JCTC Journal of Chemical Theory and Computation

'Dynamic Distance' Reaction Coordinate for Competing Bonds: Applications in Classical and Ab Initio Simulations

Christian Burisch,[†,‡] Phineus R. L. Markwick,[†,§] Nikos L. Doltsinis,[||,#] and Jürgen Schlitter*[,‡]

Lehrstuhl für Biophysik, Ruhr-Universität Bochum, ND 04, 44780 Bochum, Germany, Unité de Bioinformatique Structurale, Institut Pasteur, CNRS URA 2185, 25-28 Rue du Dr. Roux, 75015 Paris, France, and Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

Received July 11, 2007

Abstract: A versatile reaction coordinate, the "dynamic distance", is introduced for the study of reactions involving the rupture and formation of a series of chemical bonds or contacts. The dynamic distance is a mass-weighted mean of selected distances. When implemented as a generalized constraint, the dynamic distance is particularly suited for driving activated processes by controlled increase during a simulation. As a single constraint acting upon multiple degrees of freedom, the sequence of events along the resulting reaction pathway is determined unambiguously by the underlying energy landscape. Free energy profiles can be readily obtained from the mean constraint force. In this paper both theoretical aspects and numerical implementation are discussed, and the unique and diverse properties of this reaction coordinate are demonstrated using three examples: In the framework of Car−Parrinello molecular dynamics, we present results for the prototypical double proton-transfer reaction in formic acid dimer and the photocycle of the guanine−cytosine DNA base pair. As a classical mechanical example, the opening of the binding pocket of the enzyme rubisco is analyzed.

1. Introduction

A reaction coordinate (RC) provides a measure of the progress of an activated process, such as a chemical reaction, from an initial reactant to a final product state. The RC is usually defined in advance without prior knowledge of the actual pathway (or pathways), and so the choice of the coordinate is guided by a preliminary postulated picture of the reaction. Nevertheless, the reaction coordinate represents a valuable tool to enforce a transition away from the reactant state or toward the product state. This 'coordinate driving' approach is one of the valid methods for pathway search reviewed recently[1] which also provides a parametrization of the associated complex free energy surface. To be successful, a suitable coordinate must be constructed specifically for the problem at hand. Numerous examples can be found in the literature ranging from simple distance coordinates[2] and weighted combinations of distances[3,4] to more abstract coordination numbers[5] or even energy.[6,7] Activated processes or reactions can be driven by implementing auxiliary restraint potentials in the framework of umbrella sampling[8] or by applying holonomic constraints, as demonstrated in recent applications of targeted molecular dynamics.[9] Both these approaches enable the computation of free energy profiles using recently refined techniques[10,11] and can be applied to the study of a diverse range of systems from elementary chemical reactions to large scale conformational transitions in biological macromolecules.[12] Nevertheless, any reaction coordinate can (and usually does) provide a simplistic picture of a reaction, being a compromise between free exploration

* Corresponding author e-mail: juergen.schlitter@rub.de.
† These authors contributed equally to this work.
‡ Lehrstuhl für Biophysik, Ruhr-Universität Bochum.
§ Institut Pasteur.
|| Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum.
Present address: Department of Physics, King's College London, London WC2R 2LS, United Kingdom.

10.1021/ct700170t CCC: $40.75
Published on Web 12/08/2007

of the pathway through phase space and optimal accuracy of free energy.[13]

In this paper we present a novel reaction coordinate, the 'dynamic distance', which has been specifically designed for the study of reactions involving the rupture and/or formation of chemical bonds or contacts, such as salt bridges. This flexible reaction coordinate, formulated as a mass-weighted mean of selected interatomic distances, drives the activated process without influencing the sequence or mechanism of the events, such that the resulting reaction pathway is determined only by the underlying potential energy landscape. As a single coordinate constructed from multiple internal degrees of freedom, the dynamic distance possesses some remarkable properties, in particular its ability to automatize the search for low-energy reaction pathways and identify energetically metastable and stable states on the complex free energy surface. The dynamic distance, formulated within the general theory of reaction coordinates,[4] exhibits highly favorable mechanical and statistical properties which simplify the computation of free energies. In the following section we present the theory and general implementation of the RC. In sections 3 and 4 we demonstrate the versatility of the dynamic distance with applications employing both ab initio and classical molecular dynamics. As the choice of examples shows, while the dynamic distance is an extremely versatile constraint, it is particularly suited to the study of association and dissociation events and proton-transfer reactions, processes that play an extremely important functional role in biological systems.

2. Theory

Consider a system with $3N$ Cartesian coordinates or N position vectors in a configuration $\mathbf{x}$ given by $\mathbf{x} = (x_1....x_{3N}) = (\mathbf{r}_1....\mathbf{r}_N)$. The dynamic distance, D, defined as

$$D = \left(\sum_{\mathrm{NOP}} \frac{\mu_{ij}}{\mu^*} (\mathbf{r}_i - \mathbf{r}_j)^2 \right)^{1/2} \tag{1}$$

is the rms sum of distances between selected nonoverlapping pairs (NOP) of atoms, whose positions, $\mathbf{r}_i(t)$ and $\mathbf{r}_j(t)$, are time-dependent during the simulation. The square of each distance is weighted with the associated reduced mass, $\mu_{ij} = m_i m_j/(m_i + m_j)$, divided by an arbitrary constant mass. Setting μ^* to be the sum over all reduced masses, $\mu^* = \sum_{\mathrm{NOP}} \mu_{ij}$, the dynamic distance, D, becomes the usual rms distance if the reduced masses of all atom pairs are identical. The reaction coordinate can be employed to drive a reaction by application of a time-dependent constraint, $D = D(t)$, or to relax the system and to sample characteristic quantities at intermediate positions using scleronomic constraints (D = const). The use of D as a restraint in umbrella sampling simulations will be discussed briefly at the end of this section.

