

Influence of molecular flexibility on DNA radiosensitivity: A simulation study

D. Viduna,¹ K. Hinsén,² and G. Kneller^{2,*}

¹Faculty of Science, Charles University, Albertov 6, 128 43 Prague 2, Czech Republic

²Centre de Biophysique Moléculaire, CNRS UPR No. 4301, Rue Charles-Sadron, 45071 Orléans Cedex 2, France

(Received 14 January 2000)

Radiation damage in DNA is caused mainly by hydroxyl radicals which are generated by ionizing radiation in water and removing hydrogen atoms from the DNA chain. This damage affects certain nucleotide sequences more than others due to differences in the local structure of the DNA chains. This sequence dependence has been analyzed experimentally and calculated theoretically for a rigid DNA model. In this paper we take into account the flexibility of the DNA chain and show how it modifies the strand breakage probabilities. We use a simple harmonic model for DNA flexibility which permits the study of a long (68 base pair) fragment with modest computational effort. The essential influence of flexibility is an increased breakage probability towards the ends of the fragment, which can also be identified in the experimental data.

PACS number(s): 87.15.Aa

I. INTRODUCTION

Ionizing radiation generates hydroxyl radicals (OH[•]) in living matter which are able to produce strand breaks in DNA molecules [1]. These radicals form water molecules by removing hydrogen atoms from the DNA chain. In principle, such strand breaks can occur at all nucleotide sites along the DNA double helix, but the probability depends on the position of the target nucleotide. This has been demonstrated for B-DNA fragments irradiated in an oxygenated solution without scavengers [2], where a lower probability for strand breakage is observed for nucleotides belonging to specific sequences such as AATT and, to a lesser extent, GGTC and TCCT. The reason is the lower accessibility in this region of the hydrogens involved in strand breakage due to a different local structure of these regions. The atoms H4', H5'1, and H5'2 are the main atoms whose abstraction results in strand breakage [3,4]. The H5'1 is located on the edge of the minor groove of the DNA double helix, whereas H4' and H5'2 are more embedded inside this groove.

Recently static molecular modeling has been used to evaluate the accessibility of the hydrogen atoms in an energy minimized DNA fragment [5]. The accessibility was defined as the geometrical solvent-accessible surface of each hydrogen atom. It is then assumed that this surface is proportional to the probability for hydrogen removal. A direct correlation between the relative radioresistance of some sequences and the accessibility of the H4' and H5'2 atoms has been found in this way. An alternative measure for the strand-breakage probabilities, which is based on the theory of diffusion-controlled reactions [6,5], yields essentially the same results.

The present work is an attempt to include the effect of molecular flexibility into the calculations of hydrogen accessible surfaces. We consider a 68 base pair DNA fragment whose conformational energy is described in the harmonic approximation. Using an efficient normal mode technique, we compute the average accessible surfaces and their distri-

butions for some hydrogen sites in the sugar rings. The simulation results are confronted with experimental data for relative strand-breakage probabilities along the sequence.

II. SIMULATION METHOD

The basic quantity to be evaluated for the following study is the accessible surface of the hydrogen atoms located in the DNA sugar rings. We use the method developed by Eisenhaber *et al.* to compute solvent-accessible surfaces [7]. This algorithm calculates each atom's contribution to the solvent-accessible molecular surface. The essential parameters are the van der Waals radii of the probe particle—here the OH[•] radical—and of the atoms in the DNA molecule. For our calculations we use the van der Waals radii given in [8], taking the radius of oxygen (1.5 Å) for OH[•]. For given positions $\mathbf{R}_1, \dots, \mathbf{R}_N$ and van der Waals radii a_1, \dots, a_N for the N atoms in the DNA fragment, the accessible surfaces σ_H of the hydrogen atoms are spherical calottes with a surface $0 \leq \sigma_H \leq 4\pi a_H^2$, where a_H is the van der Waals radius of hydrogen. In a statistical description one defines the average surface for a given atom i (hydrogen or other) as

$$\langle \sigma_i \rangle = \int \dots \int d^3R_1 \dots d^3R_N P_{eq}(\mathbf{R}_1, \dots, \mathbf{R}_N) \times \sigma_i(\mathbf{R}_1, \dots, \mathbf{R}_N; \{a_j\}), \quad (1)$$

where $P_{eq}(\mathbf{R}_1, \dots, \mathbf{R}_N)$ is the equilibrium distribution function

$$P_{eq}(\mathbf{R}_1, \dots, \mathbf{R}_N) = \frac{1}{Z_c} \exp\{-\beta U(\mathbf{R}_1, \dots, \mathbf{R}_N)\}. \quad (2)$$

