155. Generic Shapes for the Conformation Analysis of Macrocyclic Structures

by **Paul R. Gerber, Klaus Gubernator,** and **Klaus Müller***

Central Research Units, *F. Hoffmann-La Roche & Co. AG,* CH-4002 Basel

(24. VI. 88)

A new algorithm for the systematic generation of conformations of macrocyclic systems is presented. The procedure is based on the concept of generic shapes that are found in such structures. These shapes are characterized by a selection of harmonics which occur in an approximate *Fourier* representation of the atomic coordinates of the rings. Following a fixed protocol, a limited set of in-plane and out-of-plane circular harmonics is used to define an ensemble of generic ring shapes. These generic shapes are used as start structures for energy minimizations by a given force-field method. To account for the possibility of having several final conformations originating from the same generic shape, the corresponding initial structure is taken several times and subjected to a randomization step before minimization. The resulting conformations that fall within a preset low-energy band are collected and screened for duplicates and enantiomers. The efficiency of this procedure (ratio between the number of accepted conformations and the total number of energy minimizations) depends on the flexibility of the macrocyclic system. The efficiency is generally quite high for very flexible rings. According to the proposed protocol, the number of generic shapes used as start structures grows as the square of $N(\ln N)$, where N is the ring size. The algorithm lends itself to conformational analyses of medium-size and large rings as well as of loops spanned between fixed structural units.

1. Introduction. – The problem of finding the low-energy conformations of ring systems is of great importance both from theoretical and experimental points of view. The number of possible conformations usually grows rapidly with increasing ring size so that, even with the most judicious modeling, we may overlook relevant conformations. Spectroscopic methods generally do not provide a complete set of data for an unambiguous determination of conformation. Furthermore, the relative energetics of conformations may depend on the nature of the environment (solvation, interactions at molecular interfaces).

A classical approach to this problem consists of a systematic exploration of all torsional angles of a dissected ring system and a subsequent examination of all structures that approximately close the ring [1]. With large ring sizes, this method generates a formidable number of non-closing geometries and requires a considerable amount of computer time for the examination of all the remaining potentially ring-closing geometries. The problem can be somewhat reduced by either an analytical solution for the geometry of the last few bonds [2] or by restricting the sets of torsional angle values by certain rules [3]. Although these measures extend the range of treatable ring sizes, they do not remove the principal problem of exponential growth of the number of trials with ring size. The enumeration of non-overlapping closing paths on the diamond lattice has been proposed [4] as an alternative. This approach is conceptually equivalent to a very restricted screening of conformation space and is limited to even-membered ring systems.

As an alternative, molecular-dynamics calculations have been utilized to force a given structure through its possible conformation space [5]. The problem with this approach, as generally with molecular-dynamics methods, is that much computer time is spent tracing all the high-frequency modes of a structure. Since large-scale conformational changes usually involve low-frequency, large-amplitude modes, which take place on slow time scales, molecular-dynamics simulations would have to be performed for quite prolonged real-time spans in order to ensure reasonably complete coverage of conformation space. Still, the problems of establishing suitable sampling protocols and measures for completeness of coverage remain unsolved.

The present work proposes a new approach that overcomes many of the above mentioned problems. Our strategy aims at enumerating all the possible generic shapes for a given ring structure, from which sets of specific ring conformations are generated by energy minimizations using a suitable force field. Generic shapes are approximate *Fourier* representations of the atomic positions in different ring conformations. Generic shapes, in general, violate small-scale requirements, such as correct bond lengths, valence and torsion angles. However, such defects are quickly removed by the energy minimizations usually without much change in overall shape. In this way, the spectrum of possible conformations can be searched fairly exhaustively. Our procedure may be viewed as an extension of the concept of pseudorotation as used to describe the conformations of small rings [6]. However, the constraint of producing correct valence geometries for the initial structures is absent in our method.

2. Description of Ring Shape. – To classify the possible generic shapes of a ring structure of N atoms, we start with a circle in a polar-coordinate representation.

$$r_n = r_o,\ \vartheta_n = \frac{\pi}{2},\ \varphi_n = \frac{2\pi}{N} n,\ n = 1, \cdots, N. \tag{1}$$

We keep in mind that the structures will be subjected to minimization by a suitable force field which will take care of the local geometric requirements (specific bond lengths, valence and torsion angles). Thus, φ_n need not be varied, since such a variation would affect mainly the initial values of the bond lengths.

Significant variations in shape originate from varying ϑ_n and r_n. The natural way to classify these variations is in terms of modes characterized by node numbers k and j [6]:

$$\begin{aligned} \vartheta_n &= \frac{\pi}{2} + \theta_j \cdot \sin\left(\frac{2\pi}{N} j \cdot n + \delta_j\right),\ 2 \leqslant j \leqslant \frac{N}{2}, \\ r_n &= r_o + R_k \cdot \sin\left(\frac{2\pi}{N} \mathrm{k} \cdot n + \rho_k\right),\ 2 \leqslant k \leqslant \frac{N}{2}. \end{aligned} \tag{2}$$

Values of k and j above $N/2$ are excluded because of the discreteness of the variable n. Furthermore, $j = 0$ as well as $k = 1$ only move the centroid, while $j = 1$ tilts the ring, and $k = 0$ corresponds to a change in scale.

