



# Protein Contact-Map Prediction using Variational Autoencoder

Presented by: Anowarul Kabir



# Background

- Protein structure prediction
  - Primary structure
  - Secondary structure
  - Tertiary or quaternary structure
- Protein primary sequence
- Distance matrix
  - Beta-beta carbon
  - Alpha-alpha for Glycine
- Contact-map
  - <12 angstrom

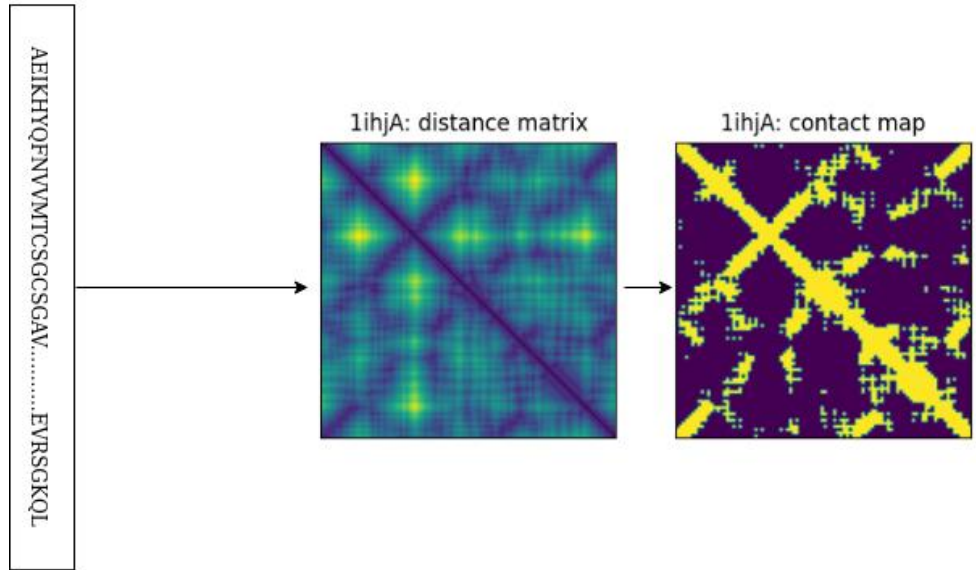


Fig 1: Primary sequence to contact-map generation

# Methodology

- Datasets
  - Train: 900
  - Validation: 300
  - Test: 300
- Model: Variational Autoencoder
- Loss computation
  - BCE
  - BCE + KLD

