J Mol Model (2004) 10:271–289
DOI 10.1007/s00894-004-0193-x

ORIGINAL PAPER

William S. York · Xiaobing Yi

CONDORR—CONstrained Dynamics of Rigid Residues: a molecular dynamics program for constrained molecules

Received: 19 January 2004 / Accepted: 7 May 2004 / Published online: 3 August 2004
© Springer-Verlag 2004

Abstract A computer program CONDORR (CONstrained Dynamics of Rigid Residues) was developed for molecular dynamics simulations of large and/or constrained molecular systems, particularly carbohydrates. CONDORR efficiently calculates molecular trajectories on the basis of 2D or 3D potential energy maps, and can generate such maps based on a simple force field. The simulations involve three translational and three rotational degrees of freedom for each rigid, asymmetrical residue in the model. Total energy and angular momentum are conserved when no stochastic or external forces are applied to the model, if the time step is kept sufficiently short. Application of Langevin dynamics allows longer time steps, providing efficient exploration of conformational space. The utility of CONDORR was demonstrated by application to a constrained polysaccharide model and to the calculation of residual dipolar couplings for a disaccharide.

Keywords Molecular dynamics · Carbohydrate · Polysaccharide · NMR

Abbreviations *Molecular dynamics:* MD · *Hard-sphere, exo-anomeric calculations:* SEA · *Metropolis Monte Carlo:* MC · *Xyloglucan endotransglucosylase:* XET · *Protein Data Bank:* PDB · *Principle axis system:* PAS · *β-Glucose archetype:* BG · *Hydroxymethyl archetype:* HM · *Advanced Micro Devices:* AMD · *Central processing unit:* CPU · *Residual dipolar coupling:* RDC · *Complex Carbohydrate Research Center:* CCRC

W. S. York (✉) · X. Yi
Complex Carbohydrate Research Center,
315 Riverbend Road, Athens, GA 30602, USA
e-mail: will@ccrc.uga.edu
Tel.: +1 706-542-4628
Fax: +1 706-542-4412

Introduction

Molecular dynamics (MD) calculations have been widely used to generate models of conformational motion of biomolecules, providing insight into the molecular processes that give rise to their important chemical and physical properties. In the most frequently employed general approach, the trajectory of each atom in the molecule is calculated based on its interactions with other individual atoms. For systems containing many atoms, this approach can become very computationally demanding as the number of atom-atom pairs, and hence the number of interactions to evaluate, increases as n^2. Thus, it is often impractical to apply this approach to large macromolecular systems such as polysaccharide aggregates, which can have hundreds of thousands of atoms.

As an alternative, atoms can be combined to form aggregates that are treated as single units. For example, the so-called "hard-sphere exo-anomeric" (HSEA) approach [1] was one of the first methods that successfully predicted the conformational properties of glycans. In this method, all the atoms in a each glycosyl residue are combined to form a rigid unit, and the geometric relationships between units are defined in terms of the glycosidic torsional angles φ and ψ and the bond angle τ. The conformations available to the glycan are defined in an "angle space", which is sampled by performing a Metropolis Monte Carlo (MMC) simulation. [2] That is, the torsional and bond angles are varied in small, random steps, and the energy of the resulting trial conformation is compared to that of the current conformation. If the energy decreases, the new conformation is accepted and the process is continued. The conformation will also be accepted if the energy increases, but only if the probability ratio (as defined by Boltzmann statistics) is less that a random number between 0 and 1. If the trial conformation is rejected, a new one is generated and tested. This approach does not produce a molecular trajectory; but rather a statistically relevant ensemble of related conformations referred to as a Markov chain. Application of MMC simulations of this type to constrained molecular systems

(such as cyclic polymers) can be problematic, as small changes in the geometry of a glycosidic bond often cause large changes in the Cartesian coordinates of distal residues, resulting in a high energy geometry that violates the molecular constraint. This produces high rates of conformation rejection, significantly slowing the progress of the calculation.

Conformational analysis of constrained glycans will undoubtedly improve our understanding of their biological functions. For example, the cell walls surrounding growing and developing plant cells consist predominantly of polysaccharide networks held together by various covalent and non-covalent crosslinks. [3] These polysaccharide networks impart tensile strength to the wall, preventing the cell from bursting under osmotic stress. Expansion of the polysaccharide networks and their incorporation of new material during cell growth involve both spontaneous (self assembly) and enzyme-catalyzed processes that are likely to be topologically complex. The hemicellulosic polysaccharide xyloglucan binds spontaneously to the surface of cellulose microfibrils in the cell wall, leading to the formation of a xyloglucan-cellulose network, which is a major load-bearing structure in the growing wall. In this network, rigid cellulose microfibrils are crosslinked by xyloglucan tethers, which are broken and reformed by an enzyme called xyloglucan endotransglucosylase (XET). [4] A comprehensive understanding of these molecular processes, including the mechanics of the xyloglucan-cellulose binding process and the effects of conformation and tensile stress on the suitability of xyloglucan tethers as XET substrates, will depend on the ability to generate dynamic models of this network. Thus, analysis of this and other biological systems would benefit by the development of alternative computational methods to examine the dynamic properties of large conformationally constrained molecular assemblies.

Toward solving this problem, we suggest an alternative approach to MD simulations, involving calculation of the trajectories of individual rigid units within the system, considering both translational and rotational motion of each unit. Here we describe CONDORR (CONstrained Dynamics of Rigid Residues), a computer program that uses this approach to simulate the molecular dynamics of large molecular systems. In the examples described here, the rigid units are individual glycosidic residues of carbohydrate polymers, which were chosen because the cyclic nature of these residues imparts localized rigidity and the energetic properties of the conformational space defined by rotation around the (flexible) glycosidic bonds has been studied previously. However, it is important to note that the computational methods used by CONDORR can be applied to other systems, such as polysaccharide or protein aggregates, where the rigid units can be much larger. The results described herein illustrate the application of the basic approach to glycans using potential energy maps based on simplified force fields, but it will be possible to obtain more accurate results efficiently with these methods upon development of better potential energy or free energy maps. The CONDORR program provides a platform that can be used to assess and utilize such energy maps, and thereby facilitate their development.

Materials and methods

The computer program CONDORR was written in C and implemented under the Linux operating system. Atomic coordinates for β-D-Glc*p* were generated starting with X-ray data [5] and performing a full geometry optimization in vacuum at 298 K using the semiempirical PM3 method with Gaussian 98 RevA.9. [6] Full geometry optimization was then performed using Density Functional Theory with B3LYP/6-311+G* at 298 K. Dihedral angles are specified using a heavy-atom nomenclature in which, for example, φ is defined as O_5–C_1–O_1–C_x and ψ is defined as C_1–O_1–C_x–C_{x-1}.

CONDORR treats each residue as a separate object, with a defined mass and moment of inertia tensor, simulating the molecular trajectories of these rigid objects with six degrees of freedom. Residues are generated from archetypes (Fig. 1) defined using data contained in a standard Protein Data Bank (PDB) file. Exchangeable protons, which are inherently mobile, are not included in the archetype definition. If a mobile group, such as the exocyclic hydroxymethyl group of a hexopyranosyl residue, is included in the PDB file that specifies an archetype, its rotation will not be calculated during the MD calculation. Alternatively, each mobile group can be put into a separate PDB file and treated as a separate residue that undergoes an independent, but constrained trajectory (Fig. 1).

Covalent linkages between residues are defined by specifying a "shared atom" that is included in the archetype definitions of both residues. For example, the glycosidic link between two (1→4)-linked residues is established by specifying that O1 of one residue is identical to O4 of the other residue (Fig. 1). The two residues are properly oriented based on bond geometry and the coordinates of the centers of mass of the two residues are adjusted so as to superimpose these two (nominally distinct) atoms. Subsequent movement and/or rotation of the residues would cause shared atoms to diverge, but their superposition is maintained by imposition of a user-specified harmonic potential. Another (usually weaker) harmonic potential maintains an appropriate distance between the two atoms (e.g., C1 of Glc*p* residue **2** and C4 of Glc*p* residue **1** in Fig. 1) that are directly attached to the shared atom (e.g., the glycosidic oxygen). Together, these two potentials establish an interresidue linkage and impose a flexible constraint upon the bond angle at the shared atom.

The harmonic potentials just described have no effect on rotation about the C1–O1 or O*x*–C*x* bond, specified by the dihedral angles φ and ψ, respectively (Fig. 1). In actual molecules, φ and ψ rotations are constrained by various forces, including those due to van der Waals interactions, electrostatic interactions, solvent effects, and the exo-anomeric effect. As indicated by the success of hard-sphere exo-anomeric (HSEA) calculations in predicting the overall features of neutral oligosaccharides [1], the dominant forces governing glycosidic bond rotations stem from van der Waals interactions and the exo-anomeric effect, which significantly limit the conformational space available to the glycosidic linkage [7]. Therefore, an extremely simple force field, based entirely on van der Waals interactions and torsional potentials, was implemented in order to test the rigid-residue computational approach embodied in CONDORR.

Forces and torques used by CONDORR when performing an MD calculation are calculated as the negative gradient of the potential energy with respect to a Cartesian coordinate or dihedral angle, respectively. Any combination of the following four methods is employed to calculate the contributions of glycosidic bond geometry to the potential energy. (1) Explicitly calculate the potential due to van der Waals interactions of atoms in each residue with atoms in its aglycon, in a manner similar to that of the HSEA

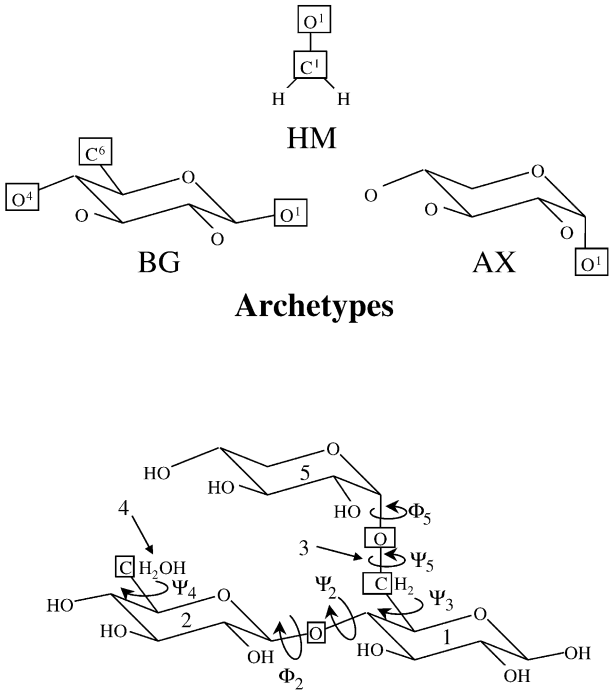

Fig. 1 Molecular models (*bottom*) are created by cloning rigid residue archetypes (*top*) and joining them together. As defined here, the archetypes AX, HM and BG respectively correspond to an α-D-Xyl*p* residue, a hydroxymethyl group, and a β-D-Glc*p* residue lacking O6, H6a and H6b. Each archetype contains atoms (indicated by boxes) that can be shared with other archetypes to form a linked structure. For example, the glycosidic link between the two D-Glc*p* residues is established by specifying that O1 of the nonreducing β-D-Glc*p* (BG) residue (**2**) is identical to O4 of the reducing Glc*p* (BG) residue (**1**). The coordinates of the two residues are adjusted so as to superimpose these two (nominally distinct) atoms. Flexible hydroxymethyl (HM) groups (**3** and **4**) are treated as separate residues, and the torsional angles (normally indicated by the symbol ω) that define their geometric relationships to the pyranosyl rings of the BG residues are specified as ψ_3 and ψ_4, respectively. The torsional angles φ_3 and φ_4, defined solely to maintain the orientation of the geminal H-atoms of the hydroxymethyl group, are not shown. (See text.) The illustrated trisaccharide is thus specified as a collection of 5 residues which are represented by 3 archetypes. Models of the disaccharide cellobiose (β-D-Glc*p*-(1→4)-D-Glc*p*) must include residues **1** and **2**, but the hydroxymethyl groups (residues **3** and **4**) can also be explicitly included in this model

approach. [1] This approach is computationally demanding, and is suitable only for short trajectories of small molecules or for the generation of potential energy maps (see below). (2) Express the potential as a user-specified function $f(\varphi)$ and/or $f(\psi)$ of one of the dihedral angles for the linkage. (3) Recall the potential associated with a particular dihedral geometry from a previously generated 2-dimensional (2D) array (i.e., a [φ, ψ] potential energy map). (4) Recall the potential associated with a particular dihedral geometry from a previously generated 3-dimensional (3D) array (i.e., a [φ, ψ, ω] map). One use of 3D potential energy maps is to facilitate the calculation of trajectories of glycans comprised of residues with mobile exocyclic groups.

In addition, the potential energy terms due to the pairwise interaction of residues that are not directly linked to each other can be applied globally (to all residue pairs) or selectively to residue pairs specified in the input file. The potential energy of these interactions is calculated either by using an all-atom approach (which, in this case, is based solely on van der Waals potentials, but which could be made more sophisticated) or by using a united-atom approach in which each residue is treated as a single, idealized sphere. Of these two methods, the all-atom approach is the more accurate, but is much more computationally demanding.

The identity and coordinates of each atom in the PDB file are used to calculate the mass and moment of inertia tensor of the archetype, which are then used as a basis for calculating the position and orientation of each residue at each time step of the simulation. The center of mass of each residue archetype (defined above) is placed at the origin of the principle axis system (PAS) defined by diagonalizing the moment of inertia matrix. The initial location and orientation of each residue are calculated (relative to the PAS atom coordinates of the archetypes) using geometric parameters specified in an input file. That is, the input file specifies (for each residue) shared atoms, atoms used to define torsional angles for the linkage, and atoms used to define bond angles, along with a residue connection table and initial torsional and bond angles. For each residue i, the orientation at time t is expressed as a matrix $\mathbf{R}_i(t)$ that rotates the residue archetype from its PAS to the MD coordinate system, in which the model is defined. Cartesian coordinates of each atom of the residue are then established by translating the center of mass of the rotated archetype to the residue's current position $\mathbf{r}_i(t)$. It should be noted that specification of the *covalent* interaction between mutually linked residues depends on the atomic coordinates of only a few atoms.

Residue trajectories are calculated using a two-part velocity-Verlet algorithm in which the position and translational velocity of each residue are treated separately from the orientation and angular velocity. That is:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\mathbf{a}_i(t)\Delta t^2}{2}$$

$$\mathbf{v}_i\left(t+\Delta t/2\right) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t)\Delta t}{2}$$

$$\mathbf{a}_i(t+\Delta t) = \frac{\mathbf{F}_i(t+\Delta t)}{m_i}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i\left(t+\Delta t/2\right) + \frac{\mathbf{a}_i(t+\Delta t)\Delta t}{2}$$

where for each residue, $\mathbf{v}_i(t)$ is the velocity, $\mathbf{a}_i(t)$ is the acceleration, $\mathbf{F}_i(t)$ is the force, and m_i is the mass.

