

# Computational Study of the Force Dependence of Phosphoryl Transfer during DNA Synthesis by a High Fidelity Polymerase

Ravindra Venkatramani and Ravi Radhakrishnan\*

Department of Bioengineering and Department of Biochemistry & Biophysics, University of Pennsylvania, 240 Skirganich, 210 S. 33rd St, Philadelphia, Pennsylvania 19104, USA

(Received 8 January 2007; published 26 February 2008)

High fidelity polymerases are efficient catalysts of phosphodiester bond formation during DNA replication or repair. We interpret molecular dynamics simulations of a polymerase bound to its substrate DNA and incoming nucleotide using a quasiharmonic model to study the effect of external forces applied to the bound DNA on the kinetics of phosphoryl transfer. The origin of the force dependence is shown to be an intriguing coupling between slow, delocalized polymerase-DNA modes and fast catalytic site motions. Using noncognate DNA substrates we show that the force dependence is context specific.

DOI: 10.1103/PhysRevLett.100.088102

PACS numbers: 87.15.A-, 87.14.G-, 87.15.La, 87.15.R-

The tremendous catalytic power of high fidelity DNA polymerases, which enables them to perform DNA base pair synthesis (BPS) with high efficiency as well as high specificity (error rate of 1 per  $10^6$ – $10^8$  base pairs synthesized) is a subject of intense experimental [1] and theoretical [2,3] investigations. During BPS a complementary nucleotide triphosphate (dNTP) is incorporated into one of the strands (denoted as primer) of a double stranded DNA molecule, opposite an unpaired base in the other strand (denoted template). The basic reaction (chemical step) is the formation of a covalent phosphodiester bond between the incoming dNTP and the terminal primer base (Fig. 1), prior to which the enzyme has to switch to an “active” state through a large scale conformational rearrangement of the fingers (or thumb) domain [1]. Subsequently, the enzyme switches back to an “inactive” state and the cycle repeats. Based on a large body of structural and biochemical data, it is known that the identity of the rate-limiting step in the BPS cycle depends not only on the type of polymerase, but also the base pairs involved in the synthesis [1]. During correct incorporation, either the conformational change preceding the chemical step (*E. coli* pol I) or the chemical step itself (pol  $\beta$ ) can be rate limiting, while for incorrect incorporation indirect evidence suggests that for several polymerases the chemical step is rate limiting. Using novel tools for single molecule experimentation [4–7] pioneering studies of polymerases replicating DNA stretched under differing tensions [8,9] show that the replication rate is highly sensitive to forces exerted on the template strand (replication rate increases with forces less than 4–7 pN and decreases for higher values of the force); this implies that the applied force affects the rate-limiting step. Theoretical models parameterized by experimental data [8,9] and atomistic simulations [10] suggest that the rate-limiting step in this case involves the rearrangement of more than one base pair. While such models are instructive the force dependence is explained purely in terms of DNA elasticity neglecting the coupling between the complex cooperative

motion of polymerase and DNA degrees of freedom at the catalytic site. In this Letter, we show that such a coupling leads to a new paradigm of force dependence of the kinetics of the chemical step in these systems. This is particularly useful for extending the experimental studies to systems with mispaired DNA substrates where the chemical step is indeed believed to be rate limiting. Through a correlation analysis of molecular dynamics (MD) simulation data, we suggest that the external force on the DNA template strand impacts the chemical step due to a coupling between the primary reactive distance for phosphoryl transfer (the  $O3'-P_\alpha$  distance in Fig. 1) and slow delocalized enzyme-DNA motions [11]. The presence of such a coupling adds to accumulating evidence of a prechemistry phase during catalysis, in which delocalized (global), and local motions of the enzyme-substrate complex driven by thermal fluctuations synergistically orchestrate the assembly of an optimal catalytic site [12,13]. We also show that the coupling is disrupted to varying extents when the

