Journal of Biomolecular NMR, 8 (1996) 136–146
ESCOM

J-Bio NMR 380

The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules

P. Luginbühl, P. Güntert, M. Billeter and K. Wüthrich*

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

Received 6 May 1996
Accepted 5 July 1996

Keywords: Molecular dynamics simulation; Structure refinement

Summary

A new program for molecular dynamics (MD) simulation and energy refinement of biological macromolecules, OPAL, is introduced. Combined with the supporting program TRAJEC for the analysis of MD trajectories, OPAL affords high efficiency and flexibility for work with different force fields, and offers a user-friendly interface and extensive trajectory analysis capabilities. Salient features are computational speeds of up to 1.5 GFlops on vector supercomputers such as the NEC SX-3, ellipsoidal boundaries to reduce the system size for studies in explicit solvents, and natural treatment of the hydrostatic pressure. Practical applications of OPAL are illustrated with MD simulations of pure water, energy minimization of the NMR structure of the mixed disulfide of a mutant *E. coli* glutaredoxin with glutathione in different solvent models, and MD simulations of a small protein, pheromone E*r*-2, using either instantaneous or time-averaged NMR restraints, or no restraints.

Introduction

Molecular dynamics simulations (reviewed in, for example, Allen and Tildesley, 1987; Van Gunsteren and Berendsen, 1990) have become a widely used tool in computer-based investigations of macromolecules, such as studies of the internal dynamics of proteins (McCammon and Harvey, 1987; Brooks III et al., 1988), the interaction of macromolecules and ligands (recently reviewed in, for example, Kollman, 1994), the hydration of proteins and DNA (e.g., Brunne et al., 1993; Billeter et al., 1996) and, along a different line, the calculation of three-dimensional structures from experimental data collected either by X-ray diffraction or by NMR spectroscopy (Brünger and Nilges, 1993). This paper describes the design and implementation of a new program, OPAL, which has the following principal features: (i) the source code is written independently of the force field used to describe the physical energy and possibly additional terms enforcing experimental data; the force field is read from a single library and can readily be modified or completely exchanged; (ii) The code is optimized for vectorizing supercomputers, so that OPAL achieves an average performance of 1.5 GFlops on a NEC SX-3 supercomputer; a version for parallel machines is in preparation; (iii) the implementation of a flexible choice of the surface boundary shape allows optimization of systems for studies in explicit solvent with a minimal number of solvent molecules; (iv) a user-friendly interface is based on a powerful command language for interactive or batch mode operation, which was previously used in the programs PROSA (Güntert et al., 1992), GARANT (Bartels et al., 1996) and the newest version of DIANA (Güntert et al., 1991; to be released). In OPAL, this user interface ensures easy handling of groups of multiple protein conformers, and of the types of input conformational constraints which typically occur in protein structure determinations by NMR spectroscopy (Wüthrich, 1986,1995). Macros can be designed for effi-

*To whom correspondence should be addressed.

Abbreviations: D, diffusion constant in cm^2/s; E*r*-2, pheromone 2 from *Euplotes raikovi*; GFlop, one billion floating point operations per second; Grx(C14S)-SG, mixed disulfide between a mutant *E. coli* glutaredoxin, with Cys^{14} replaced by Ser, and glutathione; MD, molecular dynamics; NOE, nuclear Overhauser enhancement; rmsd, root-mean-square deviation; ρ, density in g/cm^3.

Command	Explanation
amber94	read force field of Cornell et al. (1995) from library
read dg protein-DNA	read input structure from file *protein-DNA.dg*
build	build structure and initialize force field
inertia	transform coordinates to principal axes of inertia
velocity 293.0 force	set velocities based on forces at temperature 293 K
immerse water.asc	solvate structure in pre-equilibrated water bath read from file *water.asc*; get values for system variables (e.g., reference temperature and pressure)
check	ensure that all atoms are within the surface boundary
show	show all system variables (e.g., *dielectric*, *t_ref*)
groupenergy	report intramolecular and intermolecular energies
temperature	report temperature
pressure	report pressure
do i 1 100	loop for 100 MD runs of 1 ps each
md 500 10 0.002	MD run: 500 steps of 2 fs, information every 10 steps
mdstat	report average values of energy, temperature, pressure, volume
write asc snapshot$i(I3.3)	save snapshots in files numbered by loop variable *i*
end do	end of loop

Fig. 1. Macro file for OPAL used to perform a standard MD simulation. Each line contains one OPAL command on the left, and an explanation of this command on the right. This set of commands performs an MD simulation of 100 ps using a start structure read from the file *protein-DNA.dg*, where the start structure is first immersed into a preequilibrated water bath that is read from the file *water.asc*. 100 MD runs of 1 ps length each, with steps of 2 fs, are successively started, and snapshots are saved in intervals of 500 steps. Additional output, in intervals of 10 steps, reports intermediate values for the energy, the temperature, the pressure and the volume.

cient variation of the molecular system under investigation, such as changes of the temperature, or of the relative weights of different experimental constraints.

The present description of OPAL is complemented by several applications. As a test of the water model used and of the novel boundary conditions for systems with explicit solvent, we compared the results of an MD simulation of the physical properties of water with corresponding experimental data. In a second application, effects of using various different implicit and explicit solvent models in restrained energy refinements of the NMR solution structure of the mixed disulfide between the mutant *E. coli* Grx(C14S) and glutathione (Bushweller et al., 1994) are described. Finally, a comparison of restrained MD simulations using different representations of the experimental NMR constraints was performed with the NMR structure of a small protein, the pheromone E*r*-2 (Ottiger et al., 1994).

Methods

Description of the program OPAL

Well-known basic algorithms and computational techniques for MD simulations have been implemented in OPAL, such as the leap-frog and velocity Verlet algorithms to integrate Newton's equation of motion (Allen and Tildesley, 1987; Van Gunsteren and Berendsen, 1990), the control of temperature and pressure by scaling of the atom velocities, the atom coordinates and the surface of the boundary at each step of a MD simulation (Berendsen et al., 1984), and the use of internal constraints to fix bond lengths at their equilibrium value by the SHAKE procedure (Ryckaert et al., 1977). The source code of OPAL was written in standard FORTRAN-77 and has been implemented on a variety of different scalar and vector UNIX machines, i.e., SUN, SGI, IBM RS6000, DEC Alpha, Convex, Cray J-90, NEC SX-3 and NEC SX-4. In the following, some features that are specific to OPAL are briefly described.