The incremental change (for the current time slice) in the orientation of each residue is expressed as a "rotation vector"

$$\mu_i(t) = \omega_i(t)\Delta t + \frac{\alpha_i(t)\Delta t^2}{2}$$

where $\omega_i(t)$ is the angular velocity and $\alpha_i(t)$ is the angular acceleration. At each time step, $\boldsymbol{\mu}_i(t)$, whose direction corresponds to an axis of rotation and whose length corresponds to a rotation angle in radians, is converted to a *incremental* rotation matrix $\Delta\mathbf{R}_i(t)$. The overall rotation matrix $\mathbf{R}_i(t)$, which transforms archetypal atomic coordinates of residue i to the MD coordinate system, is updated by multiplication by $\Delta\mathbf{R}_i(t)$. These definitions allow the rotation matrix and angular velocity for each residue to be updated using an algorithm analogous to that used for updating translational parameters.

$$\mu_i(t) = \omega_i(t)\Delta t + \frac{\alpha_i(t)\Delta t^2}{2}$$

$$\Delta\mathbf{R}_i(t) \leftarrow \mu_i(t)$$

$$\mathbf{R}_i(t+\Delta t) = \Delta\mathbf{R}_i(t)\mathbf{R}_i(t)$$

$$\omega_i\left(t+\Delta t/2\right) = \omega_i(t) + \frac{\alpha_i(t)\Delta t}{2}$$

$$\alpha_i(t+\Delta t) = \frac{\Gamma_i(t+\Delta t)}{I_i}$$

$$\omega_i(t+\Delta t) = \omega_i\left(t+\Delta t/2\right) + \frac{\alpha_i(t+\Delta t)\Delta t}{2}$$

where, for residue i, the vector $\Gamma_i(t)$ is the torque, $\alpha_i(t)$ is the angular acceleration, and I_i is the moment of inertia. The above equations idealize each residue as a sphere, for which the moment of inertia can be expressed as a scalar quantity I_i rather than as a tensor. CONDORR also allows the angular acceleration $\alpha_i(t)$ and the updated angular velocity $\omega_i(t+\Delta t)$ to be calculated by using the diagonalized moment of inertia tensor of the residue archetype. In this mode, which is more computationally demanding, $\omega_i(t)$ and $\Gamma_i(t)$ are multiplied by $\mathbf{R}_i^{\mathrm{T}}(t)$, which rotates them to the PAS in which the archetype of residue i is defined. Euler's equations and the diagonalized moment of inertia tensor for the archetype are then used to calculate the residue's angular acceleration $\alpha_{i,\mathrm{PAS}}(t)$ in this principle axis system. Multiplication of $\alpha_{i,\mathrm{PAS}}(t)$ by $\mathbf{R}_i(t)$ generates $\alpha_i(t)$ in MD coordinates, which is then used to update $\omega_i(t)$ and calculate $\boldsymbol{\mu}_i(t)$.

CONDORR implements a Langevin dynamics [8] approach in order to control the simulation temperature and explore the conformational space available to the model polymer more efficiently. This includes calculating the effects of friction, as described by the frictional coefficients $\gamma_{i,\mathrm{trans}}$ and $\gamma_{i,\mathrm{rot}}$, which slow the motions of each residue in proportion to its translational and rotational velocity, respectively. The resulting loss of kinetic energy is balanced by stochastic forces $\mathbf{F}_{i,\mathrm{stochastic}}(t)$ and torques $\Gamma_{i,\mathrm{stochastic}}(t)$, which contribute varying amounts of kinetic energy at each time step. Thus, the total force $\mathbf{F}_i(t)$ and torque $\Gamma_i(t)$ acting on each residue are calculated as follows.

$$\mathbf{F}_i(t) = \mathbf{F}_{i,\mathrm{stochastic}}(t) + \mathbf{F}_{i,\mathrm{molecular}}(t) - \gamma_{i,\mathrm{trans}}\mathbf{v}_i(t)$$

$$\Gamma_i(t) = \Gamma_{i,\mathrm{stochastic}}(t) + \Gamma_{i,\mathrm{molecular}}(t) - \gamma_{i,\mathrm{rot}}\omega_i(t)$$

where $\mathbf{F}_{i,\mathrm{molecular}}(t)$ and $\Gamma_{i,\mathrm{molecular}}(t)$ are the force and torque on residue i due to its interactions with other residues. The magnitudes of the stochastic contributions $\mathbf{F}_{i,\mathrm{stochastic}}(t)$ and $\Gamma_{i,\mathrm{stochastic}}(t)$ are calculated according to the Fluctuation Dissipation Theorem such that a constant temperature T is maintained [8]. That is, $\mathbf{F}_{i,\mathrm{stochastic}}(t)$ and $\Gamma_{i,\mathrm{stochastic}}(t)$ are chosen from normally distributed collections of random (3-dimensional) vectors with zero average and standard deviations σ *in each dimension* of

$$\sigma_{\mathrm{trans}} = \left(\frac{2k_{\mathrm{B}}T\gamma_{\mathrm{trans}}}{\Delta t}\right)^{1/2}$$

$$\sigma_{\mathrm{rot}} = \left(\frac{2k_{\mathrm{B}}T\gamma_{\mathrm{rot}}}{\Delta t}\right)^{1/2}$$

where k_{B} is Boltzmann's constant.

The trajectory of each residue is based on the total force $\mathbf{F}_i(t)$ and torque $\Gamma_i(t)$ to which it is subjected, including those due to mutual interactions of residues (and atoms within residues, when evaluated) and those introduced as stochastic elements of the Langevin equation. As described above, residue-residue interactions can be calculated by treating each residue as an ideal sphere, ignoring the interaction of individual atoms within the residues, in which case the force on each residue is simply added to $\mathbf{F}_i(t)$, and no torque is calculated. Conversely, the force resulting from the interaction of an atom in one residue with an atom in another residue causes both atoms to accelerate, which can result in both linear and angular acceleration of the interacting residues. This is the case, for example, when a pair of "shared" atoms that defines a glycosidic linkage diverge in space. As described in the Appendix, the force $\mathbf{F}_{i,k}(t)$ acting on atom k of residue i is simply added to the total force $\mathbf{F}_i(t)$ on the residue. In addition, the torque $\Gamma_{i,k}(t)$ acting on residue i due solely to $\mathbf{F}_{i,k}(t)$ is very simply calculated as the vector product

$$\Gamma_{i,k}(t) = \mathbf{r}_{i,k} \times \mathbf{F}_{i,k}(t)$$

where $\mathbf{r}_{i,k}$ is the vector from the center of gravity of residue i to atom k of residue i. The vectors $\mathbf{F}_{i,k}(t)$ and $\Gamma_{i,k}(t)$ are evaluated at each time step and added to the total force $\mathbf{F}_i(t)$ and torque $\Gamma_i(t)$ vectors for the residue. The location, acceleration, velocity, orientation, angular acceleration, and angular velocity of each residue is then updated using the velocity-Verlet algorithm described above. As described in the Appendix, these simple equations conserve the total energy, linear momentum and angular momentum of the system.

For residue–residue interactions in which the potential energy is expressed solely as a function of a torsional angle φ, the contribution $\Gamma_{i,\phi}(t)$ to the torque is calculated by defining a fictitious atom in one of the residues, a fictitious force $\mathbf{F}_{f,\phi}(t)$, and a distance vector $\mathbf{s}$ specifying the separation of the fictitious atom from a specified atom in the other residue, as described in the Appendix. The force $\mathbf{F}_{f,\phi}(t)$ is collinear with $\mathbf{s}$ and the derivative of the potential energy with respect to $\mathbf{s}$ is related to the torsional potential gradient.

$$\left|\mathbf{F}_{f,\phi}\right| = \frac{-\mathrm{d}U}{\mathrm{d}s} = \frac{-\mathrm{d}U}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}s}$$

Defining this fictitious force allows the torsional potential gradient to be evaluated in a way analogous to that described above for forces between individual atoms in the model. That is, the total force vector $\mathbf{F}_i(t)$ for the residue is incremented by the fictitious force $\mathbf{F}_{f,\phi}(t)$ and the total torque vector $\Gamma_i(t)$ for the residue is incremented by $\Gamma_{i,\phi}(t) = \mathbf{r}_{i,f} \times \mathbf{F}_{i,\phi}(t)$, where $\mathbf{r}_{i,f}$ is the vector from the center of gravity of the residue to the fictitious atom. The (equal and opposite) force on the (real) atom in the other residue is treated just like any other force on the atom, leading to translational and rotational acceleration fothe residue. This approach conserves the total energy, angular momentum, and linear momentum of the system.

The molecular geometry of CONDORR models is calculated on the basis of information contained in several input files (Table 1). The *archetype* section of the main CONDORR input file specifies a PDB file for each residue type, along with other information that allows CONDORR to generate archetypal residues. For each archetype, the atoms involved in the formation of glycosidic bonds are also designated, including "shared" atoms and atoms that define glycosidic torsional and bond angles. In order to provide more flexibility in specifying which atoms to exclude when van der Waals forces are calculated, atoms that are one or two bonds away from each shared atom (i.e., alpha and beta atoms, respectively) are also specified. More than one residue can be constructed for each archetype defined in this way. For example, only one archetype would be listed in this section when simulating the geometry of a homopolymer such as poly-α-D-Glc*p* (amylose).

The *sequence* section of the input file (Table 1) contains information regarding the order and linkage of archetypes in the model, which constitutes a connection table. The initial torsional and glycosidic bond angles are also included in this section, along with specification of force parameters that are applied selectively to individual residues in the sequence. Together, the archetype and sequence sections of the input file provide all of the information required to generate the initial geometry of the model. If the keyword 'pdbStart' is included in the *general parameter section* of the input file (Table 1), a PDB file specifying the initial geometry of the model is generated.

CONDORR uses relatively few potential energy parameters when calculating molecular trajectories. These parameters, specified in the force field section of the main input file (Table 1), include force constants for the harmonic potentials that maintain superposition of shared atoms and appropriate bond angles about the shared atoms. Torsional potential energy functions are also specified in this section, along with any globally applied van der Waals parameters and the location of files containing 2D and 3D conformational potential maps, which can be used for the rapid calculation of molecular trajectories. The 2D maps describe the potential associated with rotations, usually expressed as torsional angles φ and ψ, about the two chemical bonds that connect a pair of residues (Fig. 1). For glycosidic linkages, the torsional angles are indexed by reference to the glycosyl residue (i.e., φ_i and ψ_i specify the geometry of the linkage between residue i and its aglycon.) CONDORR supports the creation of 2D and 3D maps by executing grid searches that use van der Waals parameters, their global or selective application to residue-residue interactions, and torsional

Table 1 Partial list of CONDORR input keywords and parameters

Keyword	Arguments	Effect
MD|noMD[a]	none	Toggle calculation of MD trajectory
langevin[a]	none	If present, use Langevin dynamics
totalTime[a]	time (ps)	Set duration of MD simulation
timeStep[a]	time step (ps)	Set time step
snapshot[a]	frequency	Set number of steps per data shapshot
temp[a]	temperature (K)	Set MD temperature
isotropic|anisotropic[a]	none	Set moment of inertia mode
pdbStart[a]	none	If present, create pdb file of starting conformation
pdbMovie[a]	none	If present, create a pdb file for each snapshot
globalContact[a]	cutoff (Å)	If present, calculate interaction potential of all residue pairs using united atom, spherical approximation
globalVDW[a]	cutoff (Å)	If present, calculate interaction potential of all residue pairs using each atom in the residues
2Dgrid[a]	linkage, pdbGrid, steps	If present, calculate 2D potential map for the specified linkage and granularity. If pdbGrid is 1, create a pdb file for each map vertex.
3Dgrid[a]	linkage1, linkage2, steps	If present, calculate 3D potential map for the specified linkages and granularity.
xWalls[a]	limit	If present, sets a potential barrier at $x=\pm$limit.
restart[a]	restart file	Restart MD using coordinates and orientations from restart file
archetypeDefinitions[b]	archetype count	Start archetype section
archetype[b]	index	Start definition of an archetype
name[b]	archeype name	Set archetype name
pdbFile[b]	pdb file, atoms	Set atomic coordinates for archetype, using specified number of atoms from pdb file
linkers[b]	linker count	Set number of linkers for the archetype
linker[b]	index, atom, atom, atom	Start definition of an archetype linker
alphaAtoms[b]	linker index, atom count, atom indices	Define alpha atoms for the linker
betaAtoms[b]	linker index, atom count, atom indices	Define beta atoms for the linker
k1[c]	k1	Set force constant for superposition of shared atoms
k2[c]	k2	Set force constant for linkage bond angle
k3[c]	k3	Set force constant for xWall interaction
2DPotentialMaps[c]	count	Start 2D map definitions
2DMap[c]	index, map file	Define a 2D map
3DPotentialMaps[c]	count	Start 3D map definitions
3DMap[c]	index, map file	Define a 3D map
torsionFunctions[c]	count	Start torsional potential function definitions
function[c]	index, count, [amplitude, frequency, phase, power] ...	Define a torsional potential function with specified number of terms, each with a specified amplitude, frequency, phase, and power
sequence[d]	count	Start sequence definition for a specified number of residues[e]
harmonic[d]	residue1, atom1, residue2, atom2, force constant, equilibrium distance	Set up a harmonic potential between two atoms
3DmapApplications[d]	count	Start specification of linkages to which 3D maps will be applied
apply3Dmap[d]	map index, linkage 1, linkage 2	Set application of 3D map to two interacting linkages

The input file is divided into four sections: general[a]; archetype[b]; force field[c]; sequence[d].

[e] The sequence parameters for each residue are specified in lines that follow the "sequence" keyword. Each of these lines includes the following parameters for a specific residue: [index, archetype-ID, aglycon-index, aglycon-link-index, tau, phi, psi, 2D-map-index, phi-torsional-function-index, psi-torsional-function-index, explicit-vdw-flag, traslation-freeze-flag, rotation-freeze-flag]. For the last six parameters, which specify how the force field will be applied to the linkage of the residue to its aglycon, a value of zero specifies no force will be applied.

potential functions, as specified in the input file. More accurate 2D and 3D maps could be generated using a wide range of techniques, including quantum-mechanical calculations and more sophisticated all-atom calculations.

CONDORR provides considerable flexibility in defining residues and their linkages. That is, a CONDORR residue can encompass a collection of polymers, a group of sugar residues within a single polymer, an individual sugar residue, or a small molecular fragment, such as the hydroxymethyl group. Therefore, the torsional angles φ and ψ that connect "residues" can have non-standard meanings. For example, two archetypes (Fig. 1) can be used to define a glucopyranosyl residue with a rotatable hydroxymethyl group. The two archetypes must first be prepared by selecting appropriate atoms from a "complete" glucoyranosyl residue, in which the exchangeable protons have been removed. The first archetype (BG, Fig. 1) includes all atoms of the glucosyl residue except H6a, H6b, and O6. Included in the BG archetype definition is specification of $C6_{BG}$ as a shared atom. The second archetype (HM) includes the three atoms that were omitted from the definition of the BG archetype (renamed $H1a_{HM}$, $H1b_{HM}$, and $O1_{HM}$), along with a "shared" carbon atom ($C1_{HM}$) that is identified with $C6_{BG}$. Superposition of the two shared atoms ($C6_{BG}$ and $C1_{HM}$) and establishment of the appropriate distance between $C5_{BG}$ and $O1_{HM}$ places the two "residues" in the proper relative position and fixes the bond angle ($C5_{BG}$–$C1_{HM}$–$O1_{HM}$). This geometry is maintained during the simulation by the imposition of harmonic potentials, as

described above. However, careful examination of this construction reveals that it does not constrain the orientation of the methylene protons ($H1a_{HM}$ and $H1b_{HM}$). Therefore, a dummy atom, corresponding to C5 of the BG archetype, is also included in the definition of the HM archetype. To maintain internal consistency within CONDORR, the torsional angle φ for the linkage between HM and BG is actually an "improper" torsion, defined as [dummy]$_{HM}$–$O1_{HM}$–$C1_{HM}$–$C5_{BG}$. Constraining φ to a value of zero (by using a torsional potential function) maintains superposition of the dummy atom with $C5_{BG}$, maintaining the proper orientation of methylene protons. The torsional angle ψ for this linkage is defined as $O1_{HM}$–$C1_{HM}$–$C5_{BG}$–$O5_{BG}$, which, in the nomenclature commonly used for glycosyl residues, is referred to as ω.

