ELSEVIER

Single-coordinate-driving method for molecular docking: Application to modeling of guest inclusion in cyclodextrin

Jaroslav Koča,*† Martin Ludin,* Serge Pérez,† and Anne Imberty†

**Laboratory of Biomolecular Structure and Dynamics and Department of Organic Chemistry, Faculty of Science, Masaryk University, Brno, Czech Republic*
†Centre de Recherches sur les Macromolecules Végétales (affiliated with Université Joseph Fourier), CNRS, Grenoble cedex 9, France

An extension of the computer program CICADA has been developed that allows us to use the single-coordinate-driving (SCD) method for flexible molecular docking. The docking procedure is composed of three independent space rotations, three independent translations, and the torsions selected by the user. One of the coordinates is driven; the other coordinates are relaxed. This procedure follows low-energy wells on the potential energy surface of the entire system. The program allows us to dock more than one ligand molecule to the receptor. We ran two test examples, docking N,N-dimethylformamide into alpha-cyclodextrin and R-phenoxypropionic acid into beta-cyclodextrin. The test examples showed that the SCD approach is able to overcome high-energy barriers and to cover the entire box within which the search is performed. The limitations of molecular dynamics docking in comparison with our approach also are discussed. The philosophy of the newly developed approach is not only to find the best dock for the receptor-ligand(s) system, but also to describe all the important binding modes and provide a good starting point for studying the dynamics within the cavity during the docking process.

Keywords: molecular docking, energy surface, single-coordinate-driving, CICADA, R-phenoxypropionic acid, cyclodextrins, N,N-dimethylformamide

Corresponding author: A. Imberty, CERMAV-CNRS, BP 53, 38041 Grenoble cedex 9, France. Tel.: 33-4-76-03-76-36; fax: 33-4-76-54-72-03.

E-mail address: anne.imberty@cermav.cnrs.fr (A. Imberty)

INTRODUCTION

Detailed analysis of molecular interactions is one of the key points to understanding and eventually controlling the fundamental processes occurring in living systems. Studying the interactions between proteins and small ligands, a process called docking, currently is a hot topic. Molecular docking is of particular interest in drug design. Two principal problems must be solved to address the issue. First, one needs to have a good energy function that properly describes the binding affinity. Second, because the solution leads to an optimization problem in multidimensional space, one must have a good search strategy to describe the ligand-receptor space. The problem of binding affinity has been addressed in the literature, from the point of view of both the potential[1] and the free energy[2] of binding.

In this paper, we focus our attention on the latter problem of search strategies. Since the introduction of DOCK,[3] the first docking program, a number of search algorithms have been proposed to solve the docking problem (for a review, see Gschwend et al.[4] and Vieth et al.[5]). The reported techniques for automated docking fall into two broad categories[6]: matching approaches and docking simulation methods. *Matching methods* create a representative model of the rigid active site usually composed of hydrogen bonding and sterically accessible sites. The geometry of the ligand is docked as a rigid body into the active site by matching both complementary geometries. Such a procedure, typically represented by the DOCK program, usually is very fast and efficient when conformational flexibility does not play a substantial role in the docking mechanism. The generally slower *simulation methods* model docking processes in a

Journal of Molecular Graphics and Modelling 18, 108–118, 2000

655 Avenue of the Americas, New York, NY 10010

1093-3263/00/$–see front matter
PII S1093-3263(00)00042-5

more precise manner. They allow for translations, rotations, and conformational changes on the ligand; therefore, they provide a more detailed picture of the docking mechanism.

From a mathematical point of view, the docking problem is principally a search in the multidimensional space of the ligand and the receptor. As such, the problem is the same as that appearing several years ago in conformational analysis. It originally was thought that a molecule could adopt only one conformation. This idea led to the development and use of algorithms to search for the global minimum on the energy surface. A number of examples later showed that flexible molecules can adopt several important conformations or conformational families. This led to the development of algorithms to solve the so-called multiple minima problem. This situation also is found in the field of molecular docking, where many techniques known from conformational analysis have been employed. Among these, simulated annealing Monte Carlo,[7,8] molecular dynamics,[9] and evolutionary algorithms[10] have been used successfully. Several different techniques have been developed, most notably in the field of genetic algorithms, which belong to the group of evolutionary algorithms.[2,11–14] It was shown recently that each approach has advantages and disadvantages, so the development of new strategies still is necessary.[5] In the present paper, we focus our attention on application of the single-coordinate-driving (SCD) method in this field.

The SCD method originally was used for conformational analysis by Ginzburg et al.[15] and Lugovsky et al.[16] It was further elaborated and developed to cover the entire conformational space by Osawa and Goto,[17–19] and Ngo and Karplus.[20] We improved this method and implemented it in the computer programs DAISY[21] and CICADA.[22] The latter program has been interfaced with MM2,[23] MMX,[24] MM3,[25] and AMBER[26] molecular mechanics. We also developed the SCD-SA method, in which SCD is combined with simulated annealing.[27,28] SCD and SCD-SA have been used successfully in the conformational analysis of almost 100 molecules, including pheromones,[29,30] carbohydrates,[31–34] peptides,[35–38] and oligonucleotides.[39] All results were reviewed by Koča.[40] Because SCD and SCD-SA provide very detailed pictures of the low-energy parts of multidimensional space (which are comparable with the grid search results[31,40] but with more efficient calculation), we decided to extend the idea and not only use it for analyzing the conformational space. We also chose to include the possibility of having more than one ligand, because this would allow one to take into account the presence of a third compound, for example, an ion, in the binding site. In this paper, we describe the application of SCD to the problem of flexible molecular docking or, in a more general sense, to the problem of how two or more molecules may interact when subsequent conformational changes take place.

METHODS

Space Operations

We consider that the analyzed system is composed of two or more molecules. One molecule will be called the receptor; the remaining molecules will be called ligands. It is necessary to ensure that the ligands can move with respect to the receptor and with respect to each other. This implies that we need to perform the rotations and translations of each ligand in a certain coordinate system where the receptor does not change its space position. At the same time, we must allow for the conformational flexibility of the receptor and all the ligands.

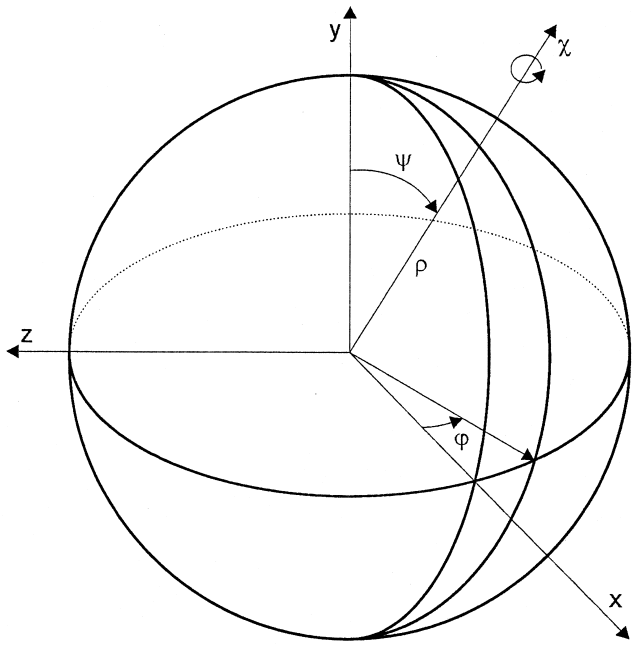

Figure 1. All rotation operations in the space can transparently be expressed by the spherical polar coordinates ρ, ϕ, ψ and the "spin" χ. The distance ρ of a point from the origin of the coordinate system is used to obtain the Cartesian coordinates of the point from its ϕ, ψ, ρ coordinates. See text for details.

