

## Structure and recognition of sheared tandem G x A base pairs associated with human centromere DNA sequence at atomic resolution

Y.G. Gao<sup>1</sup>, H. Robinson<sup>1</sup>, R. Sanishvili<sup>2</sup>, A. Joachimiak<sup>2</sup>, and A.H. Wang<sup>1</sup>

<sup>1</sup>*Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, IL 61801 USA*

<sup>2</sup>*Structural Biology Center, Argonne National Laboratory, 9700 S. Cass Ave., Argonne IL 60439 USA*

G x A mismatched base pairs are frequently found in nucleic acids. Human centromere DNA sequences contain unusual repeating motifs found in the human chromosome (e.g., (GAATG)<sub>n</sub> x (CATTC)<sub>n</sub>). The purine-rich strand of this repeating pentamer sequence forms duplex and hairpin structures with unusual stability. The high stability of these structures is contributed to by the "sheared" G x A base pairs, which present a novel recognition surface for ligands and proteins. We have solved the crystal structure of d(CCGAATGAGG) in which the centromere core sequence motif GAATG is embedded by the multiple-wavelength anomalous diffraction (MAD) method. Three crystal forms were refined to near-atomic resolution. The structures reveal the detailed conformation of tandem G x A base pairs whose unique hydrogen-bonding surface has interesting interactions with bases, hydrated magnesium ions, cobalt(III)hexaammine, spermine, and water molecules. The results are relevant to the understanding of the structure associated with the human centromere sequence in particular and G x A base pairs in nucleic acids (including RNA, like ribozyme) in general.