The user interface was adapted from other programs developed in our laboratory, in particular PROSA (Güntert et al., 1992). Commands are issued either interactively or via a macro file. An example for a set of commands that read the starting coordinates of a macromolecular structure, surround it with water molecules, build up the corresponding force field, and perform a complete MD run, is given in Fig. 1. For each command, on-line help is available, and a set of standard macros is provided with OPAL. Thus, the MD run of Fig. 1 can be performed by issuing a single command.

The force field parameters are read from a library file (written in a format corresponding to that used for the

DIANA program (Güntert et al., 1991)). Therefore, the source code of OPAL is applicable with different force fields, provided that these can be expressed in the format of Eq. 1:

$$E = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}(1 + \cos(n\phi - \gamma)) + \sum_{i<j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right) + \sum_{\text{H-bonds}} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \quad (1)$$

The force field is initialized for a given molecular structure by a single command ('build' in Fig. 1). In addition to the physical energy terms of Eq. 1, pseudo-energy terms representing NMR restraints, i.e., terms enforcing distance limits (Nilges et al., 1988; Billeter et al., 1990a) and dihedral angle constraints (Widmer et al., 1989) may be applied either as instantaneous constraints or as time-averaged constraints (Torda et al., 1990,1993); in the latter approach, scalar coupling constants can be used directly in the input and are averaged according to corresponding Karplus-type relations (Karplus, 1959). In addition, there is the possibility of restraining individual atom positions by pseudo-energy terms. At present, libraries for two versions of the AMBER force field are available, i.e., AMBER91 (Weiner et al., 1986; Howard and Kollman, 1992) and AMBER94 (Cornell et al., 1995).

The program includes special facilities for MD simulations and energy minimization in explicit water. A single command solvates the solute either in a rectangular box, a spherical or an ellipsoidal containment of water molecules. Alternatively, a water bath can be preequilibrated to a certain temperature and pressure prior to immersing the solute, whereby coordinates and velocities of the water molecules and system variables such as the reference temperature, reference hydrostatic pressure, boundary conditions and MD simulation parameters are read from a file (see Fig. 1). The model used for representation of the water molecules is TIP3P (Jorgensen et al., 1983). The intramolecular water potential can be represented either by bond length terms for O–H bonds and a bond angle term for the H–O–H bond angle, or by bond length terms for the two O–H bonds and a virtual H–H bond. The latter parametrization is recommended when SHAKE (Ryckaert et al., 1977) is applied.

The integrity of the system studied is enforced in OPAL as a repulsive energy term which prevents the molecules from escaping through the box-shaped, spherical or ellipsoidal surface boundary of the system. The boundary energy term is also used to maintain constant pressure in the system, which is for this purpose defined as the norm of the first derivative of the boundary energy per wall area. It has the following functional form:

$$E = \begin{cases} K\dfrac{(d_0 - d)^m}{d^n} & ; \text{if } d < d_0 \\ 0 & ; \text{otherwise} \end{cases} \quad (2)$$

The force constant of the boundary energy, K, the thickness of the imaginary boundary wall, d_0, and the exponents m and n are defined by system variables, and d is the distance of the atom from the boundary (typically: $d_0 = 1.0$ Å, $m = 3$, $n = 1$, $K = 1$ kcal m^{-2}/mol). When using a finite boundary, the pressure does not need to be calculated from internal coordinates via the virial theorem, but can be calculated directly as force per area acting on the wall. Therefore, in contrast to the use of the virial theorem, the instantaneous pressure is never negative, and in an equilibrated system it fluctuates within a narrow range about the reference value (typically between 0 and 2 bar for a reference value of 1 bar; see Fig. 2).

Optimal vectorization properties of OPAL for use on a NEC SX-3 supercomputer are achieved by calculating all non-bonded interactions without cutoff. Therefore, no pair list for the non-bonded interactions needs to be generated, which avoids indirect addressing of the atom pairs. This procedure also avoids the cutoff noise typically produced by truncation of the non-bonded interaction, which normally results in an artifactual increase of the kinetic energy and concomitantly of the temperature (Loncharich and Brooks, 1989). OPAL attains a speed of 1.5 GFlops on a NEC SX-3 supercomputer. The number of pair interactions, which increases proportional to the square of the number of atoms, is kept to a minimum by the aforementioned choice of the appropriate shape boundary. An illustration is provided by the small proteins pheromone E*r*-2 (574 atoms; Ottiger et al., 1994) and the 63-residue N-terminal DNA-binding domain of the 434 repressor (997 atoms; Pervushin et al., 1996). These two structures were immersed in water using either a box-shaped, spherical or ellipsoidal boundary. The principal axes of inertia for E*r*-2 have the ratio 1.0:1.6:2.0, and for 434 repressor the ratio is 1.0:1.3:1.4. The numbers of water molecules needed to immerse E*r*-2 with a minimal thickness of the water layer of 6.0 Å are 1625 for the box-shaped, 2008 for the spherical and 1347 for the ellipsoidal boundary. The corresponding numbers for 434 repressor are 2091, 1557 and 1194, respectively. Thus, for E*r*-2 the box-shaped boundary would be better suited than the spherical boundary, but the opposite holds for the 434 repressor. In both cases, the ellipsoidal boundary needs substantially less water molecules than either the box-shaped or the spherical boundary.

Description of the analysis program TRAJEC

The companion program TRAJEC is a tool for the analysis of MD trajectories from OPAL. It can handle large numbers of snapshots (e.g., 2000 snapshots for a

