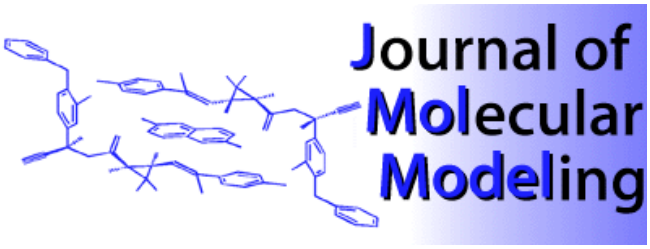
J. Mol. Model. **1999**, *5*, 46 – 62



FULL PAPER

Conformational Analysis of Tetracycline using Molecular Mechanical and Semiempirical MO-Calculations

Harald Lanig[1], Maik Gottschalk[1], Siegfried Schneider[2], and Timothy Clark[1]

[1]Computer-Chemie-Centrum, Institut für Organische Chemie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany. Tel. +49-(0)9131 85 22948; Fax +49-(0)9131 85 26565. E-mail: clark@organik.uni-erlangen.de

[2]Institut für Physikalische Chemie I, Friedrich-Alexander-Universität Erlangen-Nürnberg, Egerlandstrasse 3, D-91058 Erlangen, Germany

Received: 18 December 1998/ Accepted: 25 February 1999/ Published: 26 March 1999

Abstract A combination of gas phase molecular dynamics and systematic conformational searches with semiempirical MO-theory has been used to identify and characterise the low energy conformations of tetracycline in its neutral non-polar and zwitterionic forms and as the deprotonated zwitteranion. The conformations found have been characterised by cluster- and principal components analyses as belonging to three main groups. These are very similar for the nonpolar and zwitterionic neutral forms of tetracycline, but very different for the anion. The likely consequences of the conformational flexibilty for metal ion complexation by tetracycline are discussed.

Keywords Molecular Dynamics simulations, Semiempirical MO-calculations, Conformational analysis

Introduction

It is generally assumed that the biological activity of a drug molecule depends on a single unique conformation that is only one among all possible low-energy conformations. However, this bioactive conformation is not necessarily identical with the lowest-energy conformation in solution, but must lie in an energetic range that allows it to become the most stable conformation when docked in the receptor. To understand the relationships between structure and biological activity, and in order to isolate solvent effects on the stabilities of individual conformations, it is useful to identify all possible low-energy conformations of a drug in the gas phase. This process may be divided into two parts: an algorithm to generate many different starting structures and an energy minimisation procedure to calculate the next local minimum. For the first part, many different approaches such as systematic searches [1], random search methods [2], model building approaches [3], nearest neighbour analysis [4] or genetic algorithms [5] currently exist. These algorithms all have in common that they produce non-equilibrium structures. To obtain a statistically significant sampling, a large number of different structures must be generated. If every starting structure is subjected to energy minimisation, this second part of the conformational search strategy is rate-

Correspondence to: T. Clark

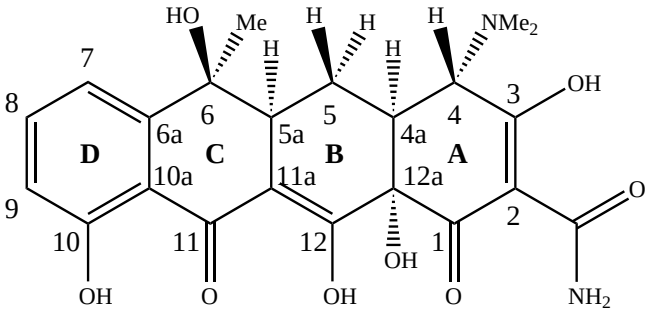


Figure 1 *Systematic atom numbering and ring labels of tetracycline used throughout the text*

determining for the entire analysis. For this reason, molecular mechanics is normally used, although quantum mechanical techniques would in many cases be preferable.

Molecular dynamics (and Monte Carlo simulations to a certain extent) can also be considered as a part of a conformational search strategy, because they are able to generate an ensemble of non-equilibrium states. One advantage of such simulation methods is that the system is able to overcome energy barriers that separate different parts of the energy hypersurface. The simulation is performed at unusually high temperatures with excess kinetic energy, which allows barrier-crossing and prevents the system from remaining trapped in a localised region of the potential surface. As well as selecting structures at regular time intervals during the molecular dynamics simulation and performing energy minimisation, minimum-energy conformations can be obtained by simulated annealing [6]. This technique is, however, also very time consuming and leads to a limited number of stable minima.

Conformations generated by various search algorithms can be very similar or even identical. If one is interested in structures (coordinates), molecular dynamics trajectories are normally analysed in a conventional way by e.g. averaging the coordinates over a period of time or extracting coordinates after regular time intervals as snapshots.

In this investigation, our aim is to explore the overall conformational space of the drug tetracycline (Tc), not just to find the global minimum. We shall use a combination of classical molecular dynamics and systematic semiempirical MO conformational searches with some novel analysis techniques.

Tc was chosen as test molecule for various reasons. Firstly, tetracycline derivatives are widely used as drugs and consequently numerous studies have been performed in order to understand the complex behaviour of this family of compounds [7–9]. It has been established that Tc's can exist in different conformations depending on the substituents, metal chelation and the solvent used [10, 11]. In neutral to acidic solutions Tc adopts a "twisted conformation"[12], where the dimethylamino group lies above the BCD ring system (for nomenclature and numbering see Figure 1). In basic or non-aqueous solutions the molecule adopts the so-called "extended

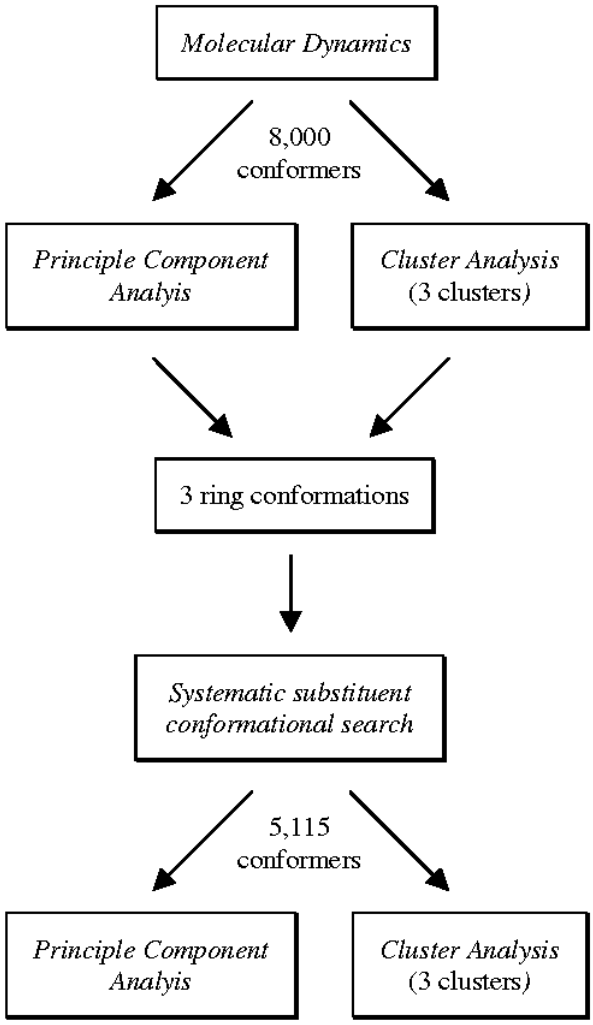


Figure 2 *Flow-chart showing the computational protocol for the neutral, zwitterionic and anionic tetracycline*

conformation", where the dimethylamino group lies below the BCD ring system [12]. This sensitivity of the molecular geometry to environmental parameters indicates that the energy hypersurface must contain several local minima with relatively low barriers between them. Secondly, it has been proposed that the equilibrium between the various conformers plays an important role in the pharmacokinetic properties of Tc's [13–15]. It is well established that Tc's are incorporated into proteins preferentially as complexes with divalent metal ions with the complexation site. The induced conformational change is still a matter of debate [7–9]. Thirdly, Tc's are molecules of intermediate size. This means that on one hand, there is enough flexibility in the molecular frame to make the determination of the conformational space a challenging problem. On the other hand, it is small enough to allow the use of quantum-chemical techniques.

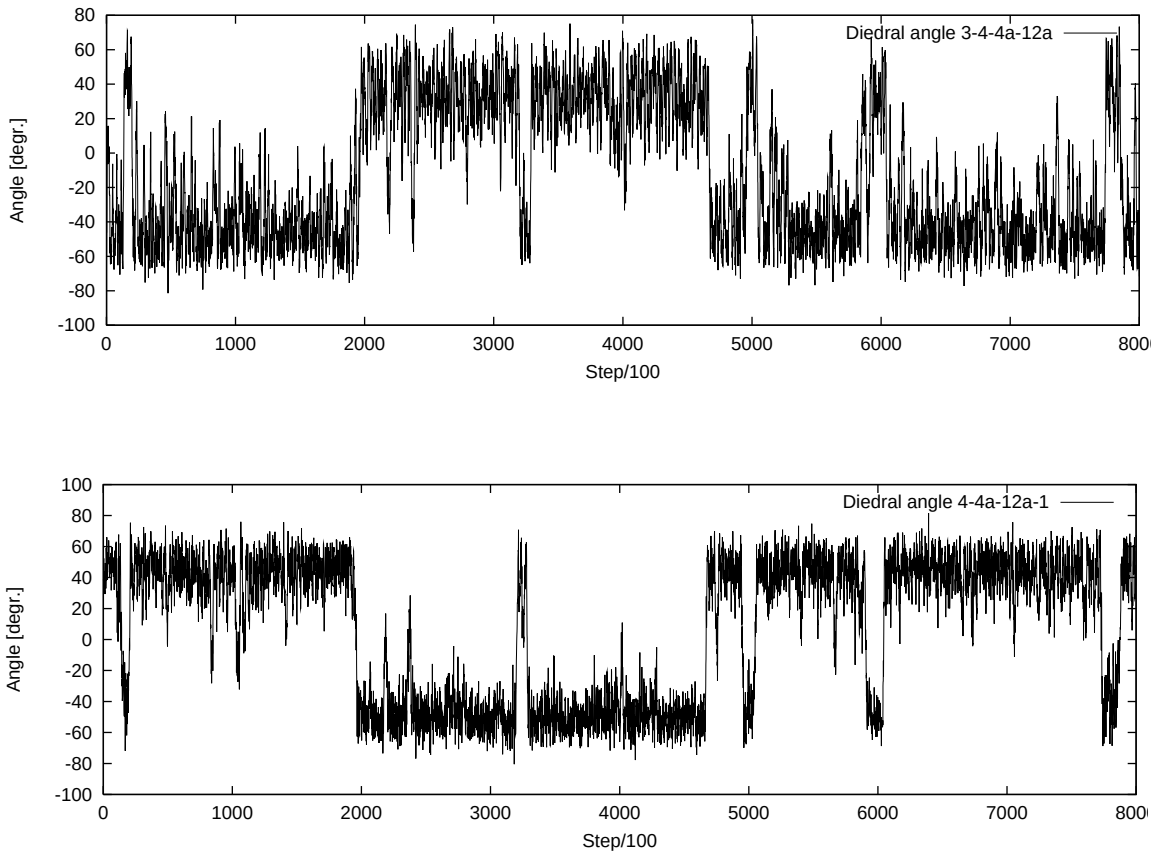


Figure 3 *Time evolution of the dihedral angles 3-4-4a-12a (1) and 4-4a-12a-1 (2) as representative examples for candidates describing stable ring flips*

