NMR structure of the rCAG repeat associated with Huntington's disease

Damian M VanEtten[1], Jonathan L. Chen[2], Matthew D. Disney[2], Ilyas Yildirim[2], and Dr Matthew A Fountain[3]
State University of New York at Fredonia
1- Department of Biology, 1 Department of Chemistry, 2- The Scripps Research Institute, Jupiter, FL, 3- Department of Chemistry and Biochemistry

Abstract

Huntington's disease is an incurable inherited autosomal-dominant neurodegenerative disorder caused by expanded CAG trinucleotide repeats in the huntingtin gene. Understanding the detailed structure of the $(CAG)_N$ repeats can help develop drugs that prevent or slow the progression of this disease.
We used NMR spectroscopy to determine the 3D structures of $(CAG)_2$ double repeat model and $(CAG)_3$ triple repeat models to understand the features of the CAG repeats containing A=A non-Watson-Crick base pairs. We collected and analyzed NMR spectra from 500 MHz and 700 MHz Bruker Avance NMR spectrometers and obtained structural restraints using SPARKY. These structural restraints were then used in AMBER 14 to generate structures that fit the NMR data. Details of the structure indicate that the two CAG repeats adopt similar 3D structures with GC base pairs on each side of the AA mismatch. The details of the structural determination process and of the CAG repeat structures will be presented.

