# The influence of base-pair tautomerism on single point mutation rates in aqueous DNA

**Alex Gheorghiu [1], Alya A. Arabi [2], Peter V. Coveney [1],[3]**
[1] *Centre for Computational Chemistry, Department of Chemistry, University College London, UK*
[2] *College of Natural and Health Science, Zayed University, Abu Dhabi, UAE*
[3] *Informatics Institute, University of Amsterdam, Netherlands*

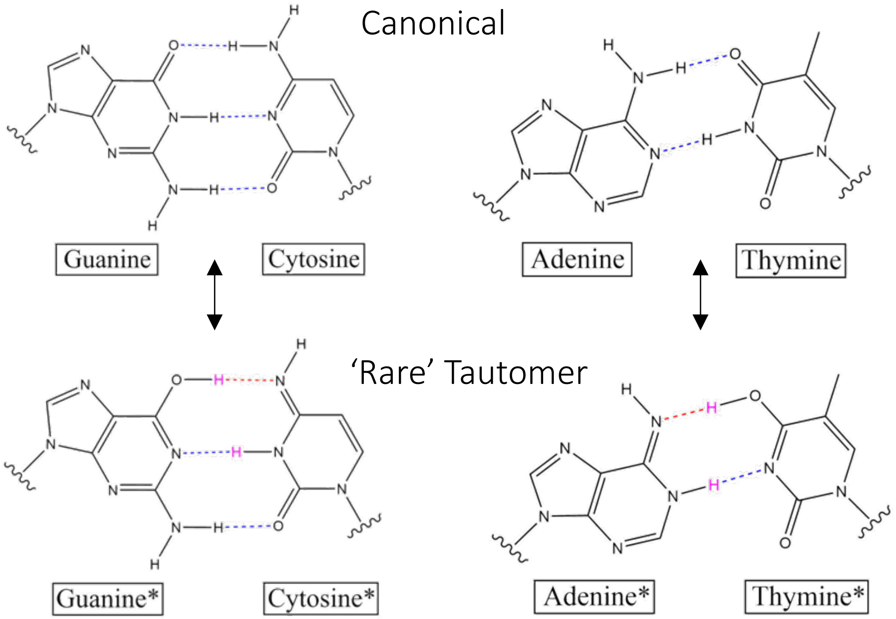
# Mutations in DNA

- External agents such as external electric fields, or other carcinogenic compounds are known to facilitate DNA mutations[1]

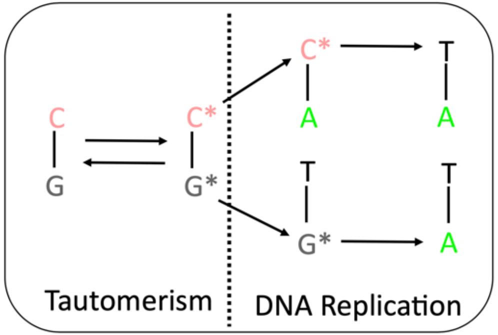
- … But what about *spontaneous mutations?*


(1) Loft, Steffen, and Henrik E. Poulsen. "Cancer risk and oxidative DNA damage in man." *Journal of molecular medicine* 74, no. 6 (1996): 297-312.

# Watson-Crick Base-Pairs and Löwdin's Hypothesis




- Protons obey the laws of quantum theory, behaving like 'wave packets'.
- Due to the quantum-mechanical tunnelling effect, there is always a small probability of proton transfer within any given base-pair.