The potential energy associated with the geometry of a glycosidic linkage (or other molecular connection) often depends on more than two torsional angles. This is most obvious for the (1→6)-linkage to a hexopyranosyl aglycon (Fig. 1), whose potential energy surface is a function of three independent variables (i.e., the three torsional angles that are usually called φ, ψ, and ω). As just described, CONDORR allows rotation about the C5-C6 bond by separating the hexopyranosyl residue into two CONDORR residues. Thus, a glycosidic (1→6)-linkage corresponds to two separate CONDORR linkages (Fig. 1). Defining the aglyconic glucose as residue i, the exocyclic hydroxymethyl group as residue i+1, and the glycosyl residue attached to O6 as residue i+2 (Fig. 1, bottom panel), the three independent variables are φ_{i+2}, ψ_{i+2}, and ψ_{i+1}, respectively. Accordingly, CONDORR can calculate 3D potential maps based on systematic variation of three torsional angles, φ and ψ of one residue and ψ of a second residue. (More sophisticated 3D maps can be generated using other methods.) As described above, φ_{i+1}, which just determines the orientation of the methylene protons, should be constrained to a value near zero, and is not considered when generating the map.

Although two residues that are consecutive in the residue sequence of the model are chosen to specify the independent variables of the 3D potential energy map for a (1→6)-linkage, 3D potential maps can also be used to describe the energetics associated with relative geometry variation for any two residues in the model. For example, the potential energy surface (φ, ψ map) describing various glycosidic bond conformations of cellobiose (β-D-Glc*p*-(1→4)-D-Glc*p*, Fig. 1) depends significantly on the orientation of the hydroxymethyl group of the aglycon. [7] CONDORR can efficiently account for this by utilizing a 3D potential map in which the independent variables are φ_2 and ψ_2 for the linkage connecting the two Glc*p* residues and ψ_3 for the linkage connecting the hydroxymethyl group to the reducing Glc*p* residue (see Fig. 1). For each component $\frac{\partial U}{\partial \varphi_i}$, $\frac{\partial U}{\partial \psi_i}$, or $\frac{\partial U}{\partial \psi_j}$ of the gradient of this potential surface, a torque is applied to the two mutually linked residues whose relative geometry is described by the corresponding torsional angle. That is to say, torques associated with two separate, interacting linkages can be generated from a single 3D map, increasing the efficiency of the calculation.

Results

Application of CONDORR to simple systems

CONDORR was tested to determine whether it generates molecular ensembles that are consistent with the force field defined in the input file. As described in the Methods section, the current version of CONDORR implements simple force fields consisting of potential energy terms corresponding to van der Waals interactions and torsional angle functions, and can generate and utilize potentials recalled from 2D or 3D potential maps. It can also utilize more sophisticated potential maps generated by external methods, such as quantum mechanical calculations or grid searches done by other molecular dynamics or molecular mechanics software. CONDORR was used to generate 2D and 3D maps based on torsional functions taken from the literature and van der Waals interactions. In this case, the exo-anomeric effect was approximated using the torsional functions of Tvaroska et al. [7] It has been shown that when these functions and Lennard-Jones 6–12 van der Waals potentials are applied to a system of rigid residues, the energetic barriers to rotation of the glycosidic bond are overestimated, and the model is too inflexibile. [7] Therefore, van der Waals potentials were calculated using a function that reproduces the attractive portion of the Lennard-Jones function, but is less steep in the repulsive portion. Initially, equations were derived to express forces acting on a residue when each of its constituent atoms is constrained by a radially symmetric harmonic potential about the its initially-defined location within the residue archetype, and the interaction of atoms in different residues is described by a Lennard-Jones 6–12 function. In this model, the "virtual movement" of atoms within each archetype leads to a potential function with a softer repulsive regime, decreasing the energetic barriers to glycosidic bond rotation. However, the profiles of such soft potential functions are very similar to van der Waals functions with the standard form

$$V_{i,j} = A\frac{r_{\min,i,j}^{m}}{r_{i,j}^{m}} + B\frac{r_{\min,i,j}^{n}}{r_{i,j}^{n}}$$

in which m<12, n<6, and $r_{\min,i,j}$ is the internuclear distance corresponding to minimum potential. (When m=12 and n=6, this function corresponds to a Lennard-Jones 6–12 potential.) Soft (m<12) potentials described by this equation are much easier to evaluate than the functions that include a radially symmetric harmonic potential term. The values of m and n are user-definable parameters within CONDORR, allowing the hardness of the residue surface to be adjusted. Typically, m was set to a value of 10.0 and n was set to 5.0.

A 2D (φ,ψ) map (180×180 points, Fig. 2a) for the β-(1→4)-linkage of cellobiose (β-D-Glc*p*-(1→4)-D-Glc*p*, Fig. 1) was generated using an exoanomeric torsional potential [7] for φ in combination with modified van der Waals potentials (m=10, n=5) for atom–atom interactions. As illustrated in Fig. 1, the model used in this case included two rigid units, i.e., β-D-Glc*p* residues (**1** and **2**), which lack O6, H6a, and H6b. The specification of "alpha atoms" (i.e., atoms that are directly attached to the shared atoms) in the archetype definition was automatically invoked by CONDORR to suppress the calculation of 1–3 interactions across the glycosidic bond, which would otherwise give rise to very high residue interaction energies and significantly degrade the reliability of the calculation. In order to eliminate bias due to the fixed orientation of the exocyclic hydroxymethyl groups, O6, H6a and H6b were not included in the archetypes used in the generation of this potential energy map (although these were included when 3D potential energy maps were generated; see below.) Conformational energy maps cal-

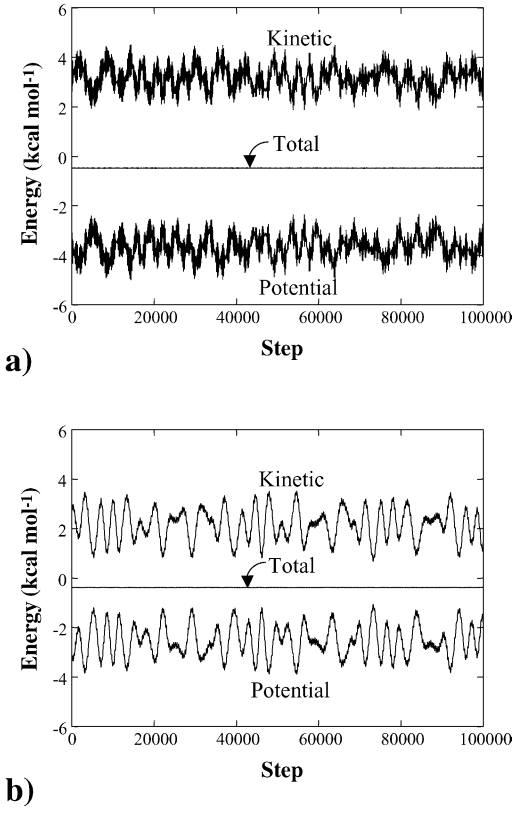

Fig. 2a, b Conservation of energy by CONDORR. Unconstrained MD simulations were performed without including stochastic (Langevin) forces while using **a** explicit van der Waals potentials or **b** the 2D potential energy map shown in Fig. 3A. The total energy (kinetic plus potential) remained nearly constant over 10^5 steps in each simulation

culated in this manner are not highly accurate, as they ignore, for example, electrostatic interactions, hydrogen bonding, and solvent effects. However, they are useful in that they direct sampling of conformational space in a way that avoids physically inaccessible regions, providing a basis for the extremely rapid evaluation of the conformational energy surface, significantly speeding up molecular dynamics calculations. Furthermore, they provide a convenient means of testing the ability of CONDORR to generate trajectories that coincide with a specified force field.

Two simulations (10 ps in 10^5 steps, 0.0001 ps/step, Fig. 2) were performed for cellobiose, the first using explicit calculation of all interresidue van der Waals interactions, and the second using the potential energies stored in a 2D (φ, ψ) map (Fig. 3a) generated by CONDORR using the same force field. Free evolution of the molecular conformation was allowed for both simulations (i.e., no stochastic Langevin forces were applied) and the total energy, linear momentum, and angular momentum were monitored. As shown in Fig. 2, the total energy was conserved while the kinetic and potential energies fluctuated. Linear and angular momentum were also conserved (data not shown). However, some drift in the total energy was observed (data not shown) when the time step was increased to 0.001 ps. Similar results were obtained regardless of whether isotropic moments of inertia or anisotropic moment of inertia tensors were used to calculate the angular acceleration of each residue. These results demonstrate that, as derived analytically in the Appendix, total energy, linear momentum, and angular momentum are conserved by the CONDORR algorithm in the limit as $\Delta t \rightarrow 0$.

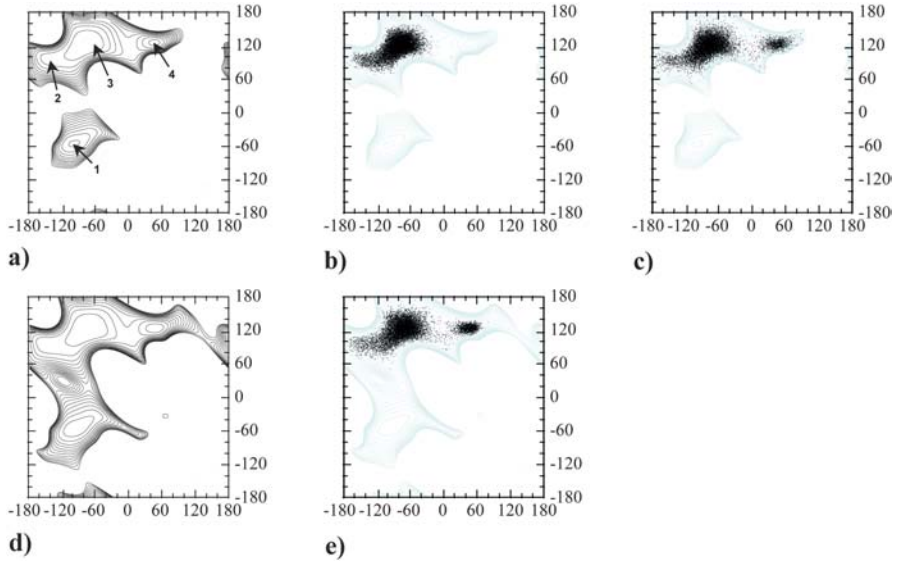

Fig. 3 a Two-dimensional potential energy surface of cellobiose generated using a simple force field consisting of van der Waals (m=10, n=5) and exoanomeric terms. Conformational families 1–4 are indicated by arrows. **b** Conformational sampling of a 10-ns MD simulation of cellobiose performed using the 2D energy surface shown in **a**. **c** Conformational sampling of a 10 ns MD simulation of cellobiose performed explicitly using the van der Waals (m=10, n=5) and exoanomeric parameters that were used to generate the 2D energy surface shown in **a**. **d** Potential surface as in **a**, but using van der Waals parameters m=9, n=4.5. **e** Conformational sampling as in **c**, but using the energy surface shown in **d**

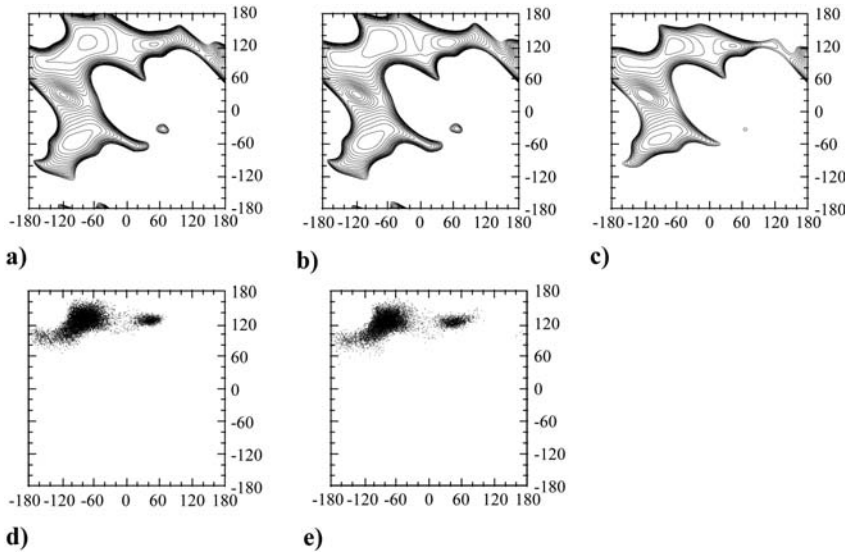

Fig. 4a–e a, b, c Planes in the 3D potential energy surface of cellobiose generated using the same simple force field illustrated in Fig. 2d (m=9, n=4.5), but including systematic variation of the hydroxymethyl group of the reducing Glc*p* residue. The three planes correspond to the gg (**a**), gt (**b**), and tg (**c**) conformations of the hydroxymethyl group. Conformational sampling of a 10 ns MD simulation of cellobiose performed using this 3D energy surface (**d**) and explicit evaluation of van der Waals (m=10, n=5) and exoanomeric parameters (**e**)

The 2D (φ, ψ) potential energy map was also used as the basis for an efficient Langevin dynamics simulation of the conformational trajectory of cellobiose (Fig. 3b). This 10-ns calculation, comprised of 10^7 steps (1 fs per step) and implementing an *anisotropic* moment of inertia tensor for both residues, was completed in under 7 min on a desktop computer equipped with a 1.1-GHz AMD Athlon CPU. A much less efficient 10-ns calculation was also performed, using the same parameters except explicitly calculating the van der Waals potentials for each time step (Fig. 3c). This calculation required 80 min when performed on the same computer. The conformational ensembles generated by these two calculations both reflected the potential energy surface calculated by grid searching, although they were not identical, due to biases inherent in the different computational methods used. That is, the trajectory calculated using explicit evaluation of van der Waals potentials exhibited greater flexibility than that calculated using the 2D potential map. The basis for this difference is that the 2D map reflects torsional potentials calculated on the basis of a *fixed* C1–O1–Cx bond angle τ and absolute superposition of the atoms shared by the two residues. Thus, although the value of τ could vary and the shared atoms could diverge in both simulations, such processes could lead to a decrease in the potential energy of a specific conformation (as defined by specific values of φ and ψ) only when explicit van der Waals interactions were calculated. That is, these processes had no effect on the torsional potential values taken from the 2D potential map, so simulations based on this map exhibited greater transition energies than did the simulation based on explicit evaluation of van der Waals interactions. This tendency to overestimate transition energies will occur when CONDORR uses torsional potential maps that are generated directly from grid searches in which the residue geometry is fixed. In order to moderate this effect, another 2D potential map was prepared with van der Waals parameters m=9 and n=4.5 (Fig. 3d), and the simulation was repeated using potentials from this map. The resulting trajectory (Fig. 3e) was very similar to that obtained using explicit van derWaals potentials (Fig. 3c). It is likely that one could obtain significantly more accurate solutions by using a more sophisticated force field to generate maps that explicitly take into account the flexibility of the residue.