Rotations All the rotation operations of the ligands are performed using spherical polar coordinates[41] as shown in Figure 1. The relationship between the Cartesian (x, y, z) and the spherical polar coordinates (ρ, ϕ, ψ) of a point in space is expressed[41] by Equation 1:

$$x = \rho \sin\psi \cos\varphi, \quad y = \rho \cos\psi, \quad z = -\rho \sin\psi \sin\varphi \tag{1}$$

where ρ is the distance of the point from the origin, and the the angles ϕ and ψ are defined as illustrated in Figure 1.

Rotation about a given axis is specified by the angle χ (designated to as "spin"). The polar coordinates ψ (elevation) and ϕ (rotation) define the space orientation of the rotation axis.

The problem to solve is to find a transformation that is expressed by the angles χ, ψ, *and* ϕ and shows the relationship among these three angles and the Cartesian coordinate system. Such a transformation[41] is shown in Equation 2:

$$\mathbf{R_{sp}} = \begin{pmatrix} c\chi + (1-c\chi)s^2\psi c^2\varphi & -s\psi s\varphi s\chi + (1-c\chi)c\psi s\psi c\varphi & -c\psi s\chi - (1-c\chi)s^2\psi c\varphi s\varphi \\ s\psi s\varphi s\chi + (1-c\chi)c\psi s\psi c\varphi & c\chi + (1-c\chi)c^2\psi & s\psi c\varphi s\chi - (1-c\chi)c\psi s\psi s\varphi \\ c\psi s\chi - (1-c\chi)s^2\psi c\varphi s\varphi & -s\psi c\varphi s\chi - (1-c\chi)c\psi s\psi s\varphi & c\chi + (1-c\chi)s^2\psi s^2\varphi \end{pmatrix} \quad (2)$$

where *s* and *c* are *sin* and *cos* functions, respectively.

This matrix $\mathbf{R_{sp}}$ is a rotation matrix in which rotation is described by the angles χ, ψ, ϕ. When the original system is rotated by the angle χ about the axis, the orientation of which is expressed by the angles ψ, ϕ, then a vector $\mathbf{A} = (\mathbf{a,b,c})$ is transformed into a new vector $\mathbf{A'} = (\mathbf{a',b',c'})$. The transformation is expressed by the matrix given in Equation 3:

$$\mathbf{A'} = \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \mathbf{R_{sp}} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{R_{sp}}\,\mathbf{A}. \quad (3)$$

The condition given in Equation 4 ensures that all possible rotation operations are considered:

$$0 \le \chi < 2\pi, \qquad 0 \le \psi < \pi, \qquad 0 \le \varphi < 2\pi. \quad (4)$$

Absolute coordinate system and translations To ensure that constraints applied on a ligand during energy minimization are really ligand constraints and do not cause rotations/translations of the entire system, it is necessary to express the coordinates of all atoms in a coordinate system in which the positions of the receptor and all the remaining ligands, except for the constrained one, remain unchanged. Such coordinates are considered to be absolute, and this system is called the absolute coordinate system (ACS). It is presumed, however, that bond length and angles will not explicitly be driven during the search. Therefore, they are excluded from ACS, which is a subset of the entire coordinate system. ACS is created such that there is one base atom selected from the receptor, i.e., A^0, and one base atom selected from each ligand, i.e., A^1, A^2, ..., A^L, where L is the number of ligands. Another two atoms, i.e., B^0 and C^0, are selected from the receptor such that both B^0 and C^0 are bonded to A^0. For each ligand *i* (i = 1, 2, ... L) we select atoms B^i and C^i applying the same rules. Because we do not expect any bond breaking/creation during the docking procedure and we expect only conformational changes, the mutual position of the atoms A^0, B^0, and C^0 will remain unchanged. The same will be valid for the angle B^0-A^0-C^0, and it also applies for all selected atoms A^i, B^i, C^i on all the ligands.

The initially used coordinate system is translated and rotated such that A^0 becomes the origin of the new coordinate system, the bond A^0B^0 is fitted to the x-axis following the rule that the x-coordinate of the atom B^0 is positive, and the atom C^0 is put in the xy-plane such that its y-coordinate is positive. The coordinates of all atoms are recalculated after each change of the coordinate system. The same applies to all translations of all ligands. Rotations of the ligand *i* are performed such that the line connecting the atoms A^iB^i is made the principal axis for the rotation, and the angles χ, ψ, ϕ are expressed within ACS as shown in Figures 1 and 2.

SCD Method

The initial idea implies that if the substantial coordinate (usually called the reaction coordinate) of the studied process is known, then one can follow the energy profile of the process just by driving the reaction coordinate and subsequently minimizing all the remaining coordinates. The problem is that the reaction coordinate in Cartesian space is a complicated combination of Cartesian coordinates of single atoms. The situation in internal coordinates space usually is much more straightforward when the internal system is properly selected. In our case, we will use the dihedral angle space combined with three

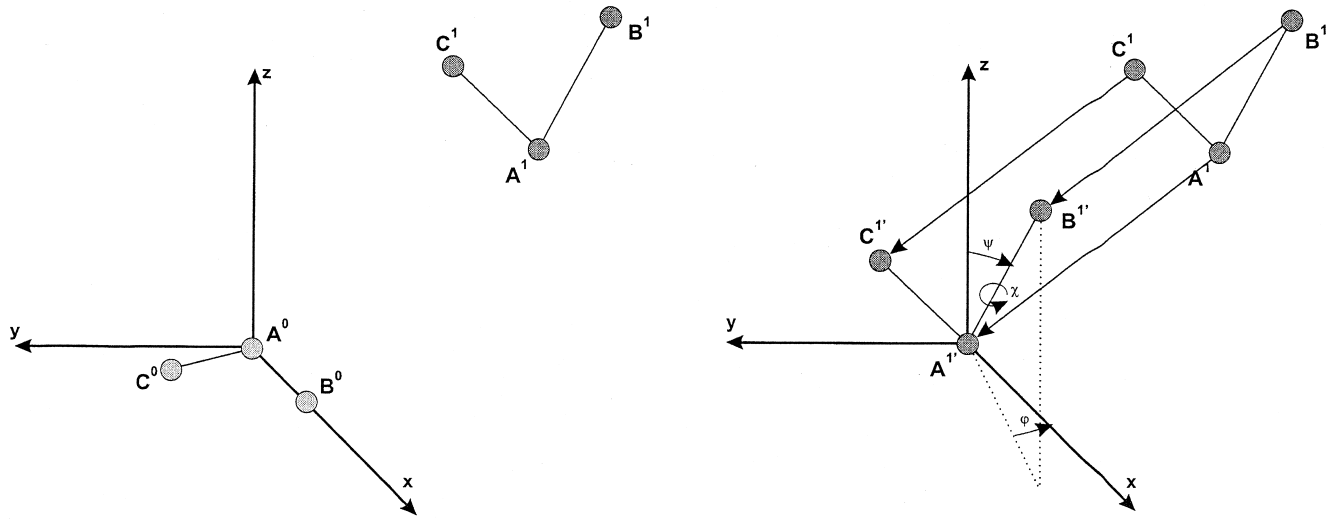

Figure 2. The absolute coordinate system is created such that A^0 becomes the origin, the bond A^0B^0 is put along the x-axis, and the atom C^0 is put in the xy-plane. The angles χ, ψ, ϕ, which characterize the rotation of the ligand, are calculated following the rule that the ligand first is translated such that A^1 is put at the origin ($A^{1'}$), then ψ is the angle between $A^{1'}B^{1'}$ and the y-axis, ϕ is the angle between the xz projection of $A^{1'}B^{1'}$ and the x-axis, and χ is the rotation around $A^{1'}B^{1'}$. See text for details.

