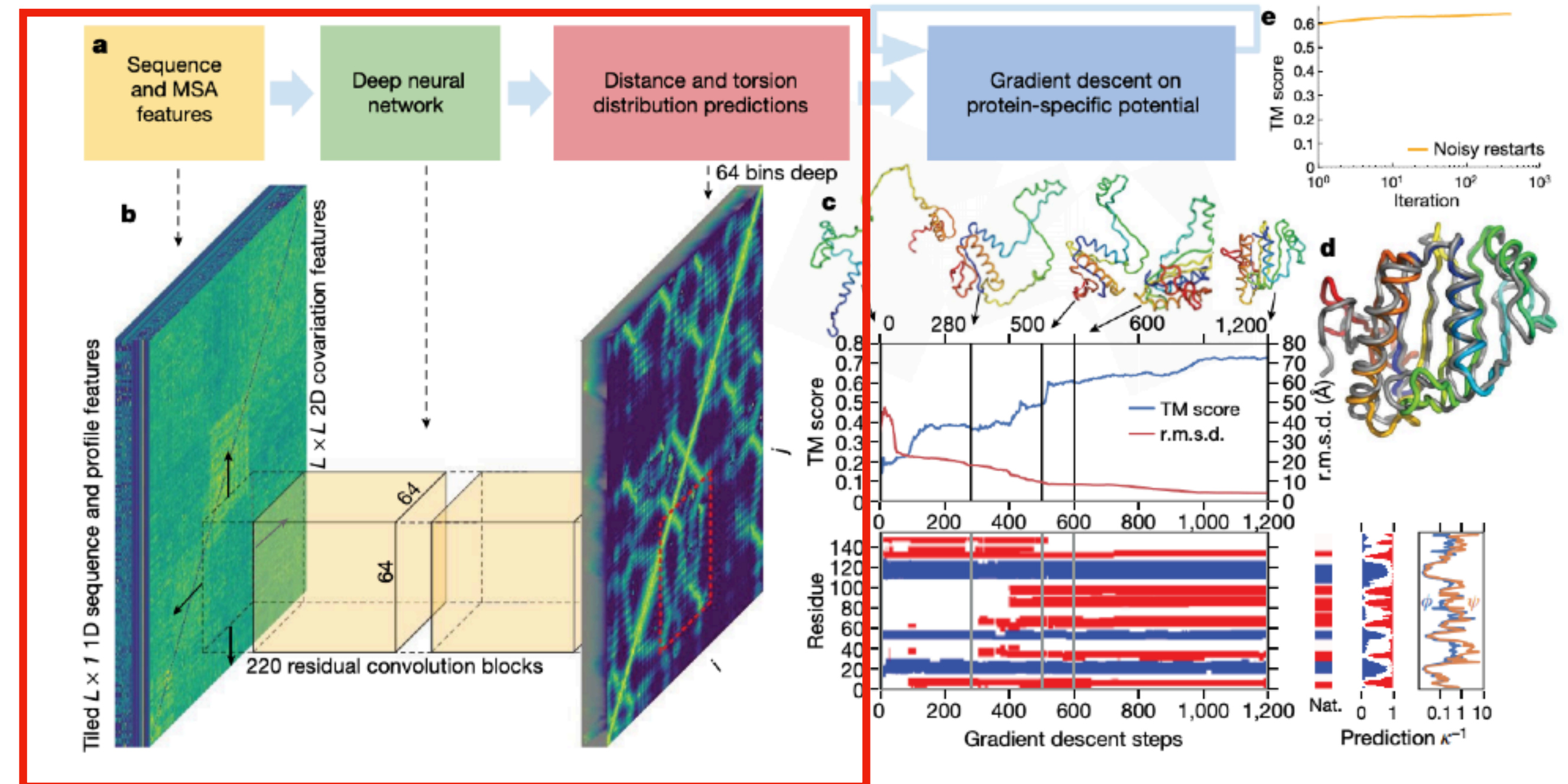


AlphaFold 1 System

- The AlphaFold 1 System [2] consists of multiple components
- Specifically, the ML sub-component contains a supervised learning task
- Input: amino acid sequence
- Output: a matrix of inter-residue distances and torsion angles



Machine learning portion of the AlphaFold 1 System

[2] Improved protein structure prediction using potentials from deep learning

Feature engineering from AA sequence

- Using domain knowledge to create numeric representations of the protein sequence
- These representations contain information that may indicate inter-residue distances
- Eg. from known protein structures, how often does residue A come in contact with residue B

QTKCEKKKCV CENCERSTYL
SERKTMKFNERDSHVVC DKTC

