Lennard-Jones Potential and Dummy Atom Settings to Overcome the AUTODOCK Limitation in Treating Flexible Ring Systems

Stefano Forli* and Maurizio Botta

Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, via Alcide de Gasperi 2, 53100 Siena, Italy

Received January 29, 2007

Here, we present a setting-up procedure of AutoDock parameters that allows the management of cycle and macrocycle flexibility during the docking process. In particular, the glue dummy atom type is introduced into calculations, and a novel empirical pseudo-Lennard-Jones potential function is applied to describe the intramolecular interactions occurring between two glue dummy atoms. The reliability of such an original protocol is tested by evaluation of 21 cyclic ligands in the corresponding binding site. As a result, the binding mode of 17 ligands is well-reproduced with respect to the X-ray crystallographic structure, with an root-mean-square deviation lower than 2 Å for 15 of them.

INTRODUCTION

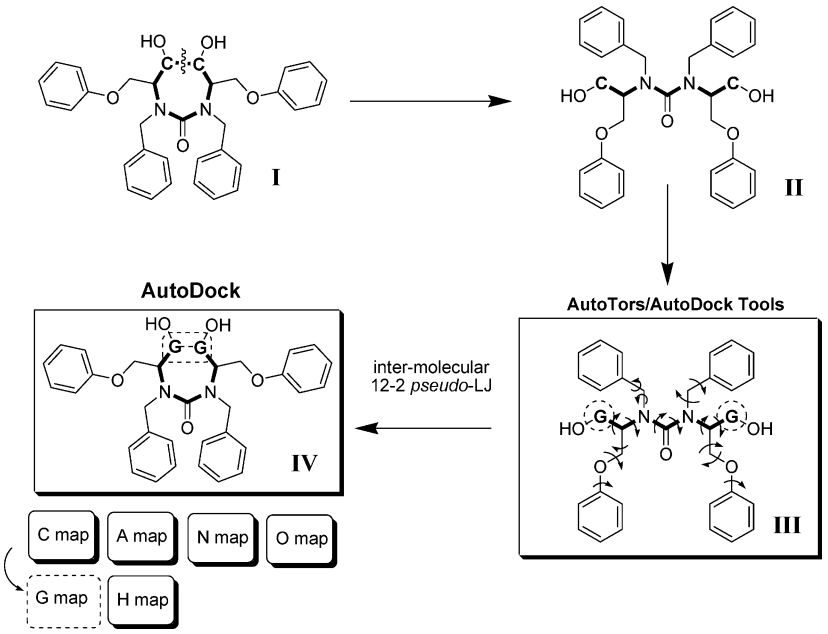
AutoDock[1] is an automated docking suite widely used to investigate the interactions between small molecules and protein or DNA macromolecules. At present, AutoDock is used in around 4000 academic, governmental, or nonprofit institutions,[2] and in a recent publication,[3] it is reported as the most cited docking software. Since version 3.0, AutoDock implemented an efficient genetic algorithm (GA) engine that is able to sample with high efficiency the conformational space of molecules to be docked, while results are evaluated using an empirical free-energy function derived from a training set of known complexes.[1] Moreover, AutoDock energy parameters can be modified to increase accuracy for custom needs or peculiar systems.[4] The suite is composed by three main programs, AutoGrid, AutoTors, and AutoDock, each program managing different steps of simulation. AutoGrid calculates interactions between ligand atoms and the protein structure and stores them in grid maps. AutoTors converts compounds to be docked from generic structure formats (PDB or mol2) to the PDBQ internal file format, describing the molecular structure as a series of branches departing from a root portion. Each branch can be allowed to rotate or not, during a docking calculation. Finally, AutoDock performs the actual docking of AutoTors formatted ligands on maps calculated by AutoGrid, using the GA engine. The whole process is a good compromise between calculation accuracy and computational speed. Despite the significant efficiency of the software in finding a reliable binding mode of ligands, AutoDock is however affected by several limitations. The most relevant of them is the treatment of the target structure as a rigid body during all calculations, often leading to unreliable results when small molecules are docked into binding sites where amino acid side-chains' movement occurs. To overcome such a limitation, though, some protocols have been developed in the past to simulate the protein flexibility together with the presence of crystallographic water molecules.[5,6] The explicit protein side-chain flexibility[2] will be implemented in the next release of AutoDock. Another limitation of the software is the inability to account for the flexibility of cyclic and macrocyclic ligands, due to the description model used for molecule structures. In particular, since endocyclic bonds are considered as not rotatable, cycles and macrocycles are treated as rigid systems. The only way to perform docking simulations involving cyclic structures is to choose a low-energy conformation as the input structure to be docked[7] or to perform many different docking calculations using multiple conformations of the same compound sampled from molecular dynamics or conformational searches.[8] This approach moreover suffer from significant limitations, especially when ligands bearing large cycles are under study or many energetically close conformations do exist. To overcome the major limitation of AutoDock software consisting of the inability to treat the flexibility of cyclic ligands, we describe here the introduction in AutoDock 3.0.5 of a new atom type (here referred to as the *glue* dummy atom) and an empirical Lennard-Jones (L-J) function that allow the docking of cyclic molecules as entirely flexible entities. The general workflow of this original protocol is outlined in Scheme 1.

Ligand Breaking. The first step of the computational protocol consists in the identification of a bond that is broken to transform the cyclic structure (compound I, Scheme 1) into the corresponding acyclic form (compound II, Scheme 1), which could be treated in the usual way by means of the AutoTors routine. In particular, all the current rotatable bonds (mainly those originally belonging to the cycle) are allowed to rotate freely, while their conformations can be explored by GA and adapted to the target binding site. Moreover, the edge C atoms previously connected by a covalent bond are renamed as G (*glue*) atoms (compound III, Scheme 1), for which appropriate parameters have to be applied (see below). In order to optimize the subsequent docking procedure, a simple guideline is proposed below to appropriately select the covalent bond to be broken:

* Corresponding author tel.: +39-0577-234307; fax: +39-0577-234333; e-mail: forli@unisi.it.

10.1021/ci700036j CCC: $37.00
Published on Web 06/22/2007

Scheme 1. Workflow of the Computational Protocol Applied to Dock Cyclic Structures as Flexible Entities



(*1) The bond resulting in the acyclic structure bearing the lower number of rotatable bonds should be privileged.* In fact, since the bond breaking increases the number of rotatable bonds to be considered during calculations, the choice of bond where the ring will be open should be made to obtain an acyclic structure with the lowest flexibility (in terms of rotatable bonds). This routine reduces complexity and calculation time without compromising accuracy. Therefore, when less flexible or partially rigid regions are present, they should not be broken. Furthermore, when other structural information is available (i.e., NMR studies), some bonds could be kept fixed.

(2) The bond should be between two carbon atoms (or two identical atom types). This is due to the fact that no maps of the *glue* atom types are calculated, but copies of the original C atom-type maps will be used for calculations. Therefore, for the sake of simplicity, to use the same map for both *glue* atoms, the choice of a bond between two atoms of the same type is suggested.

