shifts from the starting coordinates. If left uncorrected, the large forces, which may be present in crudely determined structures, could lead to a large distortions during the first few steps of a dynamics simulation.

Given the initial coordinates for the equilibration procedure, initial velocities are introduced to start the dynamics. CHARMM allows the user to specify zero velocity, which will start the system moving owing to the remaining forces, a constant kinetic energy for each atom with a randomly selected direction for the velocity, or a Gaussian distribution of velocities appropriate to a given temperature. The temperature T in molecular

dynamics is defined in terms of the kinetic energy by the relation

$$\sum_{i=1}^{N_a} \frac{m_i v_i^2}{2} = \frac{N_f k_B T}{2} \qquad (32)$$

where N_a is the number of atoms, m_i is the mass of atom i with velocity v_i, and N_f is the number of degrees of freedom (typically $3N_a$-6-N_s where N_s is the number of constraints applied by SHAKE).

Once initial coordinates and velocities are available, the system must be brought to the desired temperature. Again, there is no unique way of doing this. CHARMM provides several options. First, the velocities can be initialized at the desired temperature of the simulation. This is generally undesirable because this introduces a larger perturbation into the system; further, some additional adjustment of velocities will be required in any case. Second, the system can be heated slowly by first initializing the velocities at a small value, running molecular dynamics for some number of steps, then setting the velocities for a slightly higher temperature, and repeating until the desired temperature is reached. The velocities may be set by scaling them, which tends to preserve the current energy distribution, or by reassigning the velocities according to a newly generated Gaussian or other distribution. The second method is preferable because it tends to randomize the kinetic energy.

After the system is brought to the desired temperature, it usually still requires some equilibration to obtain a satisfactory representation of a microcanonical ensemble. Since the total energy is conserved in the absence of heating or cooling, and the system generally tends toward configurations of lower potential energy while equilibrating, the kinetic energy, and therefore the temperature, tend to increase. The user can continue to reassign velocities to further distribute the energy while maintaining the temperature. Alternately, CHARMM can monitor the temperature and adjust velocities either by reassignment or by scaling if the temperature drifts in either direction. Generally, the procedure is continued until the statistical properties of the system (temperature, fluctuations, mean positions) become independent of time for periods on the order of that to be used in analysis. Further, the system should be checked to verify that there are no local "hot spots" (e.g., in external side chains) or other deviations from equilibrium.

When the system is equilibrated, the actual simulation can begin. CHARMM will write the coordinates and velocities into files at separately controlled frequencies. The resulting trajectory is analyzed by the analysis facility which is described in the next section. In addition, the MOVIE program[66] may be used to display the trajectory on an Evans and Sutherland Picture System II.

A typical protocol for a protein of 1000 heavy atoms might involve 200 steps of steepest descents minimization, heating by 5° intervals using random Gaussian velocity assignments with 100 steps (0.1 ps) for equilibration for each interval until the desired temperature is reached. The system would now be run for 10 ps with random velocity reassignments every 0.2 ps from a Gaussian distribution at the desired temperature to remove biases from the heating procedure. An additional 10 ps period of equilibration dynamics with no external perturbation would be run and the results analyzed to determine that equilibration had been achieved. If so, the analysis period would be started and continued for the desired time interval (typically 25–100 ps).

During the course of the dynamics run, the atoms move, and, therefore, the lists of nonbonded interactions and hydrogen bonds must be regenerated. The user can specify update frequencies for each list. The choice of frequencies is determined by how much the atoms move and the switching functions. A reasonable choice of frequencies and switching functions [see Sec. II B, eqs. (13), (15), (16), and (26)] for a protein system at 300°K using SHAKE on bonds involving hydrogens with a step size of 0.001 ps would be as follows:

Nonbonded update frequency	25 steps (0.025 ps)
Hydrogen bond update frequency	10 steps (0.010 ps)
Nonbonded distance selection	8.0 Å
Off point for nonbonded switching	7.5 Å
On point for nonbonded switching	6.5 Å
Hydrogen bond distance selection	5.0 Å
Off distance for hydrogen bond	4.5 Å
On distance for hydrogen bond	3.5 Å
Hydrogen bond angle selection	90°
Off point for hydrogen bond angle	70°
On point for hydrogen bond angle	50°

Depending on the nature of system, modifications to these values may be appropriate.

A dynamics calculation generally takes much more computer time than is convenient for a single job. CHARMM provides a restart file which stores the coordinates, velocities, and other system variables so that the simulation can be suspended and resumed at will. In addition, these files minimize losses from unexpected computer down time.

C. Normal Mode Analysis

CHARMM has a generalized procedure for generating normal mode vectors and for their analysis. This approach gives an accurate local description of the potential energy surface. The normal modes can also be used to interpret IR spectra or allow one to improve the force constant parameters in an attempt to match the spectra.

Using a mass weighted form of the second-derivative matrix of the potential, the normal modes can be obtained through diagonalization. The eigenvectors correspond to the mass-weighted normal-mode displacements, and the eigenvalues are proportional to the square of frequencies. For small systems (up to 200 atoms or so), this can be performed by straightforward procedures using the Givens–Householder method,[67–71] where the matrix is stored in double precision and in upper triangular form. The treatment becomes somewhat more difficult for larger systems where virtual memory capacities are exceeded. For these systems, CHARMM can save the second-derivative matrix in a compressed form, which can then be diagonalized externally followed by a refinement step. In this manner, it is feasible to find the 300 lowest modes of a system with approximately 600 atoms.[72] For significantly larger systems (more than 1000 atoms), it becomes necessary either to reduce the size of the secular equation by removing degrees of freedom, or to apply an iterative eigenvector extraction procedure to obtain a small number of modes of interest.[73] Once found, normal modes can be analyzed as described below or saved for future reference.

CHARMM has a generalized procedure for manipulating and analyzing vectors that span the Cartesian coordinates of a molecule. The primary purpose of this facility is for use with normal-mode vectors, but it is not restricted to this use. For each vector, the following features are available:

(1) Transformation to internal coordinates— derivatives of bonds, angles, and torsions along a vector.
(2) Properties—dipole derivative, force projection, dot products with other vectors and rotational–translational modes.
(3) Surface exploration—find energy values along a vector, vectors (multidimensional), or an internal coordinate.
(4) Projection of coordinate differences—allows the projection of coordinate differences onto vectors.
(5) Coordinate trajectories—creation of a dy-

namics trajectory file for analysis (see Sec. V), or for display as a movie.
(6) Correlation functions—decompose dynamical fluctuations into separate contributions along vectors.

For most applications, these vectors are mass weighted and correspond to normal modes.

VI. THE ANALYSIS FACILITY

An essential part of dealing with calculations of the properties of large systems is an effective procedure for analyzing the results. CHARMM has an extensive facility for analyzing structural data and the results of the various calculations and options. It is designed for the large variety and quantity of data that are of interest in the study of macromolecules. Its fundamental operations are as follows:

(1) Creation, manipulation, printing, and graphical displays of tables: a data structure with the generality needed to effectively store and manipulate the variety of data generated by CHARMM.
(2) Comparison of two different structures, even if the chemical composition is different but related by partial sequence homology.
(3) Manipulation of a time series of any particular quantity calculated during a molecular dynamics calculation and the computation of correlation functions.
(4) Searches for close contacts between atoms of an entire structure or particular parts of it.

A. Analysis Tables

The facility deals with static properties and many dynamic properties by collecting the data into a data structure which we refer to as a table. The structure of a table has some elements of a relational data base[74] and of a hierarchical data base.[75] A relational data base is a two-dimensional array of data where each row is a record and each column is a data field. A subset of the fields provides an identifier, the key, so that any specific record may be retrieved. Selection of records may be based on the contents of any data field so that an arbitrary combination of records may be retrieved. A hierarchical data base is characterized by its use of a tree structure to organize the records, regardless of their internal structure.

The table hierarchy is used for the identification

of data in the table and is patterned after the organization of the PSF. A table has a list of segments where each segment contains a list of residues and each residue contains a list of cells. Each cell corresponds to a record in a relational data base and contains an arbitrarily long list of properties. Each segment is uniquely marked by the segment identifier, each residue by its residue identifier, and each cell by a tag. Therefore, the key for each cell is the concatenation of the segment identifier, residue identifier, and tag. Six classes of table are available: atom, bond, bond angle, torsion angle, improper torsion angle, and hydrogen bond. The class of table determines the construction of the cell's tag and what properties may be included in the cells.

(1) Table construction

For atom tables, each cell is tagged by either the IUPAC nomenclature or by the parameter-type code name of each atom (see Appendix 1). There are many properties that the cells in an atom table can contain, as listed below:

Static properties:

(1) Cartesian coordinates and distance from the geometric center
(2) Sequential position in PSF, useful for debugging
(3) The total potential energy per atom
(4) Magnitude of the force vector on each atom
(5) Energy and forces by empirical energy term; i.e., bond energy or force, bond angle energy or force, etc., associated with each atom; currently, 22 properties like this are available
(6) The accessible surface[76]
(7) Parameters for the van der Walls energy: radius, polarizability, and effective number of electrons
(8) The electric charge

Dynamic properties:

(1) Average coordinates and velocities (Cartesian components and magnitude) (eight possibilities)
(2) Second, third, and fourth moments about the mean of the coordinates (24 possibilities)
(3) Projection of the second through fourth moments on the principal axes of the spatial distribution[57] (24 properties)

(4) Isotropic temperature factor
(5) Anisotropy of coordinate and velocity fluctuations
(6) Average and standard deviation of the kinetic energy and temperature (four properties)
(7) Projection of the principal axes of the second moment tensor[57] onto the bond vectors

The cells of tables of bonds, bond angles, torsion angles, and improper torsions may contain a list of internal coordinate properties. The cell tag is constructed by concatenating the names of the atoms in the internal coordinates. The cells are associated with residues the same way the corresponding internal coordinates are associated with residues in the residue topology file. The static properties of internal coordinates are their energies and geometries. The dynamic properties of internal coordinates are the first through fourth moments of the time series of their energies and geometries as well as their minimum and maximum values.

Hydrogen bond tables have an organization similar to internal coordinate tables. Here, a hydrogen bond is associated with residues on the basis of the donor atom. The tags in a hydrogen bond table consist of a string which gives the hydrogen bond donor atom as well as the segment, residue, and atom of the acceptor. Since the list of hydrogen bonds is not necessarily fixed during dynamics, we are limited to static properties of hydrogen bonds for the table. Analysis of the dynamic properties of hydrogen bonds is possible using the time series and correlation functions described later.

Using the tables, we can see the value for every individual interaction in the potential except the nonbonded interactions. We have not supported nonbonded pair tables, because their shear size makes them relatively unwieldy and, consequently, of little practical use. Although a table of all nonbonded interactions is not available, there are means for looking at the details of such interactions. The close contact search can print information on particular nonbonded interactions. Also, there are atom properties defined for the van der Waals and electrostatic contributions of the energy of each atom.

(2) Table manipulation

Once a table is constructed, additional operations can be performed on it. The tables can be

supplemented with statistics, selectively trimmed by deletion, and printed. Examples of some of these operations are given in Appendix 2.

Statistics can be collected and added to the table. One can find the sum; average; second, third, and fourth moments about the mean; minimum; maximum; and root mean square of various collections of data in the table. Statistics can be collected over residues, segments, the whole structure, or arbitrarily selected subsets of cells in the table. Statistical data is added to new cells in the table so that the result of the addition operation is still a table.

Tables can be trimmed by selective deletion based on the values of any property or by any of the identifier or tags. This is useful if one is interested in only part of the data, as might be the case when statistics are added to the table.

The table can be printed in a variety of ways. The layout on the page, the sorting of the data, the omission of certain properties, the height and width of each page, and the number of significant digits in the output can all be controlled.

