Structural DNA Profiles: Single Sequence Queries

Linda Hirons,[†,‡] Eleanor J. Gardiner,[†,‡] Christopher A. Hunter,*[,†] and Peter Willett[‡]

Centre for Chemical Biology, Krebs Institute for Biomolecular Science, Department of Chemistry, University of Sheffield, Sheffield S3 7HF, United Kingdom, and Department of Information Studies, University of Sheffield, Sheffield S1 4DP, United Kingdom

Received September 9, 2005

Structural DNA profiles use the structural properties of the constituent octamers either to observe any characteristics of a single sequence that are unusual (a single sequence query) or to visualize a pattern common to a set of sequences (a multiple sequence query). They are an aid in understanding structural reasons for functional DNA activity. Profiles that answer single sequence queries are introduced and Profile Manager (a software application developed to automate profile generation) is presented. Two sequences that are similar by their nucleotide composition but are known to be very different by structure are analyzed, resulting in useful illustrations that agree with the experimental nuclear magnetic resonance structures.

INTRODUCTION

The Human Genome Project[1−3] and sequencing of genomes from other species has led to an explosion in the amount of DNA sequence information to analyze. Many sequences have unknown but presumably important functions that have been conserved by evolution.[4,5] Discovering the functional purposes of some of this so-called 'junk DNA' has been likened to searching through 'heirlooms in the attic' and finding hidden gems.[6] The development of computational tools to detect any unusual characteristics of a particular sequence that might be of functional importance will therefore be valuable.

Attempts have been made to capture the nucleotide content of a DNA sequence graphically. One such example is a path followed in two-dimensional space, where each C, T, A, and G refers to a movement north, south, west, and east, respectively.[7] These graphical walks have been extended to three dimensions by describing a sequence of bases by movements along the four vertices of a tetrahedron.[8] A Z-curve representation[9] also exists, where a curve of N points represents a sequence of length N. The x-coordinate represents the ongoing ratio of purine to pyrimidine bases, the y-coordinate represents the amino-to-keto ratio, and the z-coordinate represents weak to strong hydrogen bonding in base pairs.

The use of structural properties to identify patterns within families of sequences is justified by the observation that many very different sequences have similar structural properties.[10] This means that by looking at the information hidden within the structure, similarities between DNA sequences can be found that would otherwise be unrecognized. Observing how the structure of a sequence varies across its length not only helps predict unknown functions but also is a key to understanding the structural mechanisms involved in known functions (something that cannot be done by looking at a string of nucleotide letters). This work uses the Octamer Database[11] to describe DNA structure and answer single sequence queries via a set of single sequence profiles. Each profile gives a graphical illustration of how a particular structural parameter from the database varies across the sequence's length, with any special regions highlighted.

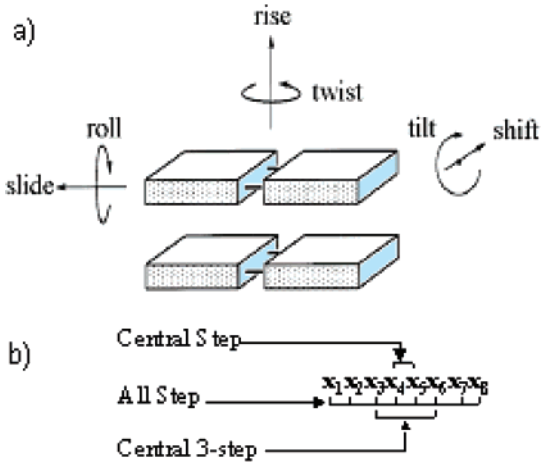

Figure 1. The step parameters. (a) Nucleotides represented by blocks, step parameters represent the movement of the top base pair relative to the bottom base pair in the directions indicated by the arrows. View from the minor groove. (b) Position of central step, all-step, and central 3-step relative to the octamer $x_1x_2x_3x_4x_5x_6x_7x_8$, where x = A, C, G, or T.

The Octamer Database[11] contains the calculated structural properties of all possible DNA octamers ($x_1x_2x_3x_4x_5x_6x_7x_8$ where x = A, C, G, or T). Its contents may be split into two categories: the properties that describe an octamer's minimum energy structure and those that measure an octamer's flexibility. The minimum energy structure is described by the step parameters and three ground state properties: the minor groove width, energy and root-mean-square deviation from a straight path (RMSD). The step parameters describe the geometry of two DNA base-pairs via six degrees of

* Corresponding author phone: (+44) 0114 2229476; e-mail c.hunter@sheffield.ac.uk.
[†] Department of Chemistry, University of Sheffield.
[‡] Department of Information Studies, University of Sheffield.

10.1021/ci050385a CCC: $33.50
Published on Web 01/25/2006

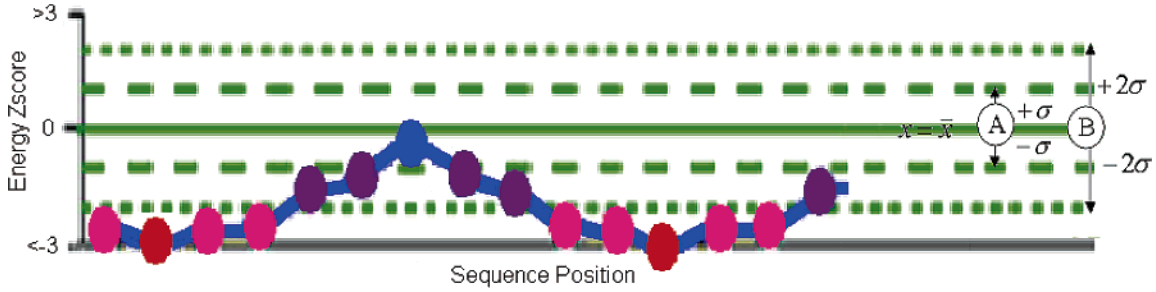

Figure 2. Example of a structural profile for energy of a sequence. Zscore boundaries marked by green lines. For a normal distribution 68% of the data falls in region A (blue dots) and 95% in region B (purple dots). Data beyond 95% is colored red.

freedom, viz, roll, twist, tilt, rise, slide, shift. The profiles use the central step (roll, twist, tilt, rise, slide, shift) and central 3-step (twist3, roll3, slide3, and shift3) parameters alone (Figure 1). Flexibility is measured by the partition coefficients that are related to the number of different conformations accessible at room temperature. Only twist and roll flexibility are considered here, since these two rotations have been recognized as important in protein−DNA recognition.[12,13] 3-step flexibility is separated out into four components that describe flexibility in terms of decreasing and increasing 3-step roll and twist from its minimum energy value ($3Q^-_{Roll}$, $3Q^+_{Roll}$, $3Q^-_{Twist}$, $3Q^+_{Twist}$). These components can be combined to give overall 3-step roll, twist, and total flexibility ($3Q_{Roll}$, $3Q_{Twist}$, $3Q_{Total}$).[11]

Structural parameter plots of DNA can be obtained from the plot.it server[14] or from DNAssist.[15] Plot.it has 45 parameters to choose from (17 of which are energy related, 16 are roll, twist, or tilt dinucleotide parameters, and 7 are flexibility measures). Smoothing options are also available. Here we introduce Profile Manager, a tool to plot the variation of structural parameters along a sequence, using structural property values obtained from the Octamer Database.

