Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 319–333
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Exploring the conformational space of cyclic peptides by a stochastic search method☆

Anwar Rayan [a], Hanoch Senderowitz [b], Amiram Goldblum [a,*]

[a] *Department of Medicinal Chemistry and Natural Products, David R. Bloom Center for Pharmacy, School of Pharmacy, The Hebrew University of Jerusalem, Jerusalem 91120, Israel*
[b] *Peptor Ltd., Kiryat Weizmann, Rehovot 78326, Israel*

Abstract

A stochastic search algorithm is applied in order to probe the conformations of cyclic peptides. The search is conducted in two stages. In the first stage, random conformations are generated and evaluated by a penalty function for ring closure ability, following a stepwise construction of each amino acid into the peptide by a random choice of one of its allowed conformations. The allowed conformational ranges of backbone dihedral angles for each amino acid have been extracted from a Data Bank of diverse proteins. Values of dihedral angles that do not contribute favorably to the scoring of ring closure are retained or discarded by a statistical test. Values are discarded up to a point from which all remaining combinations of angles are constructed, scored, sorted, and clustered. In the second stage, side chains have been added and fast optimization was applied to the set of diverse conformations in a "united atoms" approach, with the "Kollman forcefield" of Sybyl 6.8. This iterative stochastic elimination algorithm finds the global minimum and most of the best results, when compared to a full exhaustive search in appropriately sized problems. In larger problems, we compare the results to experimental structures. The root mean square deviation (RMSD) of our best results compared to crystal structures of cyclic peptides with sizes from 4 to 15 amino acids are mostly below 1.0 Å up to 8 mers and under 2.0 Å for larger cyclic peptides.

Keywords: Stochastic search algorithm; Multidimensional space; Complex combinatorial; Cyclic peptides; Structure prediction; Conformations; Flexibility; Ensembles

1. Introduction

Searching the multidimensional space of peptides and proteins in order to detect the best structures is a difficult task due to current limitations of computational approaches. In such a search, we hope to find the global minimum and the ensemble of lowest energy structures, which is mainly composed of local energy minima or energies that are close to the global minimum. While in reality these molecules achieve their equilibrium conformational ensembles in short time spans [1–5], two main obstacles hinder the prediction in silico of their three-dimensional structures. First, a search for the lowest energy conformers on the potential energy surface is considered to be a non-polynomial (NP)-hard problem for proteins [6] and peptides should not be different. Thus, the search space increases exponentially with the number of amino acids in the system. Second, a reliable and highly accurate scoring ("potential function") is required, one that should ideally distinguish between native structures and any others. In the case of known crystal structures, those have been traditionally assumed to represent the native states or at least to be close to them. This is more complicated if nuclear magnetic resonance (NMR) results are considered, because the NMR spectrum represents a Boltzmann averaged set of conformations. In the case of larger ensembles, this task is impossible: most of the accessible conformations are not known from experiment, and therefore one must find them computationally by correct physical descriptions [7–9] rather than by employing statistical potentials [10–12]. Searching space with the hope to obtain meaningful ensembles requires then to employ both a strong search method as well as a trustworthy force field. Both are under current development in many groups. This problem is even more complicated in peptides, due to the scarcity of experimental structures that should be used to calibrate both the search method and the force field. Quite a few methods were however found to be successful for cyclic and linear peptides

☆ Supplementary data associated with this article can be found at doi:10.1016/j.jmgm.2003.12.012.

* Corresponding author.
E-mail addresses: anvarr@md.huji.ac.il (A. Rayan), hanoch_s@netvision.net.il (H. Senderowitz), amiram@vms.huji.ac.il (A. Goldblum).

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2003.12.012

of relatively smaller sizes [13–20] even though they cannot guarantee the finding of a global minimum or a set of "best populations."

In a recent paper [21], we presented a novel search technique for complex combinatorial problems, and in the present paper we suggest an improvement of this technique. We demonstrate the ability of this improved search to locate ensembles of low energy conformations of cyclic peptides by casting the search over their complex, high dimensional conformational spaces into a combinatorial form. Reproducing the experimental structures is however still limited by the reliability of the scoring functions, and the complicated task of choosing and adapting a force field to this particular problem is deferred to a later stage. Here, we present the search methodology, as it applies to cyclic peptides and compare it to a full, "exhaustive" search with the same, though deficient, scoring function as employed for our stochastic search.

The quest for substituting biological peptides and proteins with cyclic peptides has been ongoing for quite a few years [8,22–32]. The conformational search of cyclic peptides [5,33,34] is interesting because they are regarded to be "conformationally restricted" analogs of bioactive linear peptides [35–37]. "Cyclization" is accompanied by the hope to achieve higher stability [36,38,39], improved affinity and specificity [8,39,40], and greater bioavailability [41]. The lower flexibility of these cyclic molecules reduces the loss of vibrational entropy upon binding, with respect to linear peptides. Also, restricted conformations could better "expose" a pharmacophore to their biological targets [42]. Cyclic peptides have been shown to possess very diverse biological activities [8,31,32,43–46] and are less prone to cleavage by proteinases [31]. It is thus widely expected that cyclic peptides could become more potent and with less side effects and longer lifetimes, compared to linear peptides. Also, cyclic peptides serve as a basis for developing peptidomimetics.

There is much interest in predicting the conformational space of cyclic peptides, and some approaches have been successful in finding the global minimum or closely related structures [41,47–50]. Currently, the conformational search of cyclic peptides is performed with a variety of methods or protocols. Most of them sample a few structures out of the whole conformational space. The sampled conformations are energy minimized (EM) [9,34] and subsequently studied by a variety of techniques such as molecular dynamics (MD) [34,51], simulated annealing (SA) [51,52], and Monte Carlo (MC) simulations [9,53,54]. The efficiency of conformational space exploration with these methods is however still limited and highly dependent on the position of the initially selected conformations in multidimensional space.

2. Methods

2.1. General

Our approach to detect conformational ensembles of cyclic peptides is based upon our algorithm, which was previously applied to the predictions of multiple loops in comparative modeling [55], of side chain rotamers [21] and of polar protons [56] in proteins. Below, we describe a modified version of the original algorithm.

The basis for any application of the stochastic algorithm is in the definition of "variables" and "variable values" for the system to be studied. The variables can be any basic parameters or "building blocks" of the system that serve to characterize it, such as bond lengths, atom–atom distances, various angles, full amino acid units or even larger stretches of amino acid units [57]. Each such variable should be represented by a set of discrete values, which may be extracted from databases, from continuous functions, from restrictions imposed by the user, or from other sources. There is no theoretical limit on the number of variables or on the number of values for each variable. However, our experience is that a smaller number of values for each variable should be preferred, while the number of the variables may be extended. A reliable "cost function" should be used for evaluating a full configuration of the system, and such a configuration is constructed by a random choice of variable values.

In our construction of cyclic peptide conformations, we used standard bond lengths and bond angles that were extracted from AMBER 4.1 [58]. Values for each of the backbone angles φ and ψ were determined from our analysis of a representative database of proteins (vide infra), and the peptide bond dihedral angle, ω, was allowed to assume a single value of 180.0° ("*trans*" conformation) for all amino acids except for proline, where it could be either180 or 0°.

2.2. Determining variable values of φ, ψ angles

A database of 1001 proteins (see supplementary material, Table 1a) was chosen out of the Protein Data Bank [59]. The requirement for inclusion in this database was a resolution $\leq$2.5 Å and a sequence identity of less than 25%. Distribution maps (see Fig. 1) were produced in steps of 12° and only those values with an occupancy above zero were further considered ("allowed variable values"). As can be seen in Fig. 1, the total number of such allowed values for each of the two angles φ and ψ depends on the type of amino acid. A smaller database of 111 proteins sharing a sequence identity of less than 17% was analyzed and gave similar results with the same "allowed values" for φ and ψ (Fig. 1). The values for Pro and Gly represent these special amino acids, which appear less in secondary structures, while Asp is a representative of most other residues that are more abundant in secondary structures.

The search is performed in two consecutive steps. First, a stochastic stage aspires to decrease, as rapidly as possible, the set of combinations from its huge initial number to a manageable number of combinations. This is achieved by assigning to each variable a random value out of its list of allowed values. A conformation is determined once