(3) Graphical display of table data

Several graphical displays of table data are available. First, data can be plotted versus the residue number or cell number from which the data came. Second, histograms of data may be generated. Third, one can make scatter plots which plot one selection of data versus another. These can reveal correlations or patterns in data. A classic example of a scatter plot is phi–psi[77] plot which shows the distribution of torsion angles in the backbone of a protein.

The normal display device for the above plots is the lineprinter. However, CHARMM can write the data necessary to drive several different plotting programs developed for use on various graphics hardware such as an Evans and Sutherland Picture System II and a Tektronix bed plotter. In addition, CHARMM can generate plotting commands to draw the complete molecule or arbitrarily selected atoms. A single-atom property from a table can be displayed by using it as a scale factor for the size of the spheres used to draw atoms. Similarly, three-atom properties from a table may be used to specify vectors emanating from each atom. Hydrogen bonds may also be plotted.

An interface to GRAMPS[78] is being developed which will provide CHARMM an interactive capability for displaying its computations.

B. Comparisons

Comparisons are an essential element of analysis, and a general scheme is provided. To allow completely general comparisons, every data structure that CHARMM uses is duplicated during comparison and may be modified or replaced. To build a table of differences between two calculations, CHARMM calculates data required for two tables and uses the difference of corresponding entries to build the actual table. Any property available for any table class may be used in these difference tables.

The program must know which atom in the comparison structure is equivalent to an atom in the main structure. In the event that the structures are identical, the problem is trivial; atoms and internal coordinates correspond identically to one another. However, if the structures are different, CHARMM matches the atoms which correspond.

The process of finding the correspondence of atoms in two different structures follows the hierarchy of the PSF. The segments are matched either by matching segments which have the same identifier or by the user's specification. Residues in matched segments are matched either by finding a homology of the residue sequence or by the user's command. Conserved residue changes are permitted in the homology-finding algorithm,[79,80] and the user may limit the range of residues over which the homology matching is performed. Atoms in matched residues may be matched either by the IUPAC nomenclature or by command.

For the atoms, a difference table of the atom class can be constructed. For the other five table classes, the correspondence of entries requires matching two lists of sets of atoms, since each internal coordinate or hydrogen bond is a set of atoms. This matching is straightforward.

The comparison command can also rotate and translate all the comparison coordinates to minimize the least square differences of the matched atoms.

An example of a comparison that can be performed would be a comparison of torsion angles in the homologous residues between the beta chain of human hemoglobin and sperm whale myoglobin. The homology would be determined for all the residues in each chain and conserved residue sets of (Asp, Glu), (Asn, Gln), (Lys, Arg), (Val, Ile, Leu, Met), (Phe, Tyr, Trp), and (Thr, Ser) could be specified to increase the number of matched residues.

C. Time Series and Time Correlation Functions

The analysis facility in CHARMM can calculate the time series for an extensive set of properties of the total system or some of its components and evaluate the time correlation functions[81,82] and spectral densities.

The basic flow of this section is as follows. The coordinate or velocity sets of a dynamics trajectory are read, and the property requested is evaluated for each set and stored. This time series $[Q(t)]$ can be manipulated to evaluate certain functionals of the time series. The resultant time series $[F(t)]$ can be integrated to obtain correlation functions. The correlation function $[C(t)]$ can further be manipulated to determine some function of the correlation function $[G(t)]$. Either the time series $[F(t)]$ or the correlation functions $[G(t)]$ can be plotted during any of these processes.

The properties for which the time series may be calculated are averages over a set of individual properties of the molecule. The individual properties can be any of the following:

(1) Position of an atom
(2) Energy or geometry of internal coordinate
(3) Vector joining two atoms
(4) Scalar product of two such vectors
(5) Scalar product of the position vectors of two atoms
(6) Fluctuations of the positions of two atoms
(7) Scalar product of the fluctuations of two atoms
(8) Velocity of an atom or the kinetic energy of an atom

The time series can also be evaluated for certain overall properties like the radius of gyration and the number density of the molecule. The ability to average over a set of atoms, vectors, or internal coordinates is helpful in defining and studying the dynamical behavior of local regions. If the correlation requested is a cross correlation between two different sets, then both of them are evaluated simultaneously and stored.

The time series can be manipulated by replacing the time series with a function of itself. The currently supported functions are as follows:

(1) Square of the time point
(2) Square root
(3) Cosine
(4) Second Legendre polynomial of the cosine of the time point
(5) Deviation from the average

(6) Ratio of each point to the average
(7) Reduction of the number of time points by averaging small sequencies of the time series together.

The simple correlation function $C(t)$ is defined by

$$C(t) = \langle \mathrm{F}(0) \cdot F'(t) \rangle = \frac{\sum\limits_{t_m} F(t_m)F'(t + t_m)}{\sum\limits_{t_m} F(t_m)F'(t_m)} \quad (33)$$

The correlation function can be evaluated using either the direct integration method or the convolution in Fourier space. The Fourier transformation and the inverse transformation are performed using the fast Fourier transform (FFT) method.[83] The use of FFT makes the latter procedure extremely efficient as the number of time points exceeds 500. Three different types of correlation functions are included:

$$C(t) = \langle F(0) \cdot F'(t) \rangle \quad (34)$$

$$C(t) = \langle \cos[F(0) - F'(t)] \rangle \quad (35)$$

$$C(t) = \langle 3[F(0) \cdot F'(t)]^2 - 1 \rangle \quad (36)$$

The $F(t)$ and $F'(t)$ correspond to the time series and are identical in the case of autocorrelation functions. In general, correlation functions should be computed for times less than half the length of the simulation and, in some cases, the accuracy is good only for much shorter periods.[84,85]

Examples of experimental quantities that can be calculated using these correlation functions are friction coefficients, IR line widths, fluorescence depolarization rates, and NMR relaxation times.[81,82,86]

Once the correlation function is calculated, a number of operations can be performed on it. The integral, the integral of the square, and the logarithm of the correlation function can be evaluated. Finally, the Fourier transform of the correlation function can be taken to obtain the spectral density.

The series of operations described above are modular. The five modules consist of obtaining a time series $[Q(t)]$, manipulating the time series $[F(t)]$, integrating to obtain the correlation function $[C(t)]$, manipulating the correlation function $[G(t)]$, and plotting these functions. This modular approach permits arbitrary combinations of these manipulations. For example, the number of types of time series $[F(t)]$ for which a correlation function can be evaluated is on the order of the product of the number of types of time series $[C(t)]$ and the

number of manipulations that can be performed on it.

D. Close Contact Searches

The analysis facility does not produce tables of nonbonded interactions, but there exists a command for looking for close contacts. One may look for all pairs of atoms in contact within some cutoff, excluding pairs which are bonded together. One may also search for contacts near a particular atom, near some spatial point, or within or near a particular residue or segment. The result of this search is a printout of all atoms which are found and can be supplemented with nonbonded energy calculations for those pairs.

VII. PROGRAM DESIGN

The design of CHARMM has a number of features distinct from most scientific programs. These features have been introduced to make the program convenient to use and to facilitate its orderly evolution. Since the methodology used in CHARMM is constantly evolving, it is essential that the program be easily modified and maintained. Transportability of the program and its data files is also desirable.

The virtual memory capacities of the VAX have been exploited in the design of CHARMM. CHARMM is a single program which does everything described in this paper. It contains about 52,000 lines of code and requires about 1 megabyte of memory for the compiled machine instructions and 1.5 megabytes for static storage. Using one program maximizes its versatility in combining various calculations while minimizing the number of intermediate files a user must save. A single program also facilitates development, because it guarantees that support and utility functions are always available. In addition, the overhead for maintaining one program is less than that for maintaining many programs of similar capabilities, as logical inconsistencies are easier to detect and testing is simplified. We maintain a set of approximately two dozen test cases which utilize most of the features of CHARMM. Such periodic testing is essential to preserving the integrity and reliability of the program.

There are some disadvantages resulting from the large size of the program. It intimidates new users, and the program cannot be run on computers lacking virtual memory without breaking it into pieces. Despite its large size, CHARMM has

been transported to an IBM 370 series computer, and portions of it have been run on a CDC 7600 and a Cray-1.

A. Program Architecture

The architecture of CHARMM can be divided into two functional parts, a highly modular command interpreter which consists of many easily separable pieces and a highly interconnected set of subroutines which provide support for the command interpreter. Figure 1 shows a schematic of this architecture, and Table I describes the various boxes in the figure.

The top-level command interpreter parses commands present in the user's input file and passes control to the appropriate routine. Thus, the user has close control of CHARMM's execution. Many of the more complex routines have command interpreters of their own which operate in a similar fashion. Nearly all of the commands accepted by CHARMM are free field and use keywords rather than position in the command to specify options. This freedom from using FORTRAN FORMAT statements simplifies the use of CHARMM and reduces the possibility of input errors.

All of the major data structures in CHARMM are stored in the top-level command interpreter or the analysis facility. This centralized store provides the mechanism for data flow through the command subroutines. In addition, it has made it easy to introduce new functions because the data required are always readily available.

The support routines provide a large set of operations useful for CHARMM. Routines exist for evaluating the various energy contributions which are called from numerous places in the program. Dynamic memory allocation is implemented as described below. A complete library of string manipulation routines is available for parsing and data manipulation. There are also routines for array manipulations, input/output, mathematical functions, etc. The availability of these routines makes developing CHARMM easier.

B. FORTRAN and FLECS

CHARMM is written in FLECS,[87] which is an extension of FORTRAN. The language is upwardly compatible with FORTRAN-77 in most respects. The FLECS source code is converted to FORTRAN with the application of the FLECS preprocessor. The preprocessor is written using FORTRAN and

is easily transported. Versions of FLECS exist on a variety of computers including IBM 360 and 370 series, CDC 6600 and 7600 series, DEC PDP-10s, PDP-11s, and Intel 8080 microprocessors.

The additional capabilities afforded by FLECS are in the handling of control flow and the use of internal procedures within a subprogram. The control flow constructs in FLECS permit us to forego the GOTO statement, which simplifies the more complex routines.[88] The internal procedures are useful in making subprograms more modular without the cost of calling external subroutines.

C. Documentation

An essential aspect of the program is good documentation on how to use CHARMM and how it is written. The documentation on its usage currently comprises some 70 000 words of text stored in a hierarchical, on-line data base. General information on the implementation of CHARMM is also kept in this data base, but implementation details are kept in the code. CHARMM has about 5000 lines of text embedded in the comments. The use of FLECS reduces the documentation requirements of the code, as FLECS makes the control flow self-evident.

D. Dynamic Storage Allocation

The ability to allocate storage for arrays during the execution of a program is very valuable. There are many routines in CHARMM which require temporary or permanent array space based on the size of the system. Without dynamic array allocation, enormous amounts of virtual memory would be wasted because all the arrays would be allocated throughout an execution of CHARMM (when most would not be needed), and because their sizes would be at their maximum limits rather than appropriate to the actual needs. We have estimated that without dynamic storage allocation CHARMM's virtual memory, requirements would exceed 10 megabytes. Such memory requirements would make a VAX into a single-user system.

Since FORTRAN does not provide dynamic storage allocation, we have implemented our own. There are two dynamic storage schemes used in CHARMM, a stack and a heap.[89]

The stack is a data structure which consists of an array, called STACK, and an integer, the stack pointer. The stack pointer serves as an index into

the array, giving the last location in stack which is in use. Initially, the stack pointer is set to zero. As storage is requested, space is allocated in one direction only. Storage is freed in inverse order to the way it was requested. When stack space is requested, the allocation routine adjusts the stack pointer and returns the index into the stack array of the word of the allocated storage. The program then can make direct reference into the stack array or, by using a subroutine call, one can map the space on the stack into an array of arbitrary type and name. Currently, the stack size is fixed at 100,000 words, and if the stack overflows, execution terminates. Generally, very large, temporary dynamic storage needs are met on the heap, because it has less constraints on its ultimate size.

