**Supplementary Information for**

# Protein Contact Map Refinement for Improving Structure Prediction Using Generative Adversarial Networks

Sai Raghavendra Maddhuri Venkata Subramaniya[1], Genki Terashi[2], Aashish Jain[1], Yuki Kagaya[3], and Daisuke Kihara[*2,1]

[1] Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA

[2] Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA

[3] Graduate School of Information Sciences, Tohoku University, Japan

* Contact: dkihara@purdue.edu

**Supplementary Table S1A. Summary of CASP13 target classification.**

| CASP ID | Domain Length | Classification |
|---|---|---|
| T0949-D1 | 139 | FM/TBM |
| T1022s1-D1 | 156 | FM |
| T1021s3-D1 | 178 | FM |
| T0992-D1 | 107 | FM/TBM |
| T1000-D2 | 431 | FM |
| T0970-D1 | 97 | FM/TBM |
| T0997-D1 | 185 | FM/TBM |
| T1005-D1 | 326 | FM/TBM |
| T1015s1-D1 | 88 | FM |
| T0953s2-D3 | 93 | FM |
| T0969-D1 | 354 | FM |
| T0953s2-D2 | 111 | FM |
| T0978-D1 | 413 | FM/TBM |
| T0981-D3 | 203 | FM/TBM |
| T0986s1-D1 | 92 | FM/TBM |
| T0987-D1 | 185 | FM |
| T1019s1-D1 | 58 | FM/TBM |
| T0953s2-D1 | 44 | FM/TBM |
| T0975-D1 | 293 | FM |
| T0980s1-D1 | 105 | FM |
| T0958-D1 | 77 | FM/TBM |
| T0987-D2 | 207 | FM |
| T0957s2-D1 | 155 | FM |
| T0989-D1 | 134 | FM |
| T1008-D1 | 77 | FM/TBM |
| T0986s2-D1 | 155 | FM |
| T1017s2-D1 | 128 | FM |
| T0990-D1 | 76 | FM |

| | | |
|---|---|---|
| T0968s1-D1 | 119 | FM |
| T0950-D1 | 342 | FM |
| T0957s1-D1 | 108 | FM |
| T1001-D1 | 139 | FM |
| T0968s2-D1 | 116 | FM |
| T1010-D1 | 210 | FM |
| T1021s3-D2 | 101 | FM |
| T0990-D2 | 231 | FM |
| T0990-D3 | 213 | FM |
| T0998-D1 | 166 | FM |
| T0989-D2 | 112 | FM |
| T0981-D2 | 80 | FM |
| T0991-D1 | 111 | FM |
| T0963-D2 | 82 | FM |
| T0960-D2 | 84 | FM |

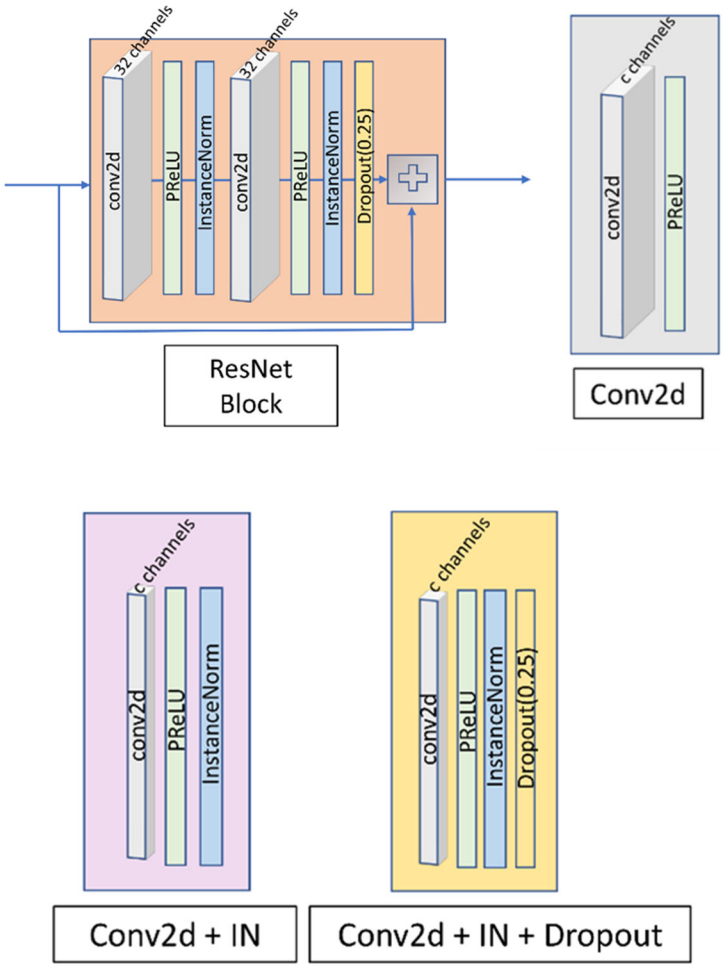
The summary of 43 contact prediction (RR) target domains in CASP13. The domain ID, length, and corresponding classification are shown. FM stands for free modeling, FM/TBM stands for free-modeling/template-based modeling. These are all the RR targets, as shown at the CASP13 RR analysis website https://predictioncenter.org/casp13/rrc_avrg_results.cgi (click both FM and FM/TBM in the checkbox). The difficulty classification is taken from the official CASP13 website, https://predictioncenter.org/casp13/domains_summary.cgi.

**Supplementary Table S1B. Summary of CASP14 target classification.**

| CASP ID | Domain Length | Classification |
|---|---|---|
| T1026-D1 | 146 | TBM-hard |
| T1027-D1 | 99 | FM |
| T1029-D1 | 125 | FM |
| T1031-D1 | 95 | FM |
| T1033-D1 | 100 | FM |
| T1035-D1 | 102 | FM/TBM |
| T1037-D1 | 404 | FM |
| T1038-D1 | 114 | FM |
| T1038-D2 | 76 | FM/TBM |
| T1039-D1 | 161 | FM |
| T1040-D1 | 130 | FM |
| T1041-D1 | 242 | FM |
| T1042-D1 | 276 | FM |
| T1043-D1 | 148 | FM |
| T1045s2-D1 | 166 | TBM-hard |
| T1046s1-D1 | 72 | FM/TBM |
| T1046s2-D1 | 141 | TBM-hard |
| T1047s1-D1 | 211 | FM |
| T1047s2-D1 | 147 | FM/TBM |
| T1047s2-D3 | 116 | FM/TBM |
| T1049-D1 | 134 | FM |
| T1052-D3 | 80 | FM/TBM |
| T1053-D1 | 405 | FM/TBM |
| T1053-D2 | 171 | FM/TBM |
| T1055-D1 | 122 | FM/TBM |
| T1058-D1 | 221 | FM/TBM |
| T1060s2-D1 | 297 | TBM-hard |
| T1061-D1 | 464 | FM/TBM |
| T1061-D2 | 271 | FM |
| T1064-D1 | 92 | FM |
| T1065s1-D1 | 119 | TBM-hard |
| T1065s2-D1 | 98 | FM/TBM |
| T1067-D1 | 221 | TBM-hard |
| T1070-D1 | 76 | FM |
| T1074-D1 | 132 | FM |
| T1078-D1 | 129 | TBM-hard |
| T1080-D1 | 133 | FM/TBM |
| T1082-D1 | 75 | FM/TBM |

| | | |
|---|---|---|
| T1083-D1 | 92 | TBM-hard |
| T1084-D1 | 71 | TBM-hard |
| T1085-D2 | 182 | FM/TBM |
| T1090-D1 | 191 | FM |
| T1093-D1 | 141 | FM |
| T1093-D3 | 106 | FM |
| T1094-D1 | 277 | TBM-hard |
| T1094-D2 | 207 | FM |
| T1096-D1 | 255 | FM |
| T1096-D2 | 171 | FM |
| T1099-D1 | 178 | TBM-hard |

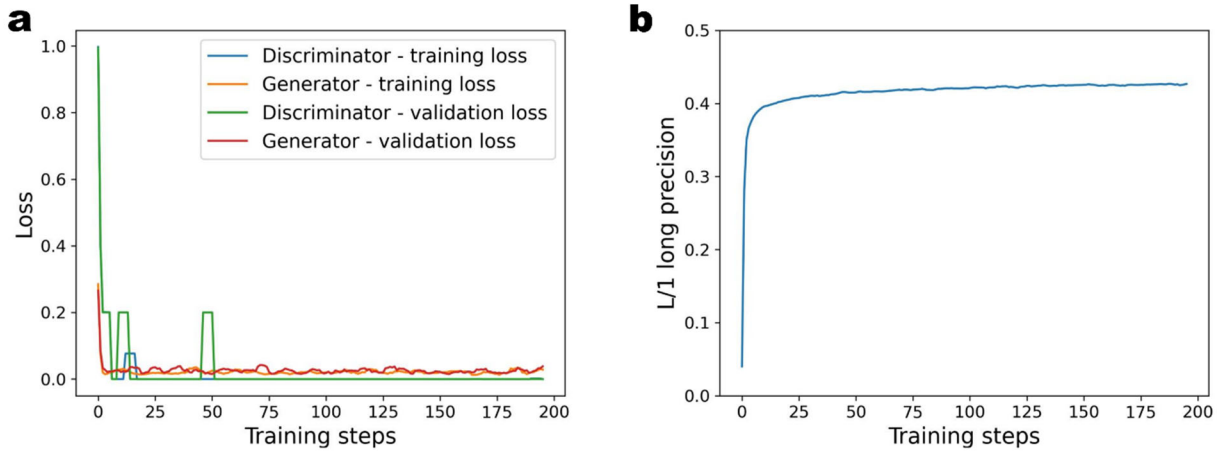
The summary of 49 contact prediction (RR) target domains in CASP14. The domain ID, length, and corresponding classification are shown. FM stands for free modeling, and TBM stands for template-based modeling. These are all the RR targets, as shown at the CASP14 RR analysis website https://predictioncenter.org/casp14/rrc_avrg_results.cgi (click both FM and FM/TBM in the checkbox). The difficulty classification is taken from the official CASP14 website, https://predictioncenter.org/casp14/domains_summary.cgi




**Supplementary Figure S1. Detailed descriptions of blocks used in ContactGAN.**

The panel above shows the ResNet block, and Conv2d block, which are used in the generator. The blue arrows indicate skip connections, which connect the input of a ResNet block. A plus sign is an operator that simply adds two matrices. Each ResNet block contains 2 convolutional layers with 32 channels and skip connections, dropout (with a dropout probability 0.25), PReLU activations, and Instance Normalization as we used a batch size of 1. In the series of the ResNet blocks, we used dilation filters of 1, 2, and 4 dilations for the first 3 blocks, respectively and repeated this unit of 3 blocks in subsequent layers. The final convolutional layer uses a smaller kernel size of 3 with 32 output channels. The last sigmoid layer outputs a contact probability for each pixel in the 2D contact map. We used a stride of 1 uniformly in all the filters in all the layers. The second panel in the top row is the conv2d block, which combines convolutional layers with 32 channels (c=32).