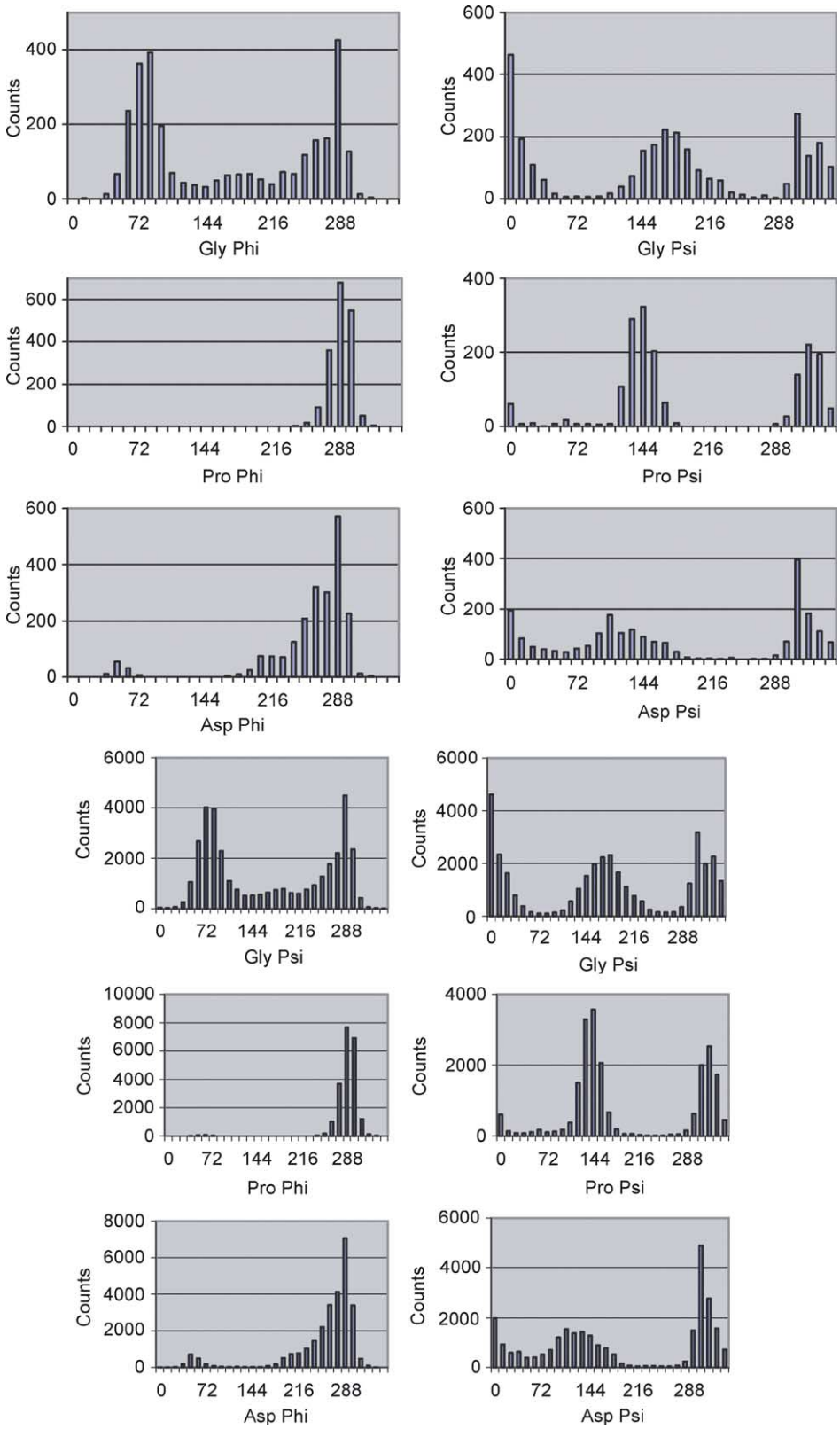

Fig. 1. Distribution maps of three residue types (Gly, Pro, and Asp). *X*-axis values are angles, computed in bins of 12°. *Y*-axis values are numbers of appearances in the database. ψ and φ have the regular meaning of rotations around the N–Cα and Cα–C bonds, respectively. Angles related to α-helix and β-sheet structures appear more frequently: ψ around 320° and φ around 300° are related to α-helix while ψ around 120° and φ around 240° are related to β-sheet domains. Ordered structures (α-helix and β-sheet domains) contain less Gly and Pro. This may be seen by comparing the histograms of Asp to those of Pro and Gly. The upper three rows are for the large database (1001 proteins) and the lower 3 rows are for the smaller database (111 proteins).

all values have been randomly determined. The conformation is scored, and the process is repeated many times to produce a sample. Variable values in this sample are statistically tested and, on the basis of such tests, they are retained or evicted. A new sample is generated and the entire procedure is iteratively repeated, each time evicting more variable values. Once the number of variables' combinations becomes manageable, an exhaustive computation stage commences, in which all the remaining variable value combinations are surveyed. In the stochastic stage, we increase the probability of keeping the variable values that contribute to the best solutions. If this is achieved, we expect to obtain a representative set of best solutions in the exhaustive stage, such that would enable to construct an "ensemble" of accessible conformations [60].

2.3. Stochastic construction of a single conformation

Each cyclic peptide has been constructed by simultaneously adding residues to both the N-terminal and C-terminal of a single "stem" amino acid. The choice of this stem residue has been arbitrarily decided to be the first residue of the linear sequence. This choice was tested for GLGGLG, by picking other residues for the stem. We obtained similar root mean square deviation (RMSD) values with these different choices. If the number of residues is even, the last residue is grown only from the N-terminal side. The addition ends by evaluating the ability to form a peptide bond at the "middle" or next to the "middle" of the cyclic structure, depending on the overall number of residues. Each additional residue is characterized by standard bond lengths and bond angles, while the dihedral angles are randomly picked out of the allowed values for each amino acid. The size of the φ and ψ set of values is reduced along the process, as conformations that do not contribute to proper closure of the cycle are discarded.

We sample n conformations X_i to begin the statistical tests. Each conformation X_i is constructed by a random choice of the dihedral angles of its m residues. We obtain $X_1 = (\varphi_{11}, \psi_{11}, \omega_{11}, \varphi_{12}, \psi_{12}, \omega_{12}, \ldots, \varphi_{1m}, \psi_{1m}, \omega_{1m}), \ldots;$ $X_n = (\varphi_{n1}, \psi_{n1}, \omega_{n1}, \varphi_{n2}, \psi_{n2}, \omega_{n2}, \ldots, \varphi_{nm}, \psi_{nm}, \omega_{nm})$, where φ_{11} is a randomly picked value of φ for the first amino acid in the first conformation, and ψ_{11} and ω_{11} are the values for ψ and ω of the same amino acid in the same conformation, etc. Thus, only backbone N, Cα, and C atoms are used for the construction of the skeleton. Once all the atoms of the backbone have been positioned for a conformation X_i, we compute its energy value E_i by applying the scoring function described below and associate the set of variable values with that energy. Thus, for the randomly picked full set of cyclic peptide conformations X_1–X_n we obtain a set of energies E_1–E_n. The total number of conformations at each iteration, in the current application, is $n = 10^5$. The energies and associated variable values are next analyzed as discussed below.

2.4. Elimination of values

We construct the distribution histogram F_E^n from all conformations. F_E^n is the set of energies of all the $n = 10^5$ sampled conformations for the cyclic peptide. We define cutoff points H and L in F_E^n. H contains the set of highest energy conformations while L contains a set of a similar size of the lowest energy conformations. In this study, the H and L sets include, each, 1000 structures.

Let us assume that all conformations of all variables in all amino acids are equally probable across F_E^n, i.e., there is no energy preference for any particular conformation. Under this assumption, I_{kp}, the number of times that conformation k of residue p is expected to appear in the H and L sets is given by $I_{kp} = 1000/N_{kp}$ where 1000 is the number of conformations in H and L, and N_{kp} is the number of conformations (values) for variable k of residue p. N_{kp} changes along the iterative process, and so is also I_{kp}. For example, if we have 10 values for the angle φ of residue m, its $I_{\varphi_m} = 100$ and in the H or L sets, such a value should thus appear about 100 times. This expectation value can then be used as a means to decide whether to eliminate or retain a value.

If the frequency of occurrence of a certain torsional value of a certain amino acid, say, φ_3 in L is lower than I_{φ_3} by a pre-defined amount, then this value is eliminated from the list of conformations for φ_3. That is because this particular conformation is found to have a low probability for contributing to the set of best (lowest scoring function value) conformations in L. The "pre-defined amount" (eviction factor, EVF) depends on the user and is usually set to between two and three. EVF values were optimized in an extensive test of a similar problem of loop closure, with the aim to seek a balance between computation time (which is much longer for numbers higher than two to three) and the conservation of the low energy population (which may be unjustifiably discarded with smaller numbers). In a similar manner, if the abundance of a value for φ_m in the H set is larger by a certain factor than its expectation value, this value is marked for removal and its abundance in L is tested. The value will be removed if its abundance in L is smaller than its expectation value by any amount. Thus, a value φ_m of a residue will be discarded only if it fulfills the demands either in L alone, or in both H and L. There are, therefore, EVF values for both the L region (EVF^L), and the H region (EVF^H).

The existence of such low (in L) or high (in H) abundance of conformations is tested for all remaining variable values (φ_i, ψ_i, and ω of proline) of each amino acid along the sequence at the end of each iteration based on a sample size of 10^5 conformations. Thus, several conformations of each residue may be discarded at each iteration. The number of conformations is thus reduced, and the total number of possible combinations for the whole system is becoming smaller. The above steps are repeated for the reduced set of values, until the number of combinations of the system becomes smaller than a pre-defined "threshold," 10^6, which marks the end of the stochastic stage. From that point on, a full

exhaustive search is conducted with all possible remaining combinations of residue conformations. The present version of the algorithm differs from the previous version mainly in its ability to evict more than a single value of each variable, in every iteration of the algorithm, due to the increased sample size at each iteration. These iterations lead from a large set of possible conformations at the beginning, to a much smaller set of solutions that are studied exhaustively and finally form the populated conformational ensemble of a molecule.

2.5. Exhaustive stage

Once there remain $M \leq 10^6$ combinations, an exhaustive search is performed and the resulting conformers are sorted based on their energies. Sorting may be done by defining the maximum number of lowest energy conformations required, or by an energy threshold, or by some other criterion.

2.6. Cost functions for constructing cyclic peptides

The prediction of cyclic peptide conformations advances in two stages. In the initial stage, a very large number of conformations is generated and evaluated. The main problem at this stage is closing the loop within the generated conformation and consequently the cost function focuses on this problem. Moreover, as this cost function is required to evaluate many structures, it is less elaborate than in the second stage, in which a much smaller set of conformations remains for full evaluation (with detailed atomic structures).