The remaining j and k values lead to characteristic shape distortions of the ring, each having a set of phase angles, δ_j and ρ_k, respectively, associated with it. The size of these sets depends on the values of j or k, on the one hand, and on the topological symmetry of the ring structure on the other hand. The first point is easily understood by considering an even-membered ring and the highest value $j = N/2$. In this case, only two distinct shapes

are possible, characterized by $\delta_j = \pm\frac{\pi}{2}$, placing even centers up and odd centers down or *vice versa*. The two cases may be identical, if the ring system has topological symmetry. Since the phases are restricted to the interval $-\pi < \delta_j \leqslant \pi$ and the spacing of values is reasonably on the order of $2\pi j/N$, there are roughly $N/j\,\delta$ values required to exhaust the relevant node positions.

Ring shapes s may then be generated in a very general manner by setting up expressions of the form

$$\vartheta_n^s = \frac{\pi}{2} + \sum_j \theta_j^s \cdot \sin\left(\frac{2\pi}{N} j \cdot n + \delta_j^s\right),$$
$$r_n^s = r_o + \sum_k R_k^s \cdot \sin\left(\frac{2\pi}{N} k \cdot n + \rho_k^s\right), \tag{3}$$

with specific values of the amplitudes θ_j^s, R_k^s and the phases δ_j^s, ρ_k^s.

3. Selection of Shape Parameters. – When producing generic ring shapes as start points for energy minimizations using a suitable force field, it makes little sense to explore all the possibilities offered by independent variation of amplitudes and phases of *Eqn. 3*. Since the minimization procedure removes and modifies most of the fine details that may be put into the start conformation, it is only necessary to select the essential shape-determining features.

We have investigated the appropriate choice of shape parameters by numerical experimentation. For economic reasons, we wanted to avoid too many start structures which eventually converge to an identical final state. However, by limiting the set of start structures too much, we may miss relevant conformations.

To study this question, we developed a program that performs a *Fourier* analysis on a given ring perimeter. With this tool, we examined a large number of ring perimeters contained in macrocyclic structures [7], as extracted from the *Cambridge Structural Data Files* [8]. From this study, it became apparent that, quite frequently, the structures show single dominant *Fourier* components in the axial and radial directions. Hence, the remaining components could be considered as minor modifications. *Fig. 1* exemplifies the notion of generic shape on a couple of experimental macrocyclic systems that are shown together with a continuous representation of corresponding shapes generated either from all or merely from the dominant harmonics as obtained from the corresponding *Fourier* analysis.

These findings led us to decide to keep only a single mode in ϑ and r for each initial shape, *i.e.* only a single j and k component in *Eqn. 3* have non-zero amplitude. At first, we even kept r_n constant and varied only ϑ_n. However, this restriction proved to be unsatisfactory, especially for large ring sizes where oval shapes were often missed. Thus, we now include radial variation in all cases (see also *Discussion*).

We restrict the variation of amplitudes along similar lines. The values of the θ amplitudes are made j-dependent *(Eqn. 4)* in order to avoid unnecessary distortions of bond lengths. The radial amplitudes are chosen independent of k, but as a simple function of N *(Eqn. 5)* to ensure acceptable ellipticities for different ring sizes.

$$\theta_j = 2/j \tag{4}$$

$$R_k/r_o = 1 - \text{Min}\left(\frac{\sqrt{2}}{2}, \frac{10}{N}\right) \tag{5}$$

The phases δ_j and ρ_k run through all the N/j and N/k values, respectively, as discussed in *Chapt. 2.* Thus, for large values of N the number of generic shapes, n_s, used as start structures, increases as *Eqn. 6*, which compares favourably with an exponential growth.

$$n_s \approx \sum_{j=2}^{N/2} \frac{N}{j} \sum_{k=2}^{N/2} \frac{N}{k} \approx (N \ln N)^2 \tag{6}$$

It is to be expected that large systems possess entire sets of conformations that result from to the same generic shape. The same shape is, therefore, used several times with a small random scatter added to the ϑ_n and r_n values each time. The algorithm proceeds to the next general shape only after a certain predefined number, f, of consecutive trials, after which the previous shape has not produced a new final structure. This randomization procedure is an additional measure to keep the set of generic shapes reasonably small.

To keep the program general enough for practical use, we allow for simple substituents on the ring. Simple chain-like substituents are restricted to a length of just one atom (Me, C=O O-atom, *etc.*), thereby excluding additional conformers due to rotations of the side chains. Furthermore, annelated rings with up to six atoms can be included, thus, allowing for proline residues and benzo derivatives. Cyclic peptides containing amino acids with long side chains have to be modeled with alanine as a substitute.

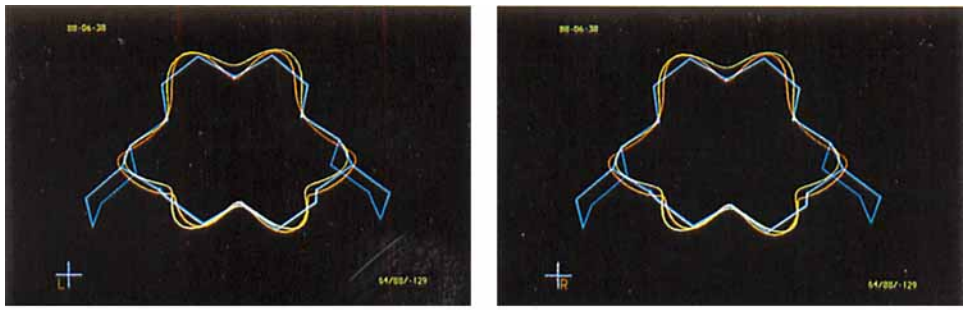
Single-atom substituents or annelated rings are added to the randomized cyclic start structures by attaching them with standard values for bond length and valence angles, as much as possible. Stereochemical constraints are taken into account by imposing corresponding terms in the force field at an initial stage of the minimization procedure. These terms are removed during the final relaxation.

