JOURNAL OF CHEMOMETRICS
J. Chemometrics 2005; **19**: 412–426
Published online 16 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.943

Bayesian conformational analysis of ring molecules through reversible jump MCMC

Kim Nolsøe[1], Mathieu Kessler[2]*, José Pérez[2] and Henrik Madsen[1]

[1]Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark
[2]Universidad Politécnica de Cartagena, Paseo Alfonso XII, Cartagena, España, Spain

Received 23 May 2005; Revised 11 August 2005; Accepted 30 September 2005

In this paper, we address the problem of classifying the conformations of *m*-membered rings using experimental observations obtained by crystal structure analysis. We formulate a model for the data generation mechanism that consists in a multidimensional mixture model. We perform inference for the proportions and the components in a Bayesian framework, implementing a Markov chain Monte Carlo (MCMC) reversible jump algorithm to obtain samples of the posterior distributions. The method is illustrated on a simulated data set and on real data corresponding to cyclo-octane structures. Copyright © 2005 John Wiley & Sons, Ltd.

KEYWORDS: cycloalcanes; reversible jump; conformational analysis; mixture model; MCMC

1. INTRODUCTION

For a given compound it is of interest to study what are the preferred geometrical conformations of the corresponding molecules. The conformational classification of structures and, in particular, the understanding of the factors that determine the molecular structure of a particular compound are important, since ideally they would allow for a rational design of complexes with specific and predictable properties; see Reference [1].

On the one hand, molecular mechanics combined with energy considerations allow us to define, for some particular structures, a given number of canonical conformations. For example the ten canonical conformations deduced by Reference [2] for cyclo-octane are represented in Figure 1. However, it is known from experimental data that some of these canonical conformations are almost never observed, and that, in contrast, some new conformations may appear. These new conformations usually appear as deformations of the canonical ones. Statistical descriptive methods have been employed as a complement to molecular mechanics computations, to detect and identify the preferred conformations in a given compound, that is to cluster the observed structures into a number of conformations. A review of different statistical methods for conformational analysis can be found in Reference [3]. These methods generally take a data analysis approach where no model is assumed for the data generation mechanism, and all the conclusions rely on the correlation structure or the similarity structure of the data. Cluster analysis and principal components analysis are examples of such methods.

In this paper, we address the problem of conformational classification of *m*-membered rings, from the observation of crystallographic data consisting in the torsion angles for a number of structures. In contrast to previously proposed methods, an essential step in our approach consists in specifying a probabilistic model for the observed sequences of torsion angles. This probabilistic model is a mixture model with an unknown number of components. We perform a Bayesian analysis by implementing the Reversible Jump Markov chain Monte Carlo (MCMC) methodology proposed by Reference [4], to obtain samples of the posterior distribution of the parameters, and infer on the conformations along with their frequencies of occurrence. We take into consideration, both in the specification of the prior distributions and in the updating steps of the MCMC algorithm, the geometrical restrictions that link the *m* torsion angles of an *m*-membered ring.

Section 2 describes the data. In Section 3, a model is formulated for the data generation mechanism, including the specification of prior distributions for the parameters. Section 4 describes the applied methodology. The results of our method on a simulated dataset and on real a dataset corresponding to cyclo-octane previously investigated by Reference [5] are presented in Section 5. Finally some conclusions are drawn in Section 1 while the Appendix contains a description of the structure of *m*-membered rings used for

*Correspondence to: M. Kessler, Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Paseo Alfonso XIII, E-30203 Cartagena , España, Spain.
E-mail: Mathieu.Kessler@upct.es
Contract/grant sponsor: European Community's Human Potential Programme; contract/grant number: HPRN-CT-2000-00100, DYNSTOCH.

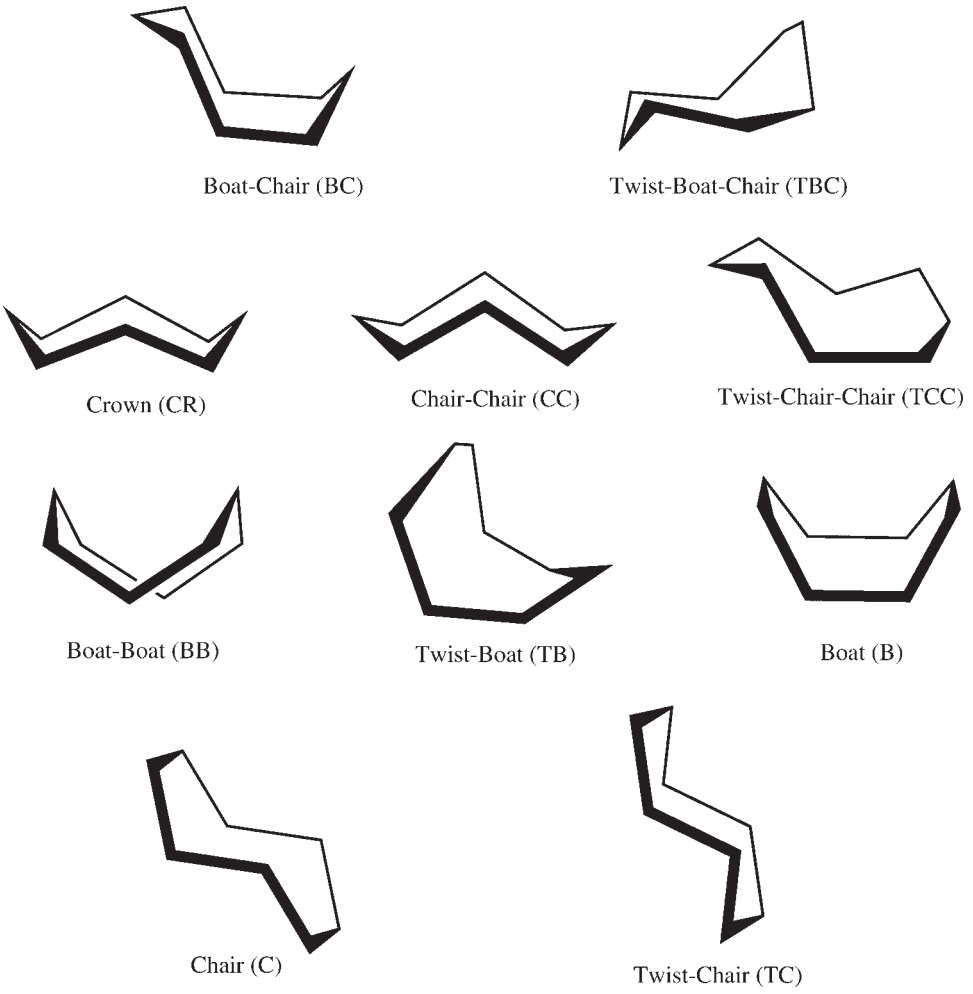

Figure 1. Canonical forms of cyclo-octane.

the MCMC algorithm, together with the expression of the acceptance probabilities associated to the different move types in the reversible jump MCMC algorithm.

2. DESCRIPTION OF THE DATA

The Cambridge Structural Database (CSD) is a powerful tool that provides chemists, access to a large amount of crystallographic structural data. Users of the CSD can retrieve the structures that match a given criterion from the database, for example all the cyclo-octane structures, and can obtain the crystallographic data that characterize each of these structures.

An important set of characteristics associated to a given m-membered ring structure consists in the m-associated torsion angles; see Appendix A.1.1 for a precise definition. These data can be retrieved from the CSD and we will assume that inference is to be made on the conformations and their frequencies of occurrence based on a sample of n structures for each of which we observe the sequence of m torsion angles. In particular, in Section 5, we will carry out the conformational classification of a sample of 31 cyclo-octane structures retrieved from the CSD. This dataset was previously analyzed by means of cluster analysis by Reference [5].

We want to emphasize some important issues related to the data: assume that we retrieve from the CSD the torsion angles corresponding to an $m = 8$ membered ring built of consecutive atoms $A_1, A_2, \ldots, A_8$.

For this single structure, the data consist in the sequence

$$(\mathrm{Tor}(A_1, A_2, A_3, A_4), \mathrm{Tor}(A_2, A_3, A_4, A_5), \ldots \\ \mathrm{Tor}(A_7, A_8, A_1, A_2), \mathrm{Tor}(A_8, A_1, A_2, A_3)) = (\tau_1, \tau_2, \ldots, \tau_7, \tau_8)$$

that is the torsion angles computed from four consecutive atoms, where the starting atom is varying through the whole ring. However, from the retrieved data it is impossible to know what was the atom that was chosen as a starting point to begin measuring the torsion angles. Concretely, this means that, for the same structure, the data could as well be $(\tau_2, \tau_3, \ldots, \tau_8, \tau_1)$, $(\tau_3, \tau_4, \ldots, \tau_1, \tau_2)$ or any of the cyclical translations of $(\tau_1, \tau_2, \ldots, \tau_7, \tau_8)$.

Moreover, the perspective from which the molecule is being measured can either be from above or from below. This implies that, for the same structure, the sequence of torsion angles can be read in a clockwise or counter-clockwise manner.

Finally, if a given conformation is present in the compound, its mirror image can also be found. As a consequence, and as described in Reference [6], the two sequences of torsion angles $(\tau_1, \tau_2, \ldots, \tau_7, \tau_8)$ and $(-\tau_1, -\tau_2, \ldots, -\tau_7, -\tau_8)$ can be met but shall be considered to correspond to equivalent conformations.

These three aspects of the retrieved data will be taken into account when we formulate our model in Section 3.

3. THE INGREDIENTS OF OUR MODEL

3.1. Notations related to vectors

To denote a d-dimensional vector $(v_1, \ldots, v_d)$ we use the notation $v_{1:d}$. If a vector v of length $k \cdot m$ is built up from binding a number k of m-dimensional blocks, we will use the matrix notation $v_{i,j}$ to denote the jth element of the ith block, the notation $v_{i,1:m}$ to denote the ith block and $v = v_{1:k,1:m}$ to denote the whole vector.

3.2. The initial model

Assume that we have retrieved n structures from the CSD, and denote by $\tau^{(1)}, \ldots, \tau^{(n)}$ the n associated m-dimensional vectors of torsion angles. We assume that they correspond to independent and identically distributed realizations from a mixture law with density $\tau \to f(\tau)$, described in Equation (1).

Consider $\tau = \tau_{1:m}$, the m torsion angles that are observed for a given structure. We assume that τ is generated from a mixture of an unknown number k of components. These k components correspond to perturbations of the conformations that can be met for the considered structures. For $c = 1, \ldots, k$, the conformation c is described through an m-dimensional vector of torsion angles $\mu_{c,1:m} = (\mu_{c,1}, \ldots, \mu_{c,m})$ and we denote by w_c its unknown frequency of occurrence for the considered kind of structure.

The observed sequence τ is then generated from the mixture density

$$f(\tau) = \sum_{c=1}^{k} w_c f(\tau, c) \tag{1}$$

where $\tau \to f(\tau, c)$ is the density of τ given the conformation c, which is described in detail below.

We are interested in estimating the number k of conformations present for the structure, the associated torsion angles $\mu_{c,1:m}$ for $c = 1, \ldots, k$, and the frequencies of occurrence $w_1, \ldots, w_k$.

3.2.1. *Description of the density of τ given the conformation $C = c$*

As described in Section 2, the output of the measurement device which yields the observed τ may correspond to a different starting point, a different direction and an opposite sign than the sequence $\mu_{c,1:m}$ associated to the conformation number c.

Consider an m-dimensional vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$. The effect of choosing a starting point s in $\{1, \ldots, m\}$, a direction d, which equals 1 if the sequence is read clockwise or -1 if it is read counter-clockwise, and choosing a sign $\delta = \pm 1$, is described through the operator $T^{s,d,\delta}$ which acts on $\boldsymbol{\mu}$ in the following way:

$$\begin{aligned} &T^{s,d,\delta}\boldsymbol{\mu} \\ &= \delta \times (\mu_s, \mu_{((s-1+d\times 1) \bmod m)+1}, \ldots, \mu_{((s-1+d\times(m-1)) \bmod m)+1}) \end{aligned} \tag{2}$$

where for any integer j, $j \bmod m$ denotes the remainder of the integer division of j by m. Notice in particular that $T^{1,1,1}\boldsymbol{\mu} = \boldsymbol{\mu}$.