The stack may be used for temporary work arrays needed only for the execution of a given subroutine. When the subroutine exits, the space is returned. In the event of nested subroutine calls where such storage is required, the nested execution of the subroutine will leave the stack in the same state it was in before it was called. Since the majority of dynamic storage allocation needs involve temporary storage, the stack is used heavily in CHARMM.

The heap provides permanent dynamic storage, as it allows an arbitrary order to the allocation and release of space. The heap consists of a large array, HEAP, and an integer. The integer, known as the free list pointer, gives the location in the heap of the head of a linked list which contains the unused storage in the heap. Initially, the free list is set so that the entire array is available. As space is requested and released, the free list is updated to indicate the current availability of storage. Data stored in the heap may be referenced in the same way as data on the stack. The initial size of the heap is set to 30,000 words. When it overflows, a request is made to the operating system for more space. Since the free storage need not be continuous, the space provided by the operating system can be easily linked into the free list. This machine dependency can be eliminated by increasing the static size of the heap to the anticipated requirements of a run.

E. Unified Data Structures

It is desirable to manipulate a large number of arrays and scalars as single entities in order to organize data and reduce the number of parameters that must be passed from one subroutine to another. For example, the routine which constructs

a specific comparison table would require roughly 500 parameters. In addition, it is desirable to allocate such arrays dynamically. We have developed a scheme for manipulating such data structures in FORTRAN where just two arrays represent a data structure consisting of an arbitrary number of arrays and scalars. The scheme also allows us to modify at will the size of any element in the data structure during execution.

The unified data structures are stored on the heap. To keep track of each element in the data structure, we use a pair of arrays, a base array, which stores the index into the heap of each element, and a length array, which stores the length. To associate each element of the base and length arrays with elements in a data structure, index variables are used. The name of an index variable is the name of the element; the value of the variable gives the index into the two arrays. Each data structure has its own set of index variables. To keep the values of these variables consistent from subroutine to subroutine, the index variables are kept in a common block and initialized before any data structures are used. By using an INCLUDE statement,* we can easily obtain the common blocks wherever they are needed and without error. If we require duplicate data structures, we need to create only another base and length array which have different names than the original. If we wish to change the size of our data structures, we allocate space on the heap for the newly sized data structure, copy everything into it, and return the space used by the old copy.

VIII. CONCLUDING DISCUSSION

CHARMM demonstrates that it is feasible to implement a molecular mechanics program with a wide variety of functions. It is general enough to treat efficiently a range of molecules from simple peptides to the hemoglobin tetramer, and systems from a single molecule to a full crystal composed of a protein and several hundred solvent molecules, while it simultaneously preserves the adaptability and maintainability essential to an academic environment where new projects and application requirements are frequently called for. In assessing the outcome of such a project, it is necessary to recognize the capabilities and limi-

tations of the current theory. This section will present a brief discussion of a few of the applications of empirical energy function modeling and some of the issues involved in the choice of the potential energy function and the method of mechanical analysis. These considerations determine the types of calculation which may be profitably persued at present and the limits in the application of these methods to other areas. We will conclude with a brief summary of some of the areas where theoretical developments are in progress to extend the capabilities of the program.

Empirical energy modeling and mechanical analysis give complete knowledge of the system under study, within the limits of the theory employed. There is an extensive experimental literature demonstrating the importance of molecular motion in the physiological,[90,91] chemical,[92,93] and physical[94] properties of biological macromolecules. Frequently, these observations show that some form of motion must be present but give little information on the details of that motion. Explicit modeling is particularly useful in these cases, giving insight into the types of motion and mechanisms involved in the activity of macromolecules. For example, the crystal structure of the myoglobin heme pocket does not permit sufficient clearance for the diffusion of ligand oxygen to the heme,[91] but modeling studies showed that energetically feasible conformational distortions would permit facile diffusion into the site.[90] In other cases, dynamical properties such as the rate of tyrosine ring flipping[95] suggest significant motion that may be of interest. Although these motions are comparatively rare (10^2–10^6/s), they may be simulated using static[24] or activated dynamical[96] methods of analysis. Vibrational properties are easily and naturally analyzed in terms of normal coordinates.[47] Solvation effects are also amenable to study and of particular interest in understanding the properties of biological molecules.[18–21] In any of these examples, the approximations of the potential energy function and mechanics employed in modeling will determine the reliability of the results.

The potential energy function chosen for CHARMM is based on internal coordinate and pairwise nonbond energy terms which allows it to be applied to either all-atom or extended-atom

---

* An INCLUDE statement causes a file to be inserted in place of the INCLUDE statement. It is a feature of VAX FORTRAN and is common on many new FORTRAN compilers. Since it is a nonstandard feature of Fortran, we have written a preprocessor that performs the action of INCLUDE prior to the compiler being called. This would allow us to transport CHARMM to another compiler which did not have this feature.

representations of macromolecules. The former is useful in systems with only a few heavy atoms. The latter greatly accelerates calculations on large systems because it reduces the number density of atomic centers by a factor roughly two and consequently the number of nonbonded contacts by a factor of 4. A simple Coulombic and 6–12 van der Waals term is used to calculate these nonbonded interactions, but even so they are the most time-consuming step in most calculations.

Several effects are neglected in this potential energy function. The harmonic form employed for the internal coordinate energies is a local approximation to the potential near minima. The functional form of the nonbond interaction neglects charge-induced dipole terms and the details of short-range repulsive interactions. Cross (valence) couplings between internal coordinate energies[28] are neglected, and three-body polarization effects[17] are not included. The form of the CHARMM potential energy reflects a balance of the requirements of computational efficiency and meaningful accuracy. There exist simpler and more restricted potentials (e.g., fixing internal degrees of freedom[23,40] or neglecting electrostatic interactions[97]) that are more efficient but less accurate, and more elaborate force fields and *ab initio* methods that are more accurate but too slow to apply to systems for which CHARMM was designed.

In most applications, the energy function is applied to an isolated system, in effect a vacuum phase calculation. The natural environment of most biological macromolecules is, however, aqueous solution where strong solvation forces are present. In some calculations, we include solvent molecules explicitly,[29,33] while in others we attempt to model some of the more important aspects of solvation by introducing electrostatic screening terms[27,98] or constraints on surface atoms. The theory of aqueous solvation is at present an approximate one, and the quantitative accuracy of calculations involving solvation effects is limited by our knowledge in this area.

Three major types of mechanical analysis are routinely employed in CHARMM: molecular dynamics, energy minimization, and vibrational analysis. Dynamics is the explicit simulation of a molecule's evolution in time over periods as long as 1 ns. This is sufficient for the analysis of local atomic fluctuations and some types of internal rearrangements and allows the calculation of entropic effects in well defined systems. Energy minimization methods are useful in finding local and, for small systems, global minimum-energy

configurations. As the low temperature limiting structures, these are of interest both in themselves and as starting points for either dynamical or vibrational analysis. Minimization methods may be combined with internal or generalized coordinate constraints to study barrier crossing or internal rearrangement in systems where dynamical analysis may not be feasible. Energy minimization methods are limited in their application to large systems by the time required to achieve convergence, as well as the lack of methods to find the global minimum-energy configuration. The analysis of molecular vibrations is useful for many systems, although for macromolecules it may be of interest to extract only a fraction of all $3N - 6$ vibrational modes. Through the combination of these methods, insight is being gained into the structure and flexibility and, consequently, the chemical properties of large molecular systems.

In addition to the mechanical analysis cited above, static structural analysis is a major use of the program. The information required varies widely, and a considerable effort has been placed on maximizing the flexibility of the analysis section.

Our work focuses on the chemistry of condensed phases, with particular emphasis on the study of macromolecular systems found in biology. The program has been employed in projects ranging from the exploration of macromolecular solvation to protein–DNA interactions and many associated studies of constituent small-molecule properties. The very large size and lack of symmetry of these systems presents us with challenging computational requirements. The methods developed to deal with these demands have application in other areas of theoretical (e.g., fluid and polymer mechanics) and experimental (e.g., crystallography, structure refinement, NMR, and other spectroscopy interpretation) study. By simulating biological macromolecules, we hope to improve our understanding of their properties and of the forces acting within them. Such knowledge will in turn help to elucidate their function and the mechanisms involved in macromolecular structure and assembly, binding site recognition, and specificity. Enzymes are among the most efficient and versatile catalysts known. The chemical and physical understanding of proteins gained through simulation will be directly applicable to understanding these unique catalysts. Combined molecular orbital and empirical energy function calculations are planned to examine the detailed interaction of molecular mechanics with electronic structure.

Nucleic acids and their transformations, which play an essential role in genetics, are being studied.

A version of CHARMM is available for distribution to non-profit institutions. Because the program is constantly evolving and continued testing by actual applications is required, the present EXPORT version has more limited capabilities than the program described in this article.

The authors wish to thank John Brady, Bruce Gelin, Jeffrey Hoch, Toshiko Ichiye, Peter Kollman, Robert Ladner, Angel Lee, Michael Levitt, Ronald Levy, J. Andrew McCammon, David Perahia, John Ramsdell, Wally Reiher, Wilfred van Gunsteren, Mark Wagman, and Arieh Warshel for helpful discussions and assistance in implementing CHARMM or its predecessor energy minimization programs. One of the authors (M.K.) particularly wishes to thank Shneior Lifson for his hospitality when the author spent a semester at the Weizmann Institute, where he was first introduced to the potential of empirical energy functions.

APPENDIX 1

A Set of CHARMM Protein Parameters

The following tables give a complete set of parameters used by CHARMM to calculate the potential energy of a system of proteins using the explicit hydrogen representation for atoms. The units used are derived from a system where the angstrom (Å) is the unit of distance, kcal/mol is the unit of energy, and atomic mass unit (amu) is the unit of mass (the AKMA system). The values correspond to PARAM4, the most recent parameter set.

Atom Classification

Name	Description
H	Hydrogen which can hydrogen bond bound to neutral atom
HC	Hydrogen which can hydrogen bond bound to charged atom
HA	Aliphatic hydrogen
CT	Aliphatic Carbon
C	Carbonyl Carbon
CH1E	An extended atom carbon with one hydrogen
CH2E	An extended atom carbon with two hydrogens
CH3E	An extended atom carbon with three hydrogens
CR1E	A extended atom carbon in an aromatic ring with one hydrogen
N	A peptide nitrogen with no hydrogens attached
NR	A nitrogen in an aromatic ring with no hydrogens
NP	A pyrole nitrogen
NH1E	An extended atom peptide nitrogen with one hydrogen
NH2E	An extended atom peptide nitrogen with two hydrogens
NH3E	An extended atom nitrogen with three hydrogens
NC2E	An extended atom charged guanidinium nitrogen with to two hydrogens
NH1	A peptide nitrogen bound to one hydrogen
NH2	A peptide nitrogen bound to two hydrogens
NH3	A nitrogen bound to three hydrogens
NC2	A charged guanidinium nitrogen bound to hydrogens
O	Carbonyl oxygen
OC	Carboxy oxygen
OH1E	Extended atom hydroxy oxygen
OH2E	Extended atom water
OH1	Hydroxy oxygen
OH2	Water oxygen
S	Sulphur
SH1E	Extended atom sulphur with one hydrogen
FE	Iron (as in heme)

In order to go from a set of atoms representing a term to the appropriate parameters, the atoms that CHARMM manipulates are chemically classified. The previous table gives the current classification of atoms.