translations (*x*-,*y*-,*z*-translation) and three fundamental rotations expressed by spherical polar coordinates as shown earlier. Translations and rotations are always considered in the absolute coordinate system introduced earlier. Because the procedure can include more than one ligand, three translation and three rotation coordinates are taken into account for each ligand. Altogether, $6L+N$ coordinates are considered for the driving procedure, where L is the number of ligands and N is the number of dihedral angles of interest. The driving procedure is illustrated on the model two-dimensional surface shown in Figure 3. Let us consider that both coordinates are angles (either dihedral or spherical polar coordinates and the spin).

The search begins in the lowest energy minimum available, i.e., point C1. In reality, it can be any minimized starting geometry. One of the coordinates available is selected. For conformational analysis, it can be any dihedral angle. For flexible docking, it also can be a coordinate describing space movement, i.e., a translation or a rotation of the ligand. Let us select an angle Φ. The selection is based on the lowest expected value of the energy barrier and is random at the beginning of the search. The selected coordinate is changed by a predefined step, ΔΦ, and a new point A is reached. The predefined step is different for ligand rotation, ligand translation, and torsion rotation. The new value of angle Φ is constrained, and all the remaining coordinates are relaxed under subsequent energy minimization. This procedure generates point B, which is on the bottom of the energy valley leading from C1 to C2. The process is repeated until the new point is lower in energy than the previous one. In such a case, a transition state has been encountered. The highest energy point T1 is stored as an approximation of the real transition state of a space movement of the ligand, a conformational change in the system, or a combination of both. This "downhill" traveling is continued until the energy increases again. This will happen at the point D. The previous point C, which is an energy minimum on the energy profile, then is fully relaxed and the true energy minimum C2 is obtained. Three points are saved for further elaboration, C1, T1, C2, together with the pathway C1-T1-C2. The minimum generated could be either a completely new one or one of those found previously. In the latter case, only the one with lower energy is retained. The procedure described is repeated for all the minima found within a predefined energy window and for all selected coordinates. Dihedral angles are rotated in both (clockwise and counterclockwise) possible directions. The same applies for space operations on ligands where each translation and rotation is subsequently performed in two directions (+/−). Because each ligand orientation depends on six degrees of freedom (three independent translations and the same number of independent rotations) and two possible directions of motion must be considered for each degree of freedom including the dihedral angles, the procedure commonly generates $2(N+6L)$ possible pathways from each minimum, where N is the number of

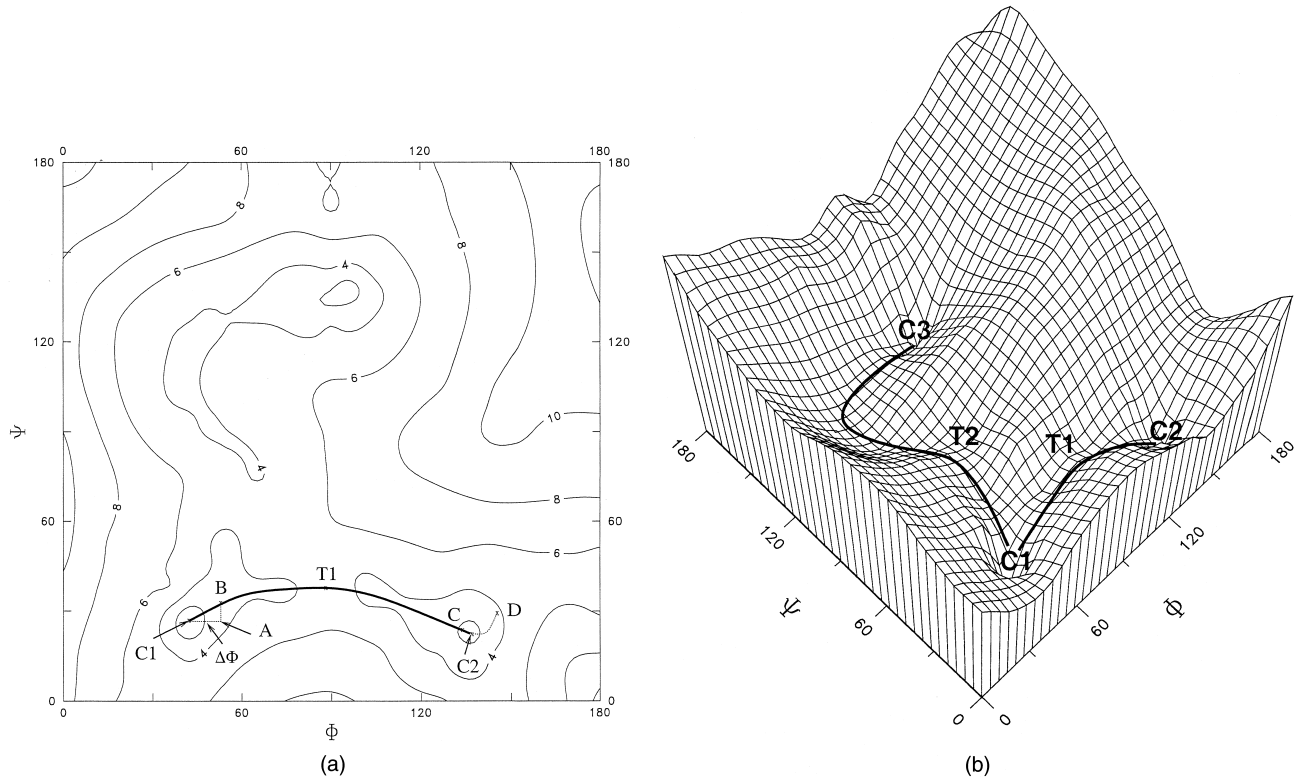

Figure 3. View of the two-dimensional energy surface (in ϕ, ψ coordinates). (a) Contour plot of the surface. (b) Example of pathways generated by the SCD approach (heavy curves). For example, when Φ is driven in the positive direction starting in the minimum C1, the pathway C1-T1-C2 is generated. By analogy, the pathway C1-T2-C3 is generated by driving ψ in the positive direction, again starting at C1. See text for details. (Reprinted from Engelsen et al.[31] with permission of Elsevier Science.)

driven dihedrals and *L* is the number of ligands. The real effect of the SCD on space operations is shown in Figure 4.

The procedure is controlled by several parameters. The most important are the box size, the set of selected space descriptors and dihedrals that have to be driven, the stop criterion E_{conf}, the identity criterion D_{min}, and the driver steps.

Translation of each ligand is allowed only within a predefined box that is created for each ligand separately. Each box is defined by six points in space (absolute coordinate system, see earlier). The *set of selected space descriptors* is chosen such that it controls space movement. It means, for example, that to speed up the docking one can only drive translations with subsequent minimization, which also will partly ensure rotations. One can even drive only one translation, such as *x*-axis movement. The same applies for space rotation operations. The *set of selected dihedrals* is chosen such that it controls the strategy of the conformational search during the docking procedure. For example, if an oligonucleotide is analyzed, the bonds of the backbone are switched on to obtain a fast preliminary conformational search, which then can be refined by switching on the rotation of small groups, such as hydroxyl groups. If it is known that conformational behavior plays a role for only a certain part of the ligand, then only the torsions of this part are switched on.