As an example, consider the canonical conformation Boat–Chair in Figure 1. As derived in Reference [2], the associated sequence of torsion angles is

$$\mu = (65.0, 44.7, -102.2, 65.0, -65.0, 102.2, 44.7, -65.0)$$

It implies that, for example,

$$T^{2,-1,1}\mu = (44.7, 65.0, -65.0, -44.7, 102.2, -65.0, 65.0, -102.2)$$

Finally, given that the structure was generated from conformation c, the observed sequence τ is obtained, for a given s, d, δ, from $T^{s,d,\delta}\mu_{c,1:m}$ after an additive perturbation $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$, which is assumed to be a Gaussian variable with covariance matrix $\sigma_c^2 \mathbf{I}d_m$.

As a conclusion, the density $\tau \to f(\tau, c)$ in Equation (1), is

$$f(\tau, c) = \frac{1}{4m} \sum_{s=1}^{m} \sum_{d=1,-1} \sum_{\delta=1,-1} f_G(\tau, T^{s,d,\delta}\mu_{c,1:m}, \sigma_c^2 \mathbf{I}d_m) \tag{3}$$

and $\tau \to f_G(\tau, T^{s,d,\delta}\mu_{c,1:m}, \sigma_c^2 \mathbf{I}d_m)$ denotes the density of the m-dimensional Gaussian law with mean $T^{s,d,\delta}\mu_{c,1:m}$ and diagonal covariance matrix $\sigma_c^2 \mathbf{I}d_m$.

3.2.2. *Density of the observations*

To sum up, the density of the observations $\tau^{(1)}, \ldots, \tau^{(n)}$ is given by

$$\prod_{i=1}^{n} \left(\sum_{c=1}^{k} w_c f(\tau^{(i)}, c) \right)$$

where $f(\tau, c)$ is defined in Equation (3).

Remark. The additive perturbation ε may be seen as a measurement error, but our modeling framework is as well inspired by the model-based clustering approach in Reference [7], where mixtures of Gaussian variables are used as a basis for clustering algoritms. In particular, the size and shape of each cluster can be modeled by imposing a given structure for the covariance matrix of the corresponding Gaussian distribution. Assuming, as we do, the covariance matrix of ε to be diagonal but allowing the variance to vary between conformations amounts to considering spherical clusters of possibly different sizes.

3.3. Extension of the parameter space

For a given conformation, the parameters $\mu_{c,1:m}$ that enter Equation (3) cannot be freely chosen: for the corresponding ring to be physically coherent, they must satisfy some restrictions. These restrictions are not easy to express, but it is necessary to have a control on them, on the one hand to specify a reasonable prior distribution on $\mu_{c,1:m}$ and on the other hand, since we will need to be able to propose reasonable candidates for $\mu_{c,1:m}$ in the MCMC algorithm.

We can guarantee that the physical restrictions are fulfilled if we extend the parameter space to include as well the bond angles and the atomic lengths of the molecule, and therefore get a complete description of the conformation number c. For a precise definition of torsion angles, bond angles and distances, see Appendix A.1.1.

Consider an m-membered ring with consecutive atoms $A_1, A_2, \ldots, A_m$. As in Appendix A.1.1, we denote by $\tau_{k-2;\,k-1;\,k;\,k+1}$ the torsion angle for the sequence of four points $A_{k-2}, A_{k-1}, A_k, A_{k+1}$, by $b_{k-1;\,k;\,k+1}$ the bond angle for the

points A_{k-1}, A_k, A_{k+1}, and by $d_{k-1;\,k}$ the distance between A_{k-1} and A_k.

For the ring $A_1, A_2, \ldots, A_m$, we can compute m torsion angles $\mu_{1:m}$, m bond angles $b_{1:m}$ and m atomic lengths $d_{1:m}$ as

$$\begin{aligned}\mu_{1:m} =&(\tau_{1;\,2;\,3;\,4}, \ldots, \tau_{m-3;\,m-2;\,m-1;\,m},\\ &\tau_{m-2;\,m-1;\,m;\,1}, \tau_{m-1;\,m;\,1;\,2}, \tau_{m;\,1;\,2;\,3})\end{aligned} \tag{4}$$

$$b_{1:m} = (b_{1;\,2;\,3}, \ldots, b_{m-2;\,m-1;\,m}, b_{m-1;\,m;\,1}, b_{m;\,1;\,2}) \tag{5}$$

$$d_{1:m} = (d_{1;\,2}, \ldots, d_{m-1;\,m}, d_{m;\,1}) \tag{6}$$

These quantities are obviously related. In fact, as described in Appendix A.1, an m-membered ring can be drawn by choosing freely the first $m-3$ torsion angles, $\mu_{1:m-3}$, the first $m-2$ bond angles, $b_{1:m-2}$, and $m-1$ atomic lengths, $d_{1:m-1}$. This fixes fully the relative positions of the m atoms in the ring. The remaining angles and lengths can be uniquely determined through the mapping

$$(\mu_{1:m}, b_{1:m}, d_{1:m}) = F(\mu_{1:m-3}, b_{1:m-2}, d_{1:m-1}) \tag{7}$$

which is described in Appendix A.1.3.

As a conclusion, once we have fixed the number k of components of our mixture model, we will consider the extended parameter to be $w_{1:k}$, $\mu_{1:k,1:m}$, $b_{1:k,1:m}$, $d_{1:k,1:m}$ and $\sigma^2_{1:k}$ although the density of the observed variables τ depends only on $w_{1:k}$, $\mu_{1:k,1:m}$, and $\sigma^2_{1:k}$; see Equations (1) and (3).

3.4. The prior distributions

3.4.1. *The parameters k, $w_{1:k}$ and $\sigma^2_{1:k}$*

For the unknown number of components, the variances of the Gaussian perturbations and the mixture proportions, we follow References [8] and [9], and choose the following priors:

$$k \sim U(1, k_{\max})$$

$$\sigma_1^2, \ldots, \sigma_k^2 \text{ are independent and } \sigma_i^2 \sim IG(\alpha_i, \beta_i), \quad i = 1, \ldots, k$$

$$(w_1, \ldots, w_k) \sim D(1, \ldots, 1)$$

where $U(1, k_{\max})$ denotes the discrete uniform distribution on the integers between 1 and $k_{\max}$, $IG(\alpha, \beta)$ denotes the Inverse Gamma distribution with parameters (α, β), and $D(\alpha_1, \ldots, \alpha_k)$ denotes the Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_k)$.

3.4.2. *The parameters μ, b and d*

Given a value of k, we have to specify our prior distribution on $\mu_{c,1:m}$, $b_{c,1:m}$, and $d_{c,1:m}$ for $c = 1, \ldots, k$ and we want it to charge only physically coherent molecules.

From Subsection 3.3, it is known that this can be done by specifying a prior distribution for $\mu_{c,1:(m-3)}$, $b_{c,1:m-2}$ and $d_{c,1:m-1}$, and deduce the remaining angles and lengths through the mapping F; see Equation (7). We therefore choose, for all $c = 1, \ldots, k$

$$\mu_{c,1:(m-3)} \sim \bigotimes_{1:(m-3)} U(-\pi, \pi) \tag{8}$$

$$b_{c,1:(m-2)} \sim N(\eta^b_{1:m-2}, \kappa^b\, \mathbf{I}d_{m-2}) 1_{[\eta^b_i \pm 2\sqrt{\kappa^b}]} \tag{9}$$

$$d_{c,1:(m-3)} \sim N(\eta^d_{1:m-1}, \kappa^d\, \mathbf{I}d_{m-3}) 1_{[\eta^d_i \pm 2\sqrt{\kappa^d}]} \tag{10}$$

and

$$(\mu_{c,1:m}, b_{c,1:m}, d_{c,1:m}) = F(\mu_{c,1:m-3}, b_{c,1:m-2}, d_{c,1:m-1}) \tag{11}$$

The choice of the prior for μ does not favor any specific conformation. On the contrary, the prior for the bond angles is meant to incorporate chemical knowledge: for the cyclo-octane dataset treated in Section 5 for example, it is known that bond angles for cyclo-octane are centered around 117 degrees, and present little variability; see for example Reference [10], p36. The indicator notation used in the prior distributions for the bond angles and the distances indicate that we truncate the normal distributions and do not consider values for which the distance to the mean is larger than two standard deviations: we discard outliers for bond angles or distances in order to charge only physically reasonable molecules.

The final model is described through the DAG graph in Figure 2.

4. THE METHODOLOGY

4.1. The Reversible Jump algorithm

We use a MCMC algorithm to obtain samples of the posterior distribution of the parameters. Given a target distribution π, such an algorithm allows to simulate trajectories of an ergodic Markov Chain with stationnary distribution π by specifying how to choose the transition of the chain. The Metropolis Hastings algorithm implements a specific way to construct the transition: given a point of the chain, a

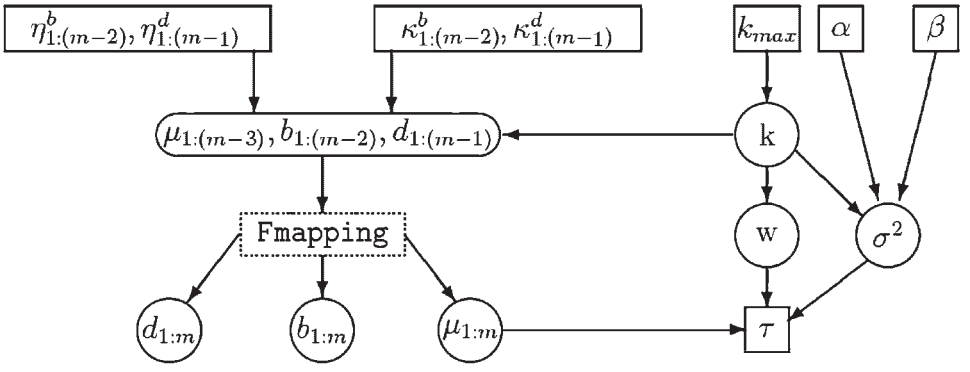

Figure 2. The DAG representation of our model.

candidate is proposed for the next point, from a user-provided candidate distribution and accepted or rejected with a specific probability. For an extensive account on MCMC methods see, for example Reference [11]. The Reversible Jump MCMC was introduced by Reference [4] as a Markov chain algorithm for varying dimension problems. It consists in a random sweep Metropolis-Hastings method adapted for general state space, which is able to jump between spaces of different dimensions. It was applied to carry out Bayesian inference in one-dimensional Gaussian-Mixture models with an unknown number of components in Reference [8] and later in Reference [9], to which we refer for more details about the implementation of the algorithm.

A complete sweep of the algorithm consists in a scanning of the following move types:

(a) updating the weights w
(b) updating the parameters $(\mu_{1:m}, b_{1:m}, d_{1:m})$
(c) updating the variance parameters $\sigma^2_{1:k}$
(d) the birth or death of a component

Notice that the dimension changing move types used for this application consist only in birth/death moves (d), and do not include the split/merge moves considered in Reference [8]. This last kind of move is not easy to implement efficiently in a multidimensional setting, we obtained rather low acceptance rates and decided to discard them in our final analysis. See [12] for a suggestion on how to implement split/merge moves for multivariate mixtures.

At each iteration of the algorithm we either perform the three fixed-k moves (a), (b), and (c) altogether or the birth/death move (d). Each of these two possibilities is decided upon with probability 0.5. Moreover, if we decide to perform move type (d), we choose either birth or death of a component with probability 0.5.

