How Stokes Shift Relaxation Reports on Poincaré Recurrences in Host Dynamics

Alexey K. Mazur
UPR9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité,
Institut de Biologie Physico-Chimique,
13, rue Pierre et Marie Curie, Paris, 75005, France

Recently it has been revealed that the equilibrium hydrogen-bond breathing dynamics of terminal base pairs in short DNA exhibit a power-law relaxation similar to that in the time-resolved Stokes shift experiments with an intercalated coumarin probe. Here a simple theory is proposed that explains the Stokes shift signal to the statistics of Poincaré recurrences in the base-pair breathing. This theory can explain the origin of the observed slow non-exponential relaxation in time-resolved Stokes shift data for DNA as well as other complex systems. It turns out that an intercalated coumarin greatly increases the breathing fluctuations in the neighboring base pairs. This motion is qualitatively similar to that in terminal residues, with the same exponent in the power-law relaxation decay. The breathing dynamics is transmitted to the photoprobe by direct contacts between aromatic π orbitals of stacked bases.

PACS numbers: 05.45.-a, 05.45.Ac, 05.45.Jn

The emission spectra of fluorescent dyes are usually red-shifted with respect to the corresponding adsorption spectra (Stokes shift). This effect occurs because, during the lifetime of the excited state, a part of the excitation energy is dissipated through interactions with the media [1]. In fluorescent photprobes, the optical energy is dissipated through interactions with the media spectra (Stokes shift). This effect occurs because, during red-shifted with respect to the corresponding adsorption in time from $10^{-7}$ to $10^{-15}$ sec, with the upper boundary set by the lifetime of the photoprobe [10]. This intriguing result was considered in many theoretical studies [15,23], but its origin remains controversial [11,20,23,24].

Recently it has been revealed by all-atom MD simulations that the hydrogen-bond (HB) breathing dynamics of terminal base pairs in short DNA exhibits a power-law relaxation typical for chaotic systems with a mixed (chaotic and regular) phase space [25]. The breathing was followed by measuring the statistics of Poincaré recurrences for distances ($R$) between atoms that form Watson-Crick (WC) H-bonds. The stopwatch was started when a given distance exceeded a certain threshold ($R_{th}$) and stopped once the boundary was crossed in the opposite direction. These events are called Poincaré recurrences. The celebrated Poincaré recurrence theorem of 1890 [26] guarantees that a dynamical trajectory with a fixed energy and bounded phase space will always return in a close vicinity of the initial state. The statistics of recurrences is defined as the probability distribution $P(\tau)$ of returns with times longer than $\tau$. For dynamical systems with hard chaos it behaves similarly to coin flipping and drops exponentially [27,28]. However, in the generic case of chaos with divided phase space where islands of integrable motion are embedded in a chaotic sea [29,30], it was established that $P(\tau)$ is described by an algebraic decay

$$P(\tau) \propto 1/\tau^\beta,$$

with the Poincaré exponent $\beta \sim 1.5$ [31,32]. A slightly smaller value $\beta \sim 1.2$ was obtained for the base-pair breathing dynamics in DNA [25]. This universal behavior, referred to as Algebraic Statistics of Poincaré Recurrences (ASPR), is due to sticking of trajectories in the vicinity of stability islands. It is possible that the power-law component in the TRSS relaxation in DNA is a manifestation of the inherent chaos of the base-pair breathing dynamics.

The response function measured in the TRSS experiments is

$$S(t) = \frac{\nu(t) - \nu(\infty)}{\nu(0) - \nu(\infty)}$$

where $\nu(t)$ is the time dependent emission frequency. Usually $S(T)$ is reduced to the equilibrium autocorrelation function of the solvation energy $E(t)$ using the linear response approximation, which gives

$$S(t) \approx C(t) = \frac{\langle \delta E(0) \delta E(t) \rangle}{\langle \delta E(0)^2 \rangle}$$

where $\delta E(t)$ denotes a deviation from the average. The numerator in Eq. [3] is naturally split into contributions from different factors. If there is a coumarin-DNA interaction characterized by a specific decay law it is represented as an additive term in Eq. [3]. In the chaos
theory, the autocorrelation function $C(\tau)$ and the probability $P(\tau)$ are related as

$$C(\tau) \propto \tau P(\tau),$$

which gives for the base-pair breathing dynamics $C(\tau) \propto 1/\tau^{0.2}$ and a very good apparent agreement between MD [25] and earlier TRSS data [10].