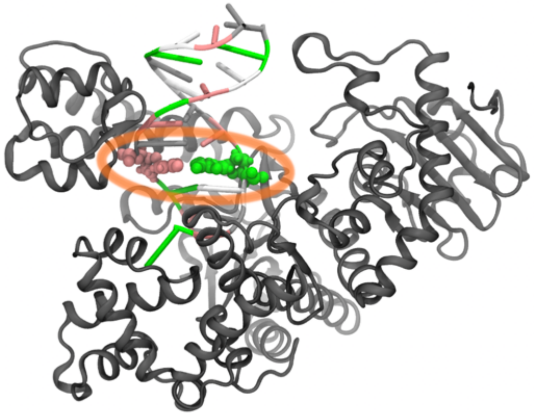



*'...this transfer of proton over distances less than 1 A may be the driving force for genetic mutations in all living organisms...'* – Löwdin[2]



[2] Löwdin, Per-Olov. "Proton tunneling in DNA and its biological implications." *Reviews of Modern Physics* 35.3 (1963): 724.

# Experimental Evidence and Previous Simulation



- DNA polymerase caught red-handed!
- Structural evidence of a C:A mismatch within crystallised DNA polymerase

- An overall **G:C → A:T** mutation bias has been reported in a variety of organisms

## QM-only gas-phase studies

- Proton transfer in G:C is energetically more favourable than in A:T[3]
- Double proton transfer; step-wise or concerted mechanism? [4]
- Little to no quantification of errors, due to lack of replicas
- A gas-phase base-pair is an idealised structure – is it truly representative?
- Provide a good starting point

[3] Jacquemin, Denis, et al. "Assessing the importance of proton transfer reactions in DNA." *Accounts of chemical research* 47.8 (2014): 2467-2474.
[4 ]Sekiya, H., & Sakota, K. (2008). Excited-state double-proton transfer in a model DNA base pair: Resolution for stepwise and concerted mechanism controversy... *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, *9*(2), 81-91.

# Considerations when modelling aqueous DNA

Reproducibility?

A realistic model; solvated, large enough DNA sample, >10,000 atoms

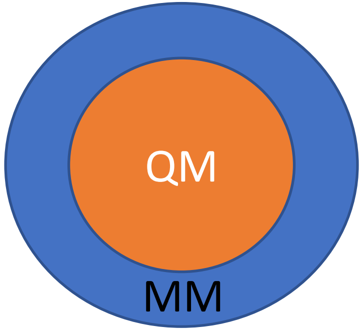
Proton transfer is both a bond breaking/ forming effect

Requires quantum mechanics (**QM**) to model: Expensive at high accuracy

DNA is a large, complex biomolecule molecule

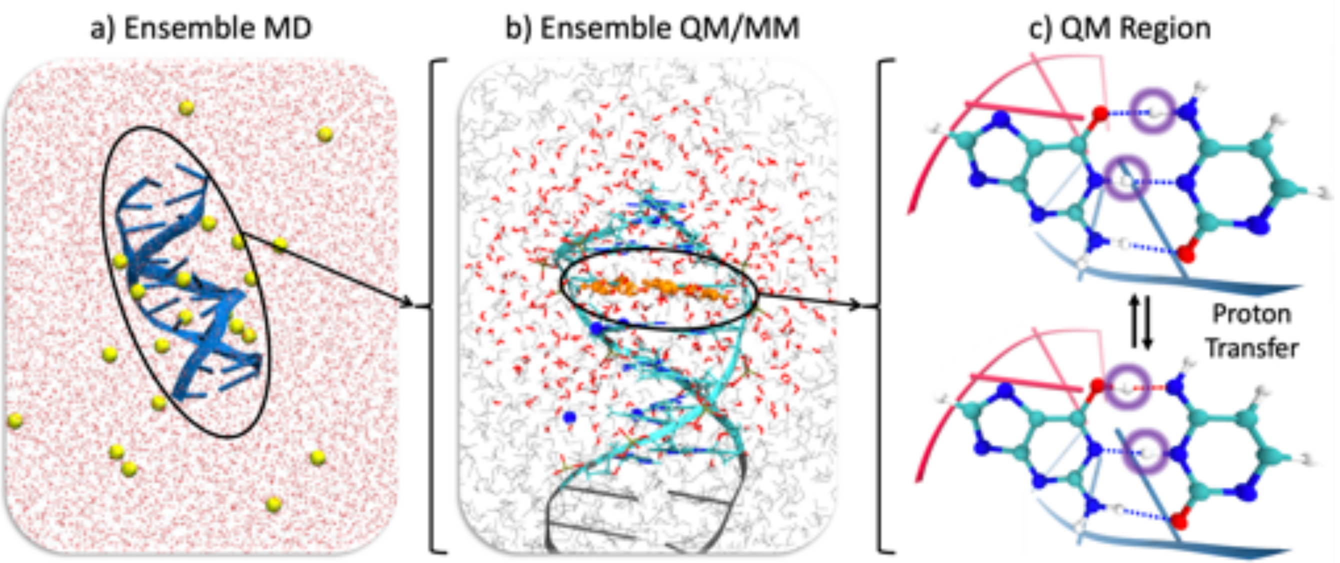
Use forcefield molecular mechanics (**MM**): Computationally cheaper