Computational methods and data preparation

Molecular dynamics simulations

Constant energy molecular dynamics simulations were carried out with AMBER 5.0 [16] in the gas phase without any explicit water molecules or periodic boundary conditions. The total simulation time was 800 ps (400 ps equilibration and 400 ps data aquisition) with an integration time step of 0.5 fs. The simulation temperature of 1200 K was achieved by coupling to a temperature bath [17] starting from 0 K with a coupling constant of 2.0 ps. The translational and rotational motions about the center of mass were removed. The atom-centred charges on the molecule were derived with the VESPA approach [18] on an AM1-optimised [19] structure of tetracycline using Vamp 6.5 [20]. A dielectric constant ε of 4 was used in the simulations to prevent destructive distortions caused by the dominant electrostatic energy contribution to the total energy during the simulation. No cut-offs were used for non-bonded interactions.

During the sampling period, every 100 time steps (0.05 ps) the current energy contributions, temperature and geometry were saved, producing an overall data set of 8000 conformations. An important aspect of the statistical analysis is the representation of the conformers in the data set. This was achieved by selecting 31 torsional angles which describe (with redundancy) all ring conformations of tetracycline. The angles were calculated using the CARNAL tool distributed with the AMBER package. The orientation of the side chains was not considered, because at this stage of the analysis the main interest is focused on the flexibility of the ring system. The complete data set consists of an array with 31 dihedal angles for 8000 molecular conformations.

Semiempirical calculations

In order to determine the conformational space of Tc, we performed several semiempirical torsion calculations, in which all dihedral angles of the polar ring substituents were systematically varied. The corresponding geometries were optimised semiempirically using the AM1-Hamiltonian [19]

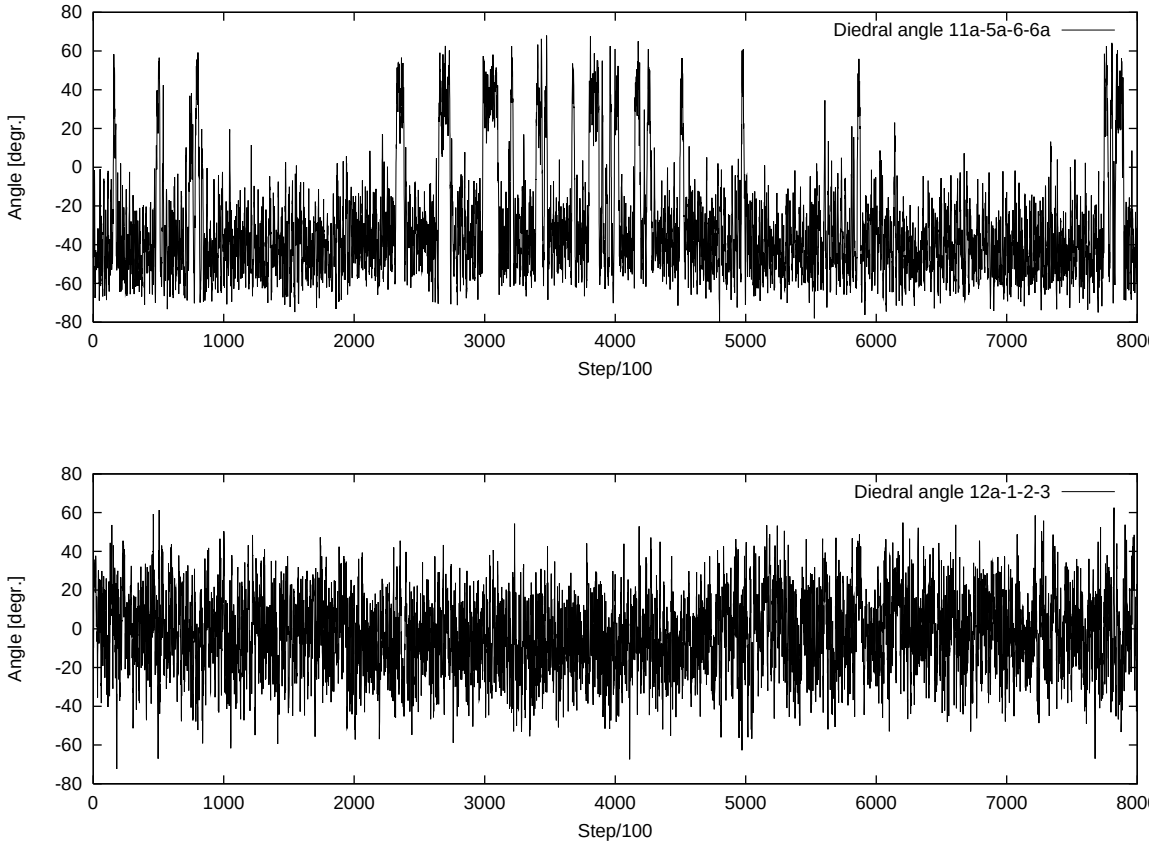


Figure 4 *Time evolution of the dihedral angles 11a-5a-6-6a and 12a-1-2-3. The upper series shows the tendency of the molecule to switch to another conformation, which is not stable at 1200 K. The lower picture visualises that the selected angle is not able to describe different ring conformations*

with a default convergence criterion of a gradient norm of 0.4 kcal·Å^{-1}. First, all of the AMBER molecular-dynamically derived cluster centres were semiempirically preoptimised with VAMP [20]. Then systematic conformational searches with full geometry optimisation were performed, where all dihedral angles between polar substituents and the Tc ring system were rotated applying fixed angle increments (Table 1). The increments were deduced from the energetically favourable orientations of the functional groups. The steric demand of the carboxamide- and dimethylamino-groups (Figure 1) led us to expect considerable influence on the ring conformations. Therefore, these groups were rotated in 90° steps.

The pK_a-values [21] show that Tc contains acidic and basic substituents. In a physiological environment (aqueous solution, pH = 7.0) Tc consists of two different species: The shift of a proton from 3-OH (pK_a = 3.1) to the 4-NMe_2-group (pK_a = 9.7) leads to a zwitterion. Additionally, the hydroxy-function 12-OH (pK_a = 7.7) may be deprotonated to give a negatively charged molecule. In order to determine the conformational space of Tc, these zwitterionic and anionic species are of interest. Starting from the neutral and uncharged AMBER-conformations, both zwitterionic species have been created by moving the hydrogen of the hydroxy-proton at carbon C_3 to the dimethylamino-nitrogen and/or by eliminating the proton of the hydroxy-function at C_{12}. All semiempirical torsion calculations and succeeding similarity and statistical examinations by principal components analysis and cluster-analytical methods, as shown in Figure 2, were identical to those used to determine the conformational space of the uncharged Tc molecule.

Processing of the simulation data

Cluster analysis

Trajectory data were processed using a cluster analysis technique [22]. This can be considered as an extension of the well known nearest neighbours analysis, which measures the pairwise distances between compounds in multidimensional space. To classify samples into clusters of similar members, a distance matrix was used. Standardisation of the input data (in this investigation based on the mean and the standard deviation of the data set) is important to avoid skewing of the

Table 1 *Functional groups attached to the tetracycline ring system and corresponding dihedral angle increments*

Functional group	incr. [deg]
C_2-$CONH_2$	90
C_3-OH	180
C_4-$N(CH_3)_2$	90
C_6-OH	120
C_{10}-OH	180
C_{12}-OH	180
C_{12a}-OH	120

resulting distances. We used an agglomeration method with Euclidian distances [23]. This simplification is necessary because of the dimensionality (31 dihedral angles) and size (8000 conformations) of the data set. All statistical analyses were performed using the program TSAR 3.1 [24].

Cluster analysis results are best examined by the construction of a dendrogram [25] showing the examined clusters of the conformers vs. their cluster-analytical diversity.

To subdivide the dendrogram into clusters, a ruler was defined as needed to extract the resulting arrangement of data into the array of dihedral angles for further examination in conjunction with a principal components analysis (see below). An important characteristic of the final clusters are their centres. Because no molecule representing exactly a centre normally exists in the data set, the structure closest to each centre was identified and is also exported to the data sheet.

Principle components analysis

The general aim of reducing the dimensionality of a dataset is to obtain a smaller, more manageable number of variables (here dihedral angles) while keeping as much of the original information as possible. This goal can be achieved in several ways. The generation of correlation matrices helps to find highly correlated pairs of variables (coefficient > 0.9) and thus redundancy in the data set. Another technique described in the literature is non-linear mapping [26], which reduces the dimensions of a data set by defining and minimising an error function followed by a conventional optimisation.

The method applied in this paper is principle components analysis (PCA) [28]. Each principal component (PC) is a linear combination of the original variables. The first PC explains the greatest variance within the data set, the second PC the next greatest variance etc. In this study, most of the variance in the data set is explained using only the first three PC's which can by visualised by a three dimensional plot, showing the projection of the data points on to the principle component vector, and coloured e.g. by the results of the cluster analysis, as we will show later.