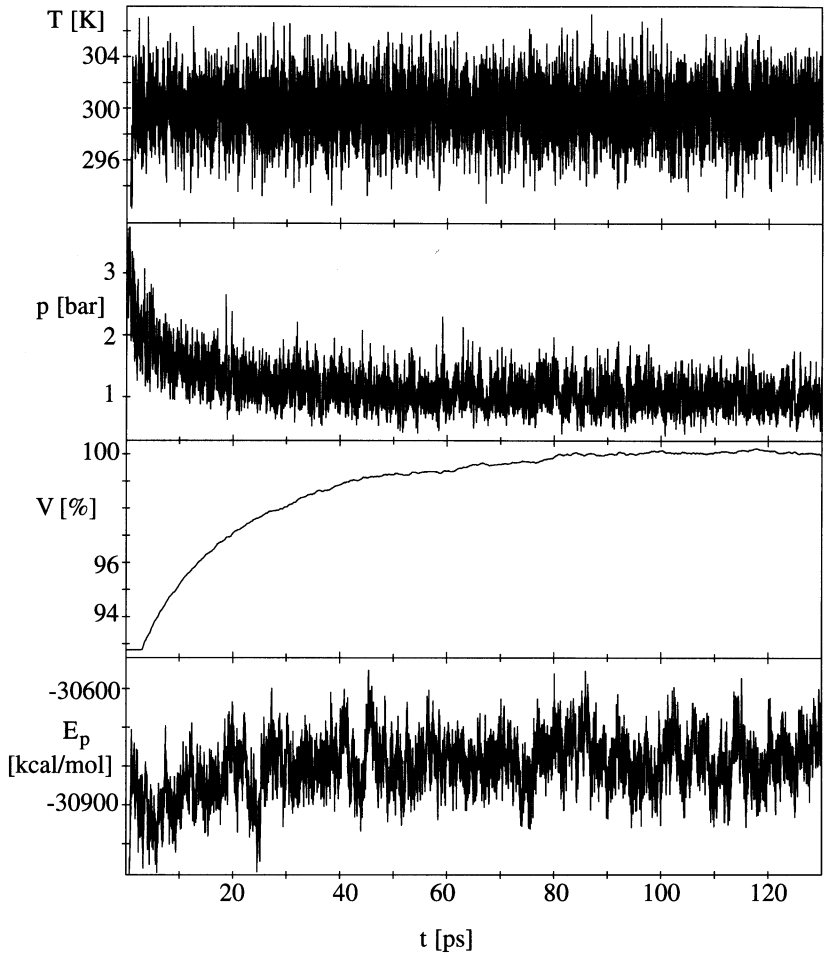

Fig. 2. Plots of the temperature, T, the pressure, p, the volume, V, and the potential energy, E_p, versus the simulation time for a system of 3318 water molecules in an ellipsoid-shaped confine (see text). The volume is plotted as the percentage of the final volume of 116 280 Å^3. The instantaneous values recorded in intervals of 0.02 ps are plotted.

system with 1579 macromolecular atoms plus 8142 water atoms (Billeter et al., 1996)) in different formats, including text or binary files. The user interface is the same as for OPAL, i.e., a series of commands can either be entered interactively, or read from a macro file. TRAJEC can superimpose all or a selected set of snapshots, and can also generate superpositions with structures from a data bank. It can further calculate structures with coordinates averaged over selected snapshots, and provide rmsd (McLachlan, 1979) or displacement values (Billeter, 1992). All such calculations can be performed with user-defined selections of atoms. Plots versus the trajectory time of rmsd values, interatomic distances, dihedral angles, order parameters, lifetimes of hydrogen bonds or other interactions, and constraint violations can efficiently be prepared, and similarly displacement values, or other parameters can be displayed versus the sequence of a macromolecule (see Figs. 5–8). The program can also tabulate different types of interactions, such as intermolecular contacts in macromolecular complexes, or contacts with water molecules in studies of hydration.

Results and Discussion

This section describes a 'calibration' of the program OPAL for MD simulations in explicit solvent water by way of a simulation of pure water. This is followed by illustrations of the practical use of the program for molecular mechanics energy minimization and MD simulation of macromolecular systems.

Physical properties of water in MD simulations with OPAL

As a test of the physical relevancy of the combined use of the AMBER force field (Weiner et al., 1986; Cornell et al., 1995), the TIP3P water model (Jorgensen et al., 1983) and the spatial confinement of the system within an ellipsoid with the boundary conditions of Eq. 2, we used OPAL to perform a 130 ps simulation of pure water. An ellipsoid with half-axes of 32.2 Å, 29.8 Å and 26.8 Å was filled with 3318 randomly oriented water molecules, the wall potential of Eq. 2 was applied, SHAKE was used for all bonds (Ryckaert et al., 1977) with the concomitant three-bond model of water (see Methods), and a dielectric

TABLE 1
COMPARISON OF EXPERIMENTAL VALUES FOR SELECTED PHYSICAL PROPERTIES OF WATER WITH THE RESULTS FROM AN OPAL SIMULATION

Parameter	OPAL[a]	Experiment
Density, ρ (g/cm^3)	1.026	0.997 ± 0.007[b]
Diffusion constant, D (cm^2/s)	5.7×10^{-5}	2.3×10^{-5} (25 °C)[b,c]
First maximum of $g_{OO}(r)$ (Å)	2.9	2.9[d]

[a] The values for the OPAL simulation are from a trajectory of 3318 water molecules using the TIP3P water model (Jorgensen et al., 1983), T = 300 K and p = 1.0 bar.
[b] Value taken from Jorgensen et al. (1983).
[c] This value was obtained at T = 298 K; for T = 288 K a diffusion constant of 1.8×10^{-5} cm^2/s was reported (Jorgensen et al., 1983).
[d] Value taken from Savage (1993).

constant of $\varepsilon = 1$ was assumed. During the 130 ps simulation the temperature, pressure, volume and potential energy were recorded in intervals of 0.02 ps, and the coordinates of all atoms were saved in intervals of 1 ps for later analysis. The initial velocities of all atoms were parallel to the forces acting on them at a system temperature of 300 K. During the initial 3 ps, when all major van der Waals violations were removed, the volume enclosed by the ellipsoidal system boundary was kept fixed. Subsequently, the volume was allowed to change in order to maintain an average pressure of 1.0 bar. The resulting time evolution of temperature, pressure, volume and potential energy is plotted in Fig. 2, using the instantaneous values measured in 0.02 ps intervals. The figure shows that an equilibration phase of about 80 ps is needed for this system.