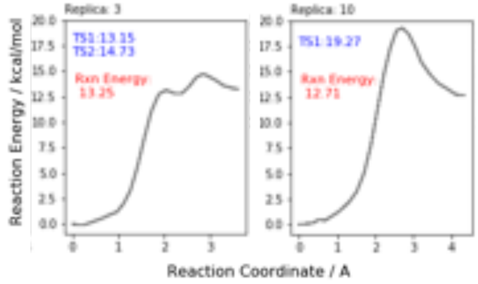
We can utilise both of these methods simultaneously[5]: (**QM/MM**)




[5] Lu, You, Zhenggang Lan, and Walter Thiel. "Hydrogen bonding regulates the monomeric nonradiative decay of adenine in DNA strands." *Angewandte Chemie International Edition* 50.30 (2011): 6864-6867.

# Modelling Proton Transfer: Our Multiscale Workflow




25 QM/MM reaction pathways mapped with nudged elastic band per base-pair

a) MD Code: NAMD
AMBER *parmbsc1* force field at 300 K, 1 atm. **10** replicas of 10 ns simulation to thermalize DNA effectively.

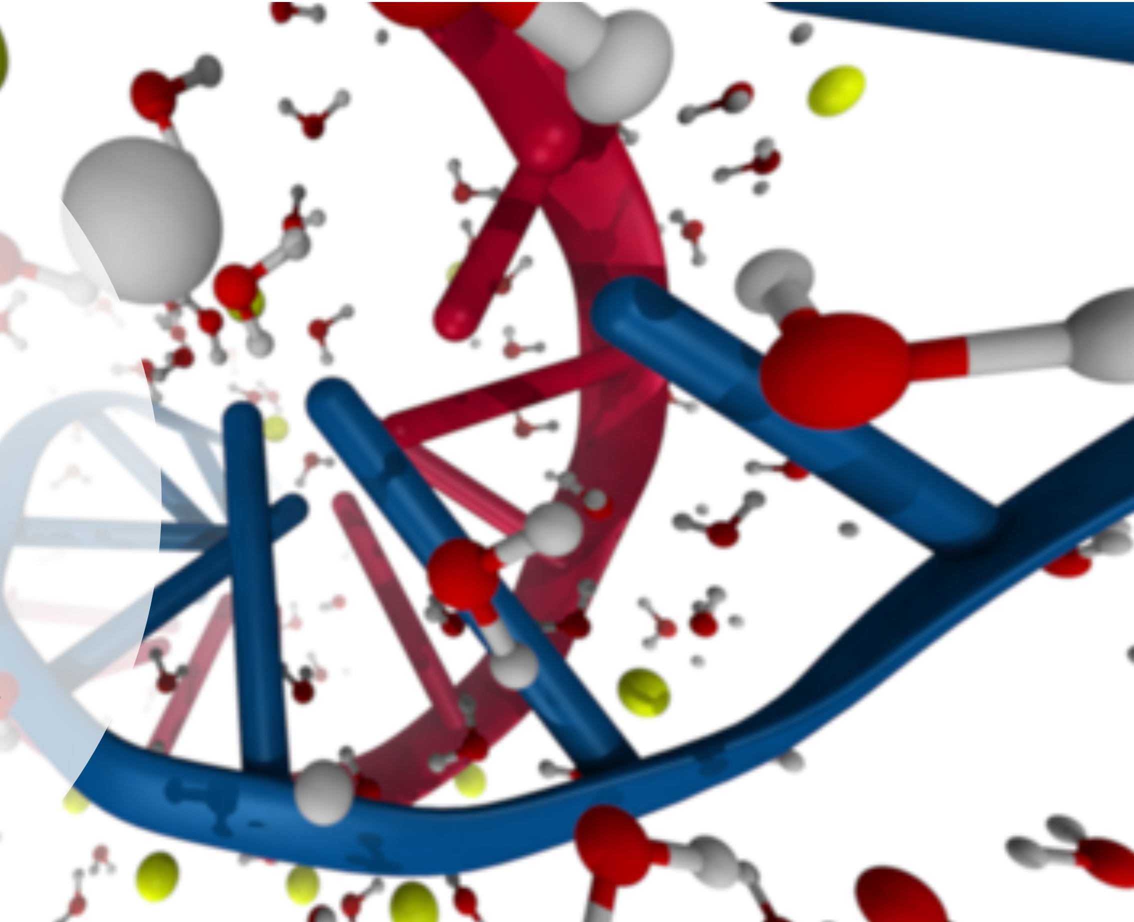
b) QM/MM Code: ChemShell
Linking NWChem (QM) with DL-POLY (MM). Ensemble QM/MM, using configurations drawn from MD*. **25 replicas** for a given base-pair.