As for move types (a) and (c), we follow Reference [9] and choose multiplicative lognormal random walks for updating the weights and the variances parameters, see also Reference [13] for a detailed description.

In the following Subsection we describe in more detail the updating of the parameters $(\mu_{1:m}, b_{1:m}, d_{1:m})$ in move type (b), which requires involved computations in order to ensure that the proposed parameters correspond to a physically coherent molecule.

4.2. Updating the parameters $(\mu_{1:m}, b_{1:m}, d_{1:m})$

Assume that the previous sample was $(\mu_{1:m}, b_{1:m}, d_{1:m})$, we propose the candidate $(\mu^*_{1:m}, b^*_{1:m}, d^*_{1:m})$, ensuring that it is associated to a physically coherent m-membered ring. To do so, we update $(\mu_{1:m-3}, b_{1:m-2}, d_{1:m-1})$ following a Gaussian Random Walk proposal,

$$\mu^*_{1:m-3} \sim N(\mu_{1:m-3}, \sigma^2_{\epsilon_\mu} \mathbf{I}d_{m-3}) \tag{12}$$

$$b^*_{1:m-2} \sim N(b_{1:m-2}, \sigma^2_{\epsilon_b} \mathbf{I}d_{m-2}) \tag{13}$$

$$d^*_{1:m-1} \sim N(d_{1:m-1}, \sigma^2_{\epsilon_d} \mathbf{I}d_{m-1}) \tag{14}$$

and deduce the remaining parameters $\mu^*_{m-2:m}$, $b^*_{m-1:m}$ and d^*_m using the F mapping described in Appendix A.1.3,

$$(\mu^*_{1:m}, b^*_{1:m}, d^*_{1:m}) = F(\mu^*_{1:m-3}, b^*_{1:m-2}, d^*_{1:m-1})$$

Following the Metropolis Hastings algorithm, we then accept the candidate with probability ρ described in Appendix A.2.

4.3. Birth and death move

We choose with probability 0.5 between the birth or the death of a component. The birth move consists in creating a new component, by first drawing $w^*_{k+1} \sim Be(1, k)$. The candidate weights are then deduced through, for $j = 1, \ldots, k$, $w^*_j = w_j(1 - w^*_{k+1})$. The remaining parameters for the new component $(\mu^*_{k+1,1:m}, b^*_{k+1,1:m}, d^*_{k+1,1:m})$ and $\sigma^{2^*}_{k+1}$ are drawn from their prior distributions, see the Appendix A.2 for details, and appended to the current state parameters.

Finally the death move consists in choosing randomly a component and deleting it from the current state, and renormalizing the weights.

The expression for the acceptance probability for these moves can be found in Subsection A.2.2.

4.4. Post processing of the output

As described in References, for example [14], [15] or [16], the invariance of the likelihood under arbitrary relabeling of the mixture components leads to symmetries in the posterior distributions of the parameters, which are therefore difficult to summarize. In particular, the posterior marginal distributions are, for permutation invariant priors, indistinguishable. One of the proposed solutions consists in imposing identifiability constraints on the parameter space through an ordering specification in the prior, for example on the weights $w_1 < \cdots < w_k$ or in the case of one-dimensional mixtures on the components means; see Reference [8]. However the choice of the identifiability constraints is not unique and, more importantly, this approach does not guarantee an acceptable solution, since the chosen identifiability constraints may not break the symmetry of the likelihood efficiently. Alternatively, post-processing algorithms of the output of the MCMC trajectories, which involve a relabeling of the mixture components, have been investigated; see References [15] or [17].

We follow a post-processing algorithm proposed by Reference [18], which intends to carry out, for each value of the parameters provided by the sampler, a relabeling of the components so that the relabeled sample *points all agree* well with each other. More concretely, if $\theta^{(1)}, \ldots, \theta^{(N)}$ are the N samples of the posterior distributions of the parameters provided by the MCMC algorithm, the aim is to find N permutations $\nu_1, \ldots, \nu_N$ which operate a relabeling of each $\theta^{(1)}, \ldots, \theta^{(N)}$ such that the permuted parameters $\nu_1(\theta^{(1)}), \ldots, \nu_N(\theta^{(N)})$ present the same ordering of the components.

The general algorithm suggested by Stephens requires the definition of a divergence $D(\theta^{(1)} \| \theta^{(2)})$ between two parameters $\theta^{(1)}$ and $\theta^{(2)}$, designed to be small when these are labeled 'the same way'. The algorithm (see Algorithm 3.1, p47, [18]), then consists, starting with some initial values for $\nu_1, \ldots, \nu_N$, in iterating the following steps until a fixed point is reached

Step 1: Choose $\hat{\theta}$ to minimize

$$\sum_{t=1}^{N} D(\nu_t(\theta^{(t)}) \| \hat{\theta})$$

Step 2: For $t = 1, \ldots, N$ choose ν_t to minimize

$$D(\nu_t(\theta^{(t)}) \| \hat{\theta})$$

Stephens suggests the following measure of divergence:

$$D(\theta^{(1)} \| \theta^{(2)}) = \sum_{i=1}^{k} \Delta\left[w_i^{(1)} N\left(.; \mu_i^{(1)}, \Sigma^{(1)}\right) \| w_i^{(2)} N\left(.; \mu_i^{(2)}, \Sigma^{(2)}\right) \right]$$

where Δ is a measure of divergence between two weighted density functions based on the Kulback Leibler divergence. In this case the two minimizations in Steps 1 and 2 above can be carried out explicitly and yield the Algorithm 3.2, p49, in Reference [18].

Notice that in our case, we have to adapt the general algorithm above in order to take into account the possibly arbitrary starting point, direction and sign of each component mean. Our algorithm therefore aims at finding, on the one hand, for each $\theta^{(t)}$, $t = 1, \ldots, N$, an 'optimal' permutation of the components labels, ν_t and, on the other hand, for each component mean, 'optimal' starting points $s_{t,1}, \ldots, s_{t,k}$ in $\{1, \ldots, m\}$, directions $d_{t,1}, \ldots, d_{t,k}$ in $\{-1, 1\}$ and signs $\delta_{t,1}, \ldots, \delta_{t,k}$ in $\{-1, 1\}$. Consider, therefore, the associated operators $T_{t,1} = T^{s_{t,1}, d_{t,1}, \delta_{t,1}}, \ldots, T_{t,k} = T^{s_{t,k}, d_{t,k}, \delta_{t,k}}$, see Equation (2), we form the global operator $T_{t,1:k} = (T_{t,1}, \ldots, T_{t,k})$ which acts block-wise on a vector $\mu_{1:k,1:m}$ of length $k \cdot m$,

$$T_{t,1:k}(\boldsymbol{\mu}_{1:k,1:m}) = (T_{t,1}\boldsymbol{\mu}_{1,1:m}, \ldots, T_{t,k}\boldsymbol{\mu}_{k,1:m})$$

The output of the post-processing algorithm is thus N relabeled samples, where a change is allowed for in the starting points, directions and signs of each conformation; see Section 2 and in particular Equation (2)

$$\theta_*^{(1)} = \left(\nu_1\left(w_{1:k}^{(1)}\right), \nu_1\left(\sigma_{1:k}^{2,(1)}\right), \nu_1\left(T_{1,1:k}\left(\boldsymbol{\mu}_{1:k,1:m}^{(1)}\right)\right) \right)$$

$$\theta_*^{(N)} = \left(\nu_N\left(w_{1:k}^{(N)}\right), \nu_N\left(\sigma_{1:k}^{2,(N)}\right), \nu_N(T_{N,1:k}\left(\boldsymbol{\mu}_{1:k,1:m}^{(N)}\right)\right) \right)$$

Notice that the permutations operators ν_t, $t = 1, \ldots, N$ act on a k-dimensional vector $\boldsymbol{v}_{1:k}$. However, the same notation is used for the operators that permute the k blocks of a $(k \cdot m)$-dimensional vector $\boldsymbol{v}_{1:k;1:m}$: $\mathbf{v}_t(v_{1:k;1:m}) = v_{\nu_t(1:k);1:m}$.

Before describing the post-processing algorithm, we introduce a last notation; for $i = 1, \ldots, k$, consider the operator p_i that picks up the ith m-dimensional block, in a $(k \cdot m)$-dimensional vector $\boldsymbol{\mu}_{1:k,1:m}$:

$$p_i(\boldsymbol{\mu}_{1:k,1:m}) = (\boldsymbol{\mu}_{i,1:m})$$

To sum up, notice in particular that $p_i(\nu_t(T_{t,1:k}(\boldsymbol{\mu}_{1:k,1:m}^{(t)})))$ is the mean of the ith -component in the relabeled $\theta^{(t)}$, where the starting point, direction and sign have been changed according to $T_{t,i}$.

Algorithm 1 below is summarizing the suggested post-processing algorithm, inspired by Reference [18], p49.

5. RESULTS

5.1. Simulated data

In this subsection, we test the performance of our method on a simulated dataset. Concretely, a three components dataset is simulated, where the three eight-membered conformations correspond to the CR, BB and TC conformations (see Figure 1) present for cyclo-octane, for which we can find the torsion angles in Reference [2]. The dataset was generated from model Equation (1) (see also Equation (3)), where the standard deviations of the perturbations were assumed to be $\sigma_1 = \sigma_2 = \sigma_3 = 10$ degrees. The total number of structures in the dataset is 60, with 10 CR structures, 20 BC structures and 30 TC structures.

In order to evaluate the benefit of taking into account the physical restrictions, through the extension of the parameter space as described in Subsection 3.3, we have also

---

Algorithm 1 Relabeling the N samples associated to k components mixture

1: Set all permutations $\nu_1, \ldots, \nu_N$ to be identity permutations and operators $T_{1,1:k}, \ldots, T_{N,1:k}$ to be identity operators.

2: Set

$$\text{for all } 1 \le i \le k, \quad \hat{w}_i = \frac{1}{N} \sum_{t=1}^{N} \nu_t\left(w_{1:k}^{(t)}\right)[i]$$

$$\text{for all } 1 \le i \le k \quad \hat{\boldsymbol{\mu}}_{i,1:m} = \frac{\sum_{t=1}^{N} \nu_t\left(w_{1:k}^{(t)}\right)[i] p_i\left(\nu_t\left(T_{t,1:k}\left(\boldsymbol{\mu}_{1:k,1:m}^{(t)}\right)\right)\right)}{\sum_{t=1}^{N} \nu_t(w_{1:k}^{(t)})[i]}$$

$$\text{for all } 1 \le i \le k, \quad \hat{\sigma}_i^2 = \frac{\sum_{t=1}^{N} \nu_t\left(w_{1:k}^{(t)}\right)[i]\left(m\left(\sigma_{\nu_t[i]}^{(t)}\right)^2 + \| p_i\left(\nu_t\left(T_{t,1:k}\left(\boldsymbol{\mu}_{1:k,1:m}^{(t)}\right)\right)\right) - \hat{\boldsymbol{\mu}}_{i,1:m} \|^2\right)}{m \sum_{t=1}^{N} \nu_t(w_{1:k}^{(t)})[i]}$$

3: For each $t = 1, \ldots, N$ choose the optimal permutation ν_t and the optimal operator T_t of the form Equation (2), minimizing

$$\begin{aligned} \sum_{i=1}^{k} \Big\{ & -\nu_t(w_{1:k})[i] \log(\hat{w}_i) - (1 - \nu_t(w_{1:k})[i]) \log(1 - \hat{w}_i) \\ & + \nu_t(w_{1:k})[i] \frac{m}{2} \log(\hat{\sigma}_i^2) . + \frac{w_{\nu_t[i]}}{2\hat{\sigma}_i^2} \left(m\sigma_{\nu_t[i]}^2 + \| p_i(\nu_t(T_{t,1:k}(\boldsymbol{\mu}_{1:k,1:m}))) - \hat{\boldsymbol{\mu}}_{i,1:m} \|^2 \right) \Big\} \end{aligned} \tag{15}$$

4: Stop when a fix point is reached, otherwise return to step (2).

implemented the naive model that would consist in ignoring these physical restrictions and the relations between the eight torsion angles and, in the MCMC algorithm, simply update $\mu_{1:m}$ following a Gaussian Random Walk proposal,

$$\mu^*_{1:m} \sim N(\mu_{1:m}, \sigma^2_{\epsilon_\mu} \mathbf{I}d_m) \tag{16}$$

as opposed to Equations (12)–(14).