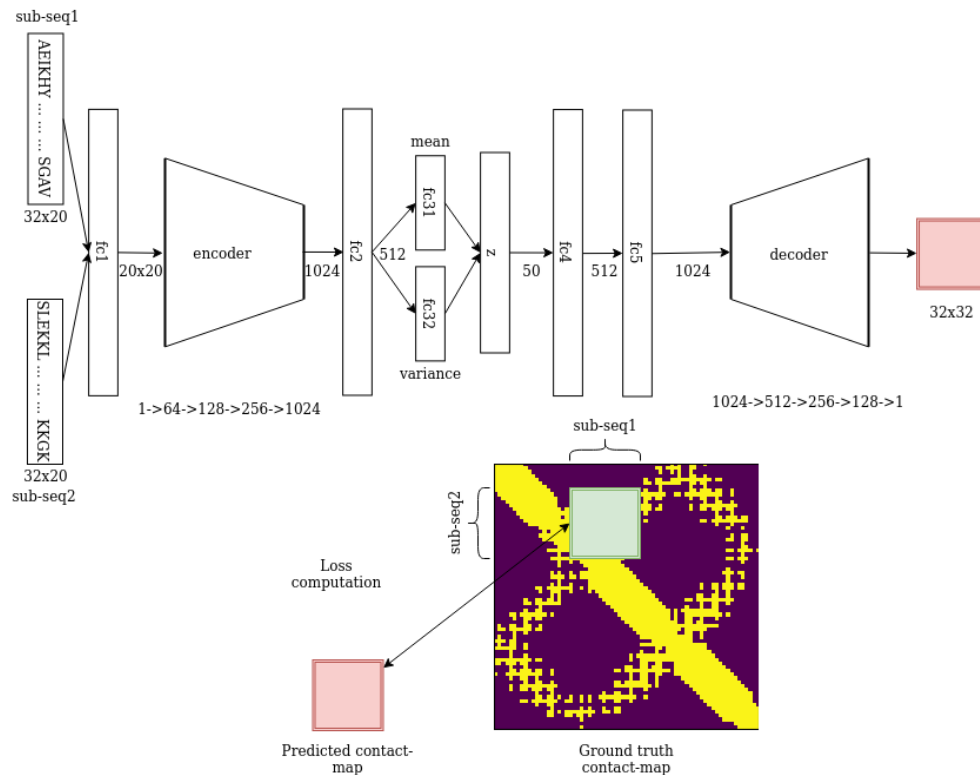


Fig 2: Model architecture and loss computation

# Models with Hyperparameters

$$BCE = L(y, y') = \sum_{i=0}^N (y_i * \log(y'_i) + (1 - y_i) * \log(1 - y'_i))$$

$$KLD = L(\mu, \sigma) = -.5 * \sum_{i=0}^N (1 + \log(\sigma_i) - \mu_i^2 - \sigma_i^2)$$

$$loss = BCE \text{ or } loss = BCE + KLD$$

Model id	Initial Learning rate	epochs	Batch size	Loss function	Final train loss
M14	1e-5	40	30	BCE	28064.9
M16	1e-6	70	40	BCE	28204.6
M17	1e-7	100	40	BCE	28234.7
M24	1e-5	40	40	BCE + KLD	28391.3
M25	1e-6	70	40	BCE + KLD	28391.3
M26	1e-7	100	40	BCE + KLD	28391.3

Table 1. Model with hyper-parameters

# Model Evaluation

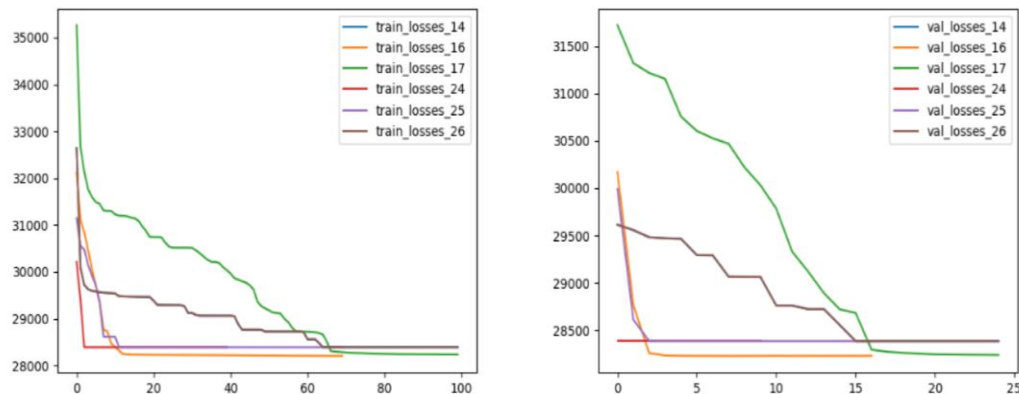


Figure 3. Train and validation losses

#	M14	M16	M17
Precision	<b>0.65637</b>	0.60495	0.27361
Recall	0.62372	0.69132	<b>0.80867</b>
f_score	<b>0.73700</b>	0.73691	0.61137

Table 2. Model evaluation

# Outputs

- Right most column: ground truth contact map
- Three predicted contact map using M14, M16 and M24 from left
- Each row corresponds to a protein