Here $U(\mathbf{R}_1, \dots, \mathbf{R}_N)$ denotes the potential energy of the system as a function of the atomic positions and Z_c is the configurational partition function which normalizes P_{eq} to one. As usual, $\beta = 1/k_B T$ is the inverse temperature divided by the Boltzmann constant.

The average (1) can be computed by Monte Carlo methods, which means that the surfaces σ_i as well as the potential energy must be evaluated for each trial conformation. If stan-

*Author to whom correspondence should be addressed. Electronic address: kneller@cnrs-orleans.fr

standard molecular mechanics force fields are used, the computation of the potential energy is not only extremely time consuming, in particular for large DNA segments, but in addition many trial moves producing close contacts or overlap configurations will be rejected. It is therefore difficult to sample the important large-amplitude motions sufficiently well. To develop a viable simulation scheme we make the assumption that the potential energy can be replaced by a harmonic approximation. Harmonic DNA models have recently been used to study DNA-ligand binding [9] and DNA flexibility for different forms of DNA [10]. One of the first applications was a study of the *B* to *Z* transition as a function of NaCl concentration [11]. In the harmonic approximation the potential energy function takes the form

$$U(\mathbf{R}_1, \dots, \mathbf{R}_N) = \frac{1}{2} \sum_{i,j} (\mathbf{R}_i - \mathbf{R}_i^{eq}) \mathbf{K}_{ij} (\mathbf{R}_j - \mathbf{R}_j^{eq}), \quad (3)$$

where the matrices \mathbf{K}_{ij} contain the force constants, i.e., the second derivatives of the potential with respect to the positions. They are evaluated at a stable equilibrium configuration, $\mathbf{R}_1^{eq}, \dots, \mathbf{R}_N^{eq}$. The potential energy (3) can be written as a quadratic form, $U = 1/2 \delta \mathbf{R} \mathbf{K} \delta \mathbf{R}$, if we introduce the $3N$ -dimensional column vectors $\mathbf{R} = (\mathbf{R}_1^T, \dots, \mathbf{R}_N^T)^T$ and $\delta \mathbf{R} = \mathbf{R} - \mathbf{R}^{eq}$, together with the $(3N \times 3N)$ matrix $\mathbf{K} = (\mathbf{K}_{ij})$. The superscript T denotes a transposition. \mathbf{K} is positive semidefinite, since \mathbf{R}^{eq} is a stable minimum energy configuration. Therefore it can be brought to diagonal form, $\mathbf{C}^T \mathbf{K} \mathbf{C} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix whose elements λ_α are non-negative. The matrix \mathbf{C} is orthogonal and its column vectors are the eigenvectors of \mathbf{K} , i.e., the normal modes of the system under consideration. If we express the displacements $\delta \mathbf{R}$ in the normal modes, $\delta \mathbf{R} = \mathbf{C} \mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_{3N})^T$, the harmonic potential energy surface takes the particularly simple form

$$\tilde{U}(x_1, \dots, x_{3N}) \approx \frac{1}{2} \sum_{\alpha=1}^{3N} \lambda_\alpha x_\alpha^2. \quad (4)$$

With Eqs. (1) and (2) the average surface can be computed via

$$\langle \sigma_i \rangle = \frac{1}{Z_c} \int \dots \int dx_1 \dots dx_{3N} \prod_{\alpha=1}^{3N} \exp \left\{ -\frac{\beta}{2} \lambda_\alpha x_\alpha^2 \right\} \times \tilde{\sigma}_i(x_1, \dots, x_{3N}; \{a_j\}), \quad (5)$$

where $\tilde{\sigma}_i(x_1, \dots, x_{3N}; \{a_j\}) = \sigma_i(\mathbf{R}_1, \dots, \mathbf{R}_N; \{a_j\})$. To compute the accessible surfaces $\tilde{\sigma}_i$ of an atom for a given set of normal mode coordinates, one calculates the Cartesian coordinates $\mathbf{R} = \mathbf{R}^{eq} + \mathbf{C} \mathbf{x}$ and then uses the algorithm from Ref. [7]. The partition function Z_c can be evaluated analytically yielding