As for the prior distributions, both models share the specification for parameters k, $\sigma^2_{1:k}$ and $w_{1:k}$; see Subsection 3.4.1. More concretely, $k_{\max} = 15$, $\sigma^2_1, \ldots, \sigma^2_k$ are i.i.d IG$(2, 40)$ which centers σ_i around 10 degrees. However, they differ in the specification of the prior for $\mu_{1:m}$.

For the full model, as described in Subsection 3.4.2, we have chosen $\mu_{1:(m-3)} \sim \otimes_{1:(m-3)} U(-\pi, \pi)$, which does not favor any conformation. The prior distribution for the bond angles was based on chemical knowledge; see Reference [10], p36, $b_{1:(m-2)} \sim \otimes_{1:(m-2)} N(117, 3^2)$. Finally we know that all atoms in the chain are the same, and we therefore assume that the distances are basically equal: $d_{1:(m-1)} \sim \otimes_{1:(m-1)} N(1, .1^2)$.

For the naive model, we have simply specified $\mu_{1:m} \sim \otimes_{1:m} U(-\pi, \pi)$, it is a prior distribution that charges also non-physically drawable molecules.

The MCMC Reversible Jump algorithm was run for 2 02 000 iterations and we kept the last 2000 points for inference. The value of the constants involved in the proposal steps were tuned so that we obtained the following reasonable acceptance rates for the different move types: for the fixed-dimension moves (a), (b), and (c) in Subsection 4.1, the acceptance rate was close to 0.5, while the birth or death moves presented an acceptance rate of approximately 6%. The mixing of the chain was satisfactory and a shorter burn-period could as well have been chosen. Our attempts with a shorter burn-in period yielded in fact similar results.

The posterior distribution for k, the number of components in the mixture, is presented in Figure 3. They both clearly favor three components in the mixture.

In Table I, the first rows contain the true values of the parameters of interest for the simulated dataset: N_c is the number of structures simulated from each conformation, w are the corresponding proportions, σ represents the standard deviation of the perturbation, while $\mu_1, \ldots, \mu_8$ are the theoretical torsion angles for the CR, BR, and TC conformations reported in Reference [2].

The estimation results for the full model, which take into account the physical restrictions through the F-mapping, as described in Subsection 4.2, are presented in the middle of the table, where the medians of the posterior distributions of the parameters $w_{1:3}$, $\sigma_{1:3}$ and $\mu_{1:3,1:8}$ are reported. Finally, the same summary for the naive model, which does not take into account the physical restrictions but uses the proposal (16), is presented in the bottom of Table I.

Figures 4 and 5 present box plot-like diagrams for the posterior distributions of the parameters of interest for both methods. Seven horizontal lines are drawn for each box, representing the percentiles 10, 15, 25, 50, 75, 85, 90. Moreover wider, thicker horizontal lines are drawn for the values of the parameter that were used to generate the simulated dataset.

As a conclusion, on this simulated dataset, the results for the full model described in this paper are very satisfying. The three components are detected and correctly estimated. On the other hand, the results are significantly better for the full model than for the naive model: the standard deviations are much better estimated and the dispersion for the posterior distribution of the means is smaller. We conclude that, even if the full model requires more involved computations, because of the evaluation of the F mapping, the improvement is worth the increase in sophistication.

5.2. Cyclo-octane dataset

In Reference [5] measurements of torsion angles for cyclo-octane were retrieved from the CSD and principal component analysis and cluster analysis were carried out. They consider in particular a dataset, labeled 8C1 in their

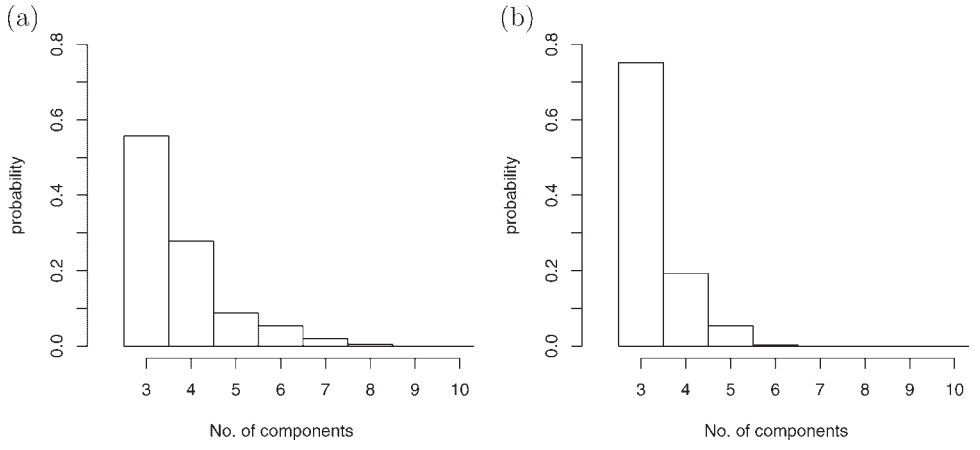

Figure 3. Posterior distribution for k, the number of components for the simulated dataset of Subsection 5.1. (a) and (b) correspond respectively to the full and naive model, that is with/without the F mapping and the physical restrictions.

Table I. True values of the parameters, and estimation results for the simulated dataset of Subsection 5.1.

N_c	w_c	σ_c	Identifiers	$\mu_{c,1}$	$\mu_{c,2}$	$\mu_{c,3}$	$\mu_{c,4}$	$\mu_{c,5}$	$\mu_{c,6}$	$\mu_{c,7}$	$\mu_{c,8}$
True values of the parameters used in the simulation of the dataset											
30	0.5	10	TC	37.3	−109.3	109.3	−37.3	−37.3	109.3	−109.3	37.3
20	0.333	10	BB	52.5	52.5	−52.5	−52.5	52.5	52.5	−52.5	−52.5
10	0.167	10	CR	87.5	−87.5	87.5	−87.5	87.5	−87.5	87.5	−87.5
Estimation results for the full model which involves the F mapping to take into account the physical restrictions											
	0.49	16.61	TC	31.67	−108.48	110.61	−36.54	−35.55	106.10	−107.96	40.62
	0.34	11.71	BB	48.46	57.55	−53.30	−53.65	53.21	58.01	−57.56	−49.41
	0.16	21.5	CR	80.94	−77.00	87.22	−95.77	90.13	−83.93	86.97	−86.60
Estimation results for the 'naive' model, that does not take into account the physical restrictions											
	0.49	34.67	TC	41.01	−109.55	108.76	−39.50	−36.55	107.60	−111.82	33.90
	0.33	42.09	BB	52.92	46.60	−59.44	−46.29	49.08	56.51	−48.31	−49.96
	0.18	61.02	CR	94.67	−71.32	107.79	−78.33	101.15	−98.18	75.83	−76.00

The upper part of the table contains the values of the parameters that were used to simulate the dataset. The column labeled N_c contains the number of structures that were simulated for each cluster, while w_c denotes the corresponding proportion. The dataset was simulated from a mixture of three conformations TC (Twisted-Chair), BB (Boat–Boat) and CR (Crown). The corresponding torsion angles are reported in the columns $\mu_{c,1}$ to $\mu_{c,8}$, while σ_c denotes the standard deviation associated to each cluster. The middle part of the table contains the medians of the posterior distributions of the parameters of interest, when the full model is used and the physical restrictions are taken into account. The bottom part of the table contains the medians of the posterior distributions of the parameters when the simple, 'naive' is used, and the physical restrictions are not taken into account

paper, consisting of 31 observations, of which 12 were classified as Boat–Chair (BC), 10 as two different components both identified as 'deformed' BC, (in between BC and Twist–Boat–Chair (TBC)) and finally 2 observations were classified as respectively Crown (CR) and Twist–Chair–Chair (TCC); 7 observations were not classified. The standard deviations within the clusters seem to be centered around 15 degrees. The results from Reference [5] are reported in Table 2: the first column contains the index of the cluster, the second column the number of elements in each cluster, the Refcode is the reference code of the structure chosen as most representative of the cluster, while the identifier column contains the kind of conformation associated to the cluster by the authors. Finally the columns $\mu_{c,1}, \ldots, \mu_{c,8}$ contain the torsion angles of the representative member of the cluster, that is the structure corresponding to the Refcode of the third column.

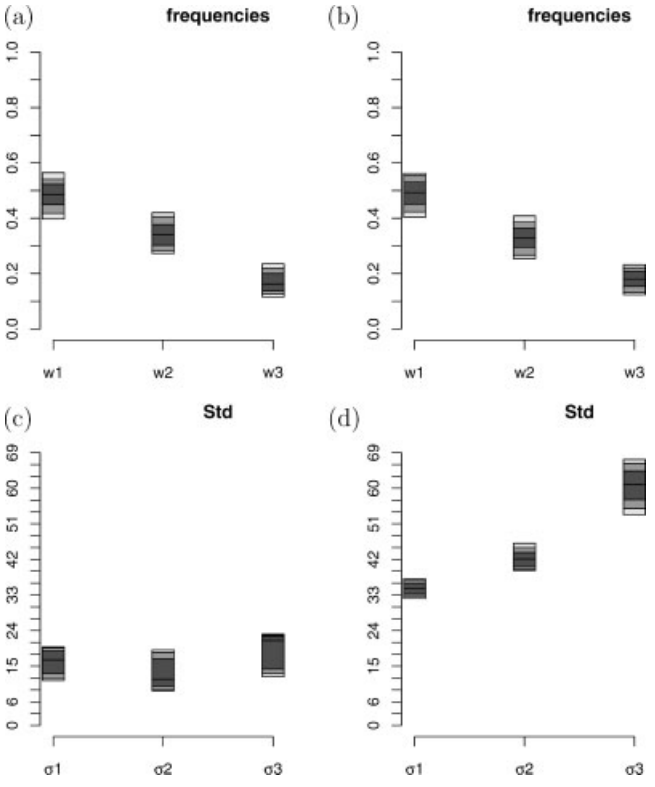

Figure 4. Simulated dataset of Subsection 5.1: Box plot-like diagrams of the posterior distributions of the proportions $w_{1:3}$ and the standard deviations $\sigma_{1:3}$ associated to the $k = 3$ detected components. The plots (a) and (c) correspond to the case when the full model is used, while the plots (b) and (d) to the case when the 'naive' model is used. Seven horizontal lines are drawn for each box, representing the percentiles 10, 15, 25, 50, 75, 85, 90.

These results will now be compared to the ones obtained by applying the method described in this paper. The prior distributions for the parameters were the same as in the previous subsection for the simulated dataset (full model), and the constants needed in the updating steps of the MCMC algorithm were tuned to obtain similar acceptance rates.

The posterior distribution for the number k of components in the mixture is presented in Figure 6 from which we deduce that $k = 6$ or $k = 7$ is the most likely value. The post-processing algorithm was therefore carried out for both values of k and no significant differences were found concerning the main groups. We present here the results obtained with $k = 7$.

In Table III the mean values are listed for comparison with Table II and in Figure 7 box plots of the posterior distributions are presented, where the 10, 15, 25, 50, 75, 85 and 90 percentiles are indicated. In the box plots where the conformations are identified as being equivalent to a conformation obtained by Reference [5] (see Table II) a horizontal black wide line has been drawn, which represents the value of a representative member of the corresponding cluster according to these authors.

