The physical determinants of the DNA conformational landscape: an analysis of the potential energy surface of single-strand dinucleotides in the conformational space of duplex DNA

Karim M. Elsawy[1,2], Michael K. Hodgson[2] and Leo S. D. Caves[1,*]

[1]Department of Biology and [2]Department of Chemistry, University of York, York YO10 5YW, UK

ABSTRACT

A multivariate analysis of the backbone and sugar torsion angles of dinucleotide fragments was used to construct a 3D principal conformational subspace (PCS) of DNA duplex crystal structures. The potential energy surface (PES) within the PCS was mapped for a single-strand dinucleotide model using an empirical energy function. The low energy regions of the surface encompass known DNA forms and also identify previously unclassified conformers. The physical determinants of the conformational landscape are found to be predominantly steric interactions within the dinucleotide backbone, with medium-dependent backbone-base electrostatic interactions serving to tune the relative stability of the different local energy minima. The fidelity of the PES to duplex DNA properties is validated through a correspondence to the conformational distribution of duplex DNA crystal structures and the reproduction of observed sequence specific propensities for the formation of A-form DNA. The utility of the PES is demonstrated through its succinct and accurate description of complex conformational processes in simulations of duplex DNA. The study suggests that stereochemical considerations of the nucleic acid backbone play a role in determining conformational preferences of DNA which is analogous to the role of local steric interactions in determining polypeptide secondary structure.

INTRODUCTION

The DNA double helix is the central icon of molecular biology: the combination of simplicity, beauty and profound explanatory power of a single molecular structure will likely never be surpassed. For over 50 years, the structure and conformation of DNA have been the subject of extensive studies which have fed into considerations of stability, mechanics, dynamics and specificity in the myriad processes in which DNA plays a vital role (1).

The observed structures of right-handed double helical DNA can be classified into two major families known as A-DNA and B-DNA (2). Watson and Crick's original model of DNA, based upon fibre diffraction data, confirmed by the single crystal X-ray structure of Drew *et al*. (3), is now referred to as the B-form. Later, another crystal structure with a distinctively different backbone conformation was discovered (4), which is now termed A-form DNA. For reference, the canonical structures of A and B-form duplex DNA are those derived from fibre diffraction analysis (5).

Single crystal X-ray diffraction analyses have revealed significant variation in the conformation of the phosphodiester backbone within duplex DNA. For example, within the B-DNA family, there exist the BI- and BII-forms (or states) which show differences in the two torsions ε (C4′–C3′–O3′–P) and ζ (C3′–O3′–P–O5′) (5). Within A-DNA, the A-form and the Crankshaft-forms differ in the torsions α (O3′–P–O5′–C5′) and γ (O5′–C5′–C4′–C3′) (6,7).

The variety of DNA conformation has been studied in statistical analyses of the backbone torsion angle distributions of the observed crystal structures. For B-DNA, significant correlations have been observed in dodecamer structures between $\chi-\delta$, $\chi-\zeta$, $\delta-\zeta$, $\varepsilon-\zeta$, $\varepsilon-\beta$ and $\zeta-\beta$ pairs of torsion angles (8). In A-DNA, correlations have been found between $\chi-\delta$, $\chi-\varepsilon$, $\chi-\alpha$, $\varepsilon-\alpha$, $\zeta-\alpha$, $\zeta-\gamma$, $\alpha-\beta$ and $\alpha-\gamma$ torsion angles pairs in tetramer structures (9). In both studies the number of data points used was limited and the focus was restricted to bivariate relationships. In an explicitly multivariate approach, a principal component analysis (PCA) of nine backbone torsion angles of dinucleotide monophosphates (DMs) fragments from duplex DNA crystal structures resulted in a 3D conformational

space which exhibited a separation of an A-DNA class, two different B-DNA classes and one so-called Crankshaft class (7). This representation was considered by the authors as a multidimensional analogue to the Ramachandran plot used in protein analysis (10). Separation of the Z-DNA structures was observed in a subsequent study (11) where the conformational subspace accommodated nine clusters corresponding to the A, B and Z-DNA substates.

Moving beyond statistical considerations, there have been several attempts to explore the conformation and energetics of DNA, using a variety of both backbone torsion angles and helicoidal parameters as effective degrees of freedom (12–15). For example, a study of the energetics of base stacking (16) replaced the backbone with methyl groups at the C1′ atoms. Using this model, the values of base-step parameters were predicted accurately provided that the values of slide, shift and propeller were assigned to observed values. In a subsequent study (17), a C1′–C1′ virtual bond model was combined with two helical degrees of freedom, slide and shift, to compute the potential energy surface (PES) of an isolated dinucleotide step. In this model, the position of the low energy regions agreed well with the geometry of observed crystal structures for only some of the base steps. For other base steps, the lack of agreement of energy landscapes with the observations was ascribed to the neglect of the effects of conformational coupling with neighbouring steps.

Therefore, the backbone in many of DNA conformational analysis studies is considered as a passive element that delineates the boundaries of the dinucleotide step conformation (18) such that it acts as no more than a constraint on the range of the conformational space accessible to the bases (19). However, for a given set of base pairs the backbone geometry is not completely defined since there are many ways in which a backbone can link a given arrangement of sequential bases (19). The sugar–phosphate backbone torsion ranges distinguish the different subclasses of DNA structures; (A-form, Crankshaft A-form) and (BI, BII) (5,6,20). Deviation from these canonical ranges was also observed in protein–DNA complexes (21). The growing evidence of the biological significance of these states (22,23) with regard to DNA packaging (24), transcription (25–27), spore UV resistance (28) and protein recognition (22,23,29–31) reaffirms the importance of the role of the backbone conformational variability. This variability, to the best of our knowledge, is not explicitly taken into account in any of the studies concerned with the preferential conformations of different dinucleotide steps.

In this study, we combine both statistical and physical methods to perform a conformational analysis of DNA. At the core of our results is the computation of the PES of a DM fragment within the conformational space spanned by duplex DNA crystal structures. To achieve this, we needed an appropriate and tractable set of degrees of freedom to define the conformational space. Extending the statistical work of Becker and Buydens (7) in a similar way to Sims and Kim (11) we generated a descriptive set of collective coordinates for a dinucleotide fragment via a PCA of backbone and sugar torsion angle parameters from a sample of duplex crystal structures. These collective coordinates (which are linear combinations of proper torsion angles) span the principal conformational subspace (PCS) of the DM. In order to investigate the physical determinants of the statistical distributions we then mapped

the PES within the PCS for all possible base combinations of DMs using an empirical energy function. The structural, energetic and thermodynamic characteristics of the resulting PESs are then discussed and compared in terms of the location and relative stability of different local energy minima and the relative contribution of the different components of the potential energy. Validation of the fidelity of the PES is presented in two ways: (i) structurally—by considering the relationship of the features of the PES with experimentally observed crystal structure distributions and (ii) thermodynamically—by computing estimates of the relative conformational propensities of base-specific DM in comparison with observed sequence-specific trends. Examples of the utility of the PES for the representation of complex conformational processes in DNA are provided through brief and illustrative analyses of kinematic and dynamic simulations of duplex DNA.

MATERIALS AND METHODS

Multivariate analysis

Conformational descriptors. The conformational space of a single-strand DM fragment was described by 19 variables made up of 9 backbone torsions and 10 sugar torsions (5 for each sugar) as depicted in Figure 1. We explicitly consider all of the individual sugar torsion angles, rather than use the conventional sugar pucker angle, P (32), for simplicity in the subsequent application of restraints in subsequent energy calculations (described below).

Data acquisition. Structural data were obtained from the Nucleic Acid Database (33) (http://ndbserver.rutgers.edu). Crystal structures of duplex A-DNA and B-DNA were selected which met the following criteria: high resolution (d_{min} < 2.5 Å); containing no base-mismatches or chemical modifications; and those that are not in complex with drugs or proteins. Z-DNA was not included due to the paucity of structural data meeting the selection criteria. As in other studies (34), we regard different crystalline environments for the same sequence as distinct observations. From the structures meeting these criteria (Supplementary Table 1), a data matrix was constructed with 1178 observations (single-strand DM fragments) for the 19 variables (torsion angles). Terminal dinucleotides were not included in this data matrix in order to minimize end effects. A single observation in the data matrix is thus represented by a vector

$$\varphi_i = (\chi_1, \delta_1, \varepsilon_1, \zeta_1, \alpha_2, \beta_2, \gamma_2, \delta_2, \chi_2, \nu_{01}, \ldots, \nu_{41}, \nu_{02}, \ldots, \nu_{42})_i.$$