Similarity calculations

For the description of the conformational space of Tc, the lowest 10 kcal·mol^{-1} range of the Boltzmann-energy distribution is of special interest. For this reason, the following examinations were restricted to this energy range. In order to eliminate energetically unfavourable conformations, all conformers of the torsion calculations lying within the lowest

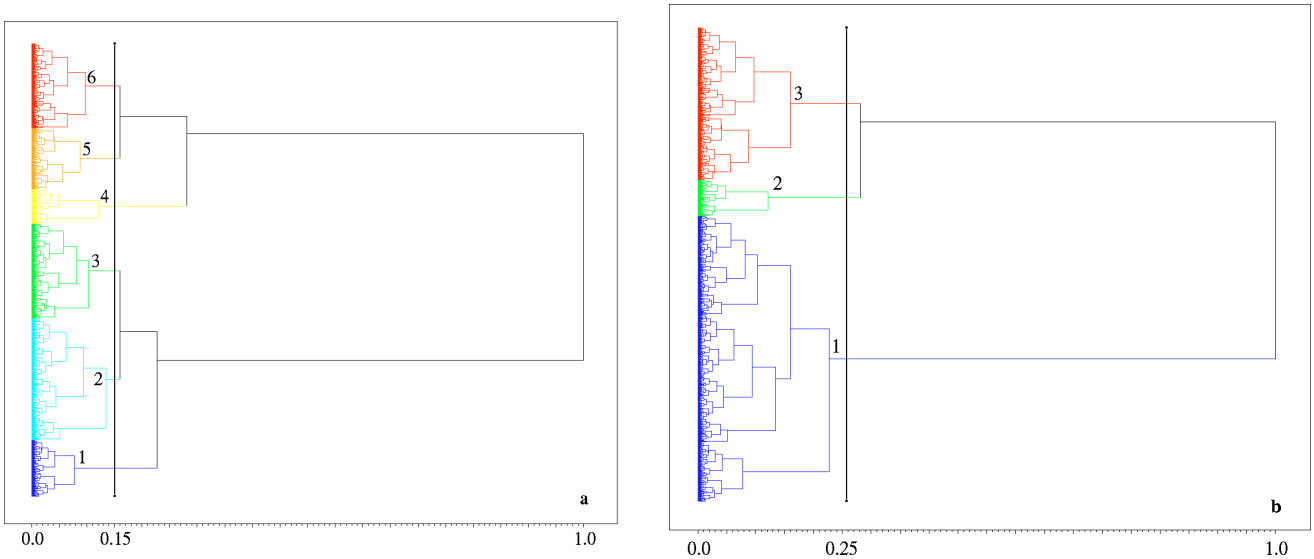


Figure 5 *Cluster analysis performed on the 14 selected dihedral angles describing ring conformations in regular time intervals during the MD-simulation. The resulting dendrogram is coloured according the division into six (a) and three clusters (b) each separated by the vertical ruler*

Table 2 *Selected dihedral angles of the data set showing the behaviour necessary for statistical analysis (see text). Lower and upper values were calculated by averaging over the respectively parts of the whole molecular dynamics trajectory. Amplitude is the difference between upper and lower. The angles in the upper block show the same principle time evolution as those in the lower block, only the sign is reversed. All data is in degrees*

Nr	Dihedral angle	Lower	Upper	Ampl.
1	3-4-4a-12a	-51	39	90
3	4a-12a-1-2	-21	39	60
4	3-4-4a-5	74	164	90
5	4-4a-5-5a	174	270	96
7	4a-12a-12-11a	-19	41	60
10	12a-4a-5-5a	-61	38	99
11	1-12a-12-11a	98	166	68
13	2-1-12a-12	210	282	72
2	4-4a-12a-1	-53	49	102
6	4-4a-12a-12	71	174	103
8	4a-5-5a-6	141	184	43
9	4a-5-5a-11a	5	53	48
12	1-12a-4a-5	183	284	101
14	5-4a-12a-12	-52	52	104

10 kcal·mol^{-1} range were refined to a gradient norm of 0.01 kcal·Å^{-1}. To eliminate identical or similar conformations, automated similarity calculations were performed with the program ASP [29]. All semiempirically optimised individuals were rigidly overlayed including all atoms, in order to generate the similarity matrix. The Carbo-index R_{AB} [30], based on the electron density of two overlayed molecules, was used to measure their similarity. The limit for eliminating conformers was R_{AB} = 0.990 (Eqn. 1),

$$R_{AB} = \frac{\int \rho_A \rho_B dV}{\sqrt{\int \rho_A^2 dV \int \rho_B^2 dV}} \quad (1)$$

where ρ_A = electron density of molecule A and ρ_B = electron density of B.

The remaining conformers were shown to be local minima by calculating their normal vibrations. Conformers with one

Table 3 *Selected dihedral angles describing the ring conformations of the molecules next to the six located cluster centers after geometry optimisation with AMBER. The reduction from six to three basic conformations due to the optimisation is indicated by the blank lines. All data is in degrees*

Cluster (id)	3-4-4a-12a	4a-5-5a-6	1-12a-4a-5
1 (3616)	42.60	133.75	177.75
2 (4426)	40.45	133.74	177.95
3 (3422)	49.33	166.84	-177.92
4 (4862)	-54.42	172.14	-70.71
5 (1804)	-51.57	171.39	-69.97
6 (6404)	-51.62	171.42	-69.98

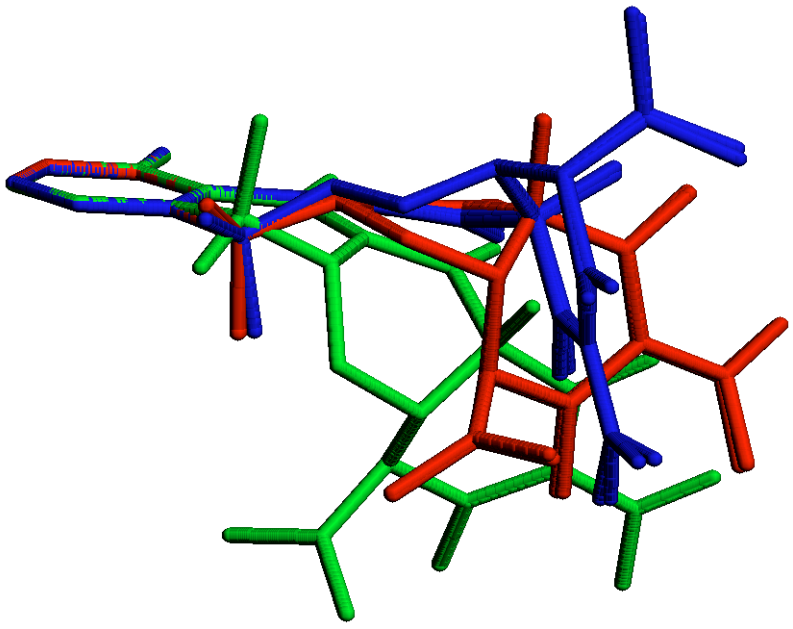


Figure 6 *Molecules identified in Figure 5a, representing the cluster centers when dividing the dendrogram into six clusters, after energy minimisation. For better visualisation, hydrogen atoms are omitted, the heavy atoms of the aromatic ring D are overlayed by a RMS-fit procedure. The initially six hot conformations reduce to three local minima on the potential energy surface, which are colour-coded according to Figure 5b. Note the large spacial area covered by the 4-NMe_2 and 2-$CONH_2$ substituents*

or more imaginary frequency were omitted, because they do not correspond to local minima on the potential energy surface. Further data reduction was then performed by statistical methods.

Results

Molecular dynamics trajectories

Dihedral angle time profiles and variable selection The molecular structure of Tc in the neutral form is shown in Figure 1. Atom numbers and ring labels (A - D) are used throughout the text. At neutral pH, the hydroxy group at C_3 is deprotonated and the adjacent 4-NMe_2 group protonated to form an intramolecular zwitterion [31]. The 12-OH group may also be deprotonated (pK_a 7.5), especially when complexing metal cations are present [21]. Nevertheless, the neutral forms of all groups were used for the molecular dynamics simulations.

Time dependent data for 31 dihedral angles 1-2-3-4, 2-3-4-4a, 3-4-4a-12a, 4-4a-12a-1, 4a-12a-1-2, 12a-1-2-3, 3-4-4a-5, 4-4a-5-5a, 4-4a-12a-12, 4a-12a-12-11a, 4a-5-5a-6, 4a-5-5a-11a, 12a-12-11a-11, 12a-12-11a-5a, 12a-4a-5-5a, 1-12a-12-11a, 1-12a-4a-5, 2-1-12a-12, 5-4a-12a-12, 5-5a-6-6a, 5-5a-11a-11, 5-5a-11a-12, 12-11a-11-10a, 12-11a-5a-6, 5a-11a-11-10a, 5a-6-6a-7, 5a-6-6a-10a, 11a-5a-6-6a, 11a-11-10a-10, 11a-11-10a-6a, and 6-5a-11a-11 were extracted from the molecular dynamics trajectory file. The system is overdetermined using all 31 dihedral angles in the following statistical analyses. Therefore, the calculation of a correlation matrix (not shown) of all variables is included within the subsequent Principal Components Analysis. This matrix clearly indicates the variable pairs with high correlation coefficients (0.0 < 0.70 low, 0.70 < 0.90 medium, 0.90 < 1.0 high).

Table 4 *Communality table for the first three out of 14 principal components. The data indicates the cumulative amount of variance of each dihedral angle that is explained by each PC. Numbers in bold-face experience the largest changes when adding a new component*

Dihedral angle	PC 1	PC 2	PC 3
3-4-4a-12a	0.7373	**0.7443**	0.9720
4-4a-12a-1	0.9604	0.9733	0.9735
4a-12a-1-2	0.7143	**0.7333**	0.9715
3-4-4a-5	0.7511	**0.7577**	0.9792
4-4a-5-5a	0.8931	0.9654	0.9654
4-4a-12a-12	0.9660	0.9746	0.9746
4a-12a-12-11a	0.7410	0.8916	0.8917
4a-5-5a-6	**0.3824**	0.9193	0.9199
4a-5-5a-11a	**0.3991**	0.9463	0.9468
12a-4a-5-5a	0.9056	0.9733	0.9733
1-12a-12-11a	**0.6944**	0.8862	0.8862
1-12a-4a-5	0.9686	0.9787	0.9793
2-1-12a-12	0.7309	**0.7406**	0.9769
5-4a-12a-12	0.9699	0.9763	0.9763