We now consider the RC as a function $\tilde{D}(\mathbf{x})$ of the Cartesian coordinates (or position vectors) and the constraint $\sigma(\mathbf{x}) = \tilde{D}(\mathbf{x}) - D = 0$. For an atom i which belongs to one of the selected atom pairs, the constraint force is given by

$$\mathbf{f}_i^{\mathrm{c}} = \lambda \frac{\partial \tilde{D}}{\partial \mathbf{r}_i} = \frac{\lambda \mu_{ij}}{D \mu^*} (\mathbf{r}_i - \mathbf{r}_j) \tag{2}$$

When the leapfrog algorithm is used to integrate Newton's equations (or more precisely the Lagrange equations of the first kind)

$$\dot{\mathbf{r}}_i = \mathbf{v}_i \text{ and } \dot{\mathbf{v}}_i = \mathbf{F}_i/m_i + \mathbf{f}_i^{\mathrm{c}}/m_i \tag{3}$$

the numerical form becomes

$$\mathbf{r}_i(t+\Delta t) = \underbrace{\mathbf{r}_i(t) + \Delta t \left(\mathbf{v}_i(t - \Delta t/2) + \Delta t \mathbf{F}_i(t-\Delta t)/m_i \right)}_{\mathbf{r}_i^*} +$$

$$\Delta t^2 \mathbf{f}_i^{\,c}(t-\Delta t)/m_i = \mathbf{r}_i^* + \delta \mathbf{r}_i \tag{4}$$

where $\mathbf{r}_i^*$ is the result of an unconstrained step in the time interval, Δt, under the influence of the force, $\mathbf{F}_i$. The Lagrange parameter $\lambda \equiv f^{\mathrm{c}}$, being the same for all atoms defined in the RC is usually called the 'constraint force' and is determined such that the constraint is satisfied. In this case, λ can be calculated directly from $\mathbf{r}_i^*$ and $\sigma(\mathbf{x})$ by means of a quadratic equation with no need for iteration. It can also be shown using (3) that there exists an analytical form for the constraint force

$$\lambda = -\frac{2K}{D} - \frac{1}{D} \sum_{\mathrm{NOP}} (\mathbf{r}_i - \mathbf{r}_j) \cdot \frac{\mathbf{F}_i m_j - \mathbf{F}_j m_i}{m_i + m_j} \tag{5}$$

where K represents the kinetic energy of the atoms involved in the constraint. This simple expression arises from the inclusion of the mass-weighting term and provides a numerical check for the constraint force calculated from the correction in (4).

Mechanical Properties. The dynamic distance formulated as described above possesses several favorable mechanical properties. First, as the RC is a function of interatomic distances which are internal coordinates, application of the constraint induces neither rotation nor translation of the system. Second, the mass-weighting procedure ensures the homogeneous action of the constraint across the system. This can be readily proven: Using the constraint forces (2) and the definition of the reduced mass, one finds that for the change of a distance due to the action of the constraint in lowest order

$$\Delta(\mathbf{r}_i - \mathbf{r}_j) = \frac{1}{2} (\Delta t)^2 \left(\frac{\mathbf{f}_i}{m_i} - \frac{\mathbf{f}_j}{m_j} \right) = \frac{1}{2} \text{ const} \cdot (\mathbf{r}_i - \mathbf{r}_j) \tag{6}$$

Obviously the relative change is the same for each atom pair, and, in particular, lighter atoms such as hydrogen are not influenced disproportionately by the constraint.

Statistical Properties. Phase space statistics can be determined by the mass-metric tensor which results in the so-called Fixman determinant[14]

$$z = \det(\mathbf{H}) = \sum \frac{1}{m_i} \left(\frac{\partial D}{\partial x_i} \right)^2 \tag{7}$$

In the present case, $z = 1/\mu^*$, which is a constant and constitutes the statistical advantage of the dynamic distance. The immediate consequence proven by Fixman[14] is a coincidence of the probability density function (pdf) of the unconstrained system in configurational space, $P(q_i,D)$, and the pdf of the system constrained to constant D, $P_c(q_i;D)$, $\{q_i,D\}$ being a complete set of generalized coordinates. As

recently shown,[11] the free energy can be obtained in a straightforward manner directly from the constraint force and the Fixman determinant. As the Fixman determinant is constant in this particular case, the relevant formula simplifies to

$$A(D) = \int \langle \lambda_D \rangle_c \mathrm{d}D \tag{8}$$

and the free energy is simply the integral over the constraint force without any correction required. Equation 8 still holds when further coordinates, such as bond lengths are constrained as long as they do not interfere with the constraint on D.[15] For activated processes along an RC, Carter et al.[16] have derived an expression for the rate and have shown that the Fixman determinant determines the effective mass associated with the RC chosen. For the dynamic distance, the rate, k, for the escape from a stable state (around a minimum) between D_0 and $D^{\ddagger}$ over a barrier (at a maximum of free energy) at $D^{\ddagger}$ is

$$k = \kappa k^{\mathrm{TST}} = \kappa \sqrt{\frac{k_{\mathrm{B}}T}{2\pi\mu^*}}\, P(D^{\ddagger}) / \int_{D_0}^{D^{\ddagger}} P(D)\mathrm{d}D \tag{9}$$

where k^{TST} is the rate given by transition state theory (TST), κ is the transmission coefficient, and $P(D)$ is the one-dimensional pdf related to the free energy, $A(D)$, by

$$P(D) = \mathrm{const} \cdot \exp(-A(D)/k_{\mathrm{B}}T) \tag{10}$$

Apparently the statistical advantage of a constant Fixman determinant is essentially due to mass-weighting and has wide consequences stated in eqs 8 and 9. The expression 9 for the rate depends on the definition of the RC but not on the computation of free energy profiles from the constraint force according to eq 7. Therefore, eq 9 also holds for profiles calculated by umbrella sampling[8] or umbrella integration.[10,17] Forces derived from the umbrella restraint potential, $\sigma^2(\mathbf{x}) = (\tilde{D}(\mathbf{x}) - D)^2$, induce neither rotation nor translation when employing the dynamic distance.