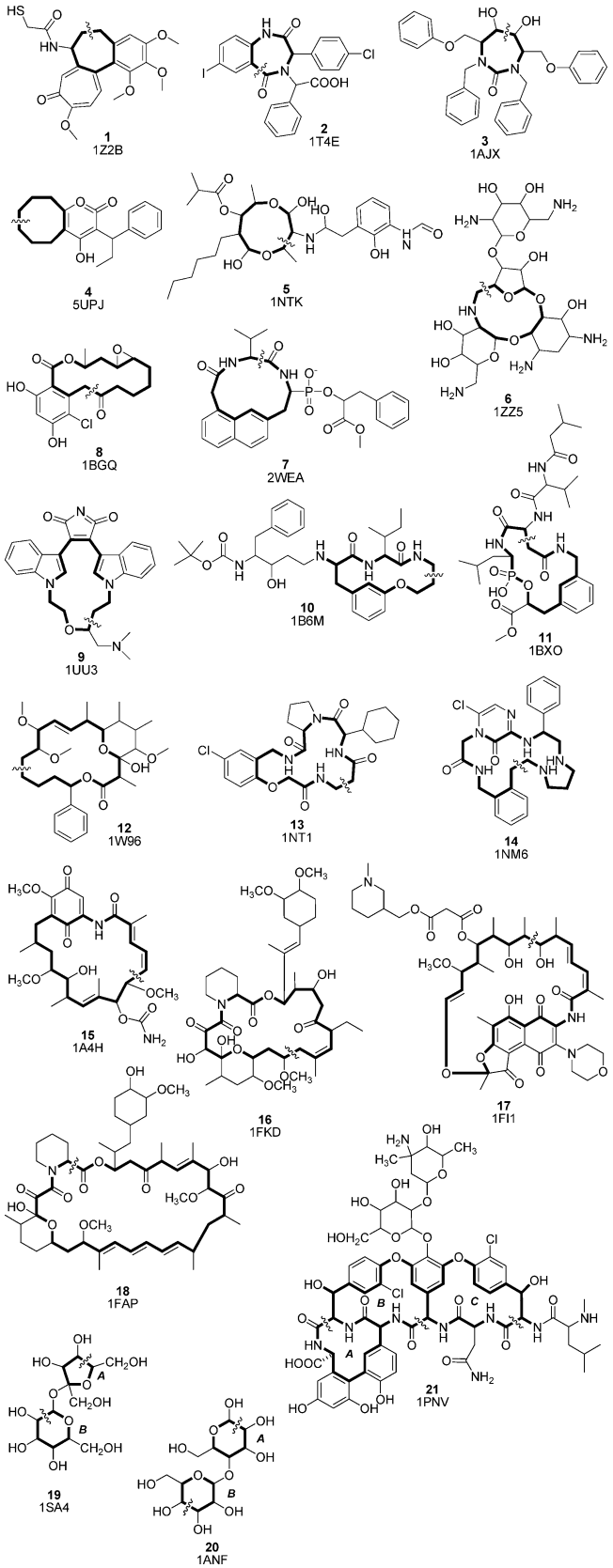
(3) Whenever possible, nonchiral carbon atoms should be preferred. In fact, because of the lack of directionality in the interaction definition of glue atoms, preservation of the original chirality is not guaranteed. Nevertheless, none of these guidelines is a mandatory requirement, even if it is suitable. Consequently, they can be neglected whenever required or when no other choices are available (i.e., no achiral carbon atoms are present, as in compounds **19** and **20**, Chart 1).

Ring Closure and Modified Lennard-Jones Potential. The next step of the protocol is the application of a constraint forcing the *glue* atoms to locate themselves at a distance compatible to a bond length, thus transforming, during the calculation, the acyclic molecular structure into a conformation mimicking the original cyclic structure (compound IV, Scheme 1). In this context, AutoDock provides a constraint function (set by the "constraint" keyword) that only allows the rejection of final solutions which do not satisfy a user-defined distance criterion. In opposition, our aim is the application of a force that bias GA to find solutions evolving toward ring closure during docking calculations. Setting up a docking process with AutoDock requires that a panel of the internal pairwise nonbonded energy be explicitly defined for all the atom types of the ligand, providing numerical values of exponents and parameters for Lennard-Jones potential.

$$V(r) \approx \frac{\frac{m}{n-m}\epsilon r_{eqm}^{n}}{r^{n}} - \frac{\frac{n}{n-m}\epsilon r_{eqm}^{m}}{r^{m}} \qquad (1)$$

Equation 1 represent the simplified Lennard-Jones potential implemented in AutoDock obtained after rearrangement of the original Lennard-Jones equation,[9] where the first term represents the repulsive component and the second term is the attractive component. Usually, $n = 12$ and $m = 6$ values are applied to model van der Waals' interactions, and $n = 12$ and $m = 10$ values are applied to model hydrogen bonds. The remaining parameters depend on the type of nonbonded atoms: r_{eqm} represents the internuclear equilibrium distance, and ϵ represents the potential well depth at the equilibrium internuclear separation. To use the glue atom type for our purpose, L-J parameters are customized using the "intnbp_r_eps" keyword to define the new G atom type. In particular, since G atoms are meant to mimic a couple of bonded carbon atoms belonging to a ring, the optimal equilibrium distance should approximately correspond to the distance between two carbon nuclei in a covalent bond (1.2−1.5 Å); accordingly, the r_{eqm} value is set to 1.5 Å. Moreover, to make the G−G interaction extremely favorable, the well depth ϵ is raised to 10, about 3 orders of magnitude higher with respect to the value used for C−C van der Waals interaction. However, preliminary calculations showed that these parameters give unsatisfactory results in terms of ring closure. Consequently, n and m values of the L-J equation are also varied, leading to an empirical pseudo-Lennard-Jones equation obtained by a reduction of the attractive term exponent of the van der Vaals L-J from 6 to 2 and characterized by a smoother profile with respect to typical 12−10 or 12−6 equations (see Figure 1). This new L-J

Chart 1. Compounds Considered in This Study



equation (a 12−2 function with $\epsilon = 10$ and $r_{eqm} = 1.5$) is characterized by a dramatically strong and long-range attraction force. In fact, its effect measured at a 20 Å distance is comparable to the effect of the hydrogen bond 12−10 Lennard-Jones potential calculated at around a 2 Å distance. Figure 1 shows Lennard-Jones equations used in AutoDock to define the potential energy for van der Waals interactions (12−6 potential) and hydrogen-bond interactions (12−10) in comparison to the custom 12−2 pseudo-L-J function.

Docking Setup for Linear Structure and GA Parameters. For docking purposes, G atoms are considered as carbon atoms, in terms of interactions with all non-G atoms, in intermolecular (with target atoms) as well as in intramolecular interactions (with other atoms present in the ligand). Therefore, to calculate intermolecular interactions involving G atoms, the C map is provided to the software, omitting the calculation of a custom grid map for the G atoms. In a similar way, the C atom parameters are used to describe the intramolecular interactions between G atoms and the remaining atoms, while the G−G interaction is described with the pseudo-L-J 12−2 potential function further implemented with custom ϵ and r_{eqm} values. The rest of the docking parameter file is tuned following general procedures for AutoDock calculations, with special care given to parameters necessary to counterbalance the increase of flexibility due to the cycle opening. In particular, the Lamarckian genetic algorithm (LGA) offers more advantages with respect to the pure GA, because of the possibility of a further refinement of results and the chance to transmit improvements to the offspring.[1] For this purpose, some good advice could be to increase the local search frequency for the pseudo-Solis and Wet algorithm, thus increasing the probability of a "local energy minimization" of a partial result. This favors the G−G distance to reach the equilibrium value, encouraging the ring closure process. Nevertheless, this aspect could be neglected, obtaining more approximately closed structures but saving computational time, because of the eventual refinement of results as commonly suggested.[9] Other docking parameters can be modified by following the AutoDock user guide. A docking parameter file modified to include parameters for the G atoms is reported as an example in Table 1.

VALIDATION

To find experimental data to be used for validating this new protocol, the Brookhaven Protein Data Bank is filtered, searching for cyclic ligand structures suitable for docking calculations. With use of the on-line interface of the database, a search is performed to collect crystallographic structures containing 7- to 30-membered cyclic compounds with at least two adjacent carbon atoms (see Materials and Methods). As a result (updated on August 16, 2006), 337 nonunique entries are found. The frequency distribution of entries with respect to the ring size is depicted in Figure 2.

Entries containing metal complexes, transition state structures, and covalently bound ligands are rejected, as well as complexes where several ligands are present in the same binding site (i.e., cofactors, ATP, vis-à-vis binding, nonwater solvent molecules, etc.). The remaining structures are visually inspected, leading to the arbitrary choice of a subset of 21 representative entries whose ligand structures are selected to be tested. The structures of ligands docked in this study are depicted in Chart 1. For each of them, molecular target structures are obtained by removing all heteroatoms (see Case Studies for exceptions) and processing the resulting structures with the AutoDock Tools[10] suite. Moreover, grid maps manually centered on the binding site are calculated for required atom types, using the default resolution (0.375 Å;