SINGLE SEQUENCE PROFILES

Consider a single DNA sequence and a single structural parameter. The first step in generating a structural profile is to convert the nucleotide sequence of length N into its consecutive overlapping (N-7) octamer sequences. For example, the 10-letter sequence AACTTTGGTC is converted into 3 octamers: AACTTTGG, ACTTTGGT, and CTTTGGTC. The chosen parameter's values are then retrieved from the octamer database for these octamer units. They are then each converted into a Zscore value that measures the importance/significance of a particular value of a parameter

$$\text{Zscore} = \frac{x - \bar{x}}{\sigma} \qquad (1)$$

where x is a particular value under consideration, $\bar{x}$ is the mean of the parameter across the population of all possible octamers, and σ is the population standard deviation.

A profile is then constructed by plotting the Zscore values against the sequence length, as illustrated in Figure 2. Cutoffs of <-3 and >3 at the minimum and maximum of the Zscore scale are used for visual purposes when comparing several profiles. Any parameters that fall outside this range are assigned values of −3 or +3 accordingly. The Zscore is the number of standard deviations a value is from the parameter's population mean. When considering a normal distribution, there is a 68% chance that a value will fall within plus or minus one standard deviation of the mean (region A in Figure 2), and a 95% chance that a value will fall within plus or minus two standard deviations (region B in Figure 2). Therefore any value that falls outside of these two boundaries (marked by green lines on a profile) is significantly different from average. Each value along a sequence has been color coded, to highlight any special regions. At the two extremes, blue means average (within region A) and red means special (outside region B). Bright red values are used for the extreme values, 3 and −3. Intermediate values, those between one and two standard deviations from the mean, are shown in purple. Important features can be seen easily at a glance by focusing solely upon the shades of red (and, to a lesser extent, purple).

The computational methods used to generate the Octamer Database were parametrized and validated using high-resolution structural information from the X-ray crystal structures of DNA oligomers.[16] In conformational searches on DNA oligomers for which X-ray crystal structure information was available, the global energy minimum structure corresponded to the experimental structure in 60−70% of the examples studied. An important assumption in the approach developed here is that the structural properties of a longer DNA sequence can be reasonably well described by the corresponding properties of a series of overlapping octamer fragments. To test the validity of this hypothesis, we can compare experimental structural profiles for oligomers for which X-ray crystal structure information is available with those generated from the Octamer Database.

Figure 3 shows the minimum energy roll profile for 10 dodecamers that were used in the original benchmarking exercise.[16] The agreement between the experimental and calculated profiles is excellent for 6 of the 10 examples. The three examples at the bottom of Figure 3, that all show significant differences between calculation and experiment, are cases where the X-ray crystal structure does not correspond to the global energy minimum structure. There are several possible explanations for these discrepancies: crystal packing can perturb the structure from the preferred solution conformation; the crystallization process samples one of several possible low energy conformations, and there are deficiencies in the force-field used. However, for the examples where a full conformational search performs well, the structure profiles generated from a set of overlapping