Introduction

The gene for Huntington's disease is called huntingtin and is composed of multiple CAG repeats on chromosome 4 (1). More than 40 CAG repeats are needed for the disease to be fully penetrant (1). The large amounts of repeats cause the protein product of huntingtin to misfold which is the root of the neurodegenerative disorder (1). Understanding the structure of the CAG repeats will allow for accurate drug binding to inhibit the translation of the misfolded proteins (2).
NMR is a powerful tool for determining the structure of biological molecules. Multidimensional NMR techniques like the nuclear Overhauser Effect (NOE) can be used to measure the distance between protons that are less than 5A. Experiments in HDO allows us to determine the hydrogen bonding between base pairs.
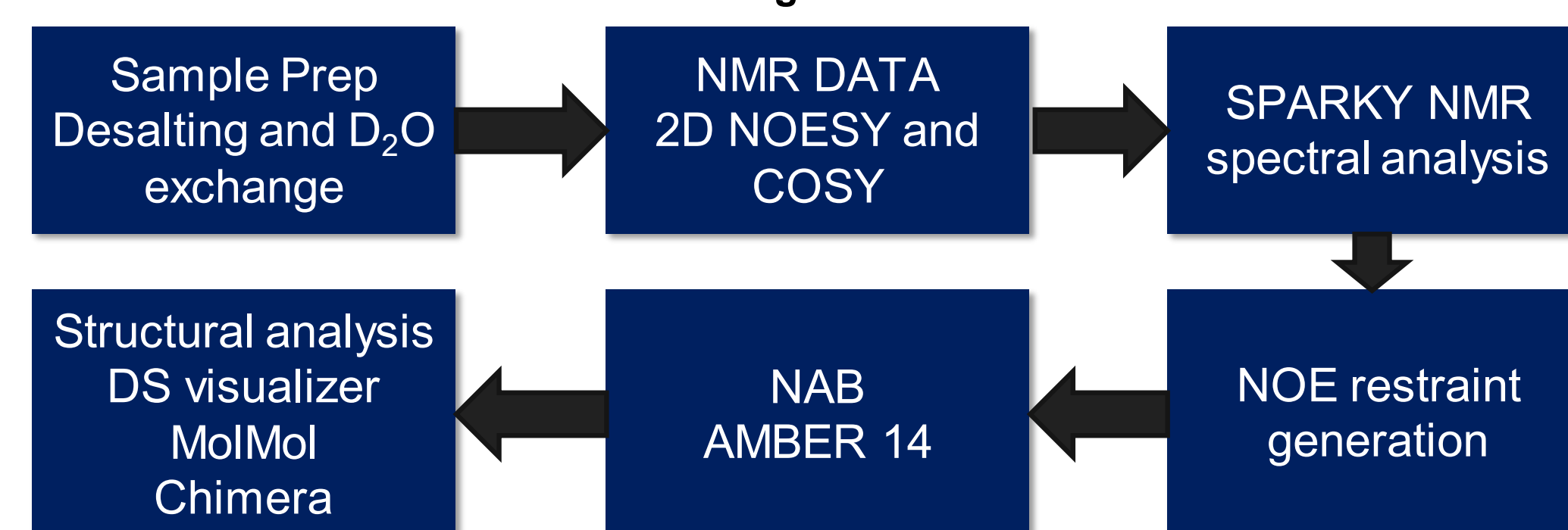