Physical properties of water that are amenable to experimental measurements and can also be extracted from the MD trajectory are compared in Table 1, where only snapshots from the equilibrated phase of the MD simulation from 81 to 130 ps were used for the comparison. Figure 3 further illustrates the radial dependence of the density, ρ, and the radial distribution function, $g_{OO}(r)$, in particular their behavior near the surface confines. It is seen that the density of the water in the OPAL simulation is slightly higher than the experimental value. It remains constant almost up to the outer confines defined by Eq. 2; only within the last 5% of the half-axes, i.e., within about 1.5 Å from the 'wall', does the density drop to values between 1.02 and 0.93 g/cm3 (Fig. 3A). The diffusion constant of water molecules calculated by OPAL, D, exceeds the experimental value by a factor of about 2.5 (Table 1). Note that D is highly sensitive to temperature changes. The lower viscosity of water in the OPAL simulation may explain that motions observed in pure water and in systems with macromolecules and water tend to be somewhat faster than expected from corresponding NMR data (Billeter et al., 1996). The first maximum of the radial distribution function, $g_{OO}(r)$, matches very well with the experimental value measured by neutron diffraction (Narten et al., 1967; Savage, 1993); no second maximum is observed in the simulation (Fig. 3B). This result is in line with a structural model where, on average at any given time, each water molecule forms three hydrogen

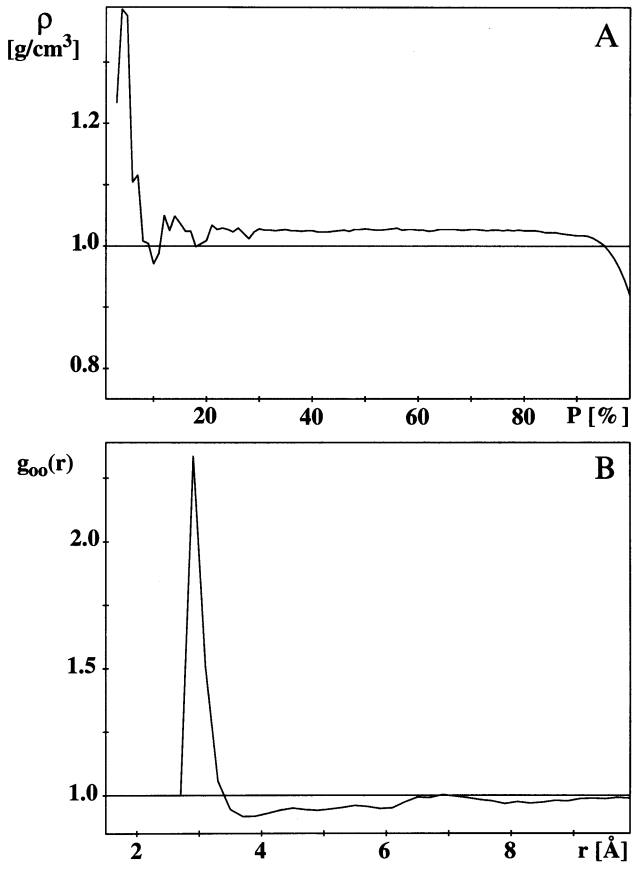

Fig. 3. (A) Variation of the water density, ρ, from the central region of the ellipsoid-shaped sample to the surface confines. The values plotted were obtained from concentric ellipsoidal shells, with the principal axes of successive shells increasing in steps of 1% of the length of the corresponding axes of the complete system. The horizontal axis shows the percentage P of the half-axes of the outer confines of the shells with respect to those of the ellipsoid that represents the surface bounds of the complete system studied. The fluctuations for P < 30% are a consequence of the small number of water molecules contained in these ellipsoids. (B) Radial distribution function, $g_{OO}(r)$, of the distance between the oxygen atoms of all pairs of water molecules, r (Narten et al., 1967). All values reported in this figure are the averages over 50 snapshots recorded in 1 ps intervals during the trajectory from 81–130 ps (see Fig. 2).

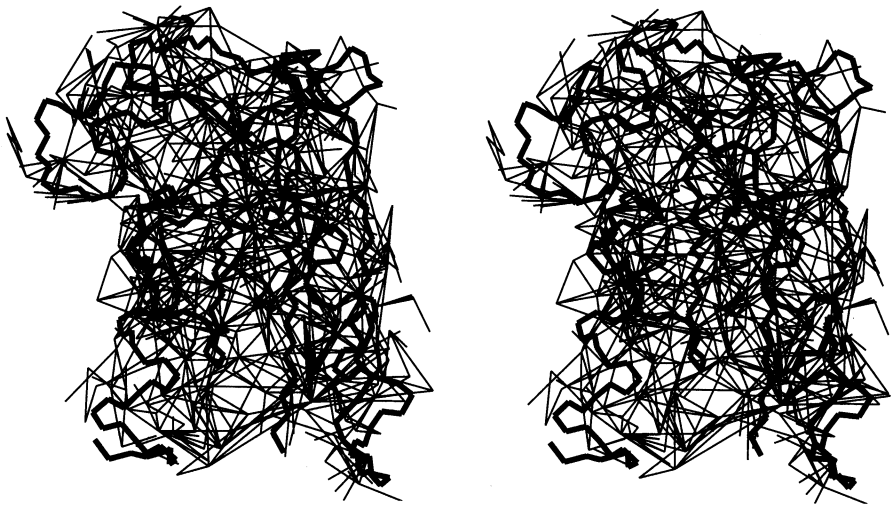

Fig. 4. Visual display of the NOE upper distance constraints used as input for the structure calculation of Grx(C14S)–SG (Bushweller et al., 1994). A stereoview of the polypeptide backbone of the conformer with the lowest residual DIANA target function value is displayed with heavy lines. Each thin line connects the positions of two atoms between which a distance constraint was observed.

bonds to neighboring molecules, with hydrogen-bond lifetimes shorter than the time resolution of 1 ps of the snapshots used to document the OPAL simulation.

Energy refinement with and without explicit water

Three different solvent models were used in restrained energy refinements of the 20 DIANA conformers of the

