nature

https://doi.org/10.1038/s41586-019-1923-7

Supplementary information

Improved protein structure prediction using potentials from deep learning

In the format provided by the authors and unedited

Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver,

Koray Kavukcuoglu & Demis Hassabis

Supplementary information for "Improved protein structure prediction using potentials from deep learning" Senior et al.

This document contains equations referenced by the Methods section.

Distance potentials The basic distance potential is computed as a sum over all residue pairs of the likelihood of the inter-residue distances:

$$V_{\text{distance}}(\mathbf{x}) = -\sum_{i,j,\ i\neq j} \log P(d_{ij} \mid \mathcal{S}, \text{MSA}(\mathcal{S})). \tag{1}$$

The distance potential with a reference state becomes:

$$V_{\text{distance}}(\mathbf{x}) = -\sum_{i,j,\ i\neq j} \log P(d_{ij} \mid \mathcal{S}, \text{MSA}(\mathcal{S})) - \log P(d_{ij} \mid \text{length}, \delta_{\alpha\beta}). \tag{2}$$

The torsions are modelled with a von Mises distribution for each residue:

$$V_{\text{torsion}}(\boldsymbol{\phi}, \boldsymbol{\psi}) = -\sum_{i} \log p_{\text{vonMises}}(\phi_i, \psi_i \mid \mathcal{S}, \text{MSA}(\mathcal{S})). \tag{3}$$

The total potential that we optimise is thus:

$$V_{\text{total}}(\boldsymbol{\phi}, \boldsymbol{\psi}) = V_{\text{distance}}(G(\boldsymbol{\phi}, \boldsymbol{\psi})) + V_{\text{torsion}}(\boldsymbol{\phi}, \boldsymbol{\psi}) + V_{\texttt{score2\_smooth}}(G(\boldsymbol{\phi}, \boldsymbol{\psi})). \tag{4}$$

The terms are weighted equally as determined by cross-validation.

Distogram lDDT lDDT is computed as follows.

$$\text{lDDT}_r = \frac{100}{4L} \sum_{t\in\{0.5,1,2,4\}} \sum_{i=1}^{L} \frac{\sum_{j,|i-j|\geq r, D_{ij}<15} \mathbb{1}(|D_{ij} - d_{ij}| < t)}{\sum_{j,|i-j|\geq r, D_{ij}<15} 1}. \tag{5}$$

We define Distogram lDDT (DLDDT) analogously:

$$\text{DLDDT}_r = \frac{100}{4L} \sum_{t\in\{0.5,1,2,4\}} \sum_{i=1}^{L} \frac{\sum_{j,|i-j|\geq r, D_{ij}<15} P(|D_{ij} - d_{ij}| < t \mid \mathcal{S}, \text{MSA}(\mathcal{S}))}{\sum_{j,|i-j|\geq r, D_{ij}<15} 1}. \tag{6}$$

Integrated gradients Given the expected value of the distance between any two residues I and J, $d^{I,J}(x)$, we can consider its derivatives with respect to the input features $x_{i,j,c}$, where i and j are residue indices and c is the feature channel index. The attribution function, as calculated using Integrated Gradients, of the expected distance between residues I and J with respect to the input features is then defined as

$$S^{I,J}_{i,j,c} = (x_{i,j,c} - x'_c) \int_{\alpha=0}^{1} d\alpha \frac{\partial d^{I,J}(\alpha x + (1-\alpha)x')}{\partial x_{i,j,c}}, \tag{7}$$

$$\text{s.t.} \sum_{i,j,c} S^{I,J}_{i,j,c} = d^{I,J}(x) - d^{I,J}(x'), \tag{8}$$

where x' is a reference set of features; in this case we average the input features spatially:

$$x'_c = \frac{1}{N^2} \sum_{i=0,j=0}^{N,N} x_{i,j,c}. \tag{9}$$

The derivatives of d can be calculated using backpropagation on the trained distogram network, and the integral over α is approximated as a numerical summation.