PconsC4: fast, free, easy, and accurate contact predictions.

Mirco Michel[1,*], David Menéndez Hurtado[1,*], and Arne Elofsson[1,†]

[1]Science for Life Laboratory and Department of Biochemistry and Biophysics, Stockholm University, Stockholm 10691, Sweden.
[†]To whom correspondence should be addressed
[*]Contributed equally.

August 2, 2018

Abstract

Motivation: Residue contact prediction was revolutionized recently by the introduction of direct coupling analysis (DCA). Further improvements, in particular for small families, have been obtained by the combination of DCA and deep learning methods. However, existing deep learning contact prediction methods often rely on a number of external programs and are therefore computationally expensive.

Results: Here, we introduce a novel contact predictor, PconsC4, which performs on par with state of the art methods. PconsC4 is heavily optimized, does not use any external programs and therefore is significantly faster and easier to use than other methods.

Availability: PconsC4 is freely available under the GPL license from `https://github.com/ElofssonLab/PconsC4`. Installation is easy using the pip command and works on any system with Python 3.5 or later and a modern GCC compiler.

Contact: arne@bioinfo.se

Introduction

To predict the structure of a protein from no other information than its sequence has been a major challenge in bioinformatics for decades. With the introduction of direct coupling analysis (DCA) to improve contact predictions (Weigt *et al.*, 2009) significant progress in protein structure prediction was reported in 2011 (Morcos *et al.*, 2011; Marks *et al.*, 2011). These methods have then been used to predict the structure of hundreds of protein families with high accuracy to an unprecedented accuracy (Ovchinnikov *et al.*, 2017). One disadvantage of DCA methods is that they require very large multiple sequence alignments to provide accurate contact predictions. This problem has been overcome by refining the initial DCA prediction with deep learning methods (Skwark *et al.*, 2014; Wang *et al.*, 2017).

Although the structure can accurately be predicted for many protein families, there still exist many families were the predictions are not sufficiently accurate. Although predictions are better for larger families, many other factors also seem to be important (Michel *et al.*, 2017a). The exact reason for what is needed to improve the predictions is not well understood. But it is not unlikely that the underlying multiple sequence alignments are not optimal. It might therefore be possible to improve the predictions if alternative multiple sequence alignments are examined or metagenomics data is included. Given the computational costs of earlier deep learning contact predictions methods it has been difficult to exhaustively examine alternative alignments for each protein family.

Here, we present a novel deep learning approach, PconsC4, that performs as well, or even better, than earlier methods. More importantly it is freely available, significantly faster and easier to install than alternative methods. This provides all users with an easy way to explore alternative multiple sequence alignments or other parameters for contact predictions.

Implementation

Below follows a short description of PconsC4, for details see the supporting information. PconsC4 is trained on a set of 2791 proteins culled from PDB and benchmarked on two datasets without any

homology to the training set. One validation set is identical to the benchmark set in PconsC3 (Michel *et al.*, 2017b). Additionally, we benchmarked PconsC4 on 44 proteins from CASP12, see Tables S2-S5. Multiple sequence alignments are created using three iterations of HHblits (Remmert *et al.*, 2011) and an e-value threshold of 1.0. Other alignment methods and cut-offs as well as combinations of different alignments were tested but did not provide significant improvements.

From each position in the multiple sequence alignments 72 features are calculated and fed into the PconsC4 network. These include: 68 one-dimensional sequential features and four pairwise features; the GaussDCA score (Baldassi *et al.*, 2014), APC-corrected mutual information, normalized APC-corrected mutual information, and cross-entropy.

At the core of PconsC4 is the U-net architecture (Ronneberger *et al.*, 2015), designed for image segmentation. It is composed of a series of convolutional layers, down- and up-sampling, with shortcut connections to help convergence.

To include secondary structure information but remain independent of external predictors, we took the pre-trained network from ProQ4 (Menéndez Hurtado *et al.*, 2018), that takes all the one-dimensional inputs and predicts secondary structure, dihedral angles, and surface accessibility for each residue. PconsC4 takes the output of the second to last layer, transforms them into two-dimensional features via an outer product, and concatenates them to the rest of the inputs.

Finally, the network produces four outputs: the probability of a contact for the thresholds of 6, 8, or 10 Å, and the distance measured as S-score.

To reduce the number of dependencies and overall run time, we re-implemented GaussDCA as a Python package using Pythran (Guelton *et al.*, 2015) resulting in a speedup of a factor of three. Optimization details are in section 2 of the supplementary information.

Table 1: Performance in PPV for the L top-ranked contact with a sequence separation of > 5 residues. Results for all, small families with Meff (Baldassi *et al.*, 2014) $<$ and for short-, $(5, 12)$, medium- $(12, 23)$ and long: $(23, \infty)$ ranges. Average runtime and external dependencies.

	All	Small	Short	Med	Long	Time [s]	Deps.
PlmDCA	0.36	0.12	0.14	0.15	0.26	84	-
GaussDCA	0.34	0.12	0.14	0.15	0.25	27	-
PSICOV	0.34	0.11	0.13	0.14	0.22	114	-
PhyCMAP	0.32	0.29	0.24	0.19	0.18	718	1,3
PconsC3	0.57	0.36	0.25	0.25	0.37	2931	1,2,3,4,5,6
Metapsicov	0.59	0.39	0.27	0.26	0.36	1158	1,3,7,8,9
PconsC4	0.65	0.42	0.29	0.28	0.45	12 (56*)	-

1: PSIPRED, 2: NetSurfP, 3: PSI-BLAST, 4: PhyCMAP, 5: PlmDCA, 6: GaussDCA, 7: PSICOV, 8: Freecontact, 9: CCMpred.
* Timing without pre-loading the network.