3. Ab Initio Molecular Dynamics Applications

3a. Double Proton-Transfer Reaction in Formic Acid Dimer. The dynamic distance constraint is readily implemented within the framework of ab initio molecular dynamics. In this section we present the application of the dynamic distance constraint using Car−Parrinello molecular dynamics to study proton-transfer events and dissociation processes. We first apply the dynamic distance constraint to the study of the well-known double proton-transfer event (DPT) in the model compound formic acid dimer shown in Figure 1. For this simple example, we discuss the technical details concerning the implementation of the constraint and show how one can extract accurate free energy profiles.

Methods. All calculations were performed using the CPMD 3.4 package.[18] The formic acid dimer was placed in a periodically repeating cell with dimensions $13.25 \times 13.25 \times 13.25$ Å^3. A fictitious mass of 400 au was ascribed to the electronic degrees of freedom within the Car−Parrinello scheme. The coupled equations of motion for atomic nuclei and molecular orbitals were solved using the velocity Verlet algorithm with a time-step of 4 au. For each nuclear configuration, the Kohn−Sham equations were solved using the BLYP exchange-correlation functional.[19,20] Core electrons were treated using norm-conserving Troullier-Martins pseudopotentials,[21] and the valence electrons were expanded in a plane wave basis up to an energy cutoff of 70 Ry in all simulations performed.

The unconstrained system was first brought to thermodynamic equilibrium at 300 K using a Nosé-Hoover thermostat.[22] For the constrained CP−MD simulations, the dynamic distance constraint comprises two distances which represent the two O−H chemical bonds. The dynamic distance was initially set at a value of 1.925 au, which was the average value of the dynamic distance in a 0.5 ps unconstrained MD simulation, and subsequently was systematically increased. The chosen increment in the step size was very small in the initial stages of the reaction during the cleavage of the O−H chemical bonds. For each dynamic distance, D_i, the system was re-equilibrated before starting a 1 ps 'production run'. The simulation length employed provides sufficiently reliable average constraint forces. Free energies were calculated by numerical integration from the cumulative average of the constraint forces using eq 8. The entropy contribution to the free energy profile was calculated from the eigenvalues of

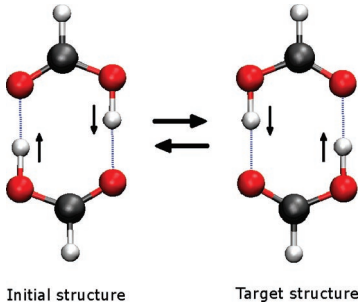

Figure 1. Illustration of initial and target structures for the double proton exchange in the formic acid dimer.

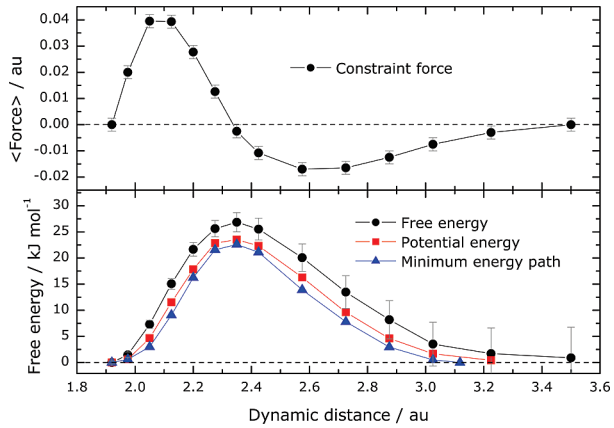

Figure 2. Average constraint force (top) and energy profiles (bottom) along the reaction coordinate for formic acid dimer. The free energy curve (black) was obtained by integration of the force curve (top). The average finite temperature potential energy is shown in red, the minimum energy profile in blue. Energies are given relative to their starting values.

the mass-weighted covariance matrix for each constrained CP−MD simulation.[23,24]

Results. Figure 2 shows the average constraint force and free energy profile along the reaction coordinate for formic acid dimer. We observe the well-known concerted double proton-transfer event. The average constraint force starts at zero and rises to a maximum at 2.1 au. This positive constraint force arises from the fact that the constraint is driving the system away from the stable configuration, as it 'pushes' the protons across the hydrogen bonds, causing the O−H chemical bonds to break. The average constraint force then falls to zero at 2.35 au, which defines the transition state for the reaction. At larger dynamic distances the average constraint force becomes negative, which represents the constraint acting to 'hold back' the protons as they try to complete the DPT reaction. The concerted nature of the reaction is represented by the single energy barrier in the free energy profile, which reaches a maximum of 26.7 kJ/mol at the transition state (a dynamic distance of 2.35 au). Quasi harmonic frequencies ω_i were calculated from the mass-weighted covariance matrix and inserted into the entropy formula[23,24]

$$S_{ho} = \sum_{i=1}^{3N-6} k_B(\hbar\omega_i/k_BT)(\exp(\hbar\omega_i/k_BT) - 1) - \ln(1 - \exp(-\hbar\omega_i/k_BT)) \quad (11)$$

After subtracting the entropic contribution, the resulting enthalpy profile for the reaction is very similar to the minimum energy path (MEP), as shown in Figure 2. The small differences arise from the fact that the MD simulation at 300 K is probing more configurational space than the MEP. This is most noticeable for larger dynamic distances, where the constraint is now acting on the newly formed hydrogen bonds. The free energy and enthalpy profiles presented in Figure 2 clearly demonstrate how well the dynamic distance constraint controls the DPT reaction, in comparison to previous constraints.[25]

To obtain correct free energy profiles, it is necessary to verify that the simulation length for each constrained CP−MD simulation provides a sufficiently reliable average constraint force. This is best achieved by monitoring the cumulative average constraint force.

Figure 3 shows the variation in the constraint force and cumulative average constraint force over a 1 ps CP−MD simulation at a dynamic distance constraint of 2.125 au. While the constraint force varies quite significantly across the trajectory from 0.01 to 0.065 au, the cumulative average constraint force converges within 1 ps to an average value of 0.039 au. The magnitude of the fluctuations in the cumulative average constraint force over the last 400 fs of the trajectory provides an estimate of the error used when calculating the free energy profile as shown in Figure 2. The rate of convergence of the cumulative average constraint force is system specific and must therefore be determined for the particular system of interest. In the case of formic acid dimer, as shown in Figure 3 a 1 ps trajectory is sufficiently long to liberate accurate converged average constraint forces.