Principal component analysis. To avoid problems in the calculation of variance for angular data (35) the data for individual torsions were adjusted in an iterative fashion such that differences from their mean values lie in the range $-180°$ to $+180°$. PCA was carried out via an eigenanalysis of the sample covariance matrix (36). The principal components (or PCs) are mutually orthogonal collective variables that maximally describe the sample variance. The elements of the PC (eigenvectors) are the coefficients of the linear combinations of the 19 torsion angles from which they are derived. Thus, the PCs [analogous to the normal coordinates derived from harmonic vibration analysis (37)] can be considered as collective
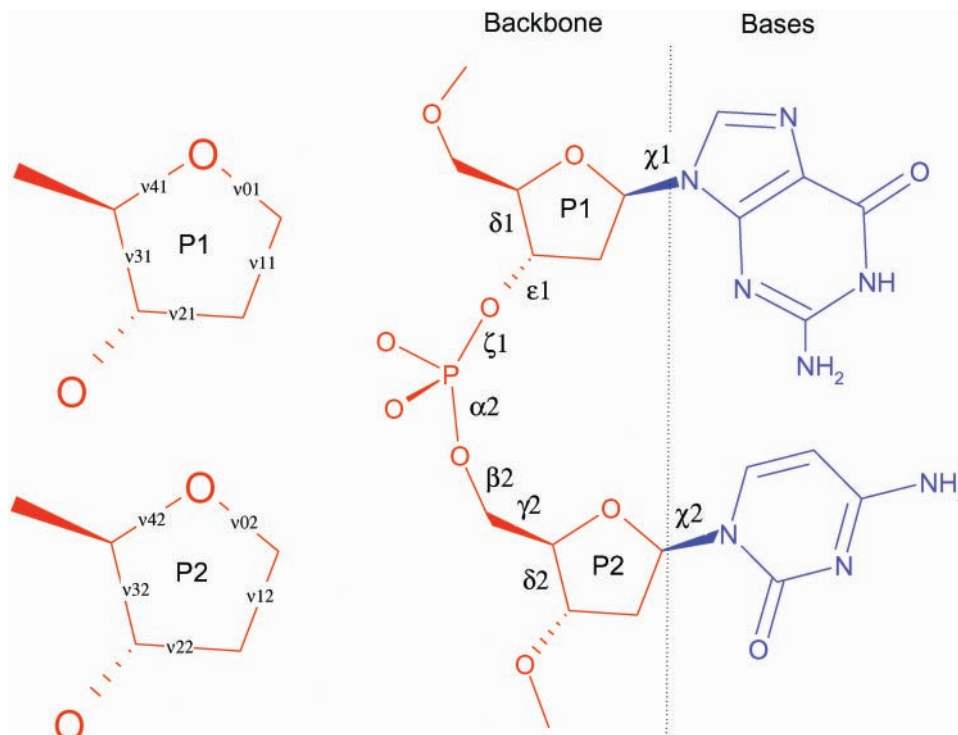
Figure 1. The torsion angle variables of the dinucleotide backbone fragment included in the PCA (see Table 1). The conformation of the sugar–phosphate backbone is conventionally defined by six torsion angles labelled α to γ, with χ describing the N-C1′ glycosyl linkage to the bases. For each sugar, there are five torsion angles v_0 to v_4 in the furanose ring [conventionally described via a pseudorotation phase angle (32)].

coordinates for describing the DNA conformational distribution via displacements from the mean torsion angles of the data matrix. The corresponding eigenvalues represent the variance of the conformational distribution along each of the collective degrees of freedom. The extent to which a given observation φ_i lies along the kth principal component $\mathbf{a}_k$ is given by the projection $z_{i,k} = \mathbf{a}_k \cdot (\varphi_i - \bar{\varphi})'$. These projections (or scores) may be used for visualizing the distribution of observations in the PCS.

Empirical energy calculations

Mapping the PES. The PES of a single-strand DM fragment within the PCS was mapped via systematic energy evaluation on a grid defined by discrete points along the first three PCs. The coordinates of a grid point $\Omega_i = (z_{i,1}, z_{i,2}, z_{i,3})$ correspond to a set of torsion angles; $\theta_i = \bar{\varphi} + \sum_{k=1}^{k=3} z_{i,k}\mathbf{a}_k$. Therefore, θ_i represents a structure reconstructed from the $\{PC1, PC2, PC3\}$ subspace whose projections along higher principal components are set to zero. Cartesian coordinates for the full DM fragment were reconstructed from these torsion angles supplemented by standard internal coordinates for other geometric parameters. The potential energy of structures was calculated using the AMBER force field (38) using the Cornell parameterization (39) implemented in the program CHARMM (40). No non-bonded interaction truncation was performed. The electrostatic terms used either a constant dielectric ($\varepsilon_r = 1$) or a distance-dependent dielectric constant ($\varepsilon_r = 4$ r_{ij}, where r_{ij} is the interatomic distance in Å) (41). The PES was mapped in 2D slices along PC3. A grid resolution of 20° in the {PC1, PC2, PC3} subspace was used over the range −300°

Table 1. Coefficients of the first three principal components derived from the distribution of torsion angles in DM fragments from duplex crystal structures (see Figure 1 for definition of torsion angles)

Torsion angle		PC1	PC2	PC3	Average structure (°)
Backbone torsions	χ_1	−0.278	0.107	−0.121	−128.130
	δ_1	−0.255	0.044	−0.151	110.845
	ε_1	0.002	−0.160	−0.499	−162.604
	ζ_1	0.260	0.094	0.632	−96.953
	α_2	−0.178	−0.710	0.125	−67.896
	β_2	0.050	0.101	0.266	172.329
	γ_2	0.163	0.609	−0.197	56.773
	δ_2	−0.248	0.068	0.153	111.143
	χ_2	−0.277	0.032	0.170	−128.360
First sugar ring torsions	v_{01}	0.133	−0.051	−0.010	−15.888
	v_{11}	−0.296	0.091	−0.048	9.990
	v_{21}	0.340	−0.096	0.082	−1.148
	v_{31}	−0.271	0.068	−0.096	−8.222
	v_{41}	0.088	−0.011	0.066	15.101
Second sugar ring torsions	v_{02}	0.126	−0.065	−0.095	−15.685
	v_{12}	−0.288	0.112	0.189	9.948
	v_{22}	0.333	−0.117	−0.208	−1.300
	v_{32}	−0.267	0.082	0.159	−7.968
	v_{42}	0.090	−0.012	−0.041	14.812
Eigenvalue/10^3 degree^2		9.430	2.367	1.671	
% Total variance		59.53	14.94	10.55	

to 300° with respect to an origin corresponding to the mean structure (Table 1). These search limits were chosen so as to encompass the range of scores spanned by the data. At each grid point, the torsion angles included in the PC basis set were restrained to the desired values using a force constant of 500 kcal mol^{-1} degree^{-2} and the system was energy

minimized using 200 steps of steepest descent followed by 2000 steps of the ABNR method (40). We note that this minimization step serves to facilitate ring closure in the sugars following Cartesian coordinate reconstruction from the internal coordinates and results in very small deviations from the desired subspace coordinate. As a consequence of the discretization, the structure and energy of any point in the PCS is considered to be the structure and energy corresponding to the closest grid point within the resolution limit of the grid.

Characterization of the features of the PES. The PES within the PCS was partitioned into a set of distinct energy valleys (42) as follows: starting from a local minimum, a set of 3D iso-energetic contours are generated at increasing discrete energy levels. An energy valley is defined by the isoenergetic surface contained within the contour level preceding the one which encapsulates another energy minimum. Thus, an energy valley is considered to extend up to the (saddle) point which serves to separate it from other minima. This process continues until no further minima are detected. The volume of each energy valley was calculated using the convex hull algorithm (43).

Thermodynamics within the PCS. The fractional population of an energy valley v can be estimated from the ratio $Q_v/\sum_{v=1}^{6} Q_v$, where Q_v is the NVT partition function given by:

$$Q_{\text{NVT}} = \sum_i g_i \exp(-E_i/k_B T) \qquad 1$$

where g_i is the degeneracy of the energy level i which is defined as the number of states with energy $E_i \pm \delta$ where δ is the uncertainty in energy. Q_v was estimated in the subspace as follows: energy levels, E_i, within each energy valley were delineated by a set of concentric isoenergetic closed surfaces in the 3D PCS corresponding to consecutive energies of 0.5 kcal mol^{-1}. The degeneracy g_i of the energy level i was estimated by the volume enclosed between the two surfaces corresponding to $E_i - 0.25$ and $E_i + 0.25$ kcal mol^{-1}. Here, we use the term 'energy level' in an *ad hoc* way since, in the classical regime, energy is continuous and not quantized. In this spirit, the (local) partition function for each valley (Equation 1) is estimated via a summation over its discrete energy levels.

RESULTS AND DISCUSSION

The principal conformational subspace

Effective dimension of the PCS. The eigenvalues resulting from diagonalization of the covariance matrix of the mean-centred data represent the amount of variance captured along each of the corresponding eigenvectors (or principal components) (44). The relative contribution of the (sorted) eigenvalues to the total variance levels off after the third or the fourth principal component and the distributions of scores along the fourth, fifth and six principal components are unimodal and therefore do not contribute significantly to additional clustering of the data (data not shown). As the first three principal components capture almost 85% of the variance (Table 1), further analysis was restricted to this subset, which effectively describes the Principal Conformational Subspace of the Dinucleotide Monophosphate (PCSDM) system.

Geometric character of the collective coordinates. In the average structure (the mean of the conformational distribution) the torsion angles are distributed symmetrically between the two nucleotides such that $\chi_1 \approx \chi_2$, $\delta_1 \approx \delta_2$ and also the sugar puckers of the two sugar rings assume virtually the same conformation. Each of the principal components is a linear combination of 19 torsion angles and may be considered as a collective coordinate representing a set of concerted displacements away from the mean structure $\bar{\varphi}$. The geometric character of each principal component is determined by the magnitude and sign of its coefficients (loadings), relating to the size and phase of its torsion angle displacements, respectively (Table 1).