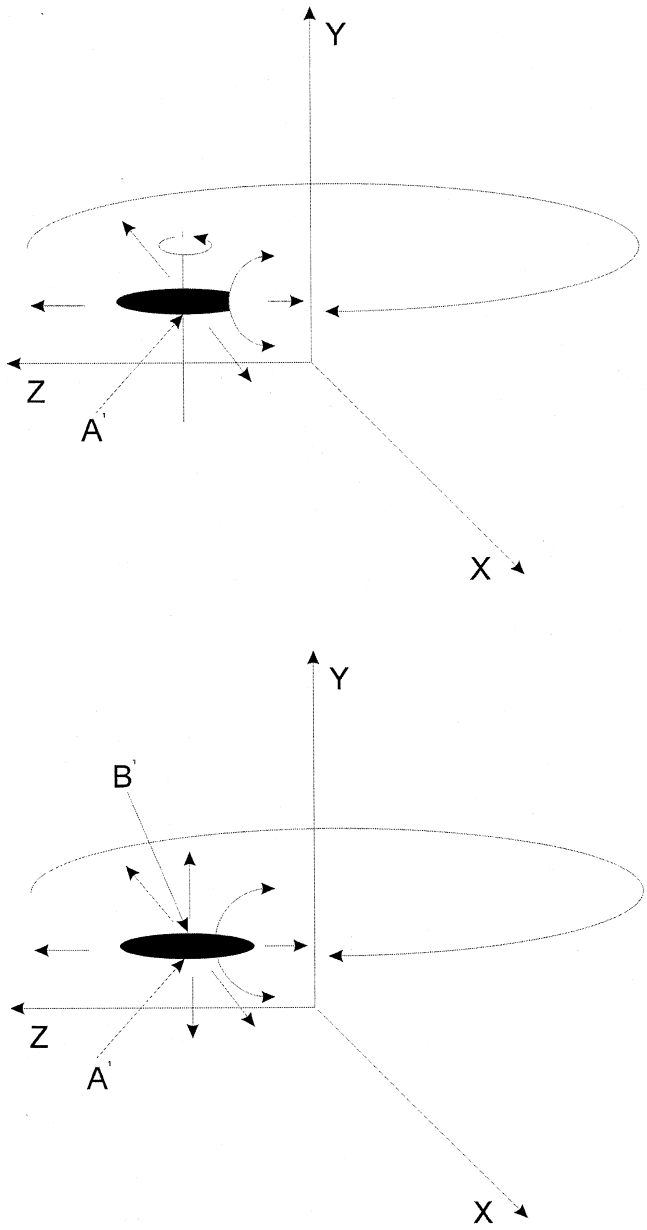

Figure 4. Real effect of SCD on space operations. (a) When the y-translation is constrained at a certain value, the atom A^1 of the ligand is allowed, during the energy minimization, to move only within the plane perpendicular to the y-axis in y_0 (the plane is described by the equation $y = y_0$), but the ligand itself is allowed to undergo any rotation around the atom A^1. (b) When the rotation about A^1B^1 (spin) is constrained, the ligand is allowed to undergo all translations and both ψ and ϕ rotations.

The stop criterion E_{conf} is based on the relative energy of nuclear configurations. When the described procedure is performed for all minima with relative energy less than E_{conf}, then the algorithm stops. The *identity criterion* is used to determine if two configurations are identical. The identity is measured by differences in space rotations, translations, and dihedral angles. Space rotations and translations are calculated with respect to the absolute coordinate system. The *driver step* determines how fine the search is. This is typically between 15 and 30° for space rotations and for dihedral angles of open chain parts of molecules, about 5° for dihedral angles within rings, and about 0.2 Å for translations. It is important to note that the driver step is not a regular grid increment, as all local minima are fully relaxed. Obviously, a smaller driver step will ensure smoother and more detailed analysis of the energy surface. For more details about SCD and SCD-SA methods, see the literature[22,27,28,40].

Implementation

The described approach has been implemented as a part of the computer program CICADA, which originally was developed for conformational analysis.[22] It has been written in FORTRAN 77. All space operations have been included in the energy gradient calculations to create a tool that would allow for space constraints. CICADA has been interfaced with the TINKER package[42] by Ponder. TINKER was selected because it includes several force fields and molecular dynamics, which is necessary for application of the SCD-SA approach.

The input data of the new version of CICADA includes at least one starting nuclear configuration and the parameters described earlier. The calculated system can be composed of one or more molecules. When only one molecule is present, no docking can be performed. When more than one molecule is entered, one of them must be called the receptor (host) and all the remaining ones are called ligands. One may switch on only certain coordinates to be driven. When all dihedral angles are switched off, (pseudo)rigid docking is performed, which means that conformations of the molecules in the system can only be changed by the minimization procedure, but the dihedral angles are not driven. The same applies to space translations and rotationsm, which are considered separately for each ligand. The restart of the job is allowed, for example, after changing the active parameters of the driving procedure. This means, for instance, that one can only switch on dihedrals on the ligand to