The cost function that is employed in the initial stage of the search was developed in order to allow for the construction of a large set of diverse "closed conformations" of cyclic peptides. The energy of each potential conformation is given by a penalty term for improper loop closure together with a partial contribution from VdW and H-bond energies:

$$E = E_{\text{penalty}} + (E_{\text{VdW}} + E_{\text{HB}})F \tag{1}$$

E_{penalty} evaluates the quality of the closure between the two parts constructed by extending both N- and C-termini according to four criteria: the deviation in peptide bond, the deviations in the two bond angles and the deviation in the (peptide) dihedral angle at the meeting point. The general type of function that was chosen for all four criteria in order to minimize deviations from standard parameters is

$$E_{\text{penalty}} = \left(\frac{X}{Y}\right)^5 - 1, \quad \text{for X} > \text{Y} \qquad \text{or}$$
$$E_{\text{penalty}} = \left(\frac{Y}{X}\right)^5 - 1, \quad \text{for Y} > \text{X} \tag{2}$$

where X is the current value of the variable and Y is its standard value.

For the peptide bond length, $Y = 1.32$ Å (the standard length of a peptide bond). For the two bond angles ($C\alpha^i$–C^i–N^{i+1} for the N-terminal residue i, C^i–N^{i+1}–$C\alpha^{i+1}$ for the C-terminal residue i+1), the standard values are 122.0 and 117.0, respectively. For the peptide dihedral angle ($C\alpha^i$–C^i–N^{i+1}–$C\alpha^{i+1}$), the criteria are somewhat more complex. Space was divided into four quarters, two for *trans* and two for *cis* peptide conformations. ω is the observed peptide dihedral angle and the standard dihedral angles are 180° for *trans* and 0° for *cis*.

For *trans*:

$$E_{\text{penalty}} = \left(\frac{180}{\omega}\right)^5 - 1, \qquad \text{if } 90 < \omega < 180$$
$$E_{\text{penalty}} = \left(\frac{180}{360-\omega}\right)^5 - 1, \quad \text{if } 180 < \omega < 270 \tag{3}$$

For *cis* (considered only for prolines):

$$E_{\text{penalty}} = \left(\frac{180}{180-\omega}\right)^5 - 1, \qquad \text{if} \quad 0 < \omega < 90$$
$$E_{\text{penalty}} = \left(\frac{180}{\omega-180}\right)^5 - 1, \quad \text{if} \quad 270 < \omega < 360 \tag{4}$$

The power of five was found to give better results than smaller numbers that were examined, while larger ones penalize deviations too much.

E_{VdW} is a "classical" pairwise non bonding energy function:

$$E_{\text{VdW}} = \varepsilon_{ij}\left(\left(\frac{V_{ij}^{12}}{R_{ij}^{12}}\right) - \left(\frac{2V_{ij}^{6}}{R_{ij}^{6}}\right)\right) \tag{5}$$

where ε_{ij} has been taken from the AMBER force field, V_{ij} the Van der Waals (VdW) parameter for atoms i and j in that force field, and R_{ij} is the distance from atom i to atom j. E_{HB} is composed of a VdW component and an electrostatic component (with $\varepsilon = r$):

$$E_{\text{HB}} = \left(\frac{A}{R_{ij}^{12}} - \frac{B}{R_{ij}^{10}}\right) + \frac{TQ_{\text{H}}Q_{\text{O}}}{R_{\text{HO}}^2} \tag{6}$$

A is the repulsion parameter for the two atoms H and O and equals 7557, while B is their attractive polarizability parameter and is set to 2385 [58]. Q_{X} is the partial charge of atom X (taken from the AMBER force field), R the distance in Å, and $T = 332.06$ is a conversion factor for transforming an energy expression in units of electron partial charges and distance in Å into kcal/mol.

The total energy is the sum of the components describing the quality of loop closure together with contributions from non-bonded interactions multiplied by factor (which has been set to 0.05 in this study).

2.7. Clustering the ensemble

The clustering of conformations was performed at the end of the exhaustive stage, in order to further reduce the number of conformations while keeping the maximal variance. Clustering was performed by using a root mean square deviation criterion for backbone atoms. All the conformations

whose structures are similar, within the given RMSD, to the global minimum are first discarded from the database of low energy conformations in the large ensemble of each peptide. The next remaining low energy conformation is selected as the representative of a second "family of conformations" and conformations that are close to it are discarded. This process is continued iteratively, until no more conformations are left in the set of each peptide. The RMSD values vary with the size of the peptide. We have used the following values for cutoff in each size of cyclic peptide:

$$\mathrm{RMSD_{cutoff}} = 0.4\,\text{Å}, \quad N \leq 6$$
$$\mathrm{RMSD_{cutoff}} = 0.7\,\text{Å}, \quad 7 \leq N \leq 10$$
$$\mathrm{RMSD_{cutoff}} = 1.2\,\text{Å}, \quad 10 \leq N \leq 12$$
$$\mathrm{RMSD_{cutoff}} = 1.5\,\text{Å}, \quad 13 \leq N \leq 15$$

After the clustering, a large number of conformations, depending on the length and composition of the cyclic peptide, was retained and processed further by adding the side chains and minimizing the full structure.

2.8. Construction of the full structure

In the next step, the Sybyl 6.8 package [61] was employed. Side chains and polar hydrogens were added to the lowest energy member from each cluster of each peptide, which was subsequently minimized using the Kollman forcefield with united atoms. A distance dependent dielectric was chosen and set to $\varepsilon = r$ (as in Eq. (6) for the previous stage). The steepest descent minimization was limited to 100 iterations for each of the peptides of length up to 10 residues, and 200 iterations for the longer ones. These artificial limits were helpful in reducing the overall time for obtaining improved structures with low RMSD values in comparison to experiments.

3. Results

The program that we developed for exploring the conformations of cyclic peptides (CPCS, cyclic peptides' conformational search) is capable of scanning the conformational space of cyclic peptides with 4–20 residues. It detects the *n* best conformations of the backbone plus Cβ (except for glycine). The program can deal with L and D conformations of amino acids and also tests options of *cis*/*trans* peptide bonds for any proline in the sequence.

Two main issues were examined for this project: the reliability and efficiency of the search method, and its ability to predict real structures. The reliability was mainly tested by comparing a full, exhaustive search to a stochastic search, for the same cyclic peptides. The ability to reproduce real structures was tested with respect to X-ray results for cyclic peptides of widely varying sizes.

3.1. Comparison of stochastic and exhaustive results

The new stochastic search technique was first applied to detect the lowest energy conformations in the conformational space of a cyclic tetrapeptide $(Gly)_4$ and the results were compared to a "full" exhaustive search of all possible combinations for that cyclic peptide. The set of accessible conformations for each of the backbone dihedral angles was similar in both searches.

A conformation of a cyclic peptide with four amino acids is determined here by nine dihedral angle variables (three for each amino acid—φ, ψ, and ω). No variables are required for the stem residue which is common to all conformations. The peptide ω angle Cα–C–N–Cα was fixed at 180° for glycines, and thus only six torsions were varied. Analysis of the distribution maps from the protein database (see Section 2) revealed that each of the φ and ψ angles in glycine may assume 30 allowed values. The full number of combinations in this small system is thus $30^6 = 7.29 \times 10^8$ conformations, consisting of all the combinations of allowed dihedral angles for the three residues. A full exhaustive search required 3.5 days on a Silicon Graphics Octane R12000.

The stochastic search method as outlined above has been applied to the same molecule. The 1000 low energy structures obtained by both methods were then compared. Parameters of the stochastic experiments were varied in order to find the best conditions, i.e., those that achieve the maximum overlap between the stochastic search and the full exhaustive search. Results of such comparisons are presented in Table 1. We varied the size of the sample between 10^5 and 10^6 conformations, the EVF values for the L and H regions, and the threshold (that determines when the search is switched from stochastic to exhaustive). The parameters determine the time (roughly proportional to the number of search iterations in Table 1) and the percent of best structures found in comparison to the full exhaustive search. Run no. 10 in Table 1 required less than 27 min, out of which the two stochastic iterations took less than 12 min. Threshold values, indicating the switching point between the stochastic and exhaustive searches are calculated by multiplying all the remaining numbers of variable values. Since the program switches to an exhaustive search once it finds a number of combinations that is lower than the requested "threshold" (which cannot be met exactly during the run), the table contains different numbers for the thresholds in different runs, even though similar thresholds were requested in some of them.

Some of the results in Table 1 display a lower success in predicting the best 1000 structures. In order to be able to predict all the best structures that were found in the exhaustive search, it is necessary to determine which values of each of the variables appear among the best 1000 solutions from the exhaustive search and then multiply their numbers in order to obtain the minimal number of combinations that are essential for predicting this same (exact "best" results) set by any other search method. For the cyclic $(Gly)_4$ system, this number is 1,368,900. Therefore, any stochastic search

Table 1
Conditions for stochastic search of the best conformations of $(Gly)_4$

Run no.	Sample size in thousands	EVFL	EVFH	Threshold	Number of iterations	Percentage of best 1000 structures
1	999	3		4681638	2	100
2	199	3		4801680	4	99.5
3	99	3		4715172	7	98.5
4	199	3	4	6036849	3	100
5	199	3	4	1245090	4	97.2
6	199	3	4	291600	5	81.8
7	99	2	3	4169880	3	99
8	99	2	3	864360	4	90.8
9	999	2	3	1156155	2	97.2
10	199	2	2	4492800	2	98.5
11	999	2	2	635040	2	88
12	999	2	2	20600000	1	100

that evicts variables until their total number of combinations falls below that value, can predict only part of the 1000 best structures. Fig. 2 depicts the variation in the percentage of the best structures that are found for the cyclic tetrapeptide as a function of the minimal number of combinations, which is the threshold for switching from the stochastic to the exhaustive stage. If the stochastic search continues to evict variables without switching to an exhaustive search, many of the best conformations might be lost. There is thus a minimal number of combinations that must be explored exhaustively in order to retain a certain number of best conformations.