4. Energy Function and Requirements of the Minimizer. – Since the structures of the global and local minima depend on the nature of the applied potential energy function, our procedure lends itself to an assessment of existing force-field methods. However, the focus of this paper is not on the quality of the force field used, but on an assessment of the completeness in locating all low-lying energy minima on the multidimensional conformational hypersurface for a given macrocyclic system.

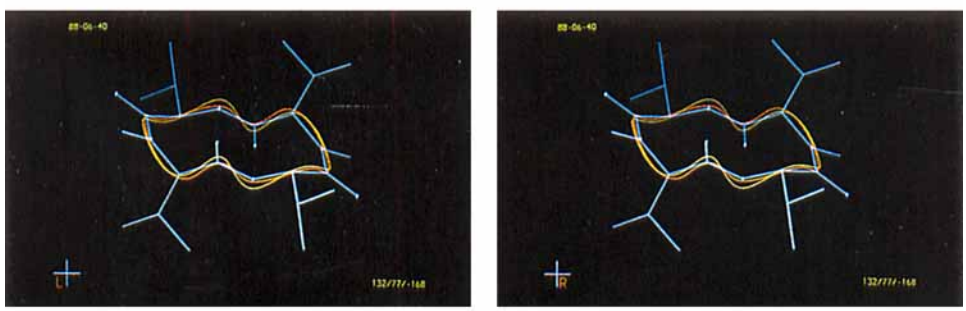
Since the start structures generated from generic ring shapes may violate the small-scale aspects of valence geometry substantially, the energy function needs terms to enforce stereochemical features such as (*Z*) or (*E*)-arrangements of double bonds or chirality at centers. Furthermore, the energy function should be free of singularities or problematic functional behavior which may compromise the minimization procedure.

The minimizer itself should stay within a given energy basin and must be able to handle highly deformed start conformations efficiently. The first requirement need not be strictly fulfilled, but our aim to find all minima excludes procedures that tend to escape from shallow secondary minima. A conjugate gradient method with a restart option [12] has proved very useful for our purpose.

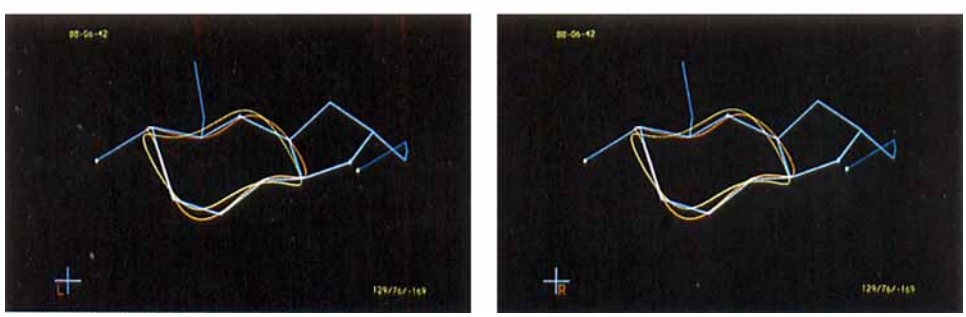
For the present investigation, we have used a special united-atom force field which is implemented in our in-house modeling system MOLOC [7]. It is based on a central-force-



R2	R3	R4	R5	R6	R7	R8	R9	·	A2	A3	A4	A5	A6	A7	A8	A9
35	3	1	0	123	0	5	0	·	365	24	21	6	19	6	11	13



R2	R3	R4	R5	R6	A2	A3	A4	A5	A6
187	0	4	0	3	0	9	0	69	0



R2	R3	R4	·	A2	A3	A4
17	8	0	·	11	67	1

Fig. 1. *The X-ray structures of the three macrocyclic derivatives isolaurallene (9-membered ring,* bottom*)* [9]*, cyclic depsipeptide (12-membered ring,* middle*)* [10]*, 18-crown-6 derivative (18-membered ring,* top*)* [1]*, displayed together with continuous representations of the corresponding* Fourier *series, using either all terms* (red) *or only the largest axial component and largest radial component* (yellow). Relative intensities of the axial and radial *Fourier* components, *An* and *Rn* ($n = 2, 3, \cdots, N/2$), respectively, are given below.