$$Z_c = \prod_{\alpha=1}^{3N} \sqrt{\frac{2\pi}{\lambda_\alpha}}. \quad (6)$$

Equation (5) shows that the computation of expectation values in the harmonic approximation amounts essentially to averaging over configurations that are obtained from Gauss-

ian random deviates. In comparison to the standard Metropolis Monte Carlo method, the sampling is optimal in the sense that all trial configurations are accepted. Although in practice only about 10% of the normal modes need to be considered in Eq. (5), their computation is still a formidable task for a large system such as our 68 base-pair DNA segment. To deal with this problem we used a simplified force field which is derived from the Amber94 potential function [12]:

$$U = \sum_{\text{bonds } ij} k_{ij} (r_{ij} - r_{ij}^{(0)})^2 + \sum_{\text{angles } ijk} k_{ijk} (\phi_{ijk} - \phi_{ijk}^{(0)})^2 + \sum_{\text{dihedrals } ijkl} k_{ijkl} \{1 + \cos(n_{ijkl} [\theta_{ijkl} - \theta_{ijkl}^{(0)}])\}^2 + \sum_{\text{all pairs } ij} U_{ij}(r_{ij}). \quad (7)$$

The constants k_{ij} , $r_{ij}^{(0)}$, k_{ijk} , $\phi_{ijk}^{(0)}$, k_{ijkl} , $\theta_{ijkl}^{(0)}$, and n_{ijkl} parametrize the energy contribution of the chemical bond structure. In comparison to the Amber94 force field we adjusted the equilibrium bond lengths, bond valence, and torsion angles such that the given input configuration is by definition an equilibrium configuration. The force constants k_{ij} , k_{ijk} , and k_{ijkl} were left unchanged. More important is the modification concerning the functional form of the force field: we replaced all nonbonded interactions (electrostatic and Lennard-Jones terms) by the unspecific harmonic deformation potential from Ref. [13],

$$U_{ij}(\mathbf{r}) = k(\mathbf{R}_{ij}^{(0)}) (|\mathbf{r}| - |\mathbf{R}_{ij}^{(0)}|)^2. \quad (8)$$

Here $\mathbf{R}_{ij}^{(0)}$ is the pair distance vector $\mathbf{R}_i - \mathbf{R}_j$ in the input configuration which is supposed to be the equilibrium configuration. The distance-dependent force constant is given by

$$k(\mathbf{r}) = c \exp \left(-\frac{|\mathbf{r}|^2}{r_0^2} \right), \quad (9)$$

and the parameter r_0 was left at the optimal value found in Ref. [13], i.e., 0.3 nm. In the same reference it has been demonstrated that the deformation force field (8) yields the correct large-amplitude and low-frequency normal modes. Combining it with the energy terms describing bonded interactions yields a good approximation for the normal modes for the whole frequency spectrum [14]. Due to the absence of long-ranged interactions, most elements of the force constant matrix \mathbf{K} are zero, and the normal modes can be computed efficiently using iterative techniques. We used the modified Lanczos algorithm for sparse matrices as implemented in the ARPACK library [15].

III. RESULTS AND DISCUSSION

We applied the method described in the preceding section to study the influence of molecular flexibility on the radiosensitivity of a 68 base pair DNA fragment. As input for our calculations we used the energy-minimized DNA structure of Ref. [5] whose sequence is *TCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAATCCACACAACATACGAGCCGG*. By construc-