The bottom panel shows blocks used in the discriminator, which contains Conv2d+IN, and Conv2D+IN+Dropout. There are 4 of these combinations connected, where channels c increases by a factor of 2 starting from 32, 64, 128, 256, then followed by Conv2d+IN with c = 512. Then, in Figure 1, finally, these are connected to conv2d with 512 channels (gray).




**Supplementary Figure S2. Loss and L/1 long precision values plotted against training steps.**

**a,** training and validation losses of the generator and discriminator networks while training 3-channel CCMpred+DeepContact+DeepCov are shown. Loss values recorded after every 1000 iterations of training with each epoch having 4962 iterations (i.e. the size of the training dataset) and are smoothed using a sliding window of size 5. **b,** L/1 long precision values on the validation dataset plotted after every 1000 iterations of training and smoothed using a sliding window of size 5. For this network with 3-channels, the network model that achieved the highest L/1 long precision on the validation set at iteration 3000 of epoch 50 was used for testing datasets.

**Supplementary Table S2A. Computational Cost for Training ContactGAN.**

| Contact Maps | Time | Specs |
|---|---|---|
| CCMpred | 7 hrs 32 mins | |
| DeepCov | 7 hrs 13 mins | |
| DeepContact | 8 hrs 01 mins | |
| CCMpred + DeepCov | 9 hrs 36 mins | 1 Nvidia GeForce RTX 2080 |
| CCMpred + DeepContact | 10 hrs 45 mins | GPU with 12GB RAM |
| DeepContact + DeepCov | 10 hrs 11 mins | |
| CCMpred + DeepContact +DeepCov | 12 hrs 18 mins | |
| trRosetta (3-channel) | 18 hrs 33 mins | 1 Nvidia TITAN RTX GPU with 24 GB RAM |

Summary of computational costs and GPU specifications used for ContactGAN training shown for different methods.

**Supplementary Table S2B. Computational Cost for Inference and Modeling**

| Targets | Length | Time (seconds) | | | Specs | | |
|---|---|---|---|---|---|---|---|
| | | Prediction | | Modeling | Inference | | Modeling |
| | | trRosetta | Other | | trRosetta | Other | |
| T0949-D1 | 139 | 4.79 | 4.74 | 355.8 | 1 Nvidia TITAN RTX GPU RAM: 24GB | 1 Nvidia GeForce RTX 2080Ti GPU RAM: 12GB | Sky Lake CPU @ 2.60GHz CPU RAM: 96 GB |
| T1000-D2 | 451 | 4.86 | 4.81 | 4744.7 | | | |
| T0969-D1 | 354 | 4.83 | 4.79 | 1930.5 | | | |
| T0953s2-D1 | 44 | 4.79 | 4.75 | 569.5 | | | |
| T1010-D1 | 210 | 4.81 | 4.74 | 362.7 | | | |

Summary of computational costs and GPU specifications used for ContactGAN prediction and CPU specifications used for protein structure modeling shown on 5 CASP13 targets of varying domain lengths. The computational time for predicting contacts by trRosetta and the maximum among the other three methods and their combinations are separately shown. Modeling was performed with pyRosetta as described in Supplementary Note 2. For modeling, the specifications are shown per node of Purdue RCAC Brown cluster (https://www.rcac.purdue.edu/compute/brown/). We use only one CPU core per node.

**Supplementary Table S3A. Summary of improvement by ContactGAN for individual methods on the validation dataset.**

| Method | L/10[a)] | | | L/5 | | | L/2 | | | L | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Short | Med | Long | Short | Med | Long | Short | Med | Long | Short | Med | Long |
| **CCMpred[b)]** | 0.362 | 0.399 | 0.450 | 0.273 | 0.300 | 0.388 | 0.177 | 0.193 | 0.278 | 0.127 | 0.133 | 0.193 |
| | 0.560 | 0.564 | 0.578 | 0.475 | 0.484 | 0.529 | 0.322 | 0.347 | 0.433 | 0.208 | 0.235 | 0.333 |
| **Dcv** | 0.711 | 0.655 | 0.677 | 0.602 | 0.566 | 0.608 | 0.406 | 0.392 | 0.472 | 0.253 | 0.265 | 0.350 |
| | 0.719 | 0.664 | 0.668 | 0.610 | 0.571 | 0.614 | 0.407 | 0.400 | 0.487 | 0.253 | 0.268 | 0.361 |
| **Dct** | 0.628 | 0.648 | 0.645 | 0.526 | 0.559 | 0.585 | 0.360 | 0.397 | 0.463 | 0.235 | 0.269 | 0.351 |
| | 0.710 | 0.677 | 0.664 | 0.605 | 0.586 | 0.611 | 0.406 | 0.417 | 0.493 | 0.253 | 0.275 | 0.374 |

Average precision values on the validation dataset of 301 proteins are shown.
**a)** L/k shows precision values when top L/k, k=10, 5, 2, and 1, contact predictions with the highest probabilities were considered. L is the length of the protein. The columns Short consider short-range contacts (residue pairs with a sequence separation of 6-11 residues), Med consider medium-range contacts (residue pairs with a sequence separation of 12-23 residues), and Long consider long-range contacts (residue pairs separated by more than 23 residues). **b)** Each result show two values: up, original average precision by the existing method; bottom, average precision of refined contact maps by ContactGAN. Dcv, DeepCov; Dct, DeepContact.

**Supplementary Table S3B. Summary of improvement by ContactGAN for individual methods on the CASP13 dataset.**

| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg[a)] | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **CCMpred** | 0.322 | 0.291 | 0.282 | 0.232 | 0.209 | 0.169 | 0.164 | 0.121 |
| | 0.451 | 0.352 | 0.430 | 0.332 | 0.350 | 0.266 | 0.287 | 0.217 |
| **Dcv** | 0.565 | 0.408 | 0.505 | 0.362 | 0.394 | 0.287 | 0.320 | 0.231 |
| | 0.633 | 0.451 | 0.584 | 0.420 | 0.457 | 0.329 | 0.368 | 0.250 |
| **Dct** | 0.652 | 0.446 | 0.601 | 0.426 | 0.475 | 0.336 | 0.382 | 0.267 |
| | 0.686 | 0.510 | 0.627 | 0.466 | 0.524 | 0.362 | 0.408 | 0.283 |

Average precision values on the CASP13 dataset of 43 protein domains are shown.
**a)** The columns Med+Lg consider medium and long-range contacts (residue pairs separated by more than 11 residues). Other notations are the same as Table S1.

**Supplementary Table S4A. Summary of improvement by ContactGAN for multi-channel methods on the validation dataset.**

| Method | L/10 | | | L/5 | | | L/2 | | | L | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Short | Med | Long | Short | Med | Long | Short | Med | Long | Short | Med | Long |
| **C+Dcv** | 0.711 | 0.655 | 0.677 | 0.602 | 0.566 | 0.608 | 0.406 | 0.392 | 0.472 | 0.253 | 0.265 | 0.350 |
| | 0.741 | 0.706 | 0.720 | 0.635 | 0.618 | 0.662 | 0.426 | 0.437 | 0.541 | 0.261 | 0.287 | 0.410 |
| **C+Dct** | 0.628 | 0.648 | 0.645 | 0.526 | 0.559 | 0.585 | 0.360 | 0.397 | 0.463 | 0.235 | 0.269 | 0.351 |
| | 0.731 | 0.691 | 0.684 | 0.627 | 0.605 | 0.637 | 0.416 | 0.433 | 0.526 | 0.259 | 0.286 | 0.405 |
| **Dcv+Dct** | 0.628 | 0.648 | 0.645 | 0.526 | 0.559 | 0.585 | 0.360 | 0.397 | 0.463 | 0.235 | 0.269 | 0.351 |
| | 0.759 | 0.709 | 0.721 | 0.650 | 0.616 | 0.663 | 0.431 | 0.439 | 0.541 | 0.265 | 0.291 | 0.411 |
| **C+Dcv+Dct** | 0.628 | 0.648 | 0.645 | 0.526 | 0.559 | 0.585 | 0.360 | 0.397 | 0.463 | 0.235 | 0.269 | 0.351 |
| | 0.766 | 0.727 | 0.738 | 0.656 | 0.640 | 0.677 | 0.438 | 0.451 | 0.555 | 0.267 | 0.294 | 0.426 |

The results shown are on the validation dataset. As input, maps from two or three methods were used. To be able to take multiple contact maps, the ContactGAN network architecture was modified to a two-channel or a three-channel GAN. C, CCMpred; Dcv, DeepCov; Dct, DeepContact. For example, C+Dcv indicates that the contact maps from CCMpred and DeepCov were used as input. The highest accuracy among the combined methods was shown as the original results.