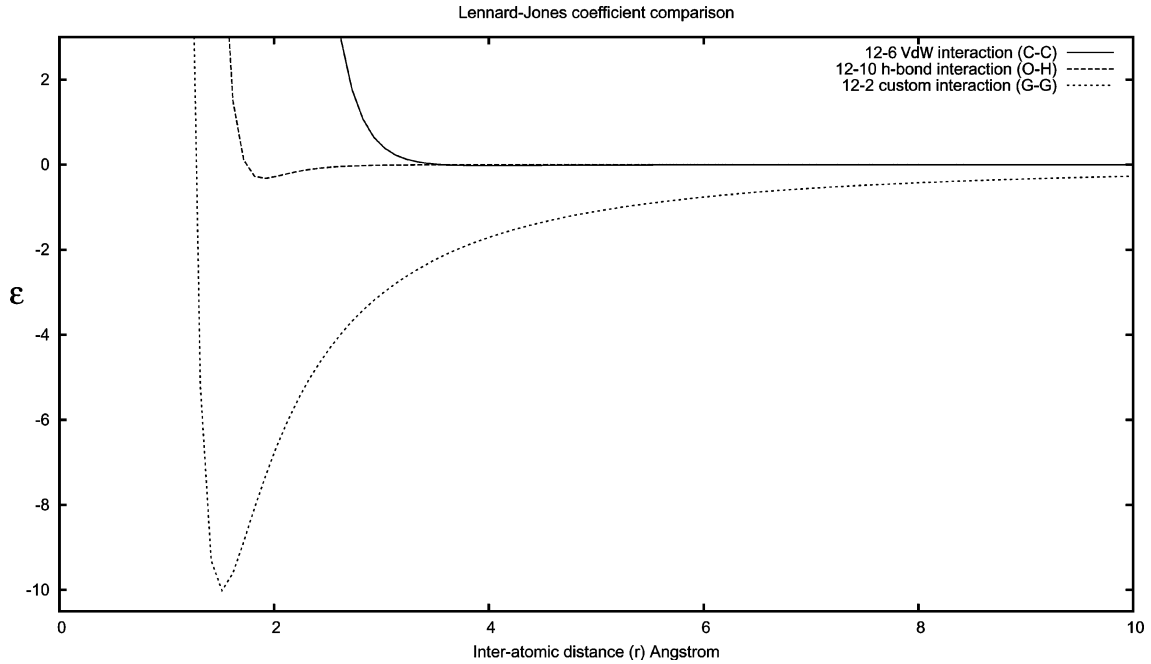


Figure 1. Graphical comparison between 12−6, 12−10, and 12−2 Lennard-Jones potentials (without the AutoDock internal smoothing function[1,9]).

Table 1. Example of a Generic Docking Parameter File Modified to Include *glue* Atom Parameters[a]

Atom Types	
types CAOHG	# atom type names[a]
Grid Maps	
map protein.C.map	# atom-specific affinity map C
map protein.A.map	# atom-specific affinity map A
map protein.O.map	# atom-specific affinity map O
map protein.H.map	# atom-specific affinity map H
map protein.C.map	# atom-specific affinity map G[a]
map protein.e.map	# electrostatics map
Intramolecular Interactions	
intnbp_r_eps 4.00 0.0222750 12 6	# C−C lj
intnbp_r_eps 4.00 0.0222750 12 6	# C−A lj
intnbp_r_eps 3.60 0.0257202 12 6	# C−O lj
intnbp_r_eps 3.00 0.0081378 12 6	# C−H lj
intnbp_r_eps 4.00 0.0222750 12 6	# C−G lj[a]
intnbp_r_eps 4.00 0.0222750 12 6	# A−A lj
intnbp_r_eps 3.60 0.0257202 12 6	# A−O lj
intnbp_r_eps 3.00 0.0081378 12 6	# A−H lj
intnbp_r_eps 4.00 0.0222750 12 6	# A−G lj[a]
intnbp_r_eps 3.20 0.0297000 12 6	# O−O lj
intnbp_r_eps 1.90 0.3280000 12 10	# O−H hb
intnbp_r_eps 3.60 0.0257202 12 6	# O−G lj[a]
intnbp_r_eps 2.00 0.0029700 12 6	# H−H lj
intnbp_r_eps 3.00 0.0081378 12 6	# H−G lj[a]
intnbp_r_eps 1.51 10.000000 12 2	# G−G plj[a]

[a] Refers to specific customizations for the G atom types.

see Materials and Methods for detailed setup). Ligand structures are imported into the software in their acyclic form, and *glue* atoms are manually added. The docking parameter files are modified manually to include parameters for *glue* atom interactions (pseudo-Lennard-Jones potential, r_{eqm}, and ϵ values). Docking parameters are varied to account for the increased system complexity due to the higher number of rotatable bonds of acyclic structures with respect to cyclic compounds. Therefore, while default values of LGA parameters are sufficient to dock ligands with a few rotatable bonds, or small binding sites, more flexible molecules or more complex binding cavities require larger populations and an increase in energy evaluations. The lowest values correspond to a population size of 50 individuals, 25×10^4 energy evaluations, and 27×10^3 generations, while more in-depth calculations require higher values such as 450 individuals, 15×10^5 energy evaluations, and 15×10^5 generations, as suggested for peptide docking.[11] Pseudo Solis and Wets algorithm parameters are always kept at default values, except for an increase in the local search probability. The maximum value used for the local search frequency is 0.26. Greater values should be better but unwise, because of the increase of calculation time required. To obtain statistically significant results, a large number of conformations (from 510 to 1275 runs) is collected for each complex by running multiple docking jobs. The calculation time varied according to degrees of freedom of the molecules, the binding site volume, and the number of runs, ranging from less than 1 h per job (i.e., ∼30−50 mins for **1**, **2**, and **4**), to several hours (i.e., **5**, **6**, **10**, **11**, **19**, and **20**), and up to several days (two jobs of ∼25 days for **21**).

RESULTS AND DISCUSSION

To exclude the possibility of chance ring closure, some test dockings are performed without including the *glue* atoms. Among those results, very few solutions are found in a conformation that can be in some way compared to the closed ring state. Moreover, during calibration of the parameters for the pseudo-LJ function, the ring closure did not occur at all, or occurred with poor results when using combinations of coefficients and parameters other than those proposed here. During the docking, the genetic algorithm is supposed to encourage an evolution of intermediate solutions with a progressively smaller separation between the *glue* atoms, aiming to the equilibrium distance r_{eqm} because of the energetic contribution of the pseudo-Lennard-Jones potential. Once closed state solutions become predominant, selection should occur by means of ligand−protein interaction energy. Although contributions from the pseudo-Lennard-Jones

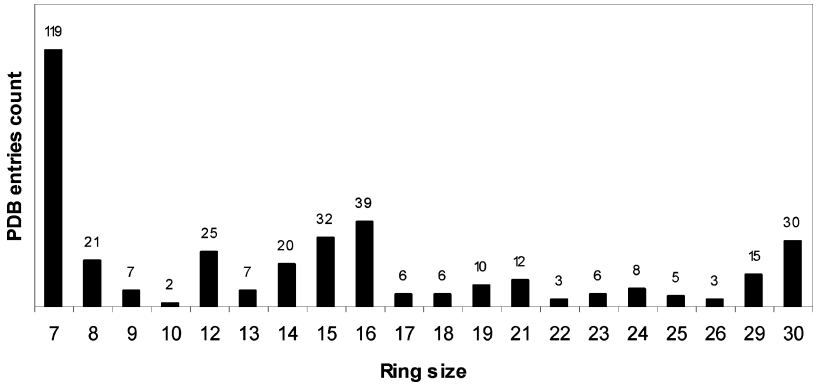


Figure 2. Distribution of cyclic molecules in PDB.

potential are large, open-form conformations are rarely found. Moreover, no bumps between the target structure and ligands are found, confirming that customization does not affect the accuracy in the AutoDock calculations.