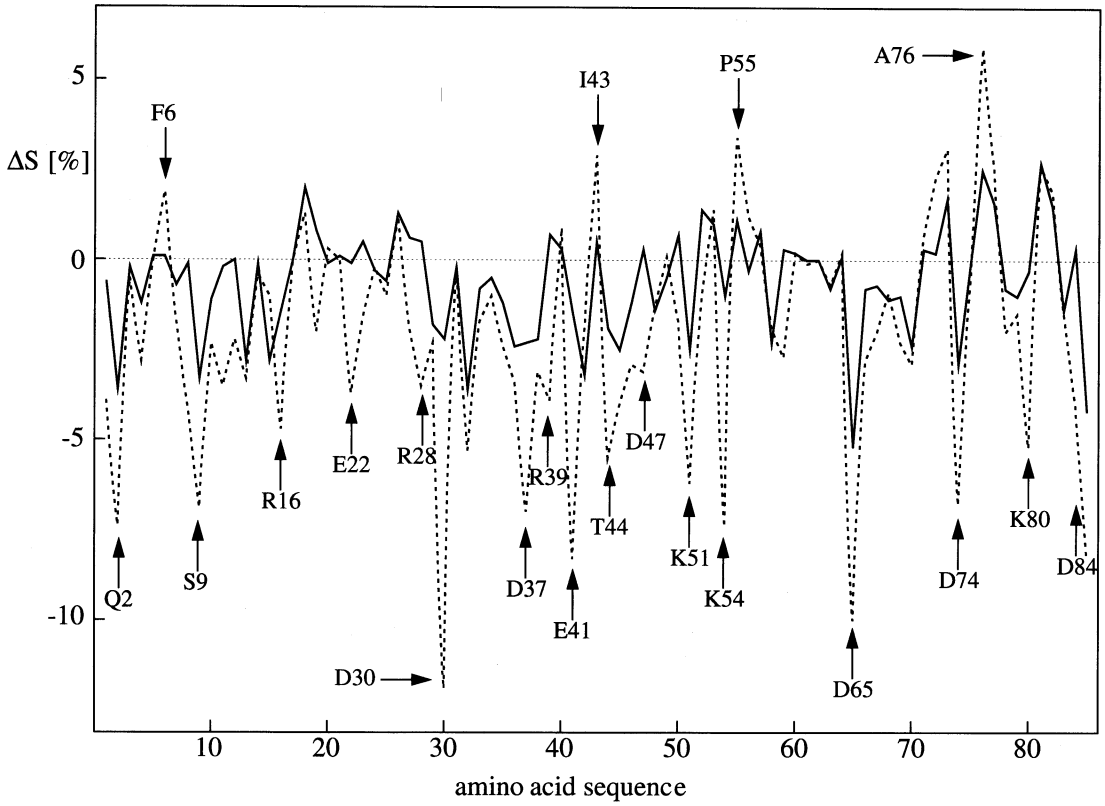

Fig. 5. Plot of the relative differences in average solvent-accessible surface area (Richmond, 1984; Billeter et al., 1990b) between the unrefined distance-geometry conformers and the conformers obtained after energy minimization with two different solvent models versus the sequence of Grx(C14S) in the mixed disulfide Grx(C14S)–SG. The sequence positions of selected hydrophilic (bottom) and hydrophobic (top) residues are indicated. Solid line: structure energy-minimized in a 6.0 Å thick water shell with constant dielectric permittivity (SC). Dotted line: structure energy-minimized in vacuo with distance-proportional dielectric permittivity (VR). Each value is the average over the 20 conformers used to represent the NMR structure (Bushweller et al., 1994).

mixed disulfide between a mutant *E. coli* glutaredoxin and glutathione, Grx(C14S)–SG, which were selected to represent the NMR structure (Bushweller et al., 1994). The 20 unrefined DIANA conformers calculated from 1010 NOE upper distance constraints and 116 scalar coupling dihedral angle constraints will be referred to as DG. Figure 4 gives an overview of the structure of this protein and of the distribution of the distance constraints used in the structure calculation, which illustrates the decrease in the density of constraints near the protein surface. The restrained energy minimizations of the 20 conformers were performed using the AMBER all-atom force field (Weiner et al., 1986) and the force constants for distance constraints and dihedral angle constraints reported by Bushweller et al. (1994). Three sets of 20 energy-minimized conformers were generated: (a) VR, with the commonly used protocol for energy-minimization in vacuo, using a distance-proportional dielectric permittivity, $\varepsilon = d/1$ Å, where d is the atom–atom distance in Å; (b) VC, with energy-minimization in vacuo, using a dielectric constant $\varepsilon = 1$; (c) SC, with refinement in a 6 Å thick shell of explicit water, using a dielectric constant $\varepsilon = 1$.

The backbone rmsd values for the 20 energy-minimized conformers relative to the mean are between 0.7 Å and 0.8 Å for each of the three protocols, which is of the same order as the corresponding value for the unrefined distance-geometry conformers. The backbone rmsd values between the mean structures obtained from the three different energy-minimization protocols are smaller than 0.5 Å, demonstrating the close similarity of the backbone conformations in the structures resulting with the use of the three different water models. The hydrogen bonds in the structures resulting from the three protocols coincide closely in the well-defined core part of the molecule, where the conformation is well determined by the high density of NOE distance constraints (Fig. 4). In particular, the populations of the 59 backbone–backbone hydrogen bonds are in good agreement, with the sole two exceptions of hydrogen bond Gly^{53} NH–O' Gln^{49}, which is highly populated only in explicit water, and Glu^{81} NH–O' Ala^{77}, which is highly populated only in the two refinements in vacuo. These hydrogen bonds are both located at the ends of helices, which appear to be sensitive to the explicit environment. Some of the 11 identified backbone–side-chain hydrogen bonds show the typical behavior seen for hydrogen bonds been pairs of backbone groups. The remaining backbone–side-chain hydrogen bonds and all 23 side-chain–side-chain hydrogen bonds are only rarely formed in the DG conformers, are lowly populated in the conformers refined in explicit water, SC, (between 0 and 7 out of the 20 conformers contained these hydrogen bonds), and are observed in 6 to 20 out of the 20 conformers VR or VC, which both were refined in vacuo. Refinements in vacuo thus tend to generate extra intramolecular hydrogen bonds and salt bridges on the protein surface. In refinements with explicit water molecules the solvent-exposed side chains can alternatively form hydrogen bonds with nearby water molecules. This is also apparent in the different average percentages of solvent-accessible surface area between the conformers VR and SC (Fig. 5): only small changes of the accessible surface, of the order of 0 to 5%, are observed upon restrained energy minimization in explicit water using a constant dielectric permittivity. In contrast, upon refinement in vacuo using a distance-dependent dielectric permittivity,