Our results coincide basically with Reference [5]: the preferred BC conformation is clearly identified (component

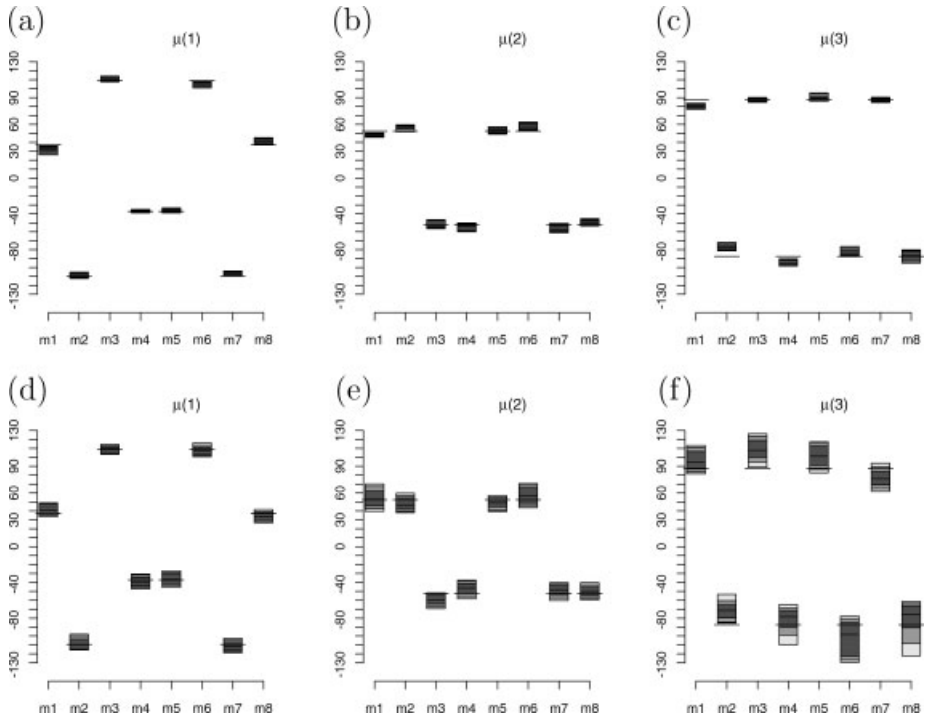

Figure 5. Simulated dataset of Subsection 5.1: Box plot-like diagrams of the posterior distributions of the components means $\mu_{c,1:8}$, for $c = 1, 2, 3$. The plots (a), (b) and (c) correspond to the case when the full model is used, while the plots (d), (e) and (f) to the case when the 'naive' model is used. Seven horizontal lines are drawn for each box, representing the percentiles 10, 15, 25, 50, 75, 85, 90. Moreover wider, thicker horizontal lines are drawn for the values of the parameter that were used to generate the simulated dataset.

Table II. Results reported in Reference [5], for the cyclo-octane dataset under study.

c	N_c	w_c	Refcode	Identifiers	$\mu_{c,1}$	$\mu_{c,2}$	$\mu_{c,3}$	$\mu_{c,4}$	$\mu_{c,5}$	$\mu_{c,6}$	$\mu_{c,7}$	$\mu_{c,8}$
1	12	0.39	BAGPII	BC	−100.9 (10)	43.2 (11)	65.0 (8)	−65.0 (8)	−43.2 (11)	100.9 (10)	−66.8 (10)	66.8 (10)
2	6	0.19	COVLUU	BC/TBC	−91.0 (9)	24.3 (18)	77.2 (10)	−57.3 (10)	−53.4 (7)	103.1 (13)	−65.3 (12)	68.9 (8)
3	4	0.13	SPTZBN	BC/TBC	−79.5 (4)	0.5 (21)	89.6 (24)	−52.4 (7)	−56.4 (13)	102.6 (32)	−71.2 (15)	76.3 (36)
4	1	0.03	DEZPUT	CR	70.4	−83.2	92.3	−73.3	63.8	−82.3	96.5	−82.0
5	1	0.03	EOCNON10	TCC	47.7	−84.7	134.4	−85.3	48.7	−82.4	124.9	−80.7

For each component index c, the number N_c of structures assigned to the component c and the corresponding proportion w_c are given. The refcode of the most representative structure of the cluster is reported, together with its observed torsion angles in the columns labeled $\mu_{c,1}$–$\mu_{c,8}$. The identifiers column represents which kind of conformation present the structures in the cluster. The number in parenthesis for the first three components represent the empirical standard deviations for each cluster

Table III. Cyclo-octane dataset of Subsection 5.2: medians of the posterior distributions for the proportions w_c, the standard deviations σ_c and the components means $\mu_{c,1}$–$\mu_{c,8}$ when $k = 7$ components are chosen

c	w_c	σ_c	Identifiers	$\mu_{c,1}$	$\mu_{c,2}$	$\mu_{c,3}$	$\mu_{c,4}$	$\mu_{c,5}$	$\mu_{c,6}$	$\mu_{c,7}$	$\mu_{c,8}$
1	0.49	9.82	BC	−98.9	39.3	67.8	−63.2	−46.0	100.8	−66.0	66.7
2	0.06	9.07	BC/TBC	−97.5	54.9	53.6	−82.4	−8.1	81.1	−71.3	66.8
3	0.11	9.26	BC/TBC	−82.5	5.1	85.8	−53.9	−57.7	100.8	−69.6	72.3
4	0.08	11.03	Crown	73.5	−94.4	92.6	−70.2	69.9	−87.2	87.7	−68.7
5	0.04	12.51	TCC	68.6	−101.0	87.5	−49.1	54.0	−93.7	89.1	−57.4
6	0.10	12.52		1.7	−61.9	101.8	−67.3	60.3	−101.0	83.8	−5.8
7	0.07	12.13		88.5	3.6	−74.5	89.8	−88.1	70.8	−4.9	−88.5

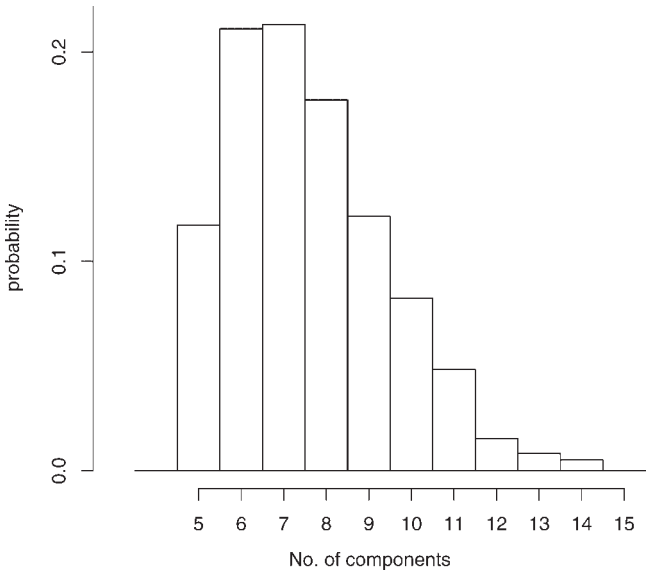

Figure 6. Cyclo-octane dataset of Subsection 5.2: posterior distribution for k the number of components.

1 in our results), and our conformations 2, 3, and 4 are similar to their clusters with refcodes COVLUU, SPTZBN, and DEZPUT (see Table II). As for the three remaining components the high dispersion of the posterior distribution for the standard deviations and the corresponding low frequencies lead us to be cautious as for their interpretation. Component no. 5, however, is somehow similar to TCC identified in Reference [5]. Notice that this dataset is particularly difficult to process since there are few observations and some of the clusters only contain one observation.

6. CONCLUSION

In this work a full Bayesian approach to conformational classification of m-membered rings is proposed, based on torsion angles measurements. We have proposed a multivariate mixture model for the data generation mechanism. We can see several advantages of a full Baye-

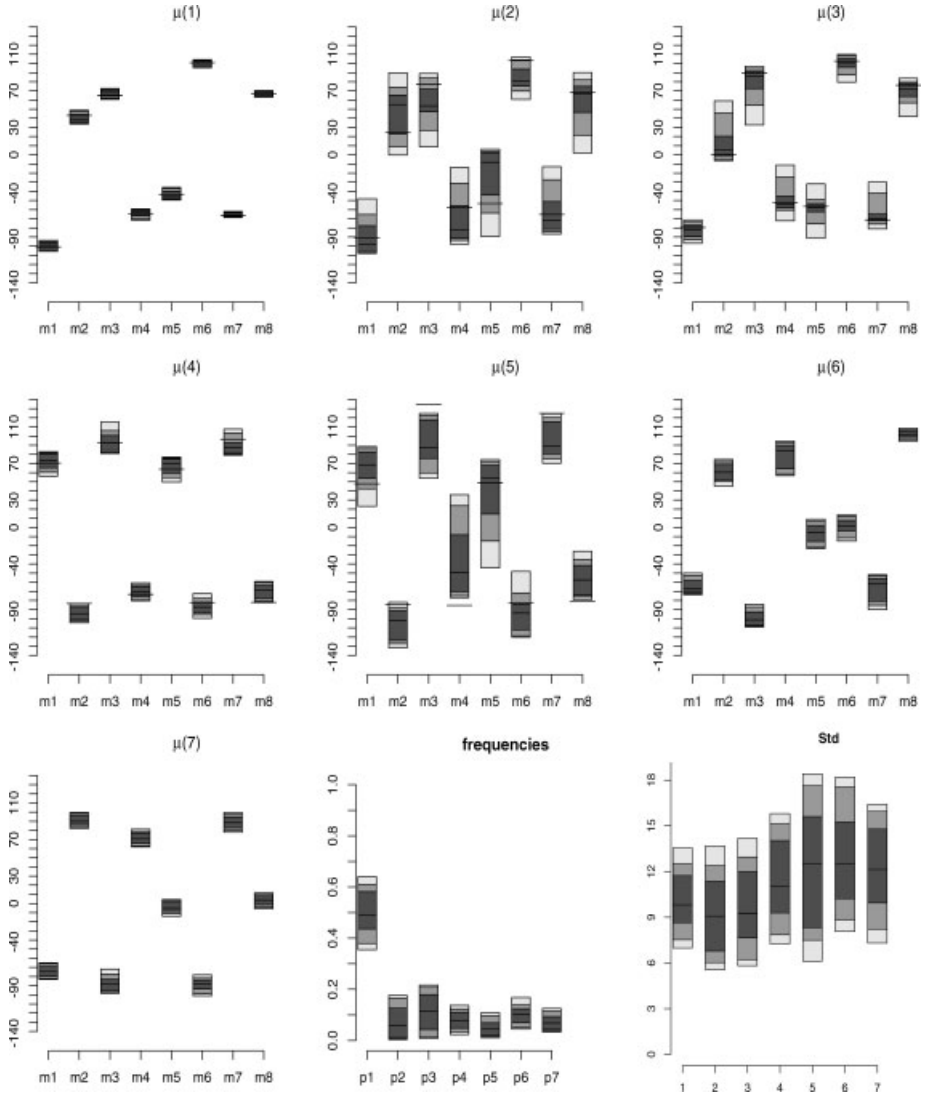

Figure 7. Cyclo-octane dataset of Subsection 5.2: Box plot-like diagrams of the posterior distributions of the means, frequencies and standard deviations associated to $k = 7$ components. For comparison with the results reported in Reference [5], see Table 2, when appropiate, a horizontal black wide line has been drawn, which represents the value of a representative member of the corresponding cluster according to these authors.

sian approach to this applied problem: on the one hand, it easily allows to take into account, through an extension of the parameter space, the physical restrictions and the intricate relations these imply between the torsion angles. On the other hand, we have been able to include chemical knowledge as part of the prior specification. Moreover, the output includes a measure of the uncertainty linked to the inference procedure, which represents a useful information for chemists.

