Thomas Vigdal
Biochem 218
Final
June 6, 2002

INSERTION SITE SIMILARITIES IN THE TC1/MARINER ELEMENT FAMILY

Abstract

We examined insertion site profiles generated from three studies of Tc1/mariner elements: Tc1 (genomic copies and experimentally induced insertions (van Luenen and Plasterk, 1994)), Tc3 (van Luenen and Plasterk, 1994) and Himar1 (Lampe et al., 1998). Through this comparison we found that bendability, protein-induced deformability and A-philicity are the most significant for determining insertion site preference. We further examined Tn5 and Ty1 insertion sites and found that Tn5 shares similarities with the Tc1/mariner elements whereas Ty1 insertions do not. Therefore, we predict that a large amount of the Tc1/mariner elements', and possibly other DNA transposons, target site specificity relies on unusual DNA structure in the area of insertion.

Introduction

Secondary structural features influence integration of DNA transposons (Kuduvalli et al., 2001; Liao et al., 2000), retroviruses (Katz et al., 1998; Muller and Varmus, 1994; Pryciak and Varmus, 1992) and retroelements (Cost and Boeke, 1998; Jurka et al., 1998). The insertional specificity is believed to exist because DNA forms an unusual or perturbed structure that allows better recognition by the transposition complex (Cost and Boeke, 1998; Katz et al., 1998; Kuduvalli et al., 2001). The recognition and binding of such structure by a transposing complex has been proposed to be the first step in transposition. In the second step, the transposase complex searches the immediate area

of bound DNA for the best insertion site based on primary DNA sequence (Davies and Hutchison, 1995; Goryshin et al., 1998).

Four structural properties of DNA examined in this paper include B-DNA twist (Gorin et al., 1995), A-philicity (Ivanov and Minchenkova, 1995; Lu et al., 2000), DNA bending (Brukner et al., 1995) and protein-induced deformability (Olson et al., 1998). B-DNA twisting or unwinding affects the ability of molecules to interact within the grooves of the DNA. These interactions allow DNA to serve as areas of binding for proteins (Gorin et al., 1995). A-philicity represents the propensity of DNA to form an A-DNA like double helix (Ivanov and Minchenkova, 1995). A-DNA has a wide and shallow minor groove that is believed to give proteins easier access to form hydrogen bonds with bases within the DNA helix. Along with A-philicity, DNA bending can lead to changes in the width and depths of the major and minor groove, again affecting a proteins access to bases within the DNA helix (Brukner et al., 1995). Protein-induced deformability measures the impact of protein binding on the topology of DNA, which in turn impacts the binding of other proteins or the action of the protein currently bound (Olson et al., 1998). Combining all of the structural information described above, one concept prevails regarding the biological role of DNA structure: spatial optimum. Spatial optimum, as defined here, refers to the DNA being in a position that requires the least amount of energy for a protein to bind and act on that particular sequence.

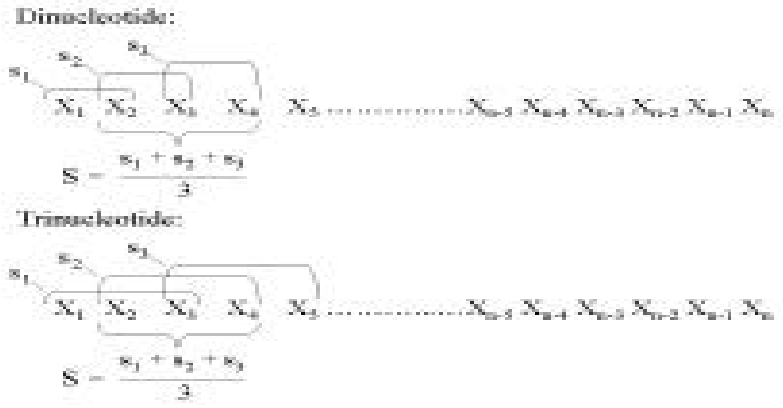
The Tc1/mariner transposon superfamily has been identified in a large range of organisms (Plasterk et al., 1999). Most Tc1/mariner elements insert into a TA dinucleotide giving the transposon an inherent specificity, yet not every TA has the same probability of insertion (van Luenen and Plasterk, 1994). The current hypothesis

