NEURAL INFORMATION PROCESSING SYSTEMS

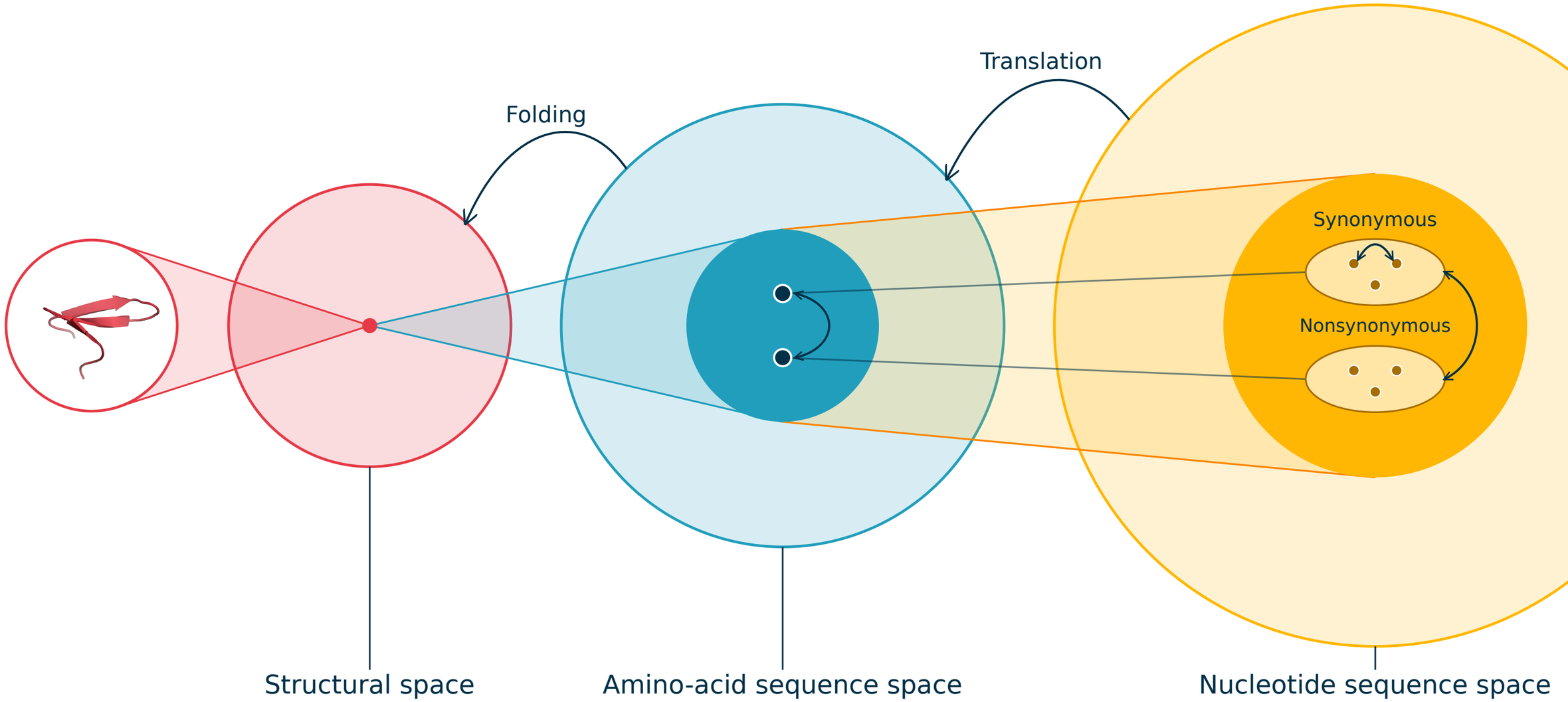
Joint protein, mRNA, and DNA sequence design and optimization with nucleotide-level Potts models

Błażej Banaszewski* Lars J. Dornfeld* Dexiong Chen Karsten Borgwardt† Lukas F. Millest†

* Equal contribution † Joint supervision



Different amino-acid and nucleotide properties



Inverse folding on the nucleotide level allows simultaneous optimization and trade-offs of both amino acid and RNA/DNA-level properties

