Computational Estimation of Energy Parameters for modified Nucleotides

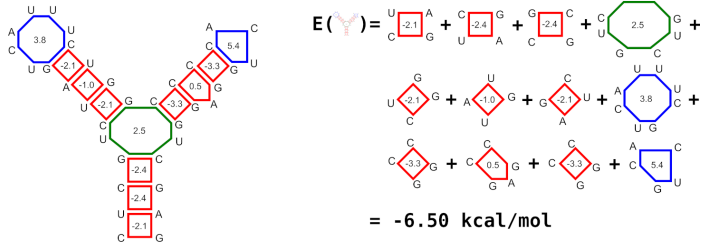
Thomas Spicher

TBI Vienna
University of Vienna

38^{th} TBI Winterseminar
Bled, February 13, 2023

universität
wien

Nearest Neighbor (NN) Model



- Secondary structures can be decomposed into loops formed by adjacent pairs
- A free energy contribution is assigned to each loop

$$E(\mathcal{S}) \approx \sum_{l \in \mathcal{S}} E(l)$$

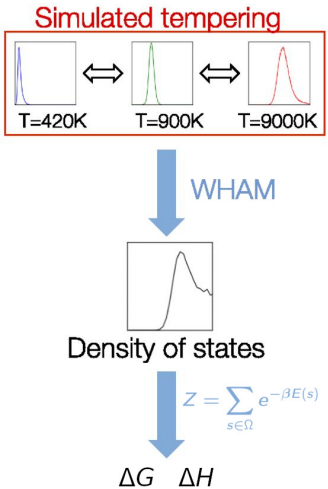
- 294 independent nearest neighbor parameters included in the Turner NNDB[1]
  - From 802 UV melting experiments
- **Virtually no energy parameters for modified bases**

[1] Turner et al., "NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure.", 2009, NAR 38, D280–D282

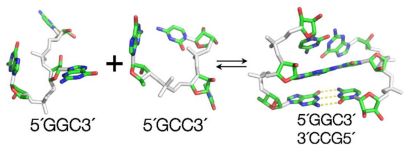
Computational Estimations of Nearest Neighbor Parameters

- Coarse grained models VS Molecular dynamics
  - No solvent
  - Less degrees of freedom
  - Lower computational costs
- Coarse grained models seem to be the best choice
  - Rosetta-RECCES
  - Coarse grained moves
  - Energy calculation at atom level

RECCES[2] Simulation



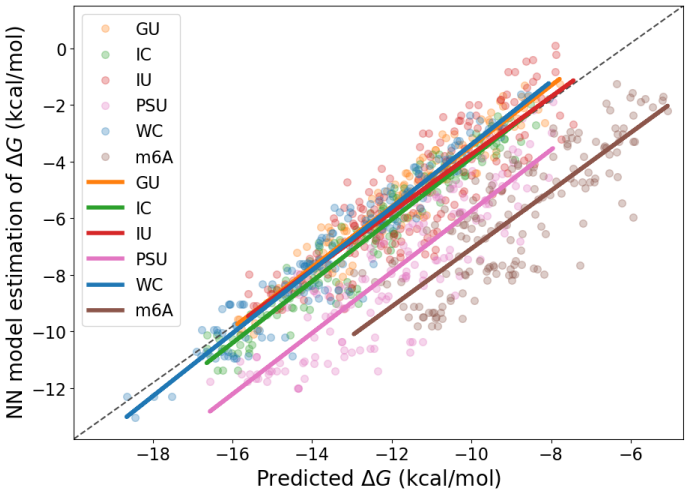
- Monte Carlo simulation
- Temperature changes during the simulation
  - To overcome energy barriers and sample the full conformation space
- Energy of each conformation determined by an energy function
  - Combination of Physics- and knowledge-based terms



$$\Delta G\begin{pmatrix}\text{5'GGC}\\ \text{3'CCG}\end{pmatrix} = G\begin{pmatrix}\text{5'GGC}\\ \text{3'CCG}\end{pmatrix} - G(\text{GGC}) - G(\text{CCG})$$

[2] Chou et al., "Blind tests of RNAnearest-neighbor energy prediction", 2016 Proceedings of the National Academy of Sciences, 113(30), 8430–8435.