A 3D (φ, ψ, ω) map (Fig. 4a–c) was also generated and used as the basis for conformational trajectory calculations of cellobiose (β-D-Glc*p*-(1→4)-D-Glc*p* Fig. 1). As illustrated in Fig. 1, the model used in this case included three rigid units: two β-D-Glc*p* residues (**1** and **2**) lacking O6, H6a, and H6b; and a hydroxymethyl group (**3**). The map consists of 180 planes, each corresponding to a φ, ψ map (180×180 points), and reflects differences in the interaction of the non-reducing Glcp residue (BG) with the HM residue as it rotates about the C5–C6 bond (Fig. 4a–c). The potential energy contributions due to interactions between the HM residue and the reducing Glcp residue were calculated using a torsional potential function obtained by fitting the rotamer populations reported by Kirshner, et al. [9] A 10 ns simulation (10^7 steps, 1 fs per step) based on the resulting 3D map was completed in approximately 11 min. The analogous calculation performed with explicit calculation of van der Waals potentials at each step required over 114 min to complete. As described above for the application of 2D potential maps, the results of these two calculations (Fig. 4d,e) showed good agreement when the van der Waals potential used for generating the 3D map was softer (m=9, n=4.5)

than that used when the van der Walls potentials were explicitly evaluated (m=10, n=5). Thus, CONDORR trajectories obtained by using 2D or 3D potential energy surfaces reflect the energetic properties of that surface, and the harmonic potentials used to define the atomic superposition and bond angles have relatively minor effects, providing that these harmonic potentials constrain these geometric paramters to values close to their equlibrium values.

Linkages in which three bonds are free to rotate

A 3D torsional potential map was also generated for the flexible α-D-Xyl*p*-(1→6)-β-D-Glc*p* linkage found in the plant polysaccharide xyloglucan. As illustrated in Fig. 1, the model used in this case included five residues: two β-D-Glc*p* residues (**1** and **2**) lacking O6, H6a, and H6b; two hydroxymethyl groups (**3** and **4**); and a terminal α-D-Xyl*p* residue (**5**). The potential at each grid point was calculated using a torsional potential function based on the rotamer populations of Kirchner et al. [9] (see above), the exoanomeric torsional potential of Tvaroska, et al., [7] and modified (m=10, n=5, see above) van der Waals potentials for the interaction of atoms in all three residues. (As in the cellobiose model described above, van der Waals interactions between the hydroxymethyl group and the β-d-Glc*p* residue to which it belongs were not included, as these are accounted for by the fitted torsional potential for rotation around the C5–C6 bond.) The φ, ψ maps constituting different planes in the 3D potential map correspond to different orientations of the xylosyl residue and as such are significantly different, as expected (data not shown).

This approach was used to calculate an MD simulation for a branched tetrasaccharide β-D-Glc*p*-(1→4)-[α-D-Xyl*p*-(1→6)]-β-D-Glc*p*-(1→4)-β-D-Glc*p*. *Two different* 3D maps were generated for this calculation. The first map, analogous to that (described above) used for the simulation of cellobiose, describes the potential energy surface for the β-D-Glc*p*-(1→4)-β-D-Glc*p* linkage, including the effects due to variation in the orientation of the hydroxymethyl group of the aglyconic Glc*p* residue. The second is analogous to the 3D map (also described above) for the α-D-Xyl*p*-(1→6)-β-D-Glc*p* linkage, but does not include the torsional function that relates potential energy to rotation about the C5-C6 bond of the β-d-Glc*p* residue. Omitting this function while generating the second 3D map avoids redundant evaluation of the torsional potential corresponding to rotation about the C5–C6 bond of the central Glc*p* residue (which bears the α-D-Xyl*p* substituent at O6).

A molecular trajectory was calculated for the tetrasaccharide based on these 3D maps just described. Explicit (m=10, n=5) van der Waals potentials were included, but *only* for the interactions of the α-D-Xyl*p* residue with the two flanking β-D-Glc*p* residues. The results suggest, as expected, that the presence of the α-D-Xyl*p* residue inhibits rotation about the C5-C6 bond of its β-D-Glc*p* aglycon. However the results suggest that this residue has only a modest effect on the conformations of the adjacent β-D-Glc*p*-(1→4)-β-D-Glc*p* linkages. The Glc residue bearing the Xyl residue remained in the initial conformation (gg), so much longer simulations (perhaps in the μs range) would be required to analyze this system fully. Such extremely long simulations should be accessible with CONDORR, and should provide meaningful results if they employ potential energy maps that represent the potential energy surface more accurately. By using 3D potential energy maps, CONDORR can significantly increase the speed of rigid-residue simulations of such branched glycans, although satisfactory accuracy will most likely require evaluation of a limited number of specified (all atom) residue-residue interactions (e.g., between atoms of the Xyl*p* residue with those of the two flanking Glc*p* residues).

Cyclic oligosaccharides

As a first step in testing the applicability of CONDORR to conformationally constrained, macrocyclic oligosaccharides, models of a linear oligosaccharide containing eight (1→4)-linked α-D-Glc*p* residues were generated. A short (0.5-ns) MD calculation, performed at simulation temperature of 50 K using a 3D potential map for α-D-Glc*p*-(1→4)-α-D-Glc*p* linkages, correctly predicted that this oligomer adopts an amylose V-helix conformation, [10] allowing its reducing and non-reducing ends to approach each other. A second 0.5-ns simulation, performed at the same simulation temperature, included a weak harmonic potential (0.1 kcal mol^{-1} Å^{-1}) between O1 of the reducing residue and O4 of the non-reducing terminal residue. Simplified, united atom residue-residue interaction potentials were also included to insure that no two residues would simultaneously occupy the same location in Cartesian space. The final conformation of the resulting trajectory was used as the starting point for a third (0.1 ns) simulation in which the harmonic potential was increased to 1.0 kcal mol^{-1} Å^{-1}. This low-temperature CONDORR calculation resulted in a conformation that was readily cyclized by specifying a glycosidic bond involving the superposition of O1 of the reducing residue and O4 of the non-reducing end residue. The conformation of the cyclized oligomer was optimized by a simulated annealing calculation (0.5 ns, 5 K), which generated a highly symmetrical conformation, with a helical structure similar to an amylose V-helix. [10] A 10-ns, 300-K simulation was performed on the cyclized model solely on the basis of the 3D potential map without explicit evaluation of any residue–residue interactions, as cyclization introduced a constraint that precluded the steric interacion of residues that were not mutually linked. The resulting trajectory was characterized by limited motion of the glucosyl residues, whose glycosidic bonds remained in the initial conformational energy wells, along with relatively rapid rotation of the hydroxymethyl groups.

Effects of external constraints, forces, and torques

Dynamic models of a linear oligosaccharide containing 20 β-(1→4)-linked D-Glc*p* residues were based on a 3D potential energy map for cellobiose (m=10, n=5), in which hydroxymethyl groups are treated as independent residues. Several 100 ns simulations were performed at a temperature of 300 K. In the absence of any external constraints, the simulations predicted such an oligomer would adopt an extended but "twisted" conformation (φ=−70, ψ=120, corresponding to conformation 3, Fig. 3a), which is distinct from that of crystalline cellulose. However, glycosidic bonds in the oligomer occasionally adopted a "folded" conformation (φ=50, ψ=120, corresponding to conformation 4, Fig. 3a). Folded conformations may allow the cellulosic backbone to fold back upon itself under certain conditions [11].

A 100-ns simulation was performed in which an equal but opposite torque (10 pN nm) was applied to each end of the oligomer, which was otherwise free to move and rotate. This torque was intended to produce the same effect as the flattening out of an initially twisted segment (not explicitly included in the model) at the non-reducing end of the polymer. Such a flattening corresponds to the transition that is likely to occur when β-(1→4)-linked glucans bind to the surface of a cellulose microfibril, [3] which is a key process in the assembly of plant cell walls. This transition would result in a net rotation of backbone segments that are not in contact with the microfibril surface, which, in the presence of rotational constrains at the other end, could lead to "over-twisting" of the polymer (i.e., additional conformational divergence from the geometry of crystalline cellulose, which is characterized by a two-fold screw axis of symmetry). The torque for each terminal residue was defined such that its vector representation remained parallel (or antiparrallel) to the residue's C1–C4 vector, ensuring that it continued to twist rather than bend the backbone regardless of the instantaneous orientation of the residue. The resulting trajectory was characterized by an increased population of glycosidic bonds in the folded conformation (Fig. 3a, well 4) and a decrease in the population of bonds in an "undertwisted" conformation (Fig. 3a, well 2). This is not surprising, as any "overtwisting" due to the torque corresponds to an increase in the value of φ and/or ψ, as would a transition from the extended (but twisted) conformation (Fig. 3a, well 3) to a folded conformation (Fig. 3a, well 4), via the relatively low energy barrier in the potential energy surface that was used. The overall result suggests that the formation of folded or partially folded structures is compatible with the binding of a conformationally constrained β-(1→4)-linked glucan to the surface of a cellulose microfibril.

Additional simulations were performed in order to examine mechanical processes that could lead to relaxation of an "overtwisted" β-(1→4)-glucan that is subjected to tensile stress. In these simulations, the reducing end residue was held at its initial Cartesian coordinates and prevented from rotating while a force along a line parallel to the initial helical axis of the backbone was applied to the non-reducing end residue, which was also prevented from rotating. These forces and constraints were intended to reproduce the mechanical situation that would occur if segments at each end of the overtwisted polymer were bound to a different cellulose microfibril and the two microfibrils were being forced apart. This corresponds to frequently cited models of the expanding plant cell wall, in which a network of cellulose microfibrils crosslinked by xyloglucan tethers [3] are subjected to tensile stress of osmotic origin. (Xyloglucans have a cellulosic backbone.) The initial conformation of each glycosidic bond in the polymer was set to (φ=−60°, ψ=145°), corresponding to modest overtwising. Extension of such an overtwisted backbone under tensile stress would be accompanied by flattening of the backbone, as the most extended structure corresponds to a flat backbone (φ=−90°, ψ=90°) with a two-fold screw axis of symmetry, as observed in crystalline cellulose I. However, under the rotational constraints imposed on the terminal residues, transition to such a flat overall conformation would require a complete 360° rotation (swiveling) about at least one glycosidic bond. Such transitions are likely to be slow for a β-(1→4)-linked glucan compared to analogous rotations of the sugar-phosphate backbone of single-stranded DNA, in which adjacent deoxyribosyl residues are connected via multiple exocyclic atoms, leading to energetic barriers to rotation that are considerably lower than those exhibited by glycosidically linked pyranosyl residues. In fact, simulations where mild tension (23 pN) was applied, the overtwisted glucn did not elongate, consistent consistent with significant barriers to rotational relaxation. The end-to-end distance in these simulations actually decreased due to frequent occurrence of folded conformations (φ=50°, ψ=120°) in the glucan chain. Transitions of a glycosidic linkage to a folded conformation were probably driven by the overall energy decrease resulting from rotations that occurred when several overtwisted segments of the backbone relaxed to the most energetically preferred conformation (φ=−70°, ψ=120°). Under greater stress (92 pN), the distance between residues at the ends of the model glucan increased in jumps, consistent with an infrequent transition having a high free energy of activation.

Examination of the glucan trajectory under the constraints just described indicated that relaxation occurred by a 360 degree rotation of the glycosidic bond nearest the nonreducing end of the chain, which although constrained to prevent rotation, was free to move translationally. This suggested that motional constraints at the other end (the reducing end) were too strict, so additional 50-ns simulations were performed in which both ends were constrained to prevent rotation, but translational motion was allowed at both ends. An equal and opposite force (92 pN) was applied at each end to introduce tensile stress without actually preventing the residues from moving. This situation may more closely resemble the mechanical situation for crosslinks in the xyloglucan–cellulose network in plant cell walls, as the total contour

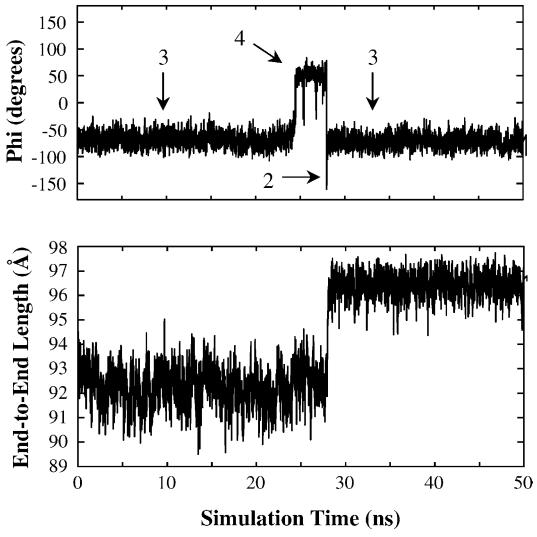

Fig. 5 Simulation of a model β-(1→4)-glucan containing 20 residues under a tensile stress of 195 pN. The end-to-end distance (*bottom*) and torsional angle φ for the glycosidic linkage between residues 16 and 17 of the sequence (*top*) undergo a simultaneous transition at about 28 ns into the simulation. The conformations adopted by the linkage are labeled according the conformational families shown in Fig. 2. That is, the linkage, initially in conformation 3, adopted conformation 4 for approximately 5 ns, and then abruptly rotated to conformation 2 for a very brief time, immediately rotating into conformation 3. This 360° swivel allowed the chain to relax and increase its end-to-end length by approximately 4.5 Å

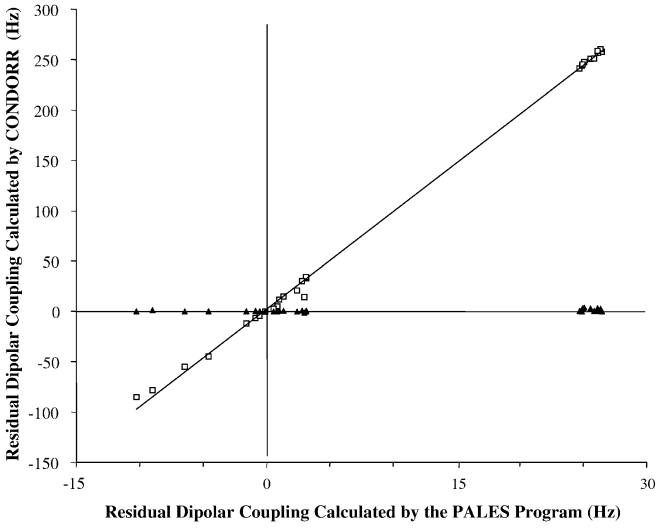

Fig. 6 Comparison of residual dipolar couplings (RDCs) calculated using orientation vectors from CONDORR simulations to values obtained by averaging RDCs obtained using the PALES program [14] operating on 10,000 conformations taken from the same simulation. Open squares indicate values taken from a CONDORR simulation that was spatially constrained to simulate anisotropic orientational ordering in a lipid-bicelle medium. Scaling effects, due to the requirement for a small distance between the restraining walls in the CONDORR simulation, result in a slope greater than one. Solid triangles represent the unconstrained (isotropic) CONDORR simulation, leading to a plot with zero slope, as expected

length of these tethers is estimated to be 30 nm (approximately 60 Glc*p* residues), [12] so central regions of the tether are likely to have significant translational freedom. Analysis of this simulation (Fig. 5) revealed that relaxation occurred via a transition pathway sequentially traversing conformations 4, 2, and 3 (Fig. 3a), with the rate limiting step being the transition from well 4 to well 2. Rotations via the alternative pathway, via conformation 1 were not observed.