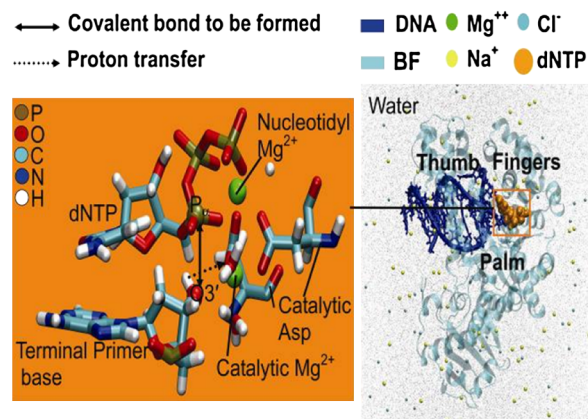


FIG. 1 (color online). (Right) Solvated BF pol/DNA/dNTP ternary complex. (Left) Catalytic site showing key elements of the phosphoryl transfer reaction. A covalent bond is formed between  $\alpha$ -phosphorous of the incoming nucleotide and the  $O3'$  oxygen of the terminal primer base.

inserted nucleotide is not complementary to the template base or when the template base is oxidatively damaged. We combine our correlation analysis with a quasiharmonic approximation to derive the context-specific force dependence of the phosphoryl transfer step for four polymerase-DNA complexes.

In this study, we have chosen to work with the large fragment from *Bacillus stearothermophilus* (Bacillus fragment or BF) even though previous single molecule data are available for the highly homologous (49% sequence identity) Klenow fragment (or KF) because BF allows us to explore four systems: correct dCTP or incorrect dATP insertion opposite an undamaged or oxidatively damaged guanine (G) template base. We denote the systems as (1) G : C, (2) G : A, (3) 8oxoG : C, and (4) 8oxoG : A. For KF, (3) and (4) are not physical because unlike BF, KF is not known to bypass the 8oxoG lesion; see Sec. S.1 and Fig. S1 in [14] for a sequence and structural comparison of BF with KF and T7 polymerases. Four fully solvated model complexes of BF pol [15] with DNA + dNTP substrate are prepared for G : C, G : A, 8oxoG : C, and 8oxoG : A systems. In the 8oxoG models G is replaced with 8-hydroxoguanine (8oxoG), one of the most prominent lesion generated by oxidative damage. We perform all atom MD simulations of fully solvated complexes using the CHARMM force field [16] and the NAMD [17] simulation package. Following standard equilibration protocol (see [14] for details on methods and analyses), 10 ns constant volume and temperature MD trajectories are obtained for the four model systems and data from the last 5 ns are utilized in the subsequent analysis.

In the single molecule experiments, a force applied to the template strand of the DNA fragment alters the replication rate. Here, using a minimal model, we specifically discuss the effect of the force on the chemical step during BPS. If A and B denote the ground and transition states for catalysis, then according to transition state theory the “equilibrium constant” for the system to switch from state A to B in the absence of any external force is (see Fig. 2):  $K_{eq}(0) \propto \exp(-\Delta G/k_B T)$ , where,  $\Delta G$  is the free energy difference between A and B [18]. A comparison of equilibrated structures (Fig. 1) with “ideal” transition state structures in similar systems [19,20] shows that the primary reactive distance, X (signifying nucleophilic attack) for catalysis is the distance between the terminal primer adenine (A) O3' atom and the  $P_\alpha$  atom of the incoming dNTP. We assume that the passage from state A to B is approximated purely by a change in the coordinate X, which is harmonic (to a first approximation) close to the catalytic ground state with a spring constant  $k_x$  and equilibrium value  $X_A$  (Fig. 2). Then, in the absence of an external force the total (free) energy cost to drive the system from state A to B is:  $W(0) = 0.5k_x[X_B(0) - X_A(0)]^2$ . Within the linear response limit, the applied force will shift the ground state equilibrium position  $X_A(F)$  and the position of the transition state  $X_B(F)$  [18]. Thus, the