Results Evaluation. While the use of AutoDock energy ranking is proven to be a reliable scoring function,[1,12−14,30] and custom potential functions do not appear to affect the overall accuracy, in this study, it is sometimes found to be insufficient for results analysis. Most of the issues are related to the intrinsic ligand complexity, because of the high flexibility (rotatable bonds are ≥ 10 in 17 of 21 compounds) and high molecular weight of some of them, causing positive final docked energies in some results. Another energy-related issue is the presence of "chance hits": clusters scarcely populated that are ranked as the best result, shifting down the clusters containing the correct result. Those "chance hit" clusters can be easily identified because of the limited number of individuals and the very low energy gap (usually less than 0.7 kcal/mol in docked energy) with the following clusters. This approximation in the energy estimation has been ascribed to actual limitations of the molecular mechanics and the discrete step in grid calculations[11] used in docking. Because of all of the aforementioned reasons, the statistical relevance of clusters can be a reliable indication in the choice of the correct result; then, the cluster population ranking is used here as a score result parameter beside the usual energetic ranking. Therefore, a pose is considered as a good result when it (*a*) matches the ring alignment to the binding site, (*b*) arranges chemical moieties in close contact with correct regions of the binding pocket, (*c*) is found among the top three cluster results, by energy or by population ranking, and (*d*) the *glue* atom pair is separated by less than 3 Å, and then the ring closure occurs. The last is not considered as restrictive filter, because, if the *glue* atom distance is bigger than the threshold but previous criteria are satisfied, the cyclization can be obtained afterward when performing the commonly recommended refinements[9] (energy minimization or molecular dynamics simulations). The results analysis is performed in two steps. In the first step, docked conformations are clustered by calibrating the root-mean-square deviation (RMSD) tolerance to the lowest value, allowing the collection of similar alignments in the same cluster. Small cluster tolerances (1−1.5 Å) are sufficient for simple compounds, while more complex results required higher values (greater than 2.5 Å), due to ligand or binding site complexity, or both. The second step is the identification of the cluster containing the X-ray binding mode, and the calculation of RMSD with respect to the X-ray conformation. Despite expectations, almost all of the docking results satisfied the imposed criteria, correctly placing the rings into the binding site and establishing known interactions. Moreover, some results show very good agreement with experimental data. The summary of all calculations performed is presented in Table 2. In the overall analysis, the cluster containing the conformation closest to the X-ray conformation is found as the top-ranked result by energy in 10 out of 21 docked complexes (**1**, **2**, **3**, **5**, **8**, **9**, **10**, **12**, **16**, and **18**) and as the most populated cluster in 11 dockings out of 21 (**2**, **3**, **5**, **8**, **9**, **10**, **12**, **16**, **18**, and **19**). In four cases (**2, 3**, **5**, and **10**), the X-ray conformation is found to be both the top-energy-ranked and most populated cluster. Altogether, in 17 dockings, the X-ray-like conformation can be found either in the best energy result or in the most populated cluster. In 15 systems, the RMSD between the top-ranked pose and the X-ray conformation is less than 2 Å, while in three cases (**1**, **4**, and **12**), the RMSD is less than 1 Å. In one case, the cocrystallized structure of molecule **17** in PDB entry 1FI1 is present with the wrong planar conformation in the piperidine ring. Flexible molecules with a high number of active torsions (≥20) are found to be more difficult to dock, while heavy molecular weight compounds are often scored with positive energetic values. Both problems are already reported in the literature.[11,14] By visual inspection of the results, unclosed ring conformations are rarely found, and the frequency is slightly related to the ring size. Generally, the most uncertainty on the ring conformation is present when there are no direct interactions between functional groups and the target structure (i.e., solvent-exposed ring portions). From a subranking analysis of the results, the poses with the lowest RMSD with respect to X-ray coordinates are seldom found ranked at a high position, except for two cases (**11** and **20**). This trend should be due to the already mentioned approximations,[11] but it can be improved with more detailed calculations (i.e., incremented grids resolution). Incrementation in the local search frequency is found to lead to a slight overall improvement in subranking. In Figure 3 are depicted some exemplifying results (**1**, **3**, **10**, and **17**), and the four compounds from case studies (**8**, **19**, **20**, and **21**) overlapped onto their respective conformations as from X-ray structures. The results obtained with this protocol are even better with respect to the usual performance in docking acyclic structures, reported in the literature for AutoDock calculations. In fact, about 80% of the X-ray crystallographic structures constituting case studies are successfully reproduced in terms of either orientation within the binding site, interactions with the binding site, or RMSD values (lower

Table 2. Summary Report of Compounds' Properties, Docking Calculations, and Final Results[a]

		ligand			docking		clustering			RMSD		
								ranking[b]				
cpd[c]	PDB ID	ring size	active torsions	heavy atoms	runs	cluster tolerance (Å)	total number (multimember clusters)	energy	population (individuals)	top cluster result RMSD[d]	lowest RMSD (subranking)[e]	ligand name
1	1Z2B	7	9	30	1020	1.5	128 (76)	1	5 (46)	0.84	0.67 (24)	colchicine
2	1T4E	7	10	32	1020	2	152 (90)	1	1 (141)	1.10	0.83 (36)	di-chloro-benzo-diazepine
3	1AJX	7	18	40	1020	2.3	310 (147)	1	1 (52)	1.81	1.32 (32)	5,6-dihydroxy-1,3 diazepan-2-one
4	5UPJ	8	8	23	1020	1	173 (80)	2	1 (392)	0.74	0.63 (193)	cyclo-octa[b]pyran-2-one
5	1NTK	9	17	38	510	2.5	90 (55)	1	1 (85)	2.25	1.87 (19)	antimycin A1
6	1ZZ5	11	20	41	510	2.5	149 (91)	54	16 (3)	4.02	2.82 (2)	neomycin
6	$1ZZ5_{wat}$	11	20	41	510	2.5	214 (112)	59	8 (7)	3.20	2.64 (6)	neomycin
7	2WEA	12	16	39	510	2	321 (90)	3	1 (16)	1.93	1.79 (14)	P3−P1 macrocyclic peptidyl
8	1BGQ	14	10	23	1020	2	103 (59)	19	6 (50)	2.06	1.66 (42)	radicicol
8	$1BGQ_{wat}$	14	10	23	1020	2	134 (83)	1	5 (45)	1.40	0.82 (10)	radicicol
9	1UU3	14	9	35	510	2	288 (84)	1	2 (24)	1.12	0.74 (18)	LY333531
10	1B6M	15	19	43	1020	2.5	314 (160)	1	1 (69)	1.32	1.14 (2)	macrocyclic peptidomimetic inhibitor 6
11	1BXO	15	20	44	510	2.5	282 (56)	3	4 (3)	1.30	1.30 (1)	PP7
12	1W96	16	16	37	510	2	214 (91)	1	2 (71)	0.96	0.86 (3)	soraphen A
13	1NT1	17	9	35	1020	2	114 (74)	6	2 (102)	1.65	0.94 (29)	cycloeptadecine
14	1NM6	19	12	38	1275	2	688 (230)	2	1 (48)	1.42	1.03 (25)	benzo-hexa-aza-cyclo-henicosine
15	1A4H	19	13	40	510	2.5	141 (86)	3	3 (14)	2.11	1.63 (7)	geldanamycin
16	1FKD	21	15	77	510	2.5	291 (112)	1	3 (7)	2.31	1.90 (4)	ascomycin
17	1FI1	24	18	67	510	2.7	319 (84)	4	1 (14)	1.73	1.72 (3)	ryfamycin CGP 4832
18	1FAP	29	17	65	510	2.5	25 (18)	1	2 (96)	2.80	1.99 (36)	rapamycin
19	1SA4	5+6	22	23	1020	2	170 (107)	3	1 (102)	1.77	1.06 (13)	sucrose
20	1ANF	6+6	22	23	1275	2	71(54)	2	1 (516)	1.41	1.00 (12)	maltose
21	1PNV	12+16+16	46	101	510	3	463 (37)	26	1 (5)	3.11	3.11 (1)	vancomycin

[a] RMSDs are calculated on heavy atoms only. [b] Ranking (by energy or by population) of cluster containing the X-ray binding mode. [c] Subscribed index indicates dockings performed in the presence of water molecules (see *Case Studies*). [d] RMSD of lowest docked energy result in cluster containing the X-ray binding mode. [e] RMSD and cluster subranking of conformation that is the most similar to the X-ray contained in the cluster.