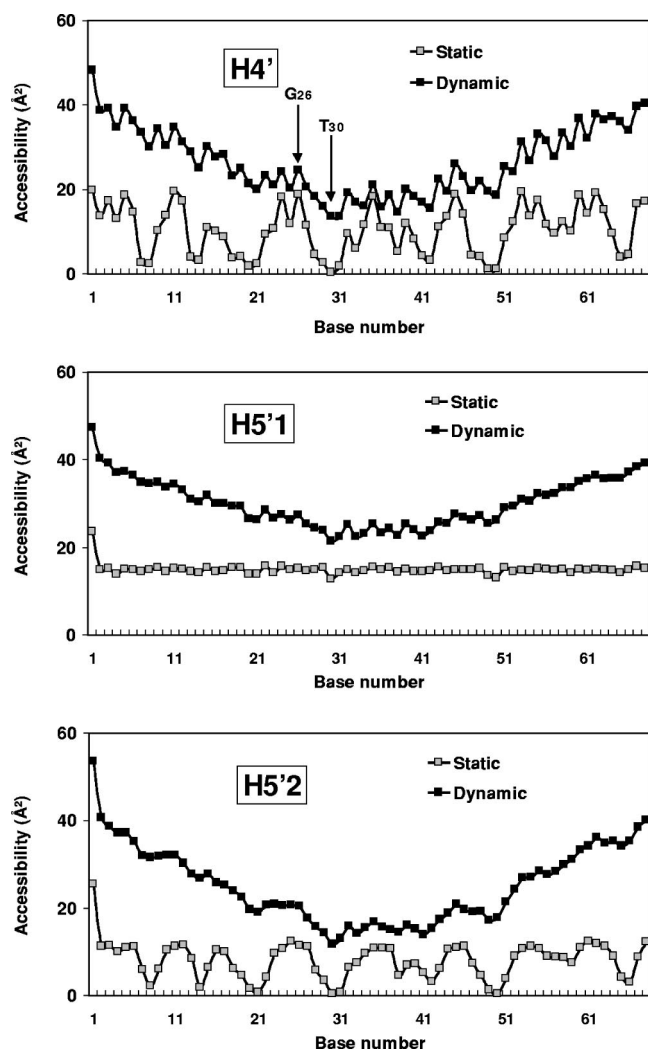


FIG. 1. Accessible surfaces of the hydrogen atoms (a) H4', (b) H5'1, and (c) H5'2 as a function of the nucleotide number for the static and the dynamic model.

tion, this configuration corresponds to a minimum of our simplified DNA force field. The normal modes were calculated with the Molecular Modeling Toolkit [16]. Friction effects resulting from the surrounding solvent need not be taken into account since they do not change the equilibrium distribution function; they are important only for the computation of time-dependent quantities [17]. In a first step we performed a full normal mode analysis of a 9-base pair DNA segment to examine how many low-frequency modes must be taken into account to obtain a stable Monte Carlo estimate for the accessible atomic surfaces of the hydrogen atoms in the sugar rings. A full normal mode analysis for a 68-base pair DNA segment would have been infeasible and moreover unnecessary, since it is well known that the important large-amplitude normal modes correspond to low-frequency motions. We found that 9% of the normal modes are sufficient to sample the hydrogen surfaces correctly. The corresponding frequency range is 0–9 THz. The subsequent Monte Carlo simulation of the 68 base pair segment was then performed with 9% of the normal modes as well.

Figures 1(a)–1(c) show the average accessible surface of the hydrogen atoms H4', H5'1, and H5'2 along the DNA molecule compared to the corresponding accessible surfaces

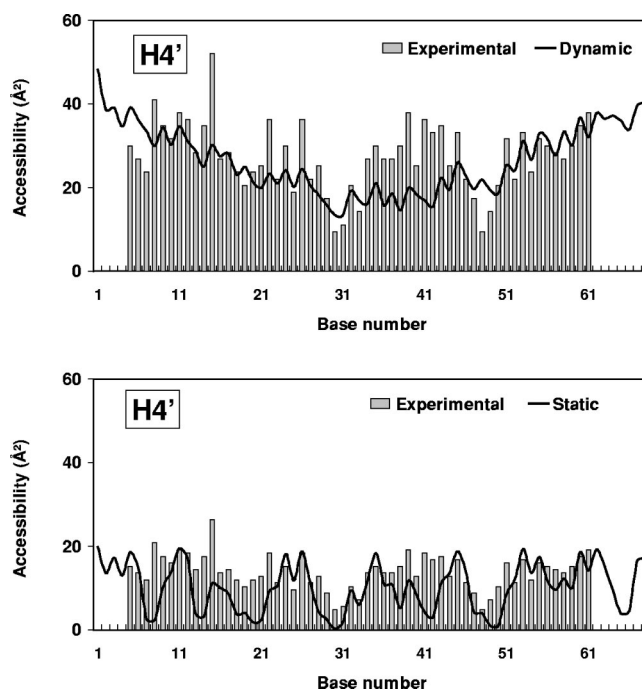


FIG. 2. Comparison between the accessible surfaces of the H4' atoms and the experimental data for relative strand-breakage probabilities from Ref. [5]. The experimental data were scaled by a factor obtained from a fit to the simulation data for the flexible model (a) and the static model (b). The quality of the two fits is very similar; see the text for details.