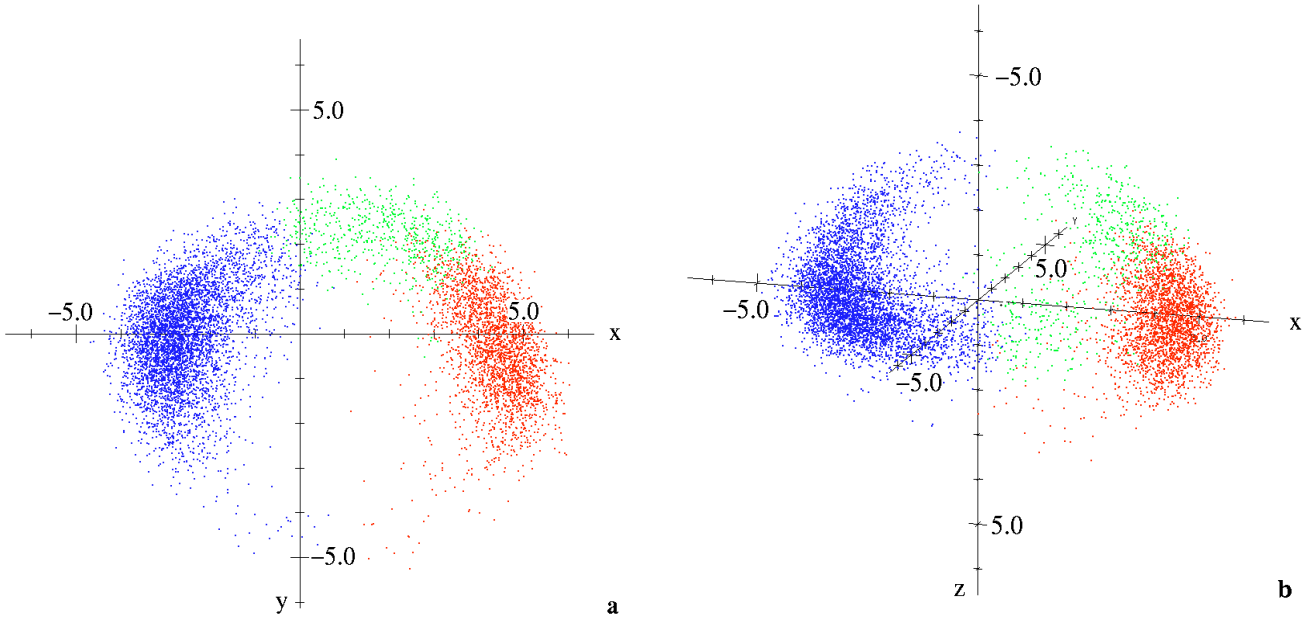


Figure 7 *Graphical representation of the first three principal components representing 96% of the diversity of 14 dihedral angles. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 5b)*

Table 5 *Results of the principle components analyses: Total variance and Eigenvalues of the neutral, zwitterionic and anionic Tc conformations*

Total variance and eigenvalues								
Neutral								
Total Variance [%]	31.7	54.1	72.2	79.0	85.3	89.4	93.2	95.0
Eigenvalue	6.7	4.7	3.8	1.4	1.3	0.9	0.8	0.4
Zwitterion neutral								
Total Variance [%]	37.4	57.2	67.8	76.1	82.1	87.2	90.5	93.8
Eigenvalue	7.8	4.2	2.2	1.7	1.3	1.1	0.7	0.7
Zwitterion anionic								
Total Variance [%]	49.2	71.0	82.7	91.8	95.7	98.2	99.4	99.9
Eigenvalue	5.9	2.6	1.4	1.1	0.5	0.3	0.1	0.1

Only two limiting conformations for tetracycline have been described in the literature [31]. These can be interconverted by rotation about the C_{4a}–C_{12a} bond. This interconversion is represented by the dihedral angles 4-4a-12a-1, 4-4a-12a-12, 1-12a-4a-5, and 5-4a-12a-12 in the data set. In principle, it is not necessary to include all four angles into the raw data set, because they describe the same dihedral interconversion. However, the high simulation temperature might allow molecular conformations that do not maintain the ideal geometry for sp^2 (trigonal-planar) and sp^3 (tetrahedral) carbon atoms. Therefore, it is useful to include more than one definition of the "same" dihedral angle in the initial table.

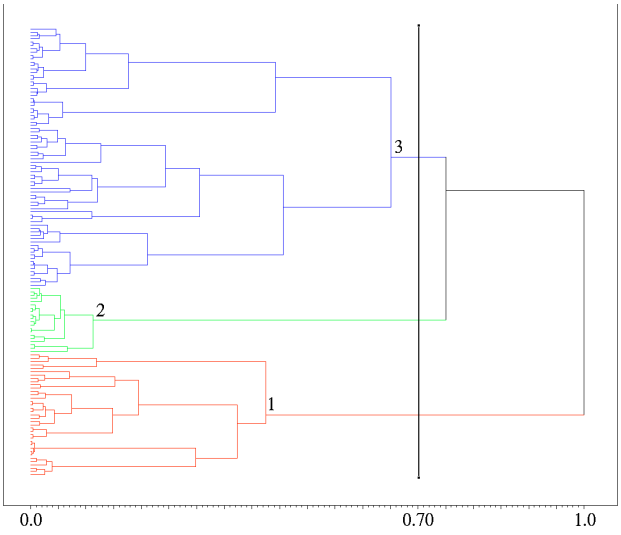


Figure 8 *Cluster-analytical results for the semiempirical calculations performed on the 21 selected dihedral angles describing stable ring conformations of the neutral Tc molecule. The resulting dendrogram is coloured according the division into three clusters by the vertical ruler*

If a plot of a given variable against time shows the establishment of a certain value for a longer time period, followed by a rapid change to another stationary value that is also stable for a longer time, this variable can be used to describe different stable conformations of the molecule. Visual inspection of all variables revealed that 14 angles 3-4-4a-12a (1), 4-4a-12a-1 (2), 4a-12a-1-2 (3), 3-4-4a-5 (4), 4-4a-5-5a (5), 4-4a-12a-12 (6), 4a-12a-12-11a (7), 4a-5-5a-6 (8), 4a-5-5a-11a (9), 12a-4a-5-5a (10), 1-12a-12-11a (11), 1-12a-4a-5 (12), 2-1-12a-12 (13), and 5-4a-12a-12 (14) show the desired behaviour.

The time series of the angles 1, 3, 4, 5, 7, 10, 11, and 13 are nearly identical, suggesting that the motions of the atoms involved are highly synchronised. Data averaging gives the values shown in Table 2, which define two principal states every dihedral angle can switch to. The first eight angles describe a synchronous flip around the central bonds 4-4a, 12a-1,

Table 6 *Communality table of the first three of 21 principle components. Bold-face communalities correspond to dihedral angles with maximum contribution to each principle component*

Dihedral angle	**Communalities**		
	PC1	**PC2**	**PC3**
4-4a-12a-1	**0.801**	0.869	0.883
4-4a-12a-12	**0.695**	0.702	0.847
12a-12-11a-11	**0.849**	0.891	0.960
1-12a-12-11a	**0.962**	0.972	0.972
4a-5-5a-11a	0.001	**0.577**	0.680
5a-6-6a-10a	0.103	**0.754**	0.764
11a-5a-6-6a	0.001	**0.850**	0.940
3-4-4a-5	0.021	0.235	**0.810**
12-11a-11-10a	0.085	0.291	**0.837**
12-11a-5a-6	0.035	0.069	**0.708**

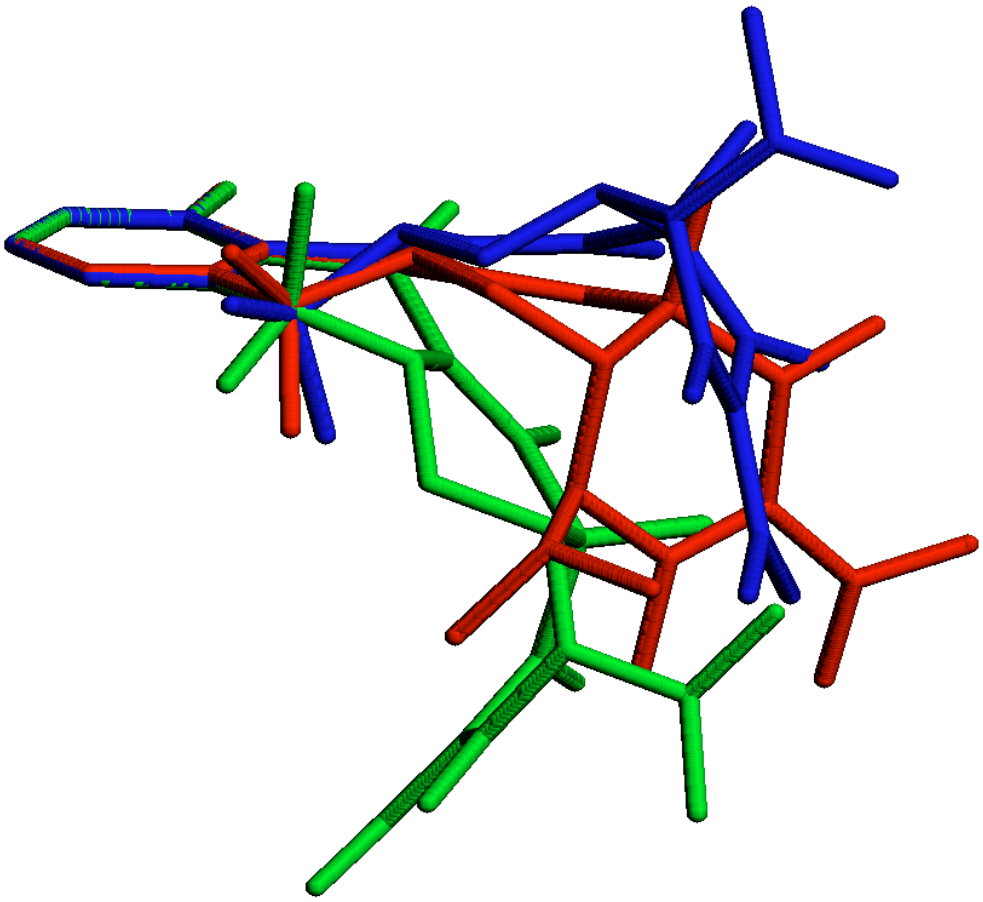


Figure 9 *Overlay of the ring systems of the three neutral Tc conformers found by principle components and cluster analytical methods of the semiempirically examined conformers. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 8)*

12a-12, 4a-5, and 1-12a, determining the conformations of the rings A and B. It is not surprising that the amplitudes of the transitions 1, 4, 5, and 10 lie in a comparable range, because no sp^2-carbon is involved in the central atoms of the torsion angle. In the case of 3, 7, 11, and 13, an sp^2-carbon is involved in the central bond, which reduces the flexibility considerably. The dihedral angles 2, 6, 8, 9, 12, and 14 show the same basic behaviour, only the sign of the amplitude is reversed. The amplitude of the angles 2, 6, 12, and 14 is even larger than those described above for the angles with no sp^2-carbon in the central bond. The small amplitude of the dihedral angles 8 and 9 is a result of the inclusion of ring C with decreased flexibility of the system in this area. Figure 3 shows as an example the time evolution of the dihedral angles 3-4-4a-12a (1) and 4-4a-12a-1 (2). The remaining dihedral angles show a completely different picture. Here, only one equilibrium value for every degree of freedom is established. The fluctuations around these values are considerably larger, and only one conformation is stable over the sampling period. Figure 4 shows the time evolution of the angles 11a-5a-6-6a and 12a-1-2-3 as examples. A special behaviour can be observed for the angles with the central bonds 5a-5, and 5a-6 (see Figure 4a). The kinetic energy is high enough that the molecule tries to switch to another conformation. In principle this should be possible, because only sp^3-atoms are involved, but the alternative conformations are not stable, at least at 1200 K. We conclude that the connection between the rings B and C is less flexible than between the rings B and A, despite the fact that more than one conformation should exist in principle. The behaviour of the angles with the central bonds 11a-5a, 12a-1, 2-3, and 3-4 shows that different conformations are not possible and therefore no additional flexibility can be observed. Because of the aromaticity of ring D no variables describing deviations from the ring planarity were included.