The loadings of the first principal component (PC1) are evenly distributed and, apart from that of ε_1, are not negligible in magnitude (Table 1). This implies that upon displacement along PC1, ε_1 is essentially stationary while all the other torsions are changing; the sugar puckers of both rings change almost in phase (see Figure 2a).
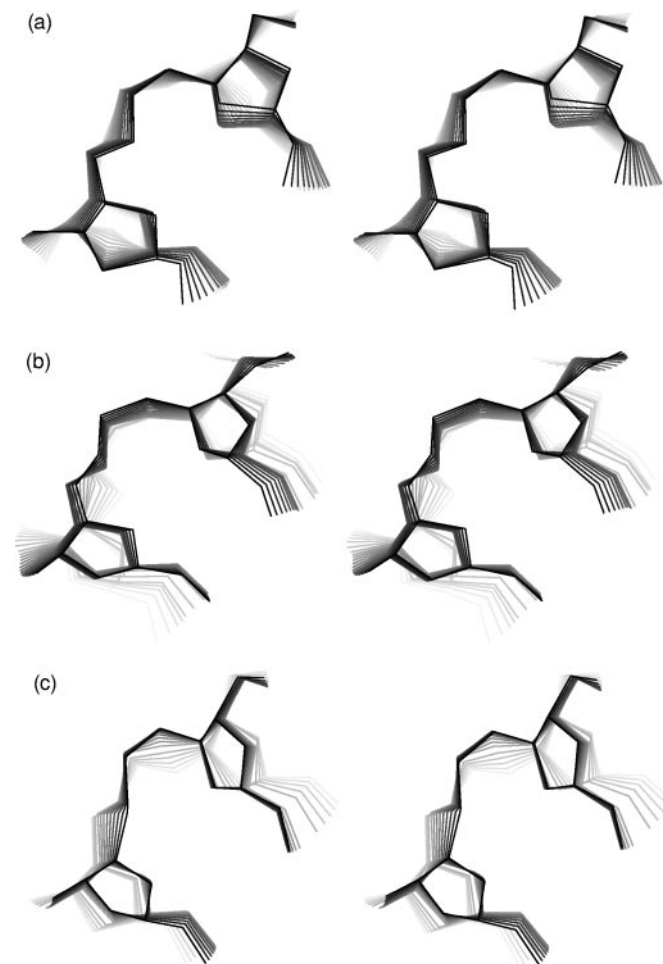

Figure 2. Stereo (wall-eyed) diagrams showing the geometric character of the principal components. The individual panels show sequential displacements along each of the first three principal components with respect to the average structure (Table 2) over the range −200° to 200° at 20° increments for (**a**) PC1 and (**b**) PC2 and from −100° to 100° at 20° increments for (**c**) PC3. The colour of the lines changes from light grey to black moving from negative to positive displacements.

For PC2, the α_2 and γ_2 torsions are dominant while their different signs reflect anti-correlated behaviour (7,9,19). The loadings of the torsions of the two sugar rings are much smaller than in PC1. Therefore, displacement along PC2 essentially involves opposing changes in α_2 and γ_2 with little change to the sugar puckers of both rings: the rest of the backbone torsions do not change significantly (see Figure 2b).

For PC3, ε_1 and ζ_1 torsions dominate; however, the loadings of the rest of the torsions are not negligible. Furthermore, the loadings of the two sugar rings no longer show the symmetrical pattern observed in the case of PC1 and PC2. Apart from ν_{41} and ν_{42}, the comparative loadings of corresponding torsions of the sugar rings are smaller in magnitude in the first ring than the second. Significantly, unlike PC1 and PC2, PC3 exhibits anti-correlated behaviour between the two sets of ring torsions. Thus, displacement along PC3 involves counterpoise changes in ε_1 and ζ_1 and in the sugar puckers of the two rings (see Figure 2c).

Given the character of the displacements, we can deduce that in reference to the mean structure: displacement along PC1 spans the changes in sugar puckers associated with A and B-form DNA; displacement along PC2 for a structure in A-form would transform it to a crankshaft-A-like structure; displacement along PC3 is more subtle, e.g. for a structure in the BI-form, it is expected to transform it to the BII-form due to the anti-correlated changes of ε_1 and ζ_1.

The PES within the PCS

The PES of a GC DM within the PCS, computed with the RDIE model, is presented as an example in Figure 3. Six low energy regions are readily apparent. The PES was partitioned into distinct energy valleys (see Materials and Methods) that were then characterized individually in terms of their structure and energy.

Structural characterization of the low energy regions of the PES. Profiles of the structural parameters of each of the energy valleys are shown in Figure 4. The sugar pucker distribution of valleys 1, 2 and 3 corresponds to the B-DNA family as their pucker ranges encompass the C2′-endo range, while those of valleys 4, 5 and 6 corresponds to the A-DNA family as their pucker ranges are encompass the C3′-endo range. Comparison of the α and γ distributions reveals that valley 4 corresponds to the A-DNA family, valley 5 corresponds to the Crank A-form (CrA) and valley 6 corresponds to A-form structures whose α and γ ranges lie between those of known A and Crank A forms (by its geometrical relation to the CrA form we name this form '−CrA').

Similarly, the $\varepsilon-\zeta$ torsion ranges of valleys 1, 2 and 3 suggest that they all correspond to the known BI-form (Figure 4). Inspection of their α and γ torsion distributions indicates that valley 2, although classified as B-form, has a Crankshaft character similar to that of valley 5. In the context of the conformational landscape of the DM (Figure 3) we name structures corresponding to valley 2 as Crank B-form (CrB).

B-form conformations with α and γ torsion ranges, such as those found in valley 2, are not observed in isolated duplex DNA structures, but have been observed in B-DNA–protein complexes (21). Interestingly, the dataset we used to construct the PCS excludes complexes and yet the PES clearly suggests the existence and location of their distinctive, non-canonical conformations. Thus, the conformational state corresponding to the CrB form of the DNA backbone, while not significantly populated in isolated structures, becomes populated in response to complex formation and therefore provides a
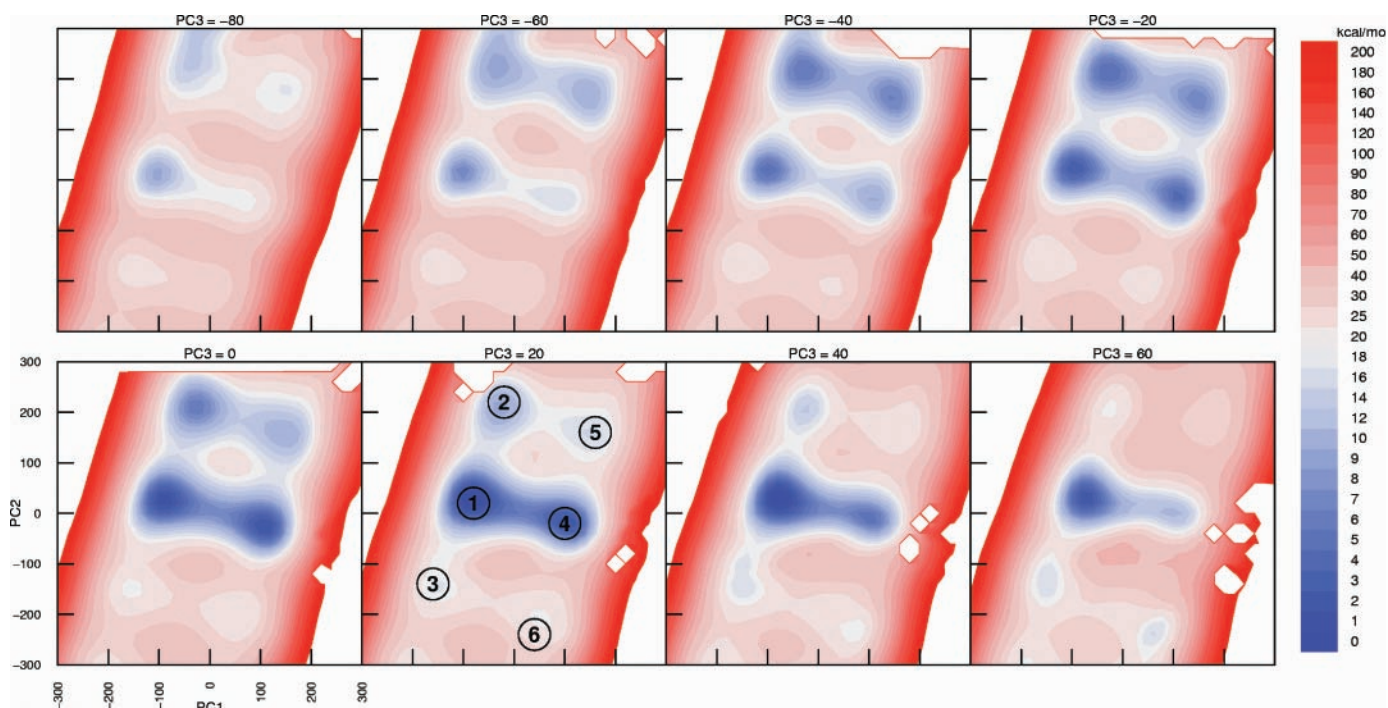