Calculation of residual dipolar couplings (RDCs)

Effective use of CONDORR is not limited to analysis of large molecules. It is also useful when analyzing the motion of small molecules that are confined to spaces with very small dimensions. For example, liquid crystalline lipid bilayers (bicelles) are frequently used to align biomolecules partially for NMR studies. [13] Under these conditions, "residual dipolar couplings" (RDCs) between pairs of nuclei in the aligned biomolecule become observable, providing information regarding its internal molecular geometry. This approach is most readily applied to rigid or quasi-rigid molecules. However, flexible molecules, such as glycans, tend to be difficult to analyze by this approach. It can be especially difficult to apply these techniques to small, highly flexible glycans for which a unique molecular coordinate system cannot be readily defined. One approach that can be used in this situation is to treat each residue as a rigid entity and compare experimentally measured RDCs for each residue to ensemble average RDCs based on a molecular dynamics calculation. The net orientation of each residue (and indeed the entire molecule) depends on molecular conformation, which is constantly changing. Nevertheless, it is possible to calculate ensemble average RDCs by simultaneous simulation of the conformation and alignment order of the molecule in solution. This can be accomplished by simulating parallel planar barriers that confine the molecule, acting like the surface of an aligned bicelle. Valid results require simulations that are long enough to allow the molecule to sample a sufficient number of orientations under the constraints of its confinement.

Very long simulations (2 μs, 100,000 snapshots) of β-methyl-cellobioside in both constrained and unconstrained conditions were performed using a 3D potential map to constrain glycosidic bond geometry. The effects of an aligning medium were simulated by specifying two walls, 12 Å apart and parallel to the *y*,*z* plane of the MD coordinate system. Atoms in the model molecule were prevented from penetrating the wall by a harmonic potential with a gradient parallel to the *x*-axis. The computational cost of these potential energy terms was considered acceptable, as it increases in proportion to the number of atoms n, rather than to n^2, which is the case when all interatomic van der Waals potentials are calculated. Inclusion of planar constraints led to a non-uniform distribution of molecular orientations, which were ana-

lyzed by explicit calculation of the Saupe order tensor [13] for 20 pairs of nuclei in the model. As a result of the way the vectors were defined, each RDC is proportional to S_{zz}, an element on the diagonal of the order tensor. As the choice of the *z*-axis in the MD axis system is arbitrary, RDCs were calculated using S_{zz} values averaged over four different orientations for the magnetic field vector ($\mathbf{B}_0$), all orthogonal to the MD *x*-axis, but at angles of 0°, 45°, 90° and 135° with respect to the MD *z*-axis. For comparison, 10,000 molecular conformations taken from the simulation were individually analyzed by the PALES program, [14] which calculates RDCs for quasi rigid molecules. After correcting for an arbitrary scaling factor, the averages of RDCs obtained by PALES analysis were in good agreement with RDCs obtained by direct analysis of the spatially constrained CONDORR trajectory (diagonal line in Fig. 6). RDCs obtained by analysis of the unconstrained simulations were close to zero (horizontal line in Fig. 6), as expected for an isotropic net orientation.

Discussion

CONDORR provides a facile method for building molecular models using rigid residues, whose atomic coordinates are taken from standard PDB files. The CONDORR program implements molecular dynamics simulations using 6 degrees of freedom for each rigid, asymmetrical residue. Total energy and angular momentum are conserved when no stochastic or external forces are applied to the model, if the time step is kept sufficiently short (less than 1 fs). Longer time steps can be used if a Langevin dynamics approach is used to maintain the model temperature. This method also allows conformational transitions to occur at a faster rate than if no stochastic forces are applied.

The general methods implemented in CONDORR were tested using glycosyl residues as rigid units. One reason for choosing this system is that glycosyl residues are relatively rigid and there has been some work in the past examining the energetics of the conformational space defined by rotation around the glycosidic bonds, which are flexible. Although significant computational savings are achieved by CONDORR, they are offset by a decrease in the accuracy of the dynamics simulations it performs. The force fields currently used by CONDORR (described herein) are extremely simple, but provide a basis for evaluating the overall properties of the algorithm in terms of its ability to generate molecular dynamics trajectories that reflect the 2D and 3D potential surfaces supplied to it. It also provides a convenient platform for assessing and applying more accurate energy surfaces, and as such will be a useful tool that will facilitate the development of such surfaces. Such improved parameterizations could include electrostatics, solvent effects, and other factors not considered in the simple force field currently used by CONDORR. In addition, the CONDORR algorithm can be used even with its present, highly simplified force field to explore the complex energy surfaces of large, constrained systems in order to locate low-energy regions that can be used as the starting points for more accurate, fully atomistic dynamics calculations.

The premise underlying CONDORR is that groups of atoms can be treated as a single entity whose position and orientation can vary with time, while its internal geometry remains constant. This is an extension of a frequently used approach of "freezing" parts of a large molecular structure while allowing other parts to undergo molecular motion. This approach is often used, for example, to model the interaction of a ligand with a larger structure, such as a protein, allowing only those portions of the protein that interact with the ligand to move, significantly decreasing the computational cost of the simulation. The disadvantage of this approach, as it is usually implemented, is that the frozen portions of the structure cannot move or rotate relative to the molecular dynamics frame of reference, which can significantly affect how the frozen parts interact with each other, with other parts of the structure, or with other molecules in the simulation. For example, two domains in a large structure that are connected by a flexible hinge could be frozen to decrease the computational costs, but doing so would not allow them to move relative to each other. This limitation can be overcome by the techniques that are implemented in CONDORR (i.e., defining each domain as a rigid unit and allow the two domains to move relative to the molecular dynamics frame, to each other, and to other components of the system). Thus, if interaction with the ligand depends on the relative orientation and position of the two domains, CONDORR methods would allow their relative positions to change, thereby modulating the interaction with the ligand. This could be done with a significant decrease in computational expense, relative to fully atomistic simulations of the system.

Although the methods implemented in CONDORR provide a convenient means of simulating the motions of quasi-rigid entities in such a hinged (or otherwise constrained) system, they do not preclude the application of an atomistic approach to other components in the system. That is, they could provide (at minimal computational cost) appropriate environmental or geometric constraints for the precise, high-level analysis of a small part of a very large dynamic system. Such hybrid calculations would still provide a significant saving in computational costs. For example, a full atomistic analysis of two domains, each with n atoms, would involve $(2n)^2$ atom–atom interactions. However, if each domain is considered rigid, a "semi-atomistic" calculation of their interaction (using a highly refined force field), in which energies are calculated for every interaction between an atom in one rigid domain and an atom in the other rigid domain, would requires the evaluation of only n^2 atom-atom interactions, a four-fold saving in computational expense. Further savings would result from parameterization (e.g., by using a potential energy map) for the interaction of the two domains. CONDORR represents an initial step in the direction of performing such calculations, which are likely to be necessary when systems are simulated at a

"meta-scale" (i.e., for large molecular assemblies that contain millions of atoms.) Although such calculations would lack a high degree of molecular detail, they could nevertheless provide important information regarding the complex interactions that are inherent in such large structures. Hybrid calculations, in which the energetics of only small portions of the system are evaluated with high precision, could be extremely valuable in understanding its overall dynamic properties. An example of such a meta-scale calculation would be a simulation of the dynamic response of the cellulose-xyloglucan network to osmotic stress in the primary walls of expanding plant cells. It is not possible to simulate this fundamentally important process using fully atomistic calculations, even with the most powerful parallel processing systems that are currently available.

Simulations using the basic CONDORR force field suggest that cellulosic polysaccharides can form folded structures under torsional stress and suggest specific pathways by which these structures can relax under longitudinal stress. It should be noted that the simulations of cellobiose and cello-oligomers reported by Hardy and Sarko [11] also suggested that folded structures spontaneously occur in the backbone of β-(1→4)-linked glucans, although these were attributed to transient residence in energy well 1, rather than well 4, as described here. Their all-atom calculations were based on more sophisticated force fields, but were limited to small oligomers that were not subjected to stress or externally applied torques. A comprehensive analysis of the behavior of β-(1→4)-linked glucans under torsional and tensile stress will depend on much longer calculations performed using more accurate force fields than those applied in this study. The computational approach embodied in the CONDORR algorithm will facilitate such simulations once more realistic energy surfaces become available.

Very long simulations (2 ms) of cellobiose were also performed on a standard PC running at 1.1 MHz. These simulations made it possible to explicitly calculate a Saupe order tensor (and the corresponding RDCs) [13] for cellobiose in an anisotropic medium. This is computationally challenging, as the model compound must undergo sufficient diffusional and rotational motion so as to adequately sample orientational space in the presence of a barrier (wall). The results obtained using CONDORR agree well with those obtained by specific application of the PALES [14] program to each of 10^4 conformers selected from the trajectory, and averaging of the resulting RDCs.

CONDORR's strength lies in its ability to use various combinations of 2D and 3D potential energy maps, along with selective application of a limited number of judiciously chosen van der Waals interactions to rapidly evaluate the conformational potential of the model. CONDORR will readily accommodate energy maps generated by methods (e.g., quantum mechanical or molecular mechanics calculations) that allow conformational relaxation of rigid units, thereby accounting for their inherent flexibility. It is likely that the calculation of such maps will itself be computational expensive. For example, one approach would be to perform a free energy calculation (based on MD or Monte Carlo methods) for each vertex in the map. Initially, this kind of map would have lower resolution (fewer data points) than those described herein. In fact, CONDORR simulations calculated using low resolution maps (15 degrees per point) are similar to those obtained with high resolution maps (4 degrees per point). Thus, highly accurate maps should provide valuable information when used as a basis for CONDORR-type calculations.

Acknowledgments The authors would like to thank Dr. Robert J. Woods and Dennis Elking of the CCRC for stimulating conversations and suggestions. This research is supported by funds from the National Science Foundation grant MCB-9974673. The DOE Center for Plant and Microbial Complex Carbohydrates is supported by U.S. Department of Energy Center for DE-FG02-93ER20097.

Appendix A: conservation of energy in CONDORR

Two point masses

Two "structureless" particles with masses m_1 and m_2 are located at coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$. Define a particle separation vector $\mathbf{s}\equiv\mathrm{r}_1-\mathrm{r}_2$ and a separation velocity $\dot{\mathbf{s}} \equiv \dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2$. The interaction force $\mathbf{F}_1$ on particle 1 is equal and opposite to the force $\mathbf{F}_2$ on particle 2, so the subscripts for the force vectors can be dropped if a force $\mathrm{F}\equiv\mathrm{F}_1=-\mathrm{F}_2$ is defined. For a pair of structureless particles, the potential energy due to their interaction depends solely on the interaction distance, so $\mathbf{F}$ is parallel to $\mathbf{s}$. To demonstrate conservation of energy, it is necessary and sufficient to show that $\mathrm{d}H/\mathrm{d}T=0$, where $H=T+U$ is the classical Hamiltonian (T is the kinetic energy and U is the potential energy).

The kinetic energy of the system is

$$T = \sum_{i=1,2} T_i = \sum_{i=1,2} \frac{m_i \dot{\mathbf{r}}_i^2}{2}$$

According to Newton's second law, $\mathbf{F}=m\mathbf{a}$, where $\mathbf{F}$ is the force and $\mathbf{a}$ is the acceleration. The acceleration of each particle is

$$\mathbf{a}_i = \ddot{\mathbf{r}}_i = \frac{\mathbf{F}}{m_i}$$

The time-derivative of the kinetic energy is

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \sum_{i=1,2} \frac{\mathrm{d}}{\mathrm{d}t} \frac{m_i \dot{\mathbf{r}}_i^2}{2} = \sum_{i=1,2} \frac{m_i}{2} \frac{\mathrm{d}\dot{\mathbf{r}}_i^2}{\mathrm{d}\dot{\mathbf{r}}_i} \frac{\mathrm{d}\dot{\mathbf{r}}_i}{\mathrm{d}t} = \sum_{i=1,2} \frac{m_i}{2} 2\dot{\mathbf{r}}_i \cdot \frac{\mathrm{d}\dot{\mathbf{r}}_i}{\mathrm{d}t}$$

$$= \sum_{i=1,2} m_i \dot{\mathbf{r}}_i \frac{\mathbf{F}_i}{m_i} = \sum_{i=1,2} \dot{\mathbf{r}}_i \cdot \mathbf{F}_i = (\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2) \cdot \mathbf{F} = \dot{\mathbf{s}} \cdot \mathbf{F}$$

The time-derivative of the potential energy is

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} \dot{\mathbf{s}}$$

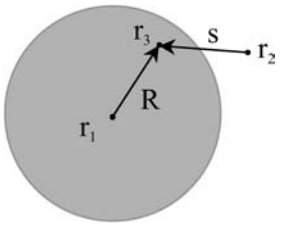

Fig. 7 Geometry of the interaction of an atom within a rigid residue with an atom outside the residue

Conservation of energy requires that

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\frac{\mathrm{d}U}{t}$$

Then,

$$\dot{\mathbf{s}} \cdot \mathbf{F} = -\frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} \cdot \dot{\mathbf{s}}$$

$$\mathbf{F} = -\frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} = -\nabla_s U \equiv -\left[\mathbf{i}\frac{\partial U}{\partial s_x} + \mathbf{j}\frac{\partial U}{\partial s_y} + \mathbf{k}\frac{\partial U}{\partial s_z}\right]$$

where ∇_s is the gradient with respect to **s**, and **i**, **j**, and **k** are unit vectors parallel to the *x*, *y* and *z*-axes. This simple, well established relationship is the basis for molecular dynamics algorithms that are limited to the interaction of "structureless" particles (i.e., point masses with no angular momentum).

The interaction of an atom within a rigid residue with an atom outside the residue

The residue is idealized as a sphere, with a mass m_1, a moment of intertia I, and a center of mass $\mathbf{r}_1$ (see Fig. 7). Here, the external atom is treated as a point mass m_2 at coordinates $\mathbf{r}_2$. The kinetic energy of the system, including that due to rotation of the residue is

$$T = \sum_{i=1,2} T_i = \frac{I\omega^2}{2} + \sum_{i=1,2} \frac{m_i \dot{\mathbf{r}}_i^2}{2}$$

The angular acceleration of the residue is $\dot{\omega} = \Gamma / I$, where Γ is a torque. The time derivative of the kinetic energy of the system can be expressed in a form analogous to that given above for two point masses. That is, the angular velocity ω can be treated the same way that the linear velocity $\dot{\mathbf{r}}$ was treated in the above derivation.