We have determined the structure of a double r(CAGCAG) and triple CAGCAGCAG repeat using NMR and molecular modeling. . A 15 nucleotide self complementary duplex containing the triple CAG was studied (Figure 1). Figure 1 also shows the base to H1' NOESY walk for the triple CAG construct base to H1' NOESY walks shows that the resonances for the CAG have very similar chemical shift values suggesting repetitive structural features.
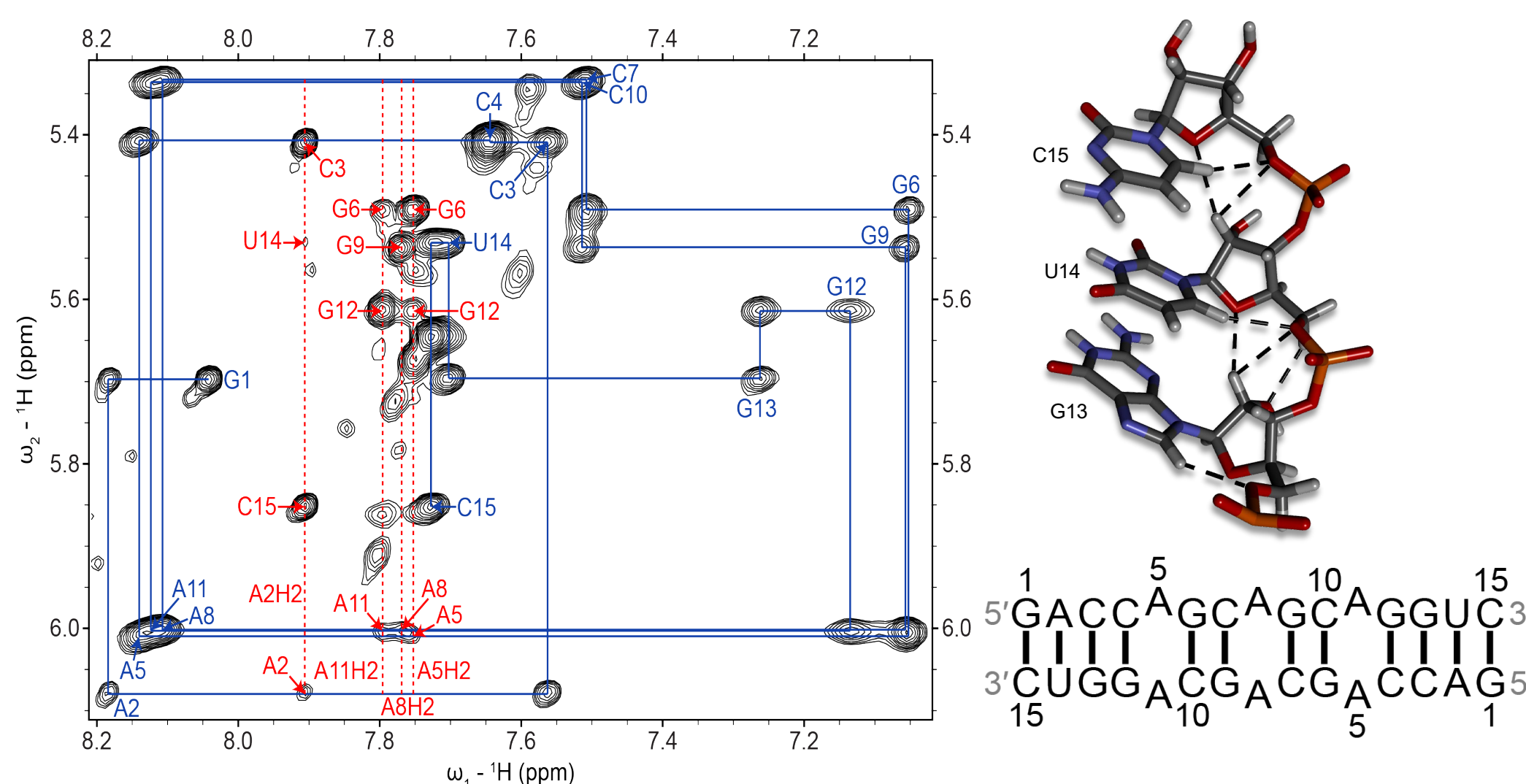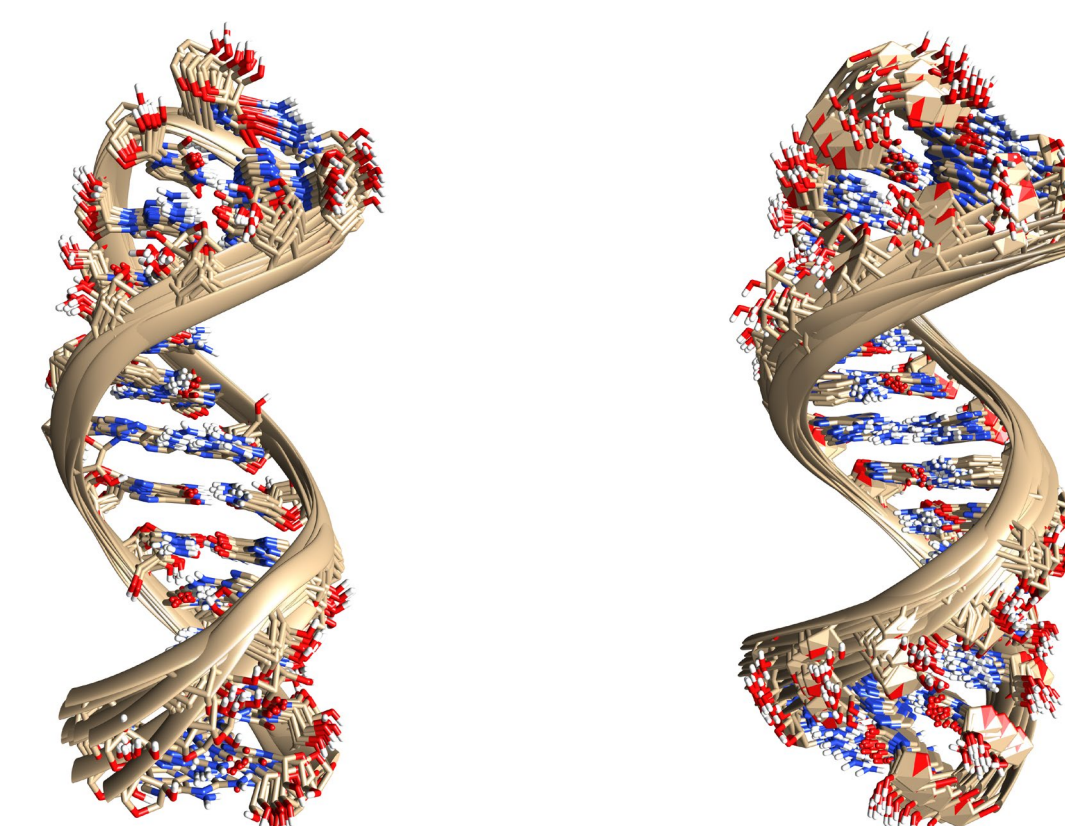
(Figure 1 $(CAG)_3$ NOESY Spectrum
This spectrum shows the proton walk from G1 to C15 in blue. The red vertical lines represent the individual adenosine H2 NOEs. Below the spectrum shows the $(CAG)_3$ sequence. The three CAG repeats are present along with their respective adenosine mismatches.