The implementation of the Reversible Jump algorithm, combined with a post-processing of the sampler output, provided quite satisfying results both for a simulated dataset and for a real dataset of cyclo-octane structures, though we did not consider a split/merge move as in Reference [8], since it turned out to be difficult to implement.

Acknowledgements

This work was supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2000-00100, DYNSTOCH. The authors are grateful to Tobias Rydén for helpful discussions.

APPENDIX

A.1. Relations between torsion angles, bond angles and distances in an *m*-membered ring

A chemical structure can be described in terms of (xyz) positions of its atoms in a Cartesian coordinate system. This description is, however, not particularly useful since the absolute coordinates are irrelevant from a chemical point of view, and instead, a description in terms of distances, bond- and torsion-angles is preferred, which moreover has the property of being invariant under rotations and translations of the structure. However for a structure consisting of m atoms, an arbitrary specification of m distances $d_{1:m}$, m bond angles $b_{1:m}$ and m torsion angles $\mu_{1:m}$ may lead to a structure with inner conflicts since the structure is uniquely described with fewer parameters.

The aim of this subsection is to describe the relations between torsion angles, bond angles and distances for an m-membered ring. An understanding of these relations allow us to suggest priors on $(\mu_{1:m}, b_{1:m}, d_{1:m})$, see Subsection 3.4.2, that only charge physically coherent molecules, or to produce samples of these parameters that fulfill the physical restrictions, see Subsection 4.2.

Some background and definitions are needed. We begin in Subsection A.1.1 by defining the concept of torsion and bond angles. In Subsection A.1.2, we explain what are the angles and distances required to build sequentially an m-membered ring. In Subsection A.1.3 we deduce an sequential procedure that expresses the coordinates, in one reference Cartesian system, of all m atoms in terms of the torsion angles, bond angles and distances and obtain in that way an operational description of the mapping F introduced in Equation (7).

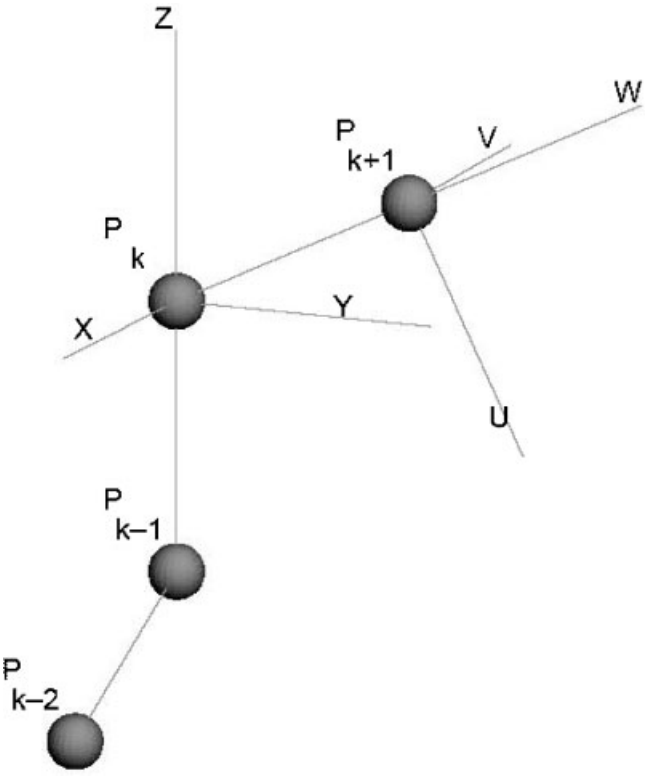

Figure 8. Two local Cartesian coordinate system are displayed. System $\mathfrak{S}_k$ determined by P_{k-2}, P_{k-1} and P_k and system $\mathfrak{S}_{k+1}$ determined by P_{k-1}, P_k and P_{k+1}.

A.1.1. Definitions and notation

Consider, for a given Cartesian system, four distinct points $P_{k-2}, P_{k-1}, P_k, P_{k+1} \in \mathbb{R}^3$. Introduce the vectors $\vec{a}, \vec{b}$ and $\vec{c}$

$$\vec{a} = \frac{P_{k-1} - P_{k-2}}{\|P_{k-1} - P_{k-2}\|}$$
$$\vec{b} = \frac{P_k - P_{k-1}}{\|P_k - P_{k-1}\|}$$
$$\vec{c} = \frac{P_{k+1} - P_k}{\|P_{k+1} - P_k\|}$$

see Figure 8.

DEFINITION A.1. *The torsion angle or dihedral angle* $\mathrm{Tor}(P_{k-2}, P_{k-1}, P_k, P_{k+1})$ *for a sequence of four points* $P_{k-2}, P_{k-1}, P_k, P_{k+1}$ *is, expressed in terms of* $\vec{a}, \vec{b}$ *and* $\vec{c}$,

$$\mathrm{Tor}(P_{k-2}, P_{k-1}, P_k, P_{k+1}) = \arg\left(\begin{pmatrix} -\vec{a}\cdot\vec{c} + \left(\vec{a}\cdot\vec{b}\right)\left(\vec{b}\cdot\vec{c}\right) \\ \vec{a}\cdot\left(\vec{b}\times\vec{c}\right) \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$$

where the arg *function denotes the planar angle between two vectors and takes its values in* $(-\pi, \pi]$. *In this Appendix, we will use the shorter notation* $\tau_{k-2;\,k-1;\,k;\,k+1}$ *for* $Tor(P_{k-2}, P_{k-1}, P_k, P_{k+1})$.

Note that the torsion angle between $\vec{a}, \vec{b}$ and $\vec{c}$ is undefined if ($\vec{b}$ and $\vec{c}$) or ($\vec{a}$ and $\vec{b}$) are parallel.

For an intuitive geometric interpretation of the torsion angle $\tau_{k-2;\,k-1;\,k;\,k+1}$ consider two planes A and B spanned by respectively P_{k-2}, P_{k-1}, P_k and P_{k-1}, P_k, P_{k+1}, the torsion angle

$\tau_{k-2;\,k-1;\,k;\,k+1}$ unites the two planes by rotating the plane A around $\vec{b}$ by an angle $\tau_{k-2;\,k-1;\,k;\,k+1} \in (-\pi;\ \pi]$.

DEFINITION A.2. *The bond angle* $b_{k-1;\,k;\,k+1}$ *between* P_{k-1}, P_k, P_{k+1} *is the numerical value of the planar angle* $\angle_{P_{k-1}P_kP_{k+1}}$, *which expressed in terms of* $\vec{b}$ *and* $\vec{c}$ *is*

$$b_{k-1;\,k;\,k+1} = \mathrm{Bind}(P_{k-1}, P_k, P_{k+1}) = |\arg\left(\vec{b}, \vec{c}\right)|$$

DEFINITION A.3. *The distance* $d_{k-1;\,k}$ *between* P_{k-1}, P_k *is the euclidian distance* $\|P_k - P_{k-1}\|$.

A.1.2. Sequential specification of a molecule through torsion angles, bond angles and distances

Assume we want to build an m-membered ring, how many torsion angles, bond angles and distances do we have to specify? To define the relative positions of the first two atoms, it is enough to specify the distance $d_{1;\,2}$. The position of the third atom will be fixed if we specify further $d_{2;\,3}$ and the bond angle $b_{1;\,2;\,3}$, while to position the fourth atom we will need $d_{2;\,4}$, $b_{2;\,3;\,4}$ and the torsion angle $\tau_{1;\,2;\,3;\,4}$. From there on, an additional atom requires the specification of one more distance, one more bond angle and one more torsion angle. The procedure is summarized in

$$P_1 \xrightarrow{d_{1;\,2}} P_2 \xrightarrow[\ ]{d_{2;\,3},\ b_{1;\,2;\,3}} P_3 \xrightarrow[[\tau_{1;\,2;\,3;\,4}]]{d_{3;\,4},\ b_{2;\,3;\,4}} P_4 \rightarrow \ldots \xrightarrow[[\tau_{k-3;\,k-2;\,k-1;\,k}]]{d_{k-1;\,k},\ b_{k-2;\,k-1;\,k}} P_k \rightarrow \ldots$$

As a conclusion, we can build sequentially an m-membered ring if we specify the first $m-1$ distances, $m-2$ bond angles and $m-3$ torsion angles.

A.1.3. Description of the F mapping

The goal of this subsection is to obtain a description of the F mapping that relates the torsion angles, bond angles and distances of an m-membered ring through

$$(\tau_{1:m}, b_{1:m}, d_{1:m}) = F(\tau_{1:m-3}, b_{1:m-2}, d_{1:m-1})$$

For any $k = 3, \ldots, m$, we associate to the sequence of three points P_{k-2}, P_{k-1} and P_k, the local Cartesian system $\mathfrak{S}_k$, with an orthonormal basis $\vec{x}^{\mathfrak{S}_k}$, $\vec{y}^{\mathfrak{S}_k}$ and $\vec{z}^{\mathfrak{S}_k}$, such that the origin is in P_k, the vectors $\overrightarrow{P_{k-1}P_k}$ and $\vec{z}^{\mathfrak{S}_k}$ are parallel and pointing in the same direction, and such that $(\overrightarrow{P_{k-2}P_{k-1}} \times \overrightarrow{P_{k-1}P_k})$ and $\vec{y}^{\mathfrak{S}_k}$ are parallel and pointing in the same direction. In Figure 8 the local Cartesian coordinate system $\mathfrak{S}_k$ is specified by (X, Y, Z).

The description of F will involve the characterization of the coordinates of the m atoms of the ring in $\mathfrak{S}_m$, in terms of the first $m-1$ distances, $m-2$ bond angles and $m-3$ torsion angles.

PROPOSITION A.1. *The coordinates of* $P_1, \ldots, P_m$ *in* $\mathfrak{S}_m$ *can be obtained sequentially by the following algorithm*

Initialization: *Consider the coordinates of* $P_1,\ P_2$ *and* P_3 *in* $\mathfrak{S}_3$:

$$P_1^{\mathfrak{S}_3} = \left(d_{1;\,2}\sin\left(b_{1;\,2;\,3}\right),\ 0,\ d_{1;\,2}\cos\left(b_{1;\,2;\,3}\right) - d_{2;\,3}\right)'$$
$$P_2^{\mathfrak{S}_3} = \left(0,\ 0,\ -d_{2;\,3}\right)'$$
$$P_3^{\mathfrak{S}_3} = (0,\ 0,\ 0)'$$

For $k = 3$ **to** $m-1$, **do:**] *From the coordinates* $(P_1^{\mathfrak{S}_k}, \ldots, P_k^{\mathfrak{S}_k})$ *of* $P_1, \ldots, P_k$ *in* $\mathfrak{S}_k$, *deduce the coordinates* $(P_1^{\mathfrak{S}_{k+1}}, \ldots, P_k^{\mathfrak{S}_{k+1}})$ *of* $P_1, \ldots, P_k$ *in* $\mathfrak{S}_{k+1}$, *through the relation, for* $i = 1, \ldots,\ k$,

$$P_i^{\mathfrak{S}_{k+1}} = \mathfrak{T}_{\mathfrak{S}_k,\mathfrak{S}_{k+1}}\left(P_i^{\mathfrak{S}_k}\right) = B_{b_{k-1;\,k;\,k+1}} A_{\tau_{k-2;\,k-1;\,k;\,k+1}} P_i^{\mathfrak{S}_k} + C_{d_{k;\,k+1}} \tag{17}$$

with

$$A_{\tau_{k-2;\,k-1;\,k;\,k+1}} = \begin{bmatrix} \cos\left(\tau_{k-2;\,k-1;\,k;\,k+1}\right) & -\sin\left(\tau_{k-2;\,k-1;\,k;\,k+1}\right) & 0 \\ \sin\left(\tau_{k-2;\,k-1;\,k;\,k+1}\right) & \cos\left(\tau_{k-2;\,k-1;\,k;\,k+1}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

whereas

$$B_{b_{k-1;\,k;\,k+1}} = \begin{bmatrix} -\cos\left(b_{k-1;\,k;\,k+1}\right) & 0 & -\sin\left(b_{k-1;\,k;\,k+1}\right) \\ 0 & 1 & 0 \\ \sin\left(b_{k-1;\,k;\,k+1}\right) & 0 & -\cos\left(b_{k-1;\,k;\,k+1}\right) \end{bmatrix}$$

and $C_{d_{k;\,k+1}} = (0,\ 0,\ -d_{k;\,k+1})'$. *Moreover set* $P_{k+1}^{\mathfrak{S}_{k+1}} = (0,0,0)'$.