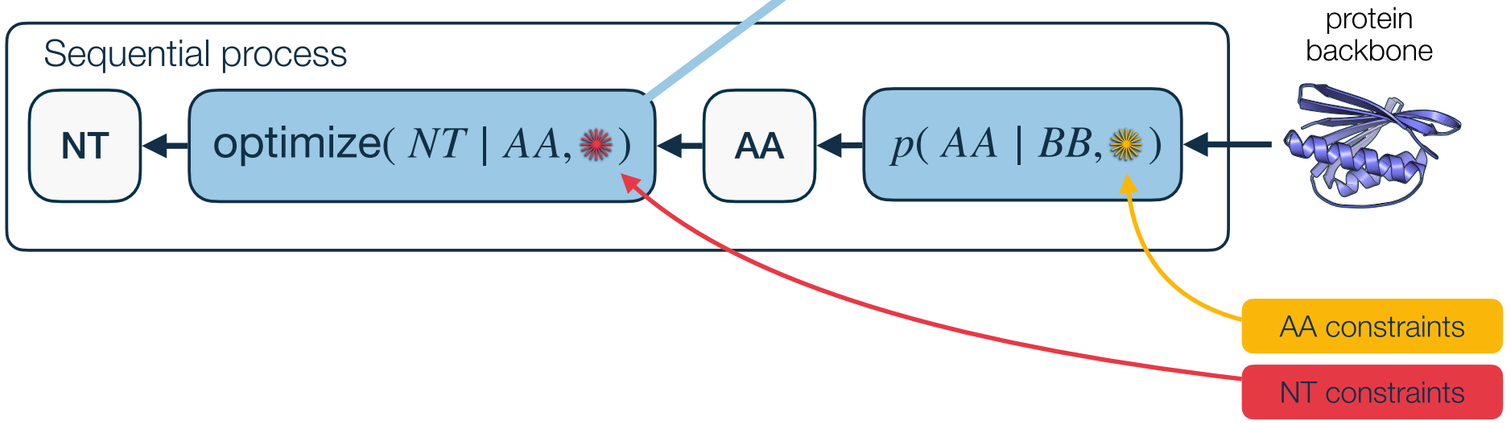
NuCaliby

Standard codon table mapping 61 codons to 20 amino acids (excl. 3 stop codons)

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
GCT GCC GCA GCG	TGT TGC	GAT GAC	GAA GAG	TTT TTC	GGT GGC GGA GGG	CAT CAC	ATT ATC ATA	AAA AAG	TTA TTG CTT CTC CTA CTG	ATG	AAT AAC	CCT CCC CCA CCG	CAA CAG	CGT CGC CGA CGG AGA AGG	TCT TCC TCA TCG AGT AGC	ACT ACC ACA ACG	GTT GTC GTA GTG	TGG	TAT TAC