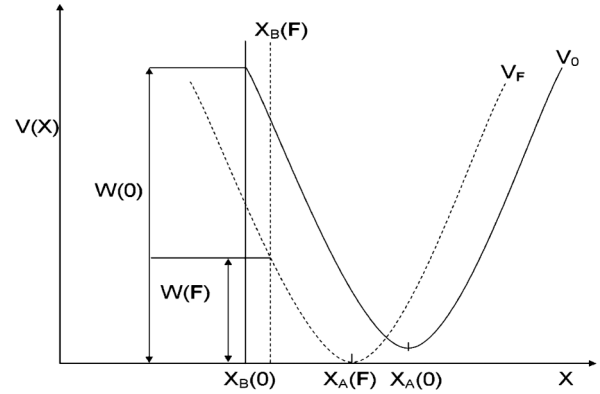


FIG. 2. The potential surface along the catalytic reactive distance  $X$  is given by  $V_0 = 0.5k_x[X - X_A(0)]^2$ . An external force  $F$  shifts the equilibrium position of the ground state A from  $X_A(0)$  to  $X_A(F)$  and that of the transition state B from  $X_B(0)$  to  $X_B(F)$  to alter the barrier for catalysis. The new potential surface for  $X$  is  $V_F = 0.5k_x[X - X_A(F)]^2 - [F_x/(2k_x)]$ , where  $F_x$  is the force along  $X$ .

change in the total (free) energy cost to transition from state A to state B, i.e.,  $\Delta W = W(0) - W(F)$ , under an applied force  $F$  acting along  $X$  is  $\Delta W(F) = W(0) - 0.5k_x[X_B(F) - X_A(F)]^2$  and the  $K_{eq}$  under force  $F$  can be written as [18]  $K_{eq}(F) \propto \exp(-[\Delta G - \Delta W(F)]/k_B T)$ .

A force applied on the template strand (as in the single molecule studies) will change  $X$  only if a coupling exists between  $X$  and the DNA template strand. We explore this coupling by carrying out a principal component analysis (PCA) [21] on an “active-site region”, including the catalytic site (Fig. 1) and comprising nonhydrogen atoms of the incoming dNTP, six residues of the DNA template strand, four residues of the DNA primer strand, the two  $Mg^{2+}$  ions, two polymerase aspartate residues D830 and D653 critical for catalysis, and bound waters at the catalytic site (see Sec. S.2 in [14] for a discussion on this choice for the subset). The plot of the variance-covariance matrix ( $\sigma$ , see Sec S.1.4 in [14]) quantifies the correlations in atomic motions in the active-site, (see Figs. 3 and S8–S11 in [14]) and the PCA of  $\sigma$  results in  $\xi = (\xi_1, \xi_2, \dots, \xi_{3N-6})$  orthogonal eigenvectors (PCs) with eigenvalues  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{3N-6})$  sorted in descending order, i.e.,  $\lambda_1 > \lambda_2 > \dots > \lambda_{3N-7} > \lambda_{3N-6}$ ; visualization of the PCs suggests how the polymerase utilizes the dynamical coupling to possibly drive the reaction (see movie S6 in [14]). Intriguingly, we find that the correlations in  $\sigma$  and hence the dynamical coupling of the PCs with  $X$  are both context-specific, see Figs. 3, S8–S11 and Table I. For the G : C system, we find that PCs that significantly influence  $X$  (Table I) are also among the top (10) PCs which account for most (>70%) of the atomic fluctuations in the MD trajectories. Such a strong coupling is disrupted to varying degrees for the three noncognate systems (Table I). The contrast in correlations in Fig. 3 (between G : C and G : A) and more generally in Figs. S8–S11 (among the four systems) determines the context-specific dynamical cou-

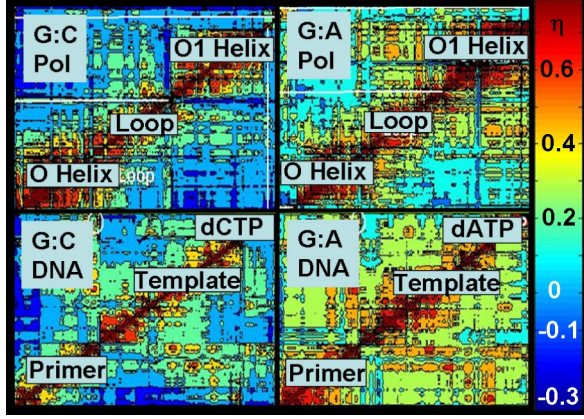


FIG. 3 (color online). Correlations between vector displacements ( $\mathbf{r} - \langle \mathbf{r} \rangle$ ) of atoms for the **G : C** and **G : A** systems; vector  $\mathbf{r}$  is from origin to the atom of interest, and  $\langle \mathbf{r} \rangle$  its average. Enlarged views for all 4 systems are provided in Figs. S8–S11 in [14].

pling in Table I, wherein the catalytic site reactive distances are coupled to global motions of the BF/DNA/dNTP complex.