Figure 3. Total potential energy slices in the principal conformational plane spanned by PC1 and PC2 for a GC dinucleotide monophosphate fragment using a radius-dependent dielectric model (RDIE) for the electrostatic term of the potential energy function. Low energy regions are labelled from 1 to 6 on the lowest energy slice.
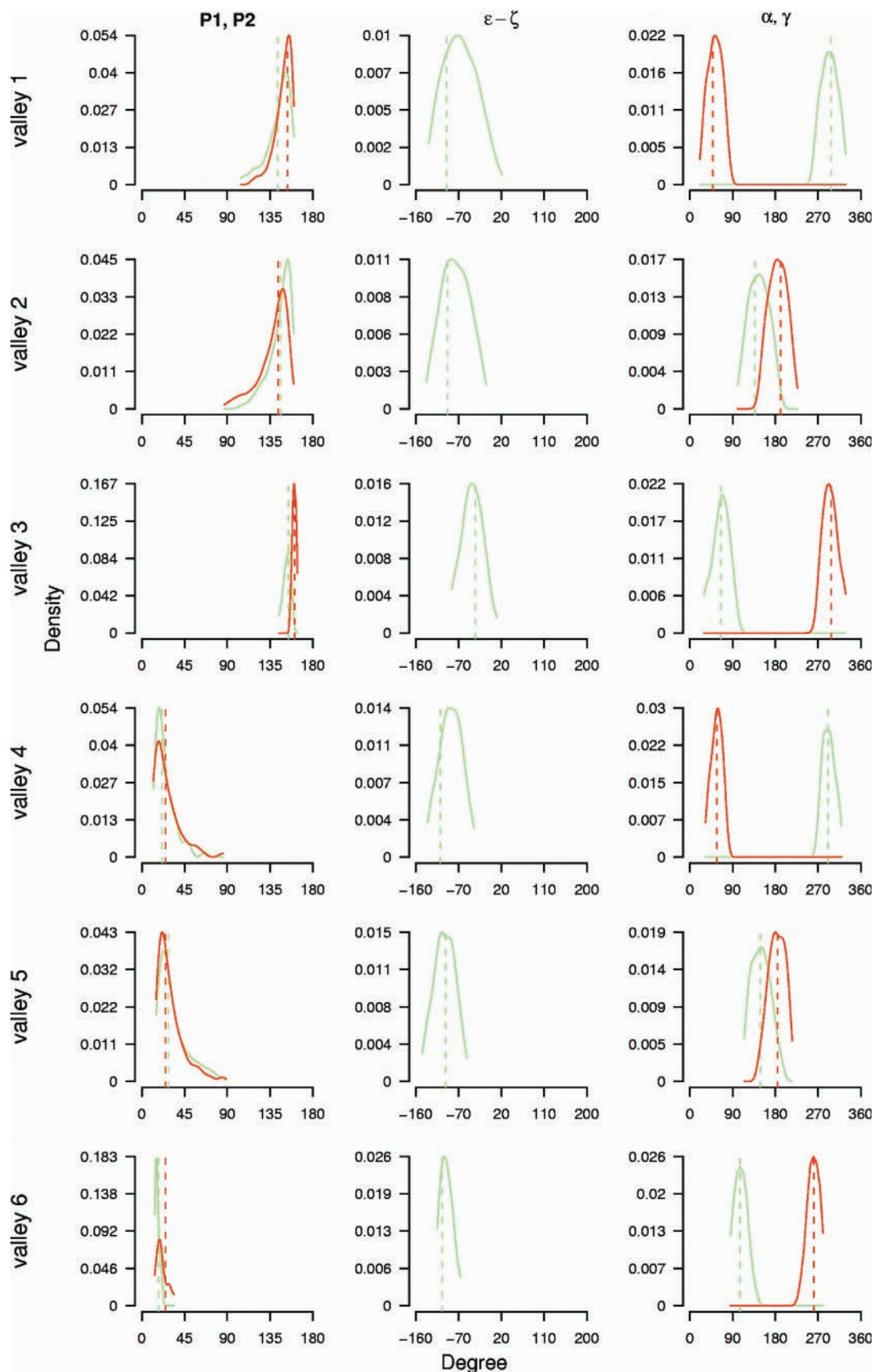
Figure 4. Distributions of the sugar psuedorotation angles P1, P2 and backbone torsion pairs; $(\varepsilon-\gamma)$, (α,γ) in each energy valley. P1 and α are shown in green while P2 and γ are in red. Positions of corresponding minima are indicated by dotted lines. Densities are based on a GC subspace PES.

clear physical interpretation of previous statistical observations (11).

The values of ε−ζ torsion differences (Figure 4) suggest that valley 1, which is almost centred at −90°, corresponds to the well-known BI-form; valley 3, which shows ε−ζ shifted by 60° from valley 1, corresponds to B-DNA structures that are intermediate between BI- and BII-forms (by its geometrical relation to the CrB form we name this form '−CrB').

Energetic determinants of the PES. To gain insight into the determinants of the general landscape of the PES, we examined the relative contributions arising from different decompositions of the empirical potential energy: The potential energy was considered in terms of steric (van der Waals + torsion angle) and electrostatic components; spatially, the interactions of the backbone and bases moieties of the DM system were considered (see Figure 1). The different contributions to the lowest energy slice of the PES in both the vacuum (CDIE) and in a distance-dependent dielectric (RDIE) models are shown in Figure 5.

Clearly, the steric interactions within the DM backbone determine the general topography of the PES supporting prior notions of their importance in determining the range of DNA conformation (45). In the RDIE model, the steric term favours the B-form over the A-form but there is little discrimination in the case of the CDIE model. This is qualitatively in agreement with the observation (46) that intramolecular interactions (van der Waals and internal energies) favour B-DNA in MD simulations of poly(G)–poly(C) DNA duplexes.

The global potential energy minimum (within the PCS) on the CDIE model PES is the A-form whereas for the RDIE model it is the BI-form (Table 2). The effect of the electrostatic component on the PES can be understood by comparing the backbone-bases and bases–bases interaction slices of the two models. The major stabilization of the A-form in the case of CDIE results from the electrostatic component of the backbone-bases interaction, which is much flatter and less discriminative in the case of RDIE. On the other hand, the bases–bases interaction in either model is rather featureless and does not serve to discriminate between the A- and B-forms.

Thermodynamics based on the PES: A- and B-form conformational equilibria. Having obtained the PES of a GC DM, an estimate was made for the equilibrium distribution of different conformational forms within the PCS at room temperature (298 K). The equilibrium thermodynamic properties depend on the relative potential energies of the local minima and the relative volumes of their associated energy valleys (47). Consideration of the potential energy difference gives an indication of the relative stability of different species but does not necessarily indicate the direction of the undergoing equilibrium process: the entropic contribution (which is estimated from the density of states of each energy valley) has an important contribution to the corresponding free-energy difference (48).

Table 2 shows the relative energetics of the minima and the relative population of the corresponding valleys. On the PES of the RDIE model (Figure 3) there exist six minima [see Table 2, using radius-dependent dielectric model (RDIE)]: the global minimum corresponds to the BI-form (within region 1 in Figure 3) while the second lowest energy minimum corresponds to A-form (region 4). The potential energy difference between the BI-form and A-form minima is 1.88 kcal/mol; however, the volume of the BI-form valley is almost twice that of the A-form. Accounting for the energy valley volume, as a measure of its entropy, the relative population of the different forms at room temperature was estimated (see Materials and Methods). The BI-form dominates the equilibrium distribution (∼96%) with only a minor contribution from the A-form [Table 2, using radius-dependent dielectric model (RDIE)]. Since the RDIE model is usually considered to be a (crude) mimic of the screening effect of an aqueous environment over short ranges (41), the relative proportions of the BI- and A-forms is in accord with the well-established preference of DNA structures to be in the BI-form in a water-rich environment (2). The populations of the other energy valleys are negligible.

It is interesting to see how changes to the electrostatic environment (i.e. the dielectric model) affect the PES and the corresponding conformational distributions. In the CDIE model [Table 2, in vacuum (CDIE)], the global energy minimum corresponds to the A-form and the next lowest minimum corresponds to the BI-form. In this case, the A-form population is dominant (∼95%) over the BI-form. This reversal of stability in the two (albeit crude) dielectric models reflects the observation that at low levels of hydration, DNA structures prefer to be in the A-form (2,49,50). Also, it is noted that changing the dielectric model not only changes the relative energetics of the local minima corresponding to the A- and B-forms but it also affects the depth and volume of the corresponding valleys. In the CDIE model, the depth of the BI-form energy valley decreases relative to its value in the RDIE model. This is not true for the A-form valley which gets deeper by almost 13% relative to the RDIE case. This indicates that the preference of DNA to be in either an A- or B-form conformation is not only driven by enthalpic stabilization (which is related to the potential energy difference of respective minima) but it is also influenced by entropic contributions which stem from the different number of states available to either forms (which is reflected in the change in the valley volume upon changing the dielectric model). Consideration of the energy valley volume has also been highlighted as important in understanding the driving forces of folding in a small polypeptide system (42).