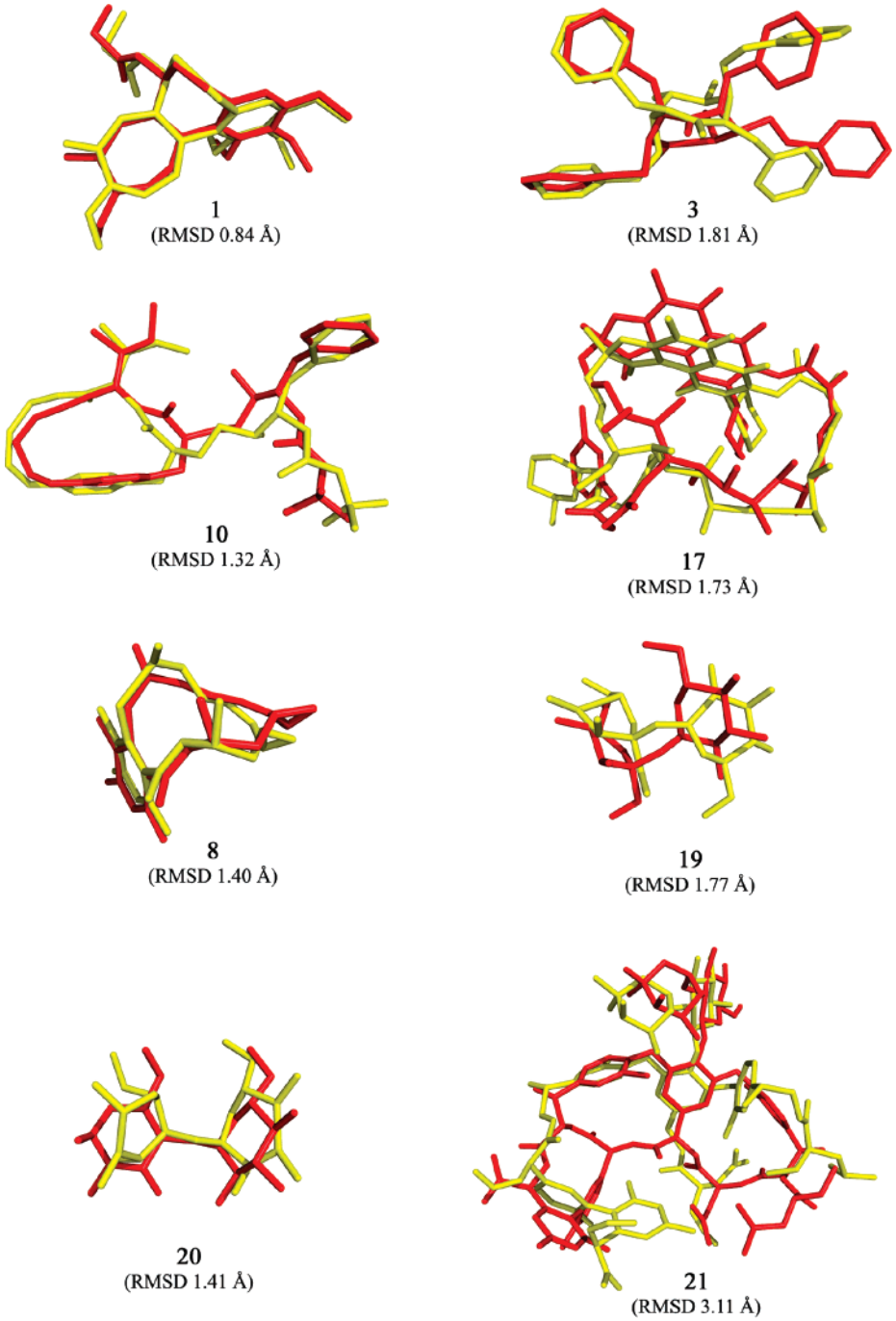


Figure 3. Some representative results (ligands **1**, **3**, **10**, and **17**) and case study compounds (ligands **8**, **19**, **20**, and **21**). X-ray crystal structures are in red; docked structures are in yellow; hydrogen bonds are not represented. RMSD values are calculated for heavy atoms only.

than 3.0 Å). As a comparison, in a recent paper, the success rate of AutoDock (with a RMSD lower than 3.0 Å) is reported to be slightly higher than 70% for calculations made on 100 protein−ligand complexes.[30]

CASE STUDIES

Water Case Study: Radicicol. Water molecules often play a very important role in the binding of small molecules into receptor structures.[15,18] In general, the water molecule position is defined when crystal structures of known ligands are obtained with high resolution, but when investigating the binding mode of novel molecules, little or no information is available about it. In such cases, suggestions about the presence and position of water molecules within a protein structure can be provided by computational methods. For example, probing the protein surface with the GRID[16] software is widely accepted as a useful tool to identify regions of profitable interactions with water molecules.[17−19] Moreover, different techniques have been proposed to include the presence of explicit water in AutoDock calculations.[5,6] In this study, we arbitrarily decided to neglect any prior knowledge of the water molecule's location, in order to simulate the unfavorable conditions in which the localization of the ligand binding site is the only information available. The presence of water molecules within the binding site is known to affect the overall accuracy of a docking protocol,

Table 3. Modified Lines in Docking Parameter File for Compound **8**

Atom Types	
types CAOHcG	# atom type names[a]
Grid Maps	
map protein.C.map	# atom-specific affinity map C
map protein.A.map	# atom-specific affinity map A
map protein.O.map	# atom-specific affinity map O
map protein.H.map	# atom-specific affinity map H
map protein.c.map	# atom-specific affinity map c
map protein.C.map	# atom-specific affinity map G[a]
map protein.e.map	
Intramolecular Interactions	
intnbp_r_eps 4.00 0.0222750 12 6	# C−C lj
intnbp_r_eps 4.00 0.0222750 12 6	# C−A lj
intnbp_r_eps 3.60 0.0257202 12 6	# C−O lj
intnbp_r_eps 3.00 0.0081378 12 6	# C−H lj
intnbp_r_eps 4.04 0.0302197 12 6	# C−c lj
intnbp_r_eps 4.00 0.0222750 12 6	# C−G lj[a]
intnbp_r_eps 4.00 0.0222750 12 6	# A−A lj
intnbp_r_eps 3.60 0.0257202 12 6	# A−O lj
intnbp_r_eps 3.00 0.0081378 12 6	# A−H lj
intnbp_r_eps 4.04 0.0302197 12 6	# A−c lj
intnbp_r_eps 4.00 0.0222750 12 6	# A−G lj[a]
intnbp_r_eps 3.20 0.0297000 12 6	# O−O lj
intnbp_r_eps 1.90 0.3280000 12 10	# O−H hb
intnbp_r_eps 3.65 0.0348827 12 6	# O−c lj
intnbp_r_eps 3.60 0.0257202 12 6	# O−G lj[a]
intnbp_r_eps 2.00 0.0029700 12 6	# H−H lj
intnbp_r_eps 3.04 0.0110335 12 6	# H−c lj
intnbp_r_eps 3.00 0.0081378 12 6	# H−G lj[a]
intnbp_r_eps 4.09 0.0409860 12 6	# c−c lj
intnbp_r_eps 4.04 0.0302197 12 6	# c−G lj[a]
intnbp_r_eps 1.51 10.000000 12 2	# G−G plj[a]

[a] Refers to specific customizations for the G atom types.

and it is also expected to influence the process for which the two G atoms move closer. To account for those effects, simulations are performed to dock radicicol (**8**) into the ATP binding site of Hsp90. The PDB entry 1BGQ contains the complex between the Hsp90 yeast chaperone protein[20] and **8**, a 14-term macrocyclic antibiotic with antitumoral activity[20] that acts as a competitor of ATP for its binding site. Interactions between Hsp90 and ATP are mediated by eight water molecules, while only two water molecules are directly involved in protein−radicicol interactions.[20] Two different parallel calculations are performed on **8**, including ($1BGQ_{wat}$) or not including (1BGQ) two water molecules (HOH_{401} and HOH_{403}) into target structures (see Materials and Methods).

Target Preprocessing. AutoGrid calculations for CAOHc atom types are performed on both 1BGQ and $1BGQ_{wat}$, obtaining a pair of map sets. The grid center is defined approximately at the center of mass of the ligand, in such a way as to embrace the whole binding site.

Ligand Preprocessing. The ligand in acyclic form is obtained by breaking bonds reported in Chart 1, allowing all rotatable bonds (10) to freely rotate during the GA calculation.

Docking and GA Parametrization. The C map is defined for the G atoms; then, intramolecular interactions are integrated with the G atoms using the usual C parameters and the pseudo-Lennard-Jones potential. The parameters used for the calculations on the Hsp90−**8** complex are reported in Table 3. The GA population size is raised to 150 individuals, with 5.5×10^5 evaluations and 3.7×10^3 generations. The local search frequency is 0.16.

Docking Analysis: No Water Molecules (1BGQ). Ring closure occurs in most of the total conformations, but all high-ranked results are very far from the binding mode found for **8** in the crystal structure. The best energy conformation engages the binding site in a way roughly similar to the X-ray pose, but the chloro-phenil ring is flipped in such a way that the chlorine atom is facing Asp_{79} instead of being in close contact with Phe_{124}. The direct consequence is the incorrect arrangement of the macrocycle, which gains closure by adopting a reverse conformation that forces the anti epoxide ring to point away from Lys_{44}. In this way, the oxygen atom is unable to form the hydrogen bond with the protonated nitrogen of Lis_{44}, lying at more than 7 Å. The correct alignment of the aromatic ring is found only in very low-energy-ranked conformations (19th cluster) that, however, lack the hydrogen bond with Lys_{44}.