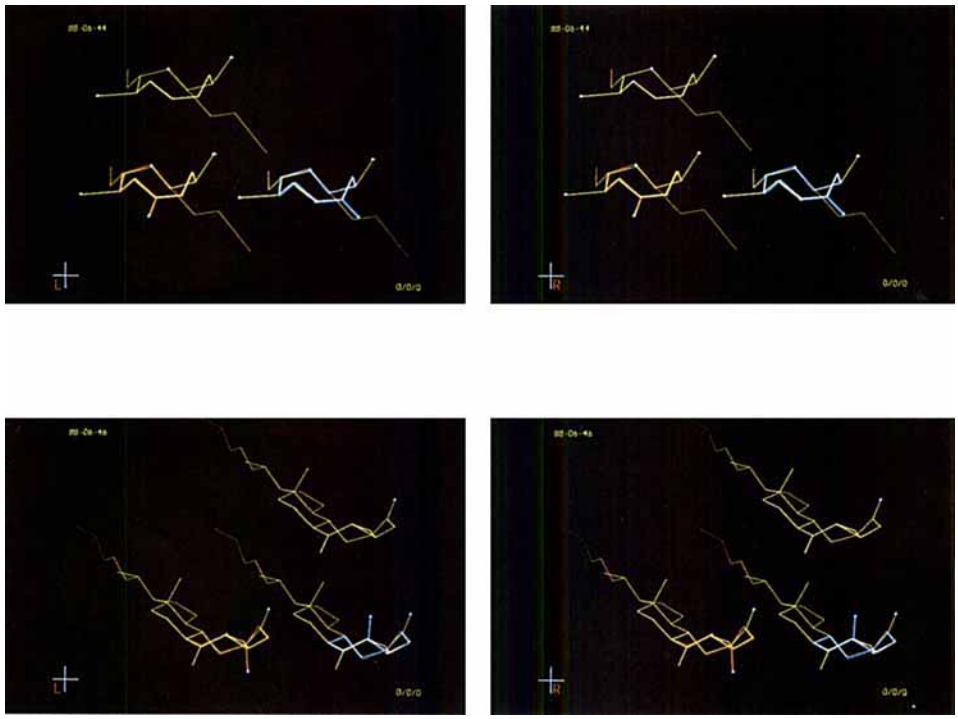
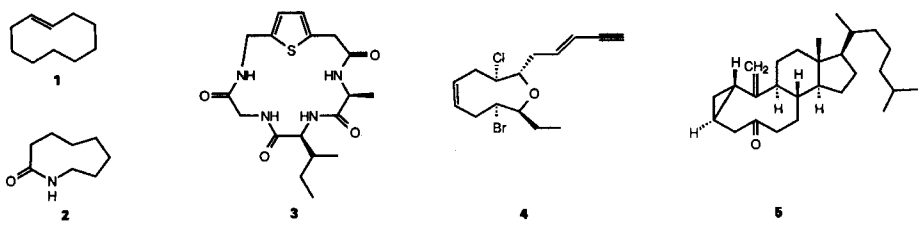


Fig. 2. Top: *Rigid-body superpositions of the X-ray structure of the (Z)-cyclononene derivative 4* [20] (yellow) *to the two computed s*-cis-*caprylolactam conformations* 2 (red) *and* 3 (blue) *of* Table 2. The two superpositions differ in the way the C=C bond is placed on top of the lactam unit. The rms deviations between corresponding ring atoms are 0.075 Å and 0.078 Å, respectively. Bottom: *Rigid-body superpositions of the X-ray structure of the 9-membered ring derivative* **5** [21] (yellow) *to the two computed s*-trans-*caprylolactam conformations* 22 (red) *and* 29 (blue) *of* Table 2. The two superpositions differ in the way the common C–C bond of the 9-membered ring and the *trans*-cyclopropyl moiety is placed on top of the lactam unit. The rms deviations between corresponding ring atoms are 0.124 and 0.116 Å, respectively.

field approximation with usual bond, valence angle, and torsion angle potentials. Non-bonded interactions include (1,4)- and (1,*n*)-type interactions ($n > 4$), containing attractive and repulsive components that both remain finite at zero distance and vanish at given threshold distances. A special pyramidality term takes care of the geometrical aspects at trigonal centers. While electrostatic interactions of the *Coulomb* type are omitted, H-bonding is accounted for by anisotropic geometric potentials. Finally, π-conjugative effects are allowed for *via* modified *Hückel*-type bond-order calculations. This force field has proved to be very efficient and reasonably accurate for qualitative modeling purposes.

5. Collection of Unique Conformations. – In our procedure, the same final structure may be produced several times from different start points. To avoid such duplicates, a newly generated structure is screened against the set of previously generated ones on the basis of both energetic and structural criteria. A new structure is kept, if its energy differs by more than a preset threshold value from the energies of previous structures. Otherwise, the new structure is exhaustively compared with all previous structures lying within a given energy threshold, by performing all possible rigid-body superpositions [13], including the mirror image where appropriate. The new structure is retained, if no satisfactory superposition can be obtained with respect to a maximum tolerance on the rms (root mean square) deviation between related atom positions. The threshold values have to be adapted according to computational accuracy considerations. To be on the safe side, an energy threshold of 0.1 kcal/mol (1 kcal = 4.187 kJ) and a maximum tolerance of 0.1 A for the rms deviation were chosen.

6. Discussion. – The proposed method is best discussed with reference to three specific examples. The first case, (*E*)-cyclodecene (**1**), allows us to comment on the generation of generic shapes, the importance of randomization, its repercussion on efficiency, and the implementation of geometrical constraints. The second case, caprylolactam (octahydro-2*H*-azonin-2-one; **2**), illustrates our method using a standard protocol in the absence of geometrical constraints. Both examples document the extensive coverage of conformation space by our method, the first by reference to an earlier conformation analysis by another force-field method, the second by reference to a large body of X-ray structure data from the literature. The third example, a macrocyclic tetrapeptide **3**, illustrates the potential of our method in combination with NMR spectroscopy.



*Example 1: (*E*)-Cyclodecene* (**1**). For cycloalkenes, a fairly extended investigation [14] of the low-energy conformations has been carried out using a consistent force-field method (CFF) [15].