Cluster analysis According to the results presented above, the values of the 14 dihedral angles describing different molecular conformations in regular time intervals were subjected to a cluster analysis. This statistical technique reveals relationships between the variables defining the 8000 individual conformations and clusters them. The results are visualised in the form of a dendrogram as shown in Figure 5. The relative horizontal position (concentrated within the first 30% of the diagram) of the vertical lines joining two clusters shows the tight geometrical relationship between the individual conformations. By setting a (in this case) vertical ruler now defining the number of clusters according to the actual ruler position is possible. For each conformation, an identifier can be copied from the dendrogram into the table of dihedral angles. Additionally, the molecule closest to each cluster centre may be identified and marked in the spreadsheet.

Table 7 *Table of the three most stable ring conformers of the neutral Tc molecule in energetically descending order. Each conformer can be described by a minimum of three selected dihedral angles given with lower and upper bounds each*

Ring conformer	Energy [kcal·mol^{-1}]	Dihedral angles [°] 3-4-4a-12a	4a-5-5a-6	12-11a-11-10a
1	-263.9	32.4	121.7	-177.6
2	-266.3	-59.0	-176.0	-86.8
3	-268.5	-57.5	172.5	-85.9
Lower bound		-59.0	-176.0	-177.6
Upper bound		32.4	172.5	-85.9

The critical step of the ongoing analysis is the selection of clusters by positioning the ruler. If the number of clusters is to small, diversity information is lost. A large number, on the other hand, may produce redundant information. Figure 5a shows for example the definition of six clusters colour-coded for easy comparison. The conformations (1200 K) next to each cluster centre were identified and subjected to AMBER energy minimisation to remove excess strain energy.

Table 3 shows the resulting values of three representative dihedral angles 3-4-4a-12a, 4a-5-5a-6, and 1-12a-4a-5. As indicated by the blank lines in the table, the six initial clusters reduce to only three clusters on geometry-optimisation of the 1200 K conformations. The reason for this behaviour is visible in the dendrogram by comparing the intersection of the vertical lines joining two clusters with the x-axis. The smaller the value (between 0.0 and 1.0), the more similar the two individual molecules or clusters joined at this point.

Clusters 5, 6 and clusters 2, 3 join at a coefficient of 0.16. The next most similar cluster 1 is able to join 2+3 with a coefficient of 0.23, explaining why the geometry optimisation leads to the same basic conformation for the centres 1, 2, and 3. The remaining cluster 4 finally joins cluster 5+6 at 0.28, explaining why the molecule representing the centre of cluster 3 optimises to different individual conformation. Based on this information, defining a vertical ruler at 0.24 gives three clusters separated in Table 3 by lines. Figure 5b shows the final clusters with the corresponding colour coding. It is important to mention that the centres of these three clusters represent the same conformations as those obtained by geometry optimisation of the six initial cluster centres. This means that each of the three clusters defines its own local minimum on the potential energy surface.

In Figure 6, the six molecules were superimposed by a RMS fit of the heavy atoms of the aromatic ring D. Hydro-

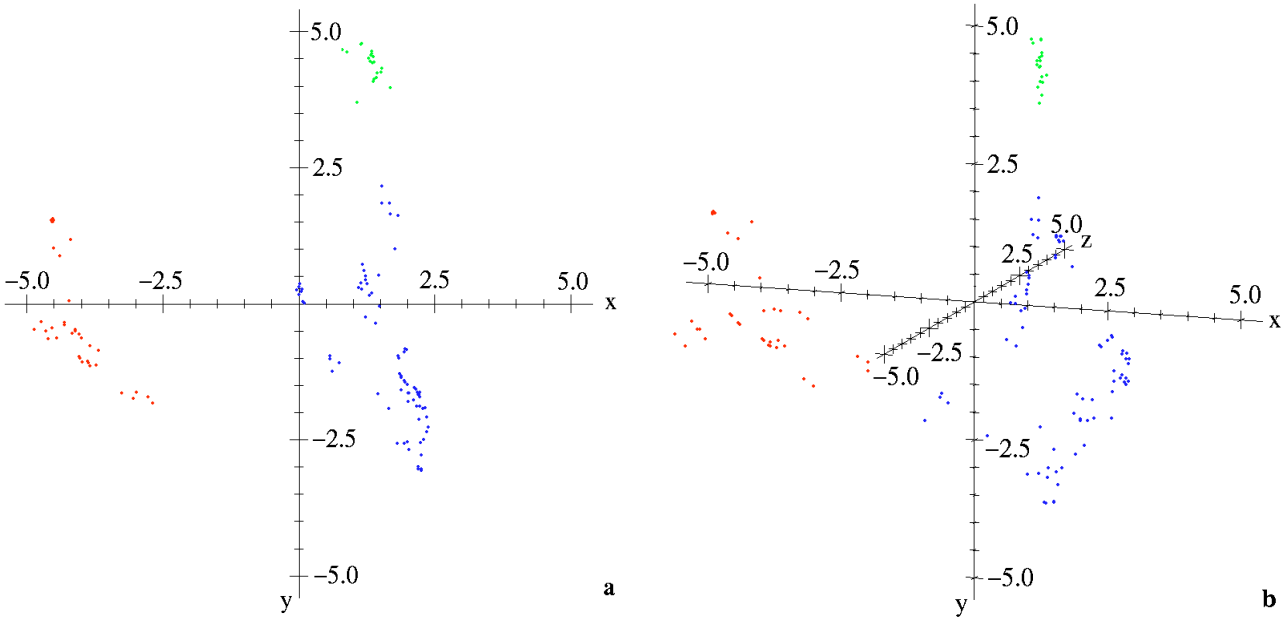


Figure 10 *Graphical representation of the first three principal components of the semiempirically examined tetracycline conformers representing 72% of the diversity of 21 dihedral angles. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 8)*

Table 8 *Table of the three most stable ring conformations of the zwitterionic Tc molecule in energetically descending order. Each conformer can be described by a minimum of three selected dihedral angles given with lower and upper bounds each*

Ring conformer	Energy [kcal·mol^{-1}]	Dihedral angles [°] 3-4-4a-12a	4a-5-5a-6	12-11a-11-10a
1	-237.9	-60.3	-176.9	-86.1
2	-242.1	63.2	104.5	-179.4
3	-244.8	-60.8	168.1	-79.1
Lower bound		-60.8	-176.9	-179.4
Upper bound		63.2	168.1	-79.1

Table 9 *Table of the three most stable ring conformers of the anionic Tc molecule in energetically descending order. Each conformer can be described by a minimum of three selected dihedral angles given with lower and upper bounds each*

Ring conformer	Energy [kcal·mol^{-1}]	Dihedral angles [°] 3-4-4a-12a	4a-5-5a-6	12-11a-11-10a
1	-279.4	75.3	149.5	173.1
2	-283.4	66.7	116.7	175.5
3	-285.7	-61.6	170.7	-77.6
Lower bound		-61.6	116.7	-77.6
Upper bound		75.3	170.7	175.5

gen atoms are not shown. The colour-coding of the molecules according to the dendrogram in Figure 5b clearly indicates that the six hot conformations reduce to three basic ring conformations, which correspond to the three clusters found above. The conformational differences are mainly located in the rings A and B. As a result, the groups 2-$CONH_2$, 3-OH, 4-NMe_2, and 12-OH, which are responsible for possible interactions with the receptor, are able to cover a wide area of receptor space because of the folded and extended ring conformations. This variability is increased even more by the fact that every group can rotate about the bond attached to the ring, which increases the range of possible interaction counter locations within the receptor site.

To summarise, the conformational space of Tc is determined by only three major ring geometries, which are nevertheless able to point the groups on ring A towards completely different spacial directions.

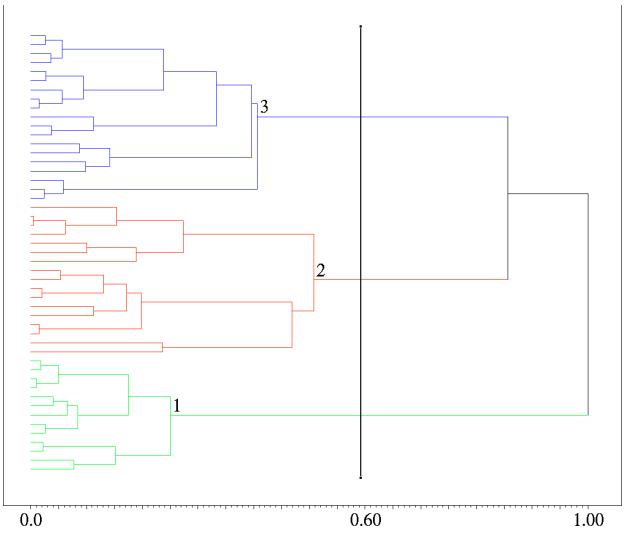


Figure 11 *Cluster analysis performed on the 21 selected dihedral angles describing stable zwitterionic ring conformations. The resulting dendrogram is coloured according the division into three clusters by the vertical ruler*

Principal components analysis The first three principal components of the 14 dihedral angles explain 77%, 12%, and 7% of the variance, respectively. Their eigenvalues are 10.81, 1.65, and 0.93. Thus, only three components explain a total variance of 96% of the complete data set. Table 4 shows the constitution of these three PC's. Defining an abitrary limit of 70%, PC1 is based on all variables except 4a-5-5a-6, 4a-5-5a-11a, and 1-12a-12-11a. The small number of excluded variables is a result of the careful variable selection, using only those dihedral angles which show the establishment of an individual conformation during a longer time period. The first two missing angles are highly correlated (0.93) and define the conformation between the rings B and C. In PC2, the influence of the variables 4a-5-5a-6, 4a-5-5a-11a, and 1-12a-12-11a is increased to improve the predictability of the model. Finally in PC3, no more variables can be included, but the individual communality coefficients raise to a lower limit of 0.89. The variables with the smallest values are 4a-12a-12-11a (0.89) and 1-12a-12-11a (0.89), both including a conformationally restricted double bond.