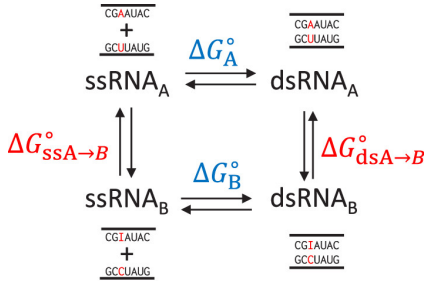
Comparison of free energy estimations



- Comparaison to estimations using existing parameters
- Duplex of 3 to 5 base pairs
- "good" correlation
- Systematic error
- Modification dependent
- Rosetta energy function poorly parameterized?
- Water plays a role in the stability of $\Psi \bullet A$ stacks

Molecular dynamics (MD)

- free energies of proteins and their ligands calculated for decades
- studies on nucleic acids rather limited
- including only inosine as modification[3]
- standard MD based on thermodynamic cycle to calculate the free energies
- impossible to simulate the chemical path
- using alchemical path
  - insert intermediate state $0 = \lambda_1, \lambda_2, ..., 1 = \lambda_n$
  - $\Delta G = \sum_i \Delta G_{i,i-1}$

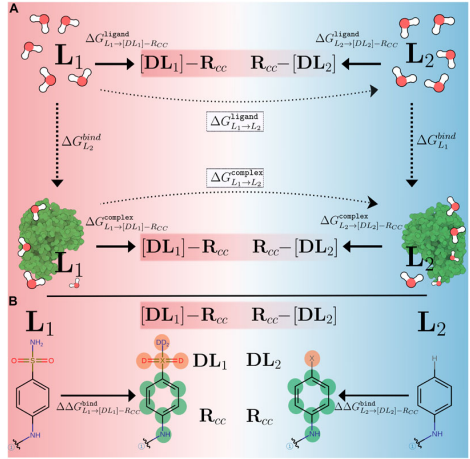


$$\Delta\Delta G^\circ_{AB} = \Delta G^\circ_{dsA\to B} - \Delta G^\circ_{ssA\to B} = \Delta G^\circ_B - \Delta G^\circ_A$$

[3] Sakuraba et al. "Free-Energy Calculation of Ribonucleic Inosines and Its Application to Nearest-Neighbor Parameters." Journal of Chemical Theory and Computation 16.9 (2020): 5923-5935.

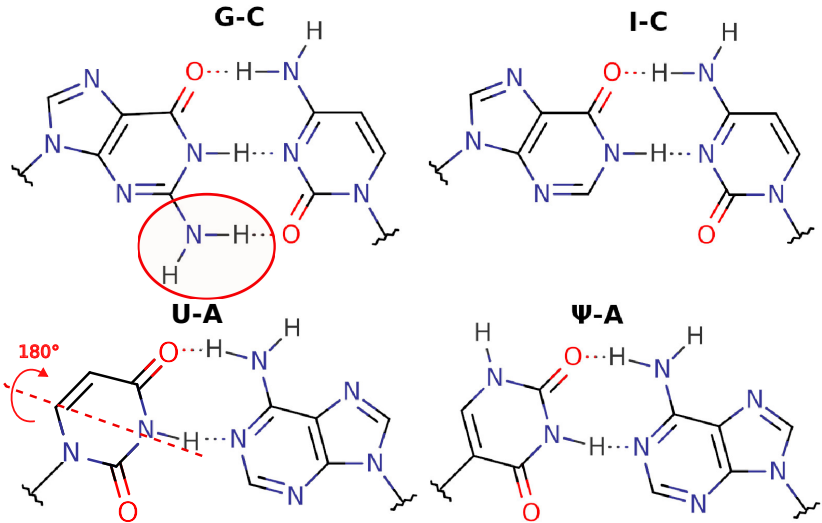
Transformato[4]