Switching from RDIE to the CDIE dielectric model has a significant impact on the potential energy landscape: the valley corresponding to Crank A is no longer present (or very shallow at <0.5 kcal/mol deep) while that corresponding to Crank B (valley 3) becomes very high in energy (∼15 kcal/mol). The disappearance of Crank A from the GC PES in vacuum is of interest as a GC dinucleotide step appears only once in the data matrix of crystal structure observations (Supplementary Table 2).

It is interesting to note that in both electrostatic models, the A- and BI-forms correspond to two distinctive energy valleys as a consequence of the existence of an energy barrier between them. This can be contrasted with the results of a computational study of the free-energy landscape of the A- to B-form conversion in aqueous solution which noted 'the presence of a wide energy well representing the global
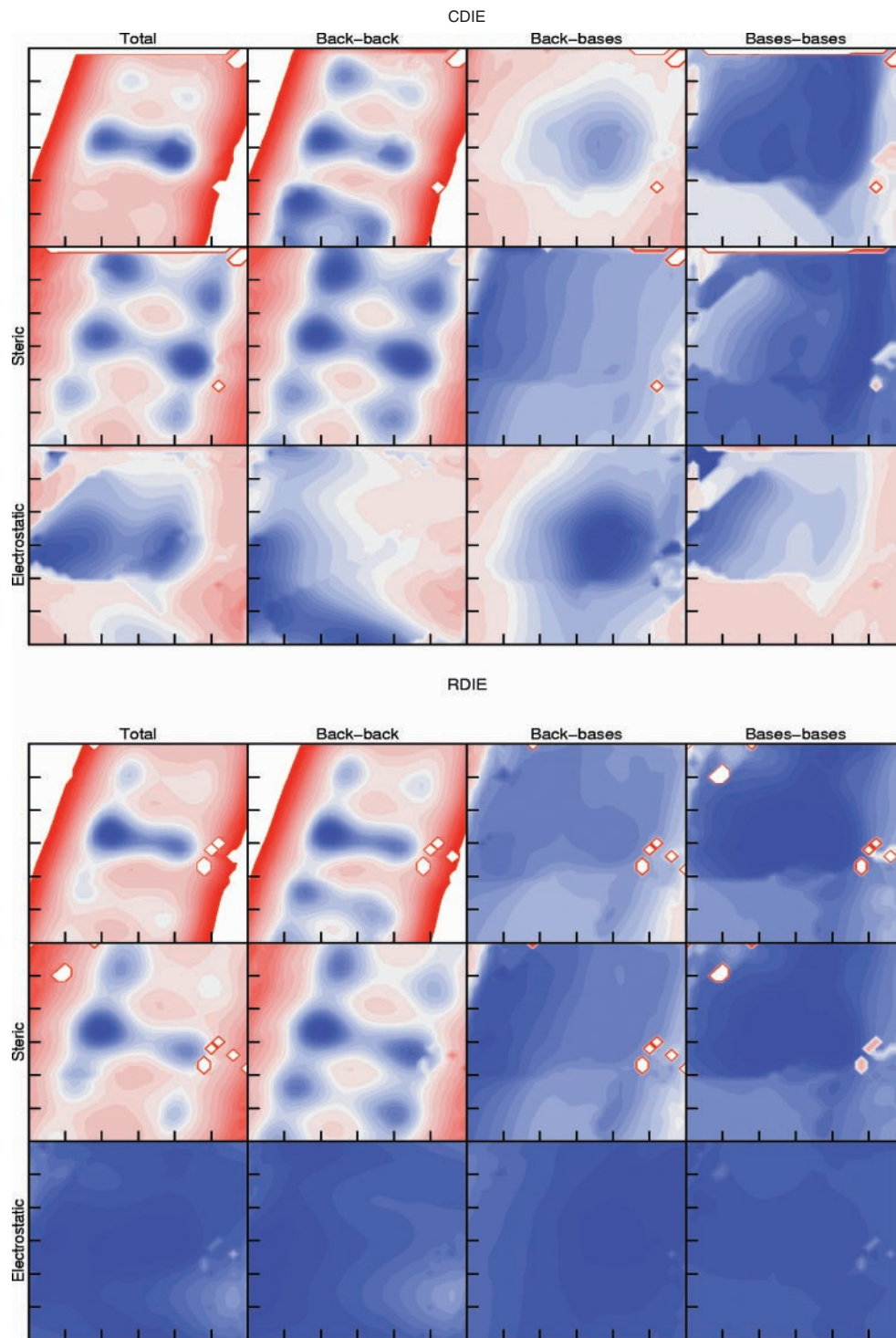
Figure 5. Decomposition of the PES of a GC dinucleotide monophosphate in the principal conformational plane (PC1–PC2) for the slice containing the lowest energy structure: for the vacuum model (CDIE top panel) and for the distance dependent dielectric model (RDIE bottom panel). Row-wise are the component energy terms while column-wise are the interactions between substructures of the system (see Figure 1): 1st column: all atoms, 2nd column: backbone self-interactions, 3rd column: backbone–base interactions, 4th column: base self-interactions.

B-form structure and the absence of a high-energy barrier separating the A- and B-forms' (51). We anticipate that the incorporation of terms relating to the free energy is likely to result in a general flattening of the PES (52). However, a detailed comparison of the topography of the potential and free-energy surfaces within the DNA conformational subspace will require more sophisticated models of the important electrostatic effects [e.g. Poisson–Boltzmann (53)] and the incorporation of conformational sampling (54), which are the subject of on-going work.

Table 2. Location, relative energetics and population of the energy valleys on the PESs of a GC dinucleotide monophosphate in the PCS derived from duplex DNA crystal structures

Minimum	Location of minima in PC subspace			Energy of the minimum[b] (kcal mol^{-1})	Valley depth[c] (kcal mol^{-1})	Valley volume (10^5 degree^3)	% Population at 298 K	Classification (DNA form)
	PC1	PC2	PC3					
(a) Using radius-dependent dielectric model (RDIE)[a]								
1	−80	20	40	0.00	5.50	4.60	96.01	BI
4	100	−20	20	1.88	4.12	2.06	3.97	A
2	−20	220	−20	5.03	5.97	6.07	0.02	CrB
5	140	160	−40	6.82	4.18	3.18	0.00	Crank A
6	40	−240	80	15.49	1.01	0.280	0.00	−CrA
3	−160	−140	40	16.02	1.48	0.913	0.00	−CrB
(b) In vacuum (CDIE)[a]								
1	−80	20	20	2.11	3.5	2.55	5.18	BI
4	100	−20	0	0.00	5.5	2.85	94.82	A
2	−20	200	−20	14.72	1.5	0.83	0.00	CrB
5[d]	—	—	—	—	—	—	—	—
6[d]	—	—	—	—	—	—	—	—
3	−160	−140	60	16.56	4.5	1.32	0.00	−CrB

[a]Minima corresponding to valleys <0.5 kcal mol^{-1} deep were ignored.
[b]Energies are relative to the global energy minimum on the respective PES.
[c]The valley depth is calculated relative to the lowest energy saddle point connecting it to neighbouring valleys.
[d]Not observed as an energy minimum.

Validation of the PES

An important step in validating the computed PES is to compare its fidelity with respect to certain observed structural and thermodynamic properties of DNA.

Distribution of observed structures on the subspace PES. Each observation in the data matrix (containing sample DM fragments from crystal structures of the A-DNA and B-DNA families) was pre-classified into known DNA conformational forms A, Crank A, BI and BII using relevant backbone torsion angle ranges (5–7). Projection of the data matrix into the PCS results in pronounced clustering into the low energy regions on the PES (Figure 6). Thus, the A-form exists mainly in region 4 (for region definitions see Figure 3) while Crank A-form exclusively occupies region 5. Few observations in our data matrix corresponding to A-DNA are found in region 6.

The distribution of the B-DNA observations in the PCS is a little more complicated. The BI-form lies in region 1 in slices at 80° > PC3 > −40°: the density of observations is maximal at PC3 = 20° and declines towards PC3 = −40° where it virtually vanishes. Interestingly, the BII-form also starts to appear in region 1, but at slices corresponding to PC3 < −40°. It is evident that the BII-form exists in a twilight region between region 1, where the BI form is centred, and a very shallow region connected to the tail of region 1 along negative PC3 (Figure 6 lower panel and Figure 3). This rather anomalous location may be the result of deficiencies in our model: either in the force field, or the representation of the environment. For example, the BII-form may be a meta-stable structure in a single strand, but becomes stable in a duplex environment. Alternatively, the BII-form may only be stable as a result of crystal packing effects (55,56). We must also not rule out artefacts in our PCS representation as a result of the loss of ~15% variance by restricting our consideration to only three PCs.

The issue of gauging the fidelity of the force field in biomolecular simulations is fraught with difficulties as it is intimately connected with the extent of sampling of the conformational space. In an extensive molecular dynamics study of tetranucleotides (57), even multi-nanosecond simulations on multiple systems were suggested to be insufficient to fully sample the conformational space of DNA. We note that our static approach, which systematically samples (by interpolation) the conformational space spanning an existing conformational distribution (in our case derived from, but not necessarily limited to, X-ray crystal structures), is complementary to the dynamical approaches.