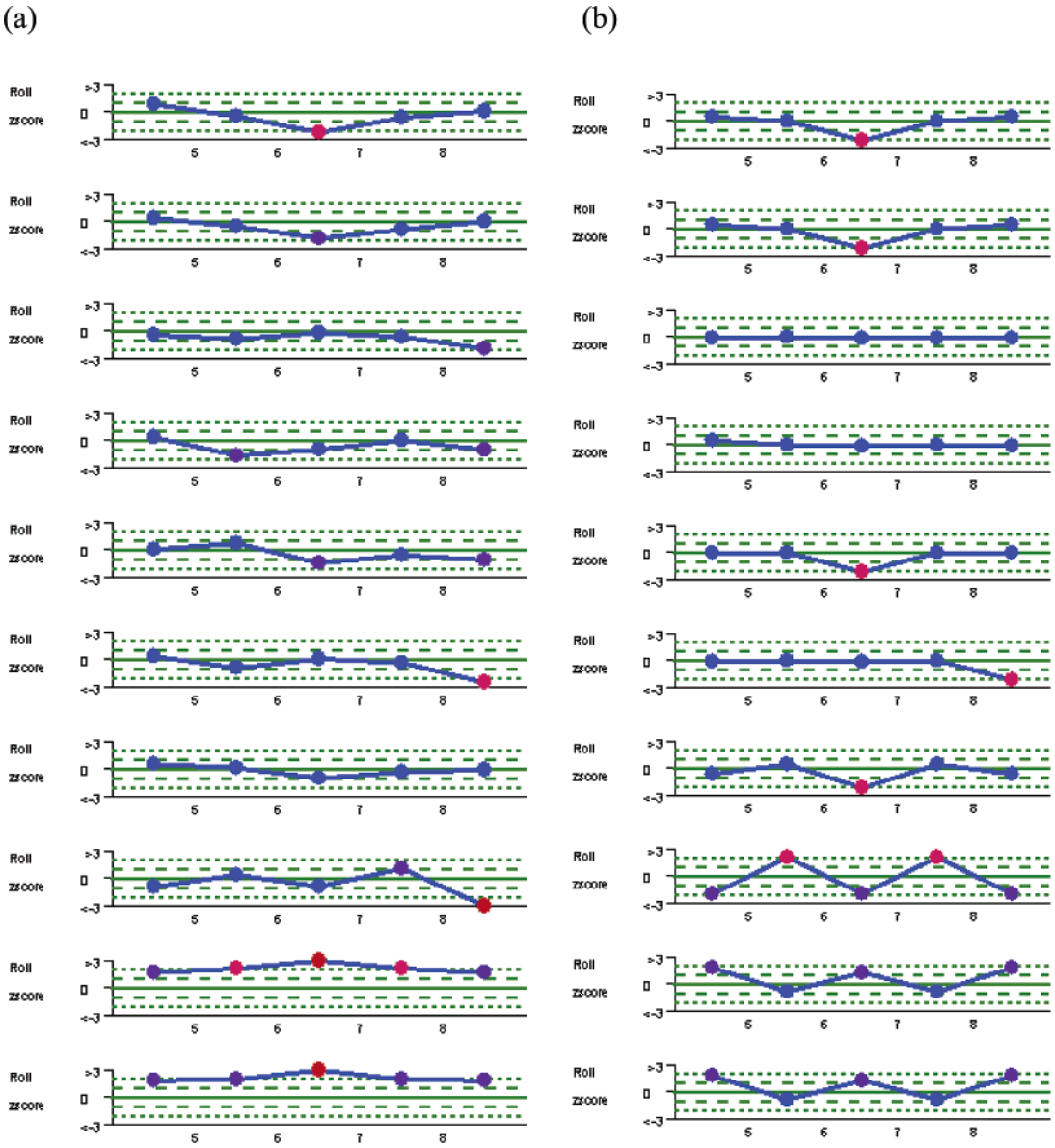

Figure 3. Roll Profiles of DNA dodecamers (pdb codes from top to bottom: bdl001, bdl029, bdl006, bdl047, bdl038, bdl015, bdl042, bdl007, adl046, adl045). (a) Profiles generated using the experimental X-ray crystal structures. (b) Profiles generated using the Octamer Database.

octamers are very similar to those found in the experimental X-ray crystal structures. The approach therefore provides a useful tool for analyzing DNA structural profiles at low computational cost. The use of these structural profiles to analyze the minimum energy structure and the flexibility of a sequence, to identify any interesting characteristics, is illustrated here using the A-tract phenomenon.

THE A-TRACT PHENOMENON

This example uses single sequence profiles to identify why two sequences that appear similar by their nucleotide composition are so different structurally (the A-tract phenomenon). An A-tract sequence is one that contains four or more adjacent adenine bases without a T−A step. The A-tract phenomenon refers to the difference between the bent A-tract structure d(A_4T_4) and the straight structure d(T_4A_4). Note that d(S) means a sequence composed of repeating units of the subsequence S.

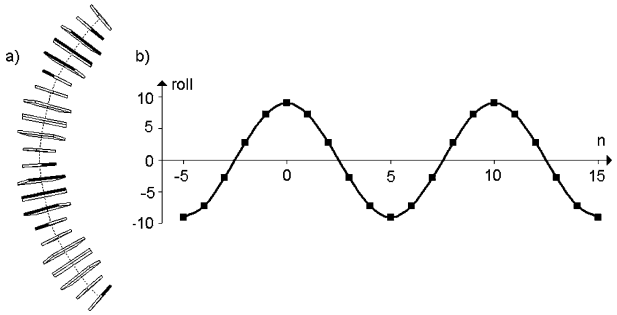

Figure 4. DNA bending. (a) Sequence with 45° bend per double-helical turn (Reprinted with permission from Calladine, C. R.; Drew, H. R. *Understanding DNA: the molecule and how it works*; p 78. Copyright 1997 Elsevier.). (b) A plot of roll angle versus step number along the sequence.

First, consider what properties of the DNA can make it bend. DNA will bend if there is a periodic roll pattern, such as that shown in Figure 4.[17] The curvature of 45° per helical

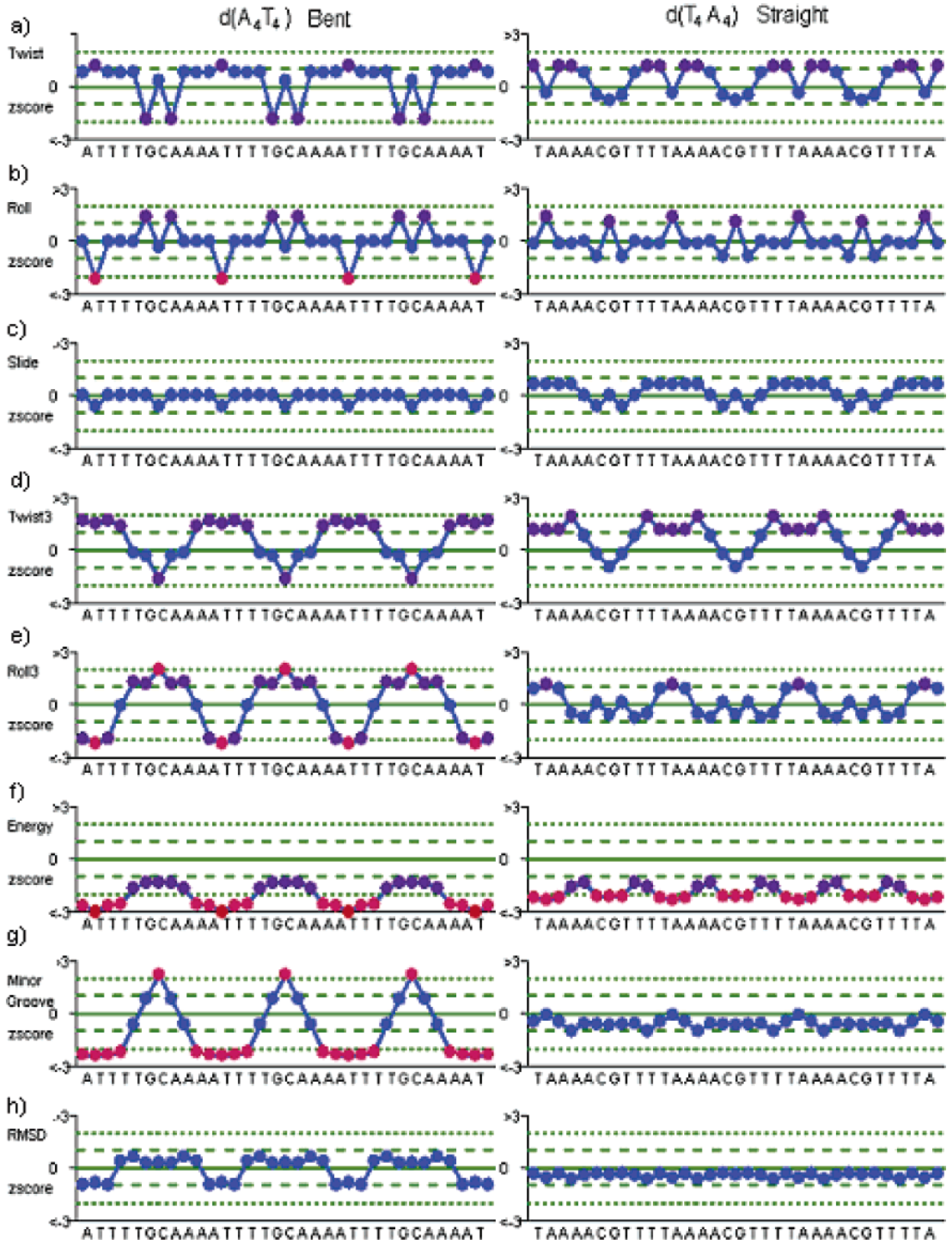

Figure 5. Structure profiles of the bent A-tract sequence d(A_4T_4) (left) and the straight d(T_4A_4) (right) sequence: (a) twist, (b) roll, (c) slide, (d) 3 step twist, (e) 3 step roll, (f) energy, (g) minor groove width, and (h) rmsd of the overall path from straight.