Bond Parameters: $E_b = k_b(r - r_0)^2$

Bond	k_B (kcal/mole-Å²)	r_0 (Å)
C - C	450.0	1.38
C - CH1E	405.0	1.52
C - CH2E	405.0	1.52
C - CH3E	405.0	1.52
C - CR1E	450.0	1.38
C - CT	405.0	1.53
C - N	471.0	1.33
C - NC2	400.0	1.33
C - NH1	471.0	1.33
C - NH2	471.0	1.33
C - NP	471.0	1.33
C - NR	471.0	1.33
C - O	580.0	1.23
C - OC	580.0	1.23
C - OH1	450.0	1.38
CH1E - CH1E	225.0	1.53
CH1E - CH2E	225.0	1.52
CH1E - CH3E	225.0	1.52
CH1E - N	422.0	1.45
CH1E - NH1	422.0	1.45
CH1E - NH2	422.0	1.45
CH1E - NH3	422.0	1.45
CH1E - OH1	400.0	1.42
CH2E - CH2E	225.0	1.52
CH2E - CH3E	225.0	1.54
CH2E - CR1E	250.0	1.45
CH2E - N	422.0	1.45
CH2E - NH1	422.0	1.45
CH2E - NH2	422.0	1.45
CH2E - NH3	422.0	1.45
CH2E - OH1	400.0	1.42
CH2E - S	450.0	1.81
CH2E - SH1E	450.0	1.81
CH3E - NH1	422.0	1.49
CH3E - S	450.0	1.77
CR1E - CR1E	450.0	1.38
CR1E - NH1	450.0	1.305
CR1E - NR	450.0	1.305
CT - CT	200.0	1.53
CT - N	422.0	1.45
CT - NC2	422.0	1.45
CT - NH1	422.0	1.45
CT - NH2	422.0	1.45
CT - NH3	422.0	1.45
CT - OH1	400.0	1.42
CT - S	450.0	1.81
FE - NP	500.0	2.09
FE - NR	65.0	1.98
H - NH1	405.0	0.98
H - NH2	405.0	0.98
H - OH1	450.0	0.96
H - OH2	450.0	0.96
HA - C	350.0	1.08
HA - CT	300.0	1.08
HC - NC2	405.0	1.00
HC - NH1	405.0	0.98
HC - NH3	405.0	1.04
OC - S	400.0	1.43
S - S	500.0	2.02

Bond Angle Parameters: $E_\theta = k_\theta(\theta - \theta_0)^2$

Angle	k_θ	θ_0
C - C - C	70.0	106.5
C - C - CH2E	65.0	126.5
C - C - CH3E	65.0	126.5
C - C - CR1E	70.0	122.5
C - C - CT	70.0	126.5
C - C - HA	40.0	120.0
C - C - NH1	65.0	109.0
C - C - NP	65.0	112.5
C - C - NR	65.0	112.5
C - C - OH1	65.0	119.0
CH1E - C - N	20.0	117.5
CH1E - C - NH1	20.0	117.5
CH1E - C - O	85.0	121.5
CH1E - C - OC	85.0	117.5
CH1E - C - OH1	85.0	120.0
CH2E - C - CR1E	70.0	121.5
CH2E - C - N	20.0	117.5
CH2E - C - NH1	20.0	117.5
CH2E - C - NH2	20.0	117.5
CH2E - C - NR	60.0	116.0
CH2E - C - O	85.0	121.6
CH2E - C - OC	85.0	118.5
CH2E - C - OH1	85.0	120.0
CH3E - C - N	20.0	117.5
CH3E - C - NH1	20.0	117.5
CH3E - C - O	85.0	121.5
CR1E - C - CR1E	65.0	120.5
CR1E - C - NH1	65.0	110.5
CR1E - C - NP	65.0	122.5
CR1E - C - NR	65.0	122.5
CR1E - C - OH1	65.0	119.0
CT - C - N	20.0	117.5
CT - C - NH1	20.0	117.5
CT - C - NH2	20.0	117.5
CT - C - O	85.0	121.5
CT - C - OC	85.0	118.5
CT - C - OH1	85.0	120.0
HA - C - NH1	40.0	120.0
HA - C - NH2	40.0	120.0
HA - C - NR	40.0	120.0
HA - C - O	85.0	121.5
N - C - O	85.0	121.0
NC2 - C - NC2	70.0	120.0
NC2 - C - NH1	70.0	120.0
NH1 - C - NR	70.0	120.0

Atoms	k_θ	θ_0
NH1 – C – O	65.0	121.0
NH2 – C – O	65.0	121.0
O – C – OH1	85.0	120.0
OC – C – OC	85.0	122.5
C – CH1E – CH1E	70.0	110.0
C – CH1E – CH2E	70.0	109.5
C – CH1E – CH3E	70.0	106.5
C – CH1E – N	45.0	111.6
C – CH1E – NH1	45.0	111.6
C – CH1E – NH2	45.0	111.6
C – CH1E – NH3	45.0	111.6
CH1E – CH1E – CH2E	45.0	112.5
CH1E – CH1E – CH3E	45.0	111.0
CH1E – CH1E – NH1	50.0	110.0
CH1E – CH1E – NH2	50.0	109.5
CH1E – CH1E – NH3	50.0	107.5
CH1E – CH1E – OH1	50.0	104.5
CH2E – CH1E – CH3E	50.0	111.5
CH2E – CH1E – N	65.0	104.0
CH2E – CH1E – NH1	65.0	110.0
CH2E – CH1E – NH2	65.0	110.0
CH2E – CH1E – NH3	65.0	110.0
CH3E – CH1E – CH3E	50.0	111.0
CH3E – CH1E – NH1	65.0	108.5
CH3E – CH1E – NH2	65.0	109.5
CH3E – CH1E – NH3	65.0	109.5
CH3E – CH1E – OH1	60.0	110.5
C – CH2E – CH1E	70.0	112.5
C – CH2E – CH2E	70.0	113.0
C – CH2E – NH1	70.0	111.6
C – CH2E – NH2	70.0	111.6
C – CH2E – NH3	70.0	111.6
CH1E – CH2E – CH1E	45.0	117.0
CH1E – CH2E – CH2E	45.0	112.5
CH1E – CH2E – CH3E	45.0	113.0
CH1E – CH2E – OH1	45.0	111.0
CH1E – CH2E – S	50.0	112.5
CH1E – CH2E – SH1E	50.0	112.5
CH2E – CH2E – CH3E	45.0	110.0
CH2E – CH2E – CH3E	45.0	111.0
CH2E – CH2E – N	65.0	105.0
CH2E – CH2E – NH1	65.0	111.0
CH2E – CH2E – NH2	65.0	109.5
CH2E – CH2E – NH3	65.0	110.5
CH2E – CH2E – S	50.0	112.5
CH3E – CH2E – OH1	45.0	111.0
C – CR1E – C	90.0	126.5
C – CR1E – CH2E	90.0	122.0
C – CR1E – CR1E	90.0	119.0
C – CR1E – NH1	90.0	109.5
C – CR1E – NR	90.0	106.5
CR1E – CR1E – CR1E	90.0	120.5
NH1 – CR1E – NH1	70.0	109.0
NH1 – CR1E – NR	70.0	109.0
C – CT – CT	70.0	109.5
C – CT – HA	70.0	109.5
C – CT – N	70.0	111.6
C – CT – NH1	70.0	111.6
C – CT – NH2	70.0	111.6
C – CT – NH3	70.0	111.6
CT – CT – CT	45.0	111.00
CT – CT – HA	40.0	109.50
CT – CT – N	65.0	105.00
CT – CT – NC2	65.0	110.00
CT – CT – NH1	65.0	110.00
CT – CT – NH2	65.0	110.00
CT – CT – NH3	65.0	110.00
CT – CT – OH1	50.0	109.50
CT – CT – S	50.0	112.50
HA – CT – HA	40.0	109.5
HA – CT – N	50.0	109.5
HA – CT – NC2	50.0	109.5
HA – CT – NH1	50.0	109.5
HA – CT – NH3	50.0	109.5
HA – CT – OH1	50.0	109.5
HA – CT – S	40.0	109.5
NP – FE – NP	50.0	90.0
NP – FE – NR	5.0	115.0
C – N – CH1E	80.0	120.0
C – N – CH2E	80.0	120.0
C – N – CT	80.0	120.0
CH1E – N – CH2E	60.0	110.0
CH1E – N – CH3E	60.0	110.0
CH2E – N – CH3E	60.0	109.5
CT – N – CT	60.0	110.0
C – NC2 – CT	80.0	120.0
C – NC2 – HC	35.0	120.0
CT – NC2 – HC	35.0	120.0
HC – NC2 – HC	40.0	120.0
C – NH1 – C	60.0	102.5
C – NH1 – CH1E	77.5	120.0
C – NH1 – CH2E	77.5	120.0
C – NH1 – CH3E	77.5	120.0
C – NH1 – CR1E	60.0	108.0
C – NH1 – CT	80.0	120.0
C – NH1 – H	30.0	120.0
CH1E – NH1 – CH3E	60.0	120.0
CH1E – NH1 – H	35.0	120.0
CH2E – NH1 – CH3E	60.0	120.0
CH2E – NH1 – H	35.0	120.0
CH3E – NH1 – H	35.0	120.0
CR1E – NH1 – CR1E	65.0	110.0
CR1E – NH1 – H	35.0	120.0
CT – NH1 – H	35.0	120.0
C – NH2 – H	30.0	120.0
CH1E – NH2 – CH2E	60.0	120.0
CH1E – NH2 – H	35.0	120.0
CH2E – NH2 – H	35.0	120.0
CT – NH2 – H	35.0	120.0
H – NH2 – H	40.0	125.0
CH1E – NH3 – CH2E	35.0	109.5
CH1E – NH3 – HC	35.0	109.5
CH2E – NH3 – HC	35.0	109.5
CT – NH3 – HC	35.0	109.5
HC – NH3 – HC	40.0	109.5
C – NP – C	70.0	102.5
C – NP – FE	50.0	128.0
C – NR – C	70.0	102.5
C – NR – CR1E	70.0	109.5
CR1E – NR – CR1E	65.0	110.0
CR1E – NR – FE	30.0	124.8
C – OH1 – H	50.0	109.5
CH1E – OH1 – H	35.0	109.5
CH2E – OH1 – H	35.0	109.5
CT – OH1 – H	35.0	109.5
H – OH2 – H	55.0	104.5
CH2E – S – CH3E	50.0	99.5
CH2E – S – S	50.0	104.2
CT – S – CT	50.0	99.5
CT – S – S	50.0	104.2
OC – S – OC	85.0	109.5

Torsion Angle Parameters: $E_\phi = |k_\phi| - k_\phi \cos(n\phi)$

Atoms	k_ϕ	n
CH1E – C – N – CH1E	10.0	2
CH2E – C – N – CH1E	10.0	2
CR1E – C – C – CR1E	5.0	2
CR1E – C – C – C	2.5	2
CR1E – C – C – NH1	2.5	2
X – C – CH1E – X	0.0	3
X – C – CH2E – X	0.0	3
X – C – CR1E – X	10.0	2
X – C – CT – X	0.0	3
X – C – N – X	8.2	2
X – C – NC2 – X	8.2	2
X – C – NH1 – X	8.2	2
X – C – NH2 – X	8.2	2
X – C – OH1 – X	1.8	2
X – CH1E – CH1E – X	-1.6	3
X – CH1E – CH2E – X	-1.6	3
X – CH1E – N – X	-0.3	3
X – CH1E – NH1 – X	-0.3	3
X – CH1E – NH2 – X	-1.8	3
X – CH1E – NH3 – X	-0.6	3
X – CH1E – OH1 – X	-0.5	3
X – CH2E – CH2E – X	-1.6	3
X – CH2E – N – X	-0.3	3
X – CH2E – NH1 – X	-0.3	3
X – CH2E – NH2 – X	-0.6	3
X – CH2E – NH3 – X	-0.6	3
X – CH2E – OH1 – X	-0.5	3
X – CH2E – S – X	-1.2	2
X – CT – CT – X	-1.6	3
X – CT – N – X	-0.3	3
X – CT – NC2 – X	-0.3	3
X – CT – NH1 – X	-0.3	3
X – CT – NH2 – X	-0.6	3
X – CT – NH3 – X	-0.6	3
X – CT – OH1 – X	-0.5	3
X – CT – S – X	-1.2	2
X – FE – NR – X	-0.05	4
X – S – S – X	-4.0	2