c) QM Code: NWChem
One base pair (~30 atoms).
QM approximation:
B3LYP+XDM/aug-cc-pvdz

* Snapshots are chosen on an average base-pair distance criteria

# Advantages to using Ensemble QM/MM

- Inclusion of explicit solvation and DNA structure - more realistic than gas-phase!
- Observe multiple reaction paths for the same base-pair
- Determine the probability of each pathway occurring
- Comment on the statistical variance of the rates for each pathway



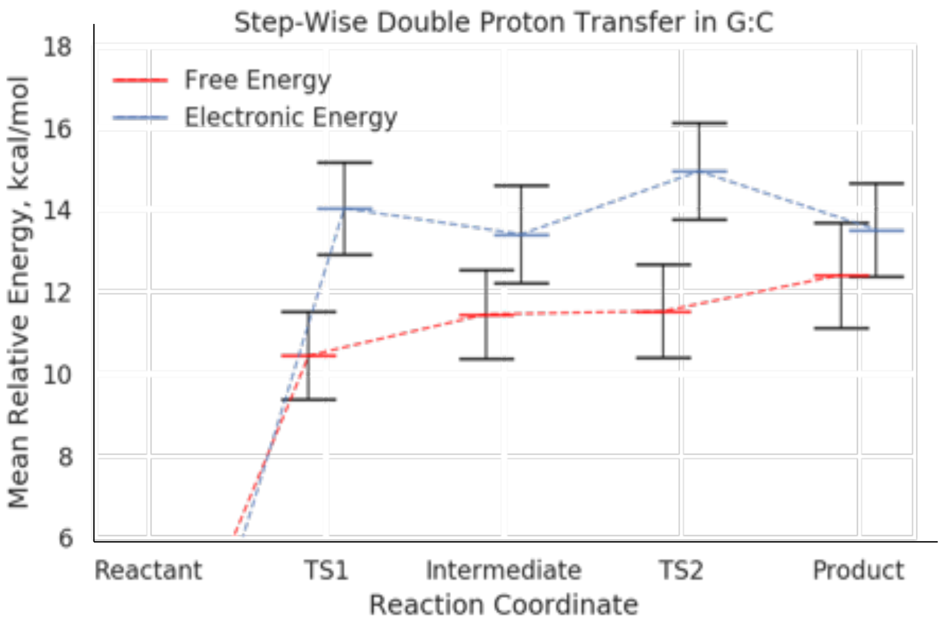
# Advantages to using Ensemble QM/MM

- Inclusion of explicit solvation and DNA structure - more realistic than gas-phase!
- Observe multiple reaction paths for the same base-pair
- Determine the probability of each pathway occurring
- Comment on the statistical variance of the rates for each pathway

| Base-Pair | No proton transfer | SPT Concerted | DPT | |
|---|---|---|---|---|
| | | | Step-Wise | Concerted |
| % of G:C replicas | 0 | 4 | 80 | 16 |
| % of A:T replicas | 56 | 36 | 0 | 8 |