turn shown occurs when roll at step n (R_n) varies as a cosine wave along the sequence (equation 2). Thus sequence-dependent values of roll and roll3 are likely to be important determinants of the A-tract phenomenon.

$$R_n = 9° \cos(36° n) \qquad (2)$$

Recent nuclear magnetic resonance (NMR) structures of d(CA_4T_4G) and d(GT_4A_4C) (referred to as d(A_4T_4) and d(T_4A_4), respectively) have identified some interesting structural characteristics,[18] the majority of which are clearly predicted by the structural profiles shown in Figure 5. Important features can be seen easily by focusing solely upon the shades of red, apparent in the roll, roll3, energy, and minor groove profiles. Here it must be emphasized that the structural profiles, although in strong agreement with the experimental structures, have been arrived at completely independently, using the calculated octamer properties.

The NMR structure reveals that AT steps of the d(A_4T_4) structure have large negative rolls, which are in phase with the large positive roll at the junction with the non-A-tract

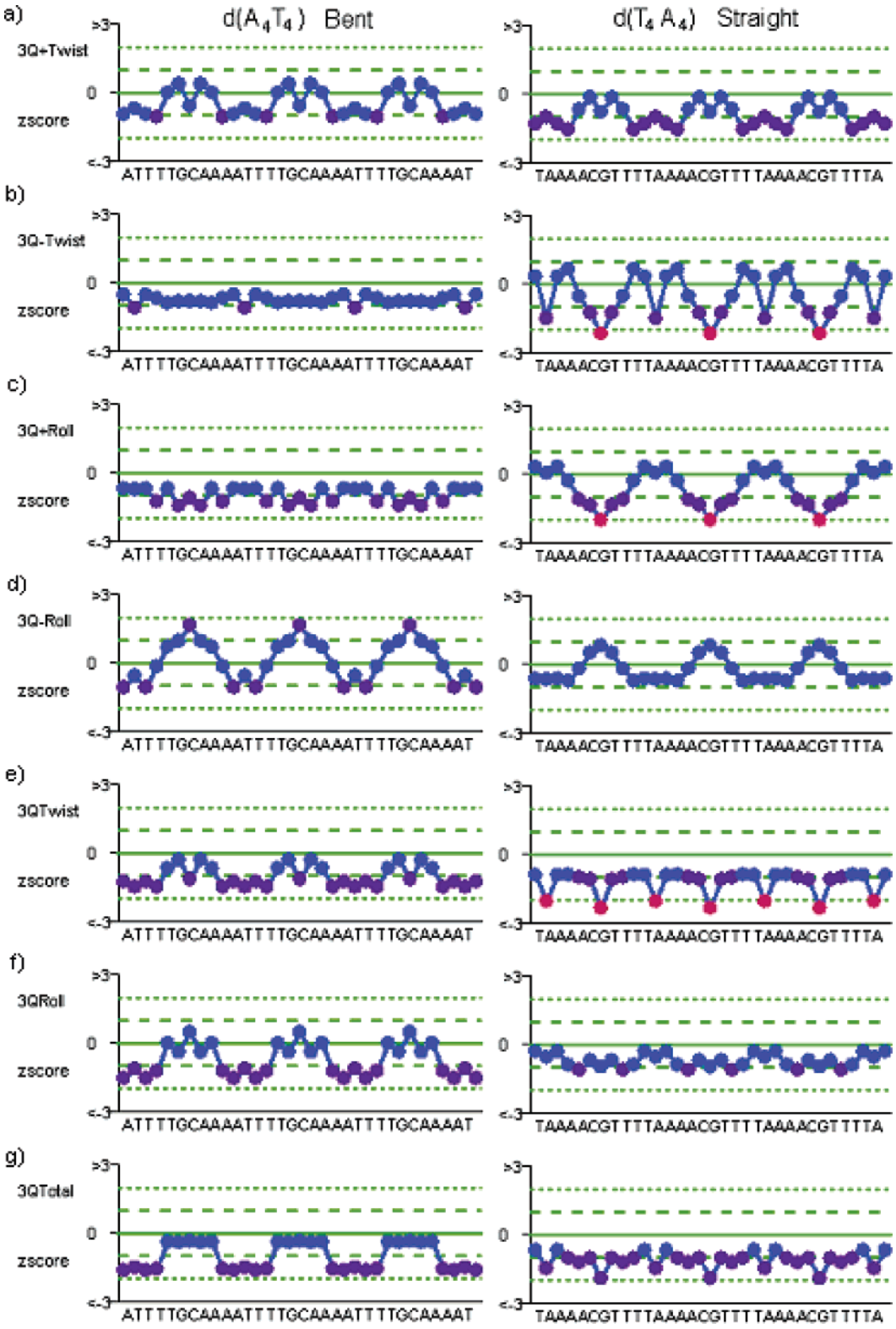

Figure 6. Flexibility profiles of the bent A-tract sequence d(A_4T_4) (left) and the straight d(T_4A_4) (right) sequence: (a) partition coefficient for increasing 3 step twist, (b) partition coefficient for decreasing 3 step twist, (c) partition coefficient for increasing 3 step roll, (d) partition coefficient for decreasing 3 step roll, (e) total partition coefficient for flexibility in 3 step twist, (f) total partition coefficient for flexibility in 3 step roll, and (g) total partition coefficient for flexibility in 3 step twist and roll.