explaining the TA bias is that a structural or sequence specific context exists which effectively recruits the Tc1/mariner transposase complex. Two members of the Tc1/mariner element family have already been implicated as having a structural preference for insertion sites other than the canonical TA (Lampe et al., 1998; van Pouderoyen et al., 1997). However the nature of these structural determinants and their relationship to the insertion site preferences of other Tc1/mariner transposons is unknown. In order to determine these structural features we undertook a comprehensive structural analysis of previously studied insertion sites of Tc1/mariner elements, Tn5 and genomic insertions of the LTR retrotransposon Ty1 applying new structural measures to them and comparing them globally.

Materials and Methods

Structural analyses

The computational structural predictions for all sequences in this study were done using a three base pair sliding window. The values of B-DNA twist, A-philicity, DNA bend and protein-induced deformability were given by previous studies of DNA crystal structure. The values we used in our predictions are available at http://www.fruitfly.org/~guochun/pins.html. The three base pair sliding window was performed as shown below.



For B-DNA twist, A-philicity and protein-induced deformability only dinucleotide values were available for determining the sequences' structure, whereas for bendability, trinucleotide values were used.

DNA sequence at the site of insertion was compared to control DNA sequences. For Tc1 and Tc3 insertions into *gpa*-2, non-insertion sites served as the control. For analysis of all other insertions, the control DNA sequences were randomly selected from genome sequences of the relevant host organisms or the plasmid targets. In order to measure if differences between the controls and the insertion site features were significant, we used MANOVA (multivariate analysis of variance). For doing our statistical analyses, we used commercial software developed by SAS Institute Inc. We chose a 90% confidence level as the cutoff for determining if a base pair was significantly different from the control data.

Results

Insertion site preference of Tc1/mariner transposons

We chose to look at insertion sites of related Tc1/mariner elements: Tc1 and Tc3 from *C. elegans*, and Himar1 from *Haematobia irritans*. These three transposons insert into a canonical TA target site. Four data sets from these elements were analyzed. The first

dataset was generated using a prototype Tc1 element (Genbank accession: X01005.1) in comparison against the *C. elegans* genome, with the 19 most similar Tc1 genomic insertions analyzed. For comparison, we also took random genomic *C. elegans* DNA sequences from chromosome I and aligned them at a TA. A second set of 65 Himar1 insertions generated by in vitro transposition into plasmid DNA (Lampe et al., 1998) were analyzed, and as a control we chose random DNA sequences from the target plasmid and aligned the sequences at a TA. In addition, 167 Tc1 and 95 Tc3 insertions into the *C. elegans gpa*-2 gene were analyzed (van Luenen and Plasterk, 1994). For a control, all non-targeted TA sites in *gpa*-2 were collected and aligned at the TA.

The profiles of the other elements are not just significantly different at the site of insertion, but are also different throughout much of the DNA profiled (Figure 1). This is shown in dramatic fashion for the profiles of Tc1 (Figure 1A) and Tc3 (Figure 1B). It appears that many of the 124 base pairs are significantly different for all properties analyzed. To a much lesser extent, the extended significance outside the insertion site is seen in Figure 1C and 1D. We suspect that this result is because the transposition complex does not select its insertion site just by local sequence topology, but also chooses its insertion site on a much broader sequence region.