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{I\omega^2}{2} + \sum_{i=1,2} \frac{\mathrm{d}}{\mathrm{d}t}\frac{m_i \dot{\mathbf{r}}_i^2}{2} = \omega \cdot \Gamma + (\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2) \cdot \mathbf{F}$$

It is important to note that this equation defines **F** as the force *on the residue at its center of mass*. (Conservation of linear momentum requires that the force on the external atom is -**F**.) Furthermore, the separation vector $\mathbf{s} \equiv \mathbf{r}_3 - \mathbf{r}_2$ is defined with reference to the two atoms rather than the residue's center of mass. Accounting for both translation and rotation, the separation velocity of the atoms is

$$\dot{\mathbf{s}} = \dot{\mathbf{r}}_3 - \dot{\mathbf{r}}_2 = [\dot{\mathbf{r}}_1 + (\omega \times \mathbf{R})] - \dot{\mathbf{r}}_2$$

where **R** is the vector from the center of mass of the residue to the interacting atom within the residue. The time derivative of the potential energy is thus

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}}\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} \cdot \dot{\mathbf{s}} = \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} \cdot [\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2 + (\omega \times \mathbf{R})]$$

Conservation of energy requires that

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\frac{\mathrm{d}U}{\mathrm{d}t}$$

$$\dot{\omega} \cdot \Gamma + (\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2) \cdot \mathbf{F} = -[\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2 + \omega \times \mathbf{R}] \cdot \frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}}$$

$$\dot{\omega} \cdot \Gamma + (\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2) \cdot \mathbf{F} = [\dot{\mathbf{r}}_1 - \dot{\mathbf{r}}_2 + \omega \times \mathbf{R}] \cdot \mathbf{F}$$

$$\omega \cdot \Gamma = \omega \times \mathbf{R} \cdot \mathbf{F}$$

$$\omega \cdot \Gamma = \omega \cdot \mathbf{R} \times \mathbf{F}$$

$$\Gamma = \mathbf{R} \times \mathbf{F}$$

where a substitution was made based on the assertion (derived above) that $-\frac{\mathrm{d}U}{\mathrm{d}\mathbf{s}} = \mathbf{F}$ and where the triple scalar product on the right hand side of the fourth equation was replaced with an equivalent expression.

These equations show that conservation of energy for the interaction of an external object with an atom within a rigid residue requires two applications of the interaction force **F** to the residue: (1) **F** must be applied to the residue at its center of mass, affecting its linear momentum; (2) a torque $\Gamma = \mathbf{R} \times \mathbf{F}$ must be applied to the residue, affecting its angular momentum. The opposite force -**F** must also be applied to the external object. It is not necessary to apply any torque to the external object if it is a point mass. However, if the external object is an atom within another rigid residue, the second residue must be treated in exactly the same way as the first, except that a force equal to **–F** must be used.

This recipe also conserves both linear and angular momentum. Conservation of linear momentum **P** is explicitly invoked in the derivation, as $\mathbf{F} \equiv \mathbf{F}_1 = -\mathbf{F}_2$. Accordingly,

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \frac{\mathrm{d}}{t}(m_1\mathbf{r}_1 + m_2\mathbf{r}_2) = m_1\frac{\mathbf{F}_1}{m_1} + m_2\frac{\mathbf{F}_2}{m_2} = \mathbf{F}_1 + \mathbf{F}_2 = 0$$

The angular momentum **L** of the system depends on the rotation of the residue ($\mathbf{L}_{\mathrm{rot}}$) and the relative motion of the objects ($\mathbf{L}_{\mathrm{trans}}$). That is,

$$\mathbf{L} = \mathbf{L}_{\mathrm{rot}} + \mathbf{L}_{\mathrm{trans}}$$

Conservation of angular momentum requires that

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{L}_{\mathrm{rot}}}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{L}_{\mathrm{trans}}}{t} = 0$$

By definition, the time derivative of $\mathbf{L}_{\mathrm{rot}}$ is $\mathrm{d}\mathbf{L}_{\mathrm{rot}}/\mathrm{d}t = \Gamma$. The time derivative of $\mathbf{L}_{\mathrm{trans}}$ can be defined relative to the center of mass **c** of the system, as illustrated in Fig. 8. The center of mass **c** is

$$\mathbf{c} = \frac{\sum_{\alpha=x,y,z} \mathbf{u}_\alpha \left(m_1 r_{1,\alpha} + m_2 r_{2,\alpha}\right)}{m_1 + m_2}$$

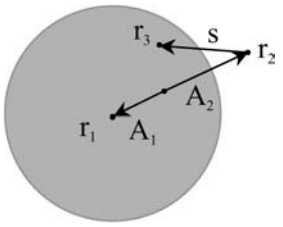

Fig. 8 Geometry defining $\mathbf{L}_{\text{trans}}$ relative to the center of mass of the system

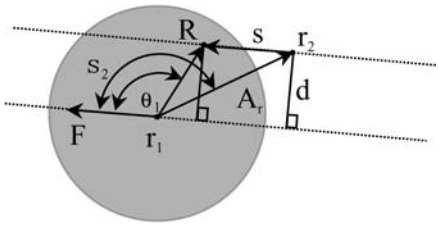

Fig. 9 Geometric representation of the conservation of angular momentum

where $\mathbf{u}_\alpha$ are unit vectors along the x, y, and z-axes. Define two vectors $A_1 \equiv \mathbf{r}_1 - \mathbf{c}$ and $A_2 \equiv \mathbf{r}_2 - \mathbf{c}$, as shown in Fig. 8. Substituting the expression for $\mathbf{c}$ into these definitions, expanding, and recollecting terms yields the following expressions.

$$\mathbf{A}_1 = \frac{m_2}{m_1 + m_2}(\mathbf{r}_1 - \mathbf{r}_2)$$

$$\mathbf{A}_2 = \frac{m_1}{m_1 + m_2}(\mathbf{r}_2 - \mathbf{r}_1)$$

Then, $\mathbf{L}_{\text{trans}}$ can be expressed

$$\begin{aligned}\mathbf{L}_{\text{trans}} &= m_1(\mathbf{A}_1 \times \dot{\mathbf{r}}_1) + m_2(\mathbf{A}_2 \times \dot{\mathbf{r}}_2) \\ &= \frac{m_1 m_2}{m_1 + m_2}[(\mathbf{r}_1 - \mathbf{r}_2) \times \dot{\mathbf{r}}_1 + (\mathbf{r}_2 - \mathbf{r}_1) \times \dot{\mathbf{r}}_2] \\ &= \mu[\mathbf{A}_r \times \dot{\mathbf{r}}_2 - \mathbf{A}_r \times \dot{\mathbf{r}}_1] \\ &= \mu \mathbf{A}_r \times (\dot{\mathbf{r}}_2 - \dot{\mathbf{r}}_1) \\ &= \mu \mathbf{A}_r \times \dot{\mathbf{A}}_r\end{aligned}$$

where a new vector $A_r \equiv \mathbf{r}_2 - \mathbf{r}_1$ has been defined and the reduced mass $\mu \equiv m_1 m_2/(m_1+m_2)$ have been substituted into the expressions. The time-derivative of $\mathbf{L}_{\text{trans}}$ is then

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{L}_{\text{trans}} &= \frac{\mathrm{d}}{\mathrm{d}t}\mu[\mathbf{A}_r \times \dot{\mathbf{A}}_r] \\ &= \mu\left[\left(\frac{\mathrm{d}\mathbf{A}_r}{\mathrm{d}t} \times \dot{\mathbf{A}}_r\right) + \left(\mathbf{A}_r \times \frac{\mathrm{d}\dot{\mathbf{A}}_r}{\mathrm{d}t}\right)\right] \\ &= \mu\left[(\dot{\mathbf{A}}_r \times \dot{\mathbf{A}}_r) + \left(\mathbf{A}_r \times \frac{\mathrm{d}\dot{\mathbf{A}}_r}{\mathrm{d}t}\right)\right] \\ &= \mu\left[\mathbf{A}_r \times \frac{\mathrm{d}\dot{\mathbf{A}}_r}{\mathrm{d}t}\right] \\ &= \mu\left[\mathbf{A}_r \times \frac{\mathrm{d}}{\mathrm{d}t}(\dot{\mathbf{r}}_2 - \dot{\mathbf{r}}_1)\right] \\ &= \mu\left[\mathbf{A}_r \times \left(\frac{\mathbf{F}_2}{m_2} - \frac{\mathbf{F}_1}{m_1}\right)\right] \\ &= -\mu\left[\mathbf{A}_r \times \left(\frac{m_1\mathbf{F}}{m_1 m_2} + \frac{m_2\mathbf{F}}{m_1 m_2}\right)\right] \\ &= -\mathbf{A}_r \times \mathbf{F}\end{aligned}$$

Here, the product rule was invoked, the cross product of vector $\dot{\mathbf{A}}_r$ with itself was identified as zero, the vector $\dot{\mathbf{A}}_r$ was expanded, the acceleration of each object was expressed in terms of $\mathbf{F}=\mathbf{F}_1=-\mathbf{F}_2$, and the reduced mass μ was cancelled. Appropriately, the expression for the time derivative of angular momentum is independent of the masses of the objects. Conservation of momentum requires that

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{L}_{\text{trans}}}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{L}_{\text{rot}}}{\mathrm{d}t} = 0$$

Substituting and recalling that $\Gamma = \mathbf{R} \times \mathbf{F}$,

$$\begin{aligned}\frac{\mathrm{d}\mathbf{L}_{\text{trans}}}{t} &= -\frac{\mathrm{d}\mathbf{L}_{\text{rot}}}{\mathrm{d}t} \\ -\mathbf{A}_r \times \mathbf{F} &= -\Gamma \\ \mathbf{A}_r \times \mathbf{F} &= \mathbf{R} \times \mathbf{F}\end{aligned}$$

The angular momentum $\mathbf{L}$ is conserved only if $\mathbf{F}$ is defined such that the last equation is true.

This can be shown geometrically with reference to Fig. 9. The force vector $\mathbf{F}$ is always parallel to the line (defined by $\mathbf{s}$) passing through the two atoms. Another line parallel to the line defined by $\mathbf{s}$ is drawn through the center of mass of the spherical residue. The vectors $\mathbf{F}$, $\mathbf{R}$ and $\mathbf{A}_r$ are all in the plane defined by these two lines, so the two cross products of interest are both normal to this plane and thereby parallel. The length of each of these cross products is equal to the distance d between the two lines. That is,

$$|\mathbf{R} \times \mathbf{F}| = |\mathbf{R}||\mathbf{F}| \sin\theta_1 = d$$

$$|\mathbf{A}_r \times \mathbf{F}| = |\mathbf{A}_r||\mathbf{F}| \sin\theta_2 = d$$

As the two vectors of interest are parallel and have the same length, they are equal, demonstrating that angular momentum is conserved.

Evaluation of torsional potentials for the interaction of two residues

The torsional angle φ is specified by four atoms, located at coordinates $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ (see Fig. 10). Atoms 1, 2, and 3 are constituents of residue 1, whose center of mass is at $\mathbf{r}_1$, and atom 4 is a constituent of residue 2, whose center of mass is at $\mathbf{r}_2$. (Two copies of the third atom

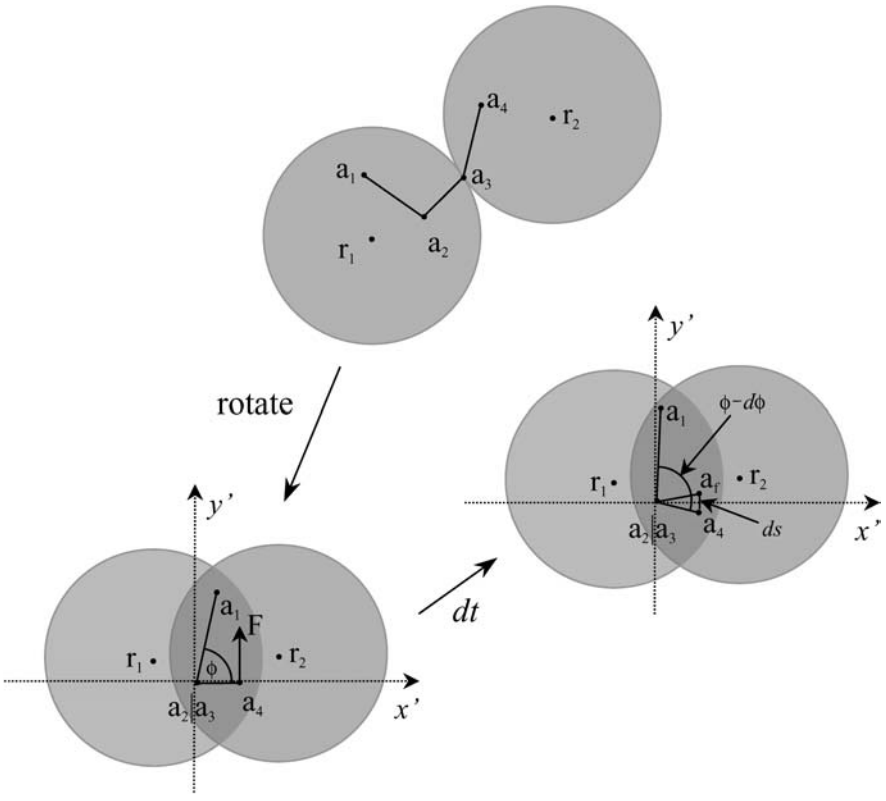

Fig. 10 Geometric representation of the fictitious atom and fictitious force used in CONDORR

exist, as this atom is shared by the two residues. However, coordinates of the atom 3 copy associated with residue 1 are used for calculating torsional angles and potentials.) As illustrated in Fig. 10, it is possible to find a coordinate transformation that results in a new system in which atom 2 is at the origin, atom 3 is on the z'-axis, and atom 4 is in the x',z'-plane. The coordinate system is chosen such that the z'-coordinate of atom 3 and the x'-coordinate of atom 4 are both positive. Unit vectors parallel to the primed axes can be calculated as follows. The vector $\mathbf{A}_3=\mathbf{a}_3-\mathbf{a}_2$ is parallel to the z'-axis, so the unit vector $\mathbf{u}_z = \mathbf{A}_3/|\mathbf{A}_3|$ defines the z'-axis. The vector $\mathbf{A}_4=\mathbf{a}_4-\mathbf{a}_3$ is in x',z'-plane, so the vector $\mathbf{A}_y=\mathbf{u}_z\times\mathbf{A}_4$ is parallel to the y'-axis and the unit vector $\mathbf{u}_y = \mathbf{A}_y/|\mathbf{A}_y|$ defines the y'-axis. The unit vector $\mathbf{u}_x=\mathbf{u}_y\times\mathbf{u}_z$ defines the x'-axis. The vector $\mathbf{A}_\mathrm{p}$ is defined as the projection of $\mathbf{A}_4$ onto the x'-axis. Vectors $\mathbf{A}_\mathrm{p}$ and $\mathbf{A}_\mathrm{y}$ have the same length l, which is

$$\begin{aligned}\ell &= |\mathbf{A}_p| = \mathbf{u}_x \cdot \mathbf{A}_4 = |\mathbf{u}_x||\mathbf{A}_4|\cos(\theta)\\ &= |\mathbf{A}_y| = |\mathbf{u}_z \times \mathbf{A}_4| = |\mathbf{u}_z||\mathbf{A}_4|\sin\left(\frac{\pi}{2}-\theta\right)\\ &= |\mathbf{u}_z||\mathbf{A}_4|\cos(\theta)\end{aligned}$$

where θ is the angle between $\mathbf{A}_4$ and the x'-axis.