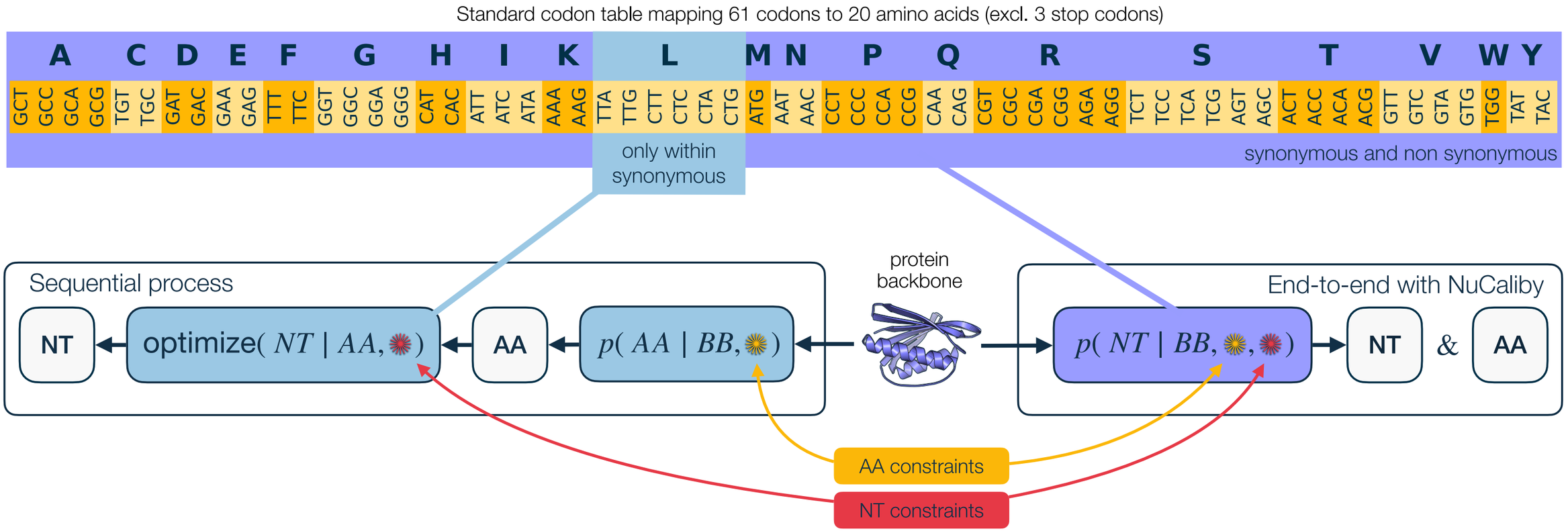
only within synonymous

synonymous and non synonymous



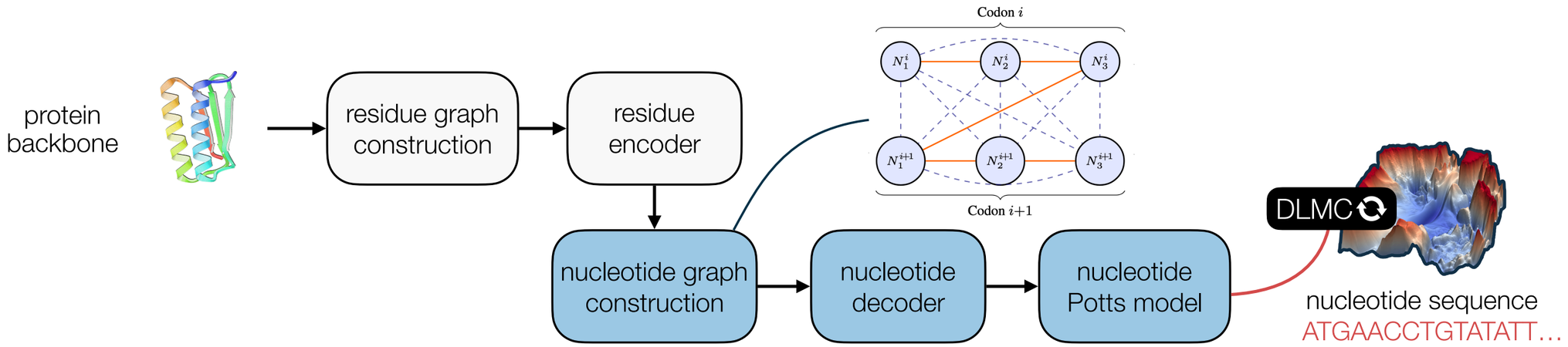
BB → protein backbone | NT → nucleotide sequence | AA → amino acid sequence

