# oxDNA3: Introducing Sequence-Specific Curvature and Elasticity into a Coarse-Grained DNA Model


Oliver Henrich [1 *], Andrea Bonato [1], Thomas Ouldridge [2], Ard Louis [3], Jonathan Doye [3], Lorenzo Rovigatti [4], Michael Matthies [5]

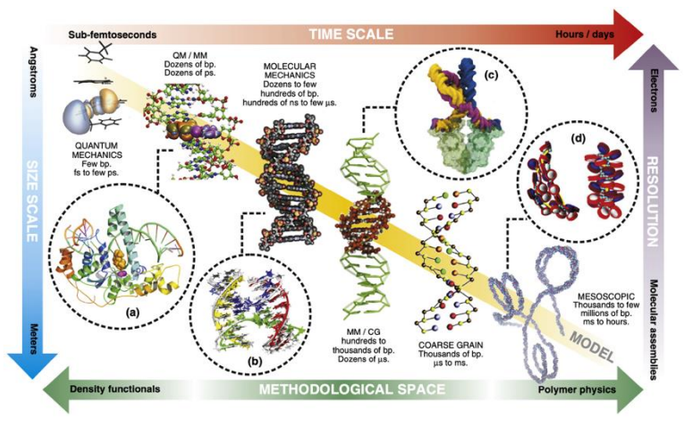
[1] Department of Physics, University of Strathclyde, Glasgow;
[2] Department of Bioengineering, Imperial College London;
[3] Rudolf Peierls Centre for Theoretical Physics & [4] Department of Chemistry, University of Oxford;
[5] Department of Physics, Sapienza University of Rome; [6] Department of Bioscience, Technical University Munich

[*] oliver.henrich@strath.ac.uk


## Computational Models of DNA

The **most detailed models on the left**

- calculate forces from **first principles**
- are **very accurate** and **board in scope**
- can **treat** only **very small systems on very short times**




The **least detailed models on the right**

- **omit** most, if not all **structural details**
- are more **limited and specific in scope**
- Can treat **very large systems on very long times**

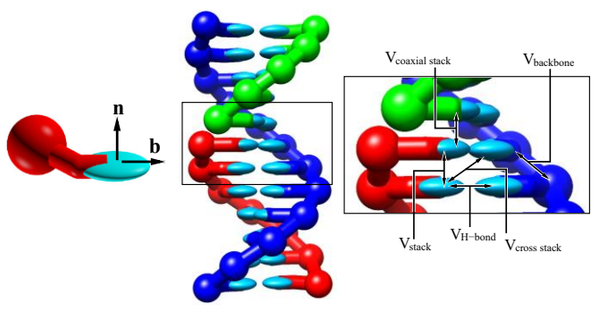
**Intermediate models at the centre,** which we develop, **strike a balance** as they

- **retain crucial structural details**, but **omit** other **less important ones** for specific purposes
- can be **quite accurate,** but have **usually** a **reduced scope**
- are **suitable** for modelling **medium sized systems** and **intermediate time scales**

## Coarse-Grained oxDNA Model

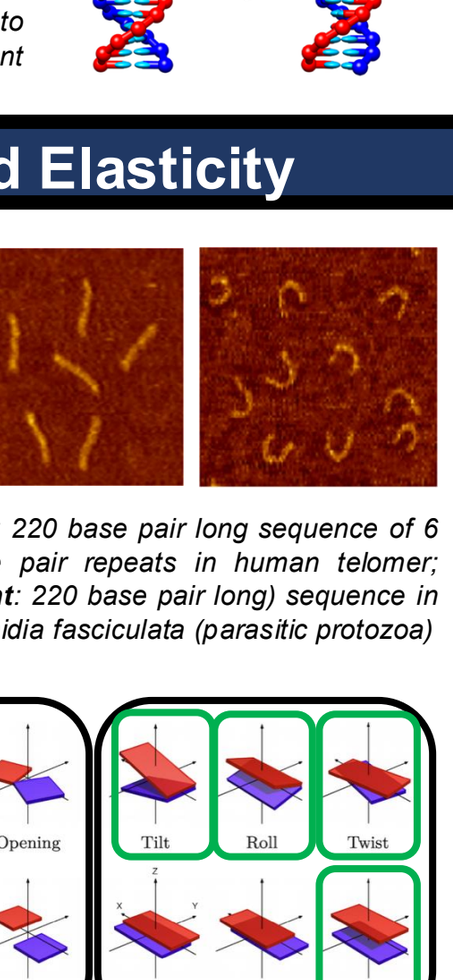
Our **oxNA suite of DNA** [1], **RNA** [2] and **DNA:RNA hybrid** [3] models is based upon what is referred to as **rigid base model**. **Each nucleotide** is modelled as a **rigid body** with **three interaction sites** for the effective interactions between the nucleotides, which we derive from much more detailed models. These are:

- **Connectivity** of the sugar-phosphate backbone
- **Excluded volume** of the nucleotides
- **Stacking**, **cross-stacking** and **coaxial stacking**
- **Hydrogen-bonding** between **complementary bases**
- **Electrostatic interaction** between negatively charged sugar-phosphate backbones




***Top and top right**: Overview of bonded and pair interactions in the oxDNA coarse-grained model of DNA.*

***Right**: Backbone interaction site at an angle (right) lead to the formation of the major and minor groove, an important structural feature of DNA.*




## Intrinsic Curvature and Elasticity

Nucleic acids are **intrinsically curved depending on the local sequence**. The sequence modulates also the **local elastic properties**, which is important for the **interaction of DNA or RNA with proteins**.

The sequence-dependent curvature and elasticity originate from the **relative orientation** of the nucleobases, here below displayed as slabs.



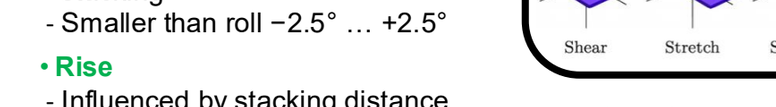
***Left**: 220 base pair long sequence of 6 base pair repeats in human telomer; **Right**: 220 base pair long) sequence in Crithidia fasciculata (parasitic protozoa)*

- **Twist**
  - Influenced by backbone length and stacking distance
  - +29° … +41°
- **Roll** (soft mode)
  - Emerges due to desire of nucleobases to overlap or to avoid a steric clash
  - −2° … +9°
- **Tilt** (stiff mode)
  - Influenced by base geometry and stacking
  - Smaller than roll −2.5° … +2.5°
- **Rise**
  - Influenced by stacking distance
  - 3.1 … 3.5 Å




*Intra-base pair coordinates*

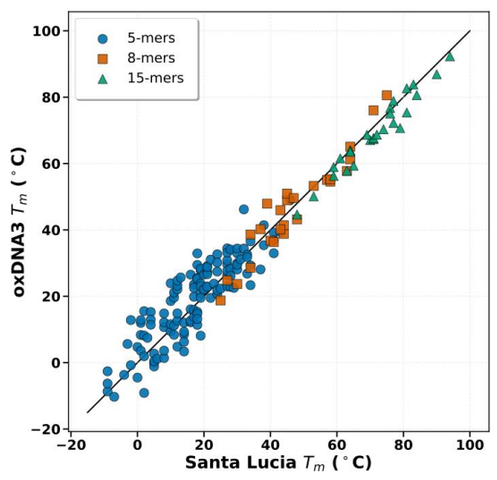
*Inter-base pair coordinates*




***Left**: The **different size** of purine and pyrimidine bases causes to either **slide (B)** or **roll (C)** within a base pair step. This effect is **physically modelled in oxDNA3**.*