In considering sequences at the insertion site, the profiles for protein-induced deformability and B-DNA twist are not consistent among the cases tested. For example, in Figure 1A, protein-induced deformability is significant above the control data at the site of insertion whereas in Figure 1B it is significant below the control data. In Figure 1A, B-DNA twist for insertion sites is lower than the random sequence, and in Figure 1C, the opposite is true with random sequence having a lower B-DNA twist. Because of this,

we believe that these three physical properties are unreliable in choosing insertion site preference. However, a uniform pattern is seen in the physical properties of both bendability and A-philicity among all elements. In the case of bendability, the DNA shows a more bent structure at insertion sites. In the case of A-philicity, insertion sites appear to have a propensity to form A-DNA either across the whole region of insertion (1A and 1C) or on one side of it (1B and 1D). This combined data suggests that insertion sites have a property beyond just the TA sequence that gives them a more bendable structure and a looser helical conformation.

Insertion site preference of non-Tc1/mariner elements

After our initial findings for the Tc1/mariner elements, we decided it may be useful to study unrelated elements to determine general features of insertion sites. We chose Tn5 because, like the Tc1/mariner elements, it uses a DDE transposase for movement. We used data from a previous study (Goryshin and Reznikoff, 1998) in which Tn5 insertions were generated *in vitro* into a plasmid. As a control we chose random DNA from the plasmid. The structural analysis of Tn5 (Figure 2A) showed a large region of significance in and around the site of insertion for all features analyzed. Two noticeable patterns around the site of insertion include an increase in bendability and an increase in the A-form DNA. In each of these cases bendability hits a maximum around the site of insertion and A-philicity hits a minimum. In the other three structural tests, the insertion site is also significant but, unlike bendability and A-philicity, they do not appear to be peaks or troughs in the 121 bp window.

Whereas the main focus of this paper has been on DNA transposons we thought it would be interesting if preferences for certain structural features were shared among LTR

retroelements. We gathered all of the insertion sites of the Ty1 elements in the complete *Saccharomyces cerevisiae* genome (Kim et al., 1998), and as a control we used random DNA taken from the *S. cerevisiae* genome. The structural properties of these insertions are different compared to the DNA transposons (Figure 2B). For example there is a peak in bending directly outside the site of insertion for Ty1 but it is not significantly different from the random DNA. A modest peak is also seen directly over the insertion site for B-DNA twist. Ty1 insertions also appear not to insert preferentially into areas of A-form DNA and also show a low amount of protein-induced deformability.

Discussion

Tc1/mariner elements share insertion site preferences

Many studies have focused on the insertion site preferences from different elements of this family (Korswagen et al., 1996; Lampe et al., 1998; van Luenen and Plasterk, 1994). Using the data gathered from two of these previous analyses for the Tc1/mariner elements, we thought it would be interesting to search for a global insertion pattern. Instead of focusing on primary sequence analysis, we focused on the search for structural motifs of insertion sites. Examining the insertion site profiles of other Tc1/mariner elements proved to show interesting conserved structural properties. The graphs in Figure 1 show that all structural properties examined (B-DNA twist, A-philicity, DNA bend and protein-induced deformability) have varying degrees of influence on targeting.

Both protein-induced deformability and A-philicity seem to be equally conserved (eight and seven base pairs out of fourteen respectively). Protein-induced deformability's role in insertion site preference is likely because of transposases direct interaction with DNA, which may allow transposase to conform the bound DNA into a spatial optimum

for strand transfer and subsequent integration of the transposon to take place. Spatial optimality may also play a role in why the Tc1/mariner elements insert into A-DNA like regions. If transposase prefers DNA in which it can make easy hydrogen bonding then A-DNA would likely be preferred. An interesting observation is that *Serratia* nuclease also prefers A-DNA like regions (Meiss et al., 1999). Whereas transposase may not perform like a nuclease on insertion, during excision it does cleave DNA. Perhaps this a distant ancestry between the two proteins in which transposase evolved to serve just one of it functions, namely excision.