Methods

Figure 2

Sample Prep Desalting and D_2O exchange → NMR DATA 2D NOESY and COSY → SPARKY NMR spectral analysis → NOE restraint generation → NAB AMBER 14 → Structural analysis DS visualizer MolMol Chimera

Results

Figure 3 shows the results of 20 structural calculations for the (CAG)2 and (CAG3) and the structural results are in TABLE I. The RSMD values for the structures were low and the helical rise of the structures agreed with one another. While the RNA had undertwisting around the A=A mismatches, little stacking was observed for the A=A mismatches.

TABLE I	$r(3\times CAG)$
No. of restraints	
All distance restraints, including hydrogen bonds	254
All NOE restraints	220
Intraresidue	132
Sequential residues	78
Long range	10
Hydrogen bond	34
Dihedral restraints	140
RMSD of experimental restraints	
Distance (Å)	1.0×10^{-3}
Dihedral (deg)	1.5
RMSD of structures for heavy atoms (Å)	
All residues	1.34 ± 0.28
Helices (excluding AA loops)	1.38 ± 0.30
AA loops	0.93 ± 0.18

$(CAG)_2$ Twelve Molecule Cluster   **Figure 3** $(CAG)_3$ Twenty Molecule Cluster

Lowest Energy Structures
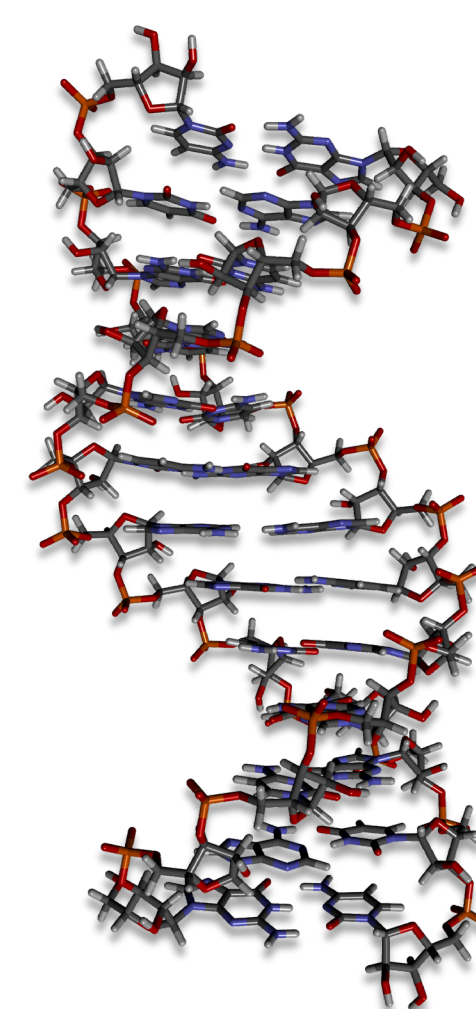
$(CAG)_2$ Structure
The structure to the left shows a representative structure from the CAG duplet. The twenty six nucleotide structure has a single adenosine mismatch between A7 and A20.
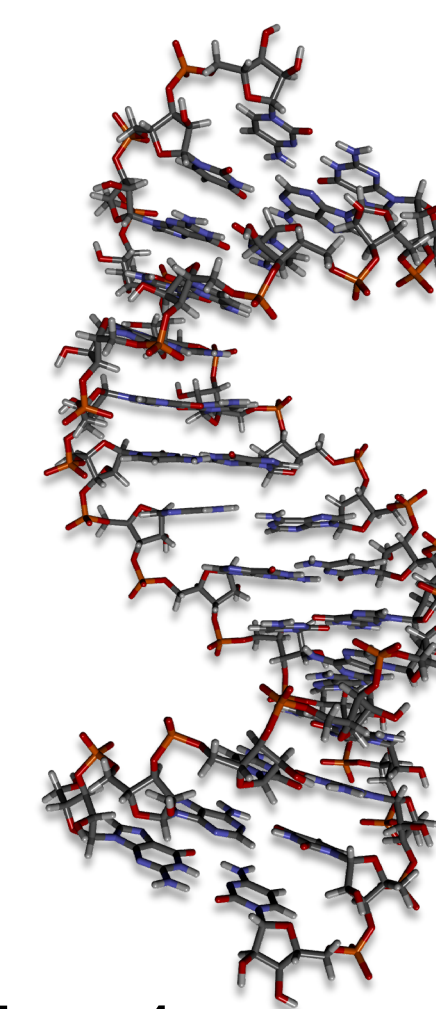
$(CAG)_3$ Structure
The CAG triplet to the left is a representative structure consisting of thirty nucleotides. There are three adenosine mismatches between A5=A26, A8=A23, and A11=A20.

