Scientific Report from a STMS visit
Cost Action P9, *Radiation Damage in Biological Systems*
Working Group 4, *Theoretical Developments for Radiation Damage*

Applicant: Maciej Haranczyk, University of Gdansk, Poland
Host: Dr. John Holliday, University of Sheffield, UK

COST STSM Reference Number: COST-STSM-P9-02569 (Period: 14/01/2007 to 11/02/2007)

Research Proposal Title: *Implementation of chemoinformatics techniques to the hybrid combinatorial-quantum mechanical method for identification of the most stable tautomers. Applications to the searches of the most stable tautomers of ionic nucleic acid bases.*

The purpose of the STSM visit was to take part in advanced training at the Chemoinformatics Group in the Department of Information Studies at the University of Sheffield. The planned training covered the following topics: (1) building and manipulating of molecular databases, (2) molecular similarity and dissimilarity calculations, and (3) molecular library diversity analysis. In parallel, my experience was extended by conducting research covering (1-3) and related to my studies of the most stable tautomers of charged nucleic acid bases (NABs), which might be important for radiation induced mutagenesis.

The main tasks carried out during the visit are: (1) building a molecular library of tautomers of NABs and analysis of each of them in terms of occurrence of substructure fragments. I developed new version of TauTGen, the tautomer generator program. It can export library of tautomers together with 2D connection tables. This data can be imported into BCI fingerprint toolkit software (Digital Chemistry), which codes 2D substructure features into Boolean arrays called fingerprints. These were later analyzed by small programs I developed; (2) writing software for performing similarity comparisons of molecules using number of different similarity coefficients; (3) writing software for performing clustering of a set of molecules basing on similarity comparisons; (4) performing substructure analysis to identify structure features that determine stability of particular tautomers.

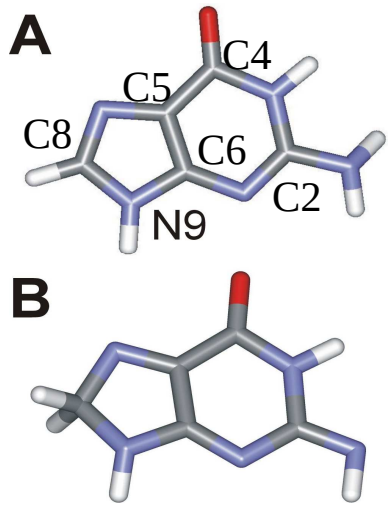

Fig. 1. Important tautomers of guanine: a) canonical tautomer of neutral guanine, b) one of adiabatically bound anions of guanine. N9 is connected to a sugar unit in DNA.

The most interesting results are related to the analysis of library of tautomers of anionic guanine. Recently, I was able to identify the most stable tautomers of anionic guanine among not-studied so far enamino-imino tautomers. I demonstrated that thirteen of these tautomers are adiabatically bound with respect to the neutral canonical tautomer (Fig. 1). During the STSM visit, I investigated structural similarity of tautomers of guanine using the chemoinformatics methods I had learnt. 2D substructure features of a set of 165 tautomers were coded into fingerprints. They were later compared using various similarity coefficients and clustered using a hierarchical aggregate group-average algorithm. I was able to find a cluster containing 24 elements including all the most stable tautomers (3% of total tautomers). When compared with the canonical tautomer, the distinct substructural features of tautomers from this cluster are additional hydrogen atoms at C8 and/or C2 atoms. By performing substructure analysis, I was able to demonstrate that another feature of adiabatically bound anions is lack of hydrogen atoms at C4, C5 and C6 carbons. These results might help us to predict proton transfer reactions induced by excess electrons attachment to guanine, as well as other NABs.

The results obtained during the visits will be presented as a poster on the 11th Electronic Computational Chemistry Conference taking place on the Internet (http://eccc.monmouth.edu) on 2nd-30th of April. We also project publication of results obtained for anionic guanine in a form of standard journal article.