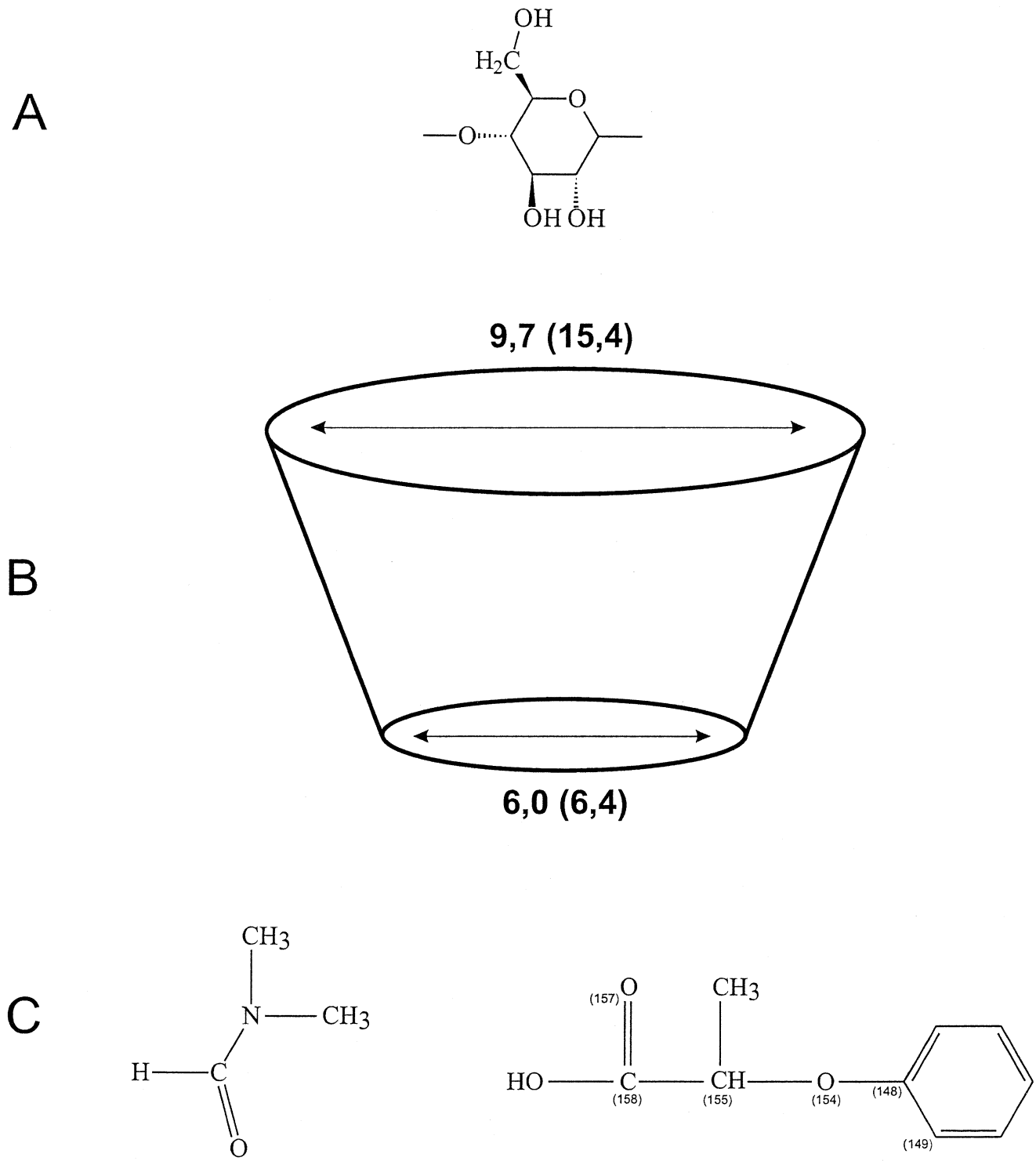

Figure 5. (A) α-D-glucose monomeric unit, of which cyclodextrins are composed. (B) Schematic representation of cyclodextrins. The numbers indicate the diameter of the torus for α and β (in parentheses) cyclodextrins (in Angstroms). (C) Ligands used for docking: N,N-dimethylformamide (DMF, left) and R-phenoxypropionic acid (RPHA, numbers in parentheses show internal atom numbering).

perform a starting inspection of the conformational behavior of the ligand in the presence of the host, and then also switch on the space operations to dock each of the (low-energy) conformers generated in the previous step. CICADA produces a set of triads made up of *energy minimum-energy maximum-energy minimum.* The *energy maximum* is an estimation of a transition state for the interconversion between the two energy minima. These triads are saved as the edges of a graph that can then be

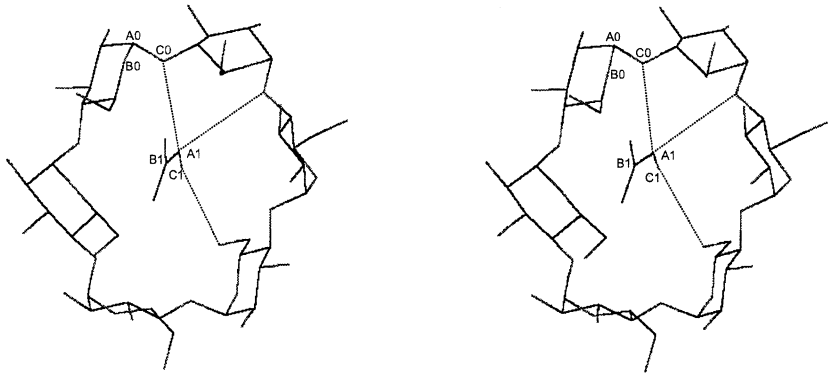

Figure 6. Stereo view of the structure of the α-cyclodextrin-N,N-dimethyformamide complex as obtained by X-ray crystallography. The absolute coordinate system is defined by atoms A^0, B^0, and C^0. Intermolecular hydrogen bonds are shown by shaded lines. See text for details.

used for an animation of Boltzmann traveling[43] along the host–ligand energy surface. For details on the procedure used to analyze and utilize the results, see a previous review.[40]

RESULTS AND DISCUSSION

We tested the methodology on two examples that fall into the chemistry of cyclodextrins. They are the cyclic oligomers of 1-4 linked α-D-glucose monomers, which have a ring structure that is torus-like in shape (Figure 5).

These compounds can include several kinds of molecules in their internal cavity, and they can be used as an environment for chemical reactions.[44,45] Cyclodextrins also have been employed in chromatography to separate constitutional isomers and enantiomers.[46,47] Altogether, these molecules are of great interest to a wide range of scientists in several fields of chemistry The first example we chose is the rigid docking of the N,N-dimethylformamide molecule (DMF) into the cyclodextrin cavity. Here, we wanted to demonstrate the ability of the software to search for all possible space orientations and all translations. The second example focused on the testing abilities of the new method for flexible docking. For this purpose, a conformationally variable phenoxypropionic acid was chosen.

All calculations were performed with the MM3[25,48,49] force field using the 1996 parameter set and a dielectric constant of 4.0 to mimic crystal conditions. The nonlinear conjugate gradient as implemented in the MINIMIZE program of the TINKER suite was used for all minimizations, with maximum minimization cycles set to 500 and the CAPPA line search termination criterion of 0.4. All other parameters were taken from the default set of the MINIMIZE program.

Docking of DMF

The chosen system is composed of an α-cyclodextrin (ACD) receptor (six monomeric units) and a DMF ligand. The starting structure for the search procedure was taken from the Cambridge Crystallographic Database,[50] access code ACDMFM.[51] The crystal structure, shown in Figure 6, reveals that DMF is mainly bound by the hydrogen bond between the hydrogen atom in the HCO group of DMF and a glycosidic oxygen of ACD, and by the hydrogen bond between the oxygen atom in the HCO group of DMF and a hydrogen atom of an OH group of ACD.

Before calculation, we randomly moved the DMF out of the cavity but kept it within the box. Figure 6 shows the definition of the absolute coordinate system (atoms A^0,B^0,C^0) and the

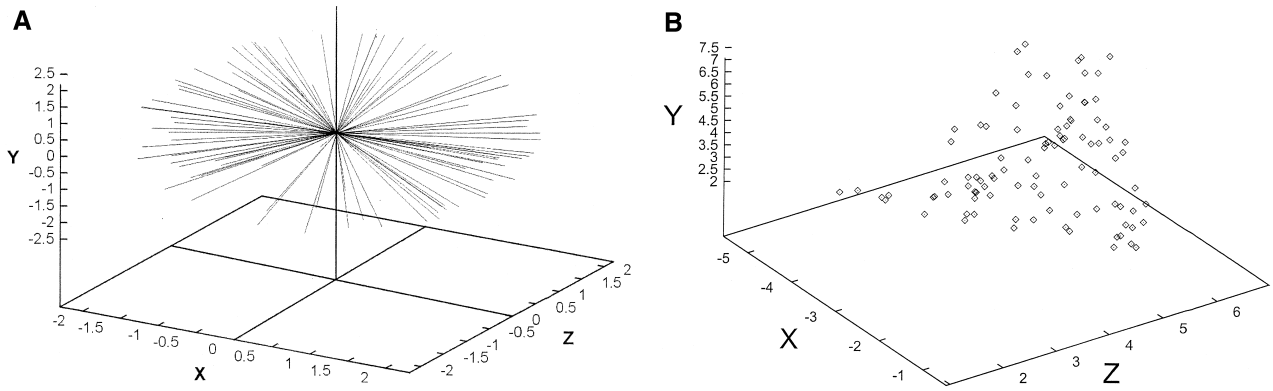

Figure 7. (a) Orientations of the vector B^1C^1 in the 100 lowest energy minima found by CICADA. The coordinates of the vector B^1C^1 are recalculated with the cyclodextrin kept fixed, and the whole system is translated such that B^1 becomes the origin of the new coordinate system. (b) Translations for the nuclear configurations used in (a).