Figure 4

Adenosine Mismatches
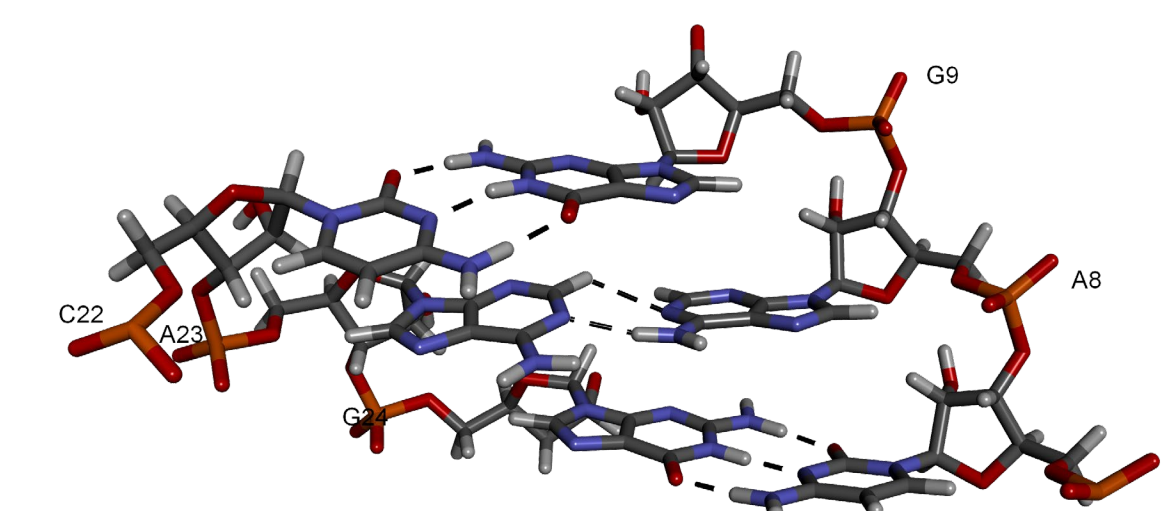
$(CAG)_3$ CAG
This CAG pair from the CAG triplet is from the center of the structure which shows the A8=A23 mismatch. The base pairing are represented by the black lines.
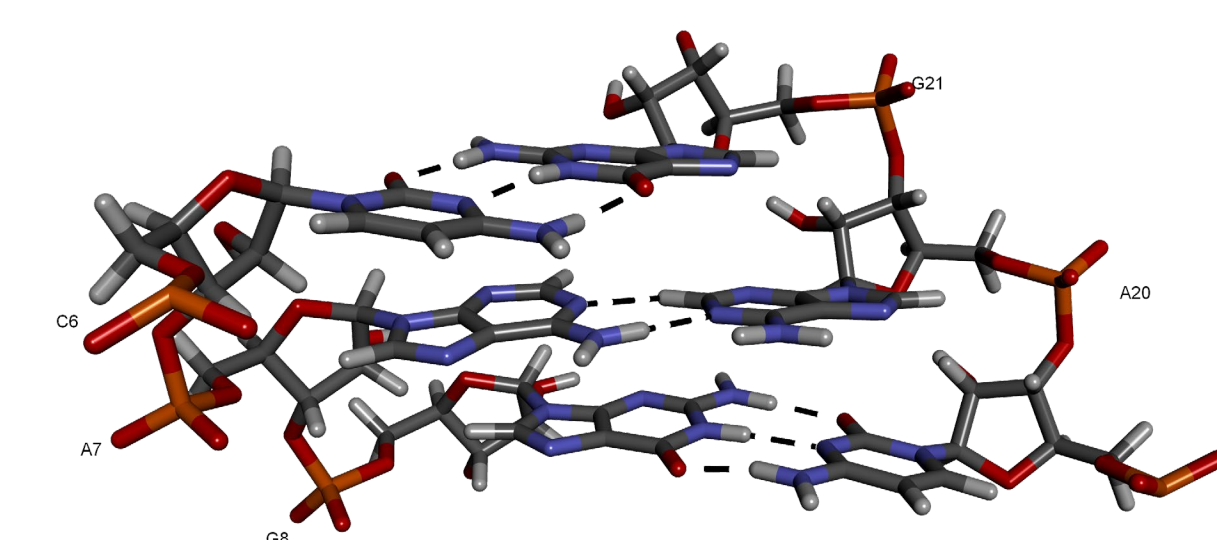
$(CAG)_2$ CAG
This CAG pair from the duplet shows the base pairing in black. The A7=A20 mismatch is located in the center of the CAG pair.