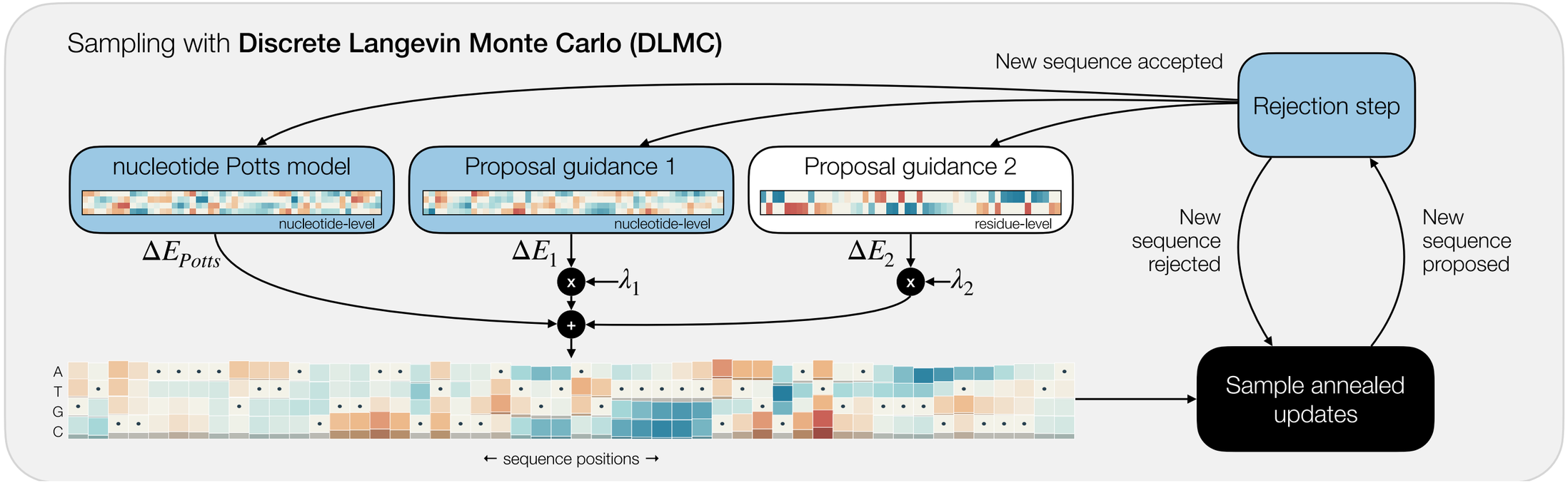
Inverse folding on the nucleotide level allows simultaneous optimization and trade-offs of both amino acid and RNA/DNA-level properties