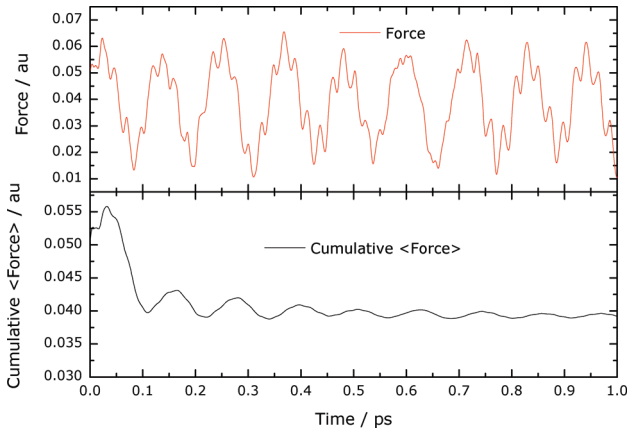

Figure 3. Trajectory of the constraint force (top) and cumulative average (bottom) as obtained in a 1 ps run at a constant reaction coordinate, $D = 2.125$ au for FAD.

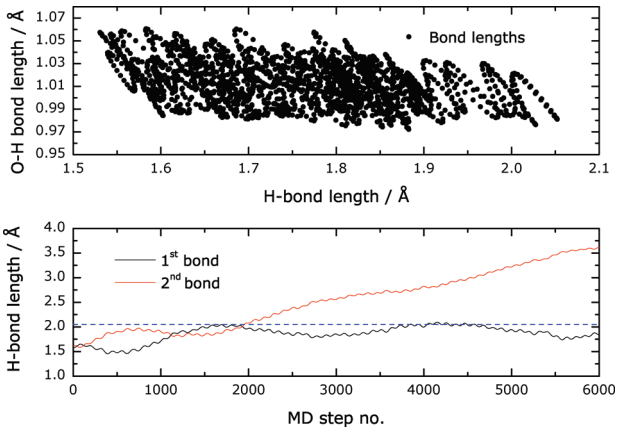

Figure 4. The lengths of a O−H chemical bond and an H-bond over a 0.5 ps unconstrained CP−MD run (top). The H-bond lengths in a constrained CP−MD run (bottom) demonstrate that one bond breaks after 2000 steps.

3b. Dissociation of Formic Acid Dimer. We extended our study of formic acid dimer to investigate the dissociation of the dimer using the dynamic distance constraint. The principal aim of this analysis was to determine whether the dissociation process is concerted or stepwise. In order to look at this event, we implemented the constraint in a slightly different way: Instead of performing a series of constrained CP−MD simulations, each at a specific constant dynamic distance, we started a simulation at a dynamic distance value of 3.025 au and systematically increased the dynamic distance by 0.0004 au per step across a single trajectory. In this case, the dynamic distance constraint comprised the two hydrogen bonds. The systematic increase in the constraint therefore drives dissociation of the dimer. The growth rate of 0.0004 au per step is small enough that the kinetic energy of the electrons remains unperturbed during the simulation. In the lower panel of Figure 4, we show the observed change in the two hydrogen bond lengths across the trajectory. Initially, both hydrogen bond lengths have a value of approximately 1.6 Å. On increasing the dynamic distance constraint over the first 2000 steps, both hydrogen bond

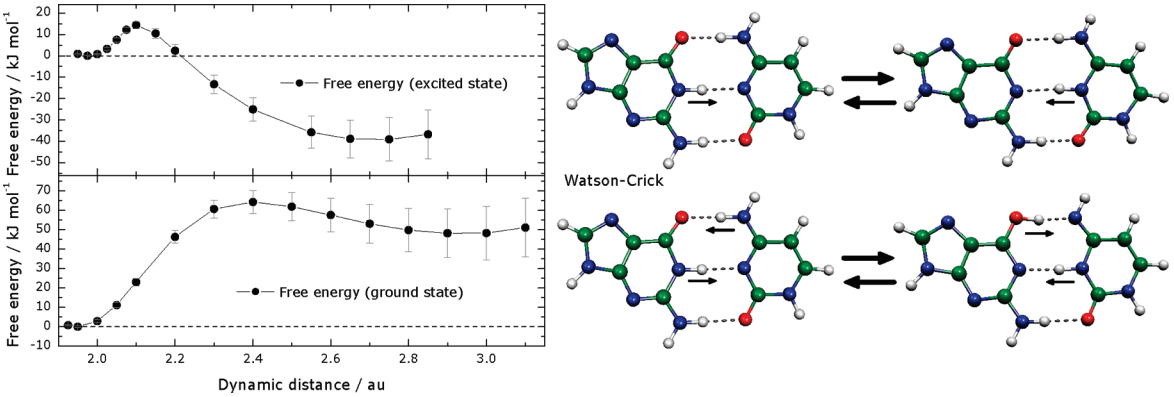

Figure 5. Free energy profiles for the ground (lower panel) and excited (upper panel) state proton-transfer reactions in the G−C base pair. The reactant and product states for the two reactions are shown graphically on the right.

lengths increase to approximately 2.0 Å. After this point, one of the hydrogen bonds breaks, and the associated hydrogen bond length increases to over 3.5 Å. The other hydrogen bond fluctuates, and its associated hydrogen bond length varies between 1.8 and 2.05 Å. The upper panel of Figure 4 shows the variation in the OH···O hydrogen bond length and the O−H chemical bond length over a short 0.5 ps unconstrained CP−MD simulation at 300 K. The hydrogen bond length varies up to 2.05 Å. These results clearly demonstrate that the dissociation process occurs in a stepwise fashion, a result that is consistent with that observed in CP−MD simulations at higher temperatures (results not shown). This simple example illustrates the versatility of the dynamic distance constraint, which can be used to probe both the DPT event and the dimer dissociation process. It also underlines the fact that the constraint does not favor or bias the reaction mechanism, be it concerted (DPT) or stepwise (dissociation).