Results and Discussion

Table 1 shows a performance comparison of different methods running on the same alignments. PconsC4 performs 14% better than PconsC3 on the benchmark dataset and 10% better than Metapsicov. These improvements are consistent across short, medium, and long range contacts and for all thresholds (Figures S1-2). Additional evaluations are also presented in the supplementary information. As shown in Figures S3-4, the predicted scores are well calibrated, i.e. the reported scores reflect the real probability for a contact to exist. This enables easy comparison of alternative alignments.

All machine learning methods perform significantly better than the DCA methods for smaller families. PconsC4 approaches maximum performance at around 10^2 effective sequences compared with 10^5 for the DCA methods (Figure S5).

The overall performance is comparable to meta-meta predictors such as DNCON2 (Adhikari *et al.*, 2018) or other deep learning methods (Wang *et al.*, 2017). However, a direct comparison is difficult as these programs are either not available for download, to slow to run for large scale experiments, or cannot be used with a specific multiple sequence alignment.

The main advantage of PconsC4 is that it is fast and easy to use. Starting from a single input alignment PconsC4 predicts the contacts in less than one minute (or 12 s if the network is preloaded).

This is more than 25 times faster than PconsC3 or Metapsicov, while still consistently outperforming them. Since PconsC4 is not tied to any specific alignment method, it can directly take advantage of any improvement on this front, such as the use of alternative alignment strategies (Buchan and Jones, 2017) or metagenomics data (Ovchinnikov *et al.*, 2017), and thus be easily integrated into pipelines.

Acknowledgements

Our thanks to Serge Guelton for his swift fixes to Pythran and to Nikos Tsardakas Renhuldt for discussion. This work was supported by grants from the Swedish Research Council (VR-NT 2016-03798 to AE). This research was conducted using the resources of High Performance Computing Center North (HPC2N).

References

Adhikari, B., Hou, J., and Cheng, J. (2018). DNCON2: improved protein contact prediction using two-level deep convolutional neural networks. *Bioinformatics*, **34**(9), 1466–1472.

Baldassi, C., Zamparo, M., Feinauer, C., Procaccini, A., Zecchina, R., Weigt, M., and Pagnani, A. (2014). Fast and accurate multivariate gaussian modeling of protein families: Predicting residue contacts and protein-interaction partners. *PLOS ONE*, **9**(3), 1–12.

Buchan, D. W. A. and Jones, D. T. (2017). Improved protein contact predictions with the metapsicov2 server in casp12. *Proteins: Structure, Function, and Bioinformatics*, **86**(S1), 78–83.

Guelton, S., Brunet, P., Amini, M., Merlini, A., Corbillon, X., and Raynaud, A. (2015). Pythran: Enabling static optimization of scientific python programs. *Computational Science & Discovery*, **8**(1), 014001.

Marks, D., Colwell, L., Sheridan, R., Hopf, T., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3d structure computed from evolutionary sequence variation. *PLoS One*, **6**(12), e28766.

Menéndez Hurtado, D., Uziela, K., and Elofsson, A. (2018). Deep transfer learning in the assessment of the quality of protein models. *ArXiv e-prints*.

Michel, M., Menendez Hurtado, D., Uziela, K., and Elofsson, A. (2017a). Large-scale structure prediction by improved contact predictions and model quality assessment. *Bioinformatics*, **33**(14), i23–i29.

Michel, M., Skwark, M. J., Menendez Hurtado, D., Ekeberg, M., and Elofsson, A. (2017b). Predicting accurate contacts in thousands of pfam domain families using pconsc3. *Bioinformatics*, **33**(18), 2859–2866.

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D., Sander, C., Zecchina, R., Onuchic, J., Hwa, T., and Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A*, **108**(49), 1293–301.

Ovchinnikov, S., Park, H., Varghese, N., Huang, P., Pavlopoulos, G., Kim, D., Kamisetty, H., Kyrpides, N., and Baker, D. (2017). Protein structure determination using metagenome sequence data. *Science*, **355**(6322), 294–298.

Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, **9**(2), 173–175.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.

Skwark, M., Raimondi, D., Michel, M., and Elofsson, A. (2014). Improved contact predictions using the recognition of protein like contact patterns. *PLoS Comput Biol*, **10**(11), e1003889.

Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017). Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol*, **13**(1), e1005324.

Weigt, M., White, R., Szurmant, H., Hoch, J., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A*, **106**(1), 67–72.

Supplementary materials to "PconsC4: fast, free, easy, and accurate contact predictions."

Mirco Michel, David Menéndez-Hurtado and Arne Elofsson

August 2, 2018

Contents

1 Results

Here we report more detailed comparisons on the test sets. First, in Table S1 we compare short, medium, and long range contacts, showing that PconsC4 is better than previous methods at all ranges. Figures S1 and S2 show the receiver operating characteristic curves.

We also include an expanded version of the Table 1, Table S1 including CASP12 and other methods. For benchmarking GaussDCA we used the Julia implementation and CCMpred for PlmDCA. The times were measured on a machine with an Intel i7-4770 CPU and a NVIDIA 1070Ti GPU for CCMpred.

PconsC4 seems to be better calibrated that earlier methods, as shown in Figures S3 and S4. This means that a predicted score of 0.2 indicates that there is a 20% chance for a contact to exist. A perfectly calibrated predictor would lie on the diagonal (dotted line).

Table S1: Precision for the top L contacts, where L is the sequence length, at different distance thresholds. Short: $(5, 12]$, medium: $(12, 23]$, long: $(23, \infty)$.

Method	Benchmark set		
	Short	Medium	Long
PlmDCA	0.14	0.15	0.26
GaussDCA	0.14	0.15	0.25
PSICOV	0.13	0.14	0.22
PhyCMAP	0.24	0.19	0.18
PconsC3	0.25	0.25	0.37
MetaPSICOV	0.27	0.26	0.36
PconsC4	0.29	0.28	0.45

Table S2: Performance in PPV for the L top-ranked contact with a sequence separation of > 5 residues, average runtime in seconds on the benchmark set, and external dependencies (except for alignment methods).