Improper Torsion Parameters: $E_\omega = k_\omega(\omega - \omega_0)^2$

Improper Torsion	k_ω (kcal/mole-radian^2)	ω_0 (degrees)
C – C – CR1E – CH2E	90.0	0.0
C – CR1E – C – CH2E	90.0	0.0
C – CR1E – CR1E – CH2E	90.0	0.0
C – CR1E – NH1 – CH2E	90.0	0.0
C – NH1 – CR1E – CH2E	90.0	0.0
C – CR1E – CR1E – OH1	150.0	0.0
C – H – H – NH2	45.0	0.0
C – OC – OC – CH1E	100.0	0.0
C – OC – OC – CH2E	100.0	0.0
C – X – X – C	25.0	0.0
C – X – X – CH2E	90.0	0.0
C – X – X – CH3E	90.0	0.0
C – X – X – CR1E	25.0	0.0
C – X – X – H	75.0	0.0
C – X – X – HA	75.0	0.0
C – X – X – NH1	100.0	0.0
C – X – X – O	100.0	0.0
C – X – X – OC	100.0	0.0
C – X – X – OH1	150.0	0.0
CH1E – X – X – CH1E	55.0	35.26439
CH1E – X – X – CH2E	55.0	35.26439
CH1E – X – X – CH3E	55.0	35.26439
CR1E – X – X – CR1E	25.0	0.0
CR1E – X – X – NH1	25.0	0.0
FE – X – X – NP	20.0	0.0
H – X – X – O	45.0	0.0
N – CH1E – CH2E – C	45.0	0.0
N – X – X – CH2E	45.0	0.0
N – X – X – CT	45.0	0.0
NC2 – X – X – CT	45.0	0.0
NC2 – X – X – HC	45.0	0.0
NH1 – X – X – CH1E	45.0	0.0
NH1 – X – X – CH2E	45.0	0.0
NH1 – X – X – CH3E	45.0	0.0
NH1 – X – X – CT	45.0	0.0
NH1 – X – X – H	45.0	0.0
NH1 – X – X – NH1	25.0	0.0
NH1 – X – X – NR	25.0	0.0
NH2 – X – X – H	45.0	0.0
NR – X – X – C	25.0	0.0
NR – X – X – CR1E	25.0	0.0
NR – X – X – CT	25.0	0.0

Nonbonded Parameters

The number of nonbonded parameters, A_{ij} and B_{ij}, is equal to $N(N + 1)/2$, where N is the number of atom types.

Since this parameter set is defined in terms of 28 atom types, there are 406 nonbonded parameters. Rather than include them, we shall show how they are generated.

The values for A_{ij} and B_{ij} in the nonbonded term is given by the following equation due to Slater and Kirkwood.[99,100]

$$B_{ij} = \frac{3}{2}\left(\frac{1}{4\pi\epsilon_0}\right)^{1/2} \frac{e\hbar m_e^{-1/2}\alpha_i\alpha_j}{(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}},$$

$$A_{ij} = \tfrac{1}{2}B_{ij}(R_i + R_j)^6 \quad \text{(A1)}$$

i,j	two atoms in an nonbonded interaction
ϵ_0	permittivity of vacuum
e	electron charge
$\hbar$	Planck's constant
m_e	electron rest mass
α_i	polarizability
N_i	effective number of outer shell electrons
R_i	van der Waals radius

The values for α_i, N_i, and R_i used by CHARMM are given in the table below.

Atom	α_i	N_i	R_i
C	1.6500	5.0	1.8000
CH1E	1.3500	6.0	1.8500
CH2E	1.7700	7.0	1.9000
CH3E	2.1700	8.0	1.9500
CR1E	2.0700	6.0	1.9000
CT	0.9800	5.0	1.8000
FE	0.0100	0.0001	0.6500
H	0.0440	1.0	0.8000
HA	0.1000	1.0	1.4680
HC	0.0440	1.0	0.8000
N	1.1000	6.0	1.6000
NC2	1.1000	6.0	1.6000
NC2E	1.7000	8.0	1.7000
NH1	1.1000	6.0	1.6000
NH1E	1.4000	7.0	1.6500
NH2	1.1000	6.0	1.6000
NH2E	1.7000	8.0	1.7000
NH3	1.1000	6.0	1.6000
NH3E	2.1300	9.0	1.7500
NP	1.1000	6.0	1.6000
NR	1.1000	6.0	1.6000
O	0.8400	6.0	1.6000
OC	2.1400	6.0	1.6000
OH1	0.8400	6.0	1.6000
OH1E	1.2000	7.0	1.6500
OH2	0.8400	6.0	1.6000
OH2E	1.2000	7.0	1.7000
S	0.3400	16.0	1.9000
SH1E	0.3400	16.0	1.9500

Hydrogen Bond Parameters

Donor	Acceptor	E_{min}	R_{min}
Nitrogen	Nitrogen	−3.00	3.0
Nitrogen	Oxygen	−3.50	2.9
Oxygen	Nitrogen	−4.00	2.85
Oxygen	Oxygen	−4.25	2.75

Note that electrostatic interactions computed separately from this term increase the magnitude of the hydrogen bond minimum energy significantly.

APPENDIX 2

A Sample CHARMM Run

An example of the use of CHARMM is presented in this appendix. In Sec. 1 below, we give the complete input file used for this example, and in the sections which follow, we present highly edited outputs that were generated by the commands. The outputs are given in the same order they appear in the command file. The run requires 48 min of CPU time on a VAX 11/780 with a floating point accelerator running in batch mode.

1. Sample Input File for CHARMM

The following is a sample input file for CHARMM that includes the information necessary to build an alanine dipeptide (which is illustrated in Fig. 3), explore the phi–psi
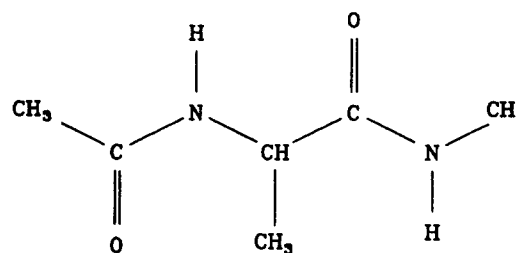
Figure 3. Structure of the alanine dipeptide, methyl alanyl acetamide.

surface with constrained minimizations, run 5 ps of dynamics trajectory and analyze the results of these calculations. Figure 4 is a contour plot of the phi–psi energy
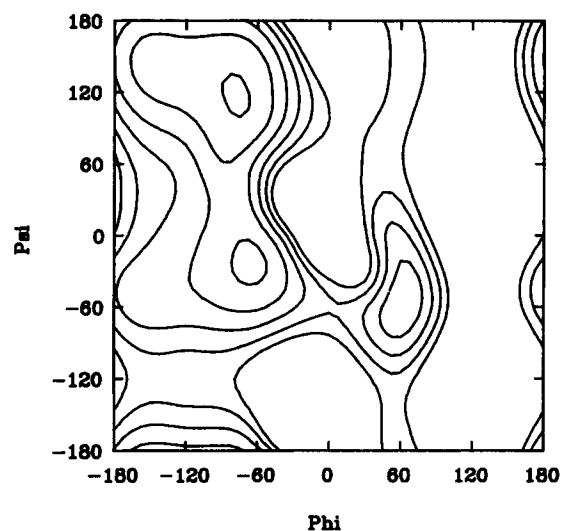
Figure 4. Contour plot of the potential energy for the alanine dipeptide as a function of the backbone dihedral angles phi and psi. The data were taken from the 12-point × 12-point example run presented in Appendix 2 and interpolated using cubic splines in both dimensions to a resolution of 50 points × 50 points. Contour levels are drawn at 1-kcal/mol intervals from the minimum-energy point in the raw data.