3c. Guanine−Cytosine DNA Base Pair: Ground-State Proton Transfer and Excited-State Coupled Proton−Electron Transfer. A further interesting aspect of the dynamic distance reaction coordinate concerns its remarkable predictive properties: Unlike atom-specific, local constraints, such as simple distance and angle constraints, which are chosen in advance in order to drive a system to a known predefined product state, the dynamic distance is a flexible collective reaction coordinate that comprises multiple internal degrees of freedom. As such, when implemented appropriately, the dynamic distance constraint automatizes the search for the lowest energy reaction pathway(s) without any specific a priori knowledge of the product state. A good example of the predictive properties of the dynamic distance constraint can be found in a recent study of irradiation-induced damage mechanisms in the guanine−cytosine (G−C) base pair,[26] and the reader is referred to this reference for computational details. For the purposes of this paper, we merely summarize the general result.

The G−C base pair possesses three interbase hydrogen bonds. Starting in the Watson−Crick geometry, there exist a large number of possible single or multiple proton-transfer reactions that can bring the system to a variety of different hydrogen-bonded tautomeric states. The dynamic distance constraint for this system was constructed using the three N−H chemical bonds involved in interbase hydrogen bonding. Similar to the method described in section 3a, we performed a series of constrained Car−Parrinello MD simulations in order to identify the lowest energy PT reactions in both the ground and the excited electronic state of the G−C base pair. The results are summarized in Figure 5. In the ground state, we observe a double proton-transfer event over a large free energy barrier (64.3 kJ/mol) leading to a meta-stable product state. In contrast to this, for the singlet excited state, we observe a single proton-transfer event over a very small free energy barrier (14.3 kJ/mol) leading to an energetically favorable charge-transfer product state. It is important to recognize that the same constraint, implemented in exactly the same manner produces two completely different reaction pathways, because the underlying potential energy surface which is calculated 'on-the-fly' during the constrained CP−MD simulations is different in the ground and excited states. For both reactions, the dynamic distance constraint finds the energetically most favorable reaction pathway, as confirmed by static post Hartree−Fock calculations.[27−29]

In these simulations, the constraint was defined using the three N−H chemical bonds for the interatomic distances. However, if the constraint is formulated with the three interbase hydrogen bonds, the average constraint force and free energy profiles look rather different: In the initial stages of the reaction the average constraint force profile rises gradually to a maximum as the system is driven toward the transition state. At the transition state, the average constraint force decreases directly to zero, and the new chemical bond(s) are formed immediately. The constraint loses control of the reaction as the flexible hydrogen bonds readily alter their geometry slightly to allow the chemical bonds to form directly while still fulfilling the conditions of the constraint. Similar behavior was also observed in the case of targeted molecular dynamics.[25] The resulting free energy barrier for the reaction is still accurately obtained by integration of the average constraint force as the constraint controls the reaction up to the transition state. Nevertheless, this simple example demonstrates that one must choose the specific interatomic distances carefully for the particular system in question in order to define a constraint that can control the reaction along the entire pathway.

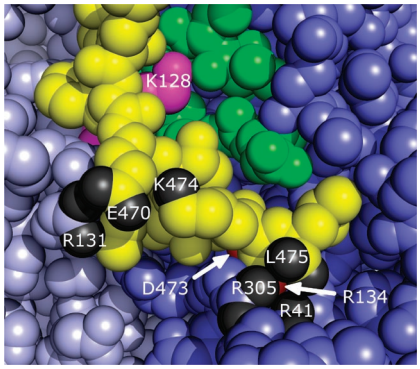

Figure 6. Three structural elements are sealing the active site of rubisco: C-terminal strand (yellow) with terminus L475, K128 (magenta) and loop 6 (green). Four ionic contacts are stabilizing the closed, active conformation of the C-terminal tail: The contacts E470-R131, L475-R41, and L475-R305 which are exposed to the solvent, and the buried bridge D473-R134 which is conserved in all rubisco homologues. The intrastrand salt bridge E470-K474 is not considered here.

4. Classical MD: Opening the Binding Pocket of Rubisco

The binding niche of rubisco is sealed by three structural elements of which the large subunit's C-terminal strand is the outermost. In the closed conformation this element stabilizes the catalytically active state of the protein (see Figure 6).

The composition of the C-terminal tail influences the substrate specificity of rubisco, which catalyzes the fixation of carbon dioxide and molecular oxygen. The time window hypothesis[30] ascribes this to the dynamics of the tail which may transiently lift off, thus interrupting catalysis. The C-terminal strand is attached to the underlying protein corpus by several ionic bridges,[31] whose specific number varies from homologue to homologue. In order to analyze the structural dynamics of the C-terminal tail, these contacts were cleaved using the dynamic distance constraint in the framework of classical MD simulation. Free energy profiles were calculated, and the contribution of each salt bridge to the stability of the enzyme's closed, active conformational state was estimated.

Methods. The four salt bridges are all of the following type: $-C-O_2^- \cdots H_2^+-N-C^{\zeta}-$ (arginine). In order to allow dynamical exchange among the carboxyl oxygens or guanine hydrogens and to avoid interference with the bond length constraints imposed on the NH_2^+ moiety, the dynamic distance of eq 1 is defined here as the rms carbon−carbon distance for each of the four salt bridges. For the classical MD simulations, the GROMACS simulation package[32] was employed with explicit SPC water and the force-field 43A1. The simulations were performed on the rubisco structure (1RBL) from the reference organism Synechococcus sp. PCC6301. The two large subunits forming a functional L2 protomer including two binding niches were treated explicitly, while the missing adjacent subunits were emulated by restraining harmonic potentials ($k_{xyz} = 250$ kJ mol^{-1} nm^{-2}) on heavy atoms involved in polar contacts to the neighboring subunits. The binding niche was modeled, and the substrate RuBP and the carbamylated K201 were parametrized as described previously.[31] Crystal waters were retained, and the protein was inserted into a simulation cell flooded with bulk water. The resulting system contained a total of 78 854 atoms. After an initial energy minimization, the system was heated and brought to equilibration during a short 200 ps MD simulation. A suitable electrostatic cutoff and reaction field were used ($r_{cp} = 0.8$ nm, $r_{cl} = 1.4$ nm, $r_{crf} = 1.4$ nm, $\epsilon_{rf} = 54$). Bonds involving hydrogen atoms were constrained using LINCS, and a time-step of 1 fs was employed. All runs were performed at a temperature of 298.15 K and a pressure of 1 bar, both regulated by a Berendsen thermostat and barostat, respectively. After equilibration, a number of conformations were extracted every 1000 ps from an unconstrained simulation to be used as starting conformations for the constrained runs.