	Benchmark set	CASP12	Time [s]	Deps.
PlmDCA (CCMpred, GPU)	0.36	0.26	84	-
PlmDCA (CCMpred, CPU)	0.36	0.26	393	-
GaussDCA (Julia)	0.34	0.25	27	-
GaussDCA (Pythran)	0.34	0.25	10	-
PSICOV	0.34	0.20	114	-
PhyCMAP	0.32	0.22	718	1,3
PconsC3	0.57	0.39	2931	1,2,3,4,5,6
Metapsicov	0.59	0.43	1158	1,3,7,8,9
PconsC4	0.65	0.46	56	-
PconsC4 (pre-loaded)	0.65	0.46	12	-

1: PSIPRED, 2: NetSurfP, 3: PSI-BLAST, 4: PhyCMAP, 5: PlmDCA, 6: GaussDCA, 7: PSICOV, 8: Freecontact, 9: CCMpred.

1.1 ROC and calibration curves

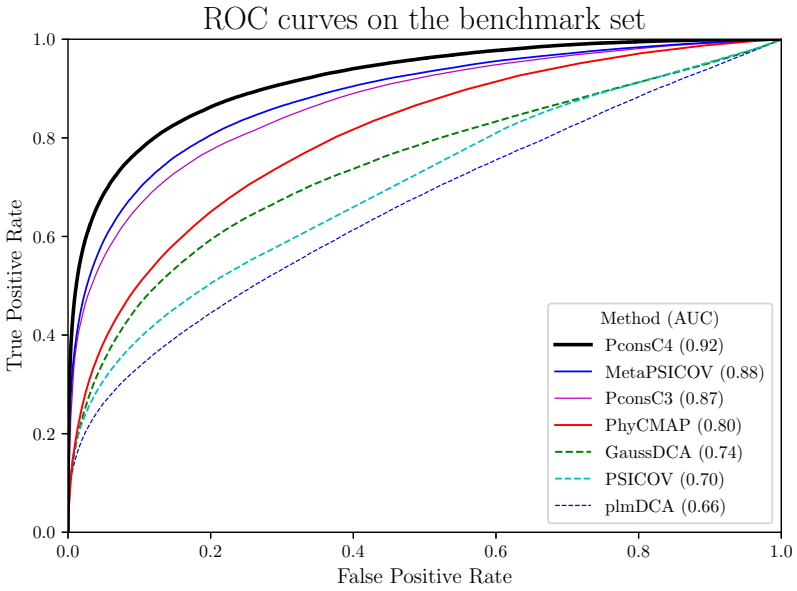


Figure S1: ROC curves on the Benchmark set

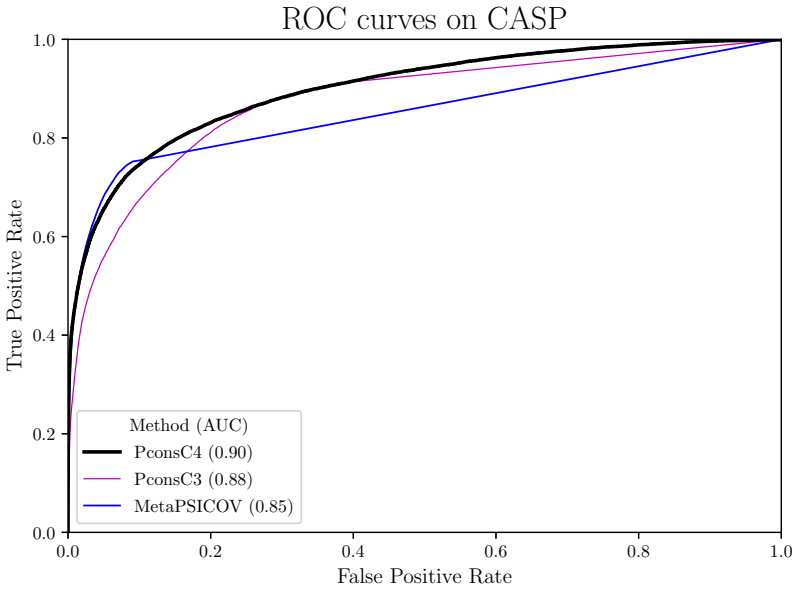


Figure S2: ROC curves on CASP12

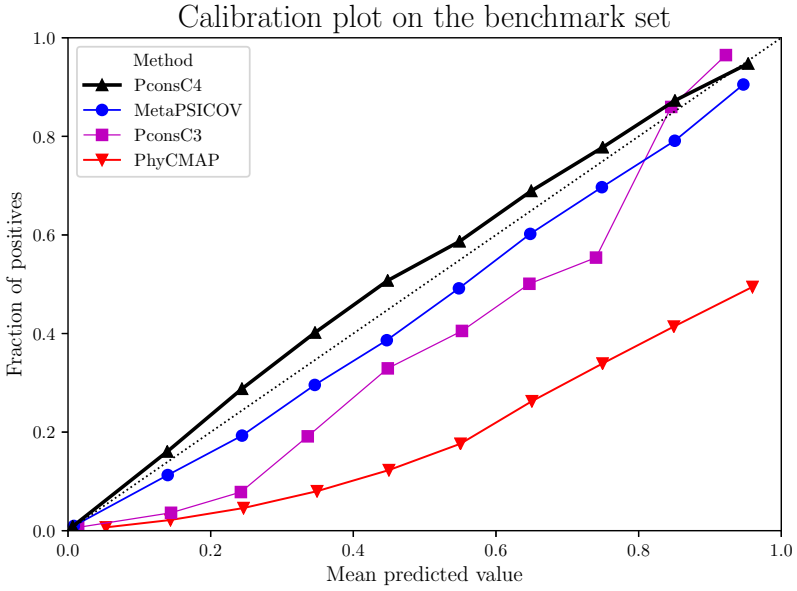


Figure S3: Calibration curves on the Benchmark set

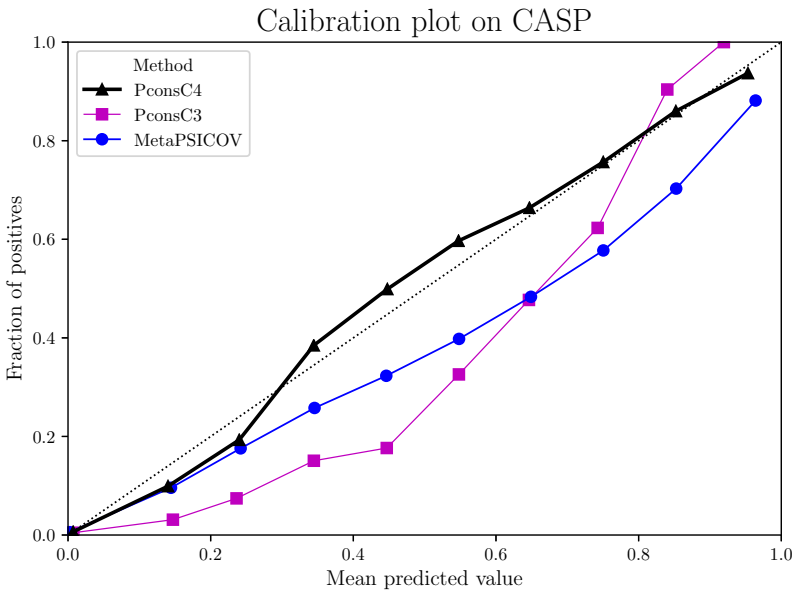


Figure S4: Calibration curves on CASP12

1.2 Precision as a function of the alignment depth

The Figure S5 shows performance vs. number of effective sequences. Here we use the same definition as in (Baldassi *et al.*, 2014a): the number of sequences that are significantly different from each other. While statistical methods, like DCA, benefit from a constant improvement for larger number of effective sequences, PconsC4 starts to plateau at around 10^2 sequences. This suggests that it is capable of learning to ignore the noise in DCA methods, even when the signal is weak.

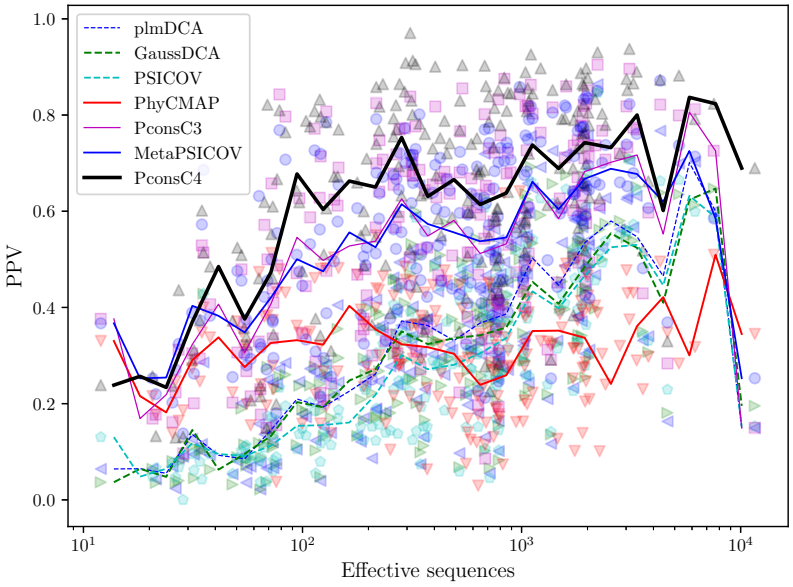


Figure S5: PPV vs. number of effective sequences on the benchmark set. The top L contacts are considered.

2 GaussDCA optimization

In the reimplementation of GaussDCA we made two differences with respect to the reference implementation that improve the speed, while keeping the results exactly the same.

2.1 GaussDCA and compressed alignments

Much of the CPU time is spent computing sequence weights, which means we need to compute all pairwise Hamming distances between sequences in the multiple sequence alignment. In a naive implementation, we store the alignment as a 2D array of int8. The compiler can then use SIMD instructions to combine several numbers to fit them in the word size of the CPU.