DNA.[18] This is in good agreement with the theoretical requirements for DNA bending, illustrated in Figure 4b. This phasing is strikingly apparent in the 3-step roll profile (Figure 5e) where the 10 base-pair periodic transitions between low and high 3-step roll in d(A_4T_4) reflect the smooth wavelike pattern of roll angles that cause DNA curvature.[17] The d(T_4A_4) sequence also shows some periodicity in 3-step roll, but the amplitude is small. Profile Manager shows that compared to d(A_4T_4), the d(T_4A_4) sequence has a relatively flat roll profile, indicating that the DNA will assume a relatively straight conformation, in agreement with the NMR structure.[18]

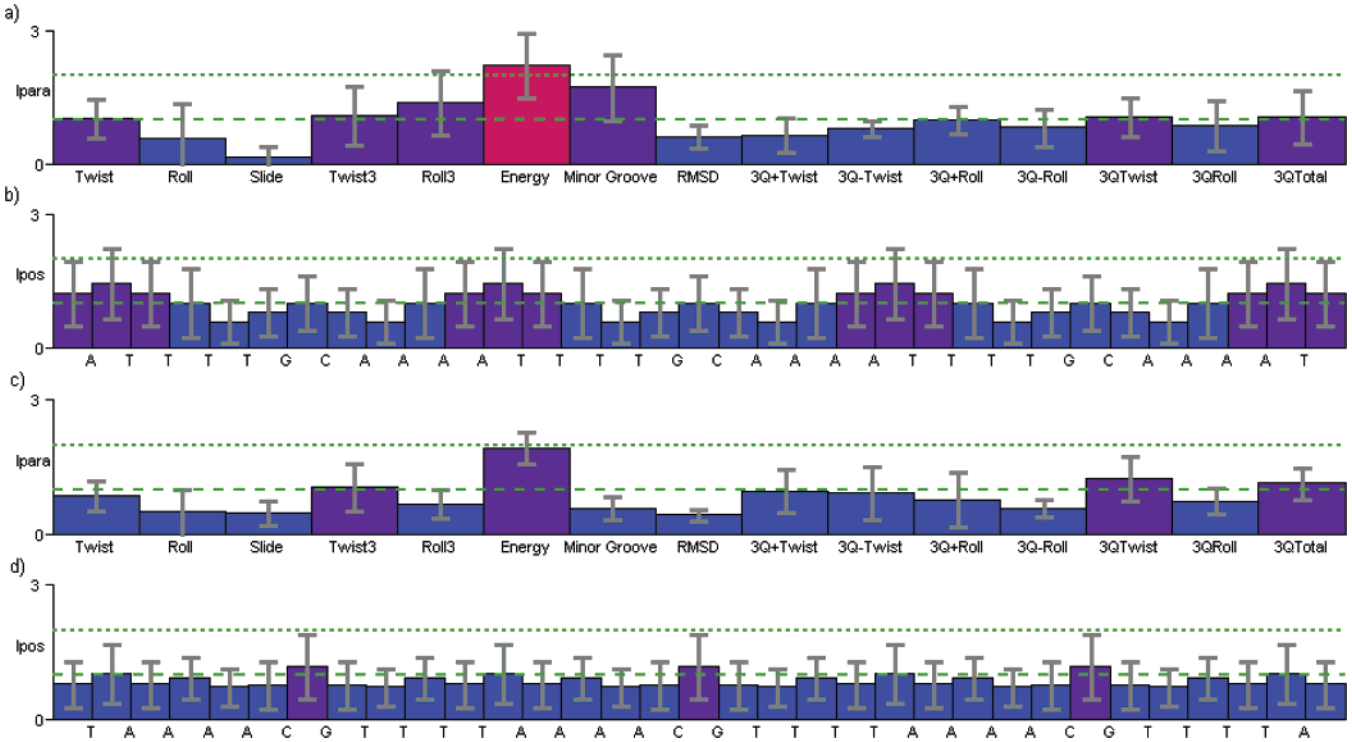

Figure 7. A-tract phenomenon summaries. (a) d(A_4T_4) parameter summary, (b) d(A_4T_4) position summary, (c) d(T_4A_4) parameter summary, and (d) d(T_4A_4) position summary.

Another important feature of the NMR structures is the "remarkable differences" between the minor groove width along the two sequences. The authors note the progressive narrowing of the d(A_4T_4) minor groove with the minimum at the AT step.[18] This feature is clearly apparent in the structural profile (Figure 5g) which is generated using the modeled structures from the octamer database. Here, narrow stretches of the minor groove, interrupted by wide grooves at each of the GC steps, are found. In comparison, there is nothing noteworthy about the groove widths along d(T_4A_4) (Figure 5g), although the NMR structure suggests that the minor groove is relatively wide at the TA step. Note that the energy of both sequences is extremely low across the majority of their lengths, meaning that these sequences are relatively stable.

Flexibility profiles, in terms of the 3-step partition coefficients, show no significantly flexible regions in either sequence (Figure 6). Significantly rigid steps can however be found in the d(T_4A_4) sequence. The CG step is rigid in the decreasing twist, increasing roll and overall twist directions (Figure 6b,c,e), and the TA step is rigid with respect to overall twist (Figure 6e). The d(A_4T_4) and d(T_4A_4) increasing twist flexibility profiles are almost identical. Looking at the $3Q_{Total}$ profiles it can be concluded that the overall flexibilities are similar (Figure 6g) but that the flexibility properties of d(A_4T_4) have a strong periodicity compared to d(T_4A_4).

SUMMARY CHARTS

The information contained within a set of profiles can be summarized by two bar charts: one displaying the general importance of each parameter (I_{para}) across the sequence's entire length and the other displaying the general importance of each position along the sequence (I_{pos}) with respect to all the parameters. In both cases the importance (bar height) is measured by averaging the appropriate Zscore magnitudes

$$I_{para} = \frac{\sum_{i=1}^{N} /\mathrm{Zscore}(p,i)/}{N}, \quad I_{pos} = \frac{\sum_{p=P_1}^{P_M} /\mathrm{Zscore}(p,i)/}{M} \qquad (3)$$

where N is the profile length, M is the number of parameters being considered, P_x is the xth parameter under consideration, and */*Zscore*(p,i)/* is the modulus of the Zscore at position i with respect to parameter p. In the summary charts (Figure 7) note the presence of green lines and color coding identical to that used in the structural profiles. Grey lines extending above and below the bars show the standard deviations of the averaged Zscores.

The parameter summary for d(A_4T_4) shows that energy is the most important parameter followed by minor groove then 3-step roll (Figure 7a). The fluctuating position importance along the repeating sequence has its maxima at the AT steps (Figure 7b). The summary for d(T_4A_4) also has energy as the most important parameter but not the minor groove or 3-step roll (Figure 7c). Twist flexibility is slightly more important than roll flexibility with CG being the most significant step (Figure 7d). The major differences between the d(T_4A_4) and d(A_4T_4) summaries are clearly located in the 3-step roll and groove parameters that are important for defining the usual properties of d(A_4T_4).