These results taken together give an overall picture of the DNA being in a spatial optimum for transposase interaction. We believe it is a combination effect that recruits the transposase with a high degree of specificity. The conserved properties for the Tc1/mariner elements were protein-induced deformability, A-philicity and bendability; with bendability being the most conserved. All of these properties could easily be seen as creating a spatial optimum in the DNA for transposase to interact with. However, without direct experimental manipulation of insertion sites it is hard to conclude whether or not artificially created areas with these structural properties will be inserted into.

Our analysis of Tn5 gives rise to speculation that DNA transposons outside of the Tc1/mariner family may also recognize similar structural features. It is hard to make definitive conclusions in this regard since we only tested one other DNA transposon. Yet, it is interesting to note these similarities, and also that neighborhood effects may influence Tn5 site choice. The structural properties of Ty1 on the other hand, tend to suggest that LTR retrotransposons select insertion sites based on different structural

properties than DNA transposons. However, this observation is based on a small data set and this study would need to be extended to make any serious claims.

References

Brukner, I., Sanchez, R., Suck, D., and Pongor, S. (1995). Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. Embo J *14*, 1812-1818.

Cost, G. J., and Boeke, J. D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry *37*, 18081-18093.

Davies, C. J., and Hutchison, C. A., 3rd (1995). Insertion site specificity of the transposon Tn3. Nucleic Acids Res *23*, 507-514.

Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. J Mol Biol *247*, 34-48.

Goryshin, I. Y., Miller, J. A., Kil, Y. V., Lanzov, V. A., and Reznikoff, W. S. (1998). Tn5/IS50 target recognition. Proc Natl Acad Sci U S A *95*, 10716-10721.

Goryshin, I. Y., and Reznikoff, W. S. (1998). Tn5 in vitro transposition. J Biol Chem *273*, 7367-7374.

Ivanov, V. I., and Minchenkova, L. E. (1995). The A-form of DNA: in search of biological role. Mol Biol *28*, 780-788.

Jurka, J., Klonowski, P., and Trifonov, E. N. (1998). Mammalian retroposons integrate at kinkable DNA sites. J Biomol Struct Dyn *15*, 717-721.

Katz, R. A., Gravuer, K., and Skalka, A. M. (1998). A preferred target DNA structure for retroviral integrase in vitro. J Biol Chem *273*, 24190-24195.

Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A., and Voytas, D. F. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res *8*, 464-478.

Korswagen, H. C., Durbin, R. M., Smits, M. T., and Plasterk, R. H. (1996). Transposon Tc1-derived, sequence-tagged sites in Caenorhabditis elegans as markers for gene mapping. Proc Natl Acad Sci U S A *93*, 14680-14685.

Kuduvalli, P. N., Rao, J. E., and Craig, N. L. (2001). Target DNA structure plays a critical role in Tn7 transposition. Embo J *20*, 924-932.

Lampe, D. J., Grant, T. E., and Robertson, H. M. (1998). Factors affecting transposition of the Himar1 mariner transposon in vitro. Genetics *149*, 179-187.

Liao, G. C., Rehm, E. J., and Rubin, G. M. (2000). Insertion site preferences of the P transposable element in Drosophila melanogaster. Proc Natl Acad Sci U S A *97*, 3347-3351.

Lu, X. J., Shakked, Z., and Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. J Mol Biol *300*, 819-840.

Meiss, G., Gast, F. U., and Pingoud, A. M. (1999). The DNA/RNA non-specific Serratia nuclease prefers double-stranded A- form nucleic acids as substrates. J Mol Biol *288*, 377-390.

Muller, H. P., and Varmus, H. E. (1994). DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. Embo J *13*, 4704-4714.

Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A *95*, 11163-11168.

Plasterk, R. H., Izsvak, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet *15*, 326-332.

Pryciak, P. M., and Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell *69*, 769-780.

van Luenen, H. G., and Plasterk, R. H. (1994). Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans. Nucleic Acids Res *22*, 262-269.

van Pouderoyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H., and Sixma, T. K. (1997). Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA. Embo J *16*, 6044-6054.