Define a fictitious atom f with coordinates $\mathbf{a}_f=\mathbf{a}_4$. A fictitious force $\mathbf{F}$ parallel to the y'-axis acting on atom f (rigidly attached to residue 1) and the equal and opposite force $-\mathbf{F}$ acting on atom 4 (rigidly attached to residue 2) for a time dt will cause atom f and atom 4 (initially coincident) to become separated by a distance ds, changing the dihedral angle φ by an increment $\mathrm{d}\phi = \frac{\mathrm{d}s}{l}$. The potential energy gradient dU/ds associated with the fictitious force can be expressed in terms of the torsional potential gradient dU/dϕ:

$$\frac{\mathrm{d}U}{\mathrm{d}s} = \frac{U}{\mathrm{d}\phi}\frac{d\phi}{\mathrm{d}s} = \frac{\mathrm{d}U}{\mathrm{d}\phi}\frac{1}{l}$$

The force $\mathbf{F}$ is parallel to $\mathbf{u}_y$, and has a magnitude of $-$dU/ds. That is,

$$\begin{aligned}\mathbf{F} &= -\frac{\mathrm{d}U}{\mathrm{d}s}\mathbf{u}_y = -\frac{\mathrm{d}U}{\mathrm{d}\phi}\frac{1}{l}\mathbf{u}_y = -\frac{\mathrm{d}U}{\mathrm{d}\phi}\frac{1}{|\mathbf{A}_y|}\frac{\mathbf{A}_y}{|\mathbf{A}_y|}\\ &= -\frac{\mathrm{d}U}{\mathrm{d}\phi}\frac{\mathbf{A}_y}{\mathbf{A}_y\cdot\mathbf{A}_y}\end{aligned}$$

where $\mathbf{A}_y=\mathbf{u}_z\times\mathbf{A}_4$, as defined above.

Fictitious forces calculated in this way can be treated as interaction forces (described above) involving pairs of atoms on different residues. (Fictitious atom f is associated with residue 1 and atom 4 is associated with residue 2.) Doing so preserves the total energy, linear momentum, and angular momentum of the system.

Atomic interaction potentials

Atomic interaction potentials have the form:

$$V = Q\varepsilon\left[\left(\frac{\sigma}{r}\right)^m - \left(\frac{\sigma}{r}\right)^n\right]$$

where Q, σ, m, and n are constants, $-\epsilon$ is the potential at its minimum value, and r is the distance between atoms.

Define a distance $r_{\min}$ where the potential is minimized. That is,V=$V_{\min}$=$-\epsilon$ when r=$r_{\min}$. Then,

$$\frac{V}{\mathrm{d}r} = Q\varepsilon\frac{\mathrm{d}}{\mathrm{d}r}\left[\left(\frac{\sigma}{r}\right)^m - \left(\frac{\sigma}{r}\right)^n\right]$$
$$= Q\varepsilon\left[-m\sigma^m\left(\frac{1}{r}\right)^{m+1} + n\sigma^n\left(\frac{1}{r}\right)^{n+1}\right]$$
$$= \frac{Q\varepsilon}{r}\left[-m\left(\frac{\sigma}{r}\right)^m + n\left(\frac{\sigma}{r}\right)^n\right]$$

Define the ratio ρ=m/n and note that the derivative is zero at the minimum,

$$\frac{\mathrm{d}V_{\min}}{r} = \frac{Q\varepsilon}{r_{\min}}\left[-\rho n\left(\frac{\sigma}{r_{\min}}\right)^{\rho n} + n\left(\frac{\sigma}{r_{\min}}\right)^n\right] = 0$$
$$\rho n\left(\frac{\sigma}{r_{\min}}\right)^{\rho n} = n\left(\frac{\sigma}{r_{\min}}\right)^n$$
$$\left(\frac{\sigma}{r_{\min}}\right)^{(\rho-1)n} = \frac{1}{\rho}$$
$$\sigma^{(\rho-1)n} = r_{\min}^{(\rho-1)n}\rho^{-1}$$
$$\sigma = r_{\min}\rho^{\frac{-1}{(\rho-1)n}}$$

Then,

$$V = Q\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\rho n} - \left(\frac{\sigma}{r}\right)^n\right]$$
$$= Q\varepsilon\left[\left(\rho^{\frac{-1}{(\rho-1)n}}\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-1}{(\rho-1)n}}\frac{r_{\min}}{r}\right)^n\right]$$
$$= Q\varepsilon\left[\left(\rho^{\frac{-\rho n}{(\rho-1)n}}\right)\left(\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-n}{(\rho-1)n}}\right)\left(\frac{r_{\min}}{r}\right)^n\right]$$
$$= Q\varepsilon\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-1}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^n\right]$$

Define ϵ such that when r=$r_{\min}$, then V=$-\epsilon$, so

$$-\varepsilon = Q\varepsilon\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right)\left(\frac{r_{\min}}{r_{\min}}\right)^{\rho n} - \left(\rho^{\frac{-1}{\rho-1}}\right)\left(\frac{r_{\min}}{r_{\min}}\right)^n\right]$$
$$\frac{-1}{Q} = \left[\left(\rho^{\frac{-\rho}{\rho-1}}\right) - \left(\rho^{\frac{-1}{\rho-1}}\right)\right]$$
$$Q = -\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right) - \left(\rho^{\frac{-1}{\rho-1}}\right)\right]^{-1}$$

The atomic interaction energies can be calculated by substituting this value of Q into the general potential expression derived above:

$$V = Q\varepsilon\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-1}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^n\right]$$
$$= \varepsilon\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right) - \left(\rho^{\frac{-1}{\rho-1}}\right)\right]^{-1}\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-1}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^n\right]$$

For example, for the familiar Lennard-Jones 6–12 potential, m=12, n=6, ρ=2, and Q=4. That is,

$$V_{6-12} = 4\varepsilon\left[\left(2^{\frac{-2}{2-1}}\right)\left(\frac{r_{\min}}{r}\right)^{12} - \left(2^{\frac{-1}{2-1}}\right)\left(\frac{r_{\min}}{r}\right)^6\right]$$
$$= 4\varepsilon\left[\frac{1}{4}\left(\frac{r_{\min}}{r}\right)^{12} - \frac{1}{2}\left(\frac{r_{\min}}{r}\right)^6\right]$$
$$= \varepsilon\left[\left(\frac{r_{\min}}{r}\right)^{12} - 2\left(\frac{r_{\min}}{r}\right)^6\right]$$

The force F due to a potential of this general type is

$$F = \frac{-\mathrm{d}V}{\mathrm{d}r}$$
$$= -Q\varepsilon\frac{\mathrm{d}}{\mathrm{d}r}\left[\left(\rho^{\frac{-\rho}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^{\rho n} - \left(\rho^{\frac{-1}{\rho-1}}\right)\left(\frac{r_{\min}}{r}\right)^n\right]$$
$$= -Q\varepsilon\left(\rho^{\frac{-\rho}{\rho-1}}\right)r_{\min}^{\rho n}\frac{\mathrm{d}r^{-\rho n}}{\mathrm{d}r} + Q\varepsilon\left(\rho^{\frac{-1}{\rho-1}}\right)r_{\min}^n\frac{\mathrm{d}r^{-n}}{\mathrm{d}r}$$
$$= -Q\varepsilon\left(\rho^{\frac{-\rho}{\rho-1}}\right)r_{\min}^{\rho n}\left(-\rho n r^{-\rho n-1}\right)$$
$$+Q\varepsilon\left(\rho^{\frac{-1}{\rho-1}}\right)r_{\min}^n\left(-nr^{-n-1}\right)$$
$$= -Q\varepsilon\left(\rho^{\frac{-\rho}{\rho-1}}\right)r_{\min}^m\left(-mr^{-m-1}\right)$$
$$+Q\varepsilon\left(\rho^{\frac{-1}{\rho-1}}\right)r_{\min}^n\left(-nr^{-n-1}\right)$$
$$= Q\varepsilon m\left(\rho^{\frac{-\rho}{\rho-1}}\right)r_{\min}^m r^{-m-1} - Q\varepsilon n\left(\rho^{\frac{-1}{\rho-1}}\right)r_{\min}^n r^{-n-1}$$

Defining the constants

$$A_F = Q\varepsilon m\left(\rho^{\frac{-\rho}{\rho-1}}\right)r_{\min}^m \quad B_F = Q\varepsilon n\left(\rho^{\frac{-1}{\rho-1}}\right)r_{\min}^n$$

makes it possible to write a simple expression for the force F:

$$F = A_F r^{-m-1} - B_F r^{-n-1}$$

Interpolation of potential surfaces and surface gradients

The energy surface $\left(U_{\Omega_1,\Omega_2}\right)$ is approximated as a function of (Ω_1, Ω_2) at discrete points (ω_i,ω_j), arranged in a square pattern, allowing the coordinates *at each vertex* to be specified using only one index (i or j). Define an inverse distance η between the vertices,

$$\eta \equiv \frac{1}{\Delta\omega} = \frac{1}{\omega_{i+1} - \omega_i} = \frac{1}{\omega_{j+1} - \omega_j}$$

Define "edge gradients" $\sigma_{a,i,j}$ along the lines connecting adjacent vertices:

$$\sigma_{1,i,j} \equiv \left(\frac{\Delta U}{\Delta\omega_i}\right)_{i,j} \equiv \frac{U_{i+1,j} - U_{i,j}}{\omega_{i+1} - \omega_i} = \left(U_{i+1,j} - U_{i,j}\right)\eta$$

$$\sigma_{2,i,j} \equiv \left(\frac{\Delta U}{\Delta \omega_j}\right)_{i,j} \equiv \frac{U_{i,j+1} - U_{i,j}}{\omega_{j+1} - \omega_j} = \left(U_{i,j+1} - U_{i,j}\right)\eta$$

For a specific point at coordinates (Ω_1,Ω_2) within the square area bounded by points (ω_i,ω_j) and (ω_{i+1},ω_{j+1}), the gradient component $(\partial U/\partial \Omega_1)$ can be approximated by calculating the values $U^{rel}_{\omega_i,\Omega_2}$ and $U^{rel}_{\omega_{i+1},\Omega_2}$ at coordinates (ω_i,Ω_2) and (ω_{i+1},Ω_2) *relative* to the value $U_{i,j}$ and making a linear extrapolation of U between these points. That is,

$$\begin{aligned}\frac{\partial U}{\partial \Omega_1} &\approx \frac{U^{rel}_{\omega_{i+1},\Omega_2} - U^{rel}_{\omega_i,\Omega_2}}{\Delta\omega_i} \\ &= \frac{\sigma_{1,i,j}\cdot\Delta\omega_i + (\Omega_2 - \omega_j)\sigma_{2,i+1,j} - (\Omega_2 - \omega_j)\sigma_{2,i,j}}{\Delta\omega_i} \\ &= \left(\frac{\sigma_{1,i,j}}{\eta} + (\Omega_2 - \omega_j)\sigma_{2,i+1,j} - (\Omega_2 - \omega_j)\sigma_{2,i,j}\right)\eta \\ &= \sigma_{1,i,j} + (\Omega_2 - \omega_j)(\sigma_{2,i+1,j} - \sigma_{2,i,j})\eta\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial U}{\partial \Omega_2} &\approx \frac{U^{rel}_{\Omega_1,\omega_{j+1}} - U^{rel}_{\Omega_1,\omega_j}}{\Delta\omega_j}\cdot \\ &\quad\cdot\left(\frac{\sigma_{2,i,j}}{\eta} + (\Omega_1 - \omega_i)\sigma_{1,i,j+1} - (\Omega_1 - \omega_i)\sigma_{2,i,j}\right)\eta \\ &= \sigma_{2,i,j} + (\Omega_1 - \omega_i)(\sigma_{1,i,j+1} - \sigma_{1,i,j})\eta\end{aligned}$$

It is possible to define another parameter $\alpha_{i,j}$ based on the definition of $\sigma_{a,i,j}$:

$$\begin{aligned}\alpha_{i,j} &\equiv \left(\sigma_{1,i,j+1} - \sigma_{1,i,j}\right) = \left(\sigma_{2,i+1,j} - \sigma_{2,i,j}\right) \\ &= \left(U_{i+1,j+1} - U_{i,j+1} - U_{i+1,j} + U_{i,j}\right)\eta\end{aligned}$$

So that, in general,

$$\frac{\partial U}{\partial \Omega_a} \approx \sigma_{a,i,j} + \left(\Omega_b - \omega_{b,c}\right)\alpha_{i,j}\eta$$

where: if a=1, then b=2 and c=j; if a=2, then b=1 and c=i.