Fig. 3 displays the results for run no. 10 in Table 1, in which 98.5% of the exhaustive results were reproduced by the stochastic search. Three lines in the figure present, from upper to lower, the stochastic results, the exhaustive (exact) results and the difference between the two in kcal/mol. That difference is zero for the first, lowest energy, 384 conformations. To find how many of the 1000 conformations of the exhaustive stage were also obtained by the stochastic search, or the "percentage of reproduction," we search for the energy of conformation number 1000 of the exhaustive search—among the stochastic search results. Assume that it has been found in position 985 of the stochastic search. Any conformation below that position in the stochastic search will also be below the 1000th conformation of the exhaustive search (failing to comply with this statement means that the full search did not get to all the combinations). Therefore, all conformations below that position of the stochastic search exist also among the 1000 conformations of the exhaustive search, and we can safely assign the number 98.5% as the reproduction level of the exhaustive search by the stochastic one.

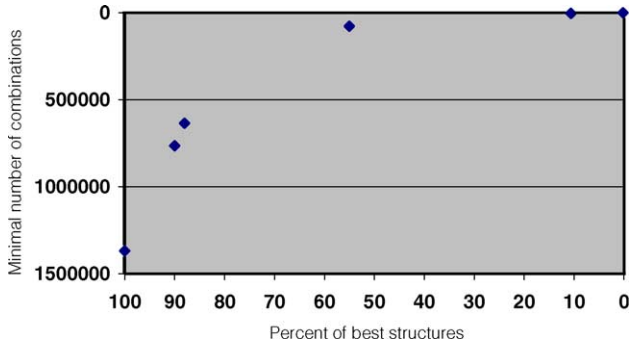

Fig. 2. The relation between the minimal number of required combinations and the prediction of best structures determines the need for switching from stochastic search to exhaustive (see text).

The crude cost function at the stochastic stage of computations is intended to improve the diversity of the resulting conformations, and therefore we cannot attach too much importance to the differences in energy between conformations at this stage. However, the difference between the two methods, at the "top" of the 1000 conformations, is only about 0.2 kcal/mol out of a total of ∼200 kcal/mol, amounting to an error of less than 0.1% in the energies. Energies between the stochastic and exhaustive computations were compared up to the fifth decimal digit. It is thus highly probable that conformations having similar energies of such accuracy are also similar in geometry.

3.2. *Efficiency of the search method*

The application of the stochastic search shortens the time for finding a set of the best solutions. In order to study the effect of increasing the number of variables on the length of the computation, we compared the number of iterations that were required to reach the exhaustive stage in cyclic peptides of increasing size. The number of iterations increases with the number of variables, in a linear rather than exponential fashion. This is displayed in Fig. 4 for cyclic peptides having from four to eight residues. The number of iterations ranged between 2 iterations for a tetrapeptide to 22 iterations for an octapeptide (similar criteria for evicting variable values were applied to all peptides). The ratio between the initial combinations for these two systems is 6.5×10^{11} (7.29×10^{8} for the tetrapeptide versus 4.78×10^{20} for the octapeptide) whereas the ratio between the number of iterations is only 11.

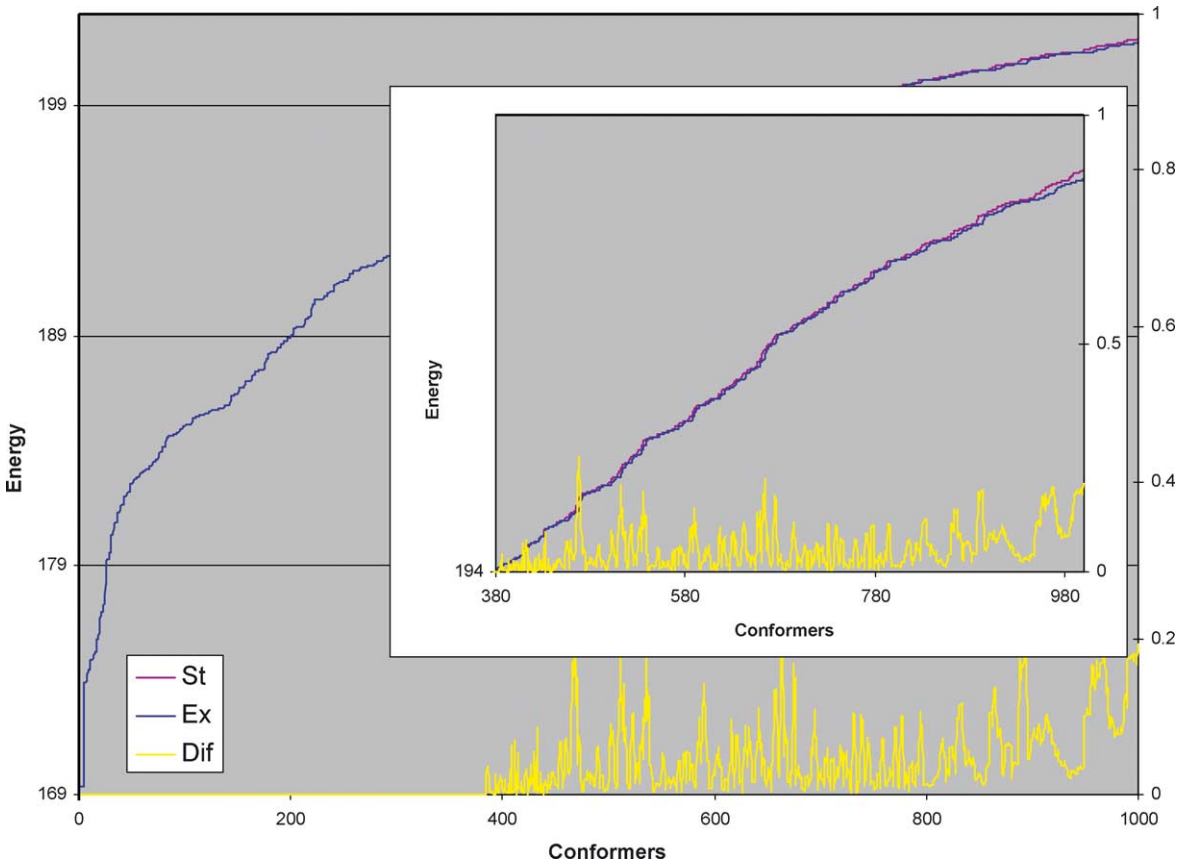

Fig. 3. Results of the stochastic search (upper curve, in the inset), exhaustive search and the difference between them (lowest line), in kcal/mol. *X*-axis values are the conformations numbers, up to 1000. *Y*-axis values are in kcal/mol. The alternative *Y*-axis on the right is the difference between stochastic and exhaustive results, in kcal/mol.

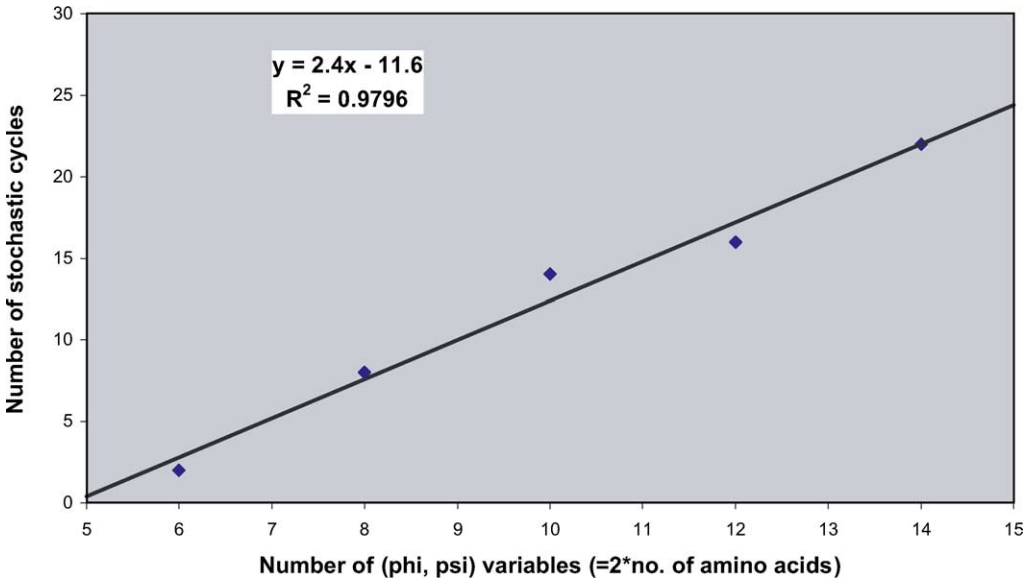

Fig. 4. Number of iterations of stochastic search (*Y*-axis) as a function of the overall number of φ and ψ variables in searching for backbone conformations of cyclic peptides.

The size of the sample is another important variable that was examined. Employing the tetrapeptide prediction, we varied the size of the samples from 10,000 to nearly 1×10^6 (see Fig. 5). The H and L test regions were left with 1000 conformations each. The lines drawn in this figure display the number of new conformations that appear among the set of 1000, as a function of changing the sample size. Ideally, we would like to achieve stability and a minimal number of changes in conformations. We find that such stability is achieved in both the L ("best") and H ("worse") regions only when sample sizes are above 10^5. With smaller samples, a large variability has been detected for both the "best" and "worst" sets (H and L).