A closer look reveals, however, that the above result probably is a casual coincidence. Indeed, Eq. (4) follows from the following reasoning. Suppose we have a step function $f(t)$ such that $f = 1$ on the intervals counted as Poincaré returns and $f = 0$ elsewhere. The unnormalized autocorrelation function is defined as

$$C(\tau) = \int_0^\infty f(t)f(t+\tau)dt.$$ 

(5)

Non-zero contributions to this integral come from returns longer than $\tau$. Duration $t > \tau$ gives a contribution $t - \tau$ while the number of such returns is $-dP(t)$, therefore,

$$C(\tau) = -\int_\tau^\infty (t-\tau)P'_t dt = -\int_\tau^\infty tP'_t dt + \int_\tau^\infty \tau P'_t dt.$$ 

Assuming that $P(\tau)$ is smooth and integrable, the above integrals readily result in

$$C(\tau) = \int_\tau^\infty \tau P(\tau)dt$$

and Eq. (4) for the power-law decay. The above derivation works for canonical chaotic systems [32, 35] because trajectories are quasi-periodic during long stays near stability islands, chaotic in between them, and different quasi-periodic intervals are uncorrelated. For HB-distances these assumptions fail, and, indeed, it was found that the spectral density of distance fluctuations is described by a power-law that does not correspond to $P(\tau)$ [25]. Even assuming that the TRSS response is a step-function of $R$, one cannot in Eq. (5) neglect correlations due to closed state dynamics which is evidently quasi-periodical.

I surmised that the base-pair breathing might not affect the measured TRSS signal directly as suggested by Eq. (2) and (3). Instead, this conventional theory can be supplemented with the following two-state model. A coumarin probe may receive a photon when the neighboring base pairs are closed ($A$) or partially opened ($B$). When a base pair is opened, new water molecules come in between the bases, which changes the preferred ion locations. This can affect the probe in various ways, for instance, by reducing the excitation lifetime or shifting the steady emission spectrum. In both cases the solvent relaxation and the red-shifted emission occur according to the conventional scenario described by Eq. (2) and (3). At the same time, transitions between states $A$ and $B$ take place.

The laser impulse highlights a subset of excited coumarins. We see only them and their number is steadily decreasing. It is reasonable to assume that states $A$ and $B$ differ by (1) the emission spectra, and/or (1a) the rates of TRSS relaxation, and/or (1b) the excitation lifetimes. Finally, (2) the photoexcitation may shift the $A \rightleftharpoons B$ equilibrium. If only (1) is true the $A \rightleftharpoons B$ dynamics should not influence the TRSS signal because the subset of excited coumarins is always in equilibrium and the average spectrum does not change. Situation (1)+(1a) is very complex because after every $A \rightleftharpoons B$ transition a new relaxation process starts form an undefined state. For simplicity, we assume that the solvent TRSS relaxation occurs on a much faster time scale [37], therefore, cases (1) and (1a) can be merged.

Now consider situation (1)+(1b). Although the $A \rightleftharpoons B$ equilibrium is maintained as a whole, in the visible subset it will be perturbed, with the average spectrum shifted. This shift may occur towards both higher and lower frequencies and result in small modulations of the TRSS profiles on the background of the overall red shift. The $A \rightleftharpoons B$ transitions within the visible subset will tend to recover the equilibrium and this dynamics should reflect upon the measured TRSS signal. For an external observer, situation (1)+(1b) is indistinguishable from (1)+(2), and we can merge (1b) and (2). Situation (1)+(2) is studied below in detail in order to understand what should be seen in the TRSS experiments in this case.

Consider the relaxation of a perturbed equilibrium in the simplest system $A \rightleftharpoons B$ with the vector of concentrations $C = \begin{pmatrix} C_A \\ C_B \end{pmatrix}$. Its kinetics is described by a linear system $\dot{C} = MC$ with matrix $M = \begin{pmatrix} -k_+ & k_- \\ k_+ & -k_- \end{pmatrix}$, eigenvalues $\lambda_1 = 0$ and $\lambda_2 = k_+ + k_-$, and unnormalized eigenvectors $e_1 = \begin{pmatrix} k_- \\ k_+ \end{pmatrix}$ and $e_2 = \begin{pmatrix} +1 \\ -1 \end{pmatrix}$, respectively. Vectors $e_1$ and $e_2$ describe the equilibrium and the perturbation, respectively. The relaxation is exponential: $C = C^0 e_1 + \delta e_2 \exp(-\lambda_2 t)$. After a perturbation the system can be divided into two components. The first one, $C^0 e_1$, is constant and invisible in experiment. The second one, $\delta e_2$, gives a contribution proportional to the spectral difference between states $A$ and $B$. It decays with the rate constant $\lambda_2$, which results in a time-dependent spectral shift measured by the TRSS method. This decay looks irreversible because vector $e_2$ is antisymmetric, which corresponds to an excess in $A$ and a hole in $B$ or vice versa. The hole is gradually closed without creating an opposite flow. If $k_- \gg k_+$ then $\lambda_2 = k_-$, that is, even if the perturbation is made by adding molecules to $A$ the relaxation looks like a faster $B \rightarrow A$ transition.

