Communication: A minimal model for the diffusion-relaxation backbone dynamics of proteins

Gerald R. Kneller, Konrad Hinsen, and Paolo Calligari

Citation: J. Chem. Phys. **136**, 191101 (2012); doi: 10.1063/1.4718380
View online: http://dx.doi.org/10.1063/1.4718380
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v136/i19
Published by the American Institute of Physics.

Communication: A minimal model for the diffusion-relaxation backbone dynamics of proteins

Gerald R. Kneller,[1,2,3,a)] Konrad Hinsen,[1,2] and Paolo Calligari[4]
[1]*Centre de Biophys. Moléculaire, CNRS, Rue Charles Sadron, 45071 Orléans, France*
[2]*Synchrotron Soleil, L'Orme de Merisiers, 91192 Gif-sur-Yvette, France*
[3]*Université d'Orléans, Chateau de la Source-Av. du Parc Floral, 45067 Orléans, France*
[4]*Département de Chimie, associé au CNRS, Ecole Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France*

(Received 19 March 2012; accepted 30 April 2012; published online 15 May 2012)

We present a model for the local diffusion-relaxation dynamics of the C_α-atoms in proteins describing both the diffusive short-time dynamics and the asymptotic long-time relaxation of the position autocorrelation functions. The relaxation rate spectra of the latter are represented by shifted gamma distributions, where the standard gamma distribution describes anomalous slow relaxation in macromolecular systems of infinite size and the shift accounts for a smallest local relaxation rate in macromolecules of finite size. The resulting autocorrelation functions are analytic for any time $t \geq 0$. Using results from a molecular dynamics simulation of lysozyme, we demonstrate that the model fits the position autocorrelation functions of the C_α-atoms exceptionally well and reveals moreover a strong correlation between the residue's solvent-accessible surface and the fitted model parameters. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4718380]

Over the recent years, the multi-scale aspect of the internal dynamics of proteins and macromolecular systems in general has attracted considerable interest on both the experimental and theoretical sides.[1–9] A characteristic feature of time correlation functions (TCFs) related to internal diffusive motions is their strongly non-exponential decay. A classical empirical model is the stretched exponential function, $c(t) = c(0)\exp(-[t/\tau]^\beta)$ $(0 < \beta \leq 1)$, which has been used to model, for example, the kinetics of protein folding.[10] The description of internal protein dynamics through a fractional Ornstein-Uhlenbeck (fOU) process[11–13] leads to the stretched Mittag-Leffler (ML) function, $c(t) = c(0)E_\beta(-[t/\tau]^\beta)$ $(0 < \beta \leq 1)$, which has proven to be a good model for the TCF of position and distance fluctuations in proteins.[3,4,14,15] The ML function is an entire function in the complex plane,[16,17] $E_\beta(z) = \sum_{n=0}^{\infty} z^n/\Gamma(1+\beta n)$, and includes the exponential function as a special case, $E_1(z) = \exp(z)$. Since the models cited above capture the essential features of the measured correlation functions by a few parameters, they have undoubtedly their merits as simple models for an *ad hoc* interpretation of experimental data. They have, however, also the unpleasant feature of becoming unphysical for short times. Due to the "stretched" argument $\propto t^\beta$ the time derivatives $c^{(n)}(0+)$ do not exist if $0 < \beta < 1$, although they should exist whenever the time evolution of the dynamical system under consideration is described by equations of motion.[18–20] Consider, for example, the diffusive motion of an atom in a dense molecular system in thermal equilibrium. Here, the dynamical variable is the deviation of the atom with respect to its mean position, $\mathbf{u} = \mathbf{x} - \langle \mathbf{x} \rangle$, and the relevant TCF is its position autocorrelation function (PACF), $c(t) = \langle \mathbf{u} \cdot \exp(\mathcal{L}t)\mathbf{u} \rangle$. The symbol $\langle \ldots \rangle$ denotes an equilibrium ensemble average and $\mathcal{L}$ is the time evolution operator for the whole system. It follows from the construction of the PACFs that $c^{(n)}(0+) = \langle \mathbf{u} \cdot \mathcal{L}^n \mathbf{u} \rangle$ are well-defined ensemble averages which should exist. For non-Hamiltonian diffusive dynamics, where $\mathcal{L}$ is a many-particle Smoluchowski operator,[21,22]

$$D_s = \frac{1}{2} \left. \frac{d\langle [\mathbf{u}(t) - \mathbf{u}(0)]^2 \rangle}{dt} \right|_{t=0} = -c^{(1)}(0+), \qquad (1)$$

defines in particular the short-time diffusion coefficient.