3.3. Comparing predictions and crystal structures of cyclic peptides

In this section, we present comparisons with experimental results for 24 molecules, out of which 7 pentapeptides were studied by Nikiforovich et al. [48] and all others were selected from a database of crystal structures.

Crystal structures were selected from the Cambridge database [62]: we limited ourselves to structures of cyclic peptides that contain various sequences of L- and D-amino acids and are all "head to tail" structures. Their length is between 5 and 15 amino acids. Each of these cyclic peptides was reconstructed by our algorithm, as described in Section 2, and a set of best results was collected.

Table 2 presents the sequences and the root mean square deviation of the backbone atoms (N, Ca, C, and O) for the predicted structure of each cyclic peptide that was found to have the lowest RMSD. In addition, the rank of this structure in the whole set of conformations, following the full energy minimization (within Sybyl 6.8), and its energy above the global minimum for that cyclic peptide are tabulated. The results have been grouped by the size of the cyclic peptides.

All the peptides up to seven residues have a best predicted conformation with RMSD $\geq$ 1.0 Å for the main-chain atoms. For all the others, the RMSD is less than 2.0 Å. However, even the predictions that deviate the most seem to have much in common with the crystal structures. Two of those are shown in Fig. 6, for the 12 mer (RMSD = 1.5 Å) and 15 mer (RMSD = 1.97 Å) cyclics. These results are reasonable despite the uncertainties in the scoring functions for cyclic peptides [47].

Nikiforovich et al. [48] employed their own technique for exploring the conformational space of seven cyclic pentapeptides with known X-ray structures (see Table 1 in their paper). In Table 2, we present our best RMSD for these pentapeptides. The average RMSD for the seven molecules is 0.70 Å by the stochastic algorithm.

Some residues, such as prolines and glycines are known to have significant effects on secondary structure elements in proteins [63–65]. We examined a set of cyclic hexapeptides that have different sequences including a varying number of prolines and glycines. The stochastic algorithm was applied to these hexapeptides, and 2000 best conformations were retained and clustered (see Section 2). The results are presented in Table 3. The number of clusters in the last column

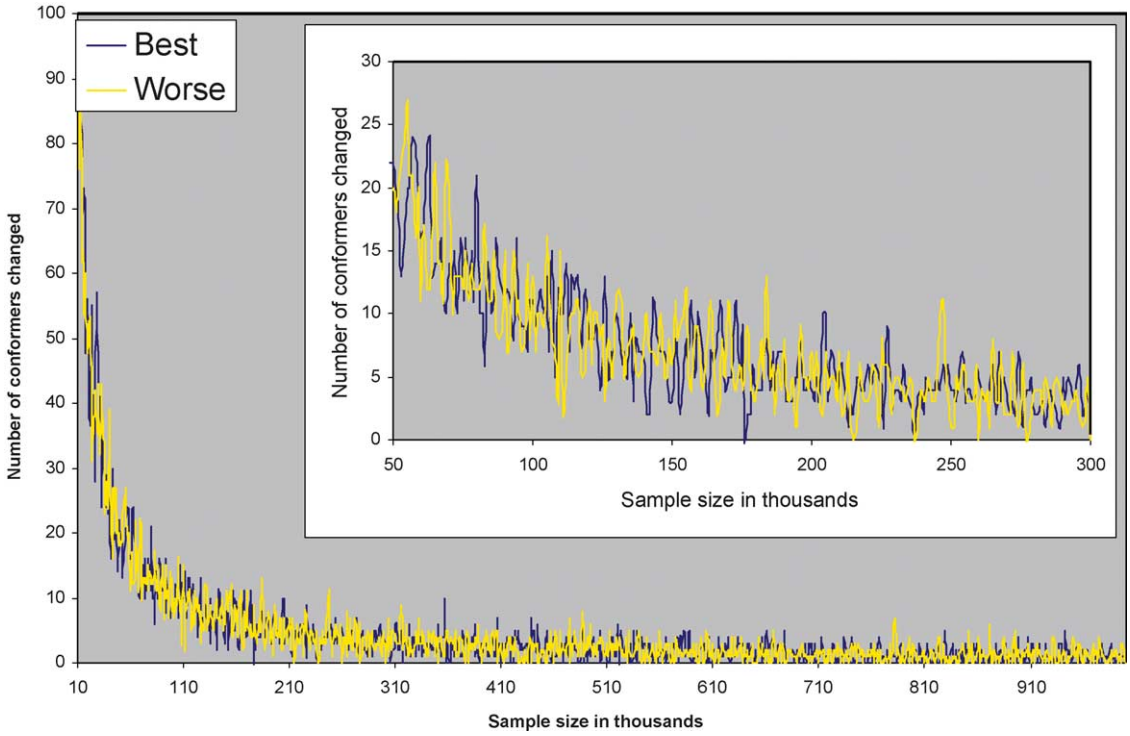

Fig. 5. Number of conformations that are modified (*Y*-axis) as a function of increasing the sample size (*X*-axis, in thousands).

Table 2
Comparing predictions to X-ray results

Sequence[a]	RMSD of backbone (N, CA, C, and O)	Cambridge data code
5AA		
GPSaP	0.40 (24, 6.0)[b]	(CGPSAQ)[c]
GPfAP	0.41 (35, 8.2)	(GICHOP)
APGfP	0.76 (190, 29.9)	(DABVIL)
GPGaP	0.78 (115, 13.3)	(CGPGAP10)
GPfGV	0.79 (22, 7.8)	(FUDWIK)
fPGaP	0.84 (51, 12.7)	(PAPGAP)
GPfGA	0.92 (82, 10.9)	(FIVSAE)
6AA		
GHGAYG	0.32 (34, 11.1)	(BAMLIK)
GGGGaa	0.45 (10, 3.1)	(GGAAGG)
GLGGLG	0.52 (61, 15.1)	(CGLEGL)
FlGfLG	0.92 (113, 16.2)	(CAHWEN)
FPaFPa	1.0 (8, 14.0)	(BIHTUH)
7AA		
SFLPVNL	0.62 (43, 15.2)	(TALVAD)
AIPFNSL	0.98 (2, 1.7)	(LACSUD)
IPIFPYP	1.0 (74, 45.1)	(JUXHAL)
8AA		
aGPFaGPF (1)	0.71 (145, 18.8)	(CEWCIQ10)
aGPFaGPF (2)	0.83 (39, 9.2)	(CACNOJ10)
aGPfaGPf (3)	1.3 (327, 28.3)	(DASXIE)
9AA		
PPFFLIILV	1.4 (59, 30.0)	(GIPKAR10)
10AA		
VPPFFVPPFF	1.1 (181, 72.3)	(DUTLAF10)
AFFPPFFVPP	1.4 (845, 2001.4)	(ANTAHC10)
PPIFVLPPYI	1.6 (168, 38.6)	(NEHYAA)
12AA		
VPGVPGVPGVPG	1.5 (18, 11.2)	(DEWFEQ)
15AA		
VPGVGVPGVGVPGVG	1.97 (88, 50.3)	(VPGVGB10)

[a] The one-letter symbols in lower case are for D-amino acid residues.
[b] In the second column, in parentheses, the first number is the rank of the best structures while the second one is the energy difference from the global minimum in kcal/mole.
[c] The name of the cyclic peptide in the Cambridge database.

Table 3
The effect of glycine and proline content on the number of clusters collected for 12 cyclo hexapeptides

Sequence	Pro	Gly	Others	Clusters
PPGPLG	3	2	1	135
FPAFPA	2	0	4	162
PVFFAG	1	1	4	215
FLGFLG	0	2	4	257
FLGLFG	0	2	4	269
GPGGPG	2	4	0	286
GLLGLL	0	2	4	307
GHGAYG	0	3	3	316
AAGGAG	0	3	3	475
AAGAGG	0	3	3	529
GGGGAA	0	4	2	553
GLGGLG	0	4	2	623

changes with the numbers of prolines and glycines. It may be seen that although the variance in some of the peptides with a similar number of glycines is quite large (when the number of Gly residues is three), the average for two, three, and four Gly residues displays a trend of increasing the number of clusters, while those for Pro display a trend of reduction in the number of clusters.

A discussion of the flexibility/rigidity of cyclic peptides [36,66,67] is deferred to another paper in which a detailed analysis and comparison of predicted to reported conformations will be presented.

In addition to the need for efficient search methods, the scoring of molecular energies of cyclic peptides by molecular mechanics is not yet of high enough quality. In addition, the case of cyclic peptides differs from proteins in that it is not expected to find that structures that are closest to the crystal conformations should necessarily have the lowest energies. Crystal packing and solvent effects probably play a larger role in cyclic peptide conformations [68,69] than in the conformations of proteins. In Table 4, we present 10 conformations of a cyclic pentapeptide, of sequence GPSaP, which have the lowest RMSD values with respect to experiment. These best results have different energies with respect to the global minimum (second column) and their ranks (out of 235 fully minimized conformations, within Sybyl 6.8) are presented in the third column. It may be seen that native like structures are not easily identified or grouped together by this force field. A similar effect is found also for the conformations of a hexapeptide, sequence GLGGLG, presented in Table 4. In this cyclic peptide, the total number of conformations was 623.