In this case the hole in $B$ is closed by the equilibrium $A \rightarrow B$ flaw which is strong because $C^0_A \gg C^0_B$ according
Now assume that $A$ and $B$ correspond to the closed and opened states, respectively, of a base pair in the vicinity of the coumarin probe. Transition $B \rightarrow A$ is not a Markov process because the probability of closing depends upon the previous history and the time spent in the open state. When the laser pulse arrives the system is in detailed balance equilibrium, that is, the $A \rightarrow B$ and $B \rightarrow A$ flows are mutually compensated due to appropriate stationary trapping time distributions in both states. One can reasonably assume that the shapes of these distributions do not change in the subset of excited coumarins. For instance, if the coumarin excitation lifetime is smaller in state $B$ this sub-population will disappear more rapidly, but the trapping time distribution is not affected because the rate of quenching does not depend upon the previous history of the opened pair. With constant trapping time distributions, the flaws are proportional to concentrations, which makes the problem equivalent to the previous linear case. A perturbed equilibrium state can be decomposed into the constant and variable components $C^0 e_1$ and $δ e_2$, respectively, and we consider the second one. At this stage the equivalence Hamiltonian equations [38] and a symplectic integrator [39, 40] with the time step of 0.01 ps [41, 42]. The integrated probability distribution $P(\tau)$ is obtained from the hole is closed, with the stationary trapping time distribution, that is, the probability density of pairs opened already for time $t$ is $P(t)dt = \nu P(\tau)d\tau$, with the total number of opened pairs $C_B(0) = \int_0^\infty P(\tau)d\tau$. Consider the case $\delta < 0$. At time $t = 0$ the opening of new base pairs stops and we have an excess $[\delta]$ of opened pairs with the trapping time distribution $\nu P(\tau)$. We are interested in the time dependence $C_B(t)$ of the number of opened pairs. The probability that a pair already opened for time $\tau$ will stay opened during time $t$ is $P(\tau + t)$. Therefore, $C_B(t) = \int_0^\infty \nu P(\tau + t)d\tau = \nu \int_0^\infty P(\tau)d\tau$. This is the required result, i. e., the response signal measured in the TRSS experiments. Interestingly, it coincides with Eq. [0]. Now consider the case $\delta > 0$. At time $t = 0$ we have a hole in the equilibrium distribution of opened pairs which is filled by a constant flow from the equilibrated pool of closed pairs. The equilibrium will be reestablished when the hole is closed, with the stationary trapping time distribution recovered. The time dependence is computed as $C_B(t) = \nu \int_0^\infty P(\tau)d\tau = C_B^0 - \nu \int_0^\infty P(\tau)d\tau$, again in agreement with Eq. [0].

This simple theory suggests that the power-law relaxation in the TRSS experiments [7] [10] can be caused by the base-pair breathing dynamics. The equilibrium population of partially opened pairs is low, which means that the rate of base-pair closing is much larger than that of opening. Therefore, the decay of fluctuations is dominated by base-pair closing and should be described by the ASPR [25]. There is, however, one important caveat. The ASPR was found only for base-pair breathing of terminal base pairs in short DNA. Similar measurements for internal pairs revealed only fast exponential relaxation [25]. In the TRSS experiments, the coumarin probe was placed inside DNA a few helical turns apart from both ends [7] [10]. At such distances a causal relationship between end fraying and the measured signal is hardly possible even assuming allosteric effects through the DNA structure. It turns out, however, that the coumarin probe intercalated into the base pair stack greatly increases the HB-breathing in its neighbors, so that the ASPR can be confirmed by the same methods.

The statistics of Poincaré recurrences was studied for the model system shown in Fig. 1. The right panel displays a pentamer duplex placed in a small water box of 456 water molecules with periodic boundaries. Eight sodium ions are added for neutralization. The sequence of this DNA is GCGCGC CG CG, where M stands for coumarin. In one strand the central base is replaced by the photoprobe and in the opposite strand the central sugar has no attached base (abasic site). This duplex is a minimal representative fragment of a longer DNA molecule used in experiments [7] [10].

