Edgar – a deep learning-based program for prediction of folding energy of nucleic acids

Artem R. Mukanov, Shamsudin S. Nasaev, Ivan I. Kuznetsov, Iosif V. Leibin, Alexander V. Veselovsky

Primary and Secondary structure

GUCCGUUCAUCCUUCGGGACGCAUGAGAUCUGACCAUGGAACGGGGGUCAGGU

...(((((..((.....))........[[[[[[[[...))))).]].]]]]]]

Edgar

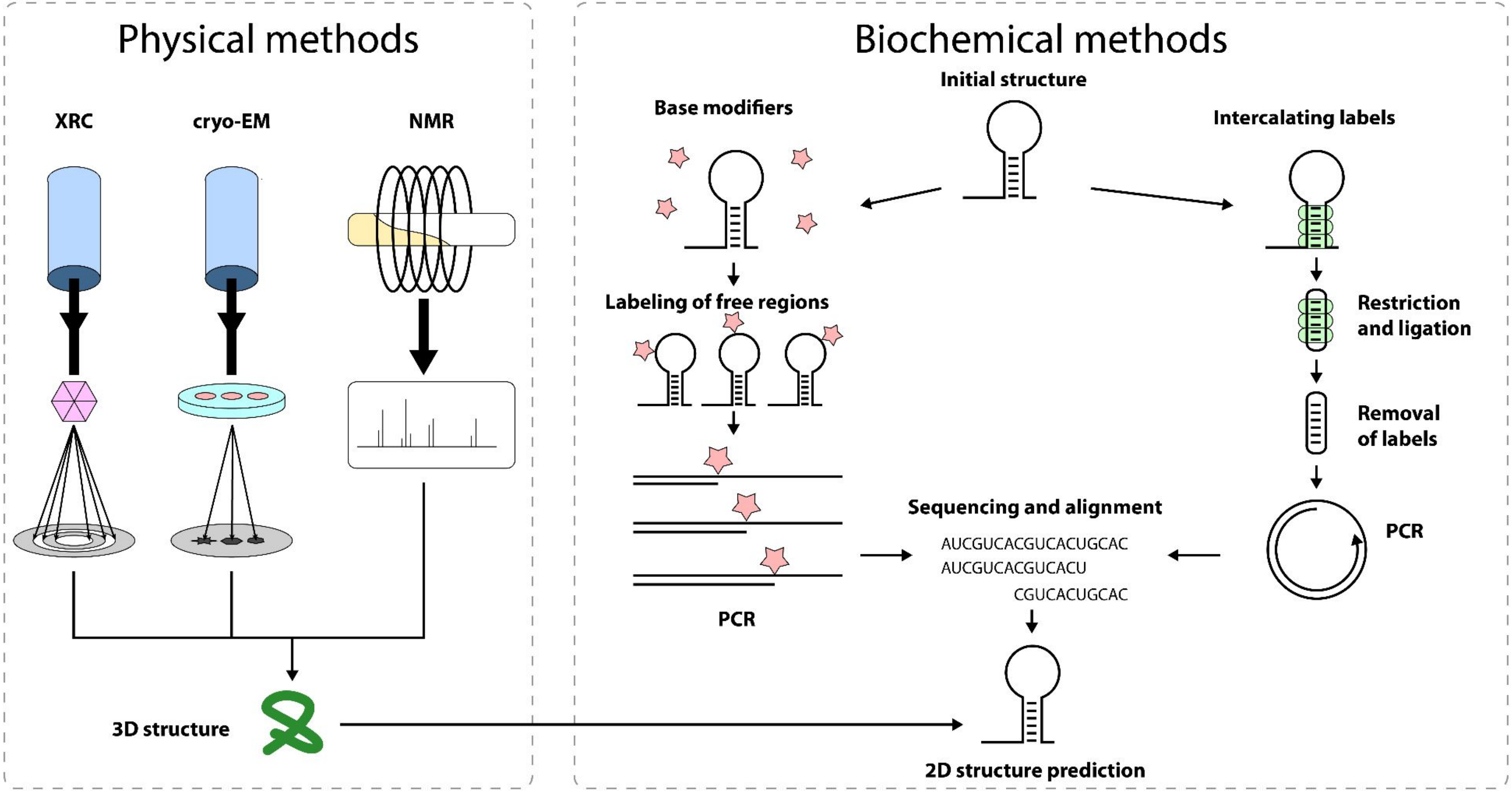
-245.91±56.37 kcal/mol

Graph of nucleotides and bonds between them