Docking Analysis: Two Water Molecules ($1BGQ_{wat}$). Accounting for water molecules during the docking of **8** into the ATP binding site of Hsp90 dramatically improves results. In fact, the conformation most similar to the X-ray conformation belongs to the best energy cluster. The population of the cluster (45 individuals) shows good statistical relevance, and ring closure occurs in all cluster members. The key interactions between ring and water molecules are found in all conformations, and they drive the correct orientation of the chloro-phenyl ring into the main cavity and the formation of the hydrogen bond between the epoxide ring and Lys_{44}. Even if some variance in the epoxide placement is found, this is one of the most stable parts of the molecule together with the aromatic ring. On the contrary, the unfunctionalized aliphatic portion of the macrocycle, that does not interact directly with any amino acid, is the most varying region. The docked conformation of **8** overlapped with the X-ray conformation is shown in Figure 3. All of this evidence suggests that water molecules play a principal role in the placement of **8** into the binding site of Hsp90, and that docking calculation accuracy is strongly influenced by their presence. Therefore, while ring closure is not an issue, the macrocycle conformation is influenced by the same factors influencing the common dockings and can be improved in the same fashion. Hsp90 is able to complex both **8** and **15.** However, due to the lower number of interactions that **8** can establish, its binding is more sensitive to the presence of water molecules. Accordingly, the binding mode of **15** is reproduced with sufficient accuracy even if no water molecules are included in the target structure (see Table 2). A similar calculation without and with water molecules is performed also for **6** (1ZZ5 and $1ZZ5_{wat}$, respectively) with very poor results in both cases. The ring closure is always achieved, but molecule arrangement with respect to the RNA target structure is completely wrong. This could be ascribed to an intrinsic limitation of the software in the docking of ligands to nucleic acids.[6,21]

POLICYCLIC SYSTEMS

When the *glue* atoms concept is extended, it can be possible to manage more than one ring per molecule by breaking more bonds and defining the appropriate number of *glue* atom-type pairs. When more than two *glue* atoms are present, each of them interacts as a carbon with the usual atom types and the nonidentical *glue* atoms in the molecule, avoiding spurious ring closures; interactions between the same type *glue* atoms are described with the 12−2 pseudo-

Table 4. The Docking Parameters Modified to Include the Two Different *glue* Atoms Defined in the Molecule **19**[a]

Atom Types	
types COHGJ	# atom type names[a]
Map List	
map protein.C.map	# atom-specific affinity map C
map protein.O.map	# atom-specific affinity map O
map protein.H.map	# atom-specific affinity map H
map protein.C.map	# atom-specific affinity map G[a]
map protein.C.map	# atom-specific affinity map J[a]
map protein.e.map	# electrostatics map
Intramolecular Interactions	
intnbp_r_eps 4.00 0.0222750 12 6	# C−C lj
intnbp_r_eps 3.60 0.0257202 12 6	# C−O lj
intnbp_r_eps 3.00 0.0081378 12 6	# C−H lj
intnbp_r_eps 4.00 0.0222750 12 6	# C−G lj[a]
intnbp_r_eps 4.00 0.0222750 12 6	# C−J lj[a]
intnbp_r_eps 3.20 0.0297000 12 6	# O−O lj
intnbp_r_eps 1.90 0.3280000 12 10	# O−H hb
intnbp_r_eps 3.60 0.0257202 12 6	# O−G lj[a]
intnbp_r_eps 3.60 0.0257202 12 6	# O−J lj[a]
intnbp_r_eps 2.00 0.0029700 12 6	# H−H lj
intnbp_r_eps 3.00 0.0081378 12 6	# H−G lj[a]
intnbp_r_eps 3.00 0.0081378 12 6	# H−J lj[a]
intnbp_r_eps 1.51 10.000000 12 2	# G−G plj[a]
intnbp_r_eps 4.00 0.0222750 12 6	# G−J lj[a]
intnbp_r_eps 1.51 10.000000 12 2	# J−J plj[a]

[a] Indicates G-atom and J-atom specific customizations.

Lennard-Jones potential. Some tests are performed to evaluate the implications of more *glue* atoms, the possible interactions between each other, and the effect of them on docking accuracy and ring closure. Moreover, the tests evaluate the possibility to manage rings smaller than seven members, like sugars.

Bicyclic System: Sucrose (5+6). The PDB entry 1SA4 contains sucrose (**19**) in complex with the human farnesyl transferase.[22] The location and conformation of the sugar within the binding site is comparable to that found in the 1LD7, 1LD8, and 1JCQ entries, showing a very stable and defined interaction. Compound **19** is expected to be an important validation tool for the computational protocol, since it includes two different noncondensed cycles of five and six members, the smallest rings considered in the study. In particular, the simulation of a five-member ring is not trivial because of the highly constrained structure.

Target Preprocessing. There are more than 20 water molecules within a 6 Å shell around the ligand, and most of them interact directly or indirectly with it, forming a big cage around the binding site. Nevertheless, for the sake of simplicity, none of them is included in the final target structure, similarly to heteroatoms (a zinc atom, a molecule of farnesyl diphosphate, and the R115777 inhibitor) that are all removed.

Ligand Preprocessing. Since all cycle carbon atoms of the molecule are chiral, there is not any choice for avoiding the problem of possible chiral inversion during ring closure. Breaking one bond for each ring (see Chart 1) leads to two pairs of glue atoms, labeled as G and J atoms for rings A and B, respectively. A total of 22 rotatable bonds are allowed to rotate.

Docking and GA Parametrization. Table 4 shows details of dpf lines that have been modified, according to the presence of *glue* atoms. Interactions between both *glue* atom types (G and J) and the protein are defined with the same C map. The intramolecular interactions table in the docking parameter file includes the pseudo-Lennard-Jones potential for G−G and J−J interactions, while the van der Waals potential is used in all other interaction pairs of G and J atoms. Therefore, G atoms interact with all other atoms in the molecule (including J atoms) as carbon atoms, except with the other G atom, for which the pseudo-Lennard-Jones potential is applied. In the same way, the J atoms act as carbon atoms with the remaining atoms (including G atoms), except with the other J atom. Such a description avoids spurious ring closures. Due to the high number of degrees of freedom resulting for the flexible linear molecule, the GA parameters are tuned iteratively with many docking tests to ensure a satisfying sampling of the conformational space. In the final GA parameter setting, each step of the evolutionary process involved a population of 400 individuals, for 1.5×10^7 evaluations and 10^7 generations. The probability of a local search is raised to 0.26. Finally, 510 docking runs are performed.

Docking Analysis. The presence of the two different *glue* atom types does not affect the ring closure: no spurious closures are found (i.e., between G−J or J−G atoms), confirming that the parametrization of multiple ring closures is reliable. The closure occurs with a very high frequency in all of the clusters, showing that the pseudo-Lennard-Jones potential is a useful tool also for treating small rings of five members. The software reproduces quite precisely the binding mode despite missing waters, but the cluster representing the X-ray binding mode is ranked as the third one by the energy score (see Table 2). Nevertheless, cluster population statistics provide useful information to identify the correct result. In fact, although the energy difference among the top three clusters is ∼0.2 kcal, the cluster containing the X-ray-like conformation is the most populated, with 102 individuals, versus 38 and 42 individuals of clusters 1 and 2, respectively. The conformation of five-membered ring A is found to be more easily reproducible, because it is very often found in the correct envelope conformation, with little difference with respect to the X-ray structure. On the other hand, the ring B hexose conformation presents some problems: most of the low energy results show a good chairlike conformation (the structure does not assume a classical chair conformation due to the partial planarization induced by the simulated G−G bond), but with the glycosidic bond rotated 180° with respect to the X-ray structure. As a consequence, the 6-hydroxy group points in the opposite direction, being in close contact with Ala_{230} and Arg_{231}, instead of being involved in the hydrogen-bond network with Gly_{282}, Gln_{285}, and Asp_{286}. Despite the good placement of rings, the incorrect orientation of ring B led to a 1.77 Å RMSD calculated on the X-ray structure of the sucrose. The first result showing correct alignment of both rings is found at the 13th subrank position. This strong deviation in the hexose ring alignment appears to be related to the lack of water molecules. Sugar hydroxyl oxygens, indeed, tend to occupy positions overlapping the coordinates of crystallized water oxygens 356, 384, 174, and 618. Moreover, many conformations of lower-ranked clusters are found to interact in the space region in between Asn_{276} and Arg_{231}, which in the X-ray structure accommodates well-solved water molecules. The chirality of centers close to the broken bond is

not always preseved but can be fixed with an additional refinement. The docked structure of **19** overlapped with the X-ray conformation is depicted in Figure 3.

Bicyclic System: Maltose (6+6). The next system investigated is the complex between maltose (glucopyranosyl-α(1-4)D-glucose, **20**) and maltodextrin binding protein (PDB ID: 1ANF). The binding site is located in the middle of the hinge region of the protein, and the ligand is reported to induce a twist movement of the two domains forming the maltodextrin receptor. Water molecules present in the crystal structure are noticeably fewer than those found in the binding site of **19**.

Target Preprocessing. All heteroatoms (including water) present in the crystal structure are removed, while partial charges and hydrogens are added by importing the structure in ADT.

Ligand Preprocessing. Ligand **20** is obtained by breaking the two bonds shown in Chart 1. In ADT, charges are assigned, nonpolar hydrogens are merged, and rotatable bonds are defined, activating a total of 22 torsions. The edge atoms are modified as G atoms for the A ring and J atoms for the B ring.