**Supplementary Table S4B. Summary of improvement by ContactGAN for multi-channel methods on the CASP13 dataset.**

| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **C+Dcv** | 0.595 | 0.472 | 0.530 | 0.391 | 0.406 | 0.303 | 0.326 | 0.238 |
| | 0.680 | 0.539 | 0.638 | 0.494 | 0.511 | 0.365 | 0.402 | 0.284 |
| **C+Dct** | 0.660 | 0.468 | 0.604 | 0.432 | 0.478 | 0.341 | 0.385 | 0.270 |
| | 0.717 | 0.501 | 0.667 | 0.459 | 0.522 | 0.371 | 0.420 | 0.299 |
| **Dcv+Dct** | 0.660 | 0.468 | 0.604 | 0.432 | 0.478 | 0.341 | 0.385 | 0.270 |
| | 0.750 | 0.554 | 0.661 | 0.497 | 0.549 | 0.406 | 0.438 | 0.310 |
| **C+Dcv+Dct** | 0.660 | 0.468 | 0.604 | 0.432 | 0.478 | 0.341 | 0.385 | 0.270 |
| | 0.696 | 0.582 | 0.651 | 0.526 | 0.540 | 0.402 | 0.437 | 0.314 |

The results shown are on the CASP13 dataset.

**Supplementary Table S5A. Summary of Improvement by ContactGAN for 3-channel trRosetta on the validation dataset.**

| Method | L/10 | | | L/5 | | | L/2 | | | L | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Short** | **Med** | **Long** | **Short** | **Med** | **Long** | **Short** | **Med** | **Long** | **Short** | **Med** | **Long** |
| **trRosetta (1e-3)** | 0.862 | 0.854 | 0.864 | 0.756 | 0.767 | 0.827 | 0.502 | 0.557 | 0.729 | 0.294 | 0.353 | 0.583 |
| **trRosetta (1e-1)** | 0.858 | 0.846 | 0.858 | 0.747 | 0.766 | 0.821 | 0.498 | 0.554 | 0.720 | 0.293 | 0.351 | 0.577 |
| **trRosetta (1)** | 0.850 | 0.837 | 0.846 | 0.741 | 0.754 | 0.807 | 0.494 | 0.544 | 0.711 | 0.292 | 0.346 | 0.567 |
| **ContactGAN[a)]** | **0.866** | **0.858** | **0.868** | **0.759** | **0.772** | **0.834** | **0.504** | **0.560** | **0.732** | **0.295** | **0.355** | **0.585** |
| **(p-value)[b)]** | 0.138 | 0.183 | 0.191 | <u>0.074</u> | **0.035** | **0.023** | **0.044** | **0.028** | **0.024** | 0.152 | **0.006** | **0.014** |
| **9-layer CGAN[c)]** | 0.860 | 0.845 | 0.858 | 0.753 | 0.760 | 0.825 | 0.499 | 0.552 | 0.728 | 0.290 | 0.352 | 0.580 |
| **G-only CGAN[d)]** | 0.847 | 0.845 | 0.844 | 0.739 | 0.758 | 0.814 | 0.490 | 0.550 | 0.722 | 0.286 | 0.346 | 0.575 |

The results shown are on the validation dataset. Top three rows show the average precision by trRosetta which used a MSA with the E-value cutoff specified in the parenthesis. 1e-3: 0.001, 1e-1: 0.1. The latter rows are the average precision of refined contact maps by ContactGAN using three trRosetta maps with the different E-value cutoffs as input.

a) ContactGAN with 18 ResNet blocks was used in the generator. For each metric, the value of the model with the best validation performance of the particular metric was shown and its p-value was evaluated. Compared with the results by trRosetta (1e-3), values are shown in **bold** if they are larger.

b) The corresponding p-value of ContactGAN[a)] results from the t-test statistics. p-values are shown in **bold** if they are statistically significant considering a p-value level of 0.05 and underlined when they are significant considering a p-value level of 0.1.

c) The same ContactGAN model as the other methods (9 ResNet blocks in the generator)

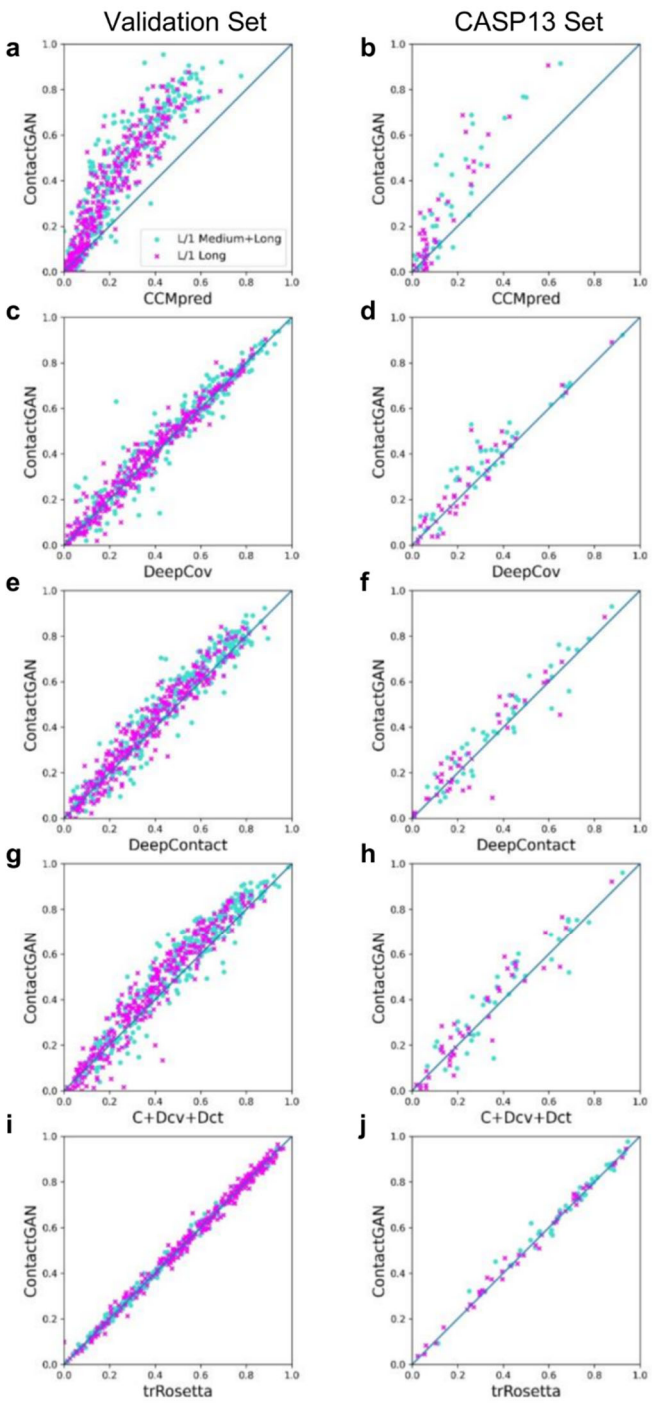
d) ContactGAN model trained only using the Generator network without using the discriminator.

**Supplementary Table S5B. Summary of Improvement by ContactGAN for 3-channel trRosetta on the CASP13 dataset.**

| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **trRosetta (1e-3)** | 0.894 | 0.806 | 0.862 | 0.767 | 0.774 | 0.653 | 0.657 | 0.510 |
| **trRosetta (1e-1)** | 0.919 | 0.816 | 0.879 | 0.751 | 0.780 | 0.637 | 0.648 | 0.500 |
| **trRosetta (1)** | 0.891 | 0.790 | 0.849 | 0.733 | 0.755 | 0.617 | 0.631 | 0.489 |
| **ContactGAN[a)]** | **0.919** | **0.840** | **0.880** | **0.776** | **0.796** | **0.663** | **0.667** | **0.516** |
| **(p-value) [b)]** | **0.035** | **0.028** | 0.063 | 0.185 | **0.000** | **0.036** | **0.017** | **0.033** |
| **9-Layer CGAN [c)]** | 0.901 | 0.803 | 0.862 | 0.741 | 0.777 | 0.640 | 0.650 | 0.512 |
| **G-only CGAN [d)]** | 0.890 | 0.811 | 0.856 | 0.751 | 0.776 | 0.639 | 0.661 | 0.506 |

The results shown are on the CASP13 dataset.

The notes for a), b), c), d) are the same as the previous table, Table S5A.