The substantial agreement of the features of the PES with the distribution of observed conformations validates its physical fidelity and relevance. What is remarkable is that the energetics of a single-strand DM model provides such a good representation of the distribution of conformers from duplex crystal structures. This agreement can be rationalized by the insight that the locations of the minima on the PES are substantially determined by steric interactions within the backbone of the single strand. This observation requires further work in terms of an equivalent treatment of the energetics of the DNA duplex, but it does suggest that the predominant focus on base–base interactions in studies of DNA energetics does not fully acknowledge the importance of local backbone stereochemical effects. In this respect, we support an analogy (7) of this (multivariate) conformational analysis of DNA with the stereochemical analysis of polypeptides (10), where local steric interactions determine the conformations observed in secondary structure. Significantly, our work goes beyond an analysis of the statistical distribution of conformers, to computing and characterizing the underlying PES from which those distributions have arisen.

Sequence dependence of the relative conformational propensities. Other combinations of DMs (16 in total) show similar topographical features to those exhibited by the GC
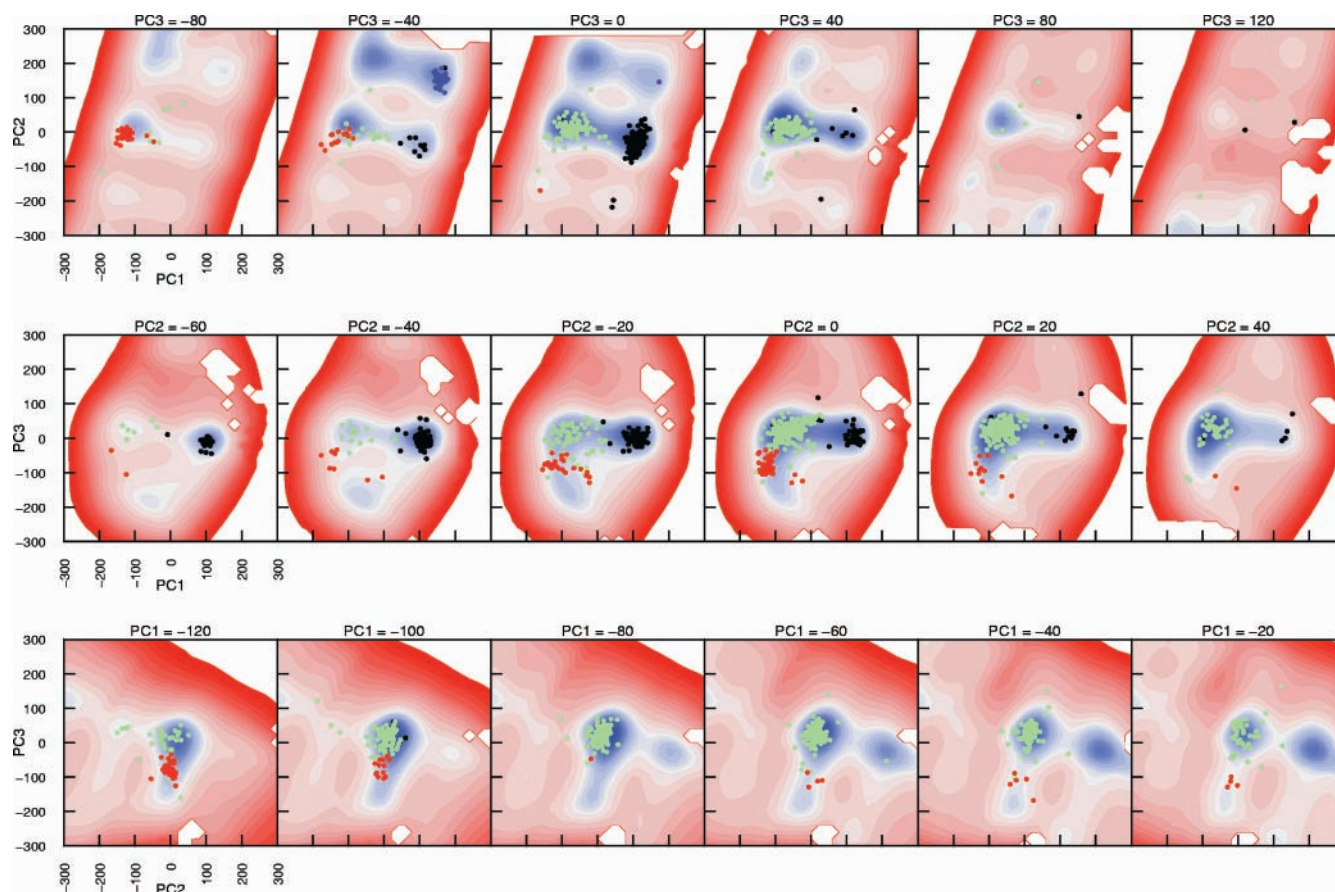
Figure 6. Selected total potential energy slices in PC1–PC2, PC1–PC3 and PC2–PC3 planes of a GC dinucleotide monophosphate fragment, with the distribution of conformers from duplex crystal structures superposed; A-form in black, Crankshaft A-form in blue, BI-form in green while BII in red.

PES. The difference between the 16 PESs can be described in terms of topographical features (relative energetics of the energy minima, the volume of the corresponding valleys) and topological features (connectivity of the energy minima). The former affect the thermodynamic properties of the system while the latter affects its kinetic behaviour (47). In the following, we will discuss the thermodynamic features, while consideration of dynamics is the subject of further work.

Inspection of the equilibrium ratio of the A-form and BI-form conformations for the RDIE model (Table 3) shows that the BI-form is dominant for all DM sequences. Notably, energy states other than the energy minimum have considerable contribution to the equilibrium ratio (Supplementary Figure 1). It is evident that the X-Pyrimidine diunclectides (X = A, T, C or G) show the highest ratios of A-form (yet still minor with respect to BI-form). The existence of A-form is thus not entirely precluded in a water-rich medium and the BI-A form equilibrium ratio is sequence specific. The larger proportion of A-form in d(AT) DM (~9%) relative to its very tiny proportion in the d(AA) DM (~1%) (Table 3) indicates that the shift of the A–B equilibrium towards the A-form (e.g. by formation of an A-form DNA–protein complex) is more probable for d(AT) than d(AA). This is in accord with the observation that runs of adenine bases are reluctant to switch to A-form while a d(AT)$_8$ tract at the 3′ end of the 5 S gene undergoes the B to A transition readily (27,58).

Table 3. The percentage equilibrium ratio of the A-form relative to the BI-form derived from the PES of the 16 possible DMs (using the RDIE model for electrostatics)

	A	G	C	T
A	0.9	0.7	3.6	8.9
G	1.3	0.4	4.0	7.9
C	0.8	0.4	5.2	8.6
T	0.5	0.1	2.3	4.4

Utility of the PES: analysis of transition pathways and dynamical processes in duplex DNA

Thus far, we have established the fidelity of the subspace PES in representing the structural and thermodynamic properties of DNA—we now turn our attention to the utility of the PES in two illustrative applications addressing conformational interconversion processes in DNA: a computation of the pathway between A- and BI-forms of DNA and an analysis of conformational sampling in molecular dynamics simulations of DNA.