definition of the ligand coordinate system (atoms A^1,B^1,C^1). The calculation was stopped after 7624 points were generated on the energy surface, of which 628 were energy minima and 780 transition states. All the energy minima were compared by X-ray geometry. The best fit exhibited a root mean square deviation (RMSD) of 0.55 and 0.30 Å calculated between CICADA results and X-ray as well as X-ray minimized structures, respectively. RMSD values were calculated over all the heavy atoms. This relatively high deviation does not reveal much about the quality of the search; instead, it yields information about the quality of the energy function for describing van der Waals and hydrogen-bonded complexes. The relative energy of the best fit (with respect to the lowest energy nuclear configuration found) is 2.2 kcal/mol. More detailed analysis of the results shows that there is no significant geometry change between ACD from the X-ray and from our calculations. It implies that the RMSD value is fully generated by the DMF orientation. Because DMF is a small molecule, it is relatively free to move, which increases RMSD. This is confirmed by the high RMSD between the X-ray and X-ray minimized configurations, which was calculated to be 0.78 Å.

To obtain more reliable information about the quality of the search, we visualized the orientations of the B^1C^1 vector (Figure 7) for the 100 lowest energy minima obtained by CICADA and recalculated them with respect to a fixed ACD. The result is shown in Figure 7.

Due to the six-fold symmetry of the guest molecule, six equivalent sites are present on ACD, and the illustration should exhibit symmetry. This is the case in Figure 7a, which shows that the space is sufficiently covered by the ligand rotations. This means that the search algorithm generates all possible orientations in the space. A similar illustration was made for the translations (Figure 7b). It also shows that the algorithm provides coverage of the entire box with translations. In summary, the test confirmed that the SCD algorithm covers all the possible rotations and translations of a rigid ligand.

All the obtained energy minima within the energy window of 5 kcal/mol were clustered into families using the FAMILY program.[52,53] The clustering was performed such that the largest RMSD within one family was less than 0.5 Å. The results are shown in Figure 8. The X-ray structure is similar to family 2 found by CICADA docking, but the orientation of the ligand is opposite. The reason it is not close to the global minimum of the search probably is because more than two molecules are present in the crystal, which influence global forces that form the final structure. There are four basic families found by CICADA. The hydrogen bonding is the substantial force stabilizing families 3 and 4. Families 1 and 2 are stabilized mainly by van der Waals forces and electrostatic interactions.

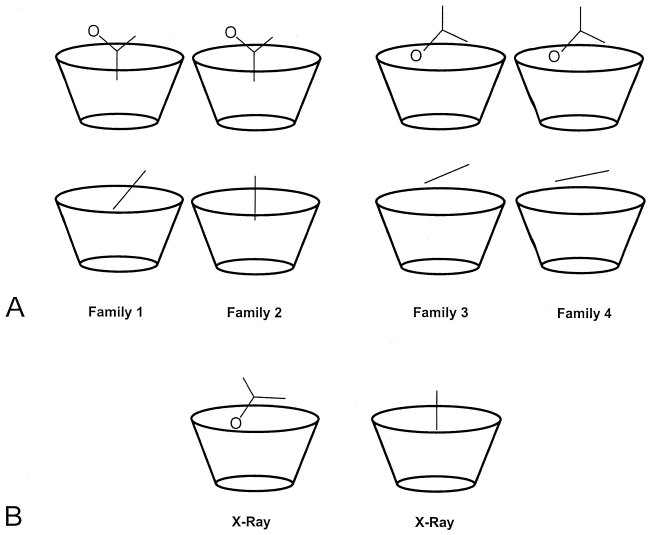

Figure 8. (A) Most populated docking modes found by CICADA for DMF/ACD docking. Relative energies of the lowest energy configuration in each single family are: 1 (0.0 kcal/mol), 2 (2.3), 3 (3.9), and 4 (4.4). Only configurations with relative energy less than 5 kcal/mol were included in the analysis. (B) Configuration of the DMF/ACD complex as obtained by X-ray crystallography[51] (relative energy before minimization 127.3 kcal/mol; after minimization 2.2 kcal/mol). Upper part shows the orientation of the CHO group of the ligand with respect to ACD. Because the molecule of DMF is planar, the bottom part shows the orientation of the plane with respect to the ACD cavity. For family 2, the orientation is almost parallel with the axis of the torus. This is not the case for the remaining families. The best fit between the X-ray minimized configuration and the CICADA results exhibits family 2 (RMSD = 0.30 Å). The same applies for the X-ray nonminimized configuration (RMSD = 0.55 Å).

Docking of R-Phenoxypropionic Acid

The second chosen system is composed of a beta-cyclodextrin (BCD) receptor (seven monomeric units) and an R-phenoxypropionic acid (Phe-O-CH(CH_3)-COOH; RPHA) ligand. In this case, all torsions of the main chain of RPHA were driven to ensure fully flexible docking. There is no crystal structure available for this complex. In this case, the starting structure for the search procedure was taken from the literature,[54] in which the authors performed several molecular dynamics (MD) runs to study the docking. The starting structure is shown in Figure 9.

This MD simulation predicted that the main driving force to form RPHA/BCD complex is the establishment of hydrogen bonds between the oxygen atom of the carboxylic function of RPHA and the OH groups on the sugar moieties, where these two functions are able to act as hydrogen bond donors and acceptors. The length of the hydrogen bonds typically is between 2.05 and 2.5 Å.

The CICADA calculation was stopped after 19,996 points were generated on the energy surface, of which 3719 were energy minima and 5604 were transition states. All the obtained energy minima within the energy window of 5 kcal/mol were clustered into families using the FAMILY program.[52,53] The clustering was performed such that the largest RMSD within one family was less than 0.5 Å. The procedure resulted in 57 families. The most populated families are shown in Figure 10 and the fundamental data collected are listed in Table 1.

Figure 10 shows that the most populated docking modes found by CICADA (Fig. 10a) are different from the equilibrium structure of the RPHA/BCD complex as generated by molecular dynamics using the MM3-96 force-field (MD, Fig. 10b).[54] Although the phenyl ring is completely out of the BCD

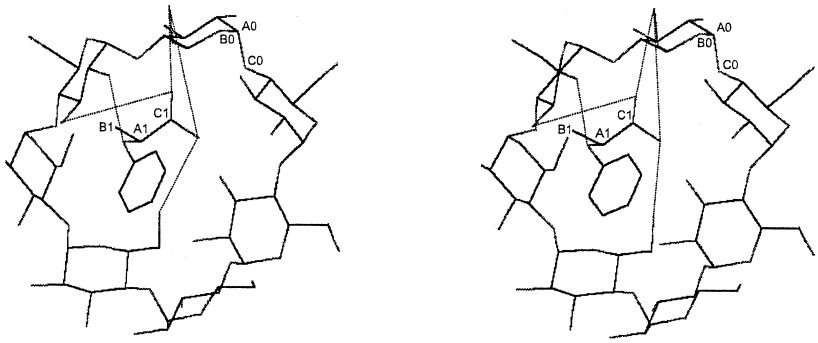

Figure 9. Structure of β-cyclodextrin and phenoxypropionic acid (Phe-O-CH(CH_3)-COOH) complex as obtained by CICADA with MM3 force field (the lowest energy dock). Intermolecular hydrogen bonds are shown by the shaded lines. See text for details.