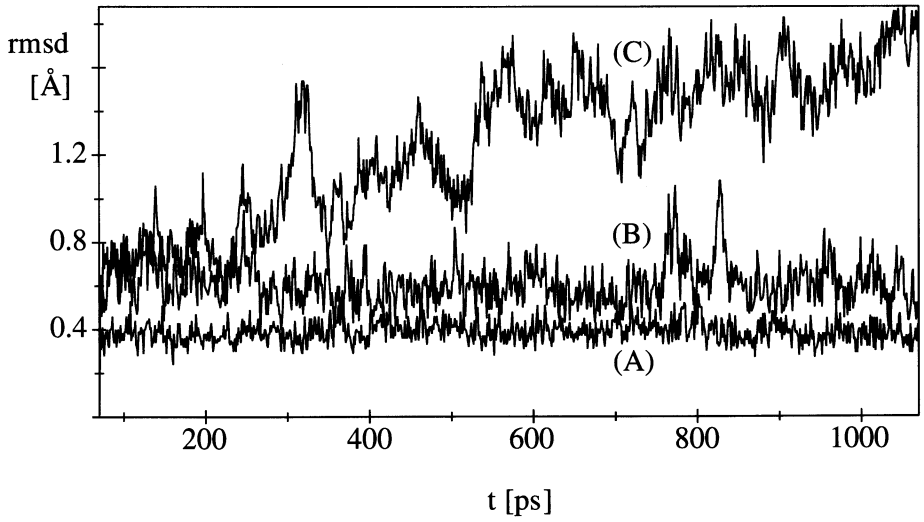

Fig. 6. Plots of the rmsd values between the E*r*-2 structures at time t and at 70 ps for the backbone atoms N, C^{α} and C' of residues 3–37 versus the simulation time, t. (A) Restrained MD simulation with instantaneous NMR constraints. (B) Restrained MD simulation with time-averaged NMR constraints. (C) Free MD simulation without NMR constraints. Using the E*r*-2 conformer with the lowest residual DIANA target function value as the starting structure, the simulation in a bath of 1708 water molecules with instantaneous NMR constraints from $t = 0$ to $t = 70$ ps was used to equilibrate the system (see text); the other simulations were started with the equilibrated structure at 70 ps (see text).

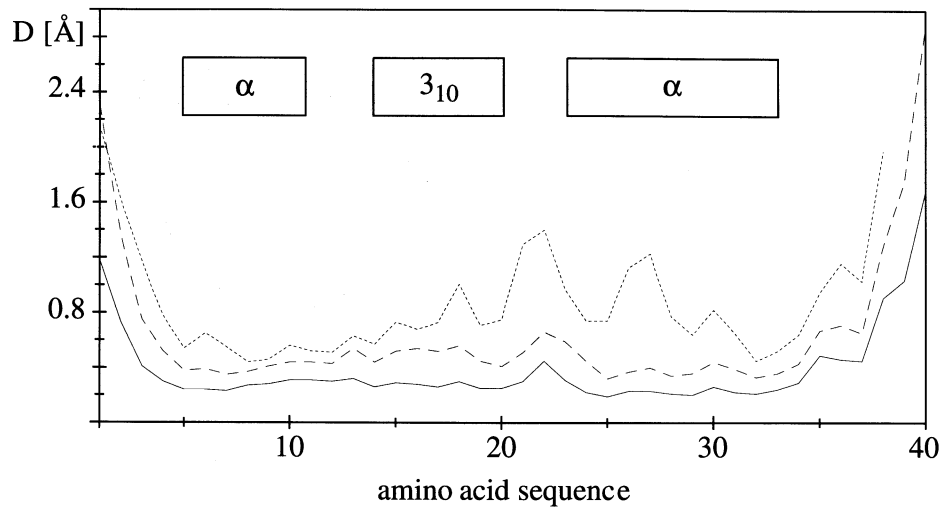

Fig. 7. Plot of the backbone displacements for the backbone atoms N, C^{α} and C' versus the Er-2 amino acid sequence. Snapshots taken in 1 ps intervals were superimposed. Solid line: restrained MD simulation with instantaneous NMR constraints; broken line: restrained MD simulation with time-averaged NMR constraints; dotted line: free MD simulation without NMR constraints. At the top the sequence locations of three helical secondary structures are indicated.

some hydrophobic side chains gain up to 6% surface accessibility, and some hydrophilic side chains lose up to 12% accessibility.

Overall, these comparisons of different refinement protocols confirm previously reported differences between explicit and implicit solvent models: (i) explicit water tends to prevent strong intra-protein interactions of polar side chains through protein–solvent interactions (Guenot and Kollman, 1992); (ii) calculations in vacuo result in hydrophilic surface groups that are underexposed to the solvent environment, and hydrophobic surface groups that are overexposed (Levitt and Sharon, 1988; Guenot and Kollman, 1992); (iii) use of a distance-dependent dielectric permittivity in vacuo to mimic the charge screening of water yields only marginal improvements when compared to the use of a fixed dielectric constant. These features, which were previously implicated from MD simulations with different software packages, are thus faithfully reproduced in molecular mechanics energy refinements by OPAL. Clearly, the use of explicit water molecules in restrained-energy refinement of NMR solution structures leads to a more realistic description of the protein surface, where one usually has a scarcity of NOE constraints when compared to the protein core.

Restrained MD simulations with different interpretations of NMR constraints

The influence of different interpretations of NMR constraints was studied with three MD simulations of the small, 40-residue protein pheromone Er-2, which forms a very stable antiparallel bundle of three helices (see also Fig. 7 below) (Ottiger et al., 1994). The DIANA conformer of Er-2 with the lowest residual target function value (Ottiger et al., 1994) was immersed in an ellipsoid containing 1708 water molecules. The temperature and pressure were fixed at 300 K and 1 bar, respectively, SHAKE was applied to constrain all covalent bond lengths, the integration steps were 2.5 fs, and the coor-

TABLE 2
NUMBER OF NOE DISTANCE CONSTRAINT VIOLATIONS FOR MD SIMULATIONS OF THE PHEROMONE Er-2 WITH DIFFERENT INTERPRETATIONS OF THE NMR CONSTRAINTS

Criteria for constraint violations[a]	MD simulations [b]		
	Instantaneous NMR constraints	Time-averaged NMR constraints	No constraints
$\langle r^{-3}\rangle^{-1/3} > b$	61	55	116
$\langle r^{-3}\rangle^{-1/3} > (b+1$ Å)	0	0	19
$\langle r^{-3}\rangle^{-1/3} \leq b$ and $r_{max} > (b+1$ Å)	0	171	147

[a] $\langle r^{-3}\rangle^{-1/3}$ denotes the effective ^1H–^1H distance corresponding to a NOE constraint obtained by time-averaging over MD trajectories of 1 ns. For a given constraint, b is the upper distance bound, and r_{max} the maximal distance value during the trajectory.

[b] The total number of NOE upper distance limits was 603. The three columns list the numbers from among these constraints that are violated for the different combinations of simulation protocols and definitions of violations.