in the static model, which was calculated from the input structure alone without any Monte Carlo sampling. Our analysis was made for one strand only, for which experimental breakage probabilities are available. The average accessible surface in the flexible model is always larger than in the static model, the difference being most pronounced towards the ends of the DNA segment. In contrast, the hydrogen accessibility pattern is partially smoothed out in the flexible model.

The parabola-shaped dependency of the hydrogen accessibilities has its counterpart in the experimental values for strand-breakage probabilities, taken from Ref. [5] and shown in Fig. 2(a). Since it is generally recognized that the H4' atom is one of the most important ones for the induction of strand breakages [5,4,3], we compare here the experimental data with the simulated accessible surfaces of that site only. The corresponding profile for the static model shown in Fig. 2(b) does not exhibit this feature and is globally more or less flat. This indicates that boundary effects play an important role for the radiosensitivity of DNA fragments. However, the experimental results were obtained with an 80-base pair fragment, whereas the calculations were performed for a fragment consisting of only the 68 central base pairs; boundary effects should thus be less pronounced in the experimental data.

The experimental data e_i are relative breakage probabilities and thus in arbitrary units. They have been scaled by a factor s to fit the exposed surface values σ_i for the dynamic and static model separately. This factor s was chosen such that the absolute error

$$A = \sqrt{\frac{1}{N} \sum_{i=1}^N (\sigma_i - s e_i)^2} \quad (10)$$

is minimized. For the dynamic model we find $s = 13.3 \text{ \AA}^2$ and $A = 6.8 \text{ \AA}^2$; the values for the static model are $s = 5.5 \text{ \AA}^2$ and $A = 5.1 \text{ \AA}^2$, i.e., very similar. However, considering the relative error,

$$R = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\sigma_i}{s e_i} - 1 \right)^2}, \quad (11)$$

we find $R = 0.37$ for the dynamic and $R = 0.52$ for the static model. Minimizing R instead of A leads to qualitatively similar results. To interpret these findings one has to take into account that the surfaces obtained from the dynamic model are systematically larger than those obtained from the static model. Consequently, the scaling factor s will also be larger to fit the experimental data, and A increases with respect to the static model. To avoid, or at least reduce such a bias, the relative error should be used to judge the quality of the fits. It is easy to see that R would be unaffected by a constant ratio between the surfaces in the dynamic and the static model, respectively. Comparison of the relative errors leads to the conclusion that the dynamic model fits the data better than the static one.

To extract more detailed information concerning the simulated hydrogen accessibilities we looked at their distribution. Figure 3 shows the probability distribution of accessible surfaces for the $\text{H4}'$ atom in the sugar rings of residue 26 and 30 in the same strand. These distributions were obtained as histograms of the surface values for the individual configurations that were sampled. We have chosen these residues since they are located sufficiently far from the ends, in order to exclude boundary effects, and because they have very different average accessibilities. According to Fig. 1(a) the hydrogen $\text{H4}'$ of residue 30 (*T*) has the lowest accessibility in the segment. The corresponding histogram is monotonically decreasing, showing that states of low accessibility are predominant—the histogram exhibits a strong skewness. In contrast, residue 26 (*G*) shows the highest accessibility in the central part of the segment [see Fig. 1(a)], and the accessibility distribution has a clear maximum at around 22 \AA^2 . Here the accessibilities are almost symmetrically distributed around the average value and the net effect of the dynamics on the accessible surface is almost zero, whereas for residue 30 the latter is considerably enhanced by flexibility. The skewness of the histogram in Fig. 3 shows that any large-amplitude motion opens channels to the $\text{H4}'$ atom that are nonexistent in the static picture.