Predicted energy of folding

nasaev.shs@gmail.com

Experimental Methods for Determining RNA Structures



Physical methods

XRC

cryo-EM

NMR

3D structure

Biochemical methods

Initial structure

Base modifiers

Labeling of free regions

PCR

Intercalating labels

Restriction and ligation

Removal of labels

PCR

Sequencing and alignment

AUCGUCACGUCACUGCAC
AUCGUCACGUCACU
CGUCACUGCAC

2D structure prediction

Secondary Structure Prediction Methods

Homology-based

CentroidAlifold

MXSCARNA

TurboFold II

DL-based*

MXFold2

AliNA

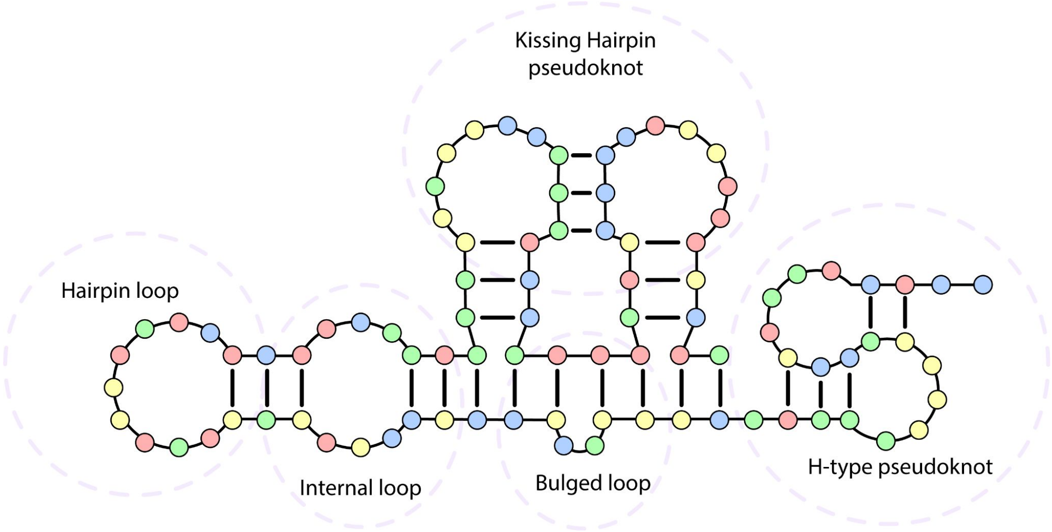
Ufold

Dynamic programming-based

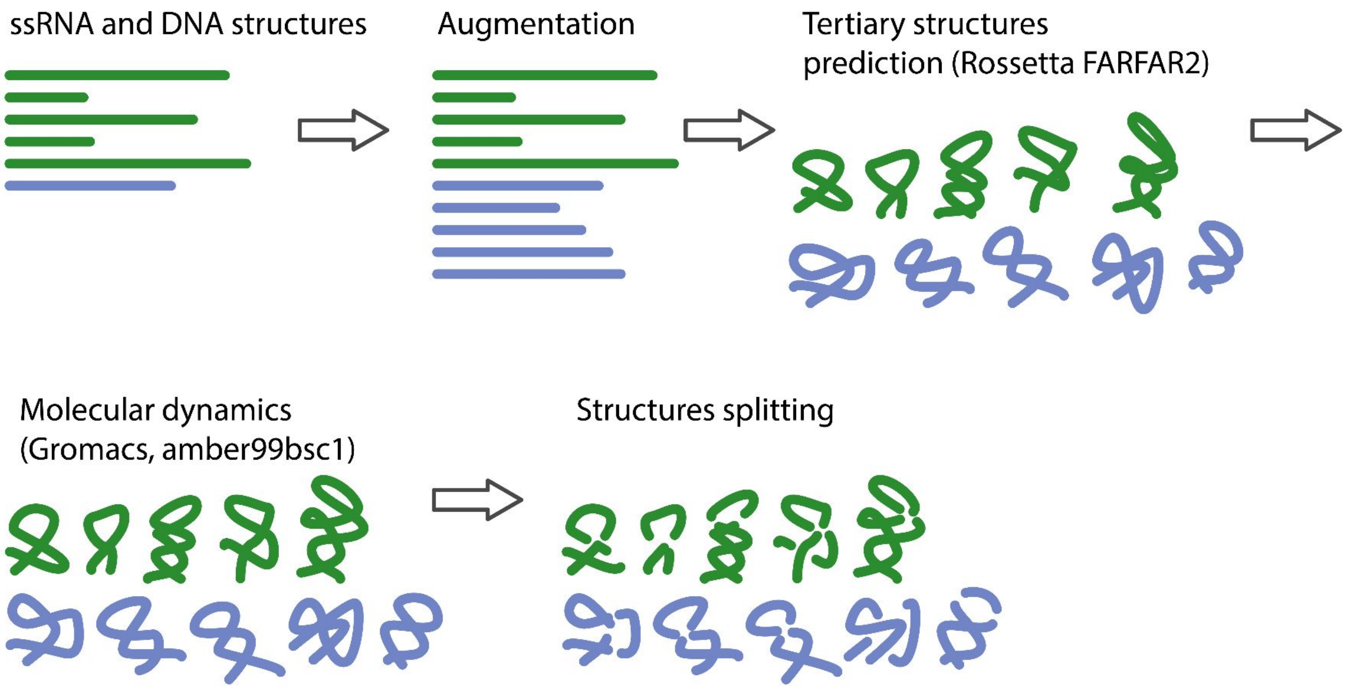
RNAFold

UNAFold

DotKnot



Data Preparation



ssRNA and DNA structures

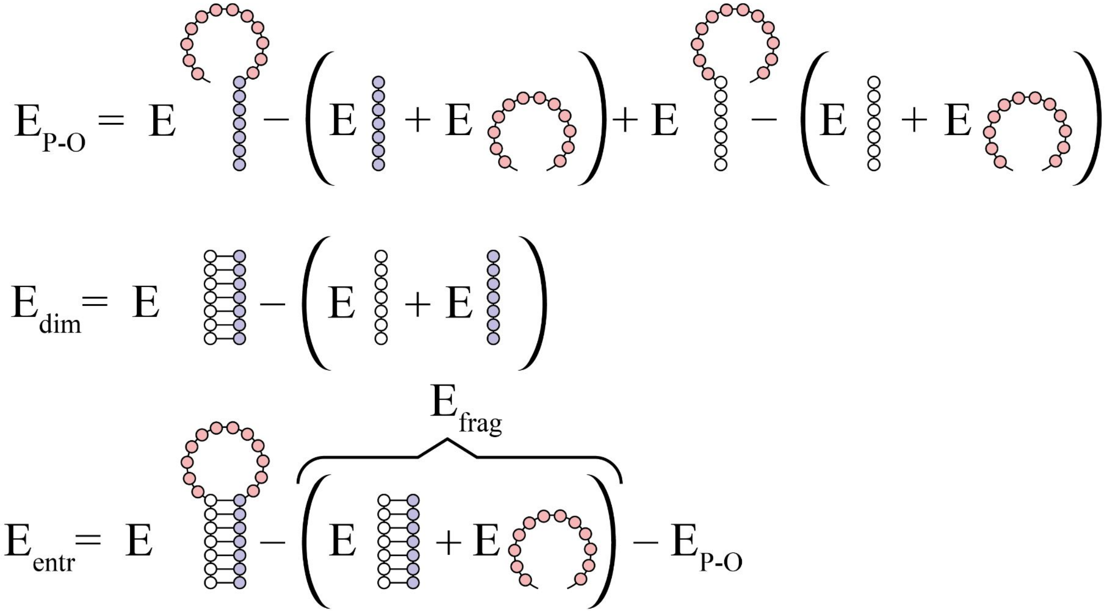
Augmentation

Tertiary structures prediction (Rossetta FARFAR2)

Molecular dynamics (Gromacs, amber99bsc1)