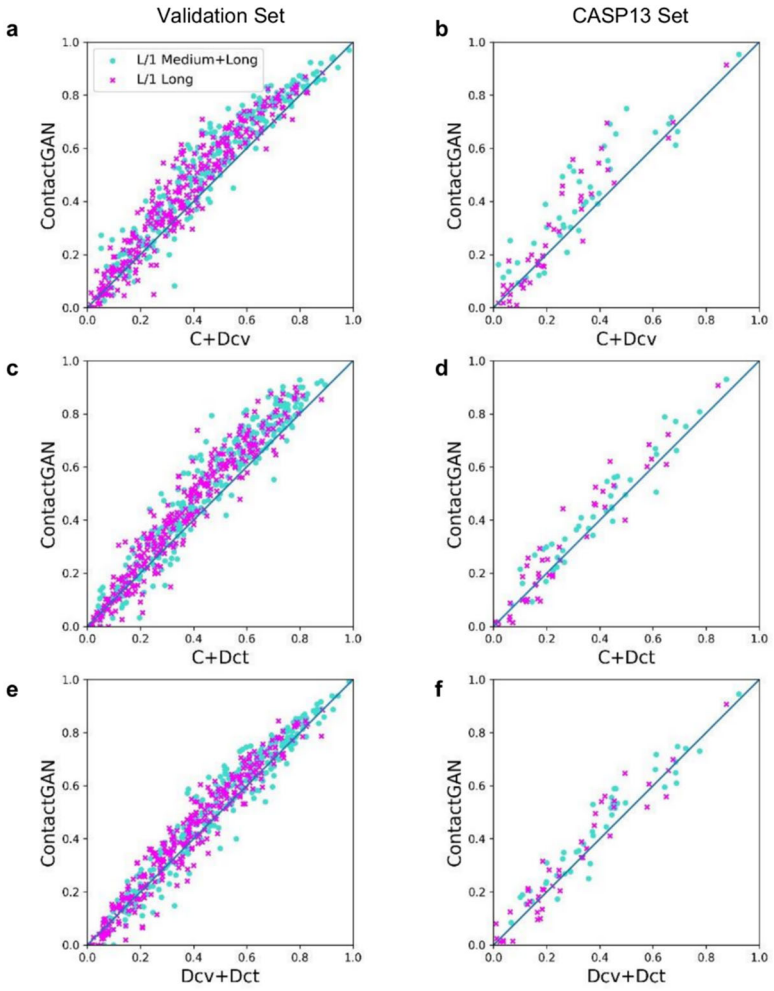



**Supplementary Figure S3A. Precision of ContactGAN relative to the original method in the validation set and the CASP13 set with 1-channel and 3-channel inputs.**

L/1 Long and Medium contacts (circles in cyan) and L/1 Long contacts (crosses in magenta) are plotted for each map in the datasets. **a,** Performance of CCMpred on the validation dataset. **b,** CCMpred on the CASP13 set. **c,** DeepCov on the validation dataset. **d,** DeepCov on the CASP13 set. **e,** DeepContact on the validation dataset. **f,** DeepContact on the CASP13 set. **g,** Performance of the three-channel ContactGAN

with CCMpred+DeepContact+DeepCov, on the validation dataset. For the x-axis, for each map the highest precision value among the three methods was plotted. **h,** The three-channel with CCMpred+DeepContact+DeepCov, on the CASP13 dataset. For the x-axis, for each map the highest precision value among the three methods was plotted. **i,** Performance of the three-channel ContactGAN with trRosetta using three E-value cutoffs, 0.001, 0.1, and 1.0 on the validation set. For the x-axis, for each map the highest precision value among the three trRosetta maps was plotted. **j,** the three-channel ContactGAN with trRosetta on the CASP13 dataset. For the x-axis, for each map the highest precision value among the three trRosetta maps was plotted.

An improvement is shown for a majority of the cases for all the contact prediction methods and their combinations. For trRosetta cases, outputs of ContactGAN improved or left unchanged precision for 63.1% and 65.1% of the maps in the validation and the CASP13 dataset, respectively.




**Supplementary Figure S3B. Precision of ContactGAN relative to the original method in the validation set and the CASP13 set with 2-channel inputs.**

L/1 Long and Medium contacts (circles in cyan) and L/1 Long contacts (crosses in magenta) are plotted for each map in the datasets. **a,** Performance of the two-channel ContactGAN with CCMpred +DeepCov, on the validation dataset. For the x-axis, for each map the highest precision value among the two methods was plotted. **b,** The two-channel CCMpred+DeepCov on the CASP13 set. **c,** CCMpred+DeepContact, on the validation dataset. **d,** CCMpred+DeepContact on the CASP13 set. **e,** DeepCov+DeepContact, on the validation dataset. **f,** DeepCov+DeepContact, on the CASP13 set.

**Supplementary Table S6A. Summary of improvement by ContactGAN for individual methods on the CASP14 dataset.**

| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg[a)] | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **CCMpred** | 0.309 | 0.289 | 0.275 | 0.247 | 0.208 | 0.179 | 0.157 | 0.128 |
| | 0.406 | 0.335 | 0.379 | 0.314 | 0.323 | 0.261 | 0.255 | 0.201 |
| **Dcv** | 0.559 | 0.475 | 0.496 | 0.417 | 0.414 | 0.331 | 0.322 | 0.253 |
| | 0.588 | 0.451 | 0.527 | 0.407 | 0.435 | 0.337 | 0.345 | 0.257 |
| **Dct** | 0.591 | 0.469 | 0.531 | 0.434 | 0.431 | 0.336 | 0.329 | 0.243 |
| | 0.608 | 0.476 | 0.551 | 0.445 | 0.457 | 0.350 | 0.352 | 0.252 |

Average precision values on the CASP14 dataset of 49 protein domains are shown.
**a)** The columns Med+Lg consider medium and long-range contacts (residue pairs separated by more than 11 residues). Other notations are the same as Supplementary Table S3.

**Supplementary Table S6B. Summary of improvement by ContactGAN for multi-channel methods on the CASP14 dataset.**

| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **C+Dcv** | 0.591 | 0.469 | 0.531 | 0.434 | 0.431 | 0.336 | 0.329 | 0.243 |
| | 0.628 | 0.540 | 0.571 | 0.483 | 0.484 | 0.375 | 0.377 | 0.275 |
| **C+Dct** | 0.559 | 0.475 | 0.496 | 0.417 | 0.414 | 0.331 | 0.322 | 0.253 |
| | 0.578 | 0.450 | 0.529 | 0.423 | 0.456 | 0.349 | 0.360 | 0.269 |
| **Dcv+Dct** | 0.559 | 0.475 | 0.496 | 0.417 | 0.414 | 0.331 | 0.322 | 0.253 |
| | 0.621 | 0.529 | 0.581 | 0.477 | 0.489 | 0.382 | 0.381 | 0.292 |
| **C+Dcv+Dct** | 0.559 | 0.475 | 0.496 | 0.417 | 0.414 | 0.331 | 0.322 | 0.253 |
| | 0.645 | 0.518 | 0.579 | 0.473 | 0.495 | 0.389 | 0.386 | 0.298 |

As input, maps from two or three methods were used. All the other notations are the same as Table S3.

**Supplementary Table S6C. Summary of Improvement by ContactGAN for 3-channel trRosetta on the CASP14 dataset.**

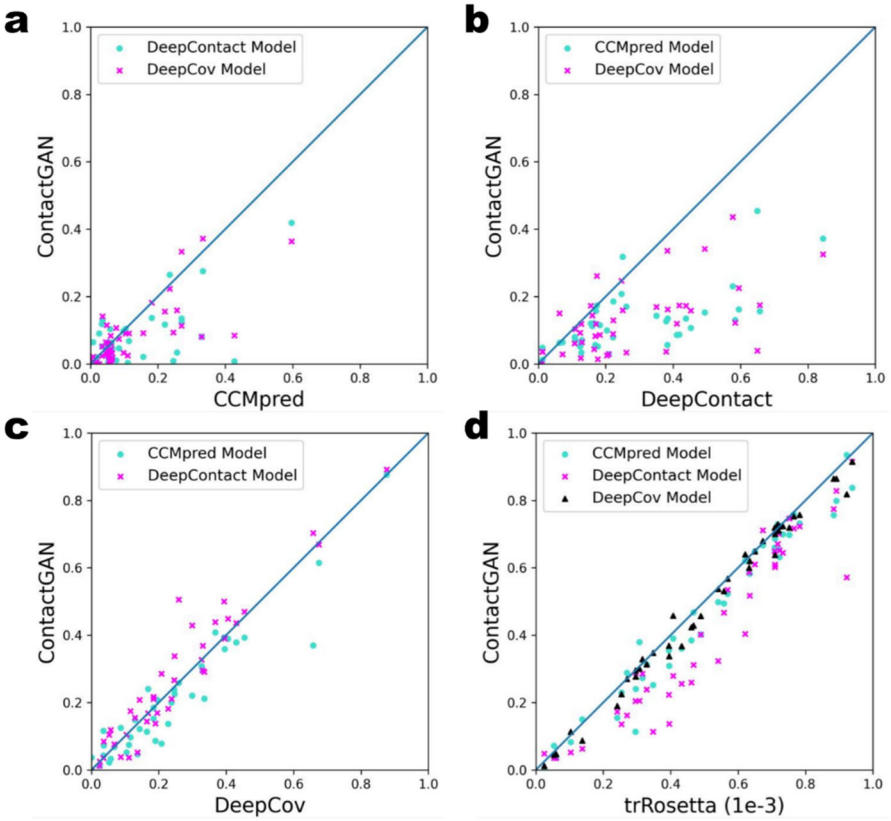
| Method | L/10 | | L/5 | | L/2 | | L | |
|---|---|---|---|---|---|---|---|---|
| | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long | Med+Lg | Long |
| **trRosetta (1e-3)** | 0.707 | 0.618 | 0.671 | 0.577 | 0.579 | 0.481 | 0.468 | 0.368 |
| **trRosetta (1e-1)** | 0.691 | 0.617 | 0.669 | 0.589 | 0.577 | 0.482 | 0.462 | 0.366 |
| **trRosetta (1)** | 0.680 | 0.614 | 0.651 | 0.585 | 0.573 | 0.477 | 0.458 | 0.362 |
| **ContactGAN[a)]** | **0.711** | **0.628** | <u>0.671</u> | **0.591** | **0.588** | **0.494** | **0.474** | **0.377** |
| **(p-value)[b)]** | 0.404 | **0.048** | 0.696 | **0.015** | **0.049** | **0.008** | **0.040** | **0.033** |
| **9-Layer CGAN[c)]** | 0.706 | 0.611 | 0.667 | 0.580 | 0.577 | 0.460 | 0.461 | 0.365 |

As input, maps from trRosetta which were generated using a MSA with the E-value cutoff specified in the parenthesis. 1e-3: 0.001, 1e-1: 0.1. All the other notations are the same as Supplementary Table S3.

a) ContactGAN with 18 ResNet blocks was used in the generator. For each metric, the value of the model with the best validation performance of the particular metric was shown and its p-value was evaluated. Compared with the results by trRosetta (1e-3), values are shown in **bold** if they are larger.