T-tests are incorporated into Profile Manager, so that sequences can be easily compared to determine whether their general levels of importance (/Zscore/ distributions) are equivalent. A useful rule[19] is that when the degrees of freedom (DOF) is much greater than 50, the distributions are significantly different if the

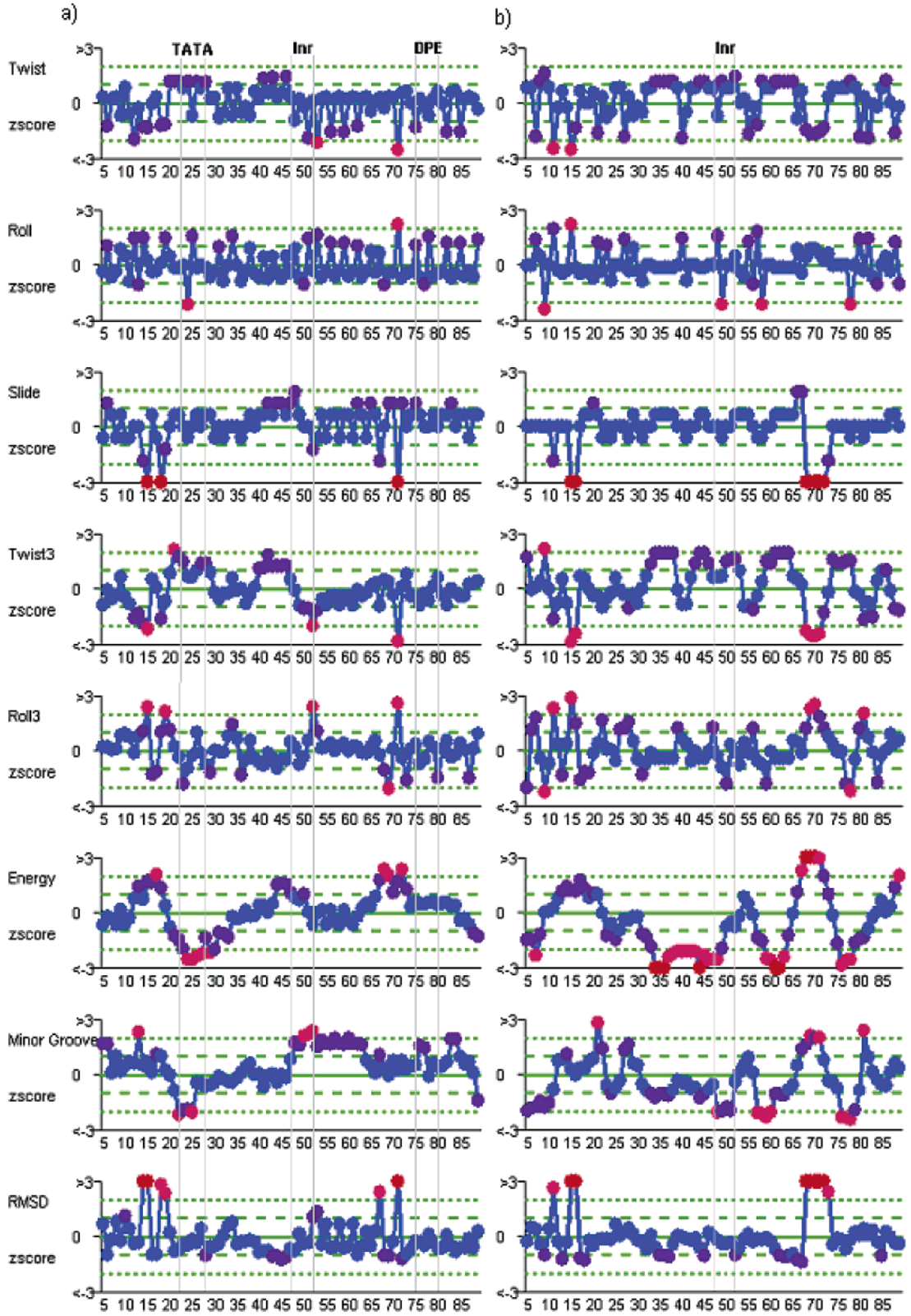

Figure 8. Structure profiles of two Drosophila promoters. TATA = TATA-box, Inr = initiator, and DPE = downstream promoter element. The experimentally determined transcription start site is at position 47: (a) the Ald promoter and (b) the 4f-rnp promoter.

magnitude of T is greater than 2.58

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad \mathrm{DOF} = \left\{ \frac{(\sigma_1^2/n_1 + \sigma_2^2/n_2)^2}{\frac{(\sigma_1^2/n_1)^2}{n_1 + 1} + \frac{(\sigma_2^2/n_2)^2}{n_2 + 1}} \right\} - 2 \tag{4}$$

where $\bar{x}_1$, σ_1^2, and n_1 are the mean, variance, and sample size of sequence 1's /Zscore/ distribution and likewise for $\bar{x}_2$, σ_2^2, and n_2 for sequence 2.

When comparing d(A_4T_4) to d(T_4A_4) in the above manner, the /Zscore/distributions are found to be significantly different with T being 3.54 and DOF being 928. T's positive value confirms that d(A_4T_4) is significantly more important

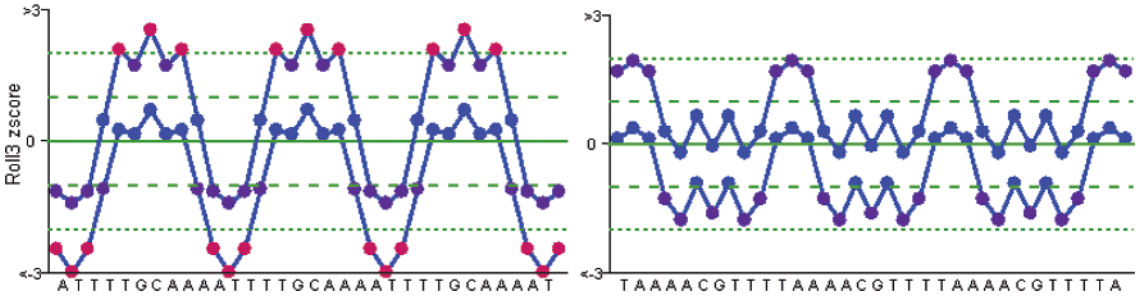

Figure 9. Structural 3-step roll tendencies of d(A_4T_4), (left) and d(T_4A_4), (right) using a probability cutoff of 0.75.