# Step-Wise Double Proton Transfer in G:C

- The most probable type of proton transfer in G:C is step-wise, 80% of the time
- Step-wise, two transition states and an intermediate

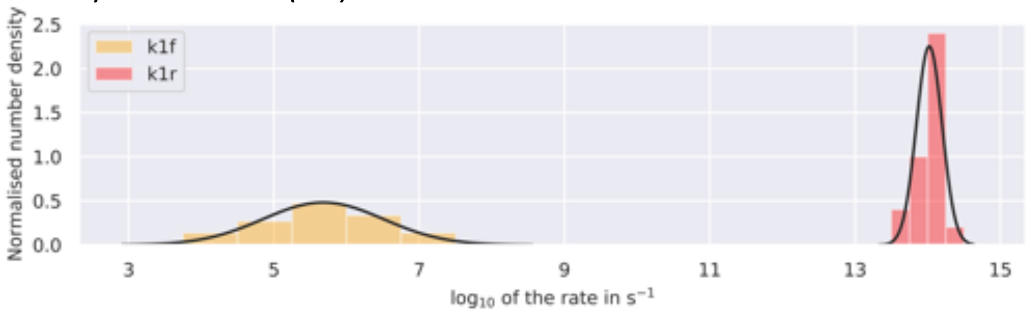



The reaction coordinate for the step-wise double proton transfer mechanism in the G:C base-pair. Error bars are standard deviation. Data points calculated from 20 replicas out of 25 total

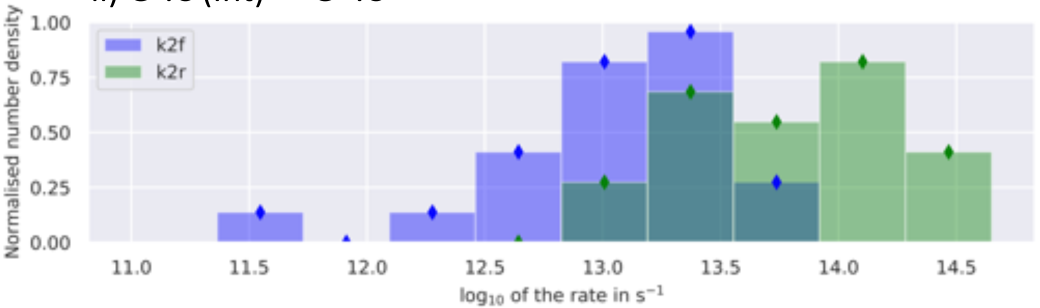
# Step-Wise Double Proton Transfer in G:C

- Rate determining step is $k_{1f}$ , many orders of magnitude slower than $k_{1r}$
- The use of ensemble QM/MM gives insight to the statistical variance
- The second step is much closer to equilibrium than the first
- Tautomer has a lifetime of ~50 fs (1/ $k_{1r}$)

i) G:C -> $G^{+}$:$C^{-}$(Int)




ii) $G^{+}$:$C^{-}$(Int) -> G*:C*




The normalized histogram of log proton transfer rates for the step-wise double proton transfer mechanism in the G:C base-pair

# Step-Wise Double Proton Transfer in G:C

$$G_{taut} = \left( \sum_{K_{DPT}} K_{eq,weighted} \right) \times G \quad (1)$$

where;

$G$ is size of genome in base-pairs ($3 \times 10^9$)