The Julia implementation of Gaussdca, (Baldassi *et al.*, 2014b), uses a manually compressed alignment: since we have only 21 states, we only need 5 bits

per symbol, so we can manually pack up to 12 amino acids in a single 64 bits int, that fits our CPU optimally. To compute the distance, we need to XOR the sequences, and count the number of 5-bit regions where it differs. This gives us a 50% increase in packing efficiency with respect to the naive 8-bit storage.

In our re-implementation we found that the naive version is actually faster, probably because it is easier for the compiler (GCC 7) to optimize. Further improvements can be obtained activating auto-vectorization (GCC flag `-ftree-vectorize`), suggesting that the code emitted by Pythran is more suitable for vectorization.

2.2 Fast estimation of the expected similarity threshold, θ, in GaussDCA

Two sequences are considered similar if their Hamming distance is below a given threshold. (Baldassi *et al.*, 2014a) introduced an automatic estimation as the average fraction of differences across the alignment, and implemented it by directly counting them, which is $\mathcal{O}(m^2)$ on the number of sequences in the alignment.

The same result can be obtained in linear time by counting the number of occurrences of each symbol in each column (called the bincount function), and computing the 2-combination $\binom{n}{2}$. For an alignment of C columns and an alphabet of size Q:

$$n_{matches} = \sum_{c}^{C} \sum_{q}^{Q} \binom{\#\text{alignment}_c = q}{2} \tag{1}$$

3 Training

3.1 Description of inputs

3.1.1 One-dimensional inputs

From the columns of the multiple sequence alignment we can compute the probability of finding each amino acid or gap, p_i. For amino acids, we compute the average frequency $\langle p_i \rangle$ as the observed frequency of the amino acid on the Uniref50 dataset. The expected probability of gap $\langle p_- \rangle$ is estimated from each alignment. We consider a total of 23 amino acid states, the usual 20, plus the gap state, plus B (asparagine or aspartic acid) and X (unknown).