- OpenMM as underlying MD programm
- Input files created by CHARMM-GUI
- Identify the maximum common substructure
- Construction of an alchemical path
  - Atoms not belonging to the common substructure are mutated to dummy atoms
  - Over a certain amount of steps the interactions of the dummy atoms are turned off
- Free energy difference to the common core calculated with Multistate Bennett Acceptance Ratio (MBAR)



[4] Karwounopoulos et al. "Relative binding free energy calculations with transformato: A molecular dynamics engine-independent tool." Frontiers in Molecular Biosciences 9 (2022)

G-C to IC and U-A to Ψ-A at 1M NaCl



$\Delta\Delta G$ MD VS. RECCES

Duplex	MD	std	RECCES	Turner
5'CAICG 3'GUCGC	2.28	0.21	1.31	2.04
5'GCAICGC 3'CGUCGCG	2.04	0.27	1.79	2.04
5'GAAGAIGCAA 3'CUUCUCCGUU	1.88	0.44	2.05	1.73
5'GAAGAICCAA 3'CUUCUCGGUU	2.03	0.49	1.80	2.04
5'GGGCUCAIGCAGCCG 3'CCCGAGUCCGUCGGC	2.68	0.18		1.54
5'CGGPUCG 3'GCCAAGC	0.87	0.92	0.89	−2.86
5'GGGCAGGPUCGCGCC 3'CCCGUCCAAGCGCGG	−0.28	2.36		−2.86

Conclusion and Outlooks

- Rosetta-RECCES
  - Good predictions for WC, $G \bullet U$, and $I \bullet C$ base pairs
  - Systematic deviation for other modifications ($\Psi \bullet A$, $m^6A \bullet U$)
- Molecular Dynamics
  - Framework still need some adaptations
  - Promising results with $I \bullet C$ base pairs
  - Use of different salt concentrations
- A combination of methods involving MD, RECCES, and experiments may represent the most suitable way
- Starting computations for a full set of stack pairs containing modifications
- Include larger loops

Acknowledgments

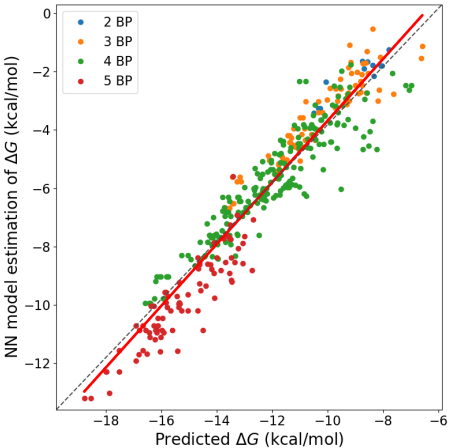
Ivo Hofacker
Ronny Lorenz
Yuliia Varenyk
Johannes Karwounopoulos
Stefan Boresch

The whole TBI team

Thank you for your attention

Comparison of free energy estimations

Canonical base pairs



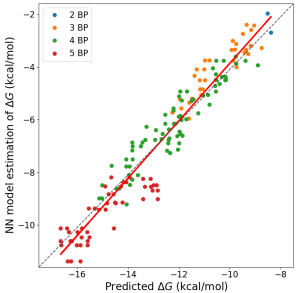
- All simulated duplexes from both methods (2 to 5 BP)

- Average offset $= 6.18$
- linear fit: $y = 1.06 \cdot x + 6.92$
- Corr. coeff. $= 0.95$

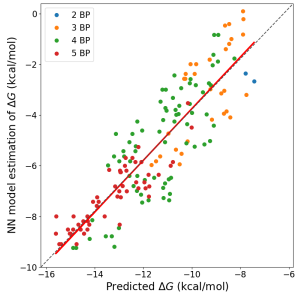
- High correlation
- Slope of the linear fit near 1
- Both models are separated by $\approx$ only a constant