Pathways and Free Energy. A pathway can be generated by modulating the RC from an initial to a final value during a so-called slow-growth simulation run. However, the computation of the free energy profile (instead of the work profile as immediately yielded by the slow growth run) requires converged mean constraint forces obtained at discrete points along the reaction pathway during a series of relaxation runs.[2,33] Due to the their rugged and heavily structured energy landscape,[34] difficulty arises when calculating free energy profiles for protein systems: In each relaxation run, the system may evade into different pathways thus rendering structurally discontinuous trajectories and useless free energy profiles.[35] To prevent such incidents a novel variant of the equidistant relaxation protocol was implemented in the rubisco simulations. The stop-and-go-like (SNG) approach integrates the slow-growth and the relaxation phases into a single simulation; the system is equilibrated for a certain period of time in a scleronomic "stop" phase ($D = \text{const}$) in which the average constraint force is calculated. The transition is then driven further in a rheonomic "go" phase ($dD/dt > 0$). This procedure is repeated for a certain number of equidistant points on the RC. The average constraint forces of the stop phases are then integrated to obtain the free energy profile of the particular reaction path.

Optimization. Before performing the production runs, it is necessary to optimize several parameters for the SNG approach, paying consideration to the available computational resources. These parameters include the number of equidistant relaxation points, p, on the RC, the equilibration time allowed at each of these points (relaxation phase: t_{stop}), and the fraction of this time period used to calculate the average constraint force (measuring phase: $t_{av} \leq t_{stop}$). The total resulting simulation time is $t_{stop} + (p - 1)(t_{go} + t_{stop})$. The RC was increased in steps of 1 nm from a starting value 0.45 nm, and the final constraint value was inferred from unconstrained long-term simulations (5−10 ns) in which the four salt bridges were observed to rupture spontaneously.

It was found that a relatively long relaxation period was required in order to obtain well-converged average constraint forces; however, the large forces observed in the initial stages of the relaxation can safely be discarded. Optimum convergence of the constraint force was achieved when considering

Table 1. Results of the Parameter Optimization in Three *ceteris paribus* Parameter Groups[a]

p	t_{tot} (fs)	t_{go} (fs)	t_{stop} (fs)	t_{av} (fs)	$\langle df^c_{cum}/dt \rangle$ (kJ mol^{-1} nm^{-1} ps^{-1})	$\langle \sigma(\langle f^c \rangle) \rangle$ (kJ mol^{-1} nm^{-1})
20	100019	1	5000	2500	**−20.48**	**9.44**
200	100199	1	500	250	−181.67	133.26
2000	101999	1	50	25	1029.91	253.88
20000	99999	1	4	2	−5086.69	85.49
20	100019	1	50000	25000	**−1.04**	**9.75**
200	100199	1	5000	2500	−2.84	32.09
2000	101999	1	500	250	−30.28	98.53
20000	99999	1	49	25	−34.58	255.08
20	10140	260	260	130	−75.31	138.29
20	101400	2600	2600	1300	−16.99	41.38
20	1014000	26000	26000	13000	**−0.42**	**14.37**

[a] Maximal distribution of simulation time on the scleronomic phases (∼0.1 ns and ∼1 ns, respectively, number of points variable), and equal distribution to the scleronomic and rheonomic phases (20 discrete points, simulation time variable). Optimal performance values are printed in boldface. 1 ns runs with an equal distribution of simulation time for stop and go phases yields the best results (last line).

only the constraint forces obtained during the latter half of the scleronomic phases ($t_{av} = 0.5t_{stop}$).

The following performance parameters were considered for further optimization: $\langle df^c_{cum}/dt \rangle$: The slope of the cumulative average of the constraint force f^c averaged during the relaxation period of length t_{av} as a criterion for the convergence of the constraint force. $\langle \sigma(\langle f^c \rangle) \rangle$: The standard error of the average constraint force averaged during the relaxation period as a criterion for the quality of the mean constraint force.

Table 1 shows the average values of these quantities for each run. The optimal parameter set was derived by the *ceteris paribus* principle, i.e., each independent variable was changed while keeping all others fixed. As optimal convergence of the constraint force at each discrete point along the reaction coordinate is the primary objective, the first approach taken was to maximize that portion of the simulation time spent on the scleronomic phases (relaxation). As a consequence, in these runs only 1 fs was spent driving the system in each rheonomic phase. The number of discrete points was varied as well as the length of each scleronomic phase accordingly. This was done in two *ceteris paribus* groups for total simulation times of ∼0.1 ns and ∼1.0 ns. In the third group, the available simulation time was equally distributed between the stop and go phases which yielded considerably better results. The total simulation time was set to ∼0.01 ns, ∼0.1 ns, and ∼1 ns. For this group, the cumulative constraint force convergence is shown in Figure 7.