Proof. Proposition A.1 is easily deduced from the following lemma:

LEMMA A.2. *Let* P *a point of* $\mathbb{R}^3$, *its coordinates in* $\mathfrak{S}_k$ and $\mathfrak{S}_{k+1}$ *are related through*

$$P^{\mathfrak{S}_{k+1}} = \mathfrak{T}_{\mathfrak{S}_k,\mathfrak{S}_{k+1}}\left(P^{\mathfrak{S}_k}\right) = B_{b_{k-1;\,k;\,k+1}} A_{\tau_{k-2;\,k-1;\,k;\,k+1}} P_i^{\mathfrak{S}_k} + C_{d_{k;\,k+1}} \tag{18}$$

where $A_{\tau_{k-2;\,k-1;\,k;\,k+1}}$, $B_{b_{k-1;\,k;\,k+1}}$ *and* $C_{d_{k;\,k+1}}$ *are defined in Proposition A.1.*

Proof. From the definition of $\mathfrak{S}_k$ and $\mathfrak{S}_{k+1}$ it is easily checked that

$$\begin{aligned} \arg\left(\vec{z}^{\mathfrak{S}_k}, \vec{z}^{\mathfrak{S}_{k+1}}\right) &= \pi - b_{k-1;\,k;\,k+1} \\ \arg\left(\vec{y}^{\mathfrak{S}_k}, \vec{y}^{\mathfrak{S}_{k+1}}\right) &= \tau_{k-2;\,k-1;\,k;\,k+1} \end{aligned},$$

see Figure 8, where $\mathfrak{S}_k$ is indicated with X, Y, Z and $\mathfrak{S}_{k+1}$ indicated with U, V, W. Hence we can transform $\mathfrak{S}_{k+1}$ into $\mathfrak{S}_k$, by the rotation of angle $-\tau_{k-2;\,k-1;\,k;\,k+1}$ around $\vec{z}^{\mathfrak{S}_k}$, next by the rotation of angle $-(\pi - b_{k-1;\,k;\,k+1})$ around $\vec{y}^{\mathfrak{S}_k}$ and finally by the translation of vector $(0, 0, -d_{k;\,k+1})'$. These transformations are presented graphically in Figure 9. Let us denote by $\mathfrak{T}_{\mathfrak{S}_k,\mathfrak{S}_{k+1}}$ the resulting transformation, composition of the two rotations and the translation. It is easy to check that, since the angles and distances are preserved by rotations and translations, the coordinates of a given point P in $\mathfrak{S}_{k+1}$ are equal to the coordinates of its image $\mathfrak{T}_{\mathfrak{S}_k,\mathfrak{S}_{k+1}}(P)$ in $\mathfrak{S}_k$. We deduce Equation (18) after checking that $A_{\tau_{k-2;\,k-1;\,k;\,k+1}}$, $B_{b_{k-1;\,k;\,k+1}}$ and $C_{d_{k;\,k+1}}$ are the matrices of the involved rotations and translation.

We are now able to provide an operational description of F which will suit our purposes:

Starting from the first $m-3$ torsion angles $\mu_{1:m-3}$, the first $m-2$ bond angles $b_{1:m-2}$ and the first $m-1$ distances $d_{1:m-1}$ that define a ring $A_1, A_2, \ldots, A_m$ (the convention used to number the torsion-, bond angles and distances in such a ring is described in (4)–(6)), use the algorithm described in Proposition A.1 to obtain the coordinates of the m atoms $A_1, A_2, \ldots, A_m$ in $\mathfrak{S}_m$. Deduce from these coordinates the remaining

$$\begin{aligned} \mu_{m-2:m} &= \left(\tau_{m-2;\,m-1;\,m;\,1}, \tau_{m-1;\,m;\,1;\,2}, \tau_{m;\,1;\,2;\,3}\right) \\ b_{m-1:m} &= \left(b_{m-1;\,m;\,1}, b_{m;\,1;\,2}\right) d_m = d_{m;\,1} \end{aligned}$$

This provides

$$(\mu_{1:m}, b_{1:m}, d_{1:m}) = F(\mu_{1:m-3}, b_{1:m-2}, d_{1:m-1}) \tag{19}$$

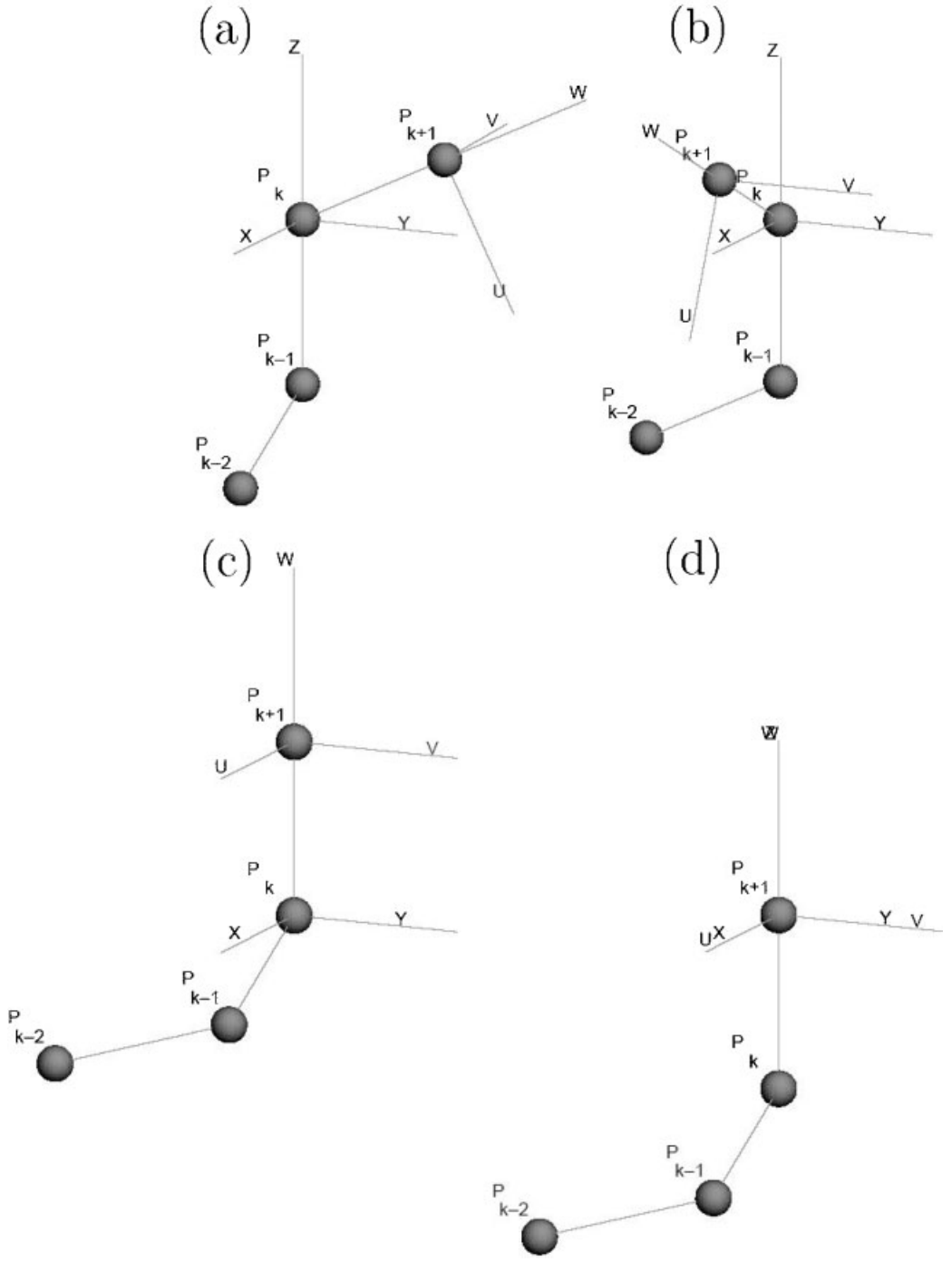

Figure 9. The decomposition of the transformation $\mathfrak{T}_{\mathfrak{S}}k, \mathfrak{S}_{k+1}$ of a structure $P_{k-2}, P_{k-1}, P_k, P_{k+1}$, described in the proof of Lemma A.2: starting from (a), the structure is rotated around $\vec{z}^{\mathfrak{S}_k}$ by $-\tau_{k-2;\,k-1;\,k;\,k+1}$, resulting in (b), next rotated by angle $-(\pi - b_{k-1;\,k;\,k+1})$ around $\vec{y}^{\mathfrak{S}_k}$ resulting in (c) and finally translated to the origin of S_k, (d).

A.2. Acceptance probabilities

We present the log acceptance probabilities $\log(\rho)$ for the different move-types in the Reversible Jump implementation for the m-membered ring structures application. The expressions follow to some extent the presentation from Reference [13].

Let l, p and f denote the likelihood, the prior- and proposal-distributions respectively. θ^* indicates the proposed state and θ is the current state of the sampler. The probability of accepting the value θ^* given that we are currently in the state θ is given by $\rho = \min(r, 1)$, where

$$\begin{aligned}\log(r) = {} & \log l(\tau_{1:n} \mid \theta^*) + \log(p(\theta^*)) - \log(f(\theta^* \mid \theta)) \\ & - \log l(\tau_{1:n} \mid \theta) - \log(p(\theta)) + \log(f(\theta \mid \theta^*))\end{aligned} \tag{20}$$

The Subsections A.2.1 and A.2.2 contain descriptions of the proposal distributions and the deduced expressions for $\log(r)$ for the fixed k moves and the birth/death moves.