BI-to-A conformational transition. The BI-to-A interconversion pathway of a duplex d(AT)$_2$ step was generated by use of the conjugate peak refinement method (59) in the TRAVEL module of CHARMM. Using local energy minima within the valleys corresponding to the BI- and A-forms as initial and final
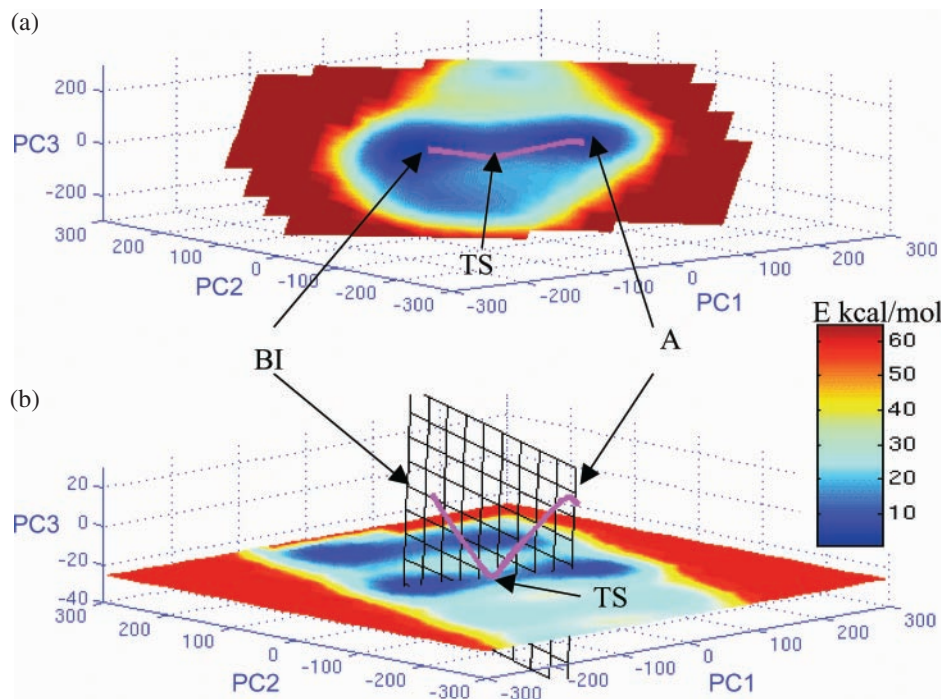
Figure 7. Trajectories of the steepest descent path of the BI-to-A interconversion computed for a duplex d(AT)$_2$ step (shown in magenta) projected onto the PES of a single-strand AT dinucleotide. (**a**) In a plane containing the two minima and the saddle points. (**b**) Alternative view of the trajectory in the PCS with the PES shown in the slice of the PC1–PC2 plane containing the transition states. The trajectories of the two individual strands of the duplex are very similar and not distinguishable at the resolution presented.
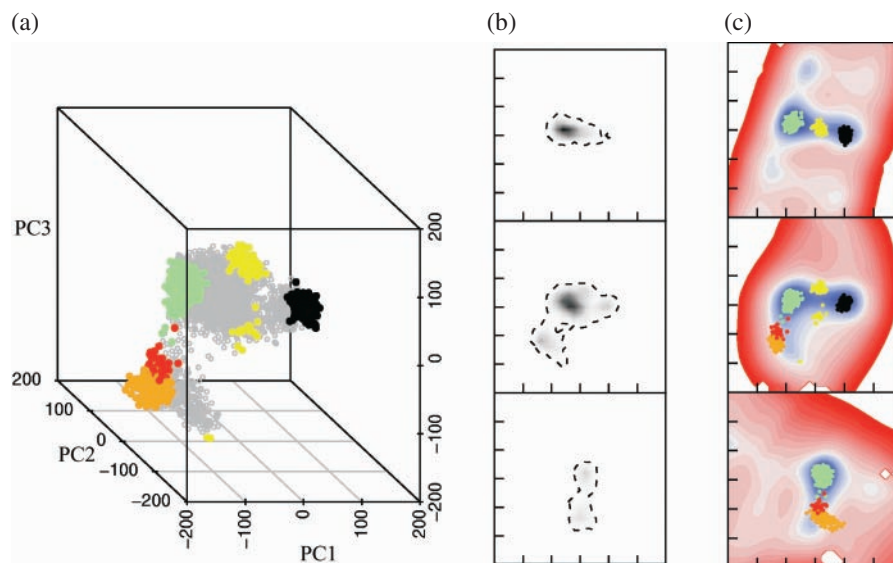

Figure 8. Visualization of a room temperature solvated molecular dynamics simulation of the duplex hexamer d(ATATAT)$_2$ initiated from the canonical B-form. Projection of the trajectory of the middle base step of 'strand a' into the PCS derived from duplex DNA crystal structures. (**a**) A scatter plot of the trajectory projection within the PCS. BI state is shown in green, A-form in black, BII-form in orange; BII-form is in red. Conformations with mixed sugar puckers (i.e. $P_A \cdot P_B$ or $P_B \cdot P_A$) in yellow. (**b**) Density plot of the trajectory projection shown in three perpendicular planes: from top to bottom; PC1–PC2 plane at PC3 = 40, PC1–PC3 plane at PC2 = 0 and PC2–PC3 plane at PC1 = −100. In each plane, the axes extend from −300 to 300 and the tick marks are placed at 20° increments. Relative densities are indicated by gradual change from black (high density) to white (zero density). The trajectory boundary (almost zero density) is indicated by a dashed line. (**c**) d(AT) PES slices in the planes indicated in (b) with the conformational states indicated in (a) superposed. The colour ramp is similar to that used in Figure 3.

structures, respectively, two intermediate saddle points were found and refined to a gradient norm of 10^{-4} kcal mol^{-1} A^{-1}.

The steepest descent path (SDP) of the duplex BI-to-A form pathway, from the perspective of the individual strand conformations, is well described in the PCS with 71.4 and 78.0% of 'strand a' and 'strand b' displacements, respectively. As expected, the best single descriptor for this BI-to-A-DNA process is PC1 which accounts for 64.3% of the displacements

for 'strand a' and 70.0% for 'strand b'. PC1 is a complex delocalized displacement (Table 1) in accord with the notion, developed in other simulation studies, that the BI-A transformation takes place over many degrees of freedom (60,61).

Projection of the trajectories of the individual strands of the SDP of the duplex onto the PES of the single-strand DM suggests that the energy landscape is both representative and descriptive of the process (Figure 7a and b). For each individual strand, the interconversion pathways are almost identical and pass through the same saddle point location located midway between the A- and BI-forms (Figure 7a and b): at this location the topography of the PES is suggestive of a first order saddle point and this was confirmed by examination of the eigen spectrum of the Hessian matrix of the duplex system corresponding to the transition state (data not shown).

Molecular dynamics trajectories. The conformational space sampled in six illustrative 1.0 ns solvated room temperature MD simulations of the duplex hexamer d(ATATAT)$_2$ is well described in the PCS derived from duplex crystal structures, with an average of 60.2% of the total variance captured over DM fragments from the individual strands. The trajectory of one of these simulations, which was initiated from a canonical B-form conformation, is projected into the PCS as shown in Figure 8. The sampling of the conformational space by the trajectory corresponds remarkably well to the energy valleys of the subspace PES; transient structures corresponding to BI-form project into valley 1 and those for the A-form project into valley 4. It is noted that the transient structures corresponding to the BII-form project into the twilight region between the BI valley and its neighbouring shallow valley (see Figure 6). Interestingly, DM conformational states with mixed A/B-form sugar puckers (i.e. $P_A \cdot P_B$ or $P_B \cdot P_A$) lie midway between A- and BI-form regions and their structure and location on the PES are reminiscent of the location of the transition states of the BI-to-A interconversion pathway discussed above. This again illustrates the descriptive capacity of the subspace PES to both dynamical and kinematic behaviour and reinforces its utility as a tool for understanding complex conformational processes of DNA.

CONCLUSIONS

PCA of torsion angles of single-strand DM fragments derived from the A- and B-family duplex DNA crystal structures confirmed that only three collective degrees of freedom, of a nominally high-dimensional space, are sufficient to describe a PCS that accounts for ~85% of the total variance of the conformational distribution in accord with previous studies (7,11).

In this work, we have demonstrated how such collective coordinates can be employed to map the underlying PES within the PCS to provide an underlying physical framework—a conformational energy landscape—in which to interpret the observed structural distributions.

The topography of the PES of single-strand DMs in the subspace spanned by the first three principal components (derived from duplex crystal structures) reveals a set of low energy regions which correspond to different known conformations of duplex DNA namely A, Crankshaft A-, BI- and BII-forms. The PES suggests the possibility of a previously unidentified B-form structure with α, γ ranges corresponding

to Crankshaft A-form, which we term the Crankshaft B-form (CrB). Such conformations are not observed in isolated duplex crystal structures, but can be found in the crystal structures of protein complexes with B-DNA (21).

An analysis of the components of the PES suggests that the conformational states observed in duplex DNA are predominantly determined by steric interactions within the backbone of a single-strand DM, with electrostatic interactions serving to fine-tune the relative stability of the different forms.

The PES was constructed using two different dielectric models namely, a distance-dependent dielectric (RDIE) which crudely mimics the screening of charges in a water-rich medium and a constant dielectric (CDIE) which represents vacuum. The different models exhibit extensive changes to the PES and result in relative preferences for the A- and B-forms in accord with the observed preference for DNA to be in the B-form in a water-rich medium, and in the A-form when under low hydration conditions (2). Furthermore, a thermodynamic analysis revealed that the preference for the B-form in a water-rich medium is not only driven by enthalpic differences, but also entropic contributions relating to the relative volumes of the different energy valleys of the PES.

A thermodynamic analysis of all 16 possible DM PESs reveals that the conformation of the BI-form is dominant for all pairs of sequences. However, the sequence-specific trends for the relative proportion of the A-form suggest that d(AT) is more *A-philic* than d(AA) which is in accord with experimental observation (27,58).

The PES of a single-strand DM provides a good description of conformational processes in duplex DNA. For example, the pathways of BI-to-A form conformational transitions are well described by the topography of the PES. Furthermore, the features of the coarse-grained, subspace PES are remarkably faithful to the fine details of conformational processes in the duplex, such as the location of the transition states.

The representative nature of the PCS and the descriptive power of the PES suggest that they will be useful in the development and comparison of interaction potentials for DNA (39) and for the evaluation of conformational sampling in molecular simulations (57,62).