For each column we can compute self-information as the vector of entries:

$$I_i = \log_2(p_i / \langle p_i \rangle), \tag{2}$$

and the partial entropies are defined as:

$$S_i = p_i \log_2(p_i / \langle p_i \rangle) \tag{3}$$

The sequence is given as one hot encoding.

3.1.2 Two-dimensional inputs

The two-dimensional inputs are mutual information (MI), normalized mutual information (NMI), cross entropy (H), and GaussDCA.

Mutual information is defined as:

$$MI(x,y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{4}$$

We also include two more inputs that are adjusted versions of MI. The first is normalized mutual information, where we think of MI as an analogue of a covariance, and NMI is the Pearson correlation coefficient. The entropies of the columns play the role of the variances:

$$NMI(x,y) = \frac{MI(x,y)}{\sqrt{S(x)S(y)}} \tag{5}$$

And lastly, cross entropy is calculated as an additive normalization of mutual information:

$$H(x,y) = S(x) + S(y) - MI(x,y) \tag{6}$$

The Average Product Correction (APC) (Dunn *et al.*, 2008) is applied to all two-dimensional inputs except for cross entropy.

3.2 Training and testing sets

PconsC4 is trained on a set of 2891 proteins culled from PDB using PISCES (Wang and Dunbrack, Jr., 2003) with a maximum sequence identity 20%, minimum resolution 2.0 Å, maximum R-factor 0.3. Furthermore, chains from the same ECOD (Cheng *et al.*, 2014) H-group as any protein in the benchmark dataset or dating from after 2016-05-01 was removed to avoid potential overlap with the test datasets. Out of these 2891 proteins, 100 randomly selected proteins were used as a validation set for optimization, see Table S3 and S4. For benchmarking, two datasets are used, the same 180 proteins, Table S5 as in (Michel *et al.*, 2017) and the 44 proteins from CASP12 with available structures, Table S6

4 Pipeline

We include a schematic of the complete pipeline, see Figure S6.

In the upper left corner are the 1D inputs being fed to the pre-trained model taken from ProQ4 (Menéndez Hurtado *et al.*, 2018). It is pre-trained to predict secondary structure and surface accessibility for each residue (golden outputs in the middle left).

The 128 output channels of the second to last layer are used to extract relevant 1D features. These are then transformed into 2D features by the outer product and combined with the remaining 2D features (GaussDCA, Mutual Information, Normalized Mutual Information, and Cross Entropy). The concatenation is then passed on to the U-net block (lower rectangle).

U-net was developed for image segmentation and combines convolutional layers, max pooling (downwards arrows) to increase the effective receptive field,

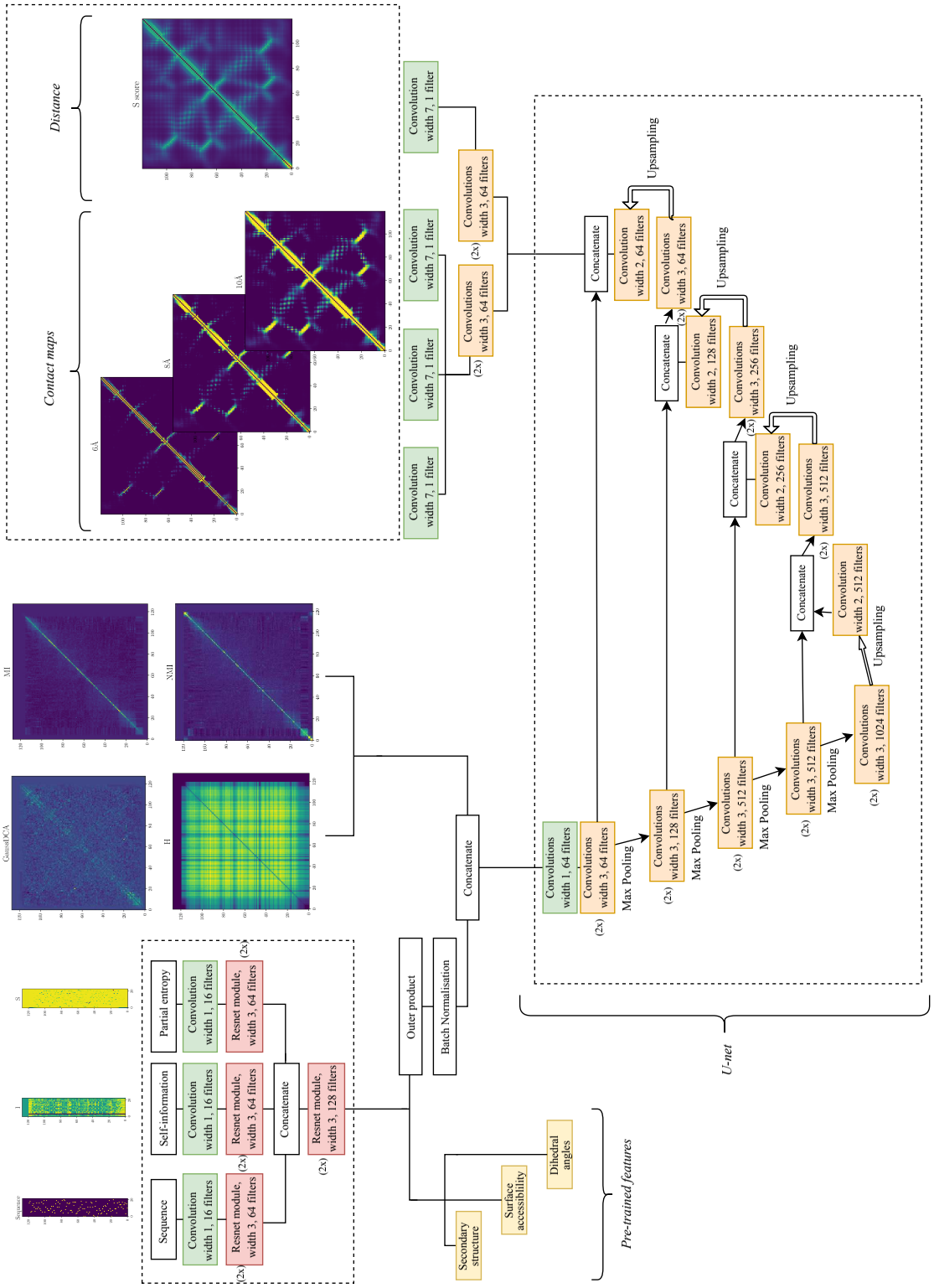
upsampling (upwards arrows) to recover the original size, and skip connections (horizontal arrows) to help convergence.