Next, we discuss how the context-specific dynamic coupling relates to the force applied to the DNA template strand to influence catalysis. Each eigenvector  $\xi_m$  represents a collective mode of motion in the active-site region of the ternary complex, with spring constant  $k_m = k_B T / \lambda_m$ , (the equality stems from the quasiharmonic approximation [21]). A force  $\mathbf{F}$  applied along the DNA helix axis as in the case of the single molecule experiments will displace each  $\xi_m$  by,  $a_m = |\mathbf{F}| \cos \theta_m / k_m$ , where,  $\theta_m$  is the projection angle of  $\mathbf{F}$  along  $\xi_m$ . Since the force is applied at the ends of the DNA template strand we project only the motions of template strand backbone (phosphate) atoms along the helix axis in calculating  $\theta_m$  for each PC, see section S.3 in [14] for further discussion. The resultant change in the active-site geometry due to the applied force  $\mathbf{F}$  along the DNA duplex axis is then  $\mathbf{R}(\mathbf{F}) - \mathbf{R}(0) = \sum_m a_m \xi_m$ , where,  $\mathbf{R} = (X_1, X_2 \dots X_{3N})$  is the vector representing the geometry of the active-site region of  $N$  atoms. Here,  $\mathbf{R}(0)$  is the ground state active-site geometry at zero force (MD simulation average), for which  $X = X_A(0)$ , and  $\mathbf{R}(\mathbf{F})$  is the new active state geometry due to  $\mathbf{F}$  for which

TABLE I. The 5 modes showing the strongest correlation with  $X$ , measured by the coefficient  $\eta$ , defined for two independent variables  $x, y$  as  $\eta_{xy} = \frac{\text{Covariance}(x,y)}{\sqrt{\text{variance}(x)}\sqrt{\text{variance}(y)}}$  with  $-1 < \eta_{xy} < 1$ .

G : C		G : A		8oxoG : C		8oxoG : A	
Mode	$\eta$	Mode	$\eta$	Mode	$\eta$	Mode	$\eta$
1	0.34	105	0.27	10	0.33	1	-0.23
3	-0.32	112	-0.22	9	0.32	20	0.21
2	0.31	54	0.21	4	0.27	47	0.18
8	0.24	115	0.19	6	-0.22	23	-0.17
5	-0.18	106	-0.19	15	0.22	71	-0.16

$X = X_A(\mathbf{F})$ . The force along  $X$  which causes a displacement,  $\Delta X_A(\mathbf{F}) = X_A(\mathbf{F}) - X_A(0)$  is:  $F_x(\mathbf{F}) = -k_x \Delta X_A(\mathbf{F})$ . The spring constants  $k_x$  can be obtained (Table II) from the distribution (histogram) of  $X$  values  $P_x$  in our ground state MD trajectories by fitting a harmonic function to the energy of the distribution  $E(X) = (1/2)k_x X^2 = -k_B T \ln(P_x)$ . Since the free energy surface projected along the reaction coordinate has a maximum at the transition state, its the negative curvature is approximated as  $k_{ts} = -\mu(\omega_{ts})^2 = -2.77 \times 10^3 \text{ pN/\AA}$ , where  $\mu$  is the reduced mass of the  $O3'-P_\alpha$  virtual bond =  $16 \times 32/48 \text{ AMU}$  and  $\omega_{ts}$  is the passage time in transition state theory,  $\omega_{ts} = -k_B T / \hbar$ . Assuming that the same force acts on  $X$  throughout the systems' passage from ground to transition state (the dynamic coupling is unaltered), the displacement of the transition state value of  $X$ ,  $\Delta X_B(\mathbf{F}) = X_B(\mathbf{F}) - X_B(0)$  is given as

$$\Delta X_B(\mathbf{F}) = \frac{F_x(\mathbf{F})}{k_{ts}} = -\frac{k_x}{k_{ts}} \Delta X_A(\mathbf{F}). \quad (1)$$