Table 4
Ten best (RMSD-wise) conformations of cyclic pentapeptide GPSaP as well as cyclic hexapeptide GLGGLG

Conformer	RMSD	ΔE (kcal/mol)	Rank
Cyclic pentapeptide GPSaP			
1	0.36	6	24
2	0.49	7.5	30
3	0.53	6.3	25
4	0.55	4.6	19
5	0.56	9	41
6	0.57	4.5	17
7	0.58	0.3	2
8	0.59	5.2	22
9	0.6	4.7	20
10	0.64	4.3	14
Cyclic hexapeptide GLGGLG			
1	0.53	15.1	61
2	0.63	17.1	117
3	0.67	25.6	399
4	0.69	23.4	321
5	0.69	16.9	108
6	0.70	19.7	207
7	0.72	21.9	275
8	0.73	18.2	162
9	0.74	15.1	62
10	0.75	21.3	256

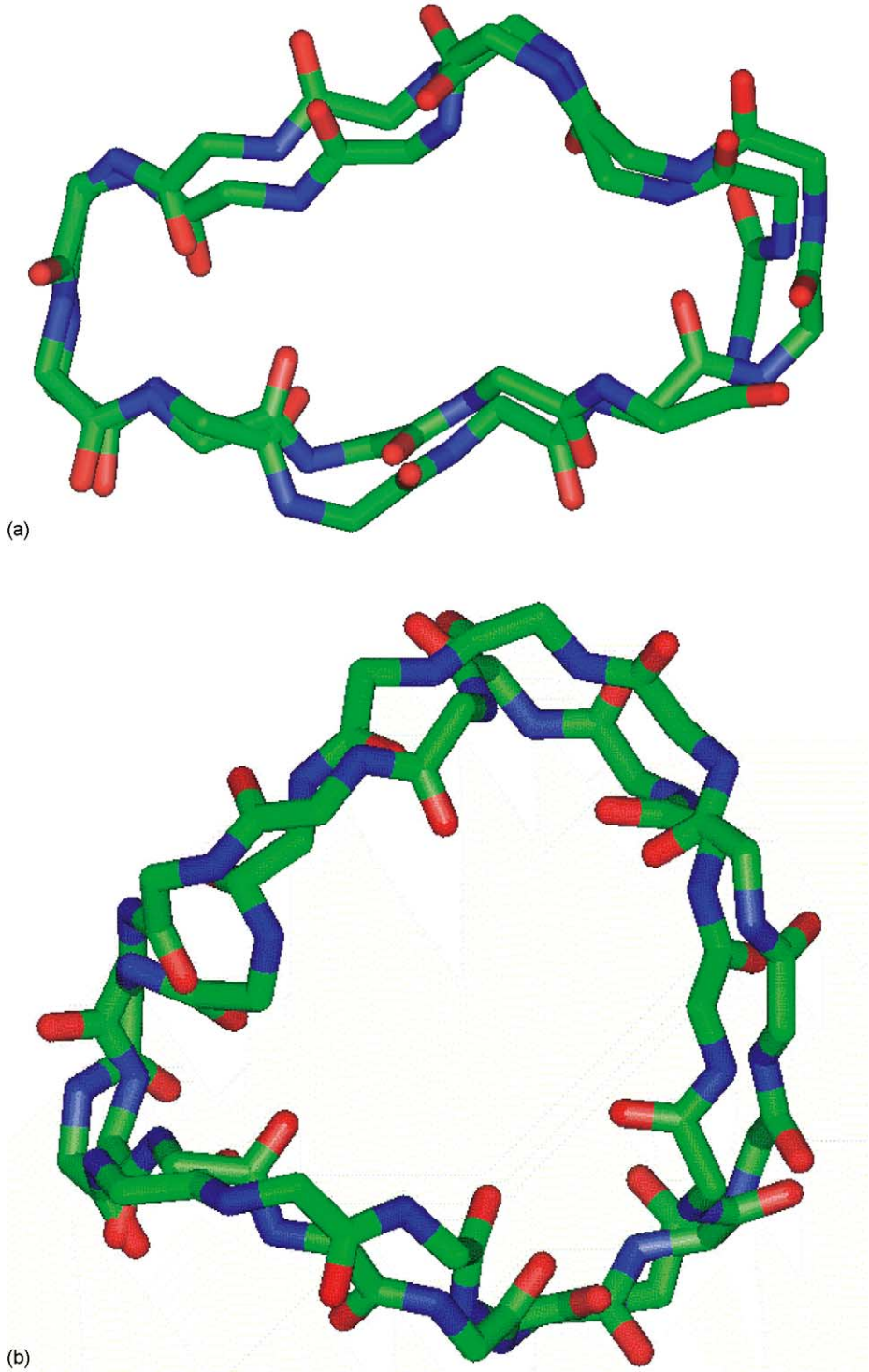

Fig. 6. Comparing the experimental to predicted results for the backbone structure of: (a) 12 mer cyclic peptide; (b) 15 mer cyclic peptide (see Table 2 for peptide sequences).

4. Discussion

Our stochastic search algorithm has been previously applied to several problems of protein structure, which were restricted by the presence of a protein template. The problem of predicting cyclic peptide conformations is somewhat different in that no external restriction exists apart from the internal link of the peptide chain onto itself and the interactions of side chains with backbone and other side chains. In this paper, we have shown that searching the conformational space of large cyclic peptides is feasible and can lead to good results. In this application of the algorithm, we divide

the task in two. The first stage constructs the cyclic backbone and determines which of the conformations may be a candidate for further exploration. In the second stage, side chains are added and a full minimization of all the viable structures takes place.

The best verification of the algorithm's reliability in searching space is achieved by comparing the stochastic search to an exhaustive one, albeit on a relatively small system due to time limits for running the exhaustive search. In such a comparison, the effect of the particular "force field" is eliminated and it purely reflects the search capabilities. We have shown, in Table 1 and in Figs. 2 and 3, that in these comparisons, the global minimum has been invariably detected, and a large set of best conformations has been reproduced. The size of the reproduced set depends, however, on the number of conformations that remains for the exhaustive stage of the calculations (the "threshold"), which cannot be exactly determined without a full exhaustive search, which is not possible for large problems. In the case of the cyclic tetrapeptide presented in Table 1, if the threshold that was determined for switching from stochastic to exhaustive search is larger than the required minimum, the stochastic search method is able to reproduce at least 98.5% of the best 1000 structures or 100% of the 384 best structures.

In large systems, in which it is impossible to obtain results from an exhaustive search and to compare them to the stochastic ones, it is impossible, at any stage, to determine the point of switching between the stochastic and exhaustive stages in order to retain a large set of best structures. What are then the chances that the best conformations are actually reproduced? In Fig. 2, the percent of best structures asymptotically approaches 0% as the minimal number of combinations (the "threshold") is reduced. However, even at very low threshold values, there is still some percent of the best structures that is reproduced, including the global minimum. It is clear then that we could possibly lose several best results. However, a partial ensemble, a subset of the best possible results, may still be discovered even with a very low threshold.

The time for finding the best conformations varies widely according to the parameters presented in Table 1 and depicted in Fig. 4. There is a highly linear correlation between the number of variables (which are directly related to the exponent of the number of combinations) and the number of stochastic iterations (which roughly determines the length of computations). We studied this relation for the set of cyclic peptides with four to eight amino acids and the linear relation strictly holds in this region. A similar relation was found previously for rotamers in proteins [21] and we may again suggest that using a parallel code should be highly beneficial for the stochastic search method and could considerably reduce the time for computation.

We probed the size of the sample (Table 1 and Fig. 5) and found that increasing the size contributes to the "stabilization" of the results, in terms of the variation in the number of conformations that form the "best set." From Fig. 5, it may be seen that, as we employed a somewhat lower number of conformations ($n = 10^5$), a larger sample could increase the stability of results even further. However, larger samples require longer calculations, and this sample size is large enough to allow subsequent statistical tests in the H and L sets. We have also tested the effect of using more stringent conditions (EVF in "H" and "L") for evicting variable values and found that a demand for values with frequencies of half (or twice, depending on H or L) the size of their statistical expectation is enough for reproducing the best results in a reasonable time. In the future, it may be useful to study the effect of varying EVF as a function of the iterations or of the number of values that remain for each variable, separately.

In the first stage of the search, the scoring of the stochastically constructed peptides enables to form a large and diverse set of accessible conformations, which are evaluated in a second stage. Therefore, the force field of the initial stage is a very crude one, which requires mostly a good "closure" of the ends of the cyclic peptide that have been positioned during the stochastic construction. The set of conformations at this stage should therefore be large enough in order to allow such a large variety. Again, the introduction of a much more rigorous force field in the second stage is still not able to find a one-to-one correspondence between the deviation from crystal structures and increasing energy (Table 4). The force field employed in the second stage was unable to distinguish between "native" and "non-native like" structures, if we may use such definitions for crystal structures of cyclic peptides.

Results for predicting the crystallographic conformations of cyclic peptides have been overall successful. According to Fiser et al. [70], a good prediction has a RMSD smaller than 1 Å, while a bad prediction has RMSD larger than 2 Å. Those that fall between the two extremes are considered to be of medium quality. Based on this approach, we succeeded (Table 2) to predict even very large peptides in adequate accuracy. We are now able to explore the conformational space of cyclic peptides and to collect a large set of low energy structures. The results could serve for studying structure activity relations (SAR) and for comparisons to experiments such as nuclear magnetic resonance. In addition, a set of low energy structures of cyclic peptides could be employed for flexible docking to proteins.

The issue of flexibility is a crucial one for cyclic peptides. We have not probed yet the flexibility of each of the predicted cyclics in comparison to NMR results or to results of molecular dynamics studies. We did demonstrate that in our search and in subsequent clustering, we find that peptides having larger numbers of proline residues are increasingly "rigid," i.e., they cluster into a relatively smaller number of groups. While, for glycine containing peptides, the opposite effect has been found. Both results for cyclic peptides are in agreement with experimental observations.