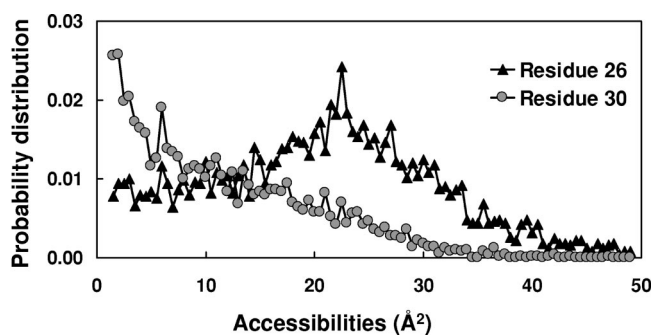


FIG. 3. Probability distribution of the accessible surface for atom $\text{H4}'$ in residues *G* 26 and *T* 30. The arrows in Fig. 1(a) indicate the locations of these residues in the DNA strand.

To summarize, we find that flexibility increases the hydrogen accessibility of DNA fragments in general, in particular towards the ends. It partially smoothes out the variations observed in the static model, yielding globally a better agreement with the experimental data. Approximate normal mode analysis permits the study of relatively large DNA fragments with an acceptable computational effort. From experience with similar calculations on proteins [14], it can be expected that our simplified force field underestimates the frequencies for slow motions and thus yields somewhat exaggerated amplitudes. The “true” accessibility values should therefore lie in between those from the static and dynamic models. In this case, the agreement between calculated and experimental can be expected to improve.

Several improvements to our very simple model can be envisaged for future studies. The hydrogen removal probabilities should be determined by electronic structure calculations instead of assuming proportionality to the solvent-accessible surface. Frictional effects must be included in order to obtain the time scales of the DNA motion and relate them to the diffusive motion time scale of the OH^\bullet radicals. Ultimately, stochastic simulations at the atomic level and with a realistic potential energy surface should be performed. However, our study shows that simple models are sufficient to obtain a qualitative picture of dynamic influences on strand breakage probabilities.

ACKNOWLEDGMENTS

The authors would like to thank Dr. M. Charlier, Dr. M. Spothem-Maurizot, Dr. D. Sy, and Dr. M. Begusova for helpful discussions and for providing the DNA input structure from Ref. [5]. D. V. wishes to thank the Centre International des Etudiants et Stagiaires for supporting this work through a grant within the bilateral cooperation program between France and the Czech Republic.

- [1] W.D. Henner, S.M. Grunberg, and W.A. Haseltine, *J. Biol. Chem.* **257**, 11 750 (1982).
- [2] J. Franchet-Beuzit, M. Spothem-Maurizot, R. Sabattier, B. Blazy-Baudras, and M. Charlier, *Biochemistry* **32**, 2104 (1993).
- [3] C. von Sonntag, in *The Chemical Basis of Radiation Biology* (Taylor and Francis, London, 1987).
- [4] B. Balasubramanian, W.K. Pogozelski, and T.D. Tullius, *Proc.*

- Natl. Acad. Sci. USA* **95**, 9738 (1998).
- [5] D. Sy, C. Savoye, M. Begusova, V. Michalik, M. Charlier, and M. Spothem-Maurizot, *Int. J. Radiat. Biol.* **72**, 147 (1997).
- [6] V. Michalik, M. Spothem-Maurizot, and M. Charlier, *J. Biomol. Struct. Dyn.* **13**, 565 (1995).
- [7] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *J. Comput. Chem.* **16**, 273 (1995).
- [8] A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).

- [9] M. Zacharias and H. Sklenar, *J. Comput. Chem.* **20**, 287 (1999).
- [10] T.H. Duong and K. Zakrzewska, *J. Comput. Chem.* **18**, 796 (1997).
- [11] A.E. Garcia and D.M. Soumpasis, *Proc. Natl. Acad. Sci. USA* **86**, 3160 (1988).
- [12] W.D. Cornell, P. Cieplak, I.R. Gould, C.I. Bayly, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [13] K. Hinsen, *Proteins* **33**, 417 (1998).
- [14] K. Hinsen and G.R. Kneller, *J. Phys. Chem.* **111**, 10 766 (1999).
- [15] R. Lehoucq, D. Sorensen, and C. Yang, ARPACK Users' Guide at <http://www.caam.rice.edu/software/ARPACK>.
- [16] K. Hinsen, *J. Comput. Chem.* (to be published).
- [17] M.C. Wang and G.E. Uhlenbeck, *Rev. Mod. Phys.* **17**, 323 (1945).