In this paper, we develop a realistic minimal model for the backbone dynamics of proteins which leads to regular PACFs for the C_α-atoms describing both the diffusive short-time dynamics and the relaxation for long times. We assume that $\mathbf{u}(t)$ is described by a stationary stochastic process and write its autocorrelation function in the form

$$c(t) = \langle \mathbf{u}^2 \rangle \psi(t/\tau), \qquad (2)$$

where $\psi(.)$ is the normalized PACF for a dimensionless time argument, with $\psi(0) = 1$, and $\tau > 0$ sets the time scale. For convenience we set $\tau = 1$ in the following. To express the multi-scale character of protein dynamics we write the PACFs as a superposition of exponential functions,

$$\psi(t) = \int_0^\infty d\lambda \; p(\lambda) \exp(-\lambda t), \qquad (3)$$

where the relaxation rate spectrum $p(\lambda)$ satisfies the normalization condition $\int_0^\infty d\lambda \; p(\lambda) = 1$, as well as $p(\lambda) \geq 0$. The moments of the relaxation rate spectrum are given by

$$\overline{\lambda^k} = \int_0^\infty d\lambda \; \lambda^k p(\lambda) = (-1)^k \psi^{(k)}(0), \qquad (4)$$

and their existence depends on the behavior of $p(\lambda)$ for $\lambda \to \infty$. For $\overline{\lambda^k}$ to exist the relaxation rate spectrum must decay

a)Electronic mail: gerald.kneller@cnrs-orleans.fr.

at least as fast as $p(\lambda) \sim \lambda^{-(k+1+\epsilon)}$, with $\epsilon > 0$. In this context it is worthwhile noting that the relaxation rate spectrum corresponding to the stretched Mittag-Leffler function, $\psi(t) = E_\beta(-t^\beta)$, which describes the PACF of the fOU process[1,14] behaves for large λ as $p_{\mathrm{fOU}}(\lambda) \sim \lambda^{-(\beta+1)}$. For this reason *none* of moments $\overline{\lambda^k}$ with $k > 0$ exist and $\psi(t)$ is non-analytic at $t = 0$.

To construct a model for $p(\lambda)$, we make the following assumptions:

(a) For a protein of finite size, the PACF of each C_α-atom is characterized by a smallest relaxation rate, $\eta_{\min}$. In thermal equilibrium, a protein performs fluctuations about its equilibrium structure, and to a first approximation protein dynamics can be described by diffusion in an effective multidimensional harmonic potential energy.[23,24] If $\omega_{\min}$ is the smallest normal frequency for this harmonic potential, then $\eta_{\min} = \gamma^{-1}\omega_{\min}^2$ is the minimal relaxation rate for the PACF of a given C_α-atom, where $\gamma > 0$ is a positive friction constant which is essentially determined by the atomic density of the nearest neighbors. With increasing system size $\omega_{\min}$ and thus $\eta_{\min}$ tend to zero.

(b) In the limit of infinite protein size, the PACF of each C_α-atom exhibits anomalous relaxation,

$$\psi(t) \overset{t\to\infty}{\sim} t^{-\beta} \quad 0 < \beta < 1, \tag{5}$$

which is characteristic for relaxation processes in large scale polymeric networks.[2] Due to the slow decay of the PACFs, the corresponding average relaxation times, $\overline{\tau}_{\exp} = \int_0^\infty dt\, \psi(t)$, diverge.

(c) The PACFs should be analytical in $t = 0$, i.e., they should representable by a Taylor series in this point. Since $\psi(t)$ is the moment generating function for $p(\lambda)$, all moments $\overline{\lambda^k}$ must exist.