The probability of step-wise double proton transfer is 0.8

$K_{eq}$ is the calculated average equilibrium constant ($1.54 \times 10^{-8}$)

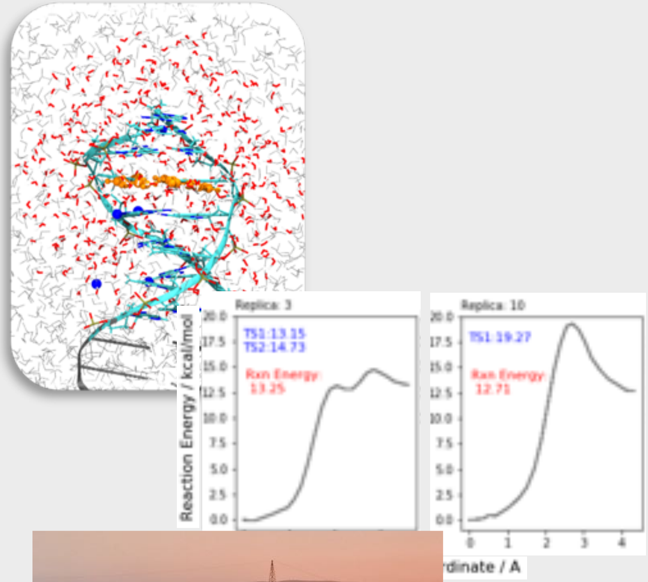
$K_{eq,weighted}$ is $K_{eq}$ multiplied by it's probability of occurrence

$G_{taut}$ is the number of tautomeric base-pairs in the genome at any given time

- Assuming;
  - Genome has 100% G:C content
  - Every tautomerism leads to mismatch
- $G_{taut}$ = 0 – 114 base-pairs per-genome
- 1 in 3.8 x $10^{-8}$ mutations per nucleotide site.
- This value is within the estimation of mutation rate per nucleotide in humans ($10^{-11}/10^{-8}$) [6]

[6] Nachman, Michael W., and Susan L. Crowell. "Estimate of the mutation rate per nucleotide in humans." *Genetics* 156.1 (2000): 297-304.

# Conclusions



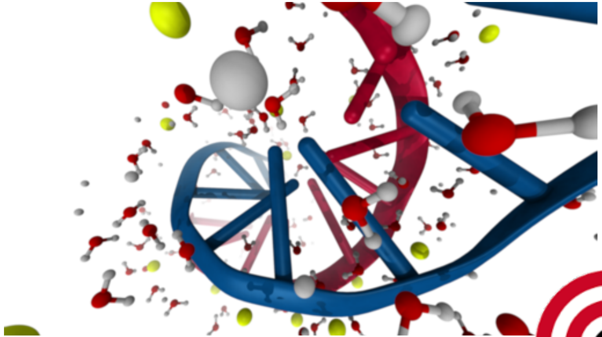

?

- Proton transfer in biological systems is very sensitive to its environment, necessitating multiscale modelling
- Multiple reaction pathways for each base-pair are observed, with varying probabilities
  - G:C double proton transfer 96% of the time
  - A:T single proton transfer 36% of the time
- The frequency of a tautomer in human genome is $< 0.001$ %
- The life-times of the tautomers are *fs*, while DNA unwinding occurs on the *ns* scale

UCL

Contact: **uccaagh@ucl.ac.uk**

# *Thanks for listening!*



Big thanks to: All the members of the CCS and Tom Keal &You Lu, who are ChemShell developers at STFC Daresbury

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 675451 (CompBioMed project). Support from the UK Engineering and Physical Sciences Research Council under the project ‘UK Consortium on Mesoscale Engineering Sciences (UKCOMES)’ (Grant No. EP/R029598/1) is gratefully acknowledged.
PhD funding received from the the Zayed Univeristy/UCL Impact Scholarship

NWCHEM
HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

archer

جامعة زايد
ZAYED UNIVERSITY

DL_Poly

ChemShell
multiscale computational chemistry