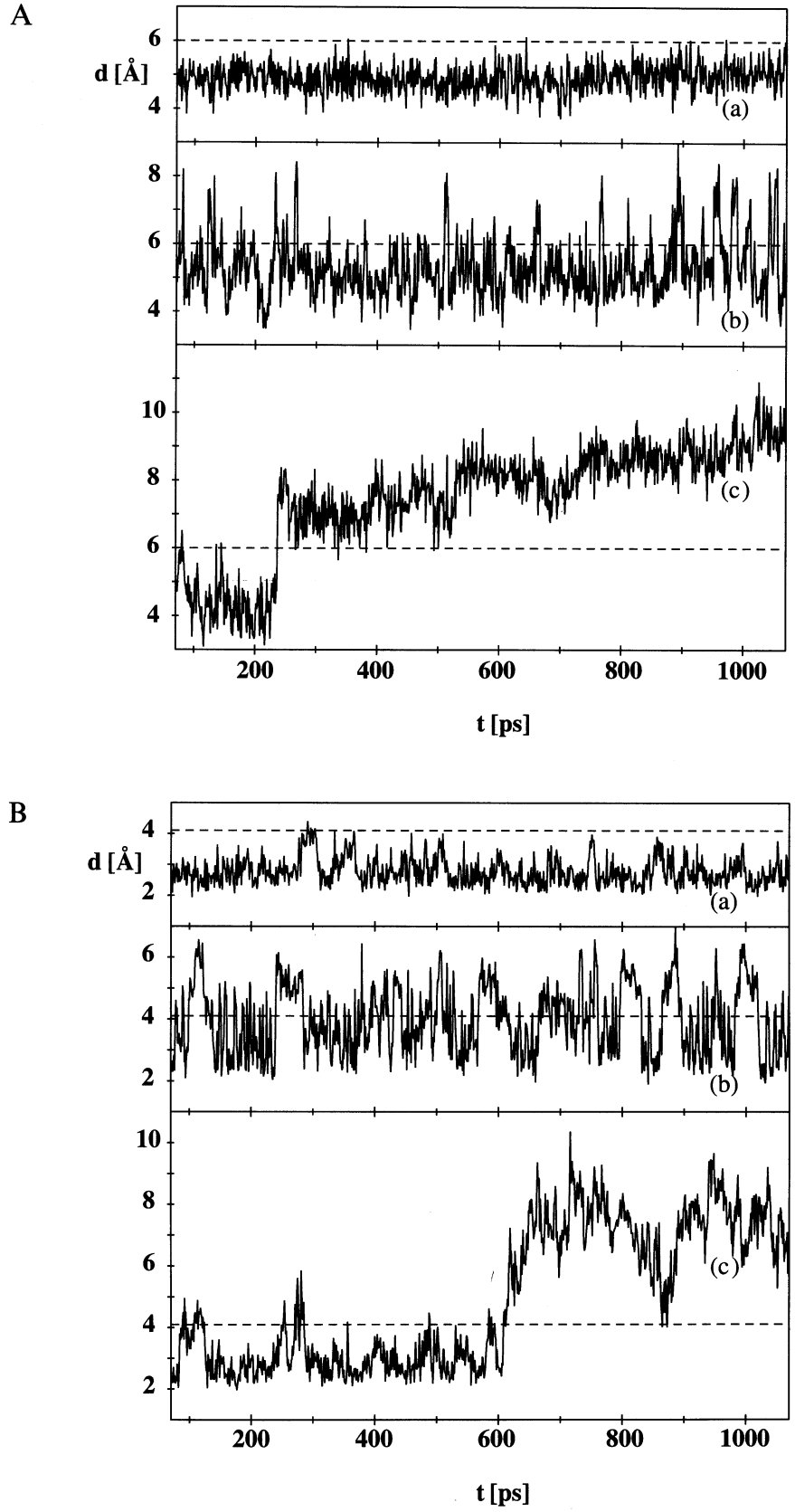

Fig. 8. Plots of selected intramolecular ^1H–^1H distances in E*r*-2 for which an upper distance limit was measured versus simulation time. (A) Distance between $H^{\beta 3}$ of Cys^5 and the δ1-methyl group of Leu^{24}, upper distance limit 6.0 Å. (B) Distance between H^{α} of Glu^{36} and $H^{\beta 3}$ of Leu^{39}, upper distance limit 4.1 Å. (a) Simulations with instantaneous constraints. (b) Simulations with time-averaged constraints. (c) Simulations without constraints. In each trace the broken line indicates the experimental upper distance constraint.

dinates of all atoms were recorded in intervals of 1 ps. The first 5 ps were run with fixed volume, thereafter the simulation was performed at a fixed pressure of 1 bar. The experimental input consisted of 603 NOE upper distance limits and 46 dihedral angle constraints (Ottiger et al., 1994). During an equilibration phase of 70 ps (Fig. 6) these NMR constraints were applied instantaneously. The rmsd value between the initial structure and the structure after 70 ps was 0.62 Å for the backbone atoms N, C^{α} and C' of residues 3–37. Starting from the structure at 70 ps, three different MD simulations were performed, each covering 1 ns. In the first run, the same instantaneous NMR constraints were used as during the equilibration phase. In the second run, time-averaged upper distance limits were applied (Torda et al., 1990), and the dihedral angle constraints were replaced by time-averaged scalar coupling constant constraints (Torda et al., 1993). The third run was a free MD simulation without NMR constraints. In Fig. 6 the rmsd values between the structures at time t and at 70 ps calculated for the backbone atoms N, C^{α} and C' of residues 3–37 are plotted versus the simulation time from 70 to 1070 ps. The rmsd value between the mean structure of the trajectory from 70 to 1070 ps and the structure at 70 ps is 0.26 Å for the simulation with instantaneous NMR constraints, and the corresponding rmsd values for the time-averaged and the free MD simulations are 0.39 Å and 1.02 Å, respectively. In the simulations with instantaneous and time-averaged NMR constraints, the rmsd stabilizes rapidly and shows only small fluctuations around an average value, whereas the rmsd reaches values larger than 1.8 Å in the free simulation. A plot of backbone displacements versus the amino acid sequence (Fig. 7) demonstrates that, when compared to the trajectory obtained with instantaneous NMR constraints, the increased extent of the conformation space covered by the trajectory obtained with time-averaged NMR constraints is distributed equally over the entire sequence. In the free trajectory, significantly increased displacements are observed from residues 18 through 40, which form part of the 3_{10}-helix, an α-helix and the C-terminal 'cap' on the three-helix bundle (Ottiger et al., 1994).