Having to construct each cyclic peptide on the basis of backbone atoms alone, how reliable is the set of best

conformations and how much does it depend on the particular amino acid content? It is clear from Table 2, in which each group of similarly sized cyclic peptides have been presented, that the content has an effect on the results. But, since we use standard bond lengths and angles and only vary dihedrals according to a database of the residues, residues that have similarly allowed regions in space will contribute exactly the same in the first stage. In the second stage, the actual side chains are added and therefore, the final results certainly reflect the type and configuration of the particular amino acid sequence.

The scanning of a very large conformational space and the attempt to detect all of the possible backbone conformations gives us an opportunity to have a reliable database of backbones for a given sequence. The importance of producing such ensembles cannot be overemphasized. In other methods, such as molecular dynamics, Monte Carlo, and simulated annealing which have been widely used for conformational analysis of cyclic peptides, the ability to detect the global minimum as well as most of the low energy structures is very limited [50]. We should still increase the reliability of the set of best conformations by improving the force fields that rank them at the different stages. Producing ensembles is an important endeavor because it is clearer today than ever before that molecular properties and molecular interactions usually take place between flexible entities that are present in appropriate proportions, and many of them should be considered for their potential contributions to the overall reaction path and mechanism [50].

In comparison to the previous version of this algorithm, the introduction of the large samples and of EVF is an important modification [21]. There, only a single variable value could be evicted in any iteration, and so the number of iterations was roughly determined by the largest number of values for a single variable. Samples were much smaller, of size 10^3. In the current version, it is possible to evict most of the values of a single variable, in one iteration, but, in order to do so, a very large sample must be constructed, at each iteration.

We have also verified that the current method of sampling and evicting is not dependent on the random number generation of seed numbers that determine the choice of variable values. We make the decisions of evicting values after analyzing a large database of conformations that were produced randomly. This reduces the effect of the variability in choosing the first seed number. In that respect, the outcome of this search strategy is somewhat different than in the previous version of this algorithm [21], in which a change of seed number affected the results considerably.

5. Conclusions

A new search technique has been devised for exploring the multidimensional space of cyclic peptides and for detecting the global minimum, as well as most of the other best (i.e., lowest energy) minima. The search algorithm is highly reliable and efficient, and thus will be further applied to other problems that require complex optimizations or conformational search for ensembles of structures. The uniqueness of the approach is its ability to reproduce a large set of best results in problems of high dimensionality, instead of searching for the global minimum alone. In fact, there could be instances in which a good ensemble will be formed, while losing the global minimum. The production of such ensembles is recognized today as an important contribution to the understanding of chemical and biological activity.

Acknowledgements

This research was supported by the Israel Science Foundation grant no. 608/02. We thank the Alex Grass Center for Drug Design and Synthesis for supporting the purchase of a workstation for this research. We wish to thank professor G.V. Nikiforovich for sending us cyclic peptide coordinates that were helpful for comparing our results to experiments.

References

[1] P. Cuniasse, I. Raynal, A. Yiotakis, V. Dive, Accounting for conformational variability in NMR structure of cyclopeptides: ensemble averaging of interproton distance and coupling constant restraints, J. Am. Chem. Soc. 119 (1997) 5239–5248.
[2] H. Li, N. Sampson, Structural analysis of fertilin beta cyclic peptide mimics that are ligands for alpha(6)beta(1) integrin, J. Pept. Res. 59 (2002) 45–54.
[3] M. Groth, J. Malicka, S. Rodziewicz-Motowidlo, C. Czaplewski, L. Klaudel, W. Wiczk, A. Liwo, Determination of conformational equilibrium of peptides in solution by NMR spectroscopy and theoretical conformational analysis: application to the calibration of mean-field solvation models, Biopolymers 60 (2001) 79–95.
[4] H. Matter, H. Kessler, Structures, dynamics, and biological-activities of 15 cyclic hexapeptide analogs of the alpha-amylase inhibitor tendamistat (Hoe-467) in solution, J. Am. Chem. Soc. 117 (1995) 3347–3359.
[5] J. Malicka, M. Groth, C. Czaplewski, W. Wiczk, A. Liwo, Conformational studies of cyclic enkephalin analogues with L- or D-proline in position 3, Biopolymers 63 (2002) 217–231.
[6] R. Unger, J. Moult, Finding the lowest free energy conformation of a protein is an NP-hard problem: proof and implications, Bull. Math Biol. 55 (1993) 1183–1198.
[7] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[8] W.J. Zhang, G.V. Nikiforovich, J. Perodin, D.E. Richard, E. Escher, G.R. Marshall, Novel cyclic analogs of angiotensin II with cyclization between positions 5 and 7: conformational and biological implications, J. Med. Chem. 39 (1996) 2738–2744.
[9] A.M. Sefler, G. Lauri, P.A. Barlett, A convenient method for determining cyclic peptide conformation from 1D H-1-NMR information, Int. J. Pept. Protein Res. 48 (1996) 129–138.

