From direct to indirect readout. DNA/Protein recognition by means of computational chemistry analysis.

Jiri Vondrasek,
Institute of Organic Chemistry and Biochemistry AS CR, v.v.i. ,
166 10 Prague 6, Czech Republic

The growing number of high-quality experimental (X-ray, NMR) structures of protein-DNA complexes has sufficient enough information to assess whether universal rules governing the DNA sequence recognition process apply. While previous studies have investigated the relative abundance of various modes of amino acid-base contacts (van der Waals contacts, hydrogen bonds), relatively little is known about the energetics of these noncovalent interactions. We have performed the first large-scale quantitative assessment of binding preferences in protein-DNA complexes by calculating the interaction energies in all 80 possible amino acid-DNA base combinations. We found that several mutual amino acid-base orientations featuring bidentate hydrogen bonds capable of unambiguous one-to-one recognition correspond to unique minima in the potential energy space of the amino acid-base pairs. A clustering algorithm revealed that these contacts form a spatially well-defined group offering relatively little conformational freedom. Various molecular mechanics force field and DFT-D ab initio calculations were performed, yielding similar results.

Consequently, representative pairs of amino acid side chains and nucleic acid bases extracted from available high-quality structures of protein DNA complexes were analyzed using a range of more accurate computational methods. CCSD(T)/CBS interaction energies were calculated for 272 pairs chosen from a large set of interactions and geometries. These reference interaction energies were used to test the MP2.5/CBS, MP2.X/ CBS, MP2-F12, DFT-D3, PM6, and Amber force field methods.

Regardless of the site of interaction, the minima found after full optimization in implicit solvent with high dielectric constant were close to the structures experimentally detected in protein-DNA complexes. According to DFT-SAPT analysis, the nature of noncovalent interactions strongly depends on the type of amino acid. The negatively charged sugar-phosphate backbone of DNA heavily influences the strength of interactions and must be included in the computational model, especially in the case of interactions with charged amino acids.