This is efficiently implemented by calculating the values of variables in the following order:

$i,\ j,\quad \sigma_{1,i,j},\quad \sigma_{2,i,j}$

and finally

$$\alpha_{i,j} = -\sigma_{1,i,j+1} - \sigma_{1,i,j} = U_{i+1,j+1} - U_{i,j+1} - \sigma_{1,i,j}$$

This can be extended to a hypersurface describing the potential U as a function of three variables. Define

$$\sigma_{1,i,j,k} \equiv \left(\frac{\Delta U}{\Delta \omega}\right)_{i,j,k} \equiv \frac{U_{i+1,j,k} - U_{i,j,k}}{\omega_{i+1} - \omega_i} = \left(U_{i+1,j,k} - U_{i,j,k}\right)\eta$$

The gradient components $(\partial U/\partial \Omega_1)$ and $(\partial U/\partial \Omega_2)$ can be calculated in any plane defined by (Ω_3=ω_k= constant) as described above. Then, for the point at coordinates (Ω_1, Ω_2, Ω_3) within the cubic solid bounded by points (ω_i,ω_j,ω_k) and (ω_{i+1},ω_{j+1},ω_{k+1}), one can trace the changes in U from the point (ω_i,ω_j,ω_k) to the points (Ω_1, Ω_2, Ω_k) and (Ω_1, Ω_2, Ω_{k+1}), making it possible to define the interpolated values $\left(U^{rel}_{\Omega_1,\Omega_2,\omega_k}\right)$ and $\left(U^{rel}_{\Omega_1,\Omega_2,\omega_{k+1}}\right)$ *relative* to the value ($U_{i,j,k}$). This makes it possible to find a linear approximation for $(\partial U/\partial \Omega_3)$ solely in terms of other, easily calculated gradient values. That is,

$$\begin{aligned}U^{rel}_{\Omega_1,\Omega_2,\omega_{3,k}} &= \left(\Omega_2 - \omega_j\right)\sigma_{2,i,j,k} + (\Omega_1 - \omega_i)\cdot \\ &\quad\cdot\left(\sigma_{1,i,j,k} + \left(\Omega_2 - \omega_j\right)\left(\sigma_{2,i+1,j,k} - \sigma_{2,i,j,k}\right)\eta\right) \\ U^{rel}_{\Omega_1,\Omega_2,\omega_{3,k+1}} &= \frac{\sigma_{3,i,j,k}}{\eta} + \left(\Omega_2 - \omega_j\right)\sigma_{2,i,j,k+1} \\ &\quad + (\Omega_1 - \omega_i)\left(\sigma_{1,i,j,k+1} + \left(\Omega_2 - \omega_j\right)\cdot\right. \\ &\quad \left.\cdot\left(\sigma_{2,i+1,j,k+1} - \sigma_{2,i,j,k+1}\right)\eta\right)\end{aligned}$$

Alternatively,

$$\begin{aligned}U^{rel}_{\Phi,\Psi,\omega_k} &= (\Omega_1 - \omega_i)\sigma_{1,i,j,k} + \left(\Omega_2 - \omega_j\right)\cdot \\ &\quad\cdot\left(\sigma_{2,i,j,k} + (\Omega_1 - \omega_i)\left(\sigma_{1,i,j+1,k} - \sigma_{1,i,j,k}\right)\eta\right) \\ U^{rel}_{\Phi,\Psi,\omega_{k+1}} &= \frac{\sigma_{3,i,j,k}}{\eta} + (\Omega_1 - \omega_i)\sigma_{1,i,j,k+1} + \left(\Omega_2 - \omega_j\right)\cdot \\ &\quad\cdot\left(\sigma_{2,i,j,k+1} + (\Omega_1 - \omega_i)\cdot\right. \\ &\quad \left.\cdot\left(\sigma_{1,i,j+1,k+1} - \sigma_{1,i,j,k+1}\right)\eta\right)\end{aligned}$$

Then,

$$\begin{aligned}\frac{\partial U}{\partial \Omega_3} &= \left(U^{rel}_{\Omega_1,\Omega_2,\omega_{k+1}} - U^{rel}_{\Omega_1,\Omega_2,\omega_k}\right)\eta \\ &= \left(\frac{\sigma_{3,i,j,k}}{\eta} + \left(\Omega_2 - \omega_j\right)\sigma_{2,i,j,k+1} + (\Omega_1 - \omega_i)\cdot\right. \\ &\quad\cdot\left(\sigma_{1,i,j,k+1} + \left(\Omega_2 - \omega_j\right)\left(\sigma_{2,i+1,j,k+1} - \sigma_{2,i,j,k+1}\right)\eta\right))\eta \\ &\quad -\left(\left(\Omega_2 - \omega_j\right)\sigma_{2,i,j,k} + (\Omega_1 - \omega_i)\left(\sigma_{1,i,j,k} + \left(\Omega_2 - \omega_j\right)\cdot\right.\right. \\ &\quad \left.\left.\cdot\left(\sigma_{2,i+1,j,k} - \sigma_{2,i,j,k}\right)\eta\right)\right)\eta \\ &= \left(\Omega_1 - \omega_{1,i}\right)\left(\Omega_2 - \omega_j\right)\cdot \\ &\quad\cdot\left(\sigma_{2,i+1,j,k+1} - \sigma_{2,i,j,k+1} - \sigma_{2,i+1,j,k} + \sigma_{2,i,j,k}\right)\eta^2 \\ &\quad +\left[\left(\Omega_2 - \omega_j\right)\left(\sigma_{2,i,j,k+1} - \sigma_{2,i,j,k}\right)\right. \\ &\quad \left.+(\Omega_1 - \omega_i)\left(\sigma_{1,i,j,k+1} - \sigma_{1,i,j,k}\right)\right]\eta + \sigma_{3,i,j,k}\end{aligned}$$

Alternatively,

$$\begin{aligned}\frac{\partial U}{\partial \Omega_3} &= \left(U^{rel}_{\Omega_1,\Omega_2,\omega_{k+1}} - U^{rel}_{\Omega_1,\Omega_2,\omega_k}\right)\eta \\ &= \left(\frac{\sigma_{3,i,j,k}}{\eta} + (\Omega_1 - \omega_i)\sigma_{1,i,j,k+1} + \left(\Omega_2 - \omega_j\right)\cdot\right. \\ &\quad\cdot\left(\sigma_{2,i,j,k+1} + (\Omega_1 - \omega_i)\left(\sigma_{1,i,j+1,k+1} - \sigma_{1,i,j,k+1}\right)\eta\right))\eta \\ &\quad -\left(\left(\Omega_1 - \omega_i\right)\sigma_{2,i,j,k} + \left(\Omega_2 - \omega_j\right)\left(\sigma_{2,i,j,k}\right.\right. \\ &\quad \left.\left.+(\Omega_1 - \omega_i)\left(\sigma_{1,i,j+1,k} - \sigma_{1,i,j,k}\right)\eta\right)\right)\eta \\ &= \left(\Omega_1 - \omega_i\right)\left(\Omega_2 - \omega_j\right)\left(\sigma_{1,i,j+1,k+1} - \sigma_{1,i,j,k+1}\right. \\ &\quad \left.-\sigma_{1,i,j+1,k} + \sigma_{1,i,j,k}\right)\eta^2 \\ &\quad +\left[\left(\Omega_2 - \omega_j\right)\left(\sigma_{2,i,j,k+1} - \sigma_{2,i,j,k}\right)\right. \\ &\quad \left.+(\Omega_1 - \omega_i)\left(\sigma_{1,i,j,k+1} - \sigma_{1,i,j,k}\right)\right]\eta + \sigma_{3,i,j,k}\end{aligned}$$

The two above expressions for $(\partial U/\partial\Omega_3)$ are equivalent as

$$\begin{aligned}&\left(U_{i+1,j+1,k+1}-U_{i,j+1,k+1}\right)\eta-\left(U_{i+1,j,k+1}-U_{i,j,k+1}\right)\eta\\&\quad-\left(U_{i+1,j+1,k}-U_{i,j+1,k}\right)\eta+\left(U_{i+1,j,k}-U_{i,j,k}\right)\eta\\&\quad=\left(\sigma_{1,i,j+1,k+1}-\sigma_{1,i,j,k+1}-\sigma_{1,i,j+1,k}+\sigma_{1,i,j,k}\right)\end{aligned}$$

and

$$\begin{aligned}&\left(U_{i+1,j+1,k+1}-U_{i+1,j,k+1}\right)\eta-\left(U_{i,j+1,k+1}-U_{i,j,k+1}\right)\eta\\&\quad-\left(U_{i+1,j+1,k}-U_{i+1,j,k}\right)\eta+\left(U_{i,j+1,k}-U_{i,j,k}\right)\eta\\&\quad=\left(\sigma_{2,i+1,j,k+1}-\sigma_{2,i,j,k+1}-\sigma_{2,i+1,j,k}+\sigma_{2,i,j,k}\right)\end{aligned}$$

and it is clear that the last two expressions are identical when stated in terms of U.

Permutation of the indices gives values for $(\partial U/\partial\Omega_1)$ and $(\partial U/\partial\Omega_2)$

$$\begin{aligned}\frac{\partial U}{\partial\Omega_1}&=\left(\Omega_2-\omega_j\right)(\Omega_3-\omega_k)\left(\sigma_{2,i+1,j,k+1}-\sigma_{2,i+1,j,k}\right.\\&\quad\left.-\sigma_{2,i,j,k+1}+\sigma_{2,i,j,k}\right)\eta^2+\left[\left(\Omega_2-\omega_j\right)\left(\sigma_{2,i+1,j,k}-\sigma_{2,i,j,k}\right)\right.\\&\quad\left.+(\Omega_3-\omega_k)\left(\sigma_{3,i+1,j,k}-\sigma_{3,i,j,k}\right)\right]\eta+\sigma_{1,i,j,k}\end{aligned}$$

$$\begin{aligned}\frac{\partial U}{\partial\Omega_2}&=(\Omega_1-\omega_i)(\Omega_3-\omega_k)\left(\sigma_{3,i+1,j+1,k}-\sigma_{3,i,j+1,k}\right.\\&\quad\left.-\sigma_{3,i+1,j,k}+\sigma_{3,i,j,k}\right)\eta^2+\left[(\Omega_1-\omega_i)\left(\sigma_{1,i,j+1,k}-\sigma_{1,i,j,k}\right)\right.\\&\quad\left.+(\Omega_3-\omega_k)\left(\sigma_{3,i,j+1,k}-\sigma_{3,i,j,k}\right)\right]\eta+\sigma_{2,i,j,k}\end{aligned}$$

Three of the six faces of the cubic solid bounded by points $(\omega_i,\omega_j,\omega_k)$ and $(\omega_{i+1},\omega_{j+1},\omega_{k+1})$ can each be associated with a parameter $\alpha_{a,i,j,k}$, which is analogous to the parameter $\alpha_{i,j}$ defined for the two-dimensional case above. That is,

$$\begin{aligned}\alpha_{1,i,j,k}&\equiv\left(\sigma_{2,i,j,k+1}-\sigma_{2,i,j,k}\right)=\left(\sigma_{3,ij+1,k}-\sigma_{3,i,j,k}\right)\\&=\left(U_{i,j+1,k+1}-U_{i,j,k+1}-U_{i,j+1,k}+U_{i,j,k}\right)\eta\end{aligned}$$

$$\begin{aligned}\alpha_{2,i,j,k}&\equiv\left(\sigma_{1,i,j,k+1}-\sigma_{1,i,j,k}\right)=\left(\sigma_{3,i+1,j,k}-\sigma_{3,i,j,k}\right)\\&=\left(U_{i+1,j,k+1}-U_{i,j,k+1}-U_{i+1,j,k}+U_{i,j,k}\right)\eta\end{aligned}$$

$$\begin{aligned}\alpha_{3,i,j,k}&\equiv\left(\sigma_{1,i,j+1,k}-\sigma_{1,i,j,k}\right)=\left(\sigma_{2,i+1,j,k}-\sigma_{2,i,j,k}\right)\\&=\left(U_{i+1,j+1,k}-U_{i,j+1,k}-U_{i+1,j,k}+U_{i,j,k}\right)\eta\end{aligned}$$

Another parameter $\beta_{i,j,k}$ can be defined by extending the definition of $\alpha_{a,i,j,k}$ to include, for example, $\alpha_{1,i+1,j,k}$, which is associated with a cubic solid adjacent to the one being considered. $\beta_{i,j,k}$ can be expressed in many different ways, some of which may be more computationally convenient:

$$\begin{aligned}\beta_{i,j,k}&=\sigma_{3,i+1,j+1,k}-\sigma_{3,i,j+1,k}-\sigma_{3,i+1,j,k}+\sigma_{3,i,j,k}\\&=\alpha_{2,i,j+1,k}-\alpha_{2,i,j,k}=\alpha_{1,i+1,j,k}-\alpha_{1,i,j,k}\\&=\sigma_{2,i+1,j,k+1}-\sigma_{2,i,j,k+1}-\sigma_{2,i+1,j,k}+\sigma_{2,i,j,k}\\&=\alpha_{3,i,j,k+1}-\alpha_{3,i,j,k}=\alpha_{1,i+1,j,k}-\alpha_{1,i,j,k}\\&=\sigma_{1,i,j+1,k+1}-\sigma_{1,i,j+1,k}-\sigma_{1,i,j,k+1}+\sigma_{1,i,j,k}\\&=\alpha_{2,i,j+1,k}-\alpha_{2,i,j,k}=\alpha_{3,i,j,k+1}-\alpha_{3,i,j,k}\\&=\left(U_{i+1,j+1,k+1}-U_{i,j+1,k+1}\right)-\left(U_{i+1,j+1,k}-U_{i,j+1,k}\right)\\&\quad-\left(U_{i+1,j,k+1}-U_{i,j,k+1}\right)+\left(U_{i+1,j,k}-U_{i,j,k}\right)\\&=\left(U_{i+1,j+1,k+1}-U_{i,j+1,k+1}\right)-\left(U_{i+1,j+1,k}-U_{i,j+1,k}\right)\\&\quad-\alpha_{2,i,j,k}=U_{i+1,j+1,k+1}-U_{i,j+1,k+1}-U_{i+1,j,k+1}\\&\quad-U_{i+1,j+1,k}+U_{i+1,j,k}+U_{i,j+1,k}+U_{i,j,k+1}-U_{i,j,k}\end{aligned}$$

Then, the components of the gradient at coordinates $(\Omega_1, \Omega_2, \Omega_3)$ can be simply expressed as

$$\begin{aligned}\frac{\partial U}{\partial\Omega_1}&=\left(\Omega_2-\omega_j\right)(\Omega_3-\omega_k)\beta_{i,j,k}\eta^2+\left[\left(\Omega_2-\omega_j\right)\alpha_{3,i,j,k}\right.\\&\quad\left.+(\Omega_3-\omega_k)\alpha_{2,i,j,k}\right]\eta+\sigma_{1,i,j,k}\end{aligned}$$

$$\begin{aligned}\frac{\partial U}{\partial\Omega_2}&=(\Omega_1-\omega_i)(\Omega_3-\omega_k)\beta_{i,j,k}\eta^2+\left[(\Omega_1-\omega_i)\alpha_{3,i,j,k}\right.\\&\quad\left.+(\Omega_3-\omega_k)\alpha_{1,i,j,k}\right]\eta+\sigma_{2,i,j,k}\end{aligned}$$

$$\begin{aligned}\frac{\partial U}{\partial\Omega_3}&=(\Omega_1-\omega_i)\left(\Omega_2-\omega_j\right)\beta_{i,j,k}\eta^2+\left[\left(\Omega_2-\omega_j\right)\alpha_{1,i,j,k}\right.\\&\quad\left.+(\Omega_1-\omega_i)\alpha_{2,i,j,k}\right]\eta+\sigma_{3,i,j,k}\end{aligned}$$

References

1. Bock K (1983) Pure Appl Chem 55:605–622
2. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) J Chem Phys 21:1087–1092
3. O'Neill MA, York WS (2003) The composition and structure of primary cell walls. In: Rose JKC (ed) The plant cell wall, Annual Plant Reviews, vol. 8. CRC Press, Boca Raton, FL, pp 1-54
4. Rose JKC, Braam J, Fry SC, Nishitani K (2002) Plant Cell Physiol 43:1421–1435
5. Chu SSC, Jeffrey GA (1968) Acta Crystallogr B24:830–838
6. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratman RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu C, Liashenko A, Piskorz P, Komaromi, I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Gonzales C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98 (Revision A.9). Gaussian, Pittsburgh, PA
7. Tvaroska I, Perez S (1986) Carbohydr Res 149:389–410
8. Beard DA (2001) A molecular modeler's guide to statistical mechanics. In: Beard DA (ed) Computational and theoretical biophysics. Biophysics Textbook Online (http://www.biophys.org/btol/), The Biophysical Society
9. Kirshner KN, Woods RJ (2001) Proc Natl Acad Sci USA 98:10541–10545
10. Nimz O, Geßler K, Uson I, Laettig S, Welfle H, Sheldrick GM, Saenger W (2003) Carbohydr Res 338:977–986
11. Hardy BJ, Sarko A (1993) J Comput Chem 14:831–847
12. McCann MC, Wells B, Roberts K (1990) J Cell Sci 96:323–334
13. Prestegard JH, Al-Hashimi HM, Tolman JR (2000) Q Rev Biophys 33:371–424
14. Zweckstetter M, Bax A (2000) J Am Chem Soc 122:3791–3792