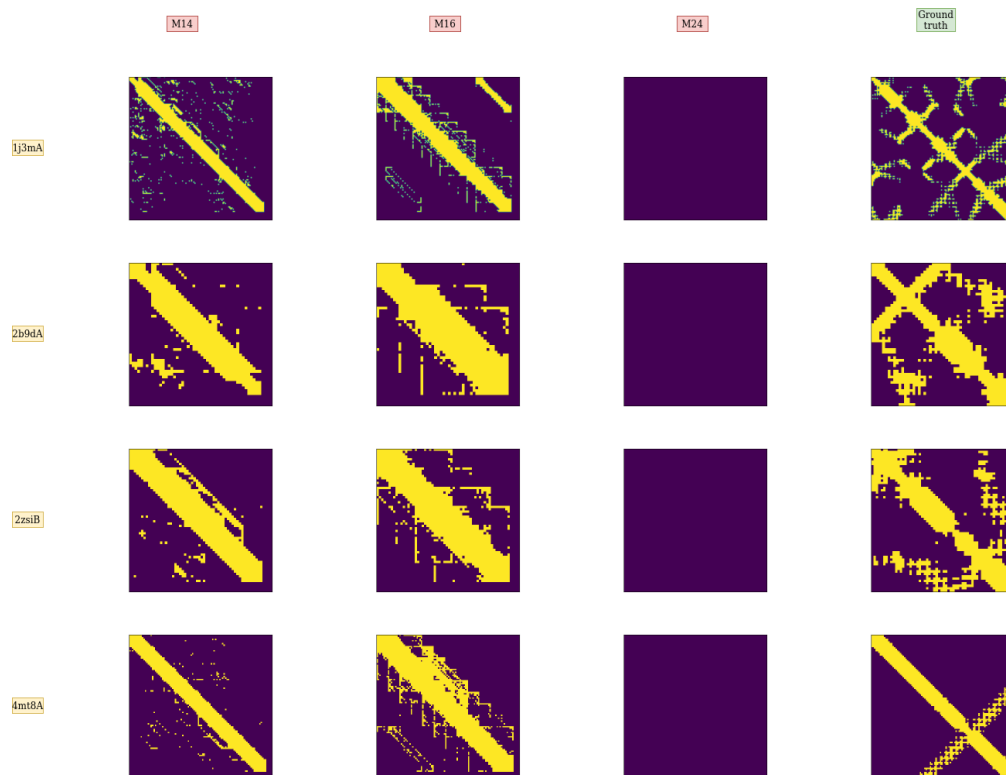


Fig 4: Sample outputs

# Miscellaneous

- Model is trained and evaluated in [GMU Argo Cluster](#).
  - 64GB Memory
  - 16GB GPU
- Project code could be found in [Github](#).
  - Github page describes how to run the code

The screenshot shows a GitHub repository interface. At the top, it displays statistics: 70 commits, 1 branch, 0 packages, 0 releases, and 1 contributor. Below this is a navigation bar with buttons for 'Branch: master', 'New pull request', 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The main content area shows a commit history table with columns for the commit message, the commit hash, and the time since the commit. The latest commit is by 'akabiraka' and is titled 'all drawio file added', made 2 days ago. Below the commit history is a section for the 'README.md' file, which includes the project title and a brief description of the project.

Commit Message	Commit Hash	Time
all drawio file added	6c995c2	2 days ago
some small changes while generating some images		4 days ago
nothing important		last month
nothing important		last month
previous work saved		26 days ago
all drawio file added		2 days ago
nothing important		last month
nothing important		last month
all inputs added		4 days ago
base pdb_id_list added		4 days ago
all output images added		4 days ago
all logs of model generation added		4 days ago
nothing important		last month
readme file edited		11 days ago
readme file edited		11 days ago

**Project title**

Protein Sequence to Contact-map Prediction using Variational-Autoencoder.