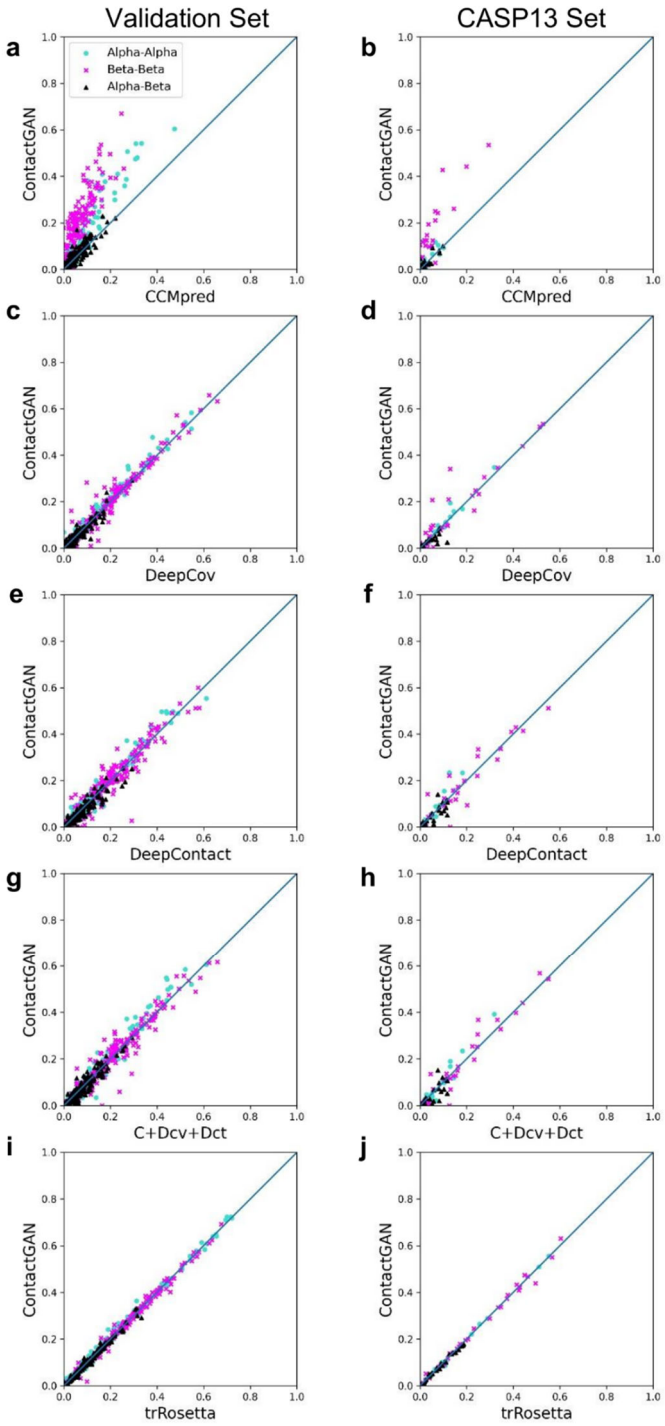
b) The corresponding p-value of ContactGAN[a)] results from the t-test statistics. p-values are shown in **bold** if they are statistically significant considering a p-value level of 0.05 and underlined when they are significant considering a p-value level of 0.1.

c) Same ContactGAN model as the other methods (9 ResNet blocks in the generator)




**Supplementary Figure S4. Precision by ContactGAN relative to the original method on the CASP13 set using different trained models**

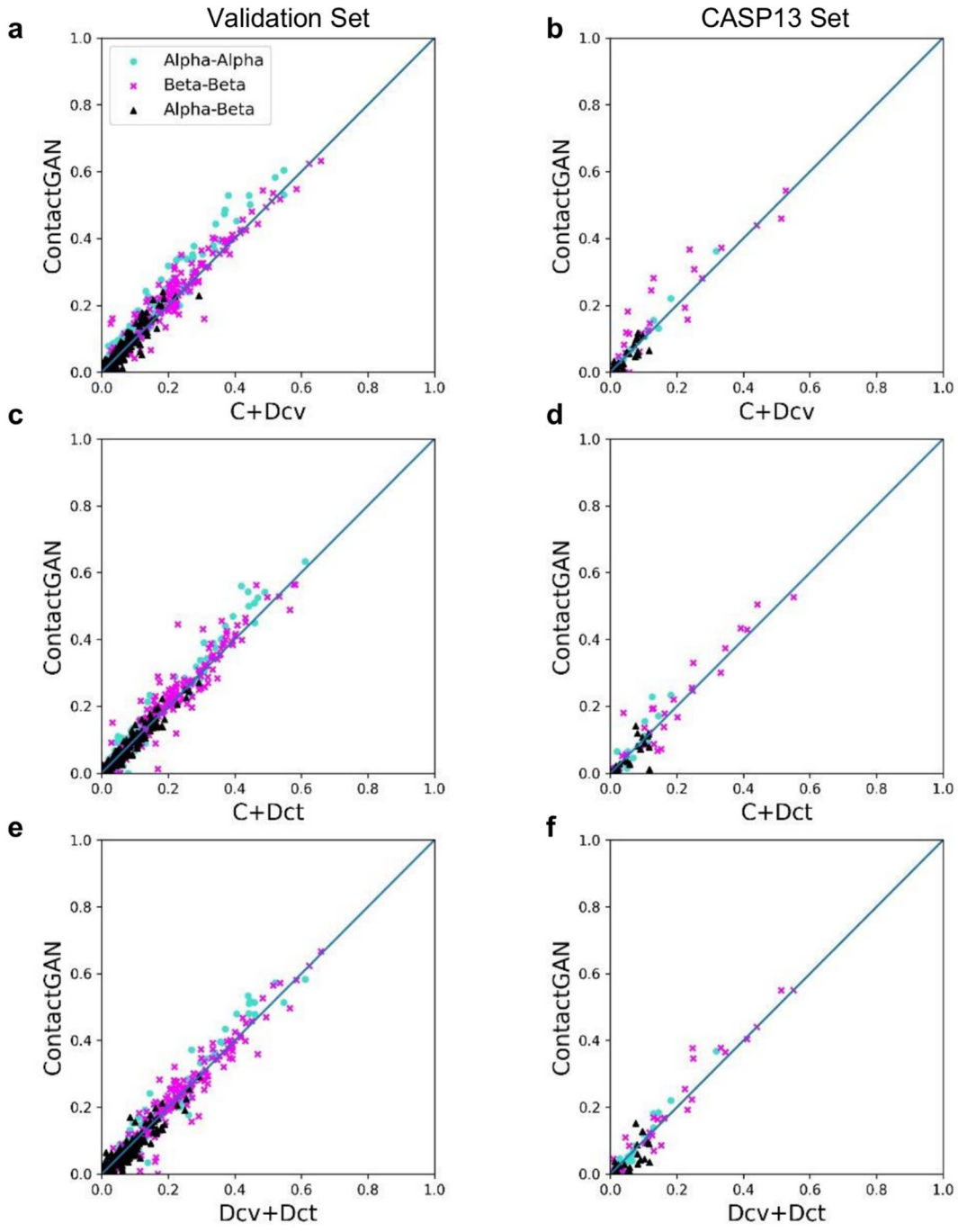
L/1 Long contact precision plotted for each map in the CASP13 dataset. **a,** Performance of ContactGAN on contact maps predicted by CCMpred. For each map in the dataset, the precision of before and after the refinement by ContactGAN was plotted. ContactGAN models were trained on maps predicted by DeepContact (circles in cyan) and on maps predicted by DeepCov (crosses in magenta). **b,** Performance on DeepContact maps using models trained on CCMpred (cyan) and DeepCov (magenta). **c,** Performance on DeepCov maps using models trained on CCMpred (cyan) and DeepContact (magenta). **d,** Performance on maps predicted by a single-channel trRosetta with an E-value of 0.001 using models trained on CCMpred (cyan), DeepContact (magenta), and DeepCov (black).




**Supplementary Figure S5A. Precision by ContactGAN for contacts between secondary structure elements with 1-channel and 3-channel inputs.**

For each map, the fraction of correct long-range contacts between residues in α-helix and α-helix (α–α; cyan), β-strand and β-strand (β–β; magenta), and α-helix and β-strand (α–β; black) among the top L/1 long predicted contacts are plotted before (x-axis) and after (y-axis) applying ContactGAN. Contacts

with a residue in loop structure are not plotted. **a,** the results on predicted maps by CCMpred on the validation dataset. The fraction of correct contacts between α–α, β–β, and α–β increased or stayed the same by ContactGAN for 82.7%, 95.0%, and 80.1% of the maps, respectively. **b,** the results on predicted maps by CCMpred on the CASP13 dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 90.7%, 88.4%, and 81.4% of the maps, respectively. **c,** the results on predicted maps by DeepCov on the validation dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 82.7%, 81.1%, and 74.8% of the maps, respectively. **d,** the results on predicted maps by DeepCov on the CASP13 dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 72.1%, 74.4%, and 60.5% of the maps, respectively. **e,** the results on predicted maps by DeepContact on the validation dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 73.8%, 75.1%, and 67.1% of the maps, respectively. **f,** the results on predicted maps by DeepContact on the CASP13 dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 86.0%, 69.8%, and 65.1% of the maps, respectively. **g,** the results on predicted maps by the three-channel ContactGAN with inputs of CCMpred, DeepCov, and DeepContact on the validation dataset. The x-axis shows the highest accuracy among the three methods for each map. Improvement or tie observed for α–α, β–β, and α–β contacts for 70.8%, 78.1%, and 72.1% of the maps, respectively. **h,** the results on predicted maps by the three-channel inputs of CCMpred, DeepCov, and DeepContact on the CASP13 dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 76.7%, 60.5%, and 65.1% of the maps, respectively. **i,** the results on predicted maps by the three-channel trRosetta (three different E-values were used, 0.001, 0.1, and 1.0) on the validation dataset. The x-axis shows the highest accuracy among the three single-channel trRosetta results for each map. Improvement or tie observed for α–α, β–β, and α–β contacts for 72.8%, 74.1%, and 70.4% of the maps, respectively. **j,** the results on predicted maps by the three-channel trRosetta on the CASP13 dataset. Improvement or tie observed for α–α, β–β, and α–β contacts for 88.4%, 72.1%, and 69.8% of the maps, respectively.