The effect of the different interpretations of the NMR constraints may be best appreciated by monitoring individual distances which are constrained by an upper distance limit (Fig. 8). The distance between $H^{\beta 3}$ of Cys^{5} and the δ1-methyl group of Leu^{24} was constrained by an upper limit of 6.0 Å in the input (Fig. 8A); similarly, there was an upper distance limit of 4.1 Å between H^{α} of Glu^{36} and $H^{\beta 3}$ of Leu^{39} (Fig. 8B). The hydrogen atoms considered have been stereospecifically assigned, and the methyl protons are represented by a pseudoatom (Wüthrich et al., 1983). During the simulation with instantaneous NMR constraints, the interatomic distances are smaller than the corresponding upper distance limits during most of the simulation time. If a constraint violation occurs, it is very small and short-lived. In the simulation with time-averaged NMR constraints, the constraint is frequently violated by up to about 2 Å. The time-averaged constraints can be satisfied by high-frequency fluctuations with much larger amplitudes than in the simulation with instantaneous NMR constraints. During the free simulation, one observes a sudden transition to a local conformation which would represent a large violation, and there is no return to a conformation that would satisfy the constraints. The amplitude of high-frequency fluctuations is only slightly larger than in the simulation with instantaneous constraints, but significantly smaller than for the time-averaged constraints. A statistical analysis of the constraint violations is given in Table 2. The number of constraint violations, i.e., the number of constraints for which the effective distance, $\langle r^{-3} \rangle^{-1/3}$, is larger than the NOE upper distance limit, corresponds to about 10% of all constraints for the simulations with instantaneous and time-averaged constraints, and about 20% for the free simulation. The number of violations was always slightly smaller for time-averaged constraints than for instantaneous constraints. The number as well as the magnitudes of the constraint violations are largest in the unconstrained simulation. Larger fluctuations are generally observed when using time-averaged constraints than with instantaneous constraints, but a comparable number of constraints is fulfilled in the time-average in both cases.

Conclusions

OPAL is a new program for molecular dynamics simulations and energy refinements of macromolecules. In our laboratory it serves two main purposes: with its high efficiency, OPAL is used for long MD runs for large systems, for example, a 2 ns trajectory of a solvated protein–DNA complex which complements experimental NMR data on the hydration of the protein–DNA interface (Billeter et al., 1996). OPAL is also routinely used for refinement of protein structures determined from NMR data by distance geometry.

OPAL is part of a suite of programs developed in our laboratory for NMR solution structure determination of biological macromolecules, which so far includes PROSA for processing of spectral data (Güntert et al., 1992), XEASY for interactive analysis of NMR spectra (Bartels et al., 1995), GARANT for automated assignment of NMR spectra (Bartels et al., 1996), HABAS for obtaining stereospecific assignments (Güntert et al., 1989), DIANA for distance geometry calculation of complete three-dimensional structures from NMR data (Güntert et al., 1991), and MOLMOL for analysis and display of macromolecular structures (Koradi et al., 1996). The OPAL output files can be displayed and analysed directly by MOLMOL.

The programs OPAL and TRAJEC are available upon request from the authors.

Acknowledgements

We thank the Centro Svizzero di Calcolo Scientifico for use of the NEC SX-3 computer, the ETH Zürich for use of the Cray Y-MP and the Cray J-90, and the Schweizerischer Nationalfonds for financial support (project 31.32033.91).

References

Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulations of Liquids*, Clarendon Press, Oxford, U.K.

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Bartels, C., Güntert, P., Billeter, M. and Wüthrich, K. (1996) *J. Comput. Chem.*, in press.

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Billeter, M., Schaumann, T., Braun, W. and Wüthrich, K. (1990a) *Biopolymers*, **29**, 695–706.

Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K. (1990b) *J. Mol. Biol.*, **214**, 183–197.

Billeter, M. (1992) *Q. Rev. Biophys.*, **25**, 325–377.

Billeter, M., Güntert, P., Luginbühl, P. and Wüthrich, K. (1996) *Cell*, **85**, 1057–1065.

Brooks III, C.L., Karplus, M. and Pettitt, B.M. (1988) *Proteins. A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Wiley, New York, NY.

Brünger, A.T. and Nilges, M. (1993) *Q. Rev. Biophys.*, **26**, 49–125.

Brunne, R.M., Liepinsh, E., Otting, G., Wüthrich, K. and Van Gunsteren, W.F. (1993) *J. Mol. Biol.*, **231**, 1040–1048.

Bushweller, J.H., Billeter, M., Holmgren, A. and Wüthrich, K. (1994) *J. Mol. Biol.*, **235**, 1585–1597.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.

Guenot, J. and Kollman, P.A. (1992) *Protein Sci.*, **1**, 1185–1205.

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.

Güntert, P., Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 517–530.

Güntert, P., Dötsch, V., Wider, G. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 619–629.

Howard, A.E. and Kollman, P.A. (1992) *Protein Sci.*, **1**, 1173–1184.

Jorgensen, W.J., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Karplus, M. (1959) *J. Phys. Chem.*, **30**, 11–15.

Kollman, P.A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 240–245.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Levitt, M. and Sharon, R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7557–7561.

Loncharich, R.J. and Brooks, B.R. (1989) *Proteins*, **6**, 32–45.

McCammon, J.A. and Harvey, S.C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, New York, NY.

McLachlan, A.D. (1979) *J. Mol. Biol.*, **128**, 49–79.

Narten, A.H., Danford, M.D. and Levy, H.A. (1967) *Discuss. Faraday Soc.*, **43**, 97–107.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Ottiger, M., Szyperski, T., Luginbühl, P., Ortenzi, C., Luporini, P., Bradshaw, R.A. and Wüthrich, K. (1994) *Protein Sci.*, **3**, 1515–1526.

Pervushin, K., Billeter, M., Siegal, G. and Wüthrich, K. (1996) *J. Mol. Biol.*, in press.

Richmond, T.J. (1984) *J. Mol. Biol.*, **178**, 63–89.

Ryckaert, J., Ciccotti, G. and Berendsen, H.C.J. (1977) *J. Comput. Phys.*, **23**, 327–341.

Savage, H.F.J. (1993) In *Water and Biological Macromolecules* (Ed., Westhof, E.), Macmillan, London, U.K.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and Van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Van Gunsteren, W.F. and Berendsen, H.C.J. (1990) *Angew. Chem.*, **102**, 1020–1055.

Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230–252.

Widmer, H., Billeter, M. and Wüthrich, K. (1989) *Proteins*, **6**, 357–371.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Wüthrich, K. (1995) *Acta Crystallogr.*, **D51**, 249–270.