Note that the displacements  $\Delta X_A(\mathbf{F})$  and  $\Delta X_B(\mathbf{F})$  have opposite signs due to positive (negative) curvatures of the free energy for ground (transition) states. In our calculations  $X_B(0) = 1.7 \text{ \AA}$ , the average value of the  $O3'-P_\alpha$  distance seen in structures of post-catalytic (product) complexes. We have used the full set of 3N-6 PC modes to calculate the displacements  $\Delta X_A(\mathbf{F})$  and  $\Delta X_B(\mathbf{F})$  for all four models and obtained the ratio  $K_{eq}(F)/K_{eq}(0) = \exp[\Delta W(F)/k_B T]$  of phosphoryl transfer rates in the presence of an external force relative to that in the zero force limit. The phosphoryl transfer rate increases with force (ratio  $> 1$ ) if  $\Delta W$  is positive, i.e., if the force reduces  $X$  from its average value and decreases (ratio  $< 1$ ) for negative  $\Delta W$ . In Fig. 4, we plot  $K_{eq}(F)/K_{eq}(0)$  versus force in the range  $|\mathbf{F}| = 1\text{--}6 \text{ pN}$  for all four models. We consider two cases (that represent extreme limits and the correct scenario is likely in-between): (a) no TSS:  $\mathbf{F}$  does not change  $X_B(0)$  i.e. that the dynamic coupling present in state  $A$  is absent (completely disrupted) in state  $B$  and (b) with TSS:  $\mathbf{F}$  changes  $X_B(0)$  to  $X_B(\mathbf{F})$  and the dynamical coupling is retained throughout the systems passage from state  $A$  to  $B$ . Based on Fig. 4, our predictions are that for three systems **G : C**, **8oxoG : C**, and **8oxoG : A** the applied force induces a net reduction in  $X$ , and hence an increase in the phosphoryl transfer rate with increasing applied force. For **G : C** incorporation in KF this trend matches the experiments although experimental evidence also shows that the chemical step is not rate limiting for **G : C**. For the **G : A** system, the opposite is true, i.e., with

TABLE II. Spring constants ( $\times 10^3 \text{ pN/\AA}$ ) for all four models estimated from MD simulation data.

Model	G : A	G : A	8oxoG : C	8oxoG : A
$k_x$	2.24	0.62	0.21	2.55



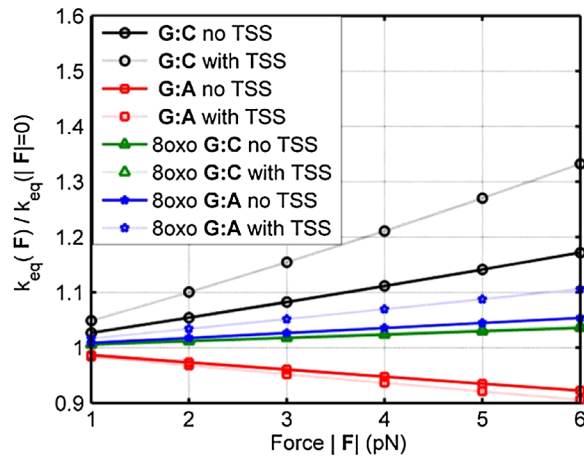


FIG. 4 (color online). Relative catalytic rate vs force on the DNA template strand. TSS: Transition state shift.

force,  $X$  increases causing a net decrease in the phosphoryl transfer rate. Thus, the increase or decrease of the relative replication rate with force is context-specific and is governed both by the dynamic coupling of the PCs with  $X$  (Table I) and the projection of the PCs along  $\mathbf{F}$  (see also section S.3 of [14]). These predictions, especially for the noncognate substrates where phosphoryl transfer is indeed rate limiting (and hence governs the overall replication rate), are new and can be directly tested by carrying out single molecule studies of DNA replication in response to weak applied forces. The dynamic coupling itself is strongest for the  $\mathbf{G} : \mathbf{C}$  system; i.e., the slow modes have a strong correlation with  $X$  (Table I). For the noncognate substrates ( $\mathbf{G} : \mathbf{A}$ ,  $8\text{oxo}\mathbf{G} : \mathbf{C}$ , and  $8\text{oxo}\mathbf{G} : \mathbf{A}$  models) the dynamic coupling appears to be disrupted to differing extents and the most significant correlations of the  $O3'-P_{\alpha}$  distance fluctuations shift to be with higher frequency (stiffer) modes (Table I). The stiffer modes by definition tend to induce smaller displacements of  $X$  and smaller shifts of the transition state [Eq. (1)]; in our PCA, the stiffer modes also show weaker projections along the applied force. The cumulative effect is a weaker ( $\mathbf{G} : \mathbf{A}$ ,  $8\text{oxo}\mathbf{G} : \mathbf{A}$ ) or insignificant ( $8\text{oxo}\mathbf{G} : \mathbf{C}$ ) dependence of the relative phosphoryl transfer rate on the applied force for the systems with noncognate substrates (Fig. 4). PCA calculations performed using extended active-site fragments show identical phosphoryl transfer rate versus force curves as reported in Fig. 4 (Sec. S.2 of [14]), which establishes that our predictions are insensitive to our choice of the active site.