cavity in the MD results (docking mode 2), it is partly inside the cavity in our results, having van der Waals contacts with the BCD (docking mode 1). In the absence of crystal structure, there is no direct validation for one method or the other. Nevertheless, our results are in good agreement with the experimental data[55,56] obtained by studying photochemical reactions in systems very similar to ours. Here, the authors showed that the ratio of *ortho* and *para* products of the photolysis of phenylesters is about 1.5:1 when BCD is not present, but the ratio is about 99:1 when BCD is used. This means that the *para* position is blocked in the latter case, indicating that the docking mode predicted by our calculations is more realistic. The probable reason why MD did not find the complexes with the phenyl ring inside the cavity is the high transition energy required for interconversion between modes 1 and 2. We calculated this barrier to be about 30 kcal/mol, which means that when one starts an MD run in the mode 2, the run probably never leaves this energy well when normal temperature is used.

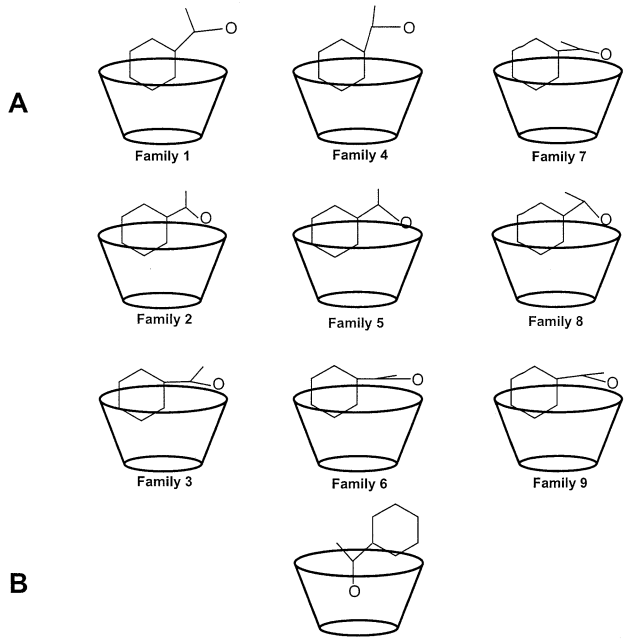

Figure 10. (A) Most populated docking modes found by CICADA for RPHA/BCD docking (docking mode 1). (B) Equilibrium configuration of the RPHA/BCD complex obtained by molecular dynamics[54] (docking mode 2, relative energy 3.5 kcal/mol).

CONCLUSIONS

We have developed software that allows us to use the SCD method for flexible molecular docking. The software has been implemented as an extension of the computer program CICADA, which originally was designed for conformational searching. The test examples, which fall into the chemistry of cyclodextrins, showed that the SCD approach is able to cover the entire box within which the search is performed. At the same time, it was determined that SCD can overcome high-energy barriers. The program is written such that more than one ligand molecule can be docked into the receptor.

The aim of the newly developed approach is not only to find the best dock for the receptor–ligand system, but also to describe all important binding modes as well as the dynamics within the cavity during the docking process. The data produced are prepared for use in a series of programs that afford

Table 1. Lowest energy nuclear configurations of the most populated docking modes found by CICADA for RPHA/BCD docking

Family	Dihedral angle (°)			Relative energy (kcal/mol)
1	168	80	40	0.0
2	−97	77	55	0.9
3	21	150	61	1.1
4	−91	79	57	1.2
5	97	79	39	1.4
6	−76	173	−47	1.6
7	110	173	−47	1.6
8	−134	−85	−48	1.8
9	70	61	−59	2.6
MD[a]	−6	84	100	3.5

Dihedral angles are denoted by: 1(149-148-154-155), 2(148-154-155-158), 3(154-155-158-157). For atom numbering see Figure 5.
[a] Values for the equilibrium configuration of the RPHA/BCD complex as obtained by molecular dynamics.[54]

detailed analysis of the system. For a review of these approaches, see Koč.[40] In the near future, the program will be tested on interactions that include biopolymers.

ACKNOWLEDGMENTS

A significant portion of the work presented here was performed during the visiting professorship of J.K. at the University of Joseph Fourrier in Grenoble. The research also was supported by The Grant Agency of The Czech Republic (Grant No. 201/98/K041) and by the Ministry of Education of The Czech Republic (Grant No. VS96095). All financial support is gratefully acknowledged. The authors also thank The Czech Academic Supercomputer Center in Brno for providing access to computer facilities. The authors are grateful to Mr. Laurence Benjamin, B.A., for language corrections.