Docking and GA Parametrization. Due to the very high similarity between **20** and **19**, the docking parameter file is almost identical to the one applied in the previous case study (see Table 4).

Docking Analysis. While the ligands are similar, the binding cleft of 1ANF is smaller than that of 1SA4, and there are a few water molecules, simplifying both the docking and eventual clustering processes. Nevertheless, the docking itself is not an easy task because of the complexity of the linear form of the molecule and the high similarity of the two rings. The first cluster contains 165 results, all of them in a closed-ring state, and it is separated by the second one by only ~0.4 kcal/mol. Visual inspection of the results shows that alignment of the ligand in the binding site is inverted, in comparison to the X-ray binding mode, aligning the A ring onto the X-ray B ring, and vice versa. The orientation of the conformations of the first cluster can be explained considering the high similarity between the two alternative binding modes, and the conformations that can be adopted by the linear structure to mimic the correct binding mode with an inverted arrangement, when adopting a closed state. The statistical relevance of the second cluster allows identification of the correct binding mode of **20**. This cluster is the most populated, containing 516 individuals with rings in the closed form. Despite the high overall RMSD value of the lowest docked energy result with respect to the X-ray coordinates, it is clear that alignment of the molecule into the binding site reproduces quite well the experimental binding mode. Unlike **19**, the missing water molecules slightly affect the ring orientation because both rings place the hydroxyl methylen groups in such a way that essential interactions are reproduced. The docked ring A is a very reliable and correctly oriented chairlike conformation. The main difference between the X-ray and docked poses is the orientation of ring B that, although it is found in a chairlike conformation, however shows the glycosidic bond in an axial instead of the correct equatorial position. Within the entire cluster, this ring presents a RMSD higher than that for ring A, probably because it is more involved in solvent interactions (six water molecules in the crystal structure) than ring A (one water molecule only). The docked structure of **20** overlapped with the X-ray conformer is represented in Figure 3.

Multicyclic System: Vancomycin (12+16+16). The protocol is also tested on a complex between vancomycin **21** and the TDP-epi-vancosaminyl transferase (PDB ID: 1PNV).[23] The core of **21** is a multicyclic system (see Chart 1 for structure and breaking points) mainly composed by two 16-membered rings and one 12-membered ring.

Target Preprocessing. The target structure is obtained by removing all heteroatoms including water molecules, because they are not involved in crucial interactions with the ligand. The grid is centered approximately on the center of mass of the ligand, and the box size is set to entirely involve the huge binding site. The grids are calculated for six atom types (CANOHc) at a default resolution (0.375 Å).

Ligand Preprocessing. Considering the presence of three bonds to be broken, parameters for G, J, and Q glue atom types are introduced in the docking parameter file, corresponding to the edge carbon atoms of rings A, B, and C, respectively (see Chart 1). Due to the very high number of possible rotatable bonds present in **21**, all amide bonds are considered in their trans configuration, resulting in a total number of 46 rotatable bonds.

Docking and GA Parametrization. The intramolecular interaction panel for **21** includes a total of nine atom types (CANOHcGJQ). GA parameters are modified according to the high system complexity to be explored, resulting in the tightest parameters used in this study: 450 as the number of individuals in the population, 5×10^6 as the number of evaluations, and 10^6 generations are simulated. The probability of a local search on an individual is 0.16.

Docking Analysis. Vancomycin represents the most complex structure analyzed in this study, characterized for example by the number of rotatable bonds largely exceeding the values for which the AutoDock scoring function has been designed and tested.[1,11] Moreover, the entropy loss contribution leads to high positive values of the estimated free energy of binding in all results, while the high number of heavy atoms present in the molecule is likely to be the cause of a rise in the internal energy of the ligand to positive values. While the energy ranking is unreliable because of the aforementioned characteristics, the clustering analysis could provide a few hints about the binding mode. Due to the very flexible structure of the ligand, finding a satisfying convergence of results, even if collecting 510 runs, is a difficult task. Increasing the cluster tolerance up to 3 Å still provides 37 clusters containing more than a single conformation. Among them, only two clusters have five conformations, and they are very poorly ranked (at the 26th and 43rd positions). By visual inspection, the latter is rejected because all of its conformations lie over the binding cavity walls in a completely extended conformation. The former contains the conformations closest to the X-ray structure (3.11 Å RMSD, see Figure 3). The ring closure did not occur at all, but the effect of the pseudo-Lennard-Jones potential effectively drives the *glue* atoms to get closer each other, even if they reach a distance higher than a bond distance. The main reason of this result could be the insufficient accuracy of the calculation performed, because the torsions to be modified to obtain the ring closure are probably too numerous to be sampled with the GA parameters used in this simulation.

On the other hand, such results encourage the use of both the protocol and the software to perform docking calculations even in very complex systems.

CONCLUSION

This work describes a computational protocol that allows the performance of docking calculations with AutoDock on cyclic structures, overcoming one of the major limitations of the software. Following this protocol, the molecule is represented as the corresponding acyclic form by breaking one or more bonds, and allowing all the bonds that originally formed the ring to be treated as completely flexible. The ring closure is obtained with a combination of a custom pseudo-Lennard-Jones potential empirically defined and a dummy atom type (the *glue* atom G). The entire customization process is performed with simple modifications of the docking parameter file and the ligand structure, and it does not require modification of the software code. The protocol is validated by docking ligands found in 21 complexes available in the Protein Data Bank, exploring cyclic molecules represented by 5- to 29-membered rings. Despite the very simple implementation of the new energy potential, the overall results returned high and unexpected accuracy. For 17 of 21 complexes, it is possible to find (by energy or cluster population analysis) the correct binding mode of the ligand into the binding site. For 15 of them, the RMSD with respect to the X-ray structure is less than 2 Å, while for three of them, it is lower than 1 Å. Moreover, the protocol has been extended to study polycyclic systems by the inclusion of more than one *glue* into calculations. The combined use of several *glue* atoms in the same molecule does not affect the ring closure efficiency, opening a way to simulate complex systems, like polysaccharides. The current protocol is expected to be a very useful tool in finding the binding mode of cyclic ligands when little or no detail is known about the interactions between the ligand and its binding site (i.e., no X-ray structure or other experimental data are available), and it can be considered as a reliable alternative to protocols to dock rigid cycles obtained from molecular dynamics sampling or conformational analysis.[7,8] Another application of using *glue* atoms is the possibility to create a continuous protein−ligand distance constraint during docking, obtaining in principle a kind of constrained docking. The attractive effect between *glue* atoms, indeed, acts during the entire GA process, and it could also be used as a constraint to force a ligand (or a portion of it) into a specific region of the target. In this case, one of the *glue* atom pairs will replace the most convenient atom of the target protein surface. This could be useful in performing a restrained docking when peculiar interactions between the ligand and regions of the target protein are known. Obviously, all of these calculations will require a careful approach in results analysis and an opportune refinement protocol to manage resulting geometries, bumps between the molecule and protein, and spurious interactions. Finally, results reported here confirm AutoDock as a very reliable and versatile piece of software, with an extreme flexibility and a wide range of possible customizations.

MATERIALS AND METHODS

Molecular representations are generated with PyMOL.[29]

Protein Data Bank Search and Filtering. The Brookhaven Protein Data Bank is screened from its Web page: Search > Search Database > Ligands > Ligands and Prosthetic Groups (via the ChemAxon PDB interface[24]), searching for entries complexed with cyclic molecules with at least a C−C bond, in the range between 7- and 30-member rings (the maximum allowed by the interface). The query is performed using a "$C_{(rn)}-C_{(rn)}$" search string, where "*rn*" represents the smallest ring size, with "*n*" ranging from 7 to 30, with a "Substructure" criterion. All systems containing vis-à-vis complexes, cofactors, or non-water solvents into the binding site and covalently bound ligands are rejected.