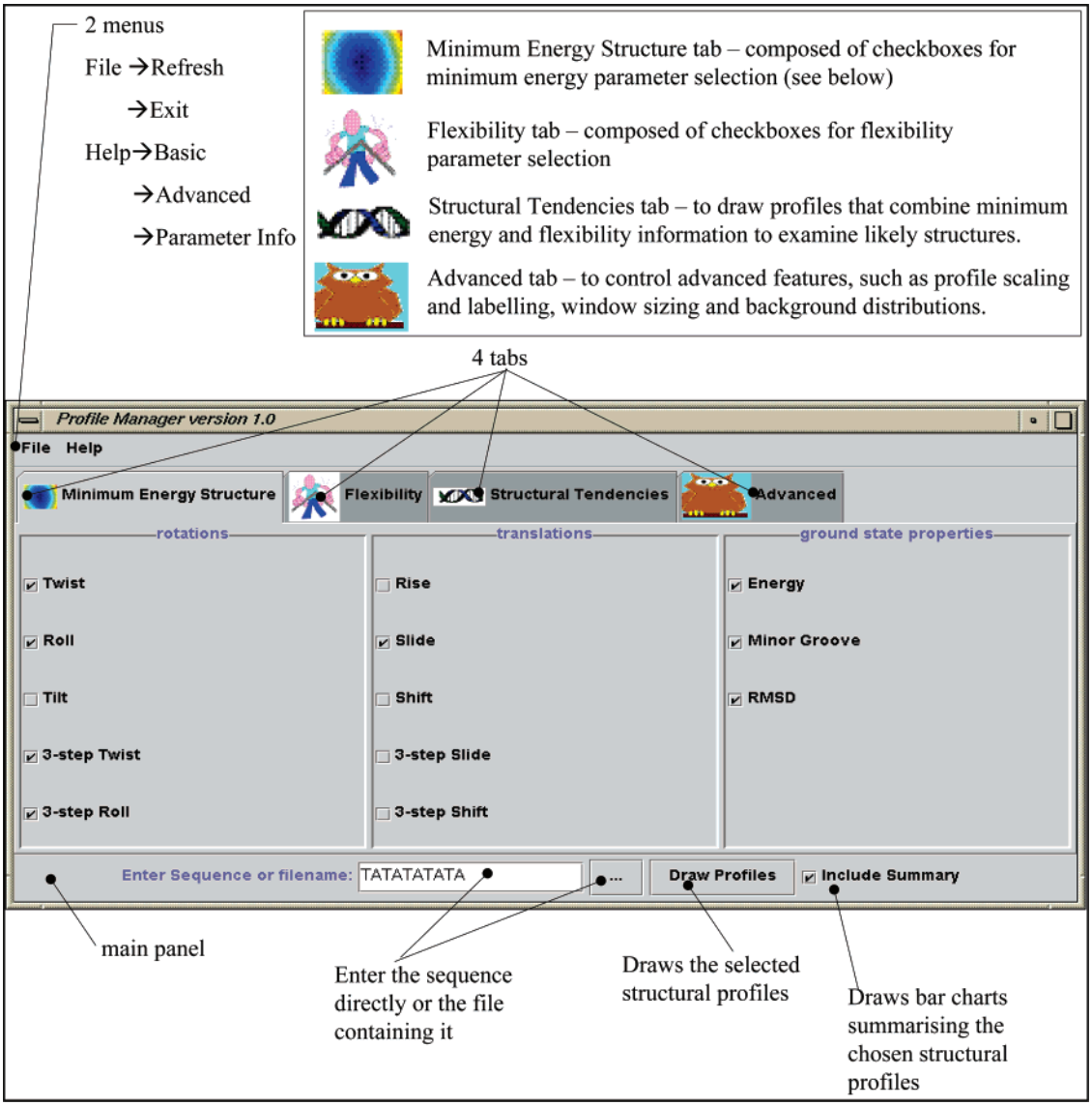

Figure 10. Profile Manager GUI and the Minimum Energy Structure tab.

than d(T_4A_4), since its Zscore magnitudes are generally higher. We can interpret this as meaning that the 3D structure of d(A_4T_4) is more extreme than that of d(T_4A_4).

ANALYSIS OF LONGER SEQUENCES

To illustrate the application of these methods to longer sequences, we have selected two sequences from the Drosophila Core Promoter Database[20] which contains 205 Drosophila Melangaster (fruit fly) promoters aligned by their experimentally determined transcription start sites. Figure 8 shows profiles for ald[21] and 4f-rnp.[22] Ald has a downstream promoter element (DPE) and a TATA box (TATA), whereas 4f-rnp possesses neither of these elements. Nevertheless, two important structural features can be seen on both the promoters when focusing upon the common red areas. Both ald and 4f-rnp have a low slide, low 3-step twist, high RMSD, and fluctuations in 3-step roll at positions 15 and 70. High energy is also a common feature at position 70. These patterns are clear when the promoters are aligned by their experimental transcription start sites, suggesting that structural features such as this could be useful in promoter recognition across different classes of promoters.

STRUCTURAL TENDENCIES

DNA is a rather flexible molecule, and the flexibility is sequence-dependent as discussed above. To allow for this feature in comparing structural profiles, the more loosely defined structural tendencies of a sequence can be described by combining minimum energy structure with flexibility. Structural probability profiles describe the range of roll or twist values that an octamer populates with a specified probability. The flexibility parameters are used to calculate the Boltzmann populations of states away from the energy minimum value in one degree increments. The upper and lower structural boundaries for a particular octamer are determined by starting at the energy minimum value and making one degree adjustments to the boundaries until the desired probability has been reached. A structural tendencies profile is drawn with a lower and upper curve, representing the structural boundaries. The space between the curves is the populated structural space. The 3-step roll structural tendencies are illustrated for the $d(A_4T_4)$ and $d(T_4A_4)$ sequences in Figure 9. A periodic roll curve is found for both sequences, but the profiles are quite different. For $d(A_4T_4)$, the linear conformation (Roll3 = 0 at all base steps) is not accessible, and the structure is always bent to some extent. Although $d(T_4A_4)$ shows some periodicity in the extreme conformations that are accessible, it can also access a perfectly linear conformation (Roll3 = 0 at all steps). Thus we would expect $d(T_4A_4)$ to exhibit a dynamic bending motion, whereas $d(A_4T_4)$ is more highly curved and permanently bent.

PROFILE MANAGER

Profile Manager is an application currently under development that aims to automate profile generation in an efficient user-friendly environment. The graphical user interface uses a tabbed window to categorize and hide advanced features, welcoming users with a simple lowest level of control (Figure 10). The four tabs refer to three profile categories (minimum energy structure, flexibility, and structural tendencies) plus an advanced tab. Separation of the parameters across a tabbed panel is essential, as it would be overwhelming to present them together, and a user may only be interested in one profile type at a given time.