The final predictions are shown on the upper right: contact maps at the three distance thresholds and distance as S-score. The branch predicting S-score is solving a regression problem with a Mean Squared Error loss, whereas the other three output branches are classifying the probability of a contact for the thresholds of 6, 8, or 10 Å, and use a cross entropy loss.

All intermediate convolutional layers are followed by a ELU (Clevert *et al.*, 2015) non linearity, Batch Normalization (Ioffe and Szegedy, 2015), and a Dropout layer (Srivastava *et al.*, 2014) with a probability of 0.1. Weights are initialized using the He distribution as described in the ResNet paper (He *et al.*, 2016). A weight decay of 10^{-12} on the L^2-norm is applied for regularization. All output layers have an appropriate function to ensure the output is in the valid range, sigmoid in our case.

The model is trained for 100 epochs (full passes over the training data) using the Adam optimizer at an initial learning rate of 0.001. The learning rate is decreased by a factor of 0.5 if the S-score training loss did not decrease over the last five epochs. The training dataset is shuffled after each epoch. The final model was selected by taking the epoch with minimum S-score loss on the validation dataset (epoch 29).

Figure S6: The example is the protein 2AV5 chain D from the benchmark set.



5 Instructions

5.1 Installation

On a machine with Python 3.5 or higher and pip, PconsC3 can be installed using the following commands:

```
pip install numpy Cython pythran
wget https://github.com/ElofssonLab/PconsC4/releases/download/0.2/\
pconsc4-0.2.tar.gz
pip install pconsc4-0.2.tar.gz
```

Once additional space is granted in PyPI (Python Package Index), a simple `pip install pconsc4` will be sufficient.

A deep learning backend compatible with Keras will also be needed. We recommend Tensorflow:

```
pip3 install -U tensorflow
```

5.2 Example usage

```
import pconsc4

# Load the deep learning model
model = pconsc4.get_pconsc4()

# It is possible to re-use on different alignments
pred_1 = pconsc4.predict(model, 'path/to/alignment1')
pred_2 = pconsc4.predict(model, 'path/to/alignment2')

# Show pred_1 on the screen:
import matplotlib.pyplot as plt
plt.imshow(pred_1['contacts']['cmap'])
plt.show()

# Save pred_2 in CASP format:
from pconsc4.utils import format_contacts_casp
print(format_contacts_casp(pred_2['contacts']['cmap'],
                           seq_2, min_sep=5))
```

5.3 Supported formats

The program accepts alignments in `.fasta`, `.a3m`, or `.aln`, without line wrapping.

6 Training and test sets

PconsC4 is trained on a set of 2891 proteins culled from PDB using PISCES (Wang and Dunbrack, Jr., 2003) on 2017-09-14 with the following criteria: maximum sequence identity 20%, minimum resolution 2.0 Å, maximum R-factor 0.3. Furthermore, chains from the same ECOD (Cheng *et al.*, 2014) H-group as any protein in the benchmark dataset or dating from after 2016-05-01 was removed to avoid potential overlap with the test datasets. Out of these 2891 proteins, 100 randomly selected proteins were used as a validation set for optimization.