One of the largest monocyclic systems considered in [14] is (*E*)-cyclodecene (**1**), for which ten conformational energy minima were located. In an exploratory investigation,

Table 1. *The Conformations of (*E*)-Cyclodecene, Calculated by the CFF Method* [14] *and by MOLOC* (this work), *Listed in Order of Ascending Relative MOLOC Energies* (in kcal/mol). Labels *M1–M10* refer to conformations reported in [14] (in order of ascending relative CFF energies). Additional conformations are labelled *A1–A9*. The five columns labelled Rf (f = 0,1, ···, 8) show how many times, for different randomization repeat factors f, individual conformations were found in runs using radial variation for generic shapes. The five columns to the right show the corresponding results for runs without radial variation. The three bottom rows list, for different runs, the total number, n_{tot} of initial structures (randomized generic shapes), the number, n, of unique final conformations, and the fraction $p = n/n_{tot}$. Each energy minimization was completed within 2 s CPU time on a *microVAX-3500*.

Conformation	Symmetry	MOLOC-E	CFF-E	R0	R1	R2	R4	R8	0	1	2	4	8
M7	*C*1	0.0	3.09	6	8	19	42	80	–	1	5	9	7
M5	*C*2	0.61	2.55	4	5	10	19	32	1	2	2	2	6
M3	*C*1	0.81	1.98	1	1	5	3	13	–	1	1	1	2
M2	*C*2	1.05	1.95	24	28	59	104	201	2	4	6	13	23
M4	*C*1	1.18	2.22	3	–	–	3	16	–	–	–	1	–
M1	*C*2	1.62	0.0	1	1	–	1	5	1	–	–	–	1
A1	*C*1	1.82	–	9	11	21	41	88	1	2	3	6	13
M6	*C*1	2.14	2.70	7	8	16	42	84	–	–	3	5	6
A2	*C*1	3.33	–	–	–	–	1	1	–	–	–	–	–
M8	*C*1	3.89	4.81	4	6	12	19	36	–	–	1	4	5
M9	*C*2	4.51	6.03	1	–	2	3	9	–	–	1	–	–
A3	*C*1	4.53	–	15	18	24	50	105	2	3	3	8	11
A4	*C*1	4.85	–	15	15	28	55	123	1	3	4	9	15
(*M10*)	(*C*2)	6.63	10.86	1	2	2	9	13	–	–	1	–	–
A5	*C*1	7.44	–	1	4	7	10	17	–	–	2	2	3
A6	*C*2	7.45	–	2	3	4	7	12	1	2	1	2	6
A7	*C*1	9.74	–	2	2	1	8	13	–	–	–	1	3
A8	*C*2	11.57	–	1	2	2	4	10	–	1	–	3	4
A9	*C*1	14.10	–	3	2	10	17	30	1	1	–	3	3
n_{tot}				100	116	222	438	888	10	20	33	69	108
n				18	16	16	19	19	8	10	13	15	15
p[%]				18	14	7	4	2	80	50	39	22	14

we varied only the ϑ_n, leaving the radial component constant. Using different randomization repeat factors, f, we collected up to 19 minima for this system *(Table 1)*. However, in each run several minima were missed. When we also varied the radial positions, most of these missed conformations were obtained for any value of f, due to the augmented set of generic shapes. In particular, an f value of four was sufficient to find all 19 minima.

Our ordering of energy levels differs from that of the highly parameterized CFF method. Nevertheless, all seven structures, reported to lie in the energy range of 0–3.5 kcal/mol [14], are also contained in our set of preferred conformations (structures *M1–M7*). Two additional dissymmetric conformations, *A1* and *A2*, are located in this energy range. Structure *A1* resembles *M4*, but represents a separate energy minimum, which is regularly located in all our runs. By contrast, structure *A2* is only rarely obtained. In the higher-energy range, we identify the two conformations *M8* and *M9* of [14], but find also several other structures, which are regularly located in all runs with radial variation. The symmetrical high-energy structure *M10* reported in [14] does not represent a minimum on our energy hypersurface, but connects two closely lying symmetry-related conformations in a very shallow double-minimum.

This example is well suited for some efficiency considerations. Increasing the number of start structures (randomized generic shapes) generally reduces the number of missed

minima with a concomitant decrease in efficiency due to an increase in redundant minimizations. This is documented in the right half of *Table 1*, where we show how many times each structure was obtained in runs with or without radial variation using different randomization repeat factors *f*. We note that certain minima are always and repeatedly found, while other require higher *f* values, and a few are found only occasionally.

We may use the percentage of unique structures found with respect to the total number of initially generated ones as an indicator of efficiency (bottom row of *Table 1*). Clearly, omission of radial variation leads to higher efficiencies at the cost of completeness. Thus, even for relatively small systems an eventual saving of CPU time would not justify a neglect of radial variation. For large ring systems, radial variation is even more crucial for obtaining a representative set of generic shapes.

Finally, we should comment on the implementation of torsional constraints enforcing (*Z*)- or (*E*)-arrangements of double bonds. We have found that such constraints should be represented by relatively soft terms so that only large deviations from the desired geometry lead to sizable restoring forces. Otherwise, initial structures may be pushed too quickly away from their original probing domains. This can be illustrated by the finding that application of a stringent (*E*)-constraint to the conformations *A2*, *M10*, and *A6* of *Table 1* forces them towards structures that relax into the conformations *M7*, *M9*, and *A4*, respectively, upon unconstrained minimization. In our present implementation, the (*Z*)/(*E*)-constraints vanish for torsional angle deviation below 40°.

Example 2: s-cis- *and s*-trans-*Caprylolactam* (*Octahydro-2*H*-azonin-2-one;* **2**). As part of an ongoing project on cyclic peptide mimetics, we performed a conformation analysis on **2**. Several X-ray structures of 9-membered lactam derivatives including the parent compound **2** are available in the literature [16–19]. Different conformations have been found for both the s-*cis*- and s-*trans*-lactams. Furthermore, there are many related 9-membered ring derivatives, such as lactones, peptides, depsipeptides, (*Z*)- and (*E*)-olefins, for which well refined X-ray structures are available from the *Cambridge Structural Data Files*.