## oxDNA3 Results and Features

### Thermodynamics



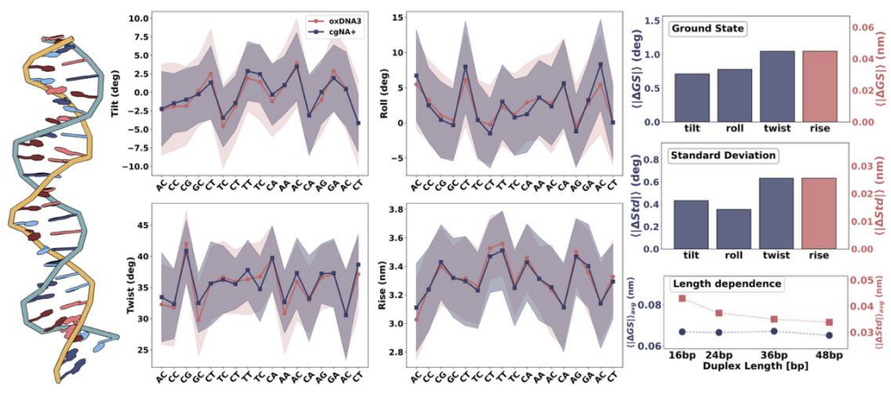

**Left: Melting temperature** in **oxDNA3** compared to predictions of the **Santa Lucia nearest-neighbour model**

Shown are results for 129 5-mer, 30 8-mer and 30 15-mers De Bruijn sequences, which include all tetramers in the training.

The **RMSD** between oxDNA3 and Santa Lucia melting temperatures is in the region of **2.9K**.

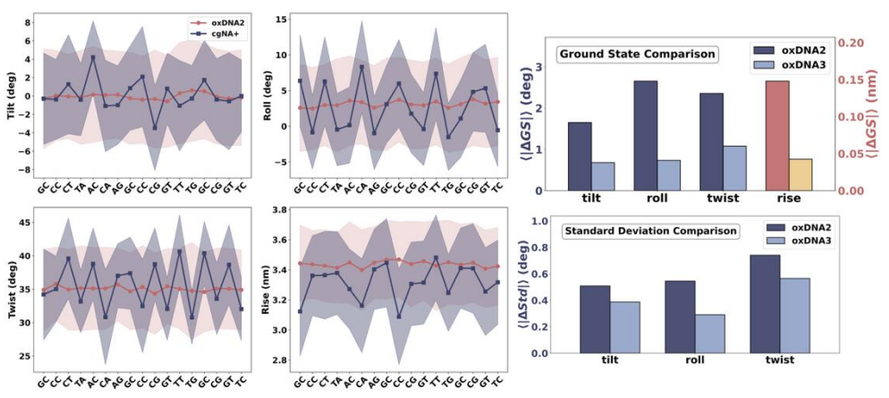
### Geometry

The geometry (**curvature** and **elasticity**) is trained on the **cgDNA+** model [4], which is itself is trained on **all-atom MD simulations** with the **parmbsc1** force field.




***oxDNA3 vs cgDNA+***

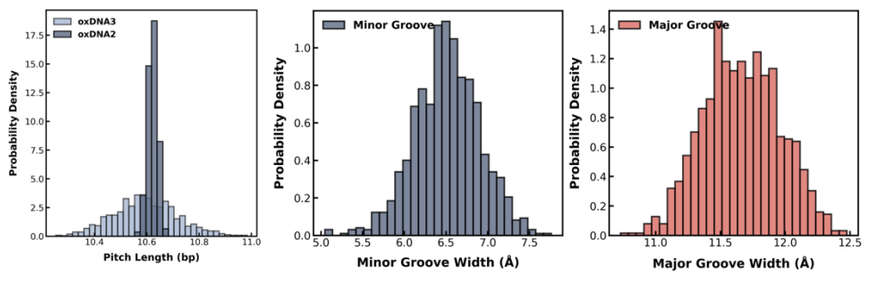
***Left**: Mean values of the **helical coordinates tilt, roll, twist** and **rise** (centre) of a random sequence (left) as well as variances (right).*