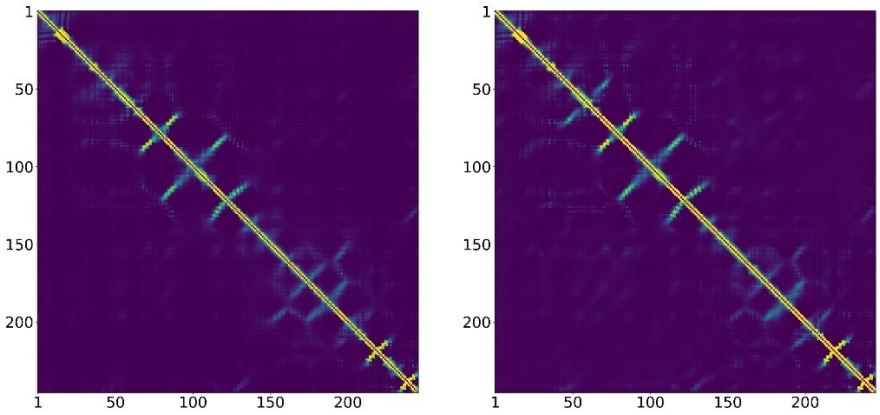



**Supplementary Figure S5B. Precision by ContactGAN for contacts between secondary structure elements with 2-channel inputs.**

For each map, the fraction of correct long-range contacts between residues in α-helix and α-helix (α–α; cyan), β-strand and β-strand (β–β; magenta), and α-helix and β-strand (α–β; black) among the top L/1 long predicted contacts are plotted before (x-axis) and after (y-axis) applying ContactGAN. Contacts with a residue in loop structure are not plotted. **a,** the results on predicted maps by the two-channel

ContactGAN with inputs of CCMpred and DeepCov on the validation dataset. The x-axis shows the highest accuracy among the three methods for each map. Improvement or tie observed for α−α, β−β, and α−β contacts for 82.4%, 84.1%, and 77.4% of the maps, respectively. **b,** the results on predicted maps by the two-channel inputs of CCMpred and DeepCov on the CASP13 dataset. Improvement or tie observed for α−α, β−β, and α−β contacts for 81.4%, 76.7%, and 69.8% of the maps, respectively. **c,** the results on predicted maps by the two-channel ContactGAN with inputs of CCMpred and DeepContact on the validation dataset. Improvement or tie observed for α−α, β−β, and α−β contacts for 74.4%, 81.1%, and 70.4% of the maps, respectively. **d,** the results on predicted maps by the two-channel inputs of CCMpred and DeepContact on the CASP13 dataset. Improvement or tie observed for α−α, β−β, and α−β contacts for 83.7%, 69.8%, and 65.1% of the maps, respectively. **e,** the results on predicted maps by the two-channel ContactGAN with inputs of DeepCov and DeepContact on the validation dataset. Improvement or tie observed for α−α, β−β, and α−β contacts for 73.8%, 75.1%, and 65.1% of the maps, respectively. **f,** the results on predicted maps by the two-channel inputs of DeepCov and DeepContact on the CASP13 dataset. Improvement or tie observed for α−α, β−β, and α−β contacts for 79.1%, 65.1%, and 53.5% of the maps, respectively.

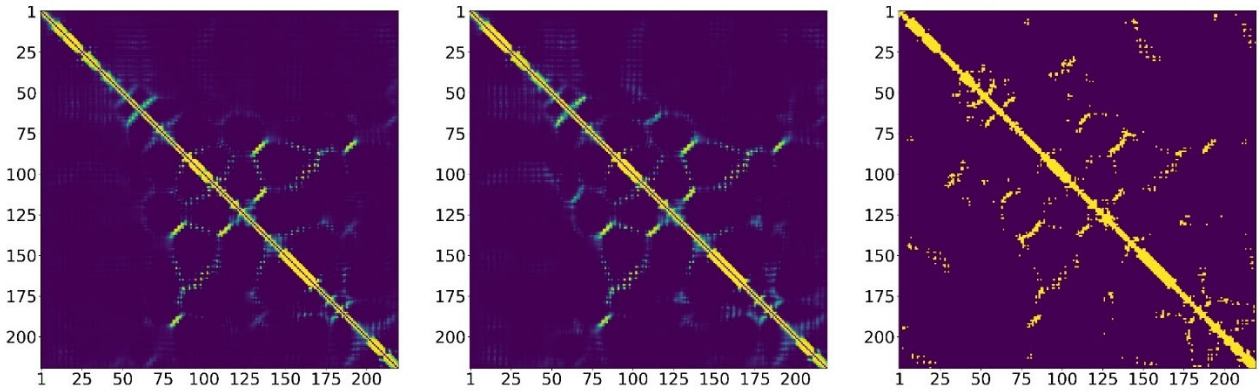
**T0989**




L/1 Med+Long precision: Before (left): 0.522; After (right): 0.619.

L/1 Long precision: Before (left): 0.328; After (right): 0.381.

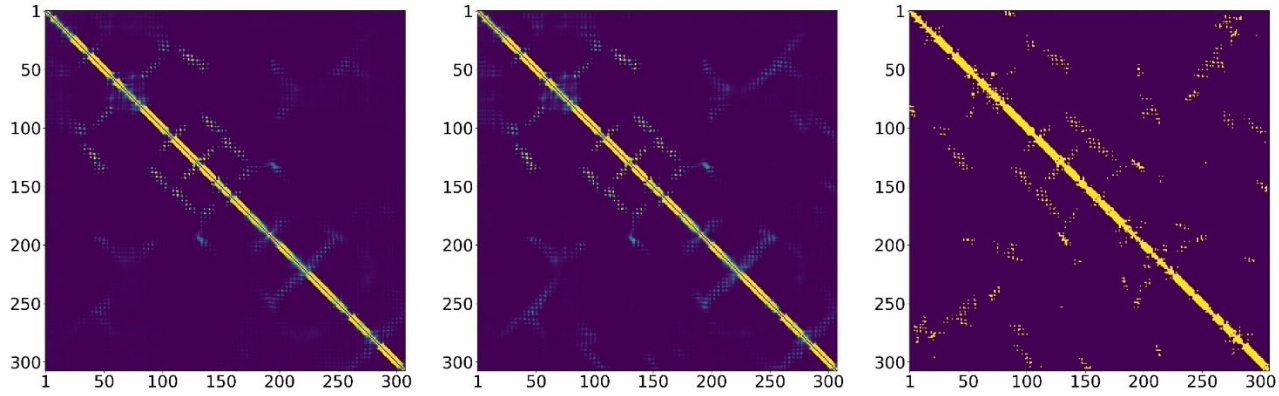
**3BO6_A**




L/1 Long precision: Before (left): 0.564; After (middle): 0.664. The right panel is the correct contact map.

L/1 Med+Long precision: Before (left): 0.650; After (middle): 0.736.

**3T6A_A**

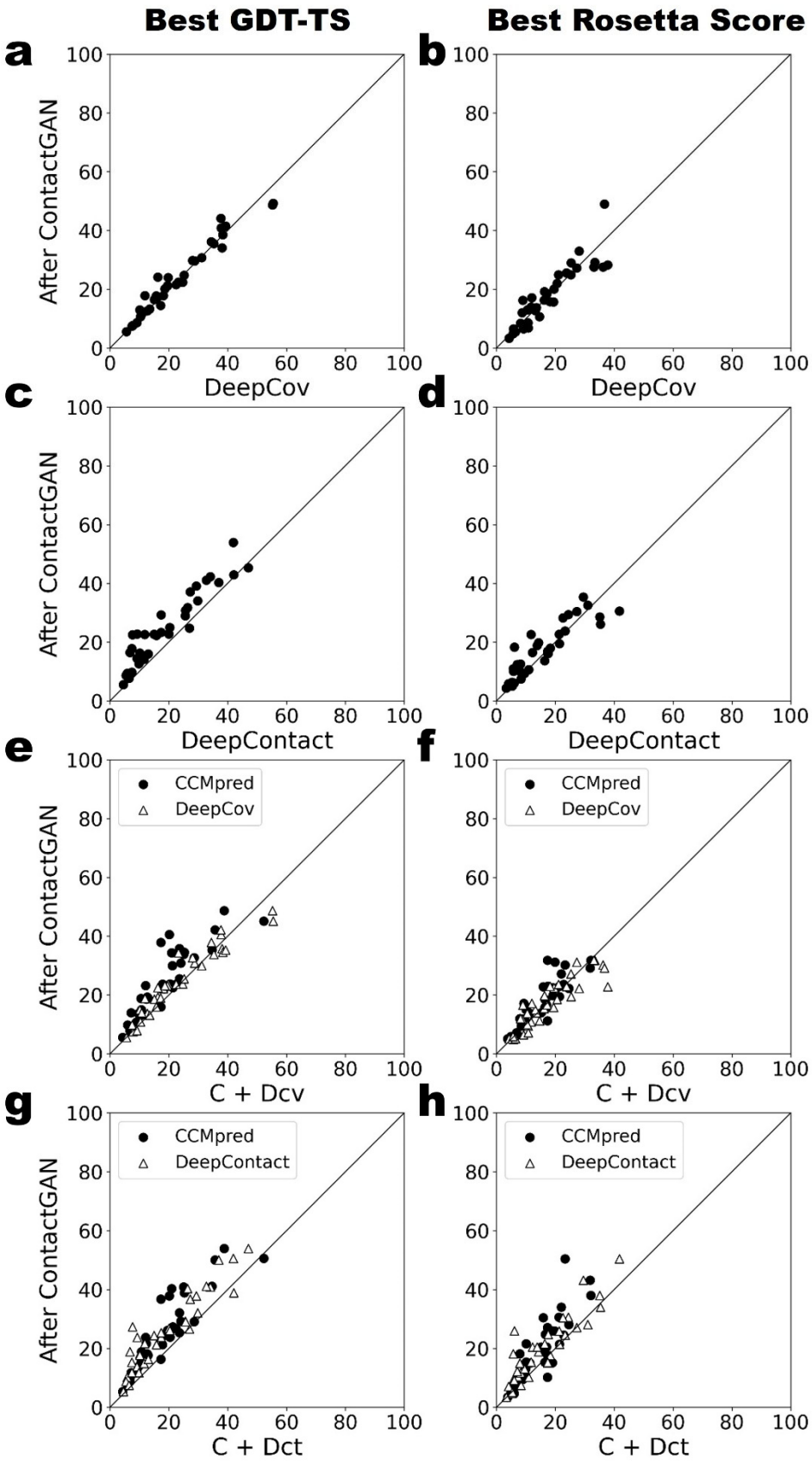



L/1 Long precision: Before (left): 0.458; After (middle): 0.523. The right panel is the correct contact map.