In conclusion, the work presented here provides a powerful and coherent framework for the analysis of DNA structure, energetics and dynamics. Further work is aimed at applying the approach to mapping the PES of duplex DNA as well as in developing the underlying methodology and extending its reach to other biomolecular systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Neidle,S. (1999) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, Oxford, New York.
2. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
3. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.
4. McCall,M., Brown,T. and Kennard,O. (1985) The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG).poly(dC). *J. Mol. Biol.*, **183**, 385–396.
5. Schneider,B., Neidle,S. and Berman,H.M. (1997) Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113–124.
6. Shakked,Z. and Rabinovich,D. (1986) The effect of the base sequence on the fine structure of the DNA double helix. *Prog. Biophys. Mol. Biol.*, **47**, 159–195.
7. Beckers,M.L.M. and Buydens,L.M.C. (1998) Multivariate analysis of a data matrix containing A-DNA and B-DNA dinucleotide monophosphate steps: multidimensional Ramachandran plots for nucleic acids. *J. Comput. Chem.*, **19**, 695–715.
8. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTCGCG. *J. Biol. Chem.*, **257**, 14686–14707.
9. Conner,B.N., Yoon,C., Dickerson,J.L. and Dickerson,R.E. (1984) Helix geometry and hydration in an A-DNA tetramer: C-C-G-G. *J. Mol. Biol.*, **174**, 663–695.
10. Ramachandran,G.N., Ramakrishan,C. and Sasisekharan,V. (1963) Stereochemistry of polypeptide-chain configurations. *J. Mol. Biol.*, **7**, 95.
11. Sims,G.E. and Kim,S.H. (2003) Global mapping of nucleic acid conformational space: dinucleoside monophosphate conformations and transition pathways among conformational classes. *Nucleic Acids Res.*, **31**, 5607–5616.
12. Flatters,D., Zakrzewska,K. and Lavery,R. (1997) Internal coordinate modeling of DNA: force field comparisons. *J. Comput. Chem.*, **18**, 1043–1055.
13. Sarai,A., Jernigan,R.L. and Mazur,J. (1996) Interdependence of conformational variables in double-helical DNA. *Biophys. J.*, **71**, 1507–1518.
14. Ulyanov,N.B., Zhurkin,V.B. and Ivanov,V.I. (1982) Analysis of the DNA conformational flexibility for the different nucleotide-sequences. *Stud. Biophys.*, **87**, 99–100.
15. Zhurkin,V.B., Poltev,V.I. and Florentev,V.L. (1980) Atom–atom potential functions for conformational calculations of nucleic-acids. *Mol. Biol.*, **14**, 882–895.
16. Hunter,C.A. and Lu,X.J. (1997) DNA base stacking interactions: a comparison of theoretical calculations with oligonucleotide crystal structures. *J. Mol. Biol.*, **265**, 603–619.
17. Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: dinucleotide conformational maps. *J. Mol. Biol.*, **295**, 71–83.
18. ElHassan,M.A. and Calladine,C.R. (1997) Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Philos. Trans. R. Soc. Lond. Ser. A-Math. Phys. Eng. Sci.*, **355**, 43–100.
19. Srivinsan,A.R. and Olson,W.K. (1987) Nucleic acid model building. The multiple backbone solutions associated with a given base morphology. *J. Biomol. Struct. Dyn.*, **4**, 895–938.
20. Prive,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.S., Kopka,M.L. and Dickerson,R.E. (1987) Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science*, **238**, 498–504.
21. Varnai,P., Djuranovic,D., Lavery,R. and Hartmann,B. (2002) Alpha/gamma transitions in the B-DNA backbone. *Nucleic Acids Res.*, **30**, 5398–5406.
22. Tisne,C., Delepierre,M. and Hartmann,B. (1999) How NF-[kappa]B can be attracted by its cognate DNA. *J. Mol. Biol.*, **293**, 139–150.
23. Schroeder,S.A., Roongta,V., Fu,J.M., Jones,C.R. and Gorenstein,D.G. (1989) Sequence-dependent variations in the 31P NMR spectra and backbone torsional angles of wild-type and mutant Lac operator fragments. *Biochemistry*, **28**, 8292–8303.
24. Dostal,L., Chen,C.Y., Wang,A.H. and Welfle,H. (2004) Partial B-to-A DNA Transition upon minor groove binding of protein Sac7d monitored by Raman spectroscopy. *Biochemistry*, **43**, 9600–9609.
25. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
26. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
27. Becker,M.M. and Wang,Z. (1989) B-A transitions within a 5 S ribosomal RNA gene are highly sequence-specific. *J. Biol. Chem.*, **264**, 4163–4167.
28. Mohr,S.C., Sokolov,N.V., He,C.M. and Setlow,P. (1991) Binding of small acid-soluble spore proteins from *Bacillus subtilis* changes the conformation of DNA from B to A. *Proc. Natl Acad. Sci. USA*, **88**, 77–81.
29. Flader,W., Wellenzohn,B., Winger,R.H., Hallbrucker,A., Mayer,E. and Liedl,K.R. (2001) B-I to B-II substrate transitions induce changes in the hydration of B-DNA, potentially mediating signal transduction from the minor to major groove. *J. Phys. Chem. B*, **105**, 10379–10387.
30. Pichler,A., Rudisser,S., Mitterbock,M., Huber,C.G., Winger,R.H., Liedl,K.R., Hallbrucker,A. and Mayer,E. (1999) Unexpected BII conformer substrate population in unoriented hydrated films of the d(CGCGAATTCGCG)2 dodecamer and of native B-DNA from salmon testes. *Biophys. J.*, **77**, 398–409.
31. Winger,R.H., Liedl,K.R., Pichler,A., Hallbrucker,A. and Mayer,E. (2000) B-DNA's B-II conformer substrate population increases with decreasing water activity. 1. A molecular dynamics study of d(CGCGAATTCGCG)(2). *J. Phys. Chem. B*, **104**, 11349–11353.
32. Altona,C. and Sundaralingam,M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of psuedorotation. *J. Am. Chem. Soc.*, **94**, 8205–8212.
33. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
34. ElHassan,M.A. and Calladine,C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*, **259**, 95–103.
35. Reijmers,T.H., Wehrens,R. and Buydens,L.M.C. (2001) Circular effects in representations of an RNA nucleotides data set in relation with principal components analysis. *Chemom. Intell. Lab. Syst.*, **56**, 61–71.
36. Jackson,J.E. (1991) *A User's Guide to Principal Components*. Wiley, New York.
37. Kitao,A. and Go,N. (1999) Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.*, **9**, 164–169.
38. Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S. and Weiner,P. (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
39. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **118**, 2309–2309.
40. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.
41. Gelin,B.R. and Karplus,M. (1975) Sidechain torsional potentials and motion of amino-acids in proteins: Bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci. USA*, **72**, 2002–2006.
42. Becker,O.M. (1997) Quantitative visualization of a macromolecular potential energy ''funnel''. *Theochem-J. Mol. Struct.*, **398**, 507–516.
43. Barber,C.B., Dobkin,D.P. and Huhdanpaa,H.T. (1996) The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, **22**, 469–483.
44. Krzanowski,W.J. (2000) *Principles of Multivariate Analysis*. Oxford University Press.
45. Norberg,J. and Nilsson,L. (1995) Temperature dependence of the stacking propensity of adenylyl-3′,5′-adenosine. *J. Phys. Chem.*, **99**, 13056–13058.
46. Cheatham,T.E., Srinivasan,J., Case,D.A. and Kollman,P.A. (1998) Molecular dynamics and continuum solvent studies of the stability of polyG–polyC and polyA–polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.*, **16**, 265–280.
47. Wales,D.J. (2003) *Energy Landscapes*. Cambridge University Press, Cambridge.
48. McQuarrie,D.A. (1973) *Statistical Thermodynamics*. University Science Books, Mill Valley, CA.

49. Cheatham,T.E. and Brooks,B.R. (1998) Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theor. Chem. Acc.*, **99**, 279–288.

50. Jose,D. and Porschke,D. (2004) Dynamics of the B-A transition of DNA double helices. *Nucleic Acids Res.*, **32**, 2251–2258.

51. Banavali,N.K. and Roux,B. (2005) Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.*, **127**, 6866–6876.

52. Brooks,C.L., Karplus,M. and Pettitt,B.M. (1988) *Proteins:A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. Wiley, New York, Chichester.

53. Gilson,M.K. (1995) Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.*, **5**, 216–223.

54. Roux,B. (1995) The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.*, **91**, 275–282.

55. Dickerson,R.E., Goodsell,D.S., Kopka,M.L. and Pjura,P.E. (1987) The Effect of crystal packing on oligonucleotide double helix structure. *J. Biomol. Struct. Dyn.*, **5**, 557–579.

56. Heinemann,U. and Hahn,M. (1992) C-C-A-G-G-C-M5c-T-G-G—helical fine-structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G. *J. Biol. Chem.*, **267**, 7332–7341.

57. Beveridge,D.L., Barreiro,G., Byun,K.S., Case,D.A., Cheatham,T.E.,III, Dixit,S.B., Giudice,E., Lankas,F., Lavery,R., Maddocks,J.H. *et al.* (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. research design and results on d(CpG) steps. *Biophys. J.*, **87**, 3799–3813.

58. Pilet,J., Blicharski,J. and Brahms,J. (1975) Conformations and structural transitions in polydeoxynucleotides. *Biochemistry*, **14**, 1869–1876.

59. Fischer,S. and Karplus,M. (1992) Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition-states in systems with many degrees of freedom. *Chem. Phys. Lett.*, **194**, 252–261.

60. Cheatham,T.E. and Kollman,P.A. (2000) Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.*, **51**, 435–471.

61. Cheatham,T.E. and Kollman,P.A. (1996) Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.*, **259**, 434–444.

62. Feig,M. and Pettitt,B.M. (1998) Structural equilibrium of DNA represented with different force fields. *Biophys. J.*, **75**, 134–149.