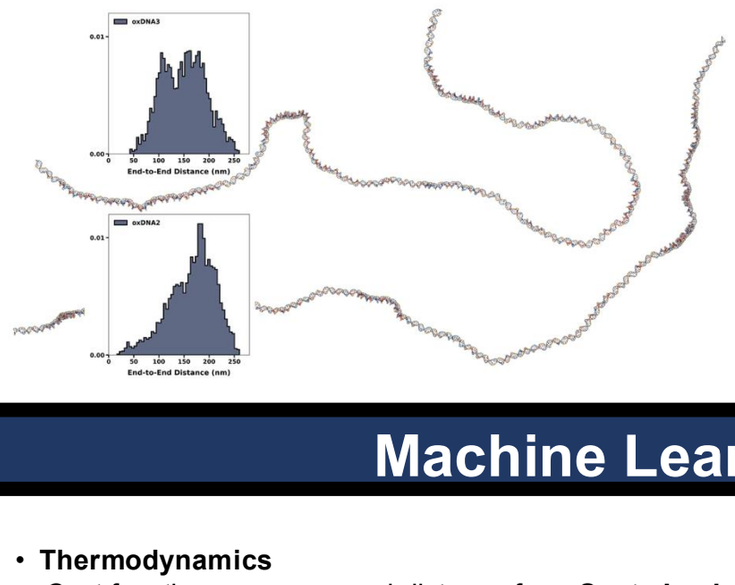

***oxDNA2 vs cgDNA+***

***Left**: Mean values of the **helical coordinates tilt, roll, twist** and **rise** (left) of the same random sequence and variances (right).*

***Right**: Histograms of **sequence-dependent pitch length** in oxDNA3 and oxDNA2 (left), **minor groove width** (centre) and **major groove width** in oxDNA3 (right).*




***oxDNA3** model (top) with **sequence-specific curvature** and **elasticity** and oxDNA2 (bottom)*




***Left**: Snapshots and distribution of the end-to-end distance of a 1,000 bp sequence from Crithidia fasciculata.*

***Sequence-averaged persistence lengths***

***oxDNA3: $l_b$ = 45 nm, $l_t$=204 nm***

***oxDNA2: $l_b$ = 44 nm, $l_t$=163 nm***

## Machine Learning

- **Thermodynamics**
  - Cost function mean squared distance from **Santa Lucia nearest-neighbour model predicted melting temperature**
- **Geometry**
  - Cost function **Kullback-Leibler divergence** of **helical coordinate distributions in oxDNA3** and **cgDNA+ (parmbsc1)**

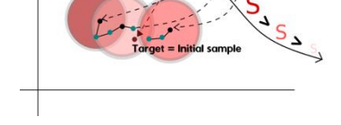
$$S[\rho^{\Gamma}_{ox}, \rho_{cgna}] = \int \rho^{\Gamma}_{ox}(\xi) \log\left(\frac{\rho^{\Gamma}_{ox}(\xi)}{\rho_{cgna}(\xi)}\right) d\xi$$

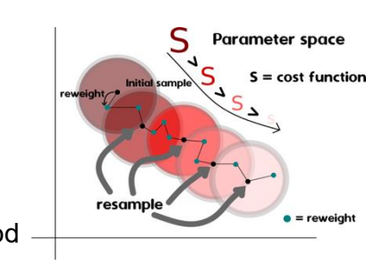
$\rho^{\Gamma}_{ox}(\xi)$ Distribution of coordinates in oxDNA

$\rho_{cgna}(\xi)$ Distribution of coordinates in cgDNA+

- $\mathcal{O}$**(3000) tuneable parameters** depending on **nucleotide tetramers**: AAAA, AAAC, AAAG, …, ACGT, …, TTTT

⇒ ML framework with PyTorch using L-BFGS quasi-Newton method




## References & Acknowledgements


[1] T. Ouldridge, A. Louis, J. Doye, *J. Chem. Phys.* **134**, 085101 (2011); B. Snodin, F. Randisi, M. Mosayebi, et al., *J. Chem. Phys.* **142**, 234901 (2015).

[2] P. Šulc, F. Romano, T. Ouldridge, J. Doye, A. Louis, *J. Chem. Phys* **140**, 235102 (2014).

[3] E. Ratajczyk, P. Šulc, A. Turberfield, J. Doye, A. Louis, *J. Chem. Phys.* **160**, 115101 (2024).

[4] R. Sharma, A. Patelli, L. De Bruin, J. Maddocks, *J. Mol. Biol.* **14**, 167978 (2023).



Engineering and Physical Sciences Research Council

Grants No. EP/N019180/2, EP/V06231X/1 & EP/Y008693/1