Table S3: PDB code and chain identification for proteins in the training set

1A1XA 1A62A 1A73A 1A76A 1AF7A 1AH7A 1ALUA 1AOCA 1AOLA 1AYOA 1B8PA 1BGCA 1BGFA 1BKRA 1BM9A 1BX4A 1BX7A 1BXYA 1BYIA 1C1KA 1C7KA 1CC8A 1CDWA 1CEOA 1CFBA 1CHDA 1CMCA 1CQYA 1CUKA 1CV8A 1CXQA 1D0QA 1D2SA 1D2TA 1D4OA 1D9CA 1DCSA 1DD3A 1DD9A 1DG6A 1DJ7A 1DK8A 1DM9A 1DMGA 1DS1A 1DUSA 1DVOA 1DXGA 1DY5A 1DZFA 1E58A 1E7LA 1EAQA 1EB6A 1EG2A 1EGWA 1EJ8A 1ELKA 1EP0A 1EYEA 1EZGA 1EZWA 1F1EA 1F32A 1F39A 1F3VA 1F86A 1F9VA 1FCQA 1FCYA 1FIPA 1FLMA 1FOBA 1FSFA 1FUKA 1FVIA 1FX2A 1FYEA 1G2RA 1G2YA 1G3PA 1G5TA 1G6XA 1G8EA 1G8QA 1GAKA 1GMXA 1GNYA 1GPRA 1GS5A 1GS9A 1GSAA 1GV9A 1GVPA 1GWMA 1H2CA 1H4XA 1H8PA 1H8UA 1H97A 1H99A 1H9MA 1HCZA 1HDOA 1HH8A 1HQ0A 1HQ1A 1HUFA 1HUWA 1HXIA 1HXNA 1HXRA 1HZ6A 1HZTA 1I1WA 1I27A 1I2KA 1I2TA 1I4JA 1I71A 1I8AA 1IAPA 1ID0A 1IFGA 1IFRA 1IGQA 1IIBA 1IJVA 1IQZA 1IRQA 1ISUA 1IZCA 1IZMA 1J0PA 1J1TA 1J27A 1J33A 1J3AA 1J5PA 1J5UA 1J5XA 1J7XA 1J8BA 1J98A 1JB3A 1JBEA 1JBZA 1JEOA 1JERA 1JF4A 1JG1A 1JHJA 1JHSA 1JI7A 1JIDA 1JL1A 1JM1A 1JNIA 1JO0A 1JOSA 1JOVA 1JR7A 1JX6A 1JYHA 1K3XA 1K4IA 1K5CA 1K77A 1K7CA 1K7JA 1K8WA 1KGDA 1KHXA 1KMTA 1KNGA 1KNMA 1KOEA 1KP6A 1KPTA 1KQ6A 1KS9A 1KT6A 1KYFA 1KZFA 1L2PA 1L3KA 1L3PA 1L6PA 1L8JA 1LC0A 1LDDA 1LFPA 1LJOA 1LKIA 1LKKA 1LMIA 1LNIA 1LS1A 1LTMA 1LUZA 1LWBA 1LXJA 1LZLA 1M2DA 1M40A 1M4LA 1M65A 1M9ZA 1MAIA 1MC2A 1MIXA 1MJ5A 1MK0A 1MKKA 1MNNA 1MSCA 1MUNA 1MY7A 1N08A 1N12A 1N1FA 1N8VA 1N9PA 1NARA 1NEPA 1NFPA 1NG6A 1NIGA 1NIJA 1NJHA 1NJRA 1NKDA 1NKOA 1NKZA 1NMYA 1NNXA 1NRGA 1NRIA 1NU0A 1NWZA 1O22A 1O4WA 1O54A 1O6DA 1O7IA 1O8XA 1OAAA 1OD3A 1ODMA 1OH0A 1OHLA 1OI7A 1OISA 1OK0A 1OKSA 1OQJA 1OU8A 1OW1A 1OZ2A 1OZ9A 1P0HA 1P3CA 1P90A 1PG6A 1PQHA 1PSRA 1PSWA 1PUJA 1PV5A 1PZ4A 1Q08A 1Q0PA 1Q2HA 1Q42A 1Q4AA 1Q4CA 1Q4EA 1Q5YA 1Q73A 1Q8CA 1Q8DA 1Q9UA 1QCSA 1QCZA 1QJ8A 1QJPA 1QNRA 1QR0A 1QSTA 1QV1A 1QW2A 1QWGA 1QZMA 1R0UA 1R29A 1R3DA 1R4XA 1R5LA 1R6DA 1R6JA 1R75A 1R7JA 1RFYA 1RG8A 1RH6A 1RI6A 1RKIA 1RL6A 1RLHA 1RMMA 1RMOA 1RO2A 1ROCA 1RSSA 1RTQA 1RTTA 1RV9A 1RXIA 1RYLA 1RYQA 1S29A 1S2OA 1S2XA 1S3CA 1S4KA 1S7KA 1S7ZA 1S9UA 1SAUA 1SBXA 1SD4A 1SDIA 1SEIA 1SENA 1SF9A 1SFPA 1SFSA 1SFUA 1SH8A 1SQHA 1SR8A 1SRAA 1SZ7A 1SZHA 1T07A 1T3YA 1T4AA 1T6SA 1T8KA 1T92A 1T95A 1T9FA 1T9IA 1TAGA 1TFEA 1TIFA 1TIGA 1TKEA 1TP6A 1TQ5A 1TQGA 1TS9A 1TT8A 1TU1A 1TUAA 1TXJA 1TXLA 1U07A 1U6TA 1U9LA 1UCDA 1UCHA 1UCRA 1UCSA 1UDVA 1UFIA 1UFYA 1UI0A 1UJ8A 1UJCA 1UKFA 1UNKA 1UNQA 1UOYA 1US0A 1USGA 1UT7A 1UUYA 1UV7A 1UX6A 1UXOA 1UZ3A 1UZKA 1V05A 1V0AA 1V2ZA 1V6PA 1V74A 1V77A 1V96A 1V9MA 1VAJA 1VBIA 1VBWA 1VCCA 1VD6A 1VGJA 1VH5A 1VHNA 1VJFA 1VJLA 1VK1A 1VK4A 1VKKA 1VL7A 1VLSA 1VLYA 1VMBA 1VMGA 1VMHA 1VP8A 1VR7A 1VR8A 1VYIA 1VYKA 1VZMA 1W0HA 1W0NA 1W1GA 1W4SA 1W53A 1WCWA 1WEHA 1WERA 1WHIA 1WHOA 1WJXA 1WKCA 1WLJA 1WMHA 1WN2A 1WNAA 1WNHA 1WOUA 1WQ6A 1WV3A 1WWCA 1WWIA 1X0TA 1X2IA 1X3KA 1X6IA 1X6OA 1X6ZA 1X8QA 1XAUA 1XBIA 1XCLA 1XDZA 1XE1A 