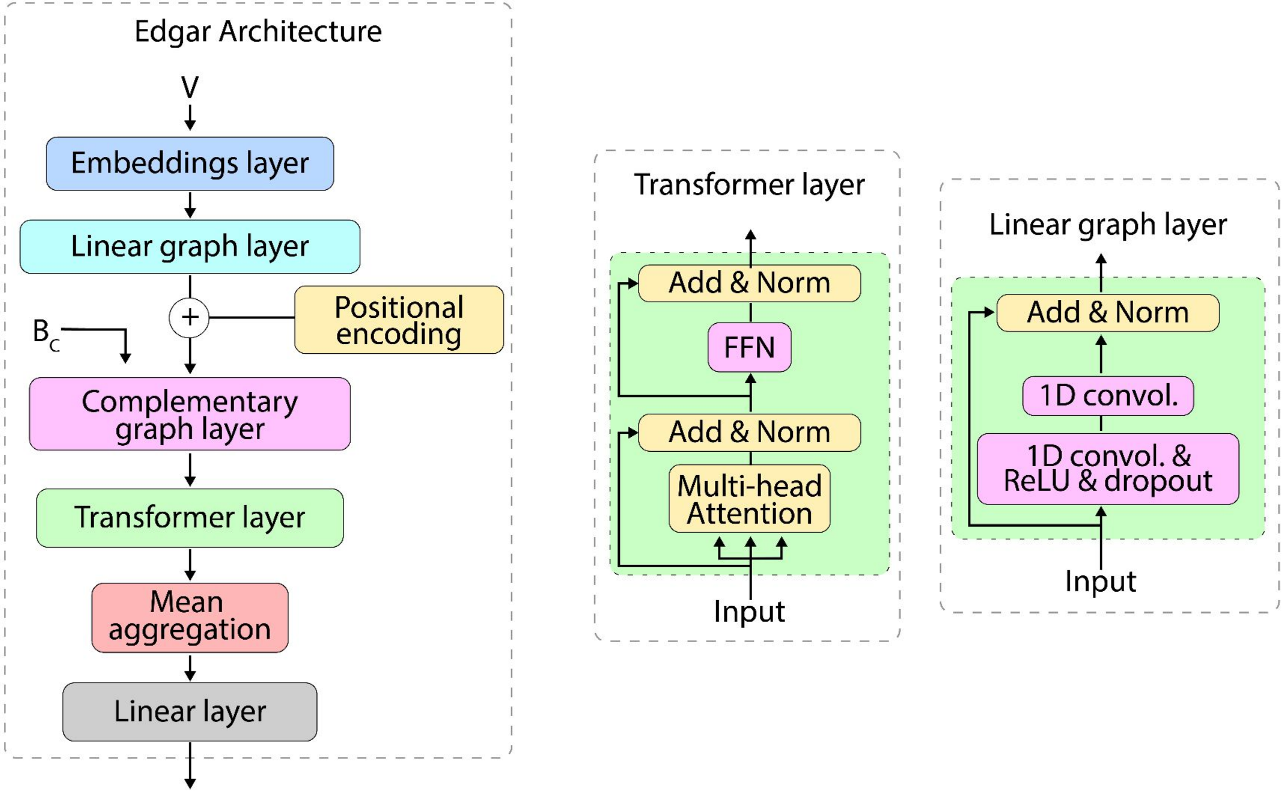
Structures splitting

Calculation of Free Energy Change



$$E_{fold} = E_{dim} + E_{entr}$$

Edgar Architecture



Edgar Architecture

V

Embeddings layer

Linear graph layer

+

Positional encoding

B_C

Complementary graph layer

Transformer layer

Mean aggregation

Linear layer

Transformer layer

Add & Norm

FFN

Add & Norm

Multi-head Attention

Input

Linear graph layer

Add & Norm

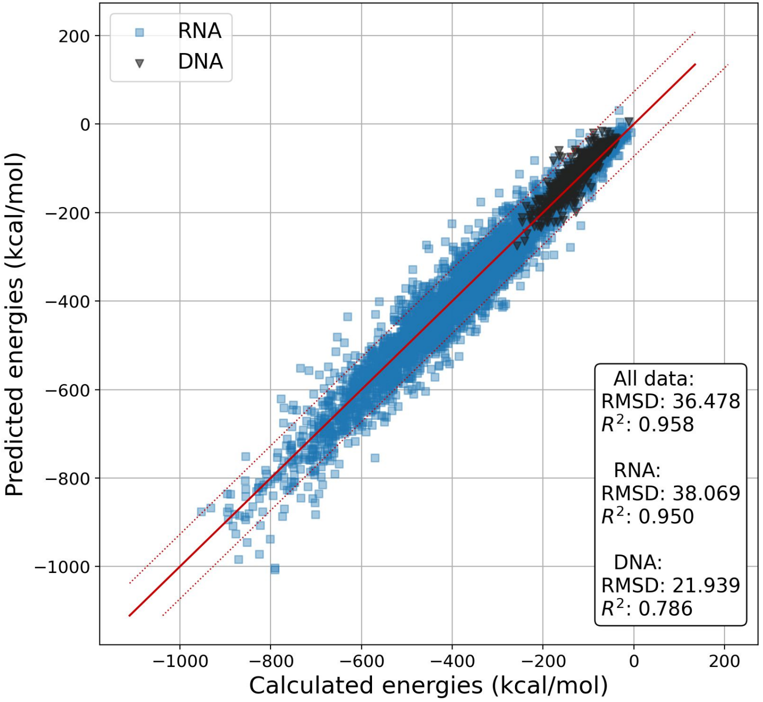
1D convol.

1D convol. & ReLU & dropout

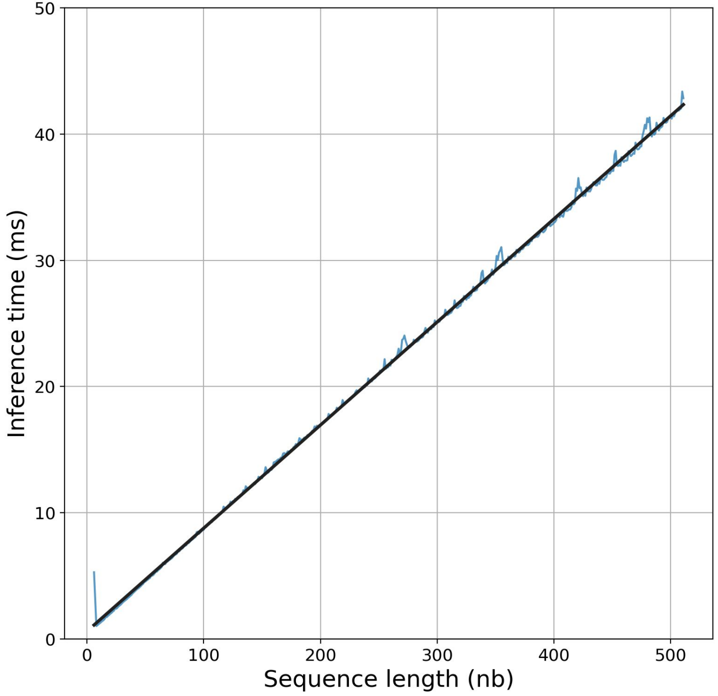
Input

The Model Quality and Inference Time

Comparison of Predicted and Calculated Energies



Evaluation of Inference Time Relative to Structure Length



Comparison of Experimental Data Sets

	Total number of structures	Proportion of favorable structures (%)
ArchiveII	2092	10.04
bpRNA-1m	11523	4.32
bpRNA-new	5197	8.4
RNAStrAlign	12388	12.0
NMR	370	**66.76**
X-ray	219	**76.71**

Edgar – a deep learning-based program for prediction of folding energy of nucleic acids

Artem R. Mukanov, Shamsudin S. Nasaev, Ivan I. Kuznetsov, Iosif V. Leibin, Alexander V. Veselovsky

Primary and Secondary structure

GUCCGUUCAUCCUUCGGGACGCAUGAGAUCUGACCAUGGAACGGGGGUCAGGU

...(((((..((.....))........[[[[[[[[...))))).]].]]]]]]

Edgar

-245.91±56.37 kcal/mol

Graph of nucleotides and bonds between them

Predicted energy of folding

nasaev.shs@gmail.com

Thank you for your attention!