Figure 7 shows two different views of a graphical representation of the first three principal components. Each of the 8000 points represents one individual conformation extracted

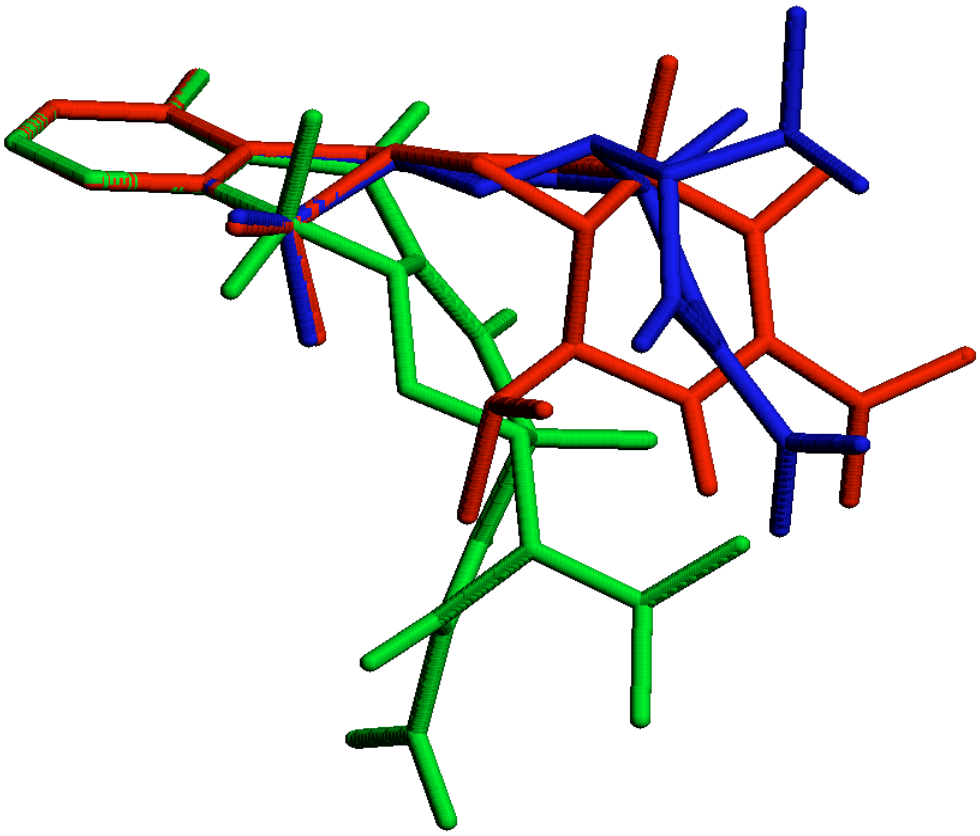


Figure 12 *Overlay of the ring systems of the three zwitterionic Tc conformers found by principle components and cluster analytical methods. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 11)*

from the molecular dynamics trajectory. The data is colour coded according to the results of the cluster analysis defined in the dendrogram (Figure 5b). The conformational similarity between the molecules revealed by the dendrogram is shown by the geometrical proximity of the corresponding points in the three dimensional space of the PCA plot. Therefore, it is easy to understand that geometry optimisations of the cluster centres 1 to 6 of the first approach result in only three basic ring conformations.

Semiempirical calculations

Tc in neutral form Starting with the ring conformations of the cluster centres from the MD-simulations, the semiempirical conformational searches with VAMP produced a total of 5,115 individuals within a heat of formation range from -239.1 to -268.5 kcal·mol^{-1}. The subset of 2,084 conformers located in the lowest 10 kcal·mol^{-1} range (-258.5 to -268.5 kcal·mol^{-1}) corresponds to 40.7% of the whole population. Identical or similar conformers were eliminated by similarity calculations with ASP [29], so that the number of individual conformations decreased from 2,084 to 136. For these 136 conformations, calculations of the normal vibrations were performed, revealing one transition state, which was eliminated. For the remaining 135 conformations, a data matrix was produced consisting of 31 different dihedral angles for each conformer. Dihedral angles with redundant information (i.e. correlation coefficients greater than 0.95) were excluded. The resulting matrix, consisting of 21 dihedral angles per conformer, was analysed by cluster-analytical methods and by generating principle components.

A cluster analysis of the dihedral angles was performed by the Ward method [32] with euclidean distance metric and z-transformed data. The results are shown as a dendrogram in Figure 8. The dendrogram has been divided into three different colour-coded clusters. The three different ring conformations, represented by the most stable member of each cluster, are shown in Figure 9 as a colour-coded overlay of the ring system of all found three ring conformations of Tc. The energetically lowest conformer is shown in dark-blue, the energetically highest in red. This overlay shows that the flatter a Tc ring conformations is, the more stable it is in the gas phase. The most significant differences between the conformers can be seen by means of the selected dihedral angles shown in Table 7. These dihedral angles are taken from the set of the 21 dihedral angles and describe the variance maxima in the first three communalities.

The first eight eigenvalues, shown in Table 5, of the PCA show two important points: Firstly, the two ring conformations of Tc described in the literature cannot be described adequately by only two dihedral angles. Thus, only 54% of the total variance can be explained by the two largest eigenvalues. Secondly, at least three principle components with eigenvalues ≥ 3.8 are necessary to describe most of the ring conformations of Tc. In this way 72 % of the conforma-

tional variance can be explained. The most important communalities of the first three principle components explaining the cumulative variance of each of the 21 dihedral angles are shown in Table 6. PC1 is based principally on the dihedral angles 4-4a-12a-1, 4-4a-12a-12, 12a-12-11a-11 and 1-12a-12-11a, indicating that the first PC is mainly dominated by twisting around bond 12a-12 between rings A and B. In PC2, the angles 4a-5-5a-11a, 5a-6-6a-10a and 11a-5a-6-6a were added. These correspond to rotation around bond 5a-6. Finally in PC3, especially the dihedral angles 3-4-4a-5, 12-11a-11-10a and 12-11a-5a-6 show a contribution to the statistical model. Both PC2 and PC3 describe the twist between rings C and D. A scatter plot of these three PCs is shown in Figure 10. In this way several agglomerations of points could be separated, pointing to three different ring conformations.

The aromatic ring D of Tc and the connected atoms C_6 and C_{11} are absolutely flat. All conformers are folded against an axis through the atoms C_6 and C_{11}. The dihedral angle 12-11a-11-10a describes the upshift of C_{12} against the flat part of the C-ring and is mainly responsible for the twist between rings B and C. The dihedral angles 3-4-4a-12a and 4a-5-5a-6 together describe the movement of the carboxamide- or the dimethylamino-group towards the benzene ring. Figure 9 shows the colour-coded overlay of the same three conformers with their substituents but without H-atoms. This figure also shows the high steric flexibility especially of ring A and its carboxamide- and dimethylamino-substituents.

The energetic diversity of the conformers found can be deduced mainly from intramolecular forces such as hydrogen bonding and delocalisation. In all stable conformers the hydroxyl function at carbon C_{10} is oriented towards the carbonyl oxygen at carbon C_{11}. The hydroxyl group at carbon C_{12} can stabilise the geometries by interaction with the carbonyl function on carbon C_{11} and the oxygen of the hydroxyl group at carbon C_{12a}. The hydroxyl group at carbon C_6 and the dimethylamino-function do not play an important part in the stabilisation of the three most stable conformers of neutral Tc. The carboxamide-group is connected to the hydroxyl function at carbon C_3 and the oxygen of the carbonyl function at carbon C_1 by two hydrogen bonds.

Tc in neutral zwitterionic form A cluster analysis of the uncharged zwitterionic Tc molecule gave the dendrogram shown in Figure 11, which shows three different colour-coded clusters. Table 5 (middle) shows the eight largest eigenvalues resulting from the PCA. The colour-coded overlay of the three most stable conformers found by principle components and cluster-analytical methods is shown in Figure 12. Again, the energetically lowest conformer is darkblue-, the energetically highest red-coloured. A plot of the first three principle components, which explain 68 % of the total variance, is shown in Figure 13. The points are colour-coded as in Figure 10, showing the same ring conformations as found in the cluster analysis. The energetically most stable individuals of each cluster are characterised by the same dihedral angles as the neutral Tc molecule (Table 8). All three conformers have been

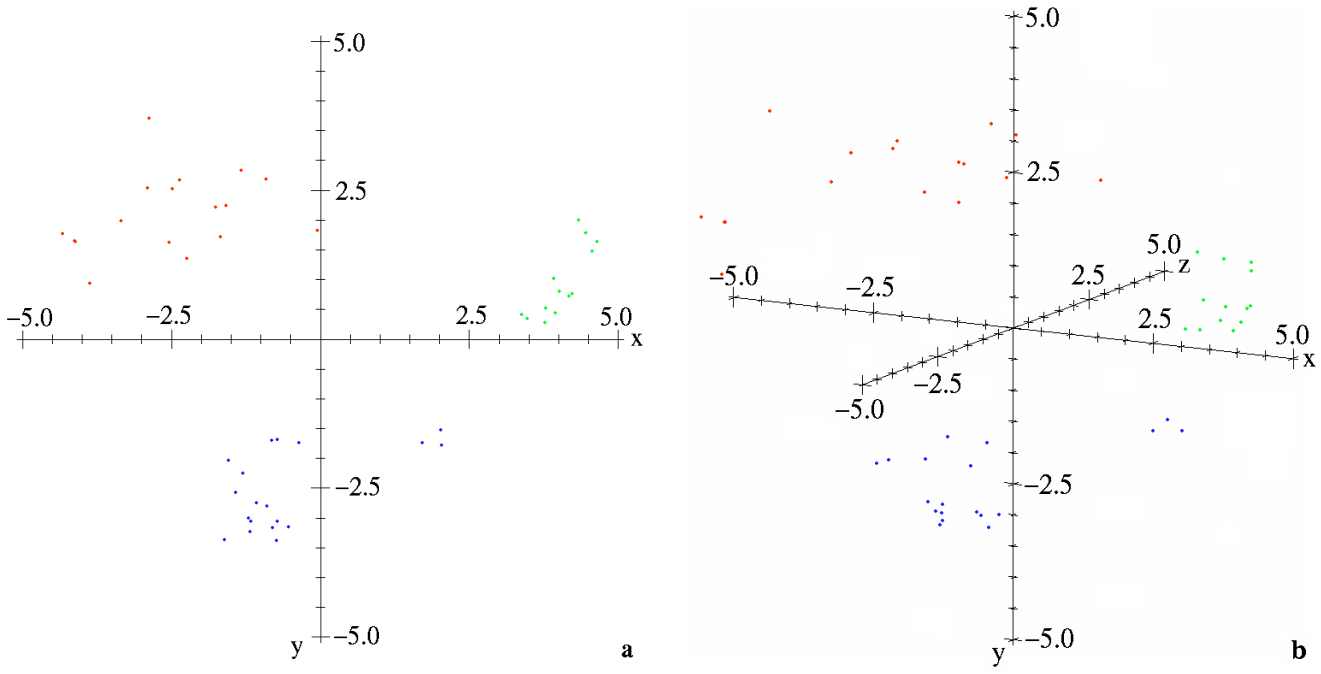


Figure 13 *Graphical representation of the first three principal components of the zwitterionic tetracycline conformers representing 68% of the diversity of 21 dihedral angles. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 11)*

Table 10 *Table of the three most stable ring conformers for the neutral, zwitterionic and anionic Tc molecules, each ordered according to their AM1-calculated heat of formation relative to the most stable conformer. The conformers are colour-coded as in the overlays and dendrograms*

Species	Distance [Å] O_{12}-O_{11}	Dihedral [°] O_{11}-C_{11}-C_{12}-C_{12}	ΔΔ H^o_f [kcal·mol^{-1}]
Neutral	2.709	1.0	0.0
	2.690	-2.8	2.2
	2.801	-36.8	4.6
Zwitterion			
	2.688	-1.7	0.0
	2.810	-32.3	2.7
	2.711	1.5	6.9
Anion			
	2.712	10.5	0.0
	2.797	28.4	2.3
	3.092	27.9	6.3

found within a heat of formation range from -237.9 and -244.8 kcal·mol^{-1}, about 25 kcal·mol^{-1} higher than the neutral uncharged Tc molecule. Thus, introducing a zwitterionic system does not change the conformational behaviour of Tc.