surface which was generated from the output of this run. Parameters are read free format from a file whose contents is given in Appendix 1.

```
* This input file builds N-methyl alanyl acetamide and uses
* minimization to generate an adiabatic phi-psi surface.
* Compares the molecule with and without electrostatic interactions.
*

READ RTF CARDS
* methyl alanyl acetamide
*
    17    1
TYPE HPRO
MASS     1 H     1.00800
MASS    11 C    12.01100
MASS    12 CH1E 13.01900
MASS    14 CH3E 15.03500
MASS    38 NH1  14.00670
MASS    51 O    15.99940

DECL  -C   -1
DECL  +N    1
DECL  +CA   2

RESI AMN    0.0
```

```
GROUP
! atom name type charge
ATOM     CL  CH3E  0.00
GROUP
ATOM     C   C    0.45
ATOM     O   O   -0.45
BOND   CL C     C +N        C O
ANGLE        CL C +N    CL C O     O C +N
DIHEDRAL   CL C +N +CA
IMPROPER    C CL +N O
ACCE    O C
IC  +N    CL    *C    O      0.0   0.0  180.0   0.0   0.0
IC  CL    C    +N   +CA      0.0   0.0  180.0   0.0   0.0
PATC  FIRS NONE LAST NONE

RESI ALA     0.0
GROUP
ATOM N    NH1   -0.35
ATOM H    H      0.25
ATOM CA   CH1E   0.10
GROUP
ATOM CB   CH3E   0.00
GROUP
ATOM C    C      0.45
ATOM O    O     -0.45
BOND    N CA     CA C      C +N      C O
BOND    N H      CA CB
ANGLE         -C N CA    N CA C     CA C +N
ANGLE         CA C O     O C +N     -C N H
ANGLE         H N CA     N CA CB    C CA CB
DIHEDRAL    -C N CA C       N CA C +N      CA C +N +CA
IMPROPER    N -C CA H        C CA +N O      CA N C CB
DONO H N -C CA
ACCE O C
IC  -C   CA   *N    H      0.0   0.0  180.0   0.0   0.0
IC  -C    N   CA    C      0.0   0.0  180.0   0.0   0.0
IC   N   CA    C   +N      0.0   0.0  180.0   0.0   0.0
IC  +N   CA   *C    O      0.0   0.0  180.0   0.0   0.0
IC  CA    C   +N  +CA      0.0   0.0  180.0   0.0   0.0
IC  CB    N  *CA    C      0.0   0.0  120.0   0.0   0.0

RESI CBX     0.0
GROUP
ATOM N    NH1   -0.35
ATOM H    H      0.25
ATOM CA   CH3E   0.10
BOND    N CA      N H
ANGLE         -C N CA    -C N H     H N CA
IMPROPER    N -C CA H
DONO H N -C CA
IC  -C   CA   *N    H      0.0   0.0  180.0   0.0   0.0
PATC  FIRS NONE LAST NONE

END
```

```
OPEN UNIT 2 READ UNFORM NAME  CD:PARAM4.MOD
READ PARAMETER FILE UNIT 2     ! Read the parameters listed
                               ! in appendix 1

READ SEQU CARD                 ! Read in the sequence.
* maa
*
3
AMN ALA CBX
GENERATE MAA SETU              ! Generate segment maa.

                               ! Now find positions for all atoms.
IC PARAMETERS                  ! Fill the IC table with values
                               ! from the parameters.
IC SEED 2 CA 2 C 3 N           ! Assign coordinate values for 3 atoms.
IC BUILD                       ! Construct remaining coordinates.
PRINT COORD
                               ! Construct nonbond and hydrogen bond
                               ! lists.  No hydrogen bonds will be
                               ! found using the very short cutoff.
NBONDS CONS EPS 2.5 CUTNB 99.0 WRNMXD 99.0
HBONDS CUTHB 0.01 CUTHBA 90.0
                               ! Now the molecule is set up, but
                               ! with a stereotypic geometry
                               ! so minimize it a bit to relieve
                               ! strains.
MINI ABNR STEP 0.01 NSTEP 100 NPRINT 20 INBFRQ 0 IHBFRQ 0 STRICT 0.05
                               ! Begin the loop to calculate the
                               ! energy contour plot in figure 4.  Each cycle
                               ! in the loop, the dihedrals are constrained
                               ! to a new set of angles and the
                               ! structure obtained from the previous
                               ! set is reminimized.
SET 1 -180.0                   !    Parameter 1 is used to store phi.
SET 2  180.0                   !    Parameter 2 is used to store psi.
LABEL LOOP
                               ! Add new Phi-Psi constraints based
                               ! on current loop variables.
CONS DIHE MAA 1 C : MAA 2 N : MAA 2 CA : MAA 2 C -
     FORCE 200.0 MIN @1
CONS DIHE MAA 2 N : MAA 2 CA : MAA 2 C : MAA 3 N -
     FORCE 200.0 MIN @2
                               ! Relax the structure to the desired
                               ! point on the Phi-Psi plot. The first
                               ! pass through this loop generated the
                               ! output in section A2.2.
MINI ABNR STEP 0.01 NSTEP 100 NPRINT 20 INBFRQ 0 IHBFRQ 0 STRICT 0.05
                               ! The final minimized energy is read
                               ! by a plotting program.
CONS CLDH                      ! Remove these dihedral constraints.
INCR 2 BY -30.                 ! And execute the loops.
```

```
IF 2 GT -170. GOTO LOOP
SET 2 180.
INCR 1 BY 30.
IF 1 LT 170. GOTO LOOP

! ——— The following input generates output for section A2.3. ———

                               ! Put molecule at a conformation near
                               ! the Phi-Psi minimum by rebuilding.
COOR INIT                      ! Reinitialize all coordinates
IC EDIT                        ! Modify the IC table for Phi-Psi angles.
     DIHE 1 C  2 N  2 CA  2 C  -75.0
     DIHE 2 N  2 CA 2 C  3 N   120.0
     END
IC SEED 2 CA 2 C 3 N           ! Again, assign coordinates for 3 atoms.
IC BUILD                       ! Construct the remaining coordinates.

                               ! Minimize to find the global minimum.
MINI ABNR TOLGRD 0.01 STEP 0.01 NSTEP 200 MINDIM 8 EIGRNG 0.001 -
NPRINT 20 INBFRQ 0 IHBFRQ 0 STRICT 0.05
                               ! Do 2 steps of Newton-Raphson
MINI NRAP NSTEP 2              ! minimization to zero the gradient.

IC FILL                        ! Fill and print internal coordinate
PRINT IC                       ! table for this conformation.

VIBRAN                         ! Vibrational analysis section.
     DIAG                      ! Diagonalize to get normal modes.
                               ! Print out information on the 2 lowest
                               ! internal modes.
     PRINT NORM MODE 7 THRU 8 INTD TEMP 300.0
     END

! ——— The following input generates output for section A2.4. ———

COOR COMPARE COPY              ! Copy coordinates to the comparison set
NBOND VDWO CUTNB 99.0          ! Remove the electrostatic energy
                               ! and forces and relax.
MINI ABNR STEP 0.01 NSTEP 200 NPRINT 20 INBFRQ 0 IHBFRQ 0 STRICT 0.05

IC FILL                        ! Print IC tables for this new conformation
PRINT IC                       ! and also its difference
IC DIFF                        ! from the previous minimized structure.

ANAL                           ! Compare the minima found with and
                               ! without electrostatics.
     COMPARE COORMATCH COOR COMPARE $ -
          NBOND COPY $ -
          HBOND COPY $ -
     BUILD DIFF ATOM R $ ENERGY $ EVDW $
     ADD STATS RMS $ COLLECT ALLTAG $ PLACE STRUCTURE $$
     PRINT TABLE COLUMN
     SELECT ALLTAGS $ PROPERTY R $
     SELECT TWO ALLTAGS $ PROPERTY ENERGY $


     2DPLOT VERTICAL -0.3 0.3 36 HORIZ 0. 0.6 60
     END

! ——— The following input generates output for section A2.5. ———

                               ! Now set up a short dynamics trajectory,
                               ! and compare the average structure
                               ! to the minimized structure stored
                               ! in the comparison coordinate set.
COOR COPY                      ! Get the coordinates of the minimum again.
NBONDS CONS EPS 2.5 CUTNB 99. WRNMXD 99. ! and turn on electrostatics.
SHAKE BONH                     ! Constrain the bonds with hydrogens.
                               ! Run 5 picoseconds of dynamics using
                               ! a time step of .001 ps.
DYNA VERLET TIMESTP 0.001 NSTEP 5000 -
     FIRSTT 300.0 FINALT 300.0 TWINDL 30.0 TWINDH 30.0 -
     INBFRQ 0 IHBFRQ 0 IEQFRQ 100 NPRINT 20 -
     IUNCRD 30 IUNVEL 31 NSAVC 4 NSAVV 4

! ——— The following input generates output for section A2.6. ———

ANAL                           ! Analyze the trajectory to get
     ACCUM COOR FIRSTU 30 NUNIT 1    ! the average coordinates.
     WRITE UNIT 32 AVECOOR TITLE AVERAGED COORDINATES $
     BUILD ATOM DYN R M2 $     ! Print out the atomic fluctuations.
     PRINT TABLE PRETTY
     END
READ COOR FILE UNIT 32
ANAL                           ! Compare the average and minimized
     COMPARE COOR COMPARE $    ! coordinates.
     BUILD DIFF ATOM R $ ENERGY $
     PRINT TABLE PRETTY PROPERTY R $
     PRINT TABLE PRETTY PROPERTY ENERGY $
     END
```

2. Minimization Output from the First Point of the Phi–Psi Surface

An example of the information generated by CHARMM during a minimization is given below. In this run, output is generated every twentieth step. The first column shows that the minimization ran 80 cycles with 82 energy evaluations. The next columns provide some information on the minimization routine. Newton–Raphson steps were successful at each point listed. The algorithm was using information from the last five steps to compute new steps, and the second derivative matrix was maintaining a full

rank of five. The total energy of the system decreased from 201 to −1.55 kcal/mol and the gradient decreased from 109 to 0.025 kcal/mol-Å. The step size decreased to 0.0002 Å/coordinate. Individual energy terms are listed in the right-hand columns.

Cycle# Eval.#	Energy	Delta E Rank Dim.	Gradient Step	Bond Angle	Dihedral Improper	vdW Elec.	Hbond Cons.
0	201.6573	−201.6573	109.1010	0.105	0.487	−0.018	0.000
1		INIT 0 0	0.0000	0.656	0.015	−5.268	205.68
20	0.2914	0.6226	6.7664	0.778	0.200	2.703	0.000
21		NR 5 5	0.0542	1.558	0.255	−5.352	0.150
40	−1.5282	0.0069	0.9196	0.397	0.036	2.308	0.000
42		NR 5 5	0.0063	0.988	0.048	−5.351	0.045
60	−1.5496	0.0000	0.2275	0.362	0.028	2.325	0.000
62		NR 5 5	0.0009	0.985	0.042	−5.340	0.048
80	−1.5509	0.0000	0.0251	0.365	0.028	2.319	0.000
82		NR 5 5	0.0002	0.990	0.041	−5.343	0.048

3. Vibrational Analysis of the Minimum

The output generated by CHARMM in analyzing the vibrational properties of the minimum on the phi–psi surface is given below. The constructed coordinates were minimized first using ABNR to get close to the minimum and then using two steps of Newton–Raphson yielding a final gradient of less than 5.0×10^{-5} kcal/mol-Å-coordinate. The internal coordinates at the minimum are listed (see Sec. IV A for the interpretation of the columns). The next section is a vibrational analysis. The energy is evaluated with its second derivatives and the full second-derivative matrix is diagonalized. The 36 normal mode frequencies are listed (including 6 rotation and translation modes at near zero frequencies). The final section lists more details about the two lowest internal modes including the mixing with rotation/translational modes, and the internal coordinate displacements that would occur in exciting each mode to 300° K. It is easily seen that the lowest internal mode of the system is primarily a rotation of the psi torsion, and the second is primarily rotation of phi.

Internal Coordinate Listing of the Minimum Energy Structure

N	I	J	K	L	R(I(J/K))	T(I(JK/KJ))	PHI	T(JKL)	R(KL)
1	2 N	1 CL	1 *C	1 O	1.3349	116.46	−179.88	121.21	1.2313
2	1 CL	1 C	2 N	2 CA	1.5198	116.46	179.62	121.80	1.4567
3	1 C	2 CA	2 *N	2 H	1.3349	121.80	179.06	119.41	0.9789
4	1 C	2 N	2 CA	2 C	1.3349	121.80	−77.78	113.91	1.5279
5	2 N	2 CA	2 C	3 N	1.4567	113.91	120.13	114.61	1.3368
6	3 N	2 CA	2 *C	2 O	1.3368	114.61	−178.48	121.89	1.2346
7	2 CA	2 C	3 N	3 CA	1.5279	114.61	178.86	122.87	1.4951
8	2 CB	2 N	2 *CA	2 C	1.5249	109.51	121.91	113.91	1.5279
9	2 C	3 CA	3 *N	3 H	1.3368	122.87	−178.49	118.94	0.9787

Vibrational Analysis of the −78,120 Minimum

Energy	Gradient	Bond angle	Dihedral improper	vdW elec.	H bond cons.
−4.5483	0.0000	0.109	0.128	−0.252	0.000
		0.746	0.031	−5.310	0.000

Normal Mode Frequencies (cm⁻¹)

1	−0.177	2	−0.104	3	−0.091	4	−0.080	5	0.109
6	0.120	7	21.834	8	41.697	9	89.516	10	169.796
11	189.588	12	228.694	13	241.877	14	303.814	15	389.458
16	400.395	17	459.610	18	585.938	19	597.810	20	628.604
21	703.685	22	820.331	23	832.500	24	878.168	25	897.909
26	1026.217	27	1119.658	28	1188.095	29	1331.274	30	1360.729
31	1599.442	32	1630.791	33	1660.991	34	1667.502	35	3194.041
36	3195.533								

Listing of the Two Lowest Internal Modes

```
VIBRATION MODE   7  FREQUENCY=   21.834244  TRANS ROT %=   0.000
EIGENVALUE=    0.040428  ENERGY DERIVATIVE=     0.082
Temperature = 300.0    MASS    NORM  STEP=    2.715215

Projection onto Internal Coordinates of the System:
```

N	I	J	K	L	R(I(J/K))	T(I(JK/KJ))	PHI	T(JKL)	R(KL)
1	2 N	1 CL	1 *C	1 O	0.0034	−0.11	0.14	0.02	0.0032
2	1 CL	1 C	2 N	2 CA	0.0045	−0.11	−0.18	0.17	0.0043
3	1 C	2 CA	2 *N	2 H	0.0034	0.17	−0.28	−0.07	0.0024
4	1 C	2 N	2 CA	2 C	0.0034	0.17	−2.97.	0.18	−0.0011
5	2 N	2 CA	2 C	3 N	0.0043	0.18	21.39	−0.23	0.0487
6	3 N	2 CA	2 *C	2 O	0.0487	−0.23	−1.32	−1.57	0.0359
7	2 CA	2 C	3 N	3 CA	−0.0011	−0.23	−0.11	−0.78	0.0023
8	2 CB	2 N	2 *CA	2 C	0.0017	−0.06	−1.16	0.18	−0.0011
9	2 C	3 CA	3 *N	3 H	0.0487	−0.78	−0.75	−1.54	0.0334

```
VIBRATION MODE   8  FREQUENCY=   41.697554  TRANS ROT %=   0.000
EIGENVALUE=    0.147443  ENERGY DERIVATIVE=     0.023
Temperature = 300.0    MASS    NORM  STEP=    1.421778
```