$\Rightarrow$ This good concordance is reflected in the accurate parameter estimations for canonical base pairs

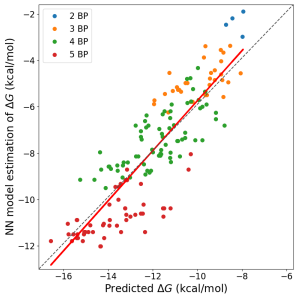
Comparison of free energy estimations



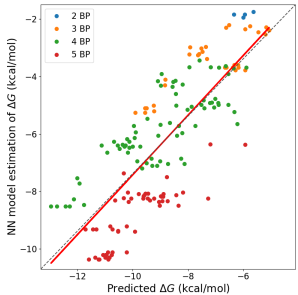
$I \bullet C$
Average offset $= 5.90$
linear fit:
$y = 1.09 \cdot x + 7.04$
corr. coeff. $= 0.96$



$I \bullet U$
Average offset $= 6.59$
linear fit:
$y = 1.12 \cdot x + 7.97$
corr. coeff. $= 0.87$



$\Psi \bullet A$
Average offset $= 4.11$
linear fit:
$y = 1.08 \cdot x + 5.05$
corr. coeff. $= 0.87$



$m^6A \bullet U$
Average offset $= 2.62$
linear fit:
$y = 1.04 \cdot x + 2.97$
corr. coeff. $= 0.80$

High Resolution Melting (HRM) experiments [5]

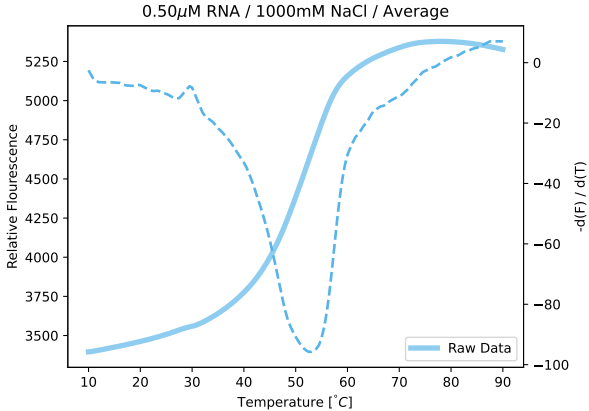
- Measures fluorescence as a function of temperature
- Derivation of thermodynamic parameters
- Good agreement with UV-melting if $T_m > 40°C$

Is this a suitable alternative to UV-melting?

- Measurements for $I - C$ pairs in Jantsch Lab:
  - Oligos of 10 base pairs
  - NaCl concentrations $0mM$, $150mM$, $1000mM$
  - Oligo concentration $0.5\mu M$
  - Melting from 10 - $85°C$ in $0.5°C$ steps
  - 3 replicates

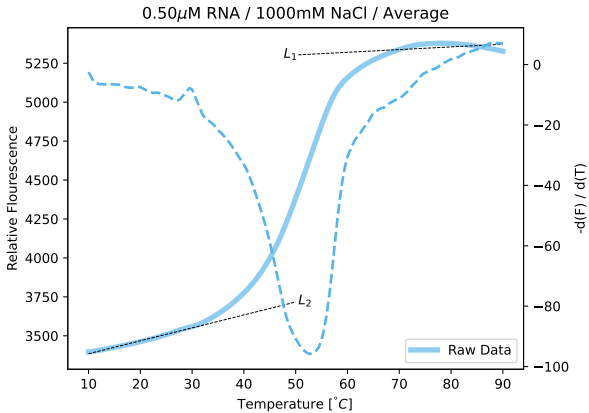
[5] Wang et al. 2016, "Assessment for Melting Temperature Measurement of Nucleic Acid by HRM", Journal of Analytical Methods in Chemistry