The results summarized in Table 1 show that in the present case, an equal distribution of simulation time between the rheonomic and scleronomic phases provides the best results in terms of the criteria defined above. While it appears sufficient to calculate the average constraint force at only a small number of points along the reaction coordinate, slow reaction coordinate modulation and long time scale equilibration phases are essential for good results. As a rule of thumb,

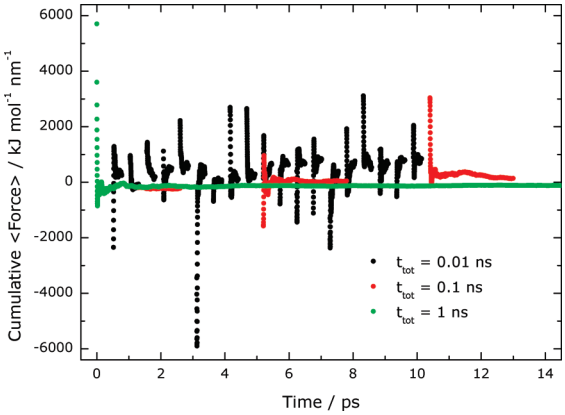

Figure 7. Convergence of the cumulative constraint force at various total simulation times (0.01, 0.1, and 1 ns). Only a sufficiently long simulation time allows convergence of the constraint force during the relaxation phase.

it is usually sufficient to check convergence of the constraint force for the starting structure only to get an impression of the necessary duration of the scleronomic phase.

In consideration of the general problem of accuracy in free energy calculations,[33] the convergence of mean forces was checked in all cases. Extensive studies on the protein showed that 1 ns runs produce reliable profiles without discontinuities with the appropriate time allocations for driving, equilibrating, and averaging. This is important because the back reaction—often recommended as a test—cannot be simulated to the same degree of accuracy for complex activated processes in large systems.

Results. Four 1 ns production runs on rubisco were performed with the optimized parameters described above. Although the free energy profiles of all simulations were consistent in terms of convergence and error of the constraint force, the specific forms of the free energy profiles were somewhat different (data not shown). We conclude that the system takes different pathways depending on the specific initial geometry. Nevertheless the sequence of rupturing events along the reaction coordinate and the relative contribution of each salt bridge to the stability of the enzyme's active conformational state were the same in all runs. A representative example is shown in Figure 8. The carboxy-terminal contacts of L475 are seen to open first; these are easily solvated and only play a minor role in stabilizing the closed conformational state at room temperature. The highly conserved bridge between D473 and R134 is cleaved last and thus is the main player, while the contact E470-R131 plays a modulating role, which makes it sensitive to specificity enhancing mutations. These results are in good agreement with previously published data based on a combined bioinformatic tools and TMD simulation approach.[31]

For the reaction path depicted in Figure 8 we calculated a free energy barrier of $\Delta A = 25.37 \pm 3.84$ kJ/mol. The error was estimated as described previously.[2] Figure 9 shows the underlying profile of the mean constraint force. Similar values of ΔA were obtained from other production run pathways, though this does not exclude the possible existence of further reaction pathways with lower activation barriers.

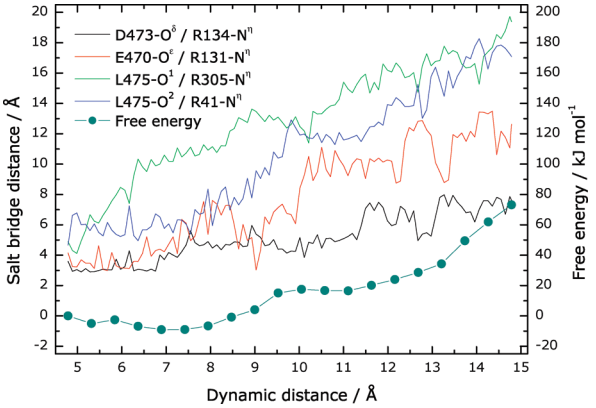

Figure 8. Free energy profile and associated distances of the four salt bridges as a function of the dynamic distance constraint. The conserved salt bridge D473/R134 is the last to open. The carboxyl-terminal contacts of L475 have a negligible effect, while the E470/R131 bridge plays a modulating role in the stability of the closed C-terminal strand.

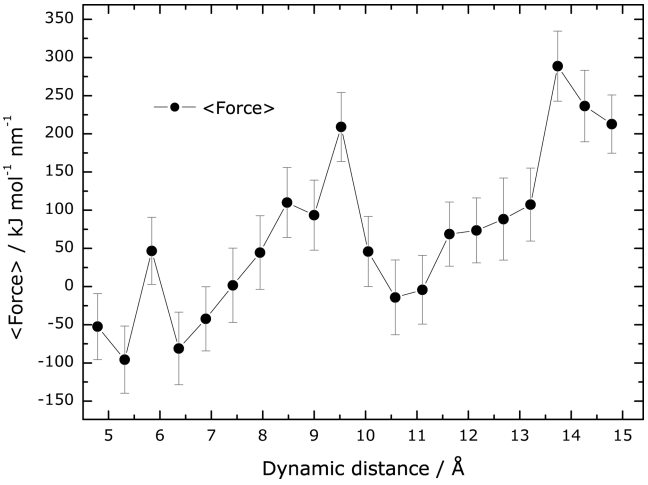

Figure 9. Profile of the average constraint force with error bars across the reaction coordinate.

To a single order of magnitude, the energy barrier is in good agreement with the nanosecond time scale of the C-terminal strand's opening in the long-term simulations already mentioned, though experimental rates are not available yet.

The path taken by the C-terminal strand during the enforced transition was confirmed by comparing an intermediate structure to results obtained from the long-term free MD simulations mentioned above. Figure 10 compares snapshots obtained from the constrained and free MD simulations at the moment when the crucial salt bridge D473/R134 is ruptured. All elements of the fluctuating system are seen to adopt comparable configurations.