[10] P.I.W. de Bakker, M.A. DePristo, D.F. Burke, T.L. Blundell, Ab initio construction of polypeptide fragments: accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the generalized born solvation model, Proteins 51 (2003) 21–40.
[11] H. Gohlke, G. Klebe, Statistical potentials and scoring functions applied to protein–ligand binding, Curr. Opin. Struct. Biol. 11 (2001) 231–235.
[12] H.H. Gan, A. Tropsha, T. Schlick, Lattice protein folding with two- and four-body statistical potentials, Proteins 43 (2001) 161–174.
[13] F. Guarnieri, H. Weinstein, Conformational memories and the exploration of biologically relevant peptide conformations: an illustration for the gonadotropin-releasing hormone, J. Am. Chem. Soc. 118 (1996) 5580–5589.
[14] F. Guarnieri, S.R. Wilson, Conformational memories and a simulated annealing program that learns—application to Ltb(4), J. Comput. Chem. 16 (1995) 648–653.
[15] S.A. Hassan, F. Guarnieri, E.L. Mehler, A new approach for folding peptides: conformational memories with a new implicit solvent model, Biophys. J. 78 (2000) 1978 (Pos Part 2).
[16] I. Kolossvary, W.C. Guida, Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides, J. Am. Chem. Soc. 118 (1996) 5011–5019.
[17] H. Senderowitz, F. Guarnieri, W.C. Still, A smart Monte Carlo technique for free-energy simulations of multiconformational molecules, Direct calculations of the conformational populations of organic-molecules, J. Am. Chem. Soc. 117 (1995) 8211–8219.
[18] H. Senderowitz, W.C. Still, MC(JBW): simple but smart Monte Carlo algorithm for free energy simulations of multiconformational molecules, J. Comput. Chem. 19 (1998) 1736–1745.
[19] J. Lee, H.A. Scheraga, S. Rackovsky, New optimization method for conformational energy calculations on polypeptides: conformational space annealing, J. Comput. Chem. 18 (1997) 1222–1232.
[20] J.M. Goodman, W.C. Still, An unbounded systematic search of conformational space, J. Comput. Chem. 12 (1991) 1110–1117.
[21] M. Glick, A. Rayan, A. Goldblum, A stochastic algorithm for global optimization and for best populations: a test case of side chains in proteins, Proc. Natl. Acad. Sci. USA 99 (2002) 703–708.
[22] S. Behrens, B. Matha, G. Bitan, C. Gilon, H. Kessler, Structure-activity relationship of the ring portion in backbone-cyclic C-terminal hexapeptide analogs of substance P, Int. J. Pept. Protein Res. 48 (1996) 569–579.
[23] J. Boer, D. Gottschling, A. Schuster, M. Semmrich, B. Holzmann, H. Kessler, Design and synthesis of potent and selective alpha(4)beta(7) integrin antagonists, J. Med. Chem. 44 (2001) 2586–2592.
[24] E.K. Bradley, S.C. Ng, R.J. Simon, D.C. Spellmeyer, Synthesis, molecular modelling, and NMR structure determination of four cyclic peptide antagonists of endothelin, Bioorg. Med. Chem. 2 (1994) 279–296.
[25] C. Breton, H. Chellil, M. Kabbaj-Benmansour, E. Carnazzi, R. Seyer, S. Phalipou, D. Morin, T. Durroux, H. Zingg, C. Barberis, B. Mouillac, Direct identification of human oxytocin receptor-binding domains using a photoactivatable cyclic peptide antagonist—comparison with the human V-1a vasopressin receptor, J. Biol. Chem. 276 (2001) 26931–26941.
[26] K. Burgess, D. Lim, S.A. Mousa, Synthesis and solution conformation of cyclo RGDRGD: a cyclic peptide with selectivity for the alpha V beta 3 receptor, J. Med. Chem. 39 (1996) 4520–4526.
[27] P. Cucchi, S. Meini, L. Quartara, A. Giolitti, S. Zappitelli, L. Rotondaro, C.A. Maggi, Interaction of linear and cyclic peptide antagonists at the human B-2 kinin receptor, Peptides 23 (2002) 1457–1463.
[28] J. Goldberg, Cyclic peptide antibiotics self-assembly required, Trends Biotechnol. 19 (2001) 379–379.
[29] R. Kasher, D.A. Oren, Y. Barda, C. Gilon, Miniaturized proteins: the backbone cyclic proteinomimetic approach, J. Mol. Biol. 292 (1999) 421–429.
[30] I.L. Karle, Aspects of peptide folding and aggregation, Acc. Chem. Res. 32 (1999) 693–701.
[31] N. Schmiedeberg, M. Schmitt, C. Rolz, V. Truffault, M. Sukopp, M. Burgle, O.G. Wilhelm, W. Schmalix, V. Magdolen, H. Kessler, Synthesis, solution structure, and biological evaluation of urokinase type plasminogen activator (uPA)-derived receptor binding domain mimetics, J. Med. Chem. 45 (2002) 4984–4994.
[32] E. Williams, G. Williams, B.J. Gour, O.W. Blaschuk, P. Doherty, A novel family of cyclic peptide antagonists suggests that *N*-cadherin specificity is determined by amino acids that flank the HAV motif, J. Biol. Chem. 275 (2000) 4007–4012.
[33] G. Zanotti, M. Saviano, G. Saviano, T. Tancredi, F. Rossi, C. Pedone, E. Benedetti, Structure of cyclic peptides: the crystal and solution conformation of cyclo(Phe–Phe–Aib–Leu–Pro), J. Pept. Res. 51 (1998) 460–466.
[34] A. Stavrakoudis, S. Makropoulou, V. Tsikaris, M. Sakarellos-Daitsiotis, C. Sakarellos, I.N. Demetropoulos, Computational screening of branched cyclic peptide motifs as potential enzyme mimetics, J. Pept. Sci. 9 (2003) 145–155.
[35] A.J. Lucke, J.D.A. Tyndall, Y. Singh, D.P. Fairlie, Designing supramolecular structures from models of cyclic peptide scaffolds with heterocyclic constraints, J. Mol. Graph. 21 (2003) 341–355.
[36] M.L. Valero, J.A. Camarero, T. Haack, M.G. Mateu, E. Domingo, E. Giralt, D. Andreu, Native-like cyclic peptide models of a viral antigenic site: finding a balance between rigidity and flexibility, J. Mol. Recognit. 13 (2000) 5–13.
[37] R.M. Kohli, C.T. Walsh, M.D. Burkart, Biomimetic synthesis and optimization of cyclic peptide antibiotics, Nature 418 (2002) 658–661.
[38] T.M. Kinsella, C.T. Ohashi, A.G. Harder, G.C. Yam, W.Q. Li, B. Peelle, E.S. Pali, M.K. Bennett, S.M. Molineaux, D.A. Anderson, E.S. Masuda, D.G. Payan, Retrovirally delivered random cyclic peptide libraries yield inhibitors of interleukin-4 signaling in human B cells, J. Biol. Chem. 277 (2002) 37512–37518.
[39] M. Hahn, D. Winkler, K. Welfle, R. Misselwitz, H. Welfle, H. Wessner, G. Zahn, C. Scholz, M. Seifert, R. Harkins, J. Schneider-Mergener, W.G. Hohne, Cross-reactive binding of cyclic peptides to an anti-TGF alpha antibody fab fragment: an X-ray structural and thermodynamic analysis, J. Mol. Biol. 314 (2001) 293–309.
[40] D.A. Horton, G.T. Bourne, M.L. Smythe, Exploring privileged structures: the combinatorial synthesis of cyclic peptides, J. Comput. Aided Mol. Des. 16 (2002) 415–430.
[41] D.P. Fairlie, G. Abbenante, D.R. March, Macrocyclic peptidomimetics—forcing peptides into bioactive conformations, Curr. Med. Chem. 2 (1995) 654–686.
[42] M. MacDonald, J. Aube, Approaches to cyclic peptide beta-turn mimics, Curr. Org. Chem. 5 (2001) 417–438.
[43] H. Kessler, R. Gratias, G. Hessler, M. Gurrath, G. Muller, Conformation of cyclic peptides. Principle concepts and the design of selectivity and superactivity in bioactive sequences by 'spatial screening', Pure Appl. Chem. 68 (1996) 1201–1205.
[44] M. Sukopp, L. Marinelli, M. Heller, T. Brandl, S.L. Goodman, R.W. Hoffman, H. Kessler, Designed beta-turn mimic based on the allylic-strain concept: evaluation of structural and biological features by incorporation into a cyclic RGD peptide (cyclo(-L-arginylglcyl-L-alpha-aspartyl-)), Helv. Chim. Acta 85 (2002) 4442–4452.
[45] P. Desai, M. Prachand, E. Coutinho, A. Saran, J. Bodi, H. Suli-Vargha, Activity and conformation of a cyclic heptapeptide possessing the message sequence His–Phe–Arg–Trp of alpha-melanotropin, Int. J. Biol. Macromol. 30 (2002) 187–195.
[46] A. Descours, K. Moehle, A. Renard, J.A. Robinson, A new family of beta-hairpin mimetics based on a trypsin inhibitor from sunflower seeds, Chem. Biochem. 3 (2002) 318–323.
[47] C. Keasar, R. Rosenfeld, Empirical modifications to the AMBER/OPLS potential for predicting the solution conformations of cyclic peptides by vacuum calculations, Fold. Des. 3 (1998) 379–388.

[48] G.V. Nikiforovich, K.E. Kover, W.J. Zhang, G.R. Marshall, Cyclopentapeptides as flexible conformational templates, J. Am. Chem. Soc. 122 (2000) 3262–3273.
[49] N. Loiseau, J.M. Gomis, J. Santolini, M. Delaforge, F. Andre, Predicting the conformational states of cyclic tetrapeptides, Biopolymers 69 (2003) 363–385.
[50] A.A. Grover, R. Kishore, Characterization of a novel type VII beta-turn conformation for a bio-active tetrapeptide rigin—a synergy between theoretical and experimental results, Eur. J. Biochem. 267 (2000) 1455–1463.
[51] K. Krause, L.F. Pineda, R. Peteranderl, S. Reissmann, Conformational properties of a cyclic peptide bradykinin B-2 receptor antagonist using experimental and theoretical methods, J. Pept. Res. 55 (2000) 63–71.
[52] C. Baysal, H. Meirovitch, Efficiency of simulated annealing for peptides with increasing geometrical restrictions, J. Comput. Chem. 20 (1999) 1659–1670.
[53] H. Morita, T. Kayashita, K. Takeya, H. Itokawa, M. Shiro, Conformation of cyclic heptapeptides: solid and solution state conformation of yunnanin A, Tetrahedron 53 (1997) 1607–1616.
[54] C. Baysal, H. Meirovitch, Free energy based populations of interconverting microstates of a cyclic peptide lead to the experimental NMR data, Biopolymers 50 (1999) 329–344.
[55] A. Rayan, E. Noy, D. Chema, A. Levitzki, A. Goldblum, Stochastic algorithm for kinase homology model construction, Curr. Med. Chem. 11 (2004) 675–692.
[56] M. Glick, A. Goldblum, A novel energy-based stochastic method for positioning polar protons in protein structures from X-rays, Proteins 38 (2000) 273–287.
[57] B. Oliva, P.A. Bates, E. Querol, F.X. Aviles, M.J.E. Sternberg, An automated classification of the structure of protein loops, J. Mol. Biol. 266 (1997) 814–830.
[58] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E.I. Cheatham, D.M. Ferguson, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER, University of California, San Francisco, Version 4.1, 1995.
[59] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[60] J.H. Viles, J.B.O. Mitchell, S.L. Gough, P.M. Doyle, C.J. Harris, P.J. Sadler, J.M. Thornton, Multiple solution conformations of the integrin-binding cyclic pentapeptide cyclo(-Ser–D-Leu–Asp–Val–Pro-) Analysis of the (phi, psi) space available to cyclic pentapeptides, Eur. J. Biochem. 242 (1996) 352–362.
[61] M. Clark, R.D. Cramer, N. Vanopdenbosch, Validation of the general-purpose tripos 5.2 force-field, J. Comput. Chem. 10 (1989) 982–1012.
[62] F.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelinkpeters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers, D.G. Watson, Cambridge crystallographic data center—computer-based search, retrieval, analysis and display of information, Acta Crystallogr. Sect. B—Struct. Commun. 35 (1979) 2331–2339.
[63] D.F. Moriarty, W. Colon, Peptide fragments from the DNA binding protein Fis: influence of proline on secondary structure, Biochemistry 41 (2002) 159.
[64] L. Brachais, H. Duclohier, C. Mayer, D. Davoust, Influence of proline-14 substitution on the secondary structure in a synthetic analog of alamethicin, Biopolymers 36 (1995) 547–558.
[65] H.R. Kricheldorf, E.T.K. Haupt, D. Muller, Secondary structure of peptides. Part 17. *Cis–trans* isomerism of solid proline-containing oligopeptides as revealed by C-13 Nmr Cp/Mas spectroscopy, Magn. Reson. Chem. 24 (1986) 41–52.
[66] P. Amodeo, G. Saviano, G. Borin, A. Calderan, P. Ruzza, T. Tancredi, Solution conformational analysis of sodium complexed [Gly6]- and [Gly9]-antamanide analogs, J. Pept. Res. 51 (1998) 180–187.
[67] I.L. Karle, Flexibility in peptide molecules and restraints imposed by hydrogen bonds, the AiB residue, and core inserts, Biopolymers 40 (1996) 157–180.
[68] B. Diblasio, A. Lombardi, X. Yang, C. Pedone, V. Pavone, Beta-Alanyl–Beta-Alanine in cyclic beta-turned peptides, Biopolymers 31 (1991) 1181–1188.
[69] I.L. Karle, Folding, aggregation and molecular recognition in peptides, Acta Crystallogr. Sect. B—Struct. Commun. 48 (1992) 341–356.
[70] A. Fiser, R.K.G. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.