The results of our conformation analysis are of particular interest here, because they document the ability of our method to reproduce experimentally observed conformations, to handle s-*cis*/s-*trans*-isomers, and to produce a fairly exhaustive set of low-energy conformations.

The conformation analysis for **2** was carried out without geometrical constraints for the amide unit and with a randomization repeat factor $f = 4$. It produced 33 structures, 21 s-*cis*- and 12 s-*trans*-lactam conformations, within an energy band of 10 kcal/mol (*Table 2*). The lowest s-*trans*-lactam structure lies approximately 4.6 kcal/mol above the global minimum, which contains a s-*cis*-amide unit. Gratifyingly, the experimentally observed conformations of both s-*cis*- and s-*trans*-lactam derivatives are reproduced within energy bands of 1 kcal/mol relative to the lowest-energy conformer of each series. The predicted structures agree fairly well with the experimental ones as judged from the rms deviations between corresponding ring atoms of 0.04–0.07 Å. Furthermore, the observed s-*trans*-lactam structures correspond to the two lowest predicted s-*trans*-lactam conformations. This is not the case for the s-*cis*-lactam derivatives, where the observed structures correspond to our predicted conformations *2* and *5*. However, our lowest-energy conformer, which has almost equal energy as conformer *2*, is found in the crystal structure of the closely related *N*-methyl derivative of thiocaprylolactam [16].

Table 2. *Conformations Generated for Caprylolactam* (**2**) *without s*-cis/*s*-trans-*Constraint for the Amide Using the Standard Protocol for Generic Shapes with a Randomization Repeat Factor* f = *4 Listed in Order of Ascending MOLOC Energies* (in kcal/mol). Conformations marked with * were missed in this run, but produced in other runs with *f* values up to 16, in which a s-*cis*- or s-*trans*-constraint was imposed. Resulting s-*cis*- and s-*trans*-amide geometries are marked by s-*c* and s-*t*, respectively. The X-ray structures of five caprylolactam derivatives (reference codes of the *Cambridge Structure Data Files* [8]) are included together with the rms deviations (in Å) between related ring atoms in these structures and corresponding calculated conformations. Ranges of rms deviations are given where more than one molecule is observed in the asymmetric unit.

Conformations	MOLOC-E	s-*c*/s-*t*	X-ray	rms deviation	Conformations	MOLOC-E	s-*c*/s-*t*
1	0.00	s-*c*	MTCAPL20	0.105	*21*	5.74	s-*t*
2	0.09	s-*c*	TMZCNO	0.064	*22*	5.76	s-*t*
3	0.15	s-*c*			*23*	5.76	s-*c*
4	0.66	s-*c*			*24*	5.93	s-*t*
5	0.76	s-*c*	CAPRLC	0.065–0.072	*25*	5.95	s-*t*
6	1.43	s-*c*			*26*	6.17	s-*t*
7	1.66	s-*c*			*27*	6.23	s-*c*
8	1.91	s-*c*			*28*	6.34	s-*c*
9	2.98	s-*c*			*29*	6.35	s-*t*
10	2.99	s-*c*			*30*	6.35	s-*c*
11	4.26	s-*c*			*31*	7.28	s-*t*
12	4.43	s-*c*			*32*	7.39	s-*t*
13	4.62	s-*t*	CAPRYL	0.064	*33*	7.40	s-*t*
14	4.85	s-*c*			*34*	7.42*	s-*t*
15	4.94*	s-*c*			*35*	7.93	s-*c*
16	5.00	s-*c*			*36*	8.24*	s-*c*
17	5.15	s-*c*			*37*	9.10*	s-*t*
18	5.16*	s-*c*			*38*	9.49*	s-*t*
19	5.48	s-*c*			*39*	9.85	s-*t*
20	5.53	s-*t*	AZCNOO	0.048–0.056			

In an attempt to further validate the predicted conformations, we examined the X-ray structures of related 9-membered ring systems containing at least one functional group that could be identified as a geometrical substitute for a s-*cis*- or s-*trans*-configured lactam unit. This is illustrated in *Fig. 2*, which shows the rigid-body superposition of the X-ray structure of a (*Z*)-cyclononene derivative, 12-epiobtusenyne (**4**) [20], with the two predicted s-*cis*-lactam conformations *2* and *3* of *Table 2*.

Since the ring conformation of *3* is dissymmetric, there are two distinct ways of superimposing a s-*cis*-peptide unit onto the (*Z*)-double bond. Both possibilities are satisfied by the low-energy lactam conformations *2* and *3* with a remarkable fit of the remaining ring atoms. In a similar manner, we have taken six other available X-ray structures of 9-membered ring derivatives with a (*Z*)-double-bond or annelated aromatic unit and found them to relate to most of the predicted s-*cis*-lactam conformations, lying within a low-energy band of 0–5 kcal/mol. Likewise, 9-membered ring derivatives containing a (*E*)-configurated double bond, annelated cyclopropane, or epoxide units can be used for comparisons with the predicted s-*trans*-lactam conformations. Interestingly, seven of the lowest eight s-*trans*-lactam conformations compare reasonably well with several of these experimental structures. One example is shown in *Fig. 2* for the superpositions of the X-ray structure of the cyclopropane derivative **5** [21], a 1,3-dehydro-5,10-secocholestenone, onto two calculated s-*trans*-lactam conformations (structures *22* and *29* of *Table 2*).

