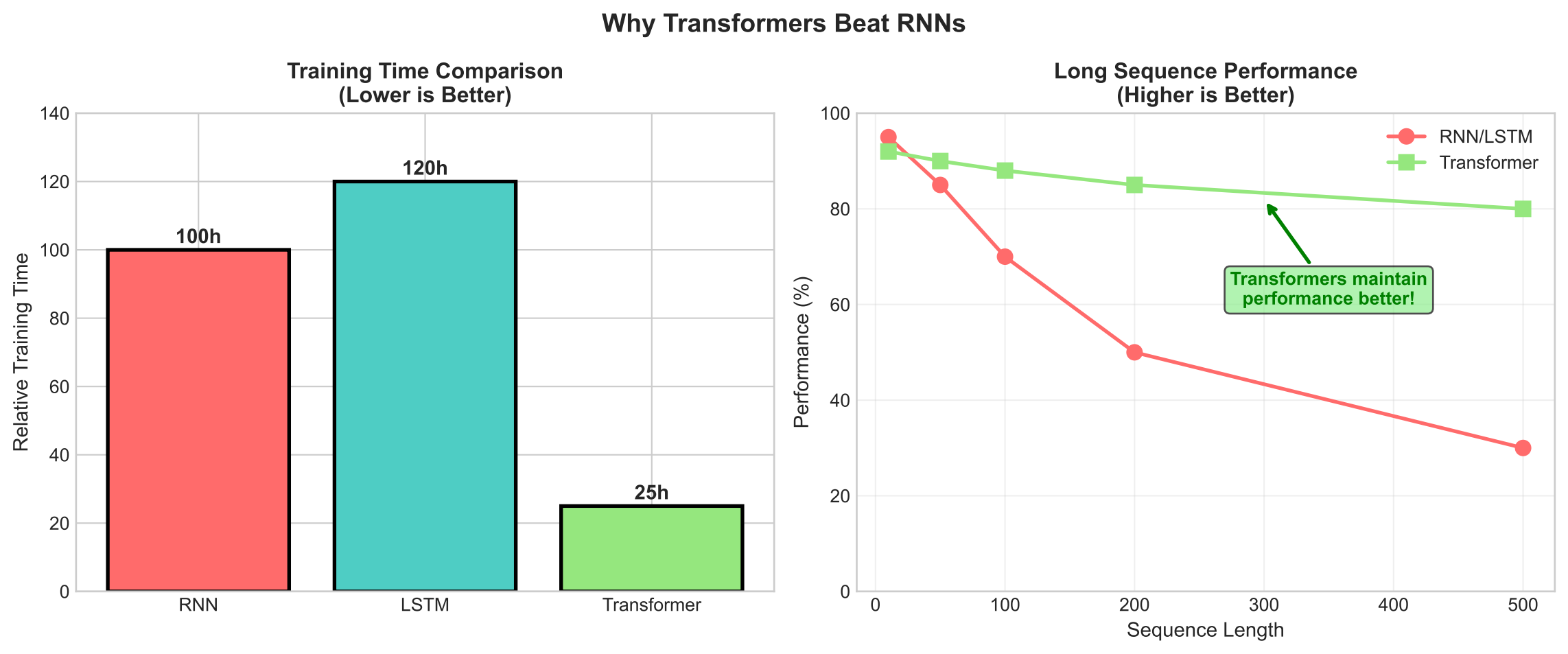
Why Transformers Beat RNNs
Training Time Comparison
(Lower is Better)
Relative Training Time
140
120
100
80
60
40
20
0
100h
120h
25h
RNN
LSTM
Transformer
Long Sequence Performance
(Higher is Better)
Performance (%)
100
80
60
40
20
0
RNN/LSTM
Transformer
Transformers maintain
performance better!
0
100
200
300
400
500
Sequence Length