

Protein – DNA interface: Structural Analysis of Contact Patches and Triplets by Specific Contact Mapping.

Mounir Errami and Valentin Ilyin¹
Email: ilyin@neu.edu

Keywords: protein-DNA interface, interface patches, triplet, recognition, Delaunay Tessellation

The interaction between protein and nucleic acids are essential for many cell processes that regulate cell cycles and vital functions. Understanding the determinants of protein / DNA recognition would lead to valuable biological, medical, and biotechnological applications. The specificity between protein and DNA molecule is very complex and far from being understood given the significant variability of protein/DNA interfaces. While many studies did focus on a pair-wise specificity rules between single residues and nucleotides [1], only few approaches have taken into account the effect of a local environment comprising sets of residues and nucleotides. In this study we present an analysis of protein DNA interfaces where DNA–protein recognition is analyzed at a contact patch and triplet levels.

A patch is defined as all successive nucleotides on a double DNA strand that have at least one contact with protein residues. The patch also includes protein residues. There are two types of specific patches: 1) patches composed of contacts occurring on the major groove of the DNA helix (MAG) and 2) those composed of contacts occurring on the minor groove (MIG) (Figure 1). These specific patches are clearly separated by the non-specific interactions occurring between protein atoms and phosphate / sugar atoms from DNA. A triplet is composed of three consequent nucleotide pairs with corresponding contacting residues. The contact definition is based on Delaunay tessellation [2] patterns where each edge represents one contact.

First we present statistical information about interface patches obtained from 294 protein-DNA structure complexes. The patches were classified and clustered according to their double DNA pattern, highlighting the important variability among the current representation of recognized DNA patterns. A deeper analysis pointed out several cases where unrelated protein structures were able to recognize identical or very close DNA patterns illustrating the more general observation among which a same pattern of DNA can be recognized in different ways by different proteins showing the complexity and the challenging aspects of protein-DNA recognition mechanisms.

In order to normalize the interface variability we have use the triplet as a standard unit. These triplets were analyzed using their Delaunay tessellation patterns turned into contact matrixes (Figure 2). Using these matrixes, we show that it is possible to define three main classes of triplets according to the direct and indirect readout contributions and nature (Figure 3). These three populations are:

- 1) MIG35, a set of MIG triplets that count on average 35 specific contacts, in which the indirect readout mechanisms mainly consists of water mediated contact.
- 2) MAG60, a set of MAG triplets that count on average 60 specific contacts, where our observation would suggest that both water and DNA flexibility contribute to the indirect mechanisms in various proportions.
- 3) MIG70, a set of few MIG triplets that count on average 70 contacts. The indirect readout mechanism consists of DNA structure changes (severe binding) with a minor or inexistent role of water. Besides, the tightness of the interaction between residues and nucleotide is illustrated by analyzing the contribution of dual groove atoms.

¹ Northeastern University – BOSTON (MA)

We have characterized these three classes according to the contribution of direct readout and indirect mechanisms. We were also able to estimate the closeness of residues and nucleotides through the analysis of the contribution of water, and the number of contacts (edges) involving the dual groove atoms. These observation correlates with the function of DNA-protein complexes and is a side effect of the interaction consequences on DNA conformation.

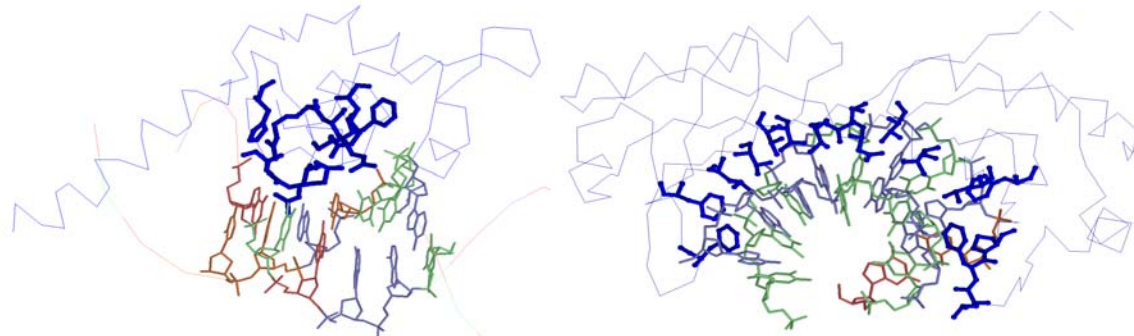


Figure 1: Left: Major Groove patch (pdb code 1kb6). Right: Minor groove patch (code 1ytb). Interface protein residues are colored in blue. Recognized nucleotides are colored as follows: A light blue, T green, C orange and G in red.

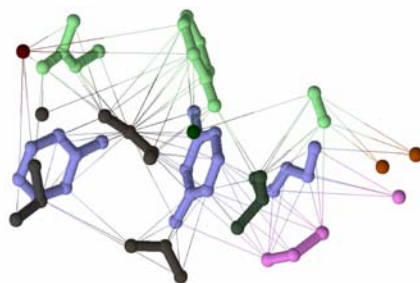


Figure 2: Delaunay tessellation pattern of a triplet (Adenine atoms are in blue, Thymine atoms in green, others are protein atoms).

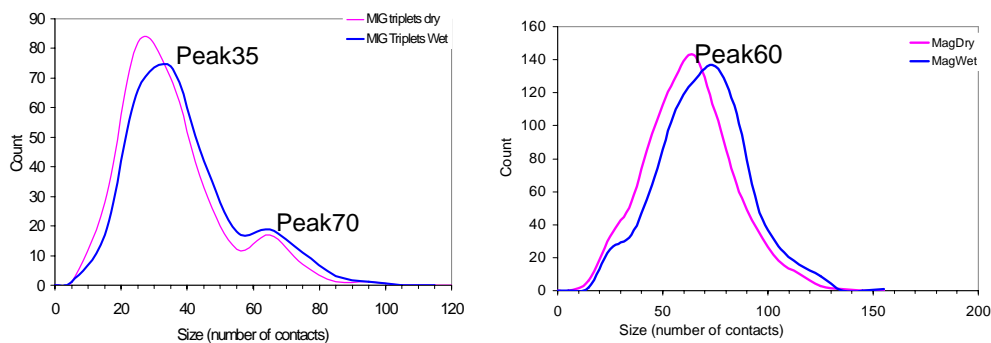


Figure 3: Left: peaks represent MIG35 and MIG70 population. Right Peak60 represent MAG60 population. Pink: triplets are considered without water mediated contacts. Blue: water mediated contact are added. Water mediates contact in MIG35 and MAG60 (a shift is observed), but seems absent from MIG70 (no shift is observed).

[1] Luscombe, N. M., R. A. Laskowski, and J. M. Thornton. "Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level." *Nucleic Acids Res.* 29.13 (2001): 2860-74.
 [2] Delaunay, B. "Sur la sphere vide." *Bull.Acad.Science USSR VII: Class.Sci.Mat.Nat.* (1934): 793-800.