We have extended this analysis [7] by using more than fifty well resolved X-ray structures from the *Cambridge Structural Data Files*. While a detailed presentation of this work is beyond the scope of the present paper, we would like to summarize some of the salient results. Most of the calculated lactam conformations can be reasonably well superimposed (rms deviations between 0.04–0.3 Å) to one or several observed 9-membered rind derivatives. There are only a few, mostly high-energy conformations for which no experimental structural counterparts with a fit better than rms deviation of 0.3 Å have been identified.

This structure superposition analysis suggests an alternative conformation-generation method in which the atomic coordinates for the generic ring structures are taken from X-ray structures of appropriate ring compounds. Such an alternative method is particularly interesting for the conformation analysis of medium-size ring systems for which a considerable amount of structural information is available. Using this alternative approach, we have been able to reproduce all 33 predicted caprylolactam (**2**) conformations, but have not generated any additional ones. Thus, we feel that this structure collection is fairly complete, although we would like to point out that it does not represent the entire set of local energy minima on our conformational hypersurface. In fact, further exploratory runs of conformation generation for **2**, imposing s-*cis*- or s-*trans*-peptide constraints and using different randomization repeat factors, up to $f = 16$, uncovered a few additional s-*cis*- and s-*trans*-lactam structures, which are also included in *Table 2* and marked by asterisks. Since all these structures lie in the upper energy band of 5–10 kcal/mol, they do not affect the low-energy patterns for s-*cis*- and s-*trans*-lactam conformations. The fact that more s-*cis*- and s-*trans*-conformations are produced by applying, respectively, a s-*cis*- and s-*trans*-peptide constraint would seem to come as a natural consequence of the imposed bias. However, it is important to note that application of too stringent s-*cis*- or s-*trans*-constraints, as implemented in one of our earlier program versions, had just the opposite effect and that geometrical constraints, therefore, should be enforced by suitably soft functional terms, as discussed in the previous example.

Example 3: Macrocyclic Tetrapeptide **3**. The macrocyclic tetrapeptide **3** has recently been studied by NMR spectroscopy and force-field calculations [22]. We have re-examined this case by a conformational analysis on the related system, in which the 1-methylpropyl side chain of Ile is replaced by a Me group, as explained in *Chapt. 3*.

Restricting all peptide moieties to a s-*trans*-geometry, our algorithm generated a total of 80 conformations in an energy range of 10 kcal/mol *(Fig. 3)*. Such a large number of low-energy conformations may be surprising at first sight. However, the rigid thiophene and peptide moieties are connected by five flexible linkages, each consisting of two single bonds with low torsional barriers. The number of located conformational minima then corresponds, on average, to slightly more than two alternative rigid-unit orientations per flexible linkage, which would seem not unreasonable in retrospect.

The density of states is relatively small in the low-energy band of 0–1.5 kcal/mol, but is relatively high in the middle range of 1.5–6 kcal/mol. Towards still higher energy values, the density drops again. This pattern may not have been anticipated, but, in our experience, seems to be quite characteristic for larger unstrained ring systems.

Fortunately, only a limited number of conformations is predicted to be significantly populated at ambient temperature so that screening of the structures against experimental data is readily accomplished.

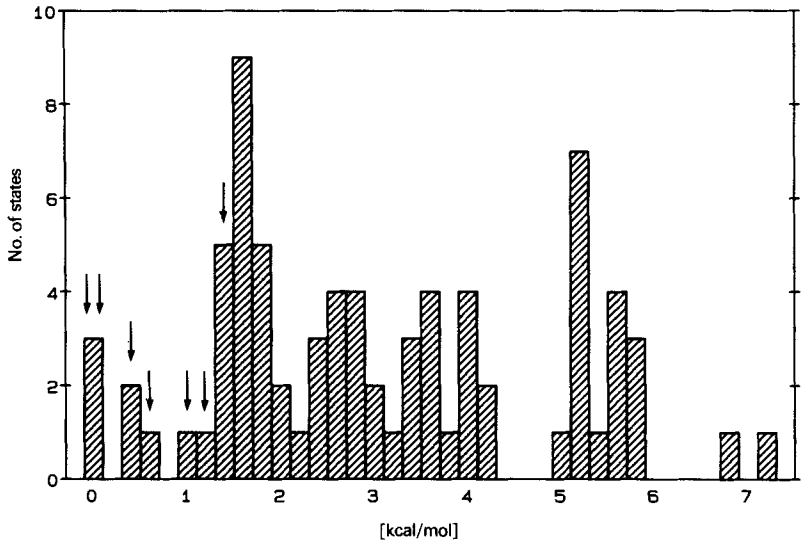


Fig. 3. *Histogram of the number of conformations in energy intervals of 0.2 kcal/mol* vs. *conformational energy, as obtained from shape-guided conformation generation for structure* **3**. Vertical arrows mark the energies of conformations that are consistent with the experimental data reported in [22].

Our analysis results in a picture which differs significantly from that obtained in [22]. There, it appeared that the lowest-lying state from molecular mechanics calculations (structure I in [22]) was unable to fulfill the experimental distance requirements, and that conformations complying with these constraints (structures II and III in [22]) had rather high energies (9 kcal/mol). However, structure I corresponds to our conformation *12*, some 1.5 kcal/mol above the minimum, while two of our three lowest-energy conformations, clustered in an energy range of 0.02 kcal/mol, correspond to structure II, differing only in the orientation of the thiophene ring. Furthermore, our conformations *5*, *6*, and *8* with relative energies of 0.4, 0.5, and 1.2 kcal/mol, respectively, also exhibit the essential features of structure II. Conformations *7* and *11* with relative energies of 1.0 and 1.3 kcal/mol, respectively, represent variants of structure III.

