

Assignment 5: Analysis of Protein/DNA Complexes with Insight and NDB: Canonical vs. Protein/Bound DNA, and DNA/Protein Interactions

In this assignment, we will study three nucleic acids structures, two of which have been crystallized with regulatory proteins: a nucleic acid dodecamer, a DNA oligomer bound to a prokaryotic protein in the helix-turn-helix (HTH) motif, and a DNA oligomer (known as the TATA-box sequence) bound to a eukaryotic transcription factor.

NDB ID code	Structure
• BDL078	DNA dodecamer
• PDR010	DNA (20 bp) bound to bacteriophage λ cI repressor
• PDT034	DNA (16 bp) bound to human TATA-Box Binding protein

1. **Structure Downloading.** Download each structure from the NDB, already explored in Assignment 4 (ndbserver.rutgers.edu:80/). Use the Archives, Atlas or Search entry points to the NDB; all are useful. In particular, the Atlas allows you to quickly view the structures. The Search entry point will be utilized later in this assignment.

Load each structure into Insight II. Separate the two complexes into DNA and Protein objects using the Modify/Unmerge command from either the **Biopolymer** or **Builder** modules. Both the DNA and Protein objects will be used later in this assignment. Use caution with the wildcard character * in selecting multiple atoms/residues/nucleotides/ proteins in Insight II. A quick test which you can use to test which object you've created with the Modify/Unmerge is to blank or blink the object with the Object/Blank or Object/Blink commands.

In order to create the B-DNA models for specified nucleic acid sequences (see below) and manipulate the downloaded structures (as in the Unmerge command), you will need to understand certain general parameters related to the structure as used in both the PDB file and Insight. These parameters refer to the DNA sequence, strand and residue labels, and so on. Explore the text in the downloaded files as well as in the structural displays. Information about relevant formatted lines in the coordinate files (such as SEQRES and ATOM) is available from the PDB web site (see Assignment 2).

2. **Canonical vs. Protein/Bound DNA Analysis.** Create an idealized B-DNA structure using the Nucleotide/Append command in the **Biopolymer** module corresponding to each of the three nucleic acid sequences in the above structures. The Nucleotide/Append command both creates new nucleic acid molecules and appends to existing molecules; when it is creat-

ing a new molecule, the Append Point is None³. You will need to use the **Nucleic_Acid/Cap** pulldown menu to replace the phosphate group with a hydroxyl group at the 5' ends of each strand.

Nucleotide/Append will create your B-DNA model, defining each strand as a separate object. Your downloaded DNAs, however, are each composed of a single object in Insight II. To properly superimpose the two structures (the task in the next part), you will need to **Modify/Merge** the two strands as one object.

Superimpose each B-DNA model upon its respective crystal structure from the NDB using the **Transform/Superimpose** pulldown menu. If your DNAs are each composed of a single object in Insight II, you will need to **Modify/Merge** the two strands as one object to properly superimpose the two structures. Use the **Heavy** option to avoid superpositioning of hydrogen atoms. After selecting the B-DNA model and the crystal structure in the selection boxes, you will need to click the **End Definition** box to **Execute the Superimpose** command. The root-mean-square deviation (RMSD) value will be printed at the bottom of your screen.

- (a) Record the RMSD values relative to idealized B-DNA for each superimposed model/structure by repeating this procedure for each crystal structure.
- (b) Use the **Search** entry point to the NDB to extract the following parameters for each base pair of the three structures: P (pseudorotation sugar pucker); the dihedral angles χ , α , β , γ , δ , ϵ , ζ ; and the helical parameters twist, tilt, roll, and propeller twist: Ω , τ , ρ , and ω , respectively.

For each conformational variable (excluding those from the end residues), calculate the average (μ) and standard deviation (σ) from the data per structure. Prepare a table in the following form:

Now discuss your results in terms of the differences noted between the protein-bound DNA and canonical B-DNA (which the BDL078 structure represents):

- (c) Which structure is most deformed from B-DNA? Which parameters display the largest changes from B-DNA (consider both μ and σ)? Based on these parameters, what is similar in the way the two complexes deform their recognition sites away from B-DNA (look for

³The **PDR010** structure has an overhanging base at each end, that is a base without a Watson-Crick partner on the other strand. (This procedure promotes crystal formation.) The recommended procedure for creating the overhanging base is to use **Nucleotide/Append** to create a 21-base-pair duplex of the correct sequence and then use **Nucleotide/Delete** command to delete one base from each strand prior to using **Nucleic_Acid/Cap**.

	BDL078	PDR010	PDT034
	μ σ	μ σ	μ σ
P			
χ			
α			
β			
γ			
δ			
ϵ			
ζ			
Ω			
τ			
ρ			
ω			
RMSD			

similarities in columns 3 and 4 above which are different from column 2)?

- (d) Are any of the changes observed localized to particular regions in the DNA? (Consider properties with μ values similar to B-DNA but large associated σ values). Plot one of these parameters as a function of position (base pair) along the DNA.
- (e) Generate a *side-by-side* picture of the three DNA structures. A recommended utility for this is the **File/Export_Plot** facility.
3. **Analysis of Interface Between Proteins and DNA.** Next, we will examine some of the interactions formed at the interface between the regulatory proteins and their DNA binding sites. Load the PDB files of each DNA/protein complex in turn and unmerge the DNA part (but leave the DNA and protein together in space).