In order to fulfill the above requirements, the relaxation rate spectrum must have the general form

$$p(\lambda; \alpha, \beta) = \theta(\lambda - \alpha) p(\lambda - \alpha; \beta), \tag{6}$$

where $\theta(.)$ is the Heaviside unit step function, α is a dimensionless minimal relaxation rate ($\alpha = \eta_{\min}\tau$), and $p(\lambda; \beta)$ must be constructed such that $\psi(t)$ has the asymptotic form (5) if $\alpha = 0$. For this purpose, one can rely on the fact that the Laplace transform of $\psi(t)$ is the Stieltjes transform of $p(\lambda)$,

$$\hat{\psi}(s) = \int_0^\infty d\mu\, \frac{p(\mu)}{s + \mu}, \tag{7}$$

$$p(\lambda) = \frac{1}{\pi} \lim_{\epsilon\to 0} \Im\{\hat{\psi}(-\lambda - i\epsilon)\}, \tag{8}$$

where $\hat{\psi}(s) = \int_0^\infty dt\, \exp(-st)\psi(t)$ ($\Re\{s\} > 0$), and on a Tauberian theorem,[25,26] according to which

$$\hat{\psi}(s) \overset{s\to 0}{\sim} \frac{\Gamma(1-\beta)}{s^{1-\beta}} \tag{9}$$

follows from relation (5) and vice versa. Combining relations (8) and (9), one can conclude that $p(\lambda; \beta)$ must have the general form

$$p(\lambda; \beta) = f(\lambda) \frac{\sin(\pi\beta)}{\pi} \frac{\Gamma(1-\beta)}{\lambda^{1-\beta}} \quad (0 < \beta < 1), \tag{10}$$

where $f(\lambda)$ is a yet undetermined function fulfilling $\lim_{\lambda\to 0} f(\lambda) = C$. The constant C must be chosen such that $\int_0^\infty p(\lambda; \beta) = 1$. We note that $\lim_{\beta\to 1} \sin(\pi\beta)\Gamma(1-\beta) = \pi$. Relation (10) is a necessary and sufficient condition for a slowly decaying PACF with the asymptotic form (5). To construct $p(\lambda)$ such that the existence of all moments $\overline{\lambda^k}$ and thus the analyticity of $\psi(t)$ in $t = 0$ is guaranteed we set

$$f(\lambda) = C \exp(-\beta\lambda). \tag{11}$$

The properly normalized relaxation rate spectrum then reads

$$p(\lambda; \beta) = \frac{\lambda^{\beta-1}\beta^\beta \exp(-\beta\lambda)}{\Gamma(\beta)}, \tag{12}$$

and $\psi(t)$ is given by

$$\psi(t; \alpha, \beta) = \frac{\exp(-\alpha t)}{(1 + t/\beta)^\beta}. \tag{13}$$

The corresponding cumulants, which are defined through

$$c_{\alpha,\beta}^{(k)} = (-1)^k \frac{d^k}{dt^k} \ln(\psi(t; \alpha, \beta))\Big|_{t=0+} \tag{14}$$

have the particularly simple form

$$c_{\alpha,\beta}^{(1)} = 1 + \alpha, \tag{15}$$

$$c_{\alpha,\beta}^{(k)} = \frac{(k-1)!}{\beta^{k-1}}, \quad (k = 2, 3, \ldots). \tag{16}$$

From the form of $\psi(t; \alpha, \beta)$ and its cumulants, one derives the consistent limits

$$\lim_{\beta\to\infty} \psi(t; \alpha, \beta) = \exp(-[1 + \alpha]t), \tag{17}$$

$$\lim_{\beta\to\infty} p(\lambda; \alpha, \beta) = \delta(\lambda - [1 + \alpha]), \tag{18}$$

and we note in this context that $\psi(t; 0, 1/(1-q))$ is the "Tsallis q-exponential" which is considered in non-extensive statistical mechanics.[27–29]