Tc in anionic form The dendrogram of the cluster analysis for Tc deprotonated at 12-OH is shown in Figure 14. It shows three different colour-coded clusters. A colour-coded overlay of the three most stable conformers found by principle components and cluster-analytical methods is shown in Figure 15. Therein, the energetically lowest conformer is darkblue-, the energetically highest red-coloured. Table 5 (bottom) shows the eight largest eigenvalues resulting from the PCA. A plot of the first three principle components, which explain 83 % of the total variance, is shown in Figure 16, showing three different colour-coded agglomerations of points corresponding to the same three ring conformations found in the cluster analysis. The energetically most stable individuals of each cluster are characterised by the same dihedral angles as the neutral Tc molecule (Table 9). All three conformers have been found within a heat of formation range between -279.4 and -285.7 kcal·mol^{-1}.

Conclusions

To investigate the flexibility of tetracycline, we have combined molecular dynamics simulations, semiempirical calculations, and statistical data analysis in order to investigate the conformational space systematically.

High temperature equilibration at 1200K over a time of 400 ps followed by a 400 ps data aquisition period generated a set of 8000 individual conformations, which were represented by 31 selected dihedral angles without considering side chain orientations. Analysing this time dependent data by clustering techniques and generating principle components, we identified three basic ring conformations which can be described by the dihedral angles 3-4-4a-12a, 4a-5-5a-6, and 1-12a-4a-5 (see Figure 1 and Table 3). These variables define the relative orientation of the rings A, B, and C within tetracycline. As shown in Figure 6, these three basic ring conformations allow the rotatable side chains to cover a large area in space where possible interaction points of a tetracycline receptor may be located. Our results extend qualitative considerations found in the literature [31], which consider only two conformations interconvertible by rotation around the central C_{4a}-C_{12a} bond.

Starting with the ring conformations of the cluster centres from the MD-simulations, semiempirical torsion calculations were performed for the uncharged, zwitterionic and anionic tetracycline molecules. The ring conformations obtained were described by non-redundant dihedral angles of the ring systems and were classified by principle components and cluster-analytical methods. For the neutral tetracycline molecule, the zwitterionic and anionic Tc conformers the clustering methods produced three different ring conformations. All ring conformers can be described adequately by the dihedral angles 3-4-4a-12a, 4a-5-5a-6 and 12-11a-11-10a (see Tables 7-9), whose influence on the conformational geometries can be seen e.g. from the overlays of the most stable conformers for the uncharged, zwitterionic and anionic tetracycline molecules (see Figures 9, 12, 15): The dihedral angles 3-4-4a-12a

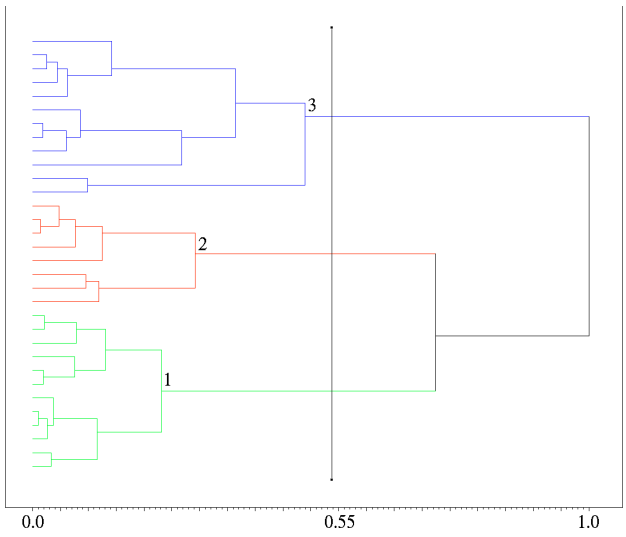


Figure 14 *Cluster analysis performed on the 21 selected dihedral angles describing stable anionic ring conformations. The resulting dendrogram is coloured according the division into three clusters by the vertical ruler*

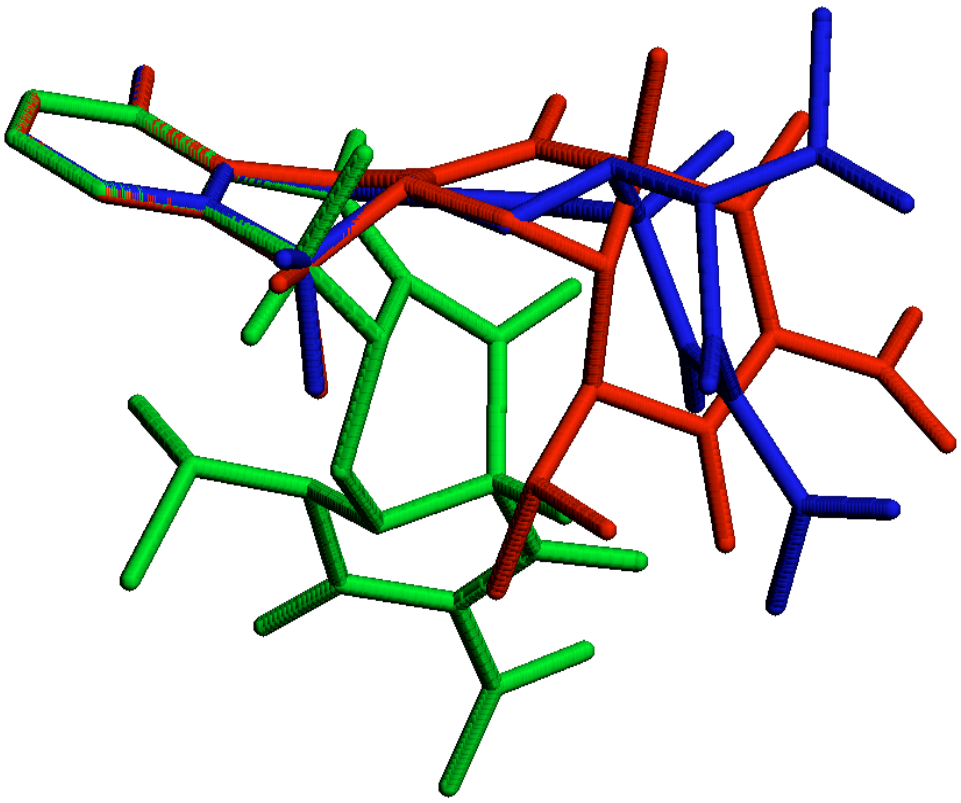


Figure 15 *Overlay of the ring systems of the three anionic Tc conformers found by principle components and cluster analytical methods. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 14)*

and 4a-5-5a-6 are mainly responsible for the movement of the carboxamide- and dimethylamino-group towards the benzene ring, whereas the angle 12-11a-11-10a describes the twist between rings B and C. Additionally, the high steric flexibility especially of ring A and its carboxamide- and dimethylamino-substituents, which is important for physiological receptor interactions, can be recognized.

A schematic two-dimensional representation of the ring systems of the conformers found for the neutral (n1 – n3), zwitterionic (z1 – z3) and anionic (a1 – a3) tetracyclines is shown in Figure 17. Additionally, for each conformation a VRML-scene is available as supplementary material. For the neutral form, the most stable conformation (n1) consists of a twist boat for the C-ring with the corners twisted up towards the ring face *syn* to the A-ring. The B-ring is best described as a flattened chair and the A-ring as a boat, as shown in Figure 17. The least stable conformation (n3) differs only in the conformation of the C-ring, which is twisted towards the face *anti* to the A-ring. The third conformation (n2), which is intermediate in energy, differs from the most stable (n1) in the conformation of the B-ring. Rather than the flattened chair found in the other two conformations, this ring adopts a twist boat conformation.

The neutral zwitterion exhibits a fully analogous unstable conformation (z3) to its nonpolar isomer. The intermediate conformation (z2) differs from its nonpolar equivalent only in a slightly different conformation of the A-ring, whereas the most stable conformer (z1) has a twist boat conformation, rather than the flattened chair found in the nonpolar form, for the central B-ring.

The anionic species also exhibits a conformation fully analogous to the red forms found for the neutral forms of Tc (a2), except that it is now intermediate in stability between the unstable conformer (a3), which corresponds closely to the most stable conformer (z1) of the neutral zwitterion. The most stable conformation of the anion (a1) represents a new conformation not found for the neutral forms. Thus, the degree of deprotonation of Tc obviously plays a significant role in determining its preferred conformation.