Melting Data Analysis



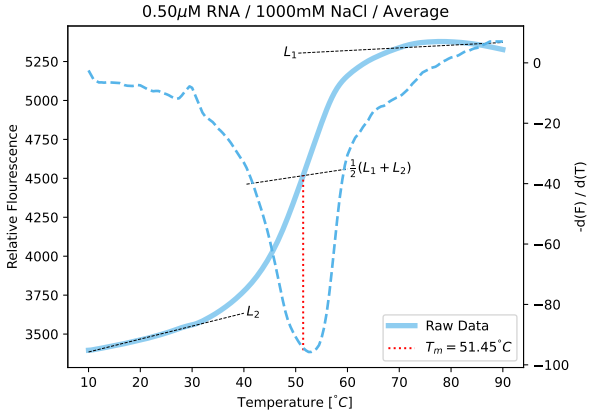
- $F(T)$ is related to the fraction of duplexes $\theta(T)$
- $A + B \rightleftarrows AB$ with $2[AB] + [A] + [B] = c_o$
- $K_a = \frac{[AB]}{[A][B]} = \frac{2\theta}{c_o(1-\theta)^2}$
- $\theta(T_m) = 0.5$

Melting Data Analysis



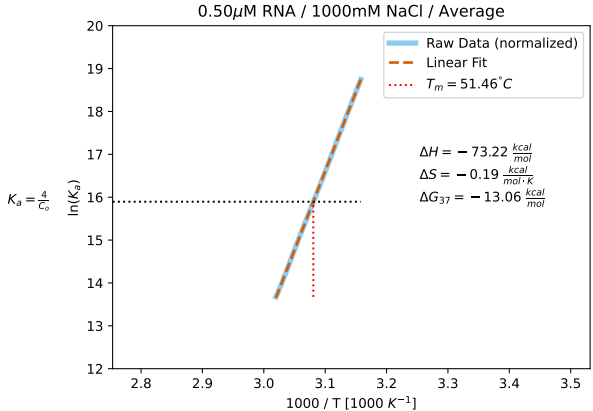
- Use $F(T)$ at low and high temperatures to estimate L_1 and L_2
- $F(T) = L_2(T) \cdot \theta(T) + L_1(T) \cdot (1 - \theta(T))$

Melting Data Analysis



What about ΔG_{37}, ΔH, and ΔS?

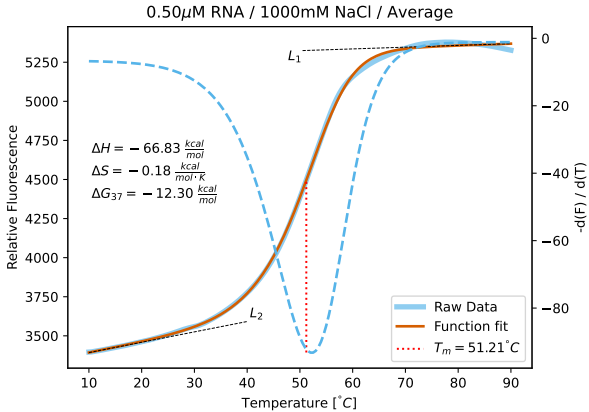
Van't Hoff Analysis



$$\theta(T) = \frac{F(T) - L_1(T)}{L_2(T) - L_1(T)} \qquad 0.1 < \theta(T) < 0.9 \qquad K_a = \frac{[AB]}{[A][B]} = \frac{2\theta}{c_o(1-\theta)^2}$$

$$\frac{1}{T_m} = \frac{\Delta S}{\Delta H} - \frac{R \cdot \ln(4/c_o)}{\Delta H} \qquad \Delta H = -m \cdot R \qquad \Delta S = \frac{\Delta H}{T_m} + R \cdot \ln(4/c_o)$$

Curve fitting



- formulate $\theta(T)$ as function of T, ΔH and ΔS using $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$ and $K_a = \frac{2\theta}{c_o(1-\theta)^2}$
- fit full function $F(T) = L_2(T) \cdot \theta(T) + L_1(T) \cdot (1 - \theta(T))$