To test our model, we performed a molecular dynamics simulation of a lysozyme molecule in water for a subsequent analysis of the C_α PACFs with our model. The simulated system was set up by starting with the initial structure 193L of the Protein Data Bank (PDB).[30] A total of 6775 water molecules were added, resulting in a system of 22 295 atoms. The simulations were performed using the NAMD program[31] with the all-atom force field AMBER99SB (Ref. 32) and with periodic boundary conditions. Electrostatic interactions were computed using the particle mesh Ewald method.[33] The integration time step was set to 1 fs and coordinates were saved every 50 fs for further analysis. After a preliminary minimization of the PDB structure, the system was first equilibrated at constant temperature (298 K) and constant pressure (1 bar) using a Langevin thermostat[34] coupled with a Nose-Hoover barostat.[35] The equilibrated system was used for a production run of 10 ns from which the PACFs of the C_α-atoms were calculated. The normalized PACFs were fitted according to

$$\frac{c(t)}{c(0)} \approx \psi(t/\tau; \alpha, \beta). \tag{19}$$

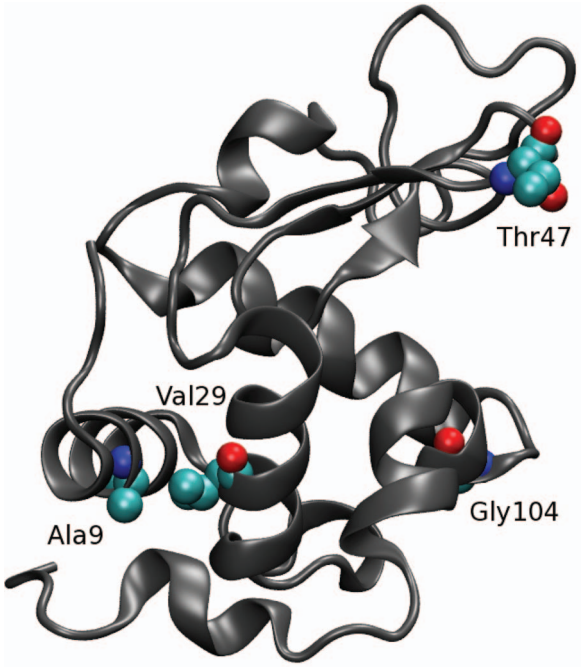

FIG. 1. Four selected residues in the lysozyme molecule.

All fits show an excellent agreement with the simulation data and we show here the results for four selected residues which have been chosen according to their exposure to solvent (see Fig. 1). Residues Ala 9 and Val 29 are buried in α-helices, whereas Thr 47 and Gly 104 are located in loop regions. Figure 2 shows the superposition of the simulated normalized PACFs (dots) with the fits of $\psi(t/\tau;\alpha,\beta)$ (solid lines) in form of a log-log plot. The fit parameters are given in the plot. A coherent view of the results is obtained by correlating the mean relaxation rate,

$$\overline{\lambda} = (1+\alpha)\tau^{-1}, \tag{20}$$

of all C_α-atoms and its spread,

$$\sigma_\lambda = (\overline{\lambda^2} - \overline{\lambda}^2)^{1/2} = \beta^{-1/2}\tau^{-1}, \tag{21}$$

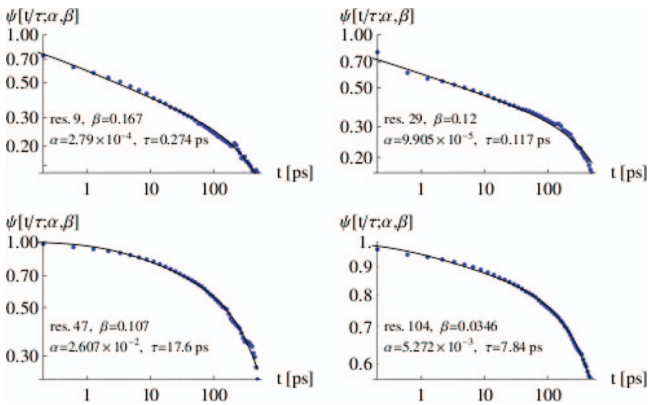

FIG. 2. Log-log plot of the simulated position autocorrelation functions for the C_α-atoms of the residues shown in Figure 1 (dots) and fits of model (13) (solid lines). For the simulated PACFs the smallest positive time argument is $t = 0.05$ ps.