Components belonging to the main panel represent frequent actions that can be seen regardless of the current active tab. The sequence or the name of the file containing the sequence can be entered into the text box. Alternatively the [...] button can be used to select the file from directory listings. The [Draw Profiles] button gets the profiles with optional summary bar charts specified by the Include Summary checkbox. The minimum energy structure parameters have been split into three groups: rotations, translations, and ground-state properties. This grouping and vertical alignment of 10 or fewer checkboxes makes the user choices easier to digest.[23]

CONCLUSION

Profile Manager v.1 is a valuable visualization tool for the analysis of DNA structure deduced from sequence. Application of single sequence queries to the A-tract phenomenon clearly illustrates the structural findings found in NMR structures.[18] Differences between the minor groove and roll3 profiles of $d(A_4T_4)$ and $d(T_4A_4)$ are striking and closely mirror the experimental results. The roll3 profiles show that $d(A_4T_4)$ bends due to a periodic roll pattern. The potential for the application of this approach in recognizing common structural features in functionally important DNA sequences has been illustrated using two Drosophila promoters.

Minimum energy structure parameters can be combined with flexibility parameters to produce structural tendency profiles that provide a more dynamic representation of DNA as a flexible molecule. Summary charts illustrate how the general importance of a sequence varies with respect to the parameters or how the importance varies across a sequence for a combination of parameters. Significant differences in the importance of two sequences can be assessed by a *t*-test. Tools to deal with pattern recognition across multiple unaligned sequences are now being developed.

SOFTWARE

The software is available upon request to the corresponding author.

ACKNOWLEDGMENT

We thank the Biotechnology and Biological Sciences Research Council for support of this work and the Wolfson Foundation for provision of computing facilities.

REFERENCES AND NOTES

(1) Collins, F. S.; Morgan, M.; Patrinos, A. The human genome project: lessons from large-scale biology. *Science* **2003**, *300* (5617), 286−290.
(2) Collins, F. S.; Green, E. D.; Guttmacher, A. E.; Guyer, M. S. A vision for the future of genomics research. *Nature* **2003**, *422* (6934), 835−847.
(3) Frazier, M. E.; Johnson, G. M.; Thomassen, D. G.; Oliver, C. E.; Patrinos, A. Realizing the potential of the genome revolution: the genomes to life program. *Science* **2003**, *300* (5617), 290−293.
(4) Woolfe, A.; Goodson, M.; Goode, D. K.; Snell, P.; McEwen, G. K.; Vavouri, T.; Smith, S. F.; North, P. et al. Highly conserved noncoding sequences are associated with vertebrate development. *PLoS Biol.* **2005**, *3* (1), 116−130.
(5) Bejerano, G.; Pheasant, M.; Makunin, I.; Stephen, S.; Kent, W. J.; Mattick, J. S.; Haussler, D. Ultraconserved elements in the human genome. *Science* **2004**, *304* (5675), 1321−1325.
(6) Johnston, M.; Stormo, G. D. Evolution: heirlooms in the attic. *Science* **2003**, *302* (5647), 997−999.
(7) Randic, M.; Vracko, M. On the similarity of DNA primary sequences. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (3), 599−606.
(8) Randic, M.; Vracko, M.; Nandy, A.; Basak, S. C. On 3-D graphical representation of DNA primary sequences and their numerical characterization. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (5), 1235−44.
(9) Zhang, C. T.; Zhang, R.; Ou, H. Y. The Z curve database: a graphic representation of genome sequences. *Bioinformatics* **2003**, *19* (5), 593−9.
(10) Gardiner, E. J.; Hunter, C. A.; Lu, X. J.; Willett, P. A structural similarity analysis of double-helical DNA. *J. Mol. Biol.* **2004**, *343* (4), 879−89.
(11) Gardiner, E. J.; Hunter, C. A.; Packer, M. J.; Palmer, D. S.; Willett, P. Sequence-dependent DNA structure: a database of octamer structural parameters. *J. Mol. Biol.* **2003**, *332* (5), 1025−1035.
(12) Koudelka, G. B.; Harbury, P.; Harrison, S. C.; Ptashne, M. DNA twisting and the affinity of bacteriophage-434 operator for bacteriophage-434 repressor. *PNAS* **1988**, *85* (13), 4633−4637.
(13) Rice, P. A.; Yang, S. W.; Mizuuchi, K.; Nash, H. A. Crystal structure of an IHF-DNA complex: a protein-induced DNA u-turn. *Cell* **1996**, *87* (7), 1295−1306.
(14) Vlahovicek, K.; Kajan, L.; Pongor, S. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* **2003**, *31* (13), 3686−7.

(15) Patterton, H. G.; Graves, S. DNAssist: the integrated editing and analysis of molecular biology sequences in windows. *Bioinformatics* **2000**, *16* (7), 652−3.
(16) Packer, M. J.; Hunter, C. A. Sequence-structure relationships in DNA oligomers: a computational approach. *J. Am. Chem. Soc.* **2001**, *123* (30), 7399−7406.
(17) Calladine, C. R.; Drew, H. R. *Understanding DNA*, 2nd ed.; Academic Press: London, 2002.
(18) Stefl, R.; Wu, H.; Ravindranathan, S.; Sklenar, V.; Feigon, J. DNA A-tract bending in three dimensions: solving the dA4T4 vs dT4A4 conundrum. *PNAS USA* **2004**, *101* (5), 1177−1182.
(19) Miller, J. C.; Miller, J. N. *Statistics for analytical chemistry*, 3rd ed.; Ellis Horwood Limited: Chichester, W. Sussex, 1994.
(20) Kutach, A. K.; Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol. Cell Biol.* **2000**, *20* (13), 4754−4764.
(21) Shaw-Lee, R.; Lissemore, J. L.; Sullivan, D. T.; Tolan, D. R. Alternative splicing of fructose 1,6-bisphosphate aldolase transcripts in Drosophila melanogaster predicts three isozymes. *J. Biol. Chem.* **1992**, *267* (6), 3959−3967.
(22) Petschek, J. P.; Scheckelhoff, M. R.; Mermer, M. J.; Vaughn, J. C. RNA editing and alternative splicing generate mRNA transcript diversity from the Drosophila 4f-rnp locus. *Gene* **1997**, *204*, 267−276.
(23) Weinschenk, S.; Jamar, P.; Yeo, S. C. *GUI design essentials for Windows 95, Windows 3.1 World Wide Web*; ed.; John Wiley & Sons: 1997.

CI050385A