L/1 Med+Long precision: Before (left):0.481; After (middle): 0.536.

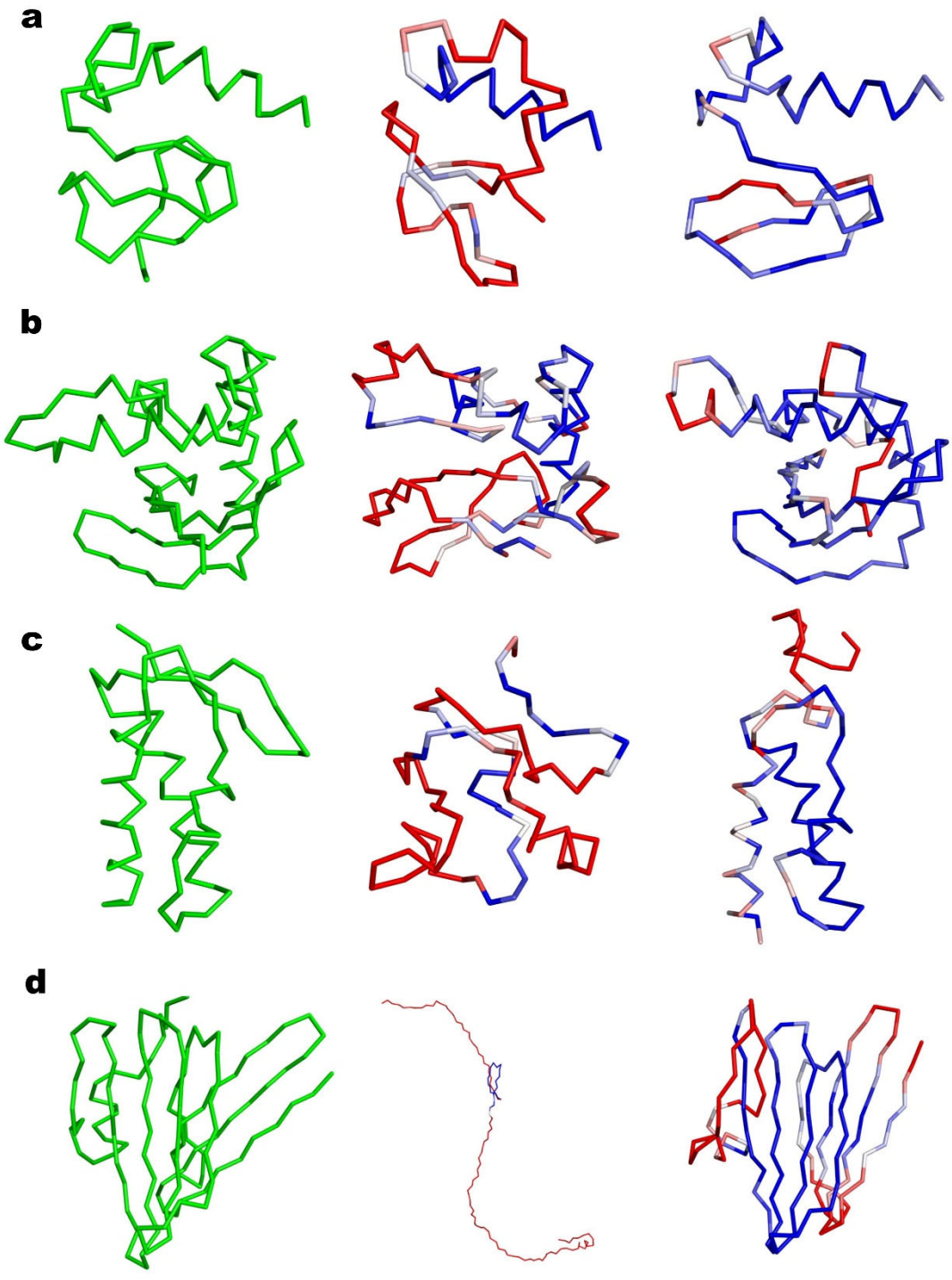
**Supplementary Figure S6. Examples of contact map improvement over maps originally predicted by trRosettta.**

These are examples where the improvement was relatively large for trRosetta contact maps. The original predictions by trRosetta do not have obvious noise. But it is observed that ContactGAN added additional correct interactions with a high probability.


Best GDT-TS
Best Rosetta Score
a
b
c
d
e
f
g
h
After ContactGAN
DeepCov
DeepContact
C + Dcv
C + Dct
CCMpred
DeepCov
DeepContact
0
20
40
60
80
100

**Supplementary Figure S7. GDT-TS comparison of protein structures generated with contact maps before and after ContactGAN.**

For each of the 35 RR target proteins in the CASP13 dataset, GDT-TS values are provided for the best GDT-TS model (the left column) and for the model with the best Rosetta score (the right column) among the 180 models generated. In the Rosetta score, the contact constraint term was omitted. GDT-TS of before (x-axis) and after (y-axis) applying ContactGAN are shown for each target. **a,** models built using maps predicted by DeepCov. The largest GDT-TS among the 180 models is reported. The number of proteins where GDT-TS improved/tie/worsened after ContactGAN are 19/4/12 (0.047), respectively. The number in the bracket indicates the P-value of the significance test conducted. **b,** models using DeepCov maps. GDT-TS of the model with the best Rosetta score is reported. Improvement/tie/worsened cases were 18/1/16 (p-value: 0.462), respectively. **c,** models with maps by DeepContact. The largest GDT-TS among the pool. 33/0/2 (p-value < 0.0001). **d,** models with maps by DeepContact. Best Rosetta score models. 22/0/13 (p-value: 0.031). **e,** a two-channel ContactGAN with CCMpred and DeepCov (Dcv). Black circles, comparison against CCMpred; triangles, against Dcv. Compared with CCMpred, 33/0/2 (p-value < 0.0001); against Dcv: 23/1/11 (p-value: 0.026). **f,** the two-channel with CCMpred and Dcv. Against CCMpred: 23/0/12 (p-value: 0.002). Against Dcv: 15/0/20 (p-value: 0.774). **g,** the two-channel with CCMpred and DeepContact (Dct). Circles, CCMpred; triangles, Dct. Against CCMpred: 33/0/2 (p-value < 0.0001). Against Dct: 33/0/2 (p-value < 0.0001). **h,** the two-channel with CCMpred and Dct. Against CCMpred; 29/1/5 (p-value < 0.0001). Against Dct: 24/1/10 (p-value < 0.0001).


a
b
c
d

**Supplementary Figure S8. Examples of protein structure models constructed with original and improved contact maps by ContactGAN.**

Left, the native structure; middle, the protein model constructed with the original contact map; right, the model constructed with the improved map. These models were selected by the Rosetta energy score. The residue contact term was removed from the energy for the selection. **a,** a CASP13 target ID T1019s1 (58 amino acids). The original map was predicted by DeepCov. The original contact map had L/1 Long precision was 0.155, which was improved to 0.190 by ContactGAN. With the improved contact map, GDT-TS of the protein model improved from 35.8 (middle) to 56.9 (right). **b,** filamentous haemagglutinin family protein, (PDB ID: 6CP9; 126 aa). The original map was predicted by trRosetta. L/1 Long precision, before: 0.407 and after: 0.466. GDT-TS, before: 28.4 (middle) and after: 44.1 (right). **c,** SCCmec type IV DNA binding protein, (PDB ID: 6BTC; 96 aa). The original map was predicted by DeepContact. L/1 Long precision, before: 0.247 and after: 0.273. GDT-TS, before: 24.7 (middle) and after: 52.3 (right). **d,** filamentous haemagglutinin family protein, (PDB ID: 6CP9; CASP13 target ID: T0968s2; 116 aa). The original map predicted by: CCMpred. L/1 Long precision, before: 0.061 and after: 0.113. GDT-TS, before: 15.9 (middle) and after: 48.0 (right).

**Supplementary Table S7. Analysis of best parameters for protein structure modeling on the CASP13 dataset.**

There are two key parameters for protein structure modeling, the probability cutoff for selecting predicted contacts and the folding scheme used in Rosetta. Here, we analyzed the performance of using different parameters.