While we do not intend to overemphasize the ordering of the low-energy states predicted by our force field, we point out that we find several conformations, including two of minimal energy, that satisfy the available experimental data.

For this example, 484 generic shapes were produced as start structures. Using our standard protocol with a randomization repeat factor $f = 4$, the number of trial structures was augmented to over 2100. The resulting set of 80 conformations, thus, represents 3.8% of the initial trial set, which again reflects on our previous discussion of completeness *vs.* efficiency. Since each energy minimization took somewhat less than 30 s CPU time on a *VAX-8700*, the whole conformational analysis was completed within 17.5 h.

7. Extension to Loops. – Our algorithm for the generation of generic ring shapes has been extended to handle loops protruding from an arbitrary substructure. For loops, the initial trial structures are gain generated according to *Eqns. 2*. However, the choice of spatial frequencies and phases, as given by *Eqns. 2*, has to be adjusted to the boundary conditions imposed by the fixed ends of the loop.

The energy-minimization procedure is also modified by imposing additional constraints during the initial phase in order to fix the loop-spanning subunit(s). Optionally, these artificial constraints can be gradually removed so that the final minimization steps will involve the whole molecular structure.

There are many applications for such a loop algorithm, some of the more prominent ones being the study of conformational induction by loop-spanning rigid templates and the modeling of surface loops on proteins.

8. Conclusions. – The principle of shape-guided conformation generation provides a fairly efficient tool for finding the low-energy conformations of macrocyclic systems on a automated basis. Low-lying conformations are found quite reliably. While we have no direct estimate of completeness, numerical experimentations on a number of test cases indicate that coverage of conformation space is quite extensive. The computational effort of about ten energy minimizations per unique final structure is quite acceptable. According to our proposed protocol, the number of generic shapes increases as the square of $N(\ln N)$. However, due to the randomization procedure, the number of actually generated initial and unique final structures may grow faster depending on the nature of the ring system under investigation.

REFERENCES

[1] G.M. Smith, Program 510 of the Quantum Chemistry Program Exchange, Dept. of Chemistry, Indiana University.

[2] M. Dygent, N. Go, H.A. Scheraga, *Macromolecules* **1975,** *8,* 750.

[3] D.P. Dolata, A.R. Leach, K. Prout, *J. Comput. Aided Mol. Des.* **1987,** *1,* 73.

[4] J.W.H.M. Uiterwijk, S. Harkema, B.W. van de Waal, F. Goebel, H.T.M. Nibbeling, *J. Chem. Soc., Perkin Trans. 2* **1983,** 1843.

[5] A.T. Hagler, *Peptides* **1985,** *7,* 213 and ref. cit. therein.

[6] J.E. Kilpatrick, K.S. Pitzer, R. Spitzer, *J. Am. Chem. Soc.* **1947,** *69,* 2483; D. Cremer, J.A. Pople, *ibid.* **1975,** *97,* 1354.

[7] P.R. Gerber, K. Mueller, unpublished results.

[8] F.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelink-Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers, D.G. Watson, *Acta Crystallogr., Sect. B* **1979,** *35,* 2331.

[9] A. Furusaki, S. Katsuragi, K. Suehiro, T. Matsumoto, *Bull. Chem. Soc. Jpn.* **1985,** *58,* 803.

[10] Z. Karimov, G.N. Tishchenko, *Kristallografiya* **1979,** *24,* 778.

[11] Yu. A. Simonov, N.F. Krasnova, A.A. Dvorkin, V.V. Yakshin, V.M. Abashkin, B.N. Laskorin, *Dokl. Akad. Nauk SSSR* **1983,** *272,* 1129.

[12] M.J.D. Powell, *Mathematical Programming* **1977,** *12,* 241.

[13] W. Kabsch, *Acta Crystallogr., Sect A* **1982,** *32,* 922; A.D. McLachlan, *ibid.* **1972,** *28,* 656; *ibid.* **1982,** *38,* 871.

[14] O. Ermer, 'Aspekte von Kraftfeldrechnungen', Wolfgang Bauer, Munich, 1981.

[15] S.R. Niketic, K. Rasmussen, 'The Consistent Force Field', Springer, Berlin, 1977, and ref. cit. therein.

[16] J.L. Flippen, *Acta Crystallogr., Sect. B* **1972,** *28,* 3618.

[17] D. van der Helm, S.E. Ealick, *Acta Crystallogr., Sect. B* **1979,** *35,* 1255.

[18] F.K. Winkler, J.D. Dunitz, *Acta Crystallogr., Sect. B* **1975,** *31,* 276, 278.

[19] M.B. Hossain, J.R. Baker, D. van der Helm, *Acta Crystallogr., Sect. B* **1981,** *37,* 575.

[20] Y. Gopichand, F.J. Schmitz, J. Shelly, A. Rahman, D. van der Helm, *J. Org. Chem.* **1981,** *46,* 5192.

[21] M.L. Mihailovic, L. Lorenc, M. Dabovic, I. Juranic, E. Wenkert, J.-M. Bernassau, M.S. Raju, A.T. McPhail, R.W. Miller, *Tetrahedron Lett.* **1979,** 4917.

[22] M. Feigel, G. Lugert, C. Heichert, *Liebigs Ann. Chem.* **1987,** 367.