Figure Legend

Figure 1: Physical properties of insertion sites. Tc1 elements were compared with random sequences for four different physical properties. All random sequences are aligned at the TA. Comparison of sequences were analyzed using MANOVA. If compared base pairs were had a confidence level higher than 90% they were marked in red. Base position 60 corresponds to the beginning of the TA insertion site. A.) Tc1 insertions in *gpa-2*; random sequences are taken from the TA sites in *gpa-2* where Tc1 did not insert. B.) Tc3 insertions in *gpa-2*; random sequences are taken from the TA sites in *gpa-2* where Tc3 did not insert. C.) Tc1 genomic insertions; random sequence is taken from chromosome one of *C. elegans*. D.) Himar1 insertions into the pBS plasmid; random sequence is taken from the pBS plasmid.

Figure 2: Structural analysis of Tn5 and Ty1. Comparison of sequences were analyzed using MANOVA. If compared base pairs had a confidence level higher than 90% they were marked with an asterisk. Base position 60 corresponds to the beginning of the insertion site. A.) Tn5 insertions into the pRETZL1 plasmid; random sequence taken from the pRETZL1 plasmid. B.) Ty1 genomic insertions; random sequence is taken from *S. cerevisiae*.

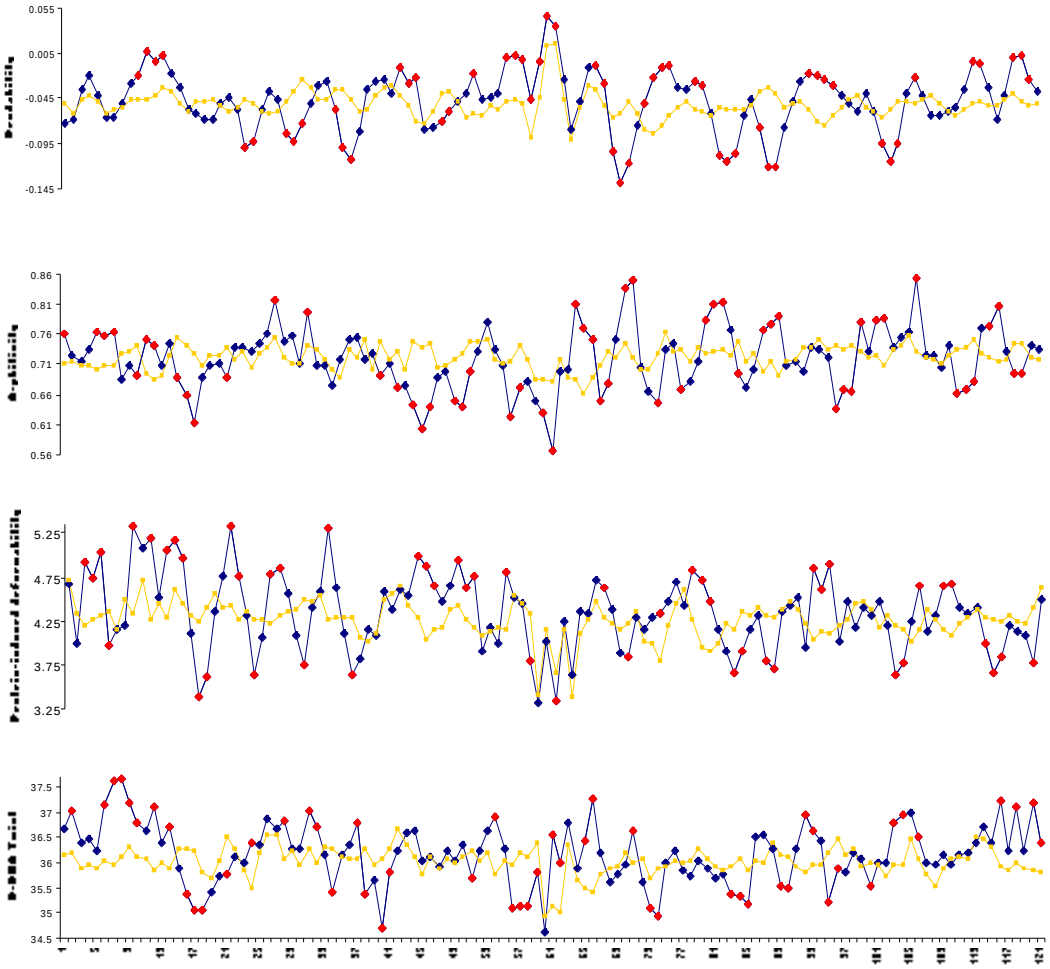
Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 1A Phyiscal properties of insertion sites.