The main tool used here is the **Subset/Interface** pulldown in the central **Viewer** module. This menu allows subsets to be defined in one molecule that satisfy a certain spatial relationship with respect to the other molecule. For example, we would like to use this menu to define subsets of atoms in the protein that are near functional groups of the DNA. A contour level of 3.5 Å is useful in this menu for defining interactions between non-hydrogen atoms, since it roughly corresponds to distances for strong interactions.

Open the **Subset/Interface** pulldown menu. You can define the **Subset Name** as you please. You can define the **Center of Subset** to be a specific functional group in DNA. For example, DNA:T:C5M refers to atom C5M of thymine's methyl group in the DNA. Define the **Search_Domain** to be the protein. The **Radius of Subset** should be set to the value 3.5.

For your reference, these are names of some DNA atoms:

- Phosphate groups: Atoms P, O1P, O2P
 - Thymine methyl groups: Atom C5M
 - Adenine amino groups: Atom N6
 - Pyrimidine carbonyls: Atom O2
 - Purine amines: Atom N3
- (a) Save listings of these subsets into output files. **Subset/List** is recommended for this task.

[Do not be alarmed if you get the error message “Invalid Comparison Object”; it simply means that the comparison could not be performed since no member of the set fulfilled the criteria. If an attempt is made to analyze all atoms **B** that are 3.5 Å from protein **A**, but all atoms of protein **A** are more than 3.51 Å from **B**, this error will occur.]

Combine the analyses of the two complexes (if you can) and construct histograms of the residue types in each subset. That is, from the listings of all residues within 3.5 Å of the atom groups above, count the number of times each residue appears (e.g., Methionine appears 60 times, Glutamate 3) and generate histogram plots as illustrated below; use the one-letter mnemonic for the amino acids. (You may also want to count the frequencies of residues grouped by type, like polar, hydrophobic, charged, etc.).

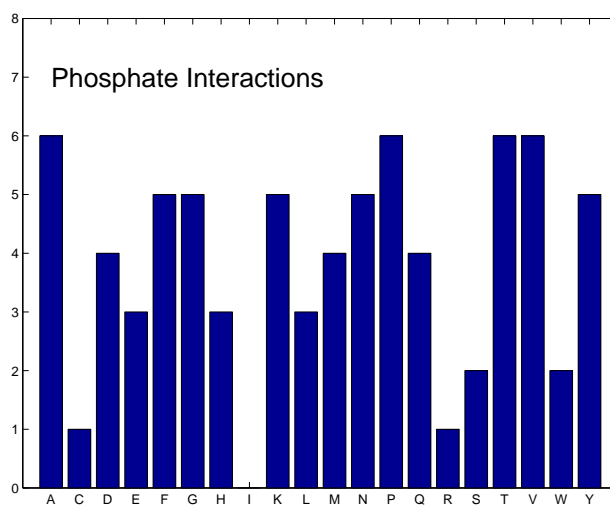


Figure D.1. Sample histogram for protein/DNA interaction analysis

- (b) Do you observe common patterns in the two complexes? Are certain amino acids likely to be found interacting with a particular functional group? What types of interactions are being formed between these nu-

cleic acid functional groups and the regulatory protein (e.g. attribute to each of the nucleic acid groups above a type of interaction such as hydrophobic, hydrogen bonding, electrostatic, intercalation/insertion motif, etc.)?

- (c) Is there anything unusual about the subsets formed between the proteins and the O2 carbonyl/N3 amine atoms?
- (d) Is there any relation between the interactions observed in these subsets and the deviations from canonical B-DNA structure observed above (i.e. how do the interactions you observe explain any of the parameter variances you diagnosed)?

Note: A trick to identify atoms/residues is via `Molecule` `Color` for assigning a color to an atom/residue. Other labeling tools such as `Molecule` `Render` and `Molecule` `Label` can similarly be used.

- (e) **Bonus Question⁴: BDL078 Homologues.** We have used the BDL078 structure as an example of B-DNA in the analysis above. However, sequence-dependent variations in local structure are also important. Therefore, a more sensitive analysis of free versus protein/bound DNA employs the same nucleotide sequence, both with and without bound proteins. There are few cases, however, in which high-resolution DNA structures are available in both the free and protein-bound states. Such analyses are illuminating and show how intrinsic DNA preferences are amplified in the DNA/protein complexes. See recent reports regarding the complex between DNA and the bovine papillomavirus E2 protein in D.M. Crothers (*Proc. Natl. Acad. Sci.* **95**:15163–15165, 1998) and H. Rozenberg *et al.* (*Proc. Natl. Acad. Sci.* **95**:15194–15199, 1998). Such analyses have not been done with our BDL078 sequence, but there are protein-DNA complexes in the NDB which are closely related to BDL078. This close relationship means that: (i) the related sequence has many similar or closely related residues to BDL078 (e.g., GGGAAAATTT is closely related to GGCATAACTT), and (ii) the protein would bind to BDL078 and this related sequence.

Determine which protein/DNA complexes these are, and briefly describe what these complexed proteins do.

Background Reading from Coursepack

- K. B. Lipkowitz, “Abuses of Molecular Mechanics. Pitfalls to Avoid”, *J. Chem. Educ.* **72**, 1070–1075 (1995).

⁴The correct solution will allow you to drop lowest homework grade in any assignment.

- S. Lifson, “Potential Energy Functions for Structural Molecular Biology”, in *Methods in Structural Molecular Biology*, pp. 359–385, D. B. Davies, W. Saenger, and S. S. Danyluk, Eds., Plenum Press, London (1981).