5. Conclusions

In this paper we have introduced a novel versatile reaction coordinate, the 'dynamic distance', which has been specifically designed for the study of reactions and activated processes involving the cleavage and/or formation of a set of bonds or contacts. The flexible reaction coordinate, formulated as a mass-weighted mean of selected distances does not bias or favor the sequence or mechanism of events

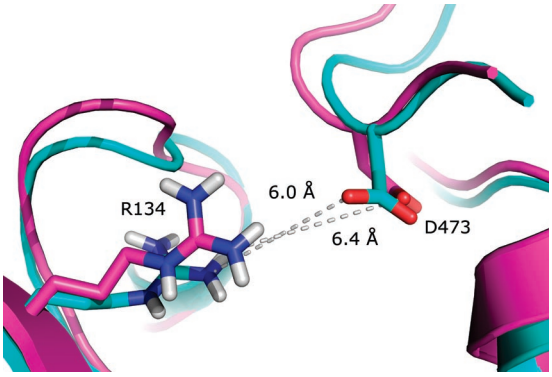

Figure 10. Comparison of structures with open C-terminal tail obtained from constrained (1 ns; dark) and free (5 ns; light) MD simulations. Both conformations are very similar, demonstrating the use of the constraint to induce accurate transitions in short simulation times.

on the resulting reaction pathway. Due to the presence of the mass-weighting term, the free energy profile for the process is readily obtained by integration over the mean constraint force without any correction, and the rheonomic constraint driving this RC represents a minimal perturbation in that it causes no net momentum or torque.

Using several examples in the framework of both ab initio and classical molecular dynamics, we have demonstrated the versatility of the dynamic distance constraint, paying particular attention to both the implementation and optimization of the RC in order to obtain accurate reaction pathways and free energy barriers. Our study of the DPT event in the prototypical model system formic acid dimer reproduces the well-known concerted reaction mechanism, and the size of the activation barrier is in full agreement with previous studies. However, the free energy profiles clearly demonstrate an increased level of control compared to alternative constraints.[25] The predictive properties of the dynamic distance have been highlighted using a recent application of the constraint to automatize the search for the lowest energy proton-transfer events in the G−C base pair in both the ground and excited state.[26] Using classical MD simulation, we considered the opening of a binding pocket in a large protein−water system ($\sim 10^5$ atoms). Proteins are known for their complex glasslike energy landscape[34] which opens a manifold of pathways for such complex processes. The constrained simulations produced both unproductive and productive pathways. The latter exhibit activation barriers and intermediate structures which compare well with available long time-scale free MD simulations. A common feature of all the simulations is the unique sequence of events when the crucial salt bridges are cleaved, which was in the focus of the present study.

Except for simple cases, there is no way to decide whether the best path was already detected by any method whatsoever. A second caveat concerns the directionality of coordinate driving methods which tend to produce different pathways during decreasing and increasing the reaction coordinate. Therefore, repeated simulations with different starting conditions and directions are suggested wherever possible for determining realistic pathways.

Among the broad range of potential applications, the dynamic distance reaction coordinate is particularly suitable for the study of important functional processes in biological systems involving association and dissociation events and proton-transfer reactions.

Acknowledgment. C.B. and J.S. acknowledge funding by the VolkswagenStiftung. One of the authors (N.L.D.) acknowledges funding by the German Science Foundation DFG with FOR 618. Another author (P.R.L.M.) acknowledges funding by the CEA. The authors are grateful to BOVILAB@RUB, EMBL-Heidelberg, NIC Jülich, and ZAIK Köln for computer time.

References

(1) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.

(2) Swegat, W.; Schlitter, J.; Kruger, P.; Wollmer, A. *Biophys. J.* **2003**, *84*, 1493−1506.

(3) Akola, J.; Jones, R. O. *J. Phys. Chem. B* **2003**, *107*, 11774−11783.

(4) Schlitter, J.; Swegat, W.; Mulders, T. *J. Mol. Model.* **2001**, *7*, 171−177.

(5) Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M. *J. Am. Chem. Soc.* **2002**, *124*, 6594−6599.

(6) Blumberger, J.; Sprik, M. *Theor. Chem. Acc.* **2006**, *115*, 113−126.

(7) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516−17524.

(8) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

(9) Schlitter, J.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291−308.

(10) Kästner, J.; Thiel, W. *J. Chem. Phys* **2005**, *123*.

(11) Schlitter, J.; Klähn, M. *J. Chem. Phys.* **2003**, *118*, 2057−2060.

(12) Ma, J. P.; Sigler, P. B.; Xu, Z. H.; Karplus, M. *J. Mol. Biol.* **2000**, *302*, 303−313.

(13) Ensing, B.; De Vivo, M.; Liu, Z. W.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73−81.

(14) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050−3053.

(15) Schlitter, J.; Klähn, M. *Mol. Phys.* **2003**, *101*, 3439−3443.

(16) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472−477.

(17) Kästner, J.; Thiel, W. *J. Chem. Phys* **2006**, *124*.

(18) Hutter, J.; Ballone, P.; Bernasconi, M.; Focher, P.; Fois, E.; Goedecker, S.; Marx, D.; Parrinello, M.; Tuckerman, M. *MPI für Festkörperforschung and IBM Zürich Research Laboratory*; Stuttgart, 2001.

(19) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(20) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(21) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993−2006.

(22) Nosé, S. *J. Chem. Phys* **1984**, *81*, 511−519.

(23) Schlitter, J. *Chem. Phys. Lett* **1993**, *215*, 617−621.

(24) Andricioaei, I.; Karplus, M. *J. Chem. Phys* **2001**, *115*, 6289−6292.

(25) Markwick, P. R. L.; Doltsinis, N. L.; Marx, D. *J. Chem. Phys* **2005**, *122*, 054112.

(26) Markwick, P. R. L.; Doltsinis, N. L.; Schlitter, J. *J. Chem. Phys.* **2007**, *126*, 45104−45107.

(27) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.

(28) Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2763−2771.

(29) Sobolewski, A. L.; Domcke, W.; Hattig, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17903−17906.

(30) Schlitter, J.; Wildner, G. F. *Photosynth. Res.* **2000**, *65*, 7−13.

(31) Burisch, C.; Wildner, G. F.; Schlitter, J. *FEBS Lett.* **2007**, *581*, 741−748.

(32) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 3.2*; 2004. www.gromacs.org (accessed March 2004).

(33) Mark, A. E.; van Helden, S. P.; Smith, P. E.; Janssen, L. H. M.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 6293−6302.

(34) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598−1603.

(35) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645−15650.

CT700170T