All-atom MD simulations were carried out as described earlier [25], under normal temperature using internal coordinate Hamiltonian equations [38] and a symplectic integrator [39, 40] with the time step of 0.01 ps [41, 42]. The integrated probability distribution $P(\tau)$ is obtained by counting the number of recurrences with duration larger than $\tau$ and normalizing it by the total number of events, that is, $P(0) = 1$ by construction. Function $P(\tau)$ is positive definite and, due to statistical averaging over a large number of crossings, stable with respect to

**FIG. 1:** (Color online) The upper part of the left panel displays the WC base pair formed by guanine and cytosine, with three H-bonds shown by thick dashed lines. The lower part shows a coumarin residue (C102). It was inserted into the DNA stack as shown in the right panel. The residue orientation in the left panel correspond to those in the DNA stack, with the major groove edges facing down. In the right panel the coumarin residue and sodium ions are shown as spheres.
fluctuations (see e.g. discussion in [55]). Importantly, these computations are trivially parallelizable, that is, the $P(\tau)$ statistics can be accumulated in a large number of independent MD trajectories. The results discussed below were obtained by using parallel computations on 129 cores for the model system that had 3557 degrees of freedom for 1665 atoms. Owing to the small system size a large total sampling volume of about 100 $\mu$s was possible. The recent version of the AMBER force field [43, 45] was used with SPC/E water [46]. The coumarin partial charges were obtained as recommended by the RESP method [47, 48]. Other details can be found in the Supplemental Material [49].

FIG. 2: (Color online) The left panel shows the statistics of Poincaré recurrences $P(\tau)$ for four different O6N4 H-bonds (see Fig. 1). Distance $R$ was measured between the oxygen and nitrogen atoms. For the two base pairs neighboring to coumarin the results are shown by solid black and dashed red lines respectively. The thinner blue solid and dotted lines display the analogous results for breathing of terminal and internal base pairs, respectively, in the absence of coumarin [25]. The dash-and-dot straight line shows the power law decay fit to the upper profile with the Poincaré exponent $\beta = 1.27$. The threshold distance is $R_{th} = 3.15$ Å in all three cases. The right panel shows the corresponding dependencies of average distances $\langle R \rangle$. The logarithms are decimal.

It was found earlier that the O6N4 H-bond is a better indicator of breathing than other H-bonds because base pairs commonly start to open from the major groove side, therefore, the O6N4 distance allows counting partial base-pair openings [25]. Fig. 2 shows results obtained for such H-bonds in the two base pairs flanking the coumarin residue in the DNA structure in Fig. 1. The central N1N3 H-bonds opened less frequently than the O6N4 H-bonds, and the N2O2 H-bonds remained always closed. The bases did not flip out. These new data are compared in Fig. 2 to analogous earlier results for breathing of terminal and internal base pairs. The threshold $R_{th} = 3.15$ Å is close to the equilibrium H-bond lengths, therefore, a large fraction of recurrences results from oscillations within the bonded ground state, which gives for small $\tau$ an exponential decay of $P(\tau)$ seen as a rapid fall in Fig. 2. Deviations towards a power-law relaxation become visible in the sup-picosecond time range, but for internal base pairs they are very small. Fig. 2 reveals that base pairs at both sides of the intercalated coumarin are perturbed, and in one of them the O6N4 H-bond is opened rather frequently. The right panel of Fig. 2 suggests that the breathing of internal and terminal base pairs is qualitatively similar, with the frequency scaled down. The exponent in the power-law decay of $P(\tau)$ is $\beta \sim 1.2$ in both cases. This motion should affect the TRSS signal quite strongly because stacked aromatic bases contact one another by $\pi$ orbitals, and optical $S0 \leftrightarrow S1$ transitions in the coumarin probe are likely to be sensitive to that.

One last counter-argument needs to be addressed. The population of partially opened base-pair states is low and perhaps insufficient to account for the experimental data quantitatively. This issue presently involves too many unknowns. On the one hand, the corresponding experimental populations are not known. On the other hand, the frequency of partial openings in MD might be underestimated. It depends upon the forcefield and the coumarin orientation in the DNA stack. There are eight possible distinct conformations of stacked coumarin and they could not all be tested. We also should keep in mind that the fluorescence quantum yield can be higher in coumarins flanked by partially opened base pairs. These issues require additional studies.

In summary, it is shown that, in contrast to traditional statistical mechanics explanations, the power-law TRSS relaxation in DNA is likely to result from chaotic dynamics of a few or even one degree of freedom of HB-breathing in the base pairs flanking the photoprobe in the stack. A simple theory is proposed that relates the base-pair breathing with the TRSS measurements. This theory is applicable to other controversial cases of slow algebraic TRSS relaxation in complex systems.

The computational resources used in this study were supported by the "Initiative d’Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01).