Preparation of Ligand. Ligand structures are obtained from X-ray coordinates then transformed in linear form and modified by random rotations, translations, and torsion modifications with Maestro version 7.5[25] and minimized with MacroModel[25] using the following parameters: an MM3 force field with no solvent model, a constant dielectric (1.0), and charges from the force field. The minimization protocol is composed by 250 cycles with an SD minimizer followed by 1000 cycles with PRCG. When ligands bear rings other than the macrocycle (i.e., glycosyl groups), they are kept fixed, as in the conformation found in the X-ray structure. AutoDock Tools[10] is used to set up both ligand and target structure parameters. Ligands are imported, automatically merging nonpolar hydrogens, assigning atom types and Gasteiger−Marsili charges,[26] and defining rotatable bonds. As a general rule, all rotatable bonds are allowed to rotate, excluding amide bonds. Once the .pdbq file is generated, corresponding atoms are manually transformed into G atoms, by simply modifying the atom-type letter in third column of the .pdbq file. Partial charges of phosphorus atoms, when present in the ligand, are modified according to Chen et al.[27]

Target Molecule Processing. Target structures are obtained from the original PDB entry file, by removing heteroatoms, ions, solvent molecules, ligands, and waters, except whenever explicitly defined (see Case Studies). When residues' alternative conformations are available, only the first is considered, and others are ignored. The entire protein setup is made by using ADT, adding nonpolar hydrogens, AMBER charges,[28] and solvation parameters. For phosphorus charges on the target (1ZZ5), guidelines from Chen et al.[27] are followed.

Water Molecules Case Study ($1BGQ_{wat}$ and $1ZZ5_{wat}$). When explicitly reported, some water molecules from the PDB structure are retained, including them in the target structure. Hydrogens are added from the Maestro building interface and are manually oriented: compound **6**, water 9; compound **8**, waters 401 and 403.

Grid Parameters. All grids are manually centered on the approximate binding site center and sized to include all residues composing the binding site. All maps are calculated with a default resolution (0.375 Å) for usual atom types only. Glue atom-type maps are obtained by a symbolic link to or creating a reference to the .dpf file pointing at the C atom maps.

Docking Protocol and .dpf Customization. The *glue* atoms are added to the atom-type list, and the intramolecular interaction table is modified to include all additional interactions related to them. Carbon atom nonbond parameters r_{eqm} and ϵ with the 12−6 Lennard Jones potential are used to describe pairwise interactions between *glue* atoms and all other atom types present in the molecule. The *glue*−*glue* (i.e., G−G) pairwise interactions are characterized by the

12−2 pseudo-Lennard Jones potential, with an r_{eqm} value of 1.5 and an ϵ of 10. When more *glue* atoms are present (see Case Studies), each one of them is described with carbon atom parameters in interactions with the usual atom types and nonidentical dummy atoms in the molecule, while *glue* atoms of the same type (i.e., G−G, J−J, and Q−Q) are described according to the 12−2 pseudo-Lennard Jones and modified r_{eqm} and ϵ values. Genetic algorithm parameters are varied according to the system complexity: the lowest values start from a population size of 50 individuals, 25×10^4 energy evaluations, and 27×10^3 generations, while more in-depth calculations required values of 450 individuals, 15×10^5 energy evaluations, and 15×10^5 generations. Pseudo Solis and Wets algorithm parameters are not changed from default values, except for the local search probability, modified to 0.26 in some calculations. To collect a large number of conformations, calculations are spliced among multiple docking jobs. A single docking job is defined with a maximum of 510 runs. Docking is performed with a recompiled AutoDock version, obtained by editing and recompiling the source code to overcome the standard binary limitations: the maximum number of atom types is 16; the maximum number of torsionals is 64; the maximum ga_runs value is 2048. The docking parameter files used here are available upon request.

Evaluation of Docking Results. Multiple docking job results are merged and clustered. The cluster tolerance RMSD of each docked complex is calibrated iteratively by visual inspection, to the lowest value that can partition similar ring alignments into the binding sites in the same cluster. The RMSD from the X-ray structure is calculated for heavy atoms only, using AutoDock with an opportune docking parameter file.[9] RMSD calculations are performed using the modified AutoDock binary version. All calculations are performed on an IBM Blade Cluster equipped with six nodes mounting a dual Intel Xeon CPU 3.06 GHz processor running Red Hat Enterprise Linux.

ACKNOWLEDGMENT

S.F. would like to thank Dr. Fabrizio Manetti for many useful discussions and precious scientific suggestions.

REFERENCES AND NOTES

(1) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(2) Further information at the AutoDock Web site: http://www.scripps.edu/mb/olson/doc/autodock/ (accessed June 2007).

(3) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−Ligand Docking: Current Status and Future Challenges. *Proteins* **2006**, *65*, 5−26.

(4) Buzko, O. V.; Bishop, A. C.; Shokat, K. M. Modified AutoDock for Accurate Docking of Protein Kinase Inhibitors. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 113−127.

(5) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.

(6) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of Aminoglycosydes to Hydrated and Flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023−1033.

(7) Mulakala, C.; Nerinckx, W.; Reilly, P. J. Docking Studies on Glycoside Hydrolase Family 47 Endoplasmic Reticulum *alpha*-(1→2) Mannosidase I to Elucidate the Pathway to the Substrate Transition State. *Carbohydr. Res.* **2006**, *341*, 2233−2245.

(8) Fang, L.; Battisti, F.; Cheng, H.; Reigan, P.; Xin Y.; Shen, J.; Ross, D.; Chan, K. K.; Martin, E. W.; Wang, P. G.; Sun, D. Enzyme Specific Activation of Benzoquinone Ansamycin Prodrugs Using HuCC49ΔCH2-*β*-Galactosidase Conjugates. *J. Med. Chem.* **2006**, *49*, 6290−6297.

(9) Morris, G.; Goodsell, D. S.; Huey, R.; Hart, W. E.; Halliday, S.; Belew, R.; Olson, A. J. *AutoDock 3.0.5 User Guide*.

(10) AutoDock Tools. http://autodock.scripps.edu/resources/adt (accessed June 2007).

(11) Heteni, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2002**, *11*, 1729−1737.

(12) Sotriffer, C. A.; Flader, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varga, J. M. Automated Docking of Ligands to Antibodies: Methods and Applications. *Methods* **2000**, *20*, 280−291.

(13) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(14) Park, H.; Lee, J.; Lee, S. Critical Assessment of the Automated AutoDock as a New Docking Tool for Virtual Screening. *Proteins* **2006**, *65*, 549−554.

(15) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 3. The Free Energy Contribution of Structural Water Molecules in HIV-1 Protease Complexes. *J. Med. Chem.* **2004**, *47*, 4507−4516.

(16) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Molecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(17) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability to Form more than Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 148−156.

(18) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding Mode Prediction of Cytochrome p450 and Thymidine Kinase Protein−Ligand Complexes by Consideration of Water and Rescoring in Automated Docking. *J. Med. Chem.* **2005**, *48*, 2308−2318.

(19) von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Michael Ryan, D.; Woods, J. M.; Bethell, R. J.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational Design of Potent Sialidase-Based Inhibitors of Influenza Virus Replication. *Nature* **1993**, *363*, 418−423.

(20) Roe, S. M.; Prodromou, C.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H. Structural Basis for Inhibition of the Hsp90 Molecular Chaperone by the Antitumor Antibiotics Radicicol and Geldanamycin. *J. Med. Chem.* **1999**, *42*, 260−266.

(21) Deterin, C.; Varani, G. Validation of Automated Docking Programs for Docking and Database Screening against RNA Drug Targets. *J. Med Chem.* **2004**, *47*, 4188−4201.

(22) Reid, T. S.; Beese, L. S. Crystal Structures of the Anticancer Clinical Candidates R115777 (Tipifarnib) and BMS-214662 Complexed with Protein Farnesyltransferase Suggest a Mechanism of FTI Selectivity. *Biochemistry* **2004**, *43*, 6877−6884.

(23) Mulichak, A. M.; Losey, H. C.; Lu, W.; Wawrzak, Z.; Walsh, C. T.; Garavito, R. M. Structure of the TDP-epi-vancosaminyltransferase GtfA from the Chloroeremomycin Biosynthetic Pathway. *Proc. Natl. Acad. Sci U.S.A.* **2003**, *100*, 9238−9243.

(24) ChemAxon Kft., Budapest, Hungary. http://chemaxon.com (accessed June 2007).

(25) *MacroModel/Maestro*; Schrodinger, LLC: Portland, OR. http://www.schrodinger.com/ (June 2007).

(26) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Thetrahedron* **1980**, *36*, 3219−3228.

(27) Chen, Y.; Kortemme, T.; Robertson, T.; Baker, D.; Varani, G. A New Hydrogen-Bonding Potential for the Design of Protein-RNA Interactions Predicts Specific Contacts and Discriminates Decoys. *Nucleic Acids Res.* **2004**, *32*, 5147−5162.

(28) Weiner, S.; Kollmann, P.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−770.

(29) PyMOL. http://pymol.sourceforge.net/ (accessed June 2007).

(30) Teramoto, R.; Fukunishi, H. Supervised Consensus Scoring for Docking and Virtual Screening. *J. Chem. Inf. Model.* **2007**, *47*, 526−534.

CI700036J