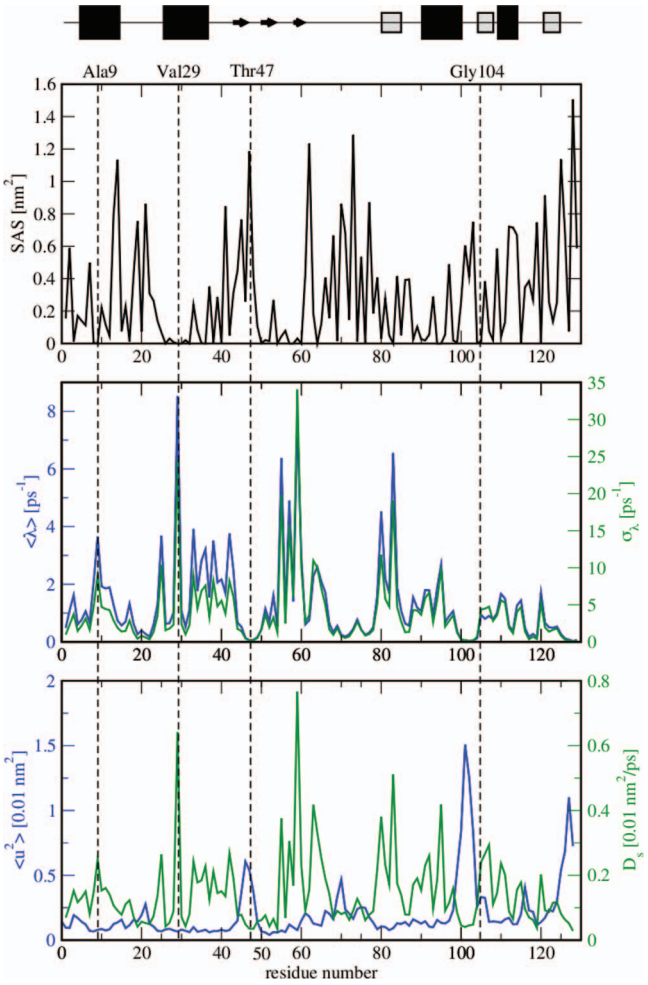

FIG. 3. Upper panel: Solvent accessible surface for the C_α-atoms in lysozyme. Middle panel: Mean relaxation rate $\overline{\lambda}$ (blue line) and corresponding standard deviation σ_λ (green line). Lower panel: Mean square position fluctuation $\langle \mathbf{u}^2 \rangle$ (blue line) and short time diffusion coefficient D_s (green line). The additional graphics on top of the figure locates the secondary structure elements. Black rectangles indicate α-helices, grey rectangles short helicoidal motifs, and arrows beta sheets.

with the solvent-accessible surface of the respective residue. Here, the total solvent-accessible surface of all atoms is considered. The fitted quantities $\overline{\lambda}$ and σ_λ for our model are, respectively, given in the upper and middle panels of Fig. 3, combining $\overline{\lambda}$ (blue line) and σ_λ (green line) in the middle panel. On top of the figure we indicate the location of secondary structure elements and the vertical lines locate the four selected residues displayed in Fig. 1. The results show that the PACFs of C_α-atoms in solvent-exposed loop regions relax one or two orders of magnitude more slowly than those buried in helices, and the spread of the relaxation rates follows exactly the same trend, which is not trivial since the mean relaxation and its spread are not described by the same parameters. The backbone relaxation dynamics in secondary structure elements is thus faster than the one in the more floppy, solvent-exposed loop regions, and has a much stronger non-exponential character. In this context, it is interesting to look at the static position fluctuations of the C_α-atoms and at the corresponding short-time diffusion coefficients, which depend on both the amplitudes of the atomic motions and on the

relaxation dynamics. It follows from relations (1), (2) and (4) that the short-time diffusion coefficient of our model is given by

$$D_s = \langle \mathbf{u}^2 \rangle \bar{\lambda}. \tag{22}$$

Both $\langle \mathbf{u}^2 \rangle$ and D_s are displayed in the lower panel of Fig. 3. As expected, the mean square position fluctuations are larger in the flexible loop regions than in secondary structure elements with higher rigidity. Astonishingly, the dynamical mobility, which is reflected in the short-time diffusion coefficient, shows the opposite trend, although $D_s \propto \langle \mathbf{u}^2 \rangle$. The reason is the dominating behavior of the mean relaxation rate $\bar{\lambda}$, which attains its peaks for C_α-atoms buried in secondary structure elements and very low values for C_α-atoms in loop regions. We explain this finding by a very fast relaxation dynamics through vibrational dephasing the first case and by a much slower segmental diffusion in the second case.