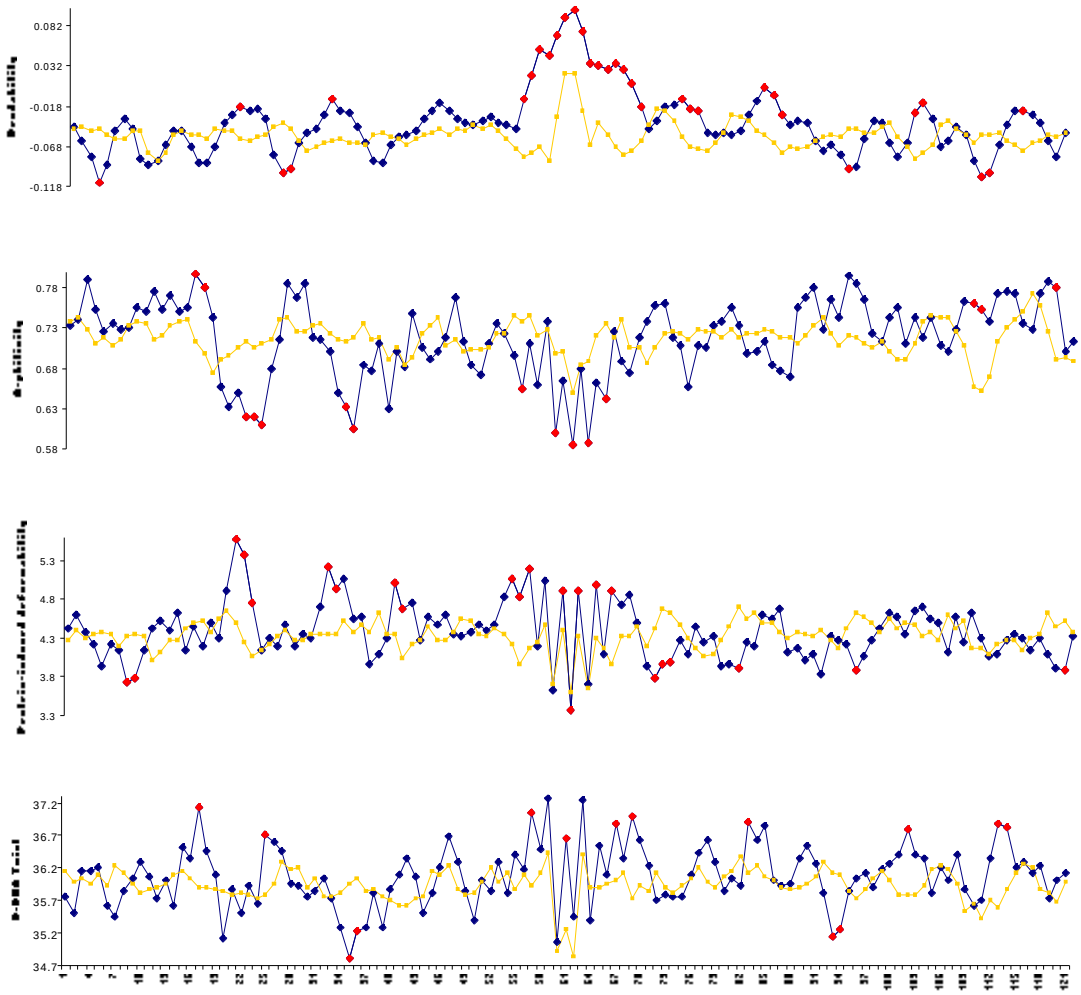
Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 1B Physical properties of insertion sites.