REFERENCES

1 Vieth, M., Hirst, J.D., Domini, B.N., Daigler, H., and Brooks, C.L. III. Assessing energy functions for flexible docking. *J. Comput. Chem.* 1998, **19,** 1612–1622
2 Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 1998, **19,** 1639–1662
3 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 1982, **161,** 269–288
4 Gschwend, D.A., Good, A.C., and Kuntz, I.D. Molecular docking towards drug discovery. *J. Mol. Recogn.* 1996, **9,** 175–186
5 Vieth, M., Hirst, J.D., Domini, B.N., Daigler, H., and Brooks, C.L. III. Assessing search strategies for flexible docking. *J. Comput. Chem.* 1998, **19,** 1623–1631
6 Rosenfeld, R., Vajda, S., and DeLisi, C. Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.* 1995, **24,** 677–700
7 Goodsell, D.S., and Olson, A.J. Automated docking of substrates to proteins by simulated annealing. *Proteins* 1990, **8,** 195–202
8 Caflisch, A., Fischer, S., and Karplus, M. Docking by Monte Carlo minimization with a solvation correction: Application to an FKBP-substrate complex. *J. Comput. Chem.* 1997, **18,** 723–743
9 Di Nola, A., Roccatano, D., and Berendsen, H.J.C. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* 1994, **19,** 174–182
10 Clark, D.E., and Westhead, D.R. Evolutionary algorithms in computer-aided molecular design. *J. Comput.-Aided Mol. Design* 1996, **10,** 337–358
11 Judson, R.S., Jaeger, E.P., and Treasurywala, A.M. A genetic algorithm based method for docking flexible molecules. *J. Mol. Struct. (Theochem)* 1994, **308,** 191–206
12 Jones, G., Willet, P., and Glen, R.C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Design* 1995, **9,** 532–549
13 Judson, R.S., Tan, Y.T., Mori, E., Melius, C., Jaeger, E.P., Treasurywala, A.M., and Mathiowetz, A. Docking flexible molecules: A case study of three proteins. *J. Comput. Chem.* 1995, **16,** 1405–1419
14 Oshiro, C.M., Kuntz, I.D., and Dixon, J.S. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Design* 1995, **9,** 113–130
15 Ginzburg, S.L., Gel'fand, I.M., Vul, E.B., and Fedorov, Y. *The valley method in the problems of X-ray structure analysis.* Nauka, Moscow, 1966
16 Lugovskoy, A.A., Dashevsky, V.G., and Kitaigorodsky, A.I. Conformation of oxygen-containing heterocycles. Isomerization paths. *Tetrahedron* 1973, **29,** 287–295
17 Goto, H., and Osawa, E. Corner flapping: A simple and fast algorithm for exhaustive generation of ring conformations. *J. Am. Chem. Soc.* 1989, **111,** 8950–8951
18 Goto, H., and Osawa, E. How many conformers are there for small n-alkanes? Consequences of asymmetric deformation in GG′ segment. *J. Chem. Soc. Perkin. Trans. 2* 1993, 187–198
19 Goto, H., Osawa, E., and Yamato, M. An efficient algorithm for searching low-energy conformers for cyclic and acyclic molecules. *Tetrahedron* 1993, **49,** 387–396
20 Ngo, J.T., and Karplus, M. Pseudosystematic conformational search. Application do cycloheptadecane. *J. Am. Chem. Soc.* 1997, **119,** 5657–5667
21 Koča, J., and Carlsen, P.H.J. DAISY, a computational method. A novel tool for a study of conformational behavior of flexible molecules. *J. Mol. Struct. (Theochem)* 1992, **257,** 105–130
22 Koča, J. Computer program CICADA—Travelling along conformational potential energy hypersurface. *J. Mol. Struct. (Theochem)* 1994, **308,** 13–24
23 Allinger, N.L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* 1977, **99,** 8127–8132
24 Gajewski, J.J., and Gilbert, K.E. MMX, 1989, Serena Software, Bloomington, IN
25 Allinger, N.L., Yuh, Y.U., and Lii, J.-H. Molecular mechanics. The MM3 force field for hydrocarbons. *J. Am. Chem. Soc.* 1989, **111,** 8551–8566
26 Pearlman, D.A., Case, D.A., Cadwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., and Kollman, P.A. AMBER, 1991, University of California., San Francisco, CA
27 Fadrná, E., and Koča, J. Single-coordinate-driving method combined with simulated annealing. An efficient tool to search conformational space. *J. Phys. Chem. B* 1997, **101,** 7863–7868
28 Fadrná, E., and Koča, J. A combination of driving method with simulated annealing to search conformational space. *J. Mol. Struct. (Theochem)* 1997, **398–399,** 523–528
29 Koča, J., and Carlsen, P.H.J. Conformational behavior of ipsdienol analogues. I. 3-methyl-2-buten-1-ol, 2-methyl-6-methylene- 1,7-octadien-3-ol and 2-methyl-6-methylene-2,7-octadien-4-on. *J. Mol. Struct.* 1991, **246,** 165–177
30 Koča, J., and Carlsen, P.H.J. Conformational behavior of insect pheromones and analogues. Part II. *J. Mol. Struct.* 1992, **268,** 263–281
31 Engelsen, S.B., Pérez, S., Braccini, I., duPenhoat, C.H., and Koča, J. Travelling on the potential energy surfaces of carbohydrates: Comparative application of an exhaustive systematic conformational search with an heuristic search. *Carbohydr. Res.* 1995, **276,** 1–29
32 Imberty, A., Mikros, E., Koča, J., Mollicone, R., Oriol,

R., and Pérez, S. Computer simulation of histo-blood group oligosaccharides. *Glycoconjug. J.* 1995, **12,** 331–349
33 Koča, J., Pérez, S., and Imberty, A. Conformational analysis and flexibility of carbohydrates using the CICADA approach with MM3. *J. Comput. Chem.* 1995, **16,** 296–310
34 Casset, F., Imberty, A., Penhoat, C.H.d., Koča, J., and Pérez, S. Validation of two conformational searching methods applied to sucrose: Simulation of NMR and chiro-optical data. *J. Mol. Structure (Theochem)* 1997, **395–396,** 211–224
35 Koča, J., and Carlsen, P.H.J. Conformational behavior and flexibility of terminally blocked alanine di- and tri-peptides. *J. Mol. Struct.* 1993, **291,** 271–286
36 Koča, J., Kriz, Z., and Carlsen, P.H.J. Computer study of conformational flexibility of 20 common amino acids. *J. Mol. Struct.* 1994, **306,** 157–164
37 Koča, J., and Carlsen, P.H.J. Conformational behavior and flexibility of Met-enkephalin. *J. Mol. Struct. (Theochem)* 1995, **337,** 17–24
38 Kriz, Z., Koča, J., and Carlsen, P.H.J. Conformational behavior and flexibility of terminally blocked cysteine and cystine. *J. Mol. Model.* 1996, **2,** 51–61
39 Fadrná, E., and Koča, J. CICADA interface with AMBER. An application on oligonucleotides and their fragments. *J. Biomol. Struct. Dynam.* 1996, **14,** 137–152
40 Koča, J. Travelling through conformational space: An approach for analyzing the conformational behaviour of flexible molecules. *Progr. Biophys. Mol. Biol.* 1998, **70,** 137–173
41 Giacovazzo, C. Crystallographic computing. In: *Fundamentals of crystallography*, C. Giacovazzo, Eds., Oxford University Press, New York, 1992 pp. 61–73
42 Ponder, J.W. TINKER, Software Tools for Molecular Design, 1998, Washington University School of Medicine, St Louis, MO
43 Koča, J. Computer simulation of conformational movement based on interconversion phenomena. *J. Mol. Struct.* 1995, **343,** 125–132
44 Komiyama, M., and Takeshige, Y. Regioselective P-O(3′) cleavage of 2′,3′-cyclic monophosphates of ribonucleosides catalyzed by β- and γ-cyclodextrins. *J. Org. Chem.* 1989, **54,** 1336–1339
45 Chung, W.S., Turro, N.J., Silver, J., and le Noble, W.J. Modification of face selectivity by inclusion in cyclodextrins. *J. Am. Chem. Soc.* 1990, **112,** 1202–1205
46 Souter, R.W. *Chromatographic separation of stereoisomers.* CRC Press, Boca Raton, 1985
47 Allenmark, S.G. *Chiral liquid chromatography.* Blackie and Son, London, 1989
48 Allinger, N.L., Rahman, M., and Lii, J.-H. A molecular mechanics force field (MM3) for alcohols and ethers. *J. Am. Chem. Soc.* 1990, **112,** 8293–8307
49 Allinger, N.L., Zhu, Z.-Q., and Chen, K. Molecular mechanics (MM3) studies of carboxylic acids and esters. *J. Am. Chem. Soc.* 1990, **112,** 6120–6133
50 Allen, F.H., and Kennard, O. 3D search and research using the Cambridge Structural Database. *Chem. Des. Automat. News* 1993, **8,** 31–37
51 Harata, K. The structure of the cyclodextrin complex. VIII. Crystal structure of α-cyclodextrin complexes with 2-pyrrolidone and N,N-dimethylformamide. *Bull. Chem. Soc. Jpn.* 1979, **52,** 2451–2459
52 Imberty, A., and Perez, S. Molecular modelling of protein-carbohydrate interactions. Understanding the specificities of two legume lectins towards oligosaccharides. *Glycobiology* 1994, **4,** 351–366
53 Kriz, Z. In house software (unpublished), 1998
54 Manunza, B., Deiana, S., Pintore, M., Delogu, G., and Gessa, C. A molecular dynamics investigation on the inclusion of chiral agrochemical molecules in beta-cyclodextrin. The complexes with dichlorprop, phenoxypropionic acid and salithion (3rd Electronic Glycoscience Conference), 1997, http://antas.agraria.uniss.it/e−papers.html
55 Griffiths, D.W., and Bender, M.L. Cycloamyloses as catalysts. *Adv. Catalysis* 1973, **23,** 209–261
56 Breslow, R. Artificial enzymes. *Science* 1982, **218,** 532–537