In this Communication, we have demonstrated the usefulness of minimal models for the diffusion-relaxation dynamics of proteins, which are based on a few observations and only basic assumptions concerning the physical nature of diffusion processes in macromolecular systems. The use of asymptotic analysis is a valuable tool in this context and properly describes the multi-scale aspect of protein dynamics. We have shown that the parameters of our model are strongly correlated with the solvent-accessible surface of the C_α-atoms, allowing for a physical interpretation of the diffusion-relaxation backbone dynamics in proteins and relating to earlier studies of protein solvent coupling (see, for example, Ref. 36). In contrast to fractional Brownian dynamics models, our model also yields a physically meaningful characterization of the short-time backbone dynamics, which is reflected in the short-time diffusion coefficients of the C_α-atoms.

P. Calligari acknowledges financial support by the Agence Nationale de Recherche (Contract No. ANR-2010-COSI-001-01).

[1] W. Glöckle and T. Nonnenmacher, Biophys. J. **68**, 46 (1995).
[2] A. Blumen, A. Gurtovenko, and S. Jespersen, J. Non-Cryst. Solids **305**, 71 (2002).
[3] H. Yang, G. Luo, P. Karnchanaphanurach, T. Louie, I. Rech, S. Cova, L. Xun, and X. Xie, Science **302**, 262 (2003).
[4] S. Kou and X. Xie, Phys. Rev. Lett. **93**, 180603 (2004).
[5] G. Kneller and K. Hinsen, J. Chem. Phys. **121**, 10278 (2004).
[6] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 98106 (2005).
[7] T. Neusius, I. Daidone, I. Sokolov, and J. Smith, Phys. Rev. Lett. **100**, 188103 (2008).
[8] O. B. Okan, A. R. Atilgan, and C. Atilgan, Biophys. J. **97**, 2080 (2009).
[9] A. Mitsutake, H. Iijima, and H. Takano, J. Chem. Phys. **135**, 164102 (2011).
[10] R. Metzler, J. Klafter, J. Jortner, and M. Volk, Chem. Phys. Lett. **293**, 477 (1998).
[11] Y. Shao, Physica D **83**, 461 (1995).
[12] R. Metzler, E. Barkai, and J. Klafter, Phys. Rev. Lett. **82**, 3563 (1999).
[13] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[14] G. Kneller, Phys. Chem. Chem. Phys. **7**, 2641 (2005).
[15] V. Calandrini, D. Abergel, and G. Kneller, J. Chem. Phys. **133**, 145101 (2010).
[16] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955).
[17] *NIST Handbook of Mathematical Functions*, edited by F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (Cambridge University Press, 2010).
[18] J. Boon and S. Yip, *Molecular Hydrodynamics* (McGraw-Hill, New York, 1980).
[19] J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, 1986).
[20] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
[21] C. Gardiner, *Handbook of Stochastic Methods*, 2nd ed., Springer Series in Synergetics Vol. 13 (Springer, Berlin, 1985).
[22] N. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992) (revised ed.).
[23] K. Hinsen, A. Petrescu, S. Dellerue, M. Bellissent-Funel, and G. Kneller, Chem. Phys. **261**, 25 (2000).
[24] K. Moritsugu and J. C. Smith, J. Phys. Chem. B **110**, 5807 (2006).
[25] J. Karamata, J. Reine Angew. Math. **1931**, 27 (1931).
[26] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. (Wiley, New York, 1971).
[27] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[28] M. Lyra and C. Tsallis, Phys. Rev. Lett. **80**, 53 (1998).
[29] J. Naudts, J. Phys.: Conf. Ser. **201**, 012003 (2010).
[30] J. Kirchmair, P. Markt, S. Distinto, D. Schuster, G. Spitzer, K. Liedl, T. Langer, and G. Wolber, J. Med. Chem. **51**, 7021 (2008).
[31] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[32] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).
[33] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[34] J. Izaguirre, D. Catarello, J. Wozniak, and R. Skeel, J. Chem. Phys. **114**, 2090 (2001).
[35] S. Nosé and M. Klein, Mol. Phys. **50**, 1055 (1983).
[36] C. Brooks and M. Karplus, J. Mol. Biol. **208**, 159 (1989).