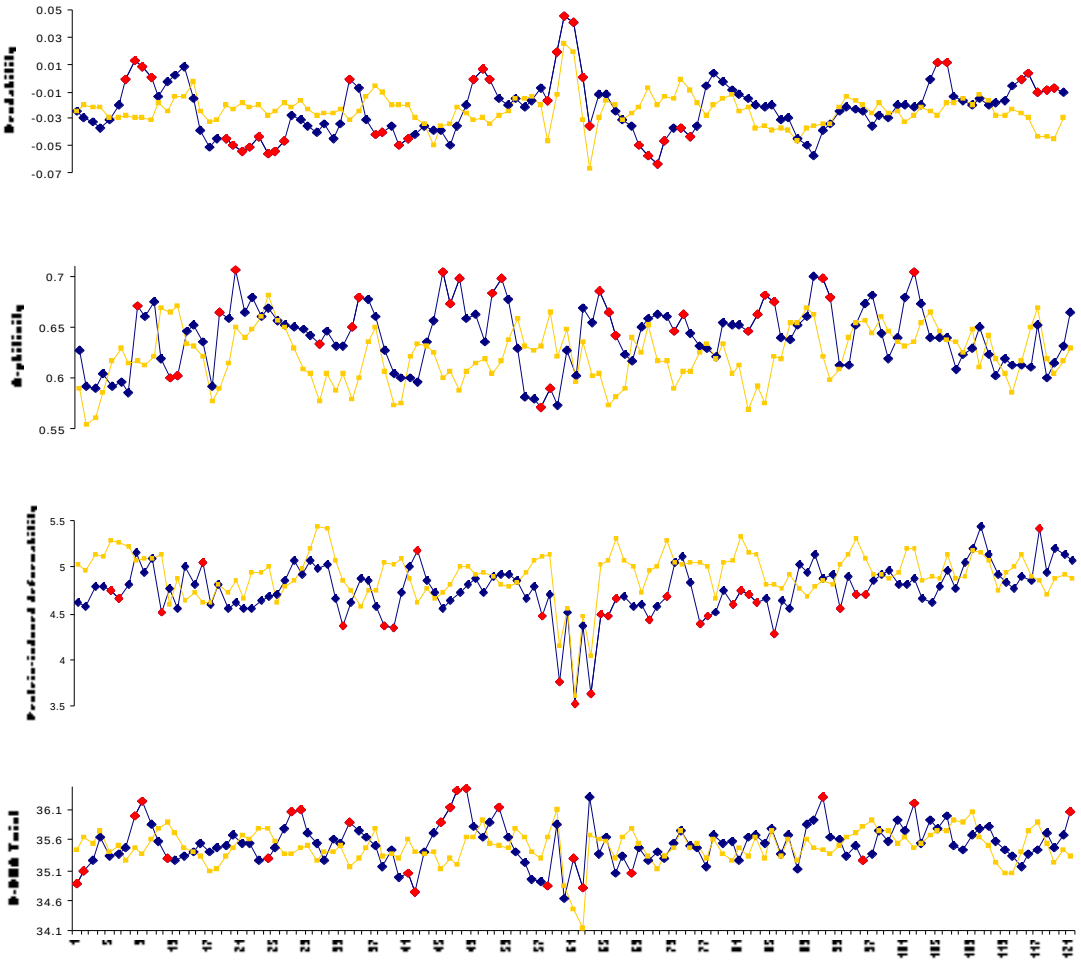
Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 1C Physical properties of insertion sites.



Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 1D Physical properties of insertion sites.

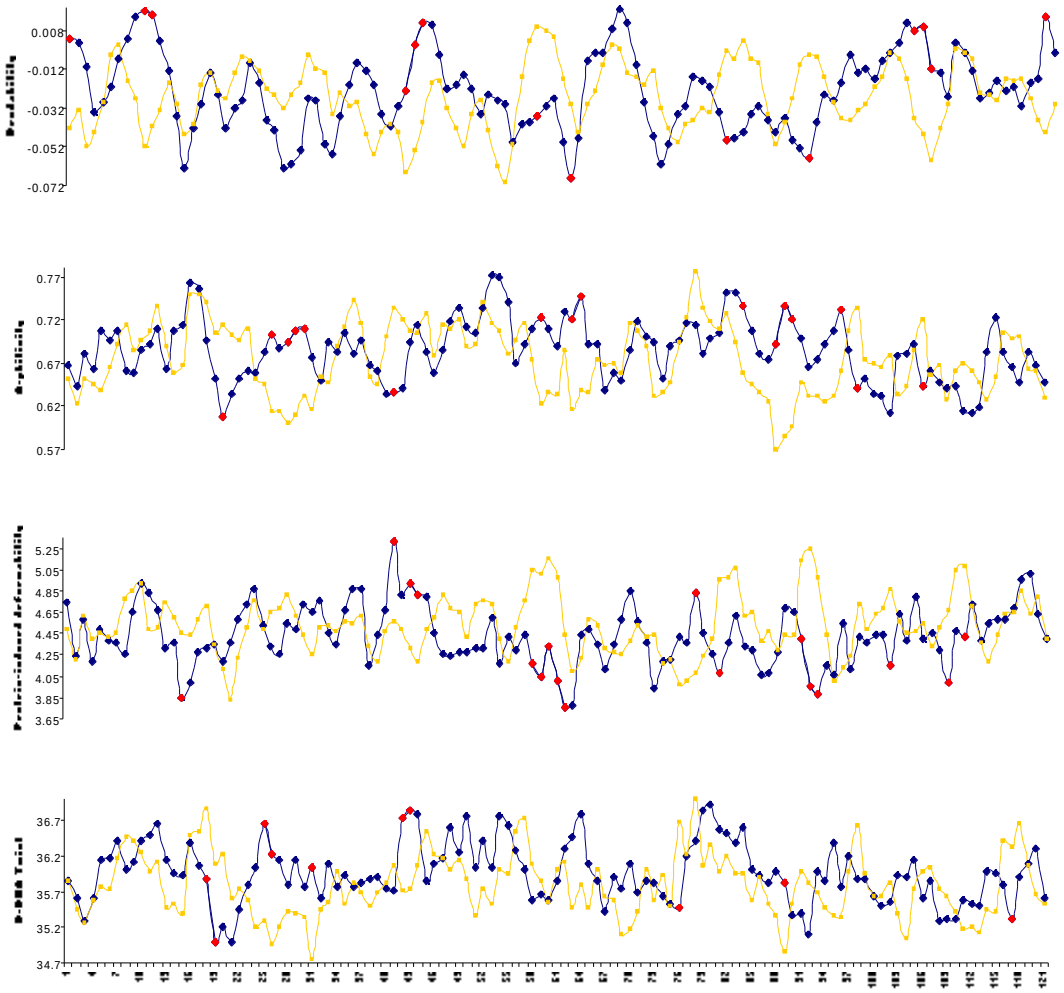
Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 2A Physical properties of insertion sites.



Legend:

Insertion DNA

Random DNA

Significant base pair

Figure 2B Physical properties of insertion sites.