A.2.1. Fixed k moves

For $w_{1:k}$:

(a) Proposal: Normalized multiplicative perturbation ⊗Log normal$(0,\eta)$, for details, see Reference 13,

$$\log\left(f\left(w_{1:k} \mid w^*_{1:k}\right)\right) - \log\left(f\left(w^*_{1:k} \mid w_{1:k}\right)\right) = \sum_{i=1}^{k} \log\left(\frac{w^*_i}{w_i}\right)$$

(b) Prior: Dirichlet ($\mathfrak{D}(\delta)$ with $\delta_{1:k-1} = 1$)

$$\log\left(p\left(w^*_{1:k}\right)\right) - \log\left(p\left(w_{1:k}\right)\right) = \sum_{i=1}^{k} (\delta_i - 1)\log\left(\frac{w^*_i}{w_i}\right) = 0$$

(c) $\log(r)$:

$$\log(r) = \log l(\tau_{1:n} \mid \theta^*) - \log l(\tau_{1:n} \mid \theta) + \sum_{i=1}^{k} \log\left(\frac{w^*_i}{w_i}\right)$$

For $(\mu_{1:k,1:m}, b_{1:k,1:m}, d_{1:k,1:m})$:

(a) Proposal: For $(\mu_{1:k,1:m-3}, b_{1:k,1:m-2}, d_{1:k,1:m-1})$, normal perturbation

$$\underset{1:k}{\otimes}\ \underset{1:(m-3)}{\otimes} \mathcal{N}\left(0, \sigma^2_{\epsilon_\mu}\right) \underset{1:k}{\otimes}\ \underset{1:(m-2)}{\otimes} \mathcal{N}\left(0, \sigma^2_{\epsilon_b}\right) \underset{1:k}{\otimes}\ \underset{1:(m-1)}{\otimes} \mathcal{N}\left(0, \sigma^2_{\epsilon_d}\right)$$

As for the remaining $(\mu^*_{1:k,m-2:m}, b^*_{1:k,m-1:m}, d^*_{1:k,m})$, they are computed from the deterministic mapping, for all $1 \le c \le k$,

$$\left(\mu^*_{c,1:m}, b^*_{c,1:m}, d^*_{c,1:m}\right) = F\left(\mu^*_{c,1:m-3}, b^*_{c,1:m-2}, d^*_{c,1:m-1}\right), \qquad (21)$$

described in Subsection A.1.3.
Moreover we add, as described in Subsection 4.2, an additional restriction on the candidate, namely it should satisfy, for all $c = 1, \ldots, k$ and $i = 1, \ldots, m$,

$$b^*_{c,i} \in \left[\eta^b_i \pm 2\sqrt{\kappa^b}\right] \quad d^*_{c,i} \in \left[\eta^d_i \pm 2\sqrt{\kappa^d}\right]$$

For candidates that satisfy Equation (21), it is easy to prove that

$$\begin{aligned} &f\left(\mu^*_{1:k,1:m}, b^*_{1:k,1:m}, d^*_{1:k,1:m} \mid \mu_{1:k,1:m}, b_{1:k,1:m}, d_{1:k,1:m}\right) \\ &= f\left(\mu^*_{1:k,1:m-3}, b^*_{1:k,1:m-2}, d^*_{1:k,1:m-1} \mid \mu_{1:k,1:m-3}, b_{1:k,1:m-2}, d_{1:k,1:m-1}\right) \end{aligned} \qquad (22)$$

Moreover,

$$\begin{aligned} &\log\left(f\left(\mu^*_{1:k,1:(m-3)}, b^*_{1:k,1:(m-2)}, d^*_{1:k,1:(m-1)} \mid \right.\right. \\ &\qquad \left.\left.\mu_{1:k,1:(m-3)}, b_{1:k,1:(m-2)}, d_{1:k,1:(m-1)}\right)\right) \\ &- \log\left(f\left(\mu_{1:k,1:(m-3)}, b_{1:k,1:(m-2)}, d_{1:k,1:(m-1)} \mid \right.\right. \\ &\left.\left.\mu^*_{1:k,1:(m-3)}, b^*_{1:k,1:(m-2)}, d^*_{1:k,1:(m-1)}\right)\right) = 0 \end{aligned}$$

(b) Prior: for $(\mu_{1:k,1:m-3}, b_{1:k,1:m-2}, d_{1:k,1:m-1})$

$$\underset{1:k}{\otimes}\ \underset{1:(m-3)}{\otimes} U(-\pi, \pi) \underset{1:k}{\otimes} \mathcal{N}\left(\eta^b_{1:(m-2)}, \kappa^b \mathbf{I}d_{m-2}\right) \underset{1:k}{\otimes} \mathcal{N}\left(\eta^d_{1:(m-1)}, \kappa^d \mathbf{I}d_{m-1}\right)$$

As for the remaining $(\mu_{1:k,m-2:m}, b_{1:k,m-1:m}, d_{1:k,m})$, they are deduced from the deterministic mapping, for all $1 \le c \le k$,

$$\left(\mu^*_{c,1:m}, b^*_{c,1:m}, d^*_{c,1:m}\right) = F\left(\mu^*_{c,1:m-3}, b^*_{c,1:m-2}, d^*_{c,1:m-1}\right)$$

described in Subsection A.1.3.

As a consequence,

$$\begin{aligned} &\log\left(p\left(\mu^*_{1:k,1:(m-3)}, b^*_{1:k,1:(m-2)}, d^*_{1:k,1:(m-1)}\right)\right) \\ &- \log\left(p\left(\mu_{1:k,1:(m-3)}, b_{1:k,1:(m-2)}, d_{1:k,1:(m-1)}\right)\right) \\ &= \sum_{c=1}^{k}\left[\sum_{i=1}^{m-2} \frac{b^2_{ci} - b^{*2}_{ci} - 2\eta^b_i(b_{ci} - b^*_{ci})}{2\kappa^b} \right. \\ &\left. + \sum_{i=1}^{m-1} \frac{d^2_{ci} - d^{*2}_{i} - 2\eta^d_i(d_{ci} - d^*_{ci})}{2\kappa^d}\right] \end{aligned}$$

(c) $\log(r)$:

$$\begin{aligned} \log(r) = \log l(\tau_{1:n} \mid \theta^*) - \log l(\tau_{1:n} \mid \theta) \\ + \sum_{c=1}^{k}\left[\sum_{i=1}^{m-2} \frac{b^2_{ci} - b^{*2}_{ci} - 2\eta^b_i(b_{ci} - b^*_{ci})}{2\kappa^b} \right. \\ \left. + \sum_{i=1}^{m-1} \frac{d^2_{ci} - d^{*2}_{ci} - 2\eta^d_i(d_{ci} - d^*_{ci})}{2\kappa^d}\right] \end{aligned}$$

For $\sigma^2_{1:k}$:

(a) Proposal: Normalized multiplicative distribution ⊗Log normal$(0,\nu)$

$$\log\left(f\left(\sigma^2_{1:k} \mid \sigma^{2*}_{1:k}\right)\right) - \log\left(f\left(\sigma^{2*}_{1:k} \mid \sigma^2_{1:k}\right)\right) = \sum_{c=1}^{k} \log\left(\frac{\sigma^{2*}_c}{\sigma^2_c}\right)$$

(b) Prior: Inverse Gamma ($\mathfrak{IG}(\alpha,\beta)$)

$$\begin{aligned} \log\left(p\left(\sigma^{2*}_{1:k}\right)\right) - \log\left(p\left(\sigma^2_{1:k}\right)\right) = (\alpha+1)\sum_{c=1}^{k} \log\left(\frac{\sigma^2_c}{\sigma^{2*}_c}\right) \\ + \frac{1}{\beta}\sum_{c=1}^{k}\left(\frac{1}{\sigma^2_c} - \frac{1}{\sigma^{2*}_c}\right) \end{aligned}$$

(c) $\log(r)$:

$$\begin{aligned} \log(r) = \log l(\tau_{1:n} \mid \theta^*) - \log l(\tau_{1:n} \mid \theta) \\ + \alpha\sum_{c=1}^{k} \log\left(\frac{\sigma^2_c}{\sigma^{2*}_c}\right) + \frac{1}{\beta}\sum_{c=1}^{k}\left(\frac{1}{\sigma^2_c} - \frac{1}{\sigma^{2*}_c}\right) \end{aligned}$$

A.2.2. Birth or death moves

Birth: For k components, the proposal for the $k+1$-th component is constructed by drawing respectively

1. $w^*_{k+1} \sim Be(1,k)$ followed by a re-normalization $w^*_{1:k+1} = ((1 - w^*_{k+1})w_{1:k}, w^*_{k+1})$.
2. For $(\mu_{k+1,1:m-3}, b_{k+1,1:m-2}, d_{k+1,1:m-1})$,

$$\underset{1:(m-3)}{\otimes} U(-\pi,\pi) \otimes \mathcal{N}\left(\eta^b_{1:(m-2)}, \kappa^b \mathbf{I}d_{m-2}\right) \otimes \mathcal{N}\left(\eta^d_{1:(m-1)}, \kappa^d \mathbf{I}d_{m-1}\right)$$

As for the remaining $(\mu_{k+1,m-2:m}, b_{k+1,m-1:m}, d_{k+1,m})$, they are deduced from the deterministic mapping,

$$\left(\mu^*_{k+1,1:m}, b^*_{k+1,1:m}, d^*_{k+1,1:m}\right) = F\left(\mu^*_{k+1,1:m-3}, b^*_{k+1,1:m-2}, d^*_{k+1,1:m-1}\right)$$

described in Subsection A.1.3.
3. $\sigma^{2*}_{k+1} \sim \mathfrak{IG}(\alpha,\beta)$.

The expression of the acceptance probability for the birth-death move is deduced as in Reference [9], $P_d(k+1)$ and $P_b(k)$ respectively being the death probability in space $k+1$ and the birth probability in space k. $P_d(k+1)$ and $P_b(k)$ were

chosen such that $P_d(k+1) = P_b(k) = c$ for $k = 1, \ldots, k_{\max}$ otherwise $P_d(1) = 0$ and $P_b(k_{\max}) = 0$

$$
\begin{aligned}
r &= \frac{(k+1)!l(\tau_{1:n} \mid \theta_{1:k+1})p(\theta_{1:k+1})}{k!l(\tau_{1:n} \mid \theta_{1:k})p(\theta_{1:k})} \frac{\frac{P_d(k+1)}{k+1}}{P_b(k)} \frac{1}{p(u)} \\
&= \frac{l(\tau_{1:n} \mid \theta_{1:k+1})}{l(\tau_{1:n} \mid \theta_{1:k})} \frac{P_d(k+1)}{P_b(k)} \frac{p(\theta_{1:k+1})}{p(\theta_{1:k})p(u)} \\
&= \frac{l(\tau_{1:n} \mid \theta_{1:k+1})}{l(\tau_{1:n} \mid \theta_{1:k})} \frac{P_d(k+1)}{P_b(k)}
\end{aligned}
$$

As for the death move, a component is chosen randomly among the existing components and is killed with probability $1/r$, where r is given above.

REFERENCES

1. Meyer TJ. Chemical approaches to artificial photosynthesis. *Acc. Chem. Res.* 1989; **22**: 163–170.
2. Hendrickson JB. Molecular geometry. vii. Modes of interconversion in the medium rings. *J. Am. Chem. Soc.* 1967; **89**: 7047–7054.
3. Zimmer M. Molecular mechanics, data and conformational analysis of first-row transition metal complexes in the cambridge strutural database. *Coord. Chem. Rev.* 2001; **212**: 133–163.
4. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**: 711–732.
5. Allen FH, Howard JAK, Pitchford NA. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. iv. cyclooctane and related eight-membered rings. *Acta Cryst.* 1996; **B52**: 882–891.
6. Mass W. *Crystal Structure Determination*. Springer-Verlag: Berlin (second edition), 2004.
7. Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 1993; **49**: 803–821.
8. Richardson S, Green PJ. Corrigendum: on Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* 1997; **59**: 731–792. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 1998; **60**: 661.
9. Cappé O, Robert CP, Rydén T. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 2003; **65**: 679–700.
10. Dunitz JD. Conformations of medium rings. In *Perspectives in Structural Chemistry*, Vol II, Dunitz JD, Ibers JA (eds). John Wiley: New York, 1968.
11. Robert CP, Casella G. *Monte Carlo statistical methods*. Springer texts in statistics (second edition). Springer-Verlag: New York, 2004.
12. Dellaportas P, Papageorgiou I. Multivariate mixtures of normals with unknown number of components. Technical report, Department of Statitics, Athens University of Economics and Finance, 2004.
13. Cappé O, Robert CP, Rydén T. Manual for ctrjmix. Technical report, Département TSI, Ecole Nationale Supérieure des Télćommunications, 2003. http://www.tsi.enst.fr/∼cappe/ctrj_mix/
14. Celeux G, Hurn M, Robert CP. Computational and inferential difficulties with mixture posterior distributions. *J. Amer. Statist. Assoc.* 2000; **95**: 957–970.
15. Stephens M. Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 2000; **62**: 795–809.
16. Cappé O, Moulines E, Rydén T. *Inference in Hidden Markov Models*. Springer texts in statistics. Springer-Verlag: New York, 2005.
17. Celeux G. Bayesian inference for mixtures, the label-switching problem. In *COMPSTAT 1998* 1998; Physica, Heidelberg, 227–232.
18. Stephens M. *Bayesian Inference for Mixtures of Normal Distributions*. PhD dissertation, Oxford, 1997.