We can, however, draw some conclusions from the present gas phase study about the influence of the conformations found on Tc‘s ability to complex metal cations. Table 10 shows the distance between oxygens 11 and 12, the most likely ligand atoms for a coordinated metal ion, and the dihedral angle between the two CO-bonds. The data, colour-coded as in the overlays and dendrograms, once again confirm that the neutral and zwitterionic forms behave very similarly. The only major change is that the green conformer, with the largest distance and twist angle is stabilised on formation of the zwitterion and that the red form is destabilised. The anion, however, shows very different behaviour. This is largely caused by the change in the conjugation that allows the C_{11a}-C_{12} bond to twist. The O – O distance, which probably affects the metal ion complexation strongly is around 2.7 Å for the flat (blue and red) conformers of the neutral and zwitterionic forms and increases to 2.8 Å for the green, twisted conformer. The anion conformers, on the other hand, cover a

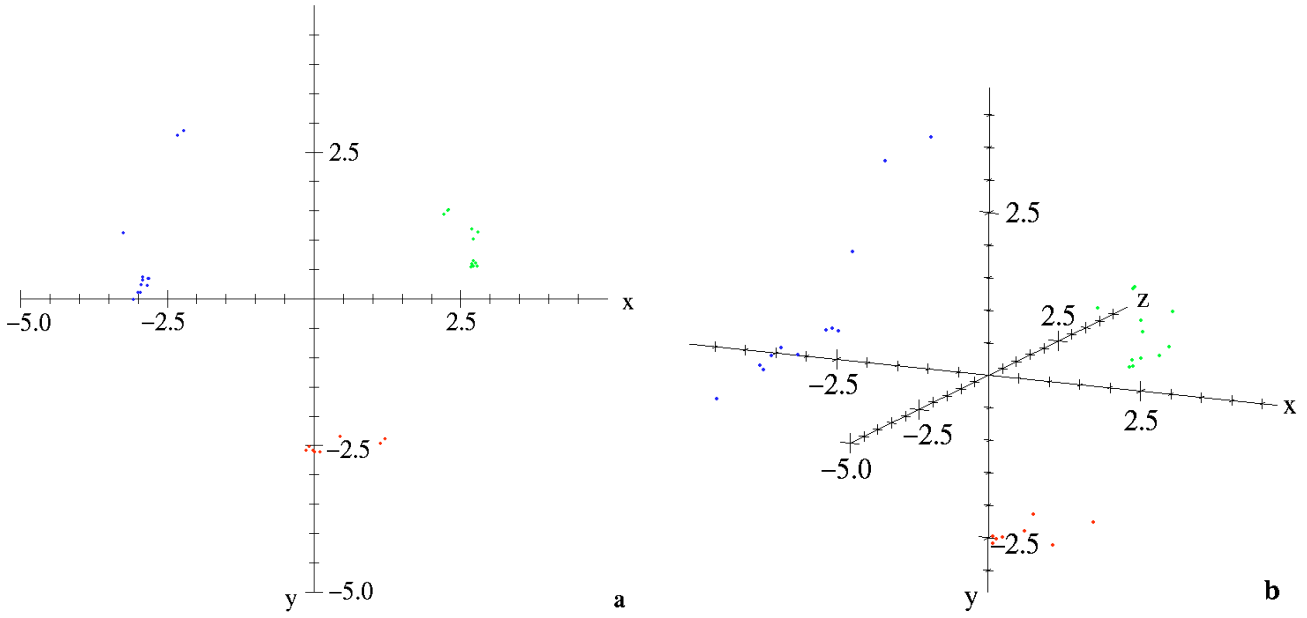


Figure 16 *Graphical representation of the first three principal components of the anionic tetracycline conformers representing 83% of the diversity of 21 dihedral angles. Colour coding is performed according to the results of the cluster analysis shown in the dendrogram (Figure 14)*

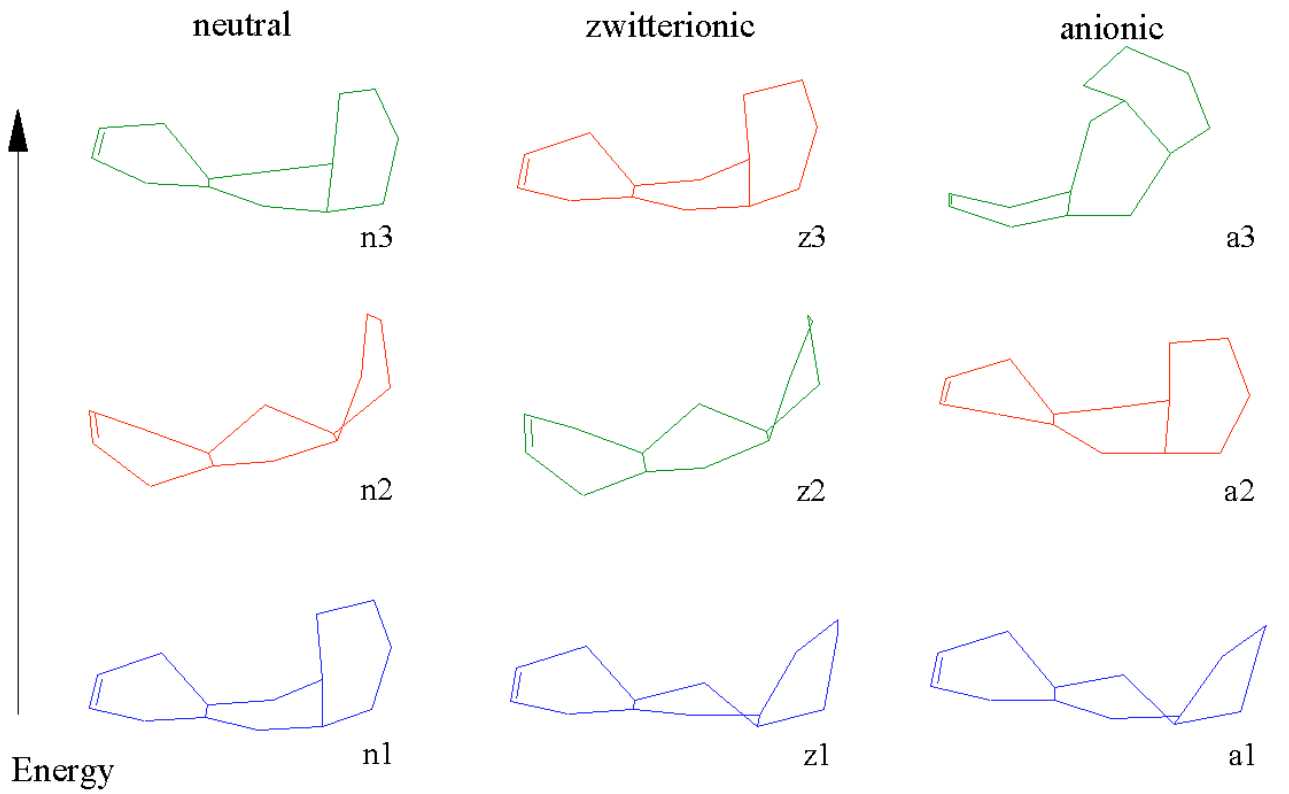


Figure 17 *Schematic two-dimensional representation of the ring systems of the conformers found for the neutral (n1 – n3), zwitterionic (z1 – z3) and anionic (a1 – a3) tetracyclines. The colour coding corresponds to the results of the cluster analyses shown in the dendrograms of Figures . All conformers are ordered according to their increasing heats of formation. Ring D and substituents have been omitted, whereas ring C can be found at the left side of the representation. All conformers are viewed along the $11_a – 5_a$ bonding*

range of 2.7 – 3.1 Å from the most stable to the least stable conformer. These results suggest that metal ions will affect the stability of the conformers and, above all, that the degree of deprotonation plays a very significant role in determining the conformational behaviour of Tc.

Future investigations on the neutral and ionic Tc molecules will therefore be performed in aqueous environment with and without addition of metal cations.

Acknowledgments This work was supported by the Fonds der Chemischen Industrie. We thank Prof. W. Hillen for helpful discussions.

Supplementary material available The ring conformations found for the neutral (n1 – n3), zwitterionic (z1 – z3) and anionic (a1 – a3) tetracycline, as shown in the schematic two-dimensional representation of Figure 17 are each available as VRML-scenes.

References

1. Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc*. **1990**, *112*, 1419.
2. Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371.
3. Leach, A. R.; Prout, K.; Dolata, D. P. *J. Comput. Aided Mol. Design* **1988**, *2*, 107.
4. Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation. Chemometrics Research Studies Series 15.* New York, Wiley, 1988.
5. Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading, Massachusetts, Addison-Wesley, 1989. McGarrah, D. B.; Judson, R. S. *J. Comput. Chem.* **1993**, *14*, 1385.
6. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
7. Mitscher, L. A. *The Chemistry of Antibiotics*; Marcel Dekker: New York, 1978, p. 91.
8. Kohn, K. W. *Nature* **1961**, *191*, 1156.
9. Dürckheimer, W. *Angew. Chem.* **1975**, *21*, 751.
10. Gulbis, J.; Everett, G. W., Jr. *Tetrahedron* **1976**, *32*, 913.
11. Mitscher, L. A.; Slater-Eng, B.; Sokoloski, T. D. *Antimicrob. Agents Chemother.* **1972**, *2*, 66.
12. Mitscher, L. A.; Bonacci, A. C.; Sokoloski, T. D. *Antimicrob. Agents Chemother.* **1968**, 78.
13. Hughes, L. J.; Stezowski, J. J.; Hughes, R. E. *J. Am. Chem. Soc.* **1979**, *101*, 7655.
14. Stezowski, J. J.; *J. Am. Chem. Soc.* **1977**, *99*, 1122.
15. Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*, 2nd ed. Butterworth, London, 1959, 517.
16. *AMBER 5.0*, Oxford Molecular, Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, England, 1997.
17. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
18. Beck, B.; Glen, R. C.; Clark, T. *J. Comput. Chem.* **1997**, *18*, 744.
19. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
20. *VAMP 6.5*, Rauhut, G.; Alex, A.; Chandrasekhar, J.; Steinke, T.; Sauer, W.; Beck, B.; Hutter, M.; Gedeck, P.; Clark, T.; Oxford Molecular, Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, England, 1997.
21. Stephens, C. R.; Murai, K.; Brunings, K. J. ; Woodward, R. B. *J. Am. Chem. Soc.* **1956**, *78*, 4155.
22. Aldenderfer, M. S.; Blahfield, R. K. *Cluster Analysis.* Newbury, Park, Sarge. Garland Publishing, New York, 1984.
23. Eckes, T.; Roßbach, H. *Clusteranalysen*. Stuttgart, Kohlhammer, 1980.
24. *TSAR 3.1*, Oxford Molecular, Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, England, 1997.
25. Bortz, J. *Statistik.* Heidelberg, Springer, 1993.
26. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vertterling, W. T. *Numerical Recipes in C - The Art of Scientific Computing.* Cambridge, Cambridge University Press, 1992.
27. Hotelling, H. *J. Educ. Psychol.* **1933**, *24*, 417.
28. Hotelling, H. *J. Educ. Psychol.* **1933**, *24*, 498.
29. *ASP 3.21*, Oxford Molecular, Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, England, 1997.
30. Carbo, R.; Leyda, L.; Arnan, M. *Int. J. Quant. Chem.* **1990**, *17*, 1185.
31. Gulbis, J. and Everett, G. W. Jr. *Tetrahedron* **1976**, *32*, 913.
32. Ward, H. J. *J. Am. Stat. Assoc.* **1963**, *58*, 236.