Figure 5

Adenosine Mismatch Helical Twist

$(CAG)_3$ Adenosine Mismatch Stacking Diagram
This stacking diagram shows the helical twist of an adenosine mismatch compared to an adjacent GC pair. There is little stacking shown despite the adenosine mismatching.
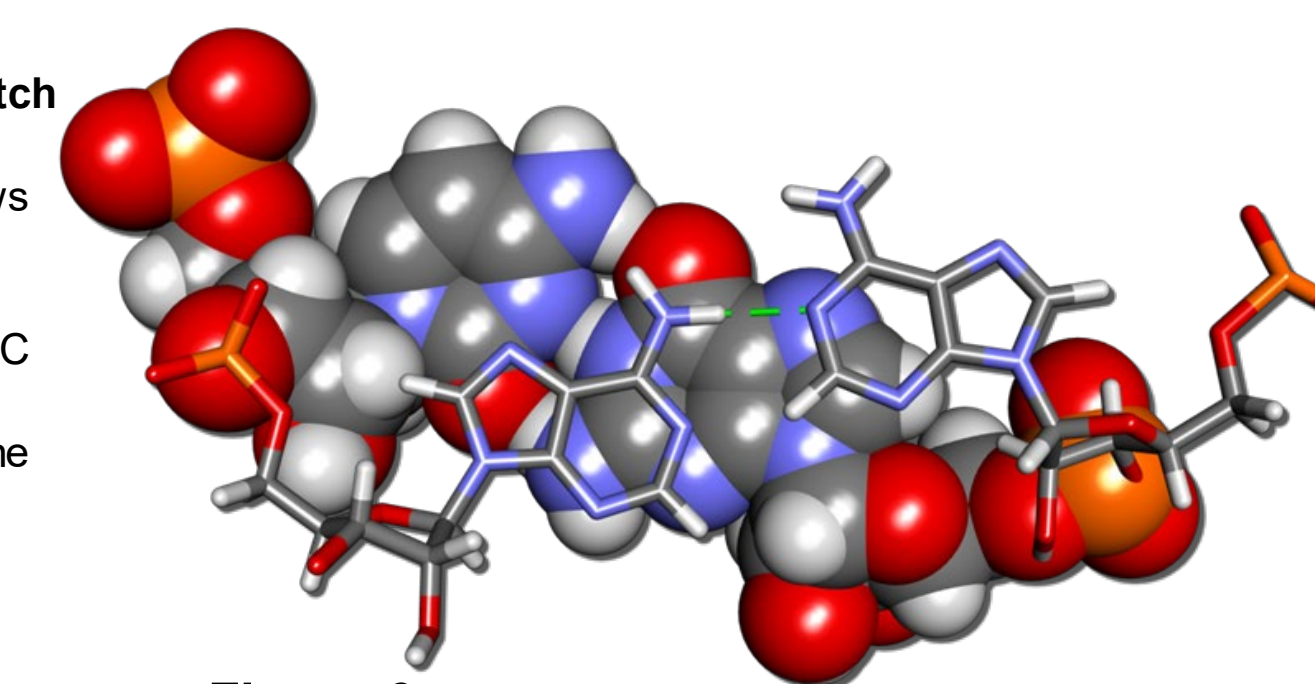
Figure 6

References

1. Ross CA, Tabrizi SJ (2011) Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol 10(1):83–98
2. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences of the United States of America, 101(19), 7287–7292.
3. Goddard, T. D., and Kneller, D. G. (2004) SPARKY, NMR Assignment and Integration Software, 3 ed., University of California, San Francisco, San Francisco.
4. Case, D. A., Betz, R. M., Botello-Smith, W., Cerutti, D. S., Cheatham, T. E. I., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, C., Lin, T., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. (2014) AMBER 14, University of California, San Francisco, San Francisco.