Projection onto Internal Coordinates of the System

N	I	J	K	L	R(I(J/K))	T(I(JK/KJ))	PHI	T(JKL)	R(KL)
1	2 N	1 CL	1 *C	1 O	0.0088	−0.73	−0.27	0.18	0.0189
2	1 CL	1 C	2 N	2 CA	0.0036	−0.73	0.72	−0.41	0.0023
3	1 C	2 CA	2 *N	2 H	0.0088	−0.41	0.07	0.01	0.0123
4	1 C	2 N	2 CA	2 C	0.0088	−0.41	11.64	0.27	0.0022
5	2 N	2 CA	2 C	3 N	0.0023	0.27	0.79	−0.25	0.0008
6	3 N	2 CA	2 *C	2 O	0.0008	−0.25	0.12	0.13	0.0033
7	2 CA	2 C	3 N	3 CA	0.0022	−0.25	0.14	−0.12	0.0010
8	2 CB	2 N	2 *CA	2 C	0.0029	0.07	0.00	0.27	0.0022
9	2 C	3 CA	3 *N	3 H	0.0008	−0.12	0.20	0.05	0.0024

4. Comparison of Structures Minimized with and without Electrostatistics

This section compares the minimum-energy structures with and without the electrostatic term in the potential. First, the internal coordinates are compared showing that even in the neutral dipeptide, significant perturbations occur when electrostatistics are neglected. The nature of these distortions is analyzed by preparing a scatter plot of the change in atom energy against the change in position. The energetic changes are seen to be not well correlated to the coordinate displacements.

Differences in Internal Coordinates

N	I	J	K	L	R(I(J/K))	T(I(JK/KJ))	PHI	T(JKL)	R(KL)
1	2 N	1 CL	1 *C	1 O	−0.0021	0.56	0.14	0.14	−0.0010
2	1 CL	1 C	2 N	2 CA	0.0	0.56	−0.21	−0.39	0.0009
3	1 C	2 CA	2 *N	2 H	−0.0021	−0.39	−0.77	0.44	−0.0012
4	1 C	2 N	2 CA	2 C	−0.0021	−0.39	−3.23	0.53	0.0032
5	2 N	2 CA	2 C	3 N	0.0009	0.53	−16.45	0.05	−0.0016
6	3 N	2 CA	2 *C	2 O	−0.0016	0.05	0.68	0.30	−0.0003
7	2 CA	2 C	3 N	3 CA	0.0032	0.05	−1.14	0.00	−0.0003
8	2 CB	2 N	2 *CA	2 C	0.0014	0.13	1.17	0.53	0.0032
9	2 C	3 CA	3 *N	3 H	−0.0016	0.00	1.51	0.28	−0.0013

Coordinate Matching

```
SUM OF SQUARE DIFFERENCES BETWEEN EQUIVALENT ATOMS IS  0.4725122
ROOT MEAN SQUARE OF DIFFERENCE IS   0.1984339

DIFFERENCES OF ATOM PROPERTIES
```

SEG RES ATOM	R	E	EVDW	SEG RES ATOM	R	E	EVDW
MAA AMN 1 CL	0.171	−0.003	−0.003	MAA ALA 2 CB	0.139	−0.119	−0.118
MAA AMN 1 C	0.102	0.097	0.031	MAA ALA 2 C	0.123	0.049	0.030
MAA AMN 1 O	0.217	−0.045	−0.047	MAA ALA 2 O	0.371	−0.166	−0.166
MAA ALA 2 N	0.033	0.086	0.061	MAA CBX 3 N	0.149	0.071	0.078
MAA ALA 2 H	0.123	0.004	0.005	MAA CBX 3 H	0.389	0.005	0.005
MAA ALA 2 CA	0.092	−0.175	−0.071	MAA CBX 3 CA	0.123	−0.027	−0.028
MAA STAT RMS	0.198	0.091	0.071				

Analysis of the Correlation between Changes in Position and Energy

```
STATISTICS ON 13 POINTS IN THE MATCHED SELECTIONS:
        MEAN       VARIANCE
X  0.1716        0.0996
Y −0.0103        0.0903
LEAST SQUARES SCALING FACTOR = −0.1135
CORRELATION COEFFICIENT = −0.3011
LINEAR REGRESSION M= −0.2731    B= 0.3662E-01
```

```
 0.283 I  —I——I——I——I——I——I——I——I——I——
 0.267 I                                      I
 0.250 I                                      I
 0.233 I                                      I
 0.217 I                                      I
 0.200 I                                      I
 0.183 I                                      I
 0.167 I                                      I
 0.150 I                                      I
 0.133 I                                      I
 0.117 I                                      I
 0.100 I                                      I
 0.083 I   *    *    *                        I
 0.067 I            *                         I
 0.050 I                                      I
 0.033 I       *                              I
 0.017 I                                      I
 0.000 I      *              *                I
-0.017 I                                      I
-0.033 I      *   *                           I
-0.050 I           *                          I
-0.067 I                                      I
-0.083 I                                      I
-0.100 I                                      I
-0.117 I                                      I
-0.133 I        *                             I
-0.150 I                                      I
-0.167 I                  *                   I
-0.183 I    *                                 I
-0.200 I                                      I
-0.217 I                                      I
-0.233 I                                      I
-0.250 I                                      I
-0.267 I                                      I
-0.283 I                                      I
-0.300 I                                      I
       —I——I——I——I——I——I——I——I——I——
       0.030 0.100 0.170 0.240 0.310 0.380 0.450 0.520
              Change in position (angstroms)
```
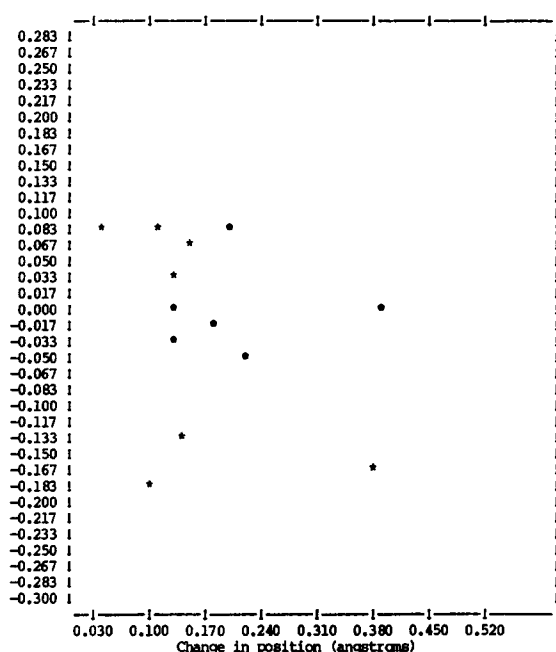
5. Output from Dynamics

An example of the output generated by CHARMM in doing dynamics is now presented. The run begins by assigning the velocities, zeroing out rotation and translation components, and beginning integration using the Verlet algorithm. The first 100 steps (0.1 ps) of the run are shown along with a summary of the averages and fluctuations in the various energy components. No kinetic energy is listed for the first step because the integrator is being initialized.

TIME (ps) (akma)	TOTAL E BOND CONSTRAINT	TOTAL KE BOND ANG VDW	TOTAL PE DIHEDRAL ELEC	Temperature IMPROPER H-BOND
0 0	0.00000			-4.550	434.764
	0.00000	0.108	0.746	0.128	0.031
		0.000	-0.252	-5.310	0.000
20 2	0.02000	7.26718	6.573	0.694	236.265
	0.409	1.744	4.088	0.194	0.176
		0.000	-0.538	-4.971	0.000
40 2	0.04000	7.24285	7.840	-0.597	281.803
	0.818	0.423	4.493	0.207	0.081
		0.000	-1.052	-4.749	0.000
60 2	0.06000	7.25134	7.795	-0.544	280.172
	1.227	0.999	3.270	0.116	0.490
		0.000	-0.778	-4.640	0.000
80 2	0.08000	7.26530	6.005	1.260	215.847
	1.636	2.379	2.659	0.223	0.253
		0.000	0.223	-4.476	0.000
100 2	0.10000	7.23765	7.484	-0.246	268.999
	2.045	0.479	1.890	0.010	0.461
		0.000	1.649	-4.735	0.000

```
AVERAGES FOR THE LAST 100 STEPS:
100  2    0.10000    7.26055    7.015    0.246    252.123
          2.045      1.524      3.325    0.157    0.368
                     0.000     -0.344   -4.783    0.000

RMS FLUCTUATIONS FOR 100 STEPS:
100  2    0.10000    0.01095    1.081    1.088    252.123
          2.045      0.636      1.158    0.068    0.246
                     0.000      0.664    0.277    0.000

DRIFT/STEP (LAST-TOTAL) : -3.52858E-05   VELOCITIES HAVE BEEN SCALED -  1.1747
E AT STEP 0             :  7.2623        OLD TEMPERATURE            -  268.999
CORR. COEFFICIENT       : -9.29934E-02   NEW TEMPERATURE            -  371.163
```

6. Analysis of the Dynamical Averaged Coordinates

We now present a comparison of the dynamics average coordinates to the minimized coordinates. First, a list of the atom displacements is printed followed by a list of the difference in atom energies. Note that the dynamics average structure energy are consistently higher than the minimized structure.

```
1250 COORD RECORDS READ FROM 1 UNITS STARTING WITH UNIT  30
RUNNING FROM STEP      4 TO      5000 SKIPPING   4 STEPS BETWEEN RECORDS
Time step was  2.0454811E-02 AKMA time units.

LISTING OF SECOND MOMENT DISTANCE DISPLACEMENTS

RESIDUE   CL    C     O     N     H     CA    CB

AMN 1   0.397 0.150 0.420
ALA 2         0.149 0.483 0.211 0.481 0.183 0.258
CBX 3                     0.301 0.615 0.388

HOMOLOGY CALLED - COST OF TRANSFORMING SEQUENCES IS      0
SUM OF SQUARE DIFFERENCES BETWEEN EQUIVALENT ATOMS IS  0.2066818
ROOT MEAN SQUARE OF DIFFERENCE IS  0.1312383
IN THE PAIRING OF PSF'S, 1 SEGMENTS WERE MATCHED,
   3 RESIDUES WERE MATCHED, AND    12 ATOMS WERE MATCHED.

DIFFERENCES OF COORDINATES

RESIDUE   CL    C     O     N     H     CA    CB

AMN 1   0.138 0.067 0.177
ALA 2         0.053 0.137 0.125 0.251 0.071 0.122
CBX 3                     0.056 0.090 0.142

DIFFERENCES OF TOTAL ATOM ENERGIES

RESIDUE   CL    C     O     N     H     CA    CB

AMN 1   0.343 1.166 0.774
ALA 2         1.305 1.151 0.852 0.359 0.212 0.031
CBX 3                     1.816 0.906 0.568
```

APPENDIX 3

Execution Time of Representative Tasks

Task	Arg-Pro Dipeptide	PTI
Generation	5.22 s	21.2 s
Energy evaluation	0.1 s	4.25 s
Normal-mode analysis	33.9 s	$\approx$24 h
Dynamics 10 ps (10^4 steps)	21.5 min	$\approx$13 h
Analysis of a 10-ps trajectory (2000 time points) computing mean position and standard deviation about the mean	41.0 s	11.8 min

The CPU times listed for the various tasks were measured for CHARMM version 16 running on a VAX 11/780 with a floating point accelerator and version 2.5 of VAX/VMS. The explicit hydrogen topology file was used to generate the two structures. The structure of PTI (bovine pancreatic trypsin inhibitor)[61] contains 580 atoms, including hydrogen bonding hydrogens. The dipeptide, Arg–Pro, which contains 27 atoms was constructed using the first two residues of PTI. In both cases, the nonbonded cutoff distance was 7.5 Å and the hydrogen bond cutoffs were 4.5 Å and 90°. Approximately 24,000 nonbonded pairs were present in PTI and 200 in Arg–Pro. The electrostatic potential was computed using the distance-dependent dielectric constant.