**Table S7A: Contact prediction probability cutoffs.**

| Method | Best GDT-TS score [a] | | | | Best Rosetta score | | | |
|---|---|---|---|---|---|---|---|---|
| | **0.1** [b] | **0.3** | **0.5** | **0.7** | **0.1** | **0.3** | **0.5** | **0.7** |
| CCMpred [c] | 2.49 | 4.25 | **2.14** | 4.66 | 5.33 | 5.26 | **4.23** | 6.16 |
| | 3.54 | 2.87 | **1.57** | 4.15 | 6.89 | 7.29 | **5.49** | 7.02 |
| DeepCov | 4.28 | **2.69** | 3.35 | 4.64 | 7.86 | **5.37** | 6.91 | 7.43 |
| | 3.82 | 2.60 | **1.51** | 3.77 | 7.81 | **5.77** | 5.90 | 7.52 |
| DeepContact | 5.30 | 1.91 | 3.22 | **0.75** | 7.11 | 4.59 | 6.14 | **3.66** |
| | 2.90 | 3.19 | **1.36** | 5.81 | 7.14 | 8.48 | 6.30 | 8.67 |
| C + Dcv [d] | 4.12 | 2.41 | **2.22** | 4.20 | 7.76 | 7.07 | **6.87** | 8.11 |
| C + Dct | 3.19 | 3.48 | **2.09** | 3.75 | 7.37 | 7.78 | **7.27** | 7.87 |
| Dcv + Dct | 3.90 | 3.33 | **1.98** | 3.32 | 9.73 | 8.00 | **6.21** | 8.50 |
| C + Dcv + Dct | 3.95 | 3.25 | **1.94** | 4.72 | 8.21 | 7.71 | **6.24** | 8.99 |
| trRosetta [e] | 2.33 | 3.87 | **2.27** | 4.67 | **6.84** | 10.1 | 7.39 | 11.3 |
| | **2.89** | 6.04 | 3.02 | 5.71 | 9.07 | 11.9 | **7.79** | 11.7 |
| | 3.66 | 5.10 | **3.18** | 6.79 | **9.22** | 10.9 | 9.66 | 11.7 |
| | 4.06 | 4.22 | **2.67** | 4.67 | 10.7 | 11.2 | **9.01** | 12.8 |

With each probability cutoff value 0.1, 0.3, 0.5 and 0.7, 60 models were generated. Thus, in total, 60x4 = 240 models were generated for each of the CASP13 35 targets. See Supplementary Note 2 for detailed description of the modeling process. For each target, we identified the best, i.e. the largest GDT-TS model among the generated 240 models (overall-best model).

**a)** On the left part of the table labeled as "Best GDT-TS score", for each target, we compared GDT-TS of the overall-best model with the GDT-TS of the best model generated under the probability cutoff (best-for-the-cutoff model). Then, the difference of the GDT-TS was averaged over the 35 targets. On the right part of the table labeled as "Best Rosetta score", we compared the overall-best model with the model selected using the Rosetta score (best-Rosetta-score model).

**b)** modeling was performed using residue contacts that satisfy four probability cut-offs, 0.1, 0.3, 0.5, and 0.7.

**c)** For each contact prediction method, two columns are shown: up, modeling using the original map by the prediction method. bottom, the average difference for models generated with refined contact maps by ContactGAN.

**d)** multi-channel (2 or 3) ContactGAN. C, CCMpred, Dcv, DeepCov; Dct, DeepContact. **e)** four values shown. Values in first three rows correspond to trRosetta with E-values of 0.001, 0.01, and 1, respectively. Values in the last column is for models generated with refined maps by the three-channel ContactGAN for trRosetta.

For each row, the best (the smallest) value among the "Best GDT-TS score" category and the "Best Rosetta score" category was highlighted in bold.

### Table S7B: Contact prediction folding schemes.

Besides the contact prediction probability cutoff concerned in this table, we used three folding protocols following the trRosetta paper (Yang et al., 2020). See Supplementary Note 2 for detailed description of the modeling process.

| Method | Best GDT-TS score | | | Best Rosetta score | | |
|---|---|---|---|---|---|---|
| | 1 [a)] | 2 | 3 | 1 | 2 | 3 |
| CCMpred | **0.69** | 0.98 | 1.66 | **3.93** | 4.06 | 4.49 |
| | **1.23** | 2.18 | 2.89 | **6.04** | 6.68 | 7.51 |
| DeepCov | **1.67** | 1.81 | 3.08 | 6.84 | **6.37** | 7.76 |
| | **1.35** | 1.89 | 3.22 | 7.31 | **6.76** | 7.31 |
| DeepContact | **1.03** | 1.61 | 1.93 | **4.16** | 4.20 | 5.03 |
| | **1.57** | 1.59 | 3.70 | 7.82 | **7.81** | 8.82 |
| C + Dcv | **1.26** | 2.29 | 3.28 | 8.02 | **7.61** | 9.05 |
| C + Dct | **1.47** | 2.37 | 3.58 | 7.18 | **7.13** | 8.90 |
| Dcv + Dct | **1.24** | 1.80 | 3.68 | **7.59** | 7.78 | 9.50 |
| C + Dcv + Dct | 1.98 | **1.60** | 3.64 | **7.46** | 8.87 | 9.32 |
| trRosetta | **1.80** | 1.98 | 5.84 | **10.3** | 10.5 | 13.6 |
| | **1.83** | 3.51 | 5.96 | **10.2** | 10.6 | 12.6 |
| | **2.31** | 2.89 | 6.80 | **10.7** | 12.2 | 14.2 |
| | **1.50** | 3.63 | 5.77 | 10.1 | **10.9** | 14.6 |

**a)** Values computed for three different folding schemes. The first scheme used the short-, the medium, and the long-range contacts progressively. The second scheme used the short and the medium-range contacts in the first run, and then the long-range contacts were added later. The third scheme used all contacts at once. Using a folding scheme, we generated 60 models each with the four probability cutoffs, 0.1, 0.3, 0.5, and 0.7; thus, 240 models were generated for each target. Then, the difference of GDT-TS between the overall-best model and the best model among the models generated under each folding scheme (best-for-the-scheme model) was computed for each CASP target and averaged over all 35 targets. Other notations are same as Table S8A.

For each row, the best (the smallest) value among the "Best GDT-TS score" category and the "Best Rosetta score" category was highlighted in bold.

From Table S7A and S7B, we can see that the 0.5 performed best among the cutoff values tested for most targets (Table S7A), while the first folding scheme worked the best (Table S7B).

**Supplementary Note 1: Hyperparameters of ContactGAN**

The GAN architecture was inspired by SRGAN (Ledig, et al., 2017). Therefore, the main architecture of ContactGAN follows SRGAN. SRGAN uses 16 ResNet blocks in the generator network, which we reduced to 9 ResNet blocks to fit to our limited GPU resource. Also, to reduce the complexity of the network, we used 32 channels in the convolutions of the ResNet blocks, where SRGAN used 64 channels. Our discriminator network consisted of 10 convolution layers as compared to 8 layers in SRGAN, but with a reduced number of channels in each layer by a factor of 2.

We experimented with learning rates of [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1] for the generator and the corresponding values of [0.0004, 0.002, 0.004, 0.02, 0.04, 0.2, 0.4] for the discriminator; thus using separate learning rates based on the Two Time-scale Update Rule (TTUR) (Heusel, et al., 2017). The best learning rates were chosen based on the performance on the validation set.

**Supplementary Note 2: Building structure models from a contact map**

To build protein structure models from the predicted contact map, we used the energy minimization protocol, *MinMover*, in pyRosetta (Chaudhury, et al., 2010). We added scoring terms that account for Cβ-Cβ contact predictions, which come from an input contact map, and backbone dihedral angle predictions, which was computed by SPOT-1D, in the energy function.

Contact predictions were represented as a flat-harmonic contact potential, *f(x)*, and added to the Rosetta energy:

$$f(x) = \begin{cases} 0 & for \quad x \leq 8.0 \\ \left(\frac{x}{4.0}\right)^2 & for \quad x > 8.0 \end{cases} \qquad (1)$$

where *x* is the Cβ-Cβ distance between a residue pair. Contact potentials were added to only for the residue pairs that have a contact probability higher than a cutoff value.

As for the backbone dihedral angle prediction, we added circular harmonic constraints as follows:

$$g(x) = \left(\frac{(ang - ang0) \; mod \; 2\pi}{0.2}\right)^2 \qquad (2)$$

where *ang* is the dihedral angle in the model, *ang0* is a predicted dihedral angle by SPOT-1D. Dihedral angle constraints were added only for residues with a probability of helix or strand that is higher than a cutoff value (0.5, 0.3, or 0.1).

We used three probability cutoff values, 0.5, 0.3, and 0.1, for predicted contacts and angles. Predicted contacts were separated into short-range ($1 \leq$ *sequence separation* $< 12$) and medium-range ($12 \leq$ *sequence separation* $< 24$), and long-range (*sequence separation* $\geq 24$). Following the trRosetta modeling protocol (Yang, et al., 2020), the three ranges of predicted contacts were used as restraints in three schemes: The first scheme used the short-, the medium, and the long-range contacts progressively. The second scheme used the short and the medium-range contacts in the first run, and then the long-range contacts were added later. The third scheme used all contacts at once. For each step in a scheme, the energy minimization by *MiniMover* was performed with a maximum iteration number of 1000. After a minimization, we applied full-atom relaxation by *FastRelax* protocol in Rosetta with a maximum iteration of 200 maximum. For each scheme, we generate 20 models from different starting models with random backbone torsion angles. In total 180 models (3 probability cutoffs * 3 modeling schemes * 20 models) were generated for each target. Supplementary Table S7 provides dependency of the parameter preference. In Figure 4 and Supplementary Figure S6, the best GDT-TS model (on the left column) and the best Rosetta energy model (on the right column) were reported for each target in the CASP13 dataset.

**Reference in the Supplementary Information**

Chaudhury, S., Lyskov, S. and Gray, J.J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* 2010;26(5):689-691.

Yang, J.*, et al.* Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences* 2020.

Ledig, C., et al. Photo-realistic single image super-resolution using a generative adversarial network. In, Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4681-4690.

Heusel, M., et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In, Adv. Neural Inf. Process. Syst.; 2017. p. 6626-6637.