BB → protein backbone | NT → nucleotide sequence | AA → amino acid sequence

A guidance signal can enter the proposal or rejection step



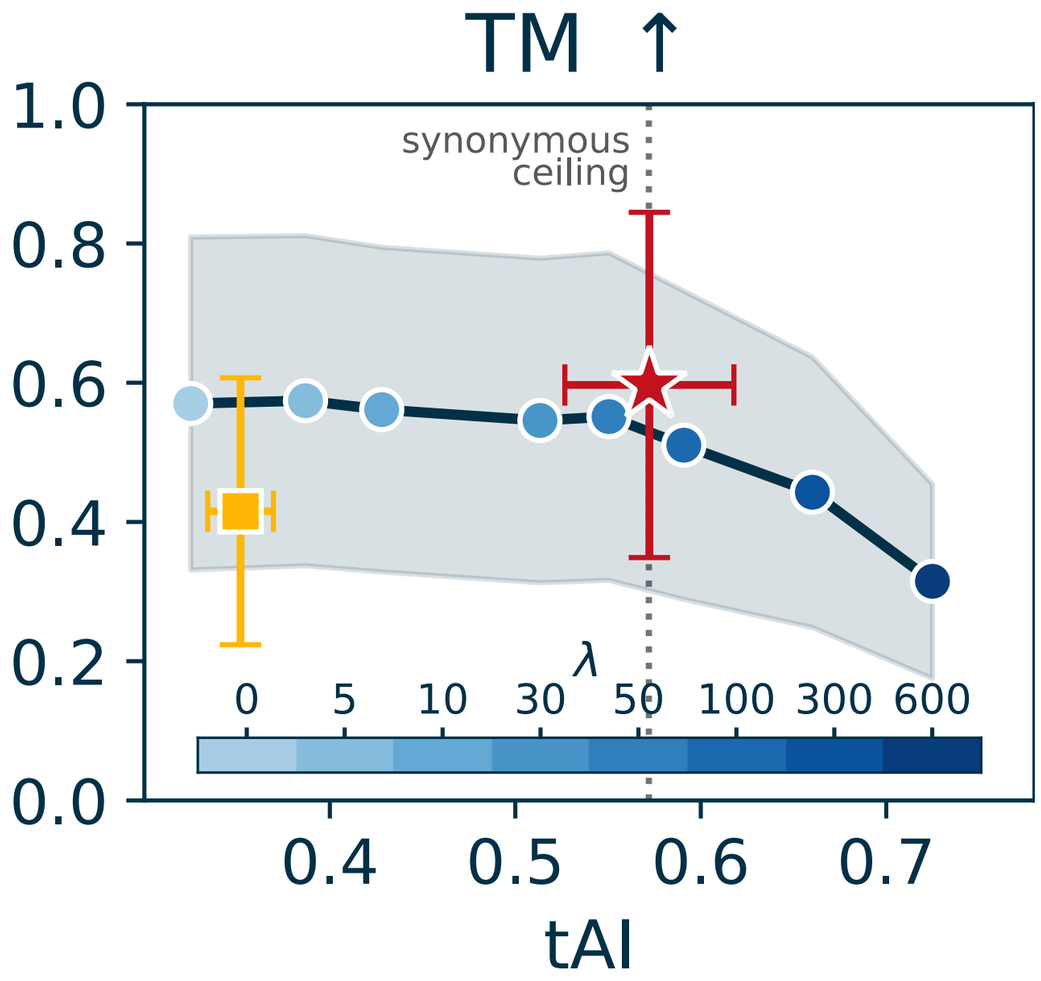


Nucleotide-level design preserves protein designability

183 test proteins, mean ± SD

Model	TM-score ↑	RMSD (Å) ↓
ProteinMPNN	0.509 ± 0.233	4.37 ± 1.81
Caliby	0.597 ± 0.248	3.85 ± 1.88
NuCaliby AA	0.593 ± 0.245	3.84 ± 1.85
NuCaliby NT	0.579 ± 0.245	3.85 ± 1.82

Trade-off between amino acid and nucleotide level properties



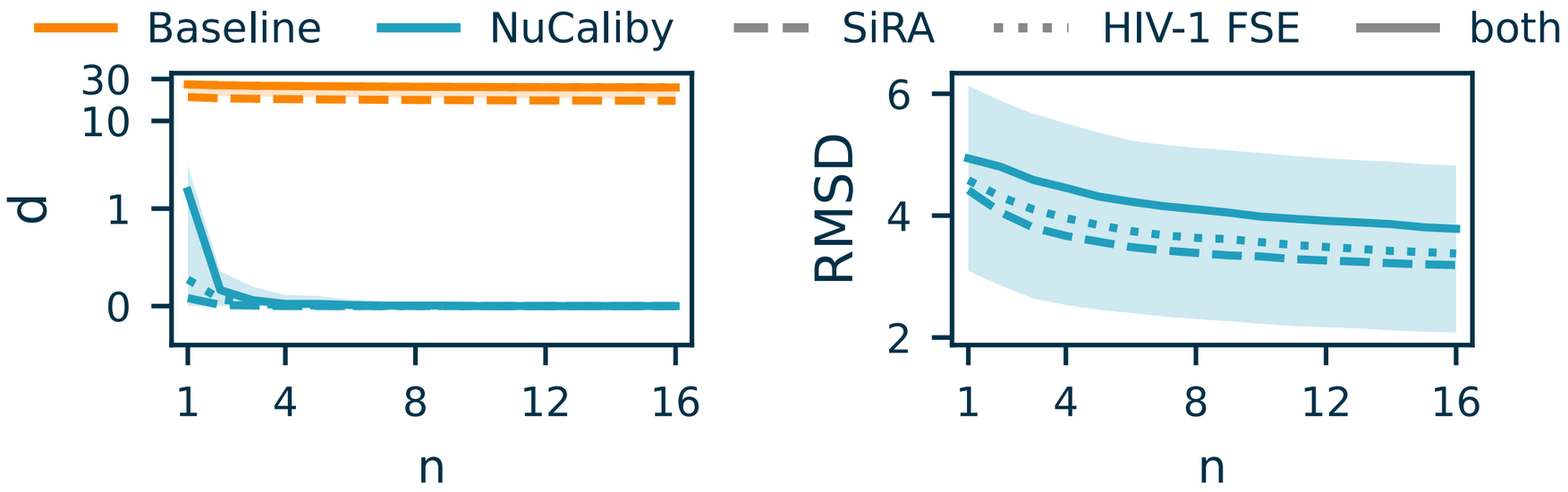
E. coli BL21(DE3)

NuCaliby$^{NT}_{tAI}$

Caliby + DNA Chisel

CodonMPNN

Include RNA motifs during sequence design



d: motif mismatches n: samples per backbone

Exact matches on all 183 backbones

Single motif: ≤2 samples Both motifs: ≤16 samples

Wet-lab experiments

8 protein backbones

24 designs per method and backbone

Up to 2 selected for synthesis

Expression in *E. coli* BL21(DE3)