References

1. J. B. Hendrickson, *J. Am. Chem. Soc.*, **83**, 4537 (1961).
2. M. Levitt and S. Lifson, *J. Mol. Biol.*, **46**, 269 (1969).
3. H. A. Scheraga, *Chem. Rev.*, **71**, 195 (1971).
4. E. M. Engler, J. D. Andose, and P. v. R. Schleyer, *J. Am. Chem. Soc.*, **95**, 8005 (1973).
5. J. A. McCammon and M. Karplus, *Annu. Rev. Phys. Chem.*, **31**, 29 (1980).
6. M. Karplus, B. R. Gelin, and J. A. McCammon, *Biophys. J.*, **32**, 628 (1980).
7. J. N. Kushick and M. Karplus, *Macromolecules*, **14**, 325 (1981).
8. W. K. Olson and J. L. Sussman, *J. Am. Chem. Soc.*,

104, 270 (1982).

9. K. B. Wiberg, *J. Am. Chem. Soc.*, 87, 1070 (1965).

10. A. Warshel, M. Levitt, and S. Lifson, *J. Mol. Spectrosc.*, 33, 84 (1970).

11. N. L. Allinger, *Adv. Phys. Org. Chem.*, 13, 1 (1976).

12. T. L. Starr and D. E. Williams, *Acta Crystallogr. Sect. A*, 33, 771 (1977).

13. G. D. Smith, V. Z. Pletnev, W. L. Duax, T. M. Balasubramanian, H. E. Bossard, E. W. Czerwinski, N. E. Kendrick, F. Mathews, and G. R. Marshall, *J. Am. Chem. Soc.*, 103, 1493 (1981).

14. D. E. Williams, *Chem. Phys.*, 55, 163, (1981).

15. F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, 60, 1545 (1974).

16. P. Lykos, Ed. *Computer Modeling of Matter*, Am. Chem. Soc. Symp. Ser. 86, Am. Chem. Soc., Washington, DC, 1978.

17. P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature*, 282, 459 (1979).

18. P. J. Rossky and M. Karplus, *J. Am. Chem. Soc.*, 101, 1913 (1979).

19. P. J. Rossky and M. Karplus, *Biopolymers*, 18, 825 (1979).

20. C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.*, 71, 2975 (1979).

21. S. Swaminathan and D. L. Beveridge, *J. Am. Chem. Soc.*, 101, 5832 (1979).

22. M. Bishop, D. Ceperley, H. L. Frisch, and M. H. Kalos, *J. Chem. Phys.*, 72, 3228 (1980).

23. W. L. Jorgensen, *J. Am. Chem. Soc.*, 103, 335 (1981).

24. B. R. Gelin and M. Karplus, *Proc. Nat. Acad. Sci. USA*, 72, 2002 (1975).

25. J. A. McCammon, B. R. Gelin, M. Karplus, and P. G. Wolynes, *Nature*, 262, 325 (1976).

26. B. Honig, A. Ray, and C. Levinthal, *Proc. Nat. Acad. Sci. USA*, 73, 1974 (1976).

27. J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature*, 267, 585 (1977).

28. A. T. Hagler and J. Moult, *Nature*, 272, 222 (1978).

29. W. F. van Gunsteren and M. Karplus, *Nature*, 293, 677 (1981).

30. P. A. Kollman, P. K. Weiner, and A. Dearing, *Biopolymers*, 20, 2583 (1981).

31. E. Clementi and G. Corongiu, *Biopolymers*, 21, 763 (1982).

32. W. K. Olson, *J. Am. Chem. Soc.*, 104, 278 (1982).

33. W. F. van Gunsteren and M. Karplus, *Biochemistry*, 21, 2259 (1982).

34. A. Warshel and M. Levitt, "QCFF/PI," *Quantum Chemistry Program Exchange* 11, 247 (1974).

35. M. J. Browman, L. M. Carruthers, K. L. Kashuba, F. A. Momany, M. S. Pottle, S. P. Rosen, S. M. Rumsey, G. F. Endres, and H. A. Scheraga, "ECEPP," *Quantum Chemistry Program Exchange* 11, 286 (1975).

36. B. R. Gelin, Ph.D. thesis, Harvard University, Cambridge, MA, 1976.

37. N. L. Allinger, "MMI/MMPI," *Quantum Chemistry Program Exchange* 11, 318 (1976).

38. E. Huler, R. Sharon, and A. Warshel, "MCA," *Quantum Chemistry Program Exchange*, 11, 325 (1977).

39. R. Potenzone, Jr., E. Cavicchi, H. J. R. Weintraub, and A. J. Hopfinger, *Comput. Chem.*, 1, 187 (1977).

40. M. J. Browman, A. W. Burgess, L. G. Dunfield, S. M. Rumsey, G. F. Endres, and H. A. Scheraga, "UNICEPP," *Quantum Chemistry Program Exchange*, 11, 361 (1978).

41. J. Snir, R. A. Nemenoff, and H. A. Scheraga, *J. Phys. Chem.*, 82, 2497 (1978).

42. D. E. Williams, "PCK5/PCK6," *Quantum Chemistry Program Exchange*, 11, 373 (1979).

43. N. L. Allinger and Y. A. Yuh, "MM2," *Quantum Chemistry Program Exchange* 13, 395 (1981).

44. P. K. Weiner and P. A. Kollman, *J. Comput. Chem.*, 2, 287 (1981).

45. B. R. Gelin and M. Karplus, *Biochemistry*, 18, 1256 (1979).

46. L. G. Dunfield, W. Burgess, and H. A. Scheraga, *J. Chem. Phys.*, 82, 2609 (1978).

47. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations*, Dover, New York, 1955.

48. W. L. Jorgensen, *J. Am. Chem. Soc.*, 103, 341 (1981).

49. W. L. Jorgensen, *J. Am. Chem. Soc.*, 103, 345 (1981).

50. S. Lifson and A. Warshel, *J. Chem. Phys.*, 49, 5116 (1968).

51. A. Warshel and S. Lifson, *J. Chem. Phys.*, 53, 582 (1970).

52. A. Warshel and M. Levitt, *J. Mol. Biol.*, 103, 227 (1976).

53. W. Reiher and M. Karplus, in preparation.

54. J. D. Bernal and R. H. Fowler, *J. Chem. Phys.*, 1, 515 (1933).

55. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, 34, 1311 (1977).

56. B. R. Brooks and M. Karplus, in preparation.

57. B. T. Willis and A. W. Pryor, *Thermal Vibrations in Crystallography*, Cambridge University Press, London, 1975.

58. R. Fletcher and C. M. Reeves, *Comput. J.*, 7, 149 (1964).

59. D. J. States and M. Karplus, in preparation.

60. D. J. States and M. Karplus, in preparation.

61. J. Deisenhofer and W. Steigemann, *Acta Crystallogr. Sect. B*, 31, 238 (1975).

62. B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, 27, 1208 (1957).

63. L. Verlet, *Phys. Rev.*, 159, 98 (1967).

64. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, England Cliff, NJ, 1971.

65. J. L. Lebowitz, J. K. Percus, and L. Verlet, *Phys. Rev.*, 153, 250 (1967).

66. R. C. Ladner, *Proc. Soc. Physico-Optical Instr. Engin.* 367, 13 (1982).

67. J. H. Wilkinson, *Comput. J.*, 1, 90 (1958).

68. J. H. Wilkinson, *Comput. J.*, 3, 23 (1960).

69. J. M. Ortega, *J. Assoc. Comput. Mach.*, 7, 260 (1960).

70. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, London 1965.

71. H. Michels and F. Prosser, *Quantum Chemistry Program Exchange*, 11, 62 (1965).

72. B. R. Brooks and M. Karplus, in preparation.

73. B. R. Brooks and M. Karplus, in preparation.

74. E. F. Codd, *Commun. Assoc. Comput. Mach.*, 13, 377 (1970).

75. D. C. Tsichritzis and F. H. Lochovsky, *Comput. Surveys*, 8, 105 (1976).

76. B. Lee and F. M. Richards, *J. Mol. Biol.*, 55, 379 (1971).

77. G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, 7, 95 (1963).

78. T. J. O'Donnell, "GRAMPS, GRAphics for the MPS, A Command Language Interpreter for Producing Computer Graphics on the Evans & Sutherland Multi-Picture System," National Resource for

Computation in Chemistry, Berkeley, CA, 1981.

79. R. A. Wagner and M. J. Fischer, *J. Assoc. Comput. Mach.*, **21**, 168 (1974).

80. S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.*, **48**, 443 (1970).

81. R. Zwanzig, *Ann. Rev. Phys. Chem.*, **16**, 67 (1965).

82. B. J. Berne and R. Pecora, *Dynamic Light Scattering*, Wiley, New York, 1976.

83. J. M. Cooley and J. W. Tukey, *Math. Comput.*, **19**, 297 (1965).

84. R. Zwanzig and N. K. Ailawadi, *Phys. Rev.*, **182**, 280 (1969).

85. S. Y. Lee and M. Karplus, in preparation.

86. R. M. Levy, J. A. McCammon, and M. Karplus, *Biophys. J.*, **32**, 628 (1980).

87. T. Beyer, "FLECS User's Manual" (1975), Computing Center, University of Oregon, Eugene, OR 97403.

88. E. W. Dijkstra, *Comm. Assoc. Comput. Mach.*, **11**, 147, (1968).

89. J. D. Ullman, *Fundamental Concepts of Programming Systems*, Addison Wesley, Reading, MA, 1976.

90. D. A. Case and M. Karplus, *J. Mol. Biol.*, **132**, 343 (1979).

91. N. Alderbing, S. S. Chan, L. Eisenstein, H. Fraunfelder, D. Good, I. C. Gunsalus, T. M. Nordland, M. F. Perutz, A. H. Reynolds, and L. B. Sorensen, *Biochemistry*, **17**, 43 (1979).

92. S. W. Englander, N. W. Downer, and H. Teitlebaum, *Ann. Rev. Biochem.*, **41**, 903 (1972).

93. C. K. Woodward and D. Hilton, *Ann. Rev. Biophys. Bioeng.*, **8**, 99 (1979).

94. G. Weber, *Adv. Protein Chem.*, **29**, 1 (1978).

95. G. H. Snyder, R. Rowan, S. Karplus, and B. D. Sykes, *Biochemistry*, **14**, 3765 (1975).

96. J. A. McCammon and M. Karplus, *Biopolymers*, **19**, 1375 (1980).

97. M. Levitt, *J. Mol. Biol.*, **145**, 251 (1981).

98. D. J. States and M. Karplus, in preparation.

99. J. C. Slater and J. G. Kirkwood, *Phys. Rev.*, **37**, 682 (1931).

100. K. S. Pitzer, in *Advances in Chemical Physics*, I. Prigogine, Ed., Interscience, New York, Vol. I, p. 59.