To summarize, subject to the well-appreciated approximations (these include imperfections in biomolecular force-fields, limited sampling in MD simulations, the linear response as well as quasiharmonic assumptions; in particular, our analysis would break down for strong forces due to anharmonicities, force induced mode mixing, and

the coupling of the force to reactive distances other than  $X$ ), we have identified a dynamic coupling between slow delocalized polymerase-DNA modes and the  $O3'-P_{\alpha}$  reactive distance for catalysis which is optimal only for the  $\mathbf{G} : \mathbf{C}$  system and is disrupted in systems with noncognate DNA + dNTP substrates. The emerging theme from our studies is that the dynamical coupling alters the free energy landscape for phosphoryl transfer; i.e., its context-specific manifestation impacts the kinetics of phosphoryl transfer differently for different substrates. Thus the optimization of the dynamic coupling machinery through evolution can be a possible additional mechanism in the polymerase to achieve error control during DNA replication.

Computational resources were provided in part by the National Partnership for Advanced Computational Infrastructure under Grant No. MCB060006.

\*Corresponding author.

rradhak@seas.upenn.edu

- [1] C.M. Joyce and S.J. Benkovic, *Biochemistry* **43**, 14 317 (2004).
- [2] J. Florian, M.F. Goodman, and A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6819 (2005).
- [3] R. Radhakrishnan and T. Schlick, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5970 (2004).
- [4] C. Bustamante, Z. Bryant, and S.B. Smith, *Nature (London)* **421**, 423 (2003).
- [5] R. Lavery *et al.*, *J. Phys. Condens. Matter* **14**, R383 (2002).
- [6] M.J. Schnitzer, K. Visscher, and S.M. Block, *Nat. Cell Biol.* **2**, 718 (2000).
- [7] Q. Guo and R. Sousa, *J. Mol. Biol.* **358**, 241 (2006).
- [8] G.J.L. Wuite *et al.*, *Nature (London)* **404**, 103 (2000).
- [9] B. Maier, D. Bensimon, and V. Croquette, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12 002 (2000).
- [10] I. Andricioaei *et al.*, *Biophys. J.* **87**, 1478 (2004).
- [11] R. Venkatramani and R. Radhakrishnan, *Proteins Struct. Funct. Bioinf.* (to be published).
- [12] S. Hammes-Schiffer and S.J. Benkovic, *Annu. Rev. Biochem.* **75**, 519 (2006).
- [13] R. Radhakrishnan *et al.*, *Biochemistry* **45**, 15 142 (2006).
- [14] See EPAPS Document No. E-PRLTAO-100-031808 for details of system preparation and methods along with supplementary figures and one supplementary movie. For more information on EPAPS, see <http://www.aip.org/pubservs/epaps.html>.
- [15] S.J. Johnson, J.S. Taylor, and L.S. Beese, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3895 (2003).
- [16] A.D. MacKerell *et al.*, *J. Phys. Chem. B* **102**, 3586 (1998).
- [17] J.C. Phillips *et al.*, *J. Comput. Chem.* **26**, 1781 (2005).
- [18] C. Bustamante *et al.*, *Annu. Rev. Biochem.* **73**, 705 (2004).
- [19] S.D. Lahiri *et al.*, *Science* **299**, 2067 (2003).
- [20] L.S. Beese and T.A. Steitz, *EMBO J.* **10**, 25 (1991).
- [21] A. Amadei *et al.*, *J. Biomol. Struct. Dyn.* **13**, 615 (1996).