1XJ3A 1XJVA 1XKRA 1XLQA 1XMKA 1XMTA 1XQOA 1XSVA 1XT5A 1XTPA 1XUBA 1XW3A 1XYIA 1Y08A 1Y5HA 1Y63A 1Y6XA 1Y71A 1Y80A 1Y88A 1Y8AA 1Y9LA 1YB3A 1YD0A 1YDIA 1YFQA 1YG9A 1YGTA 1YI9A 1YJFA 1YLEA 1YLIA 1YLXA 1YN4A 1YOZA 1YQ5A 1YQEA 1YQHA 1YQSA 1YRKA 1YWMA 1YZVA 1Z0NA 1Z0PA 1Z21A 1Z67A 1Z6MA 1Z6NA 1Z9LA 1ZARA 1ZCEA 1ZD0A 1ZDYA 1ZGKA 1ZHVA 1ZI8A 1ZK4A 1ZK5A 1ZLDA 1ZM8A 1ZMAA 1ZMIA 1ZO2A 1ZPSA 1ZS9A 1ZT3A 1ZUUA 1ZWXA 1ZXXA 1ZZKA 2A0BA 2A26A 2A6ZA 2A72A 2AGKA 2AH5A 2AIBA 2AJ6A 2ANXA 2AP3A 2ASBA 2ASKA 2ATZA 2AWLA 2AWMA 2AXCA 2AXOA 2AXWA 2AYDA 2B0AA 2B18A 2B1YA 2B3PA 2B4WA 2B8IA 2B8MA 2B97A 2B9DA 2BCQA 2BDRA 2BF5A 2BFWA 2BI0A 2BJFA 2BJNA 2BKAA 2BKFA 2BKMA 2BL8A 2BOPA 2BT9A 2BTIA 2BW0A 2BZ1A 2BZGA 2C2IA 2C3VA 2C60A 2C71A 2CAYA 2CB8A 2CCQA 2CCVA 2CG7A 2CHHA 2CIUA 2CIWA 2CJJA 2CKKA 2CO3A 2CPGA 2CS7A 2CULA 2CVBA 2CVEA 2CW9A 2CWSA 2CWYA 2CX7A 2CXAA 2CXHA 2CXYA 2CY5A 2CYJA 2CZSA 2D2EA 2D48A 2D4PA 2D4XA 2D59A 2D5MA 2D68A 2D7VA 2D81A 2DB7A 2DDXA 2DEJA 2DFAA 2DLBA 2DNJA 2DP9A 2DPMA 2DQAA 2DQLA 2DVKA 2DWKA 2DXAA 2DXQA 2DYIA 2E12A 2E1FA 2E2OA 2E3HA 2E56A 2E5YA 2E6MA 2E6XA 2E7VA 2E85A 2E8EA 2EAQA 2EBEA 2EBNA 2EFVA 2EH3A 2EHPA 2EI9A 2EJ9A 2EK0A 2EKLA 2ELAA 2ENDA 2ERFA 2ERVA 2ESSA 2ET1A 2ETJA 2EVRA 2EW0A 2F1FA 2F1NA 2F22A 2F23A 2F46A 2F69A 2F9HA 2FA5A 2FB6A 2FB9A 2FBQA 2FCWA 2FD4A 2FHTA 2FHZA 2FI1A 2FJ8A 2FL4A 2FM9A 2FMAA 2FOZA 2FQ4A 2FSJA 2FSQA 2FSRA 2FSUA 2FUEA 2FUJA 2FUPA 2FVVA 2FVYA 2FWHA 2FWTA 2FYGA 2FZPA 2G0CA 2G1UA 2G3RA 2G40A 2G7OA 2G7SA 2GA1A 2GAUA 2GENA 2GJ3A 2GJLA 2GKEA 2GKGA 2GKPA 2GMQA 2GNOA 2GOMA 2GPEA 2GPIA 2GQTA 2GS5A 2GSVA 2GU3A 2GUDA 2GUIA 2GUKA 2GXQA 2GZQA 2GZSA 2H0UA 2H1VA 2H30A 2H5PA 2H6EA 2H7OA 2H8EA 2H9WA 2HALA 2HBAA 2HC8A 2HE7A 2HHCA 2HHZA 2HINA 2HJEA 2HJNA 2HJOA 2HKVA 2HL7A 2HL9A 2HLJA 2HLYA 2HNGA 2HP7A 2HQ4A 2HQQA 2HQZA 2HRSA 2HS1A 2HU9A 2HUHA 2HUJA 2HW4A 2HX0A 2HX5A 2HXXA 2HYTA 2HZCA 2I02A 2I2CA 2I53A 2I5HA 2I5UA 2I6DA 2I8DA 2I8TA 2I9CA 2I9IA 2I9WA 2IA7A 2IAYA 2IC2A 2IC6A 2IDLA 2IGPA 2II2A 2IIHA 2IJQA 2IKKA 2ILKA 2ILRA 2IM9A 2IMFA 2IN3A 2INWA 2IP6A 2IRXA 2IU1A 2IUWA 2IVNA 2IXMA 2IY2A 2IYVA 2IZXA 2J0PA 2J22A 2J6AA 2J6BA 2J73A 2J8KA 2J97A 2J9WA 2JAYA 2JDCA 2JE3A 2JEKA 2JFRA 2JFUA 2JG6A 2JH1A 2JKUA 2JLIA 2MCMA 2NLRA 2NLVA 2NNUA 2NPNA 2NQ3A 2NR7A 2NRRA 2NS0A 2NSAA 2NSFA 2NSZA 2NUHA 2NUJA 2NVHA 2NWFA 2NWHA 2NX2A 2NXFA 2NZCA 2O0AA 2O0MA 2O1QA 2O24A 2O2XA 2O30A 2O38A 2O4AA 2O4DA 2O4TA 2O5HA 2O6LA 2O7AA 2O8NA 2O8PA 2O8QA 2O90A 2O9SA 2OA2A 2OAFA 2OB5A 2OBLA 2OCTA 2OD0A 2OD5A 2OEBA 2OEEA 2OF3A 2OFCA 2OFZA 2OHWA 2OIXA 2OJHA 2OKMA 2OKTA 2OKUA 2OMDA 2OMLA 2OO3A 2OOCA 2OOKA 2OPCA 2OQBA 2OQZA 2OS0A 2OSOA 2OU3A 2OU5A 2OU6A 2OV0A 2OVGA 2OVJA 2OVSA 2OXLA 2OXNA 2OXOA 2OY9A 2OYAA 2OYNA 2OYRA 2OZEA 2OZHA 2OZJA 2OZNA 2OZTA 2P08A 2P0NA 2P17A 2P2VA 2P38A 2P4FA 2P51A 2P58A 2P5DA 2P5KA 2P65A 2P67A 2P6VA 2P84A 2P9WA 2PA7A 2PAGA 2PC1A 2PEBA 2PETA 2PFIA 2PH0A 2PIEA 2PK8A 2PMAA 2PN1A 2PN2A 2PORA 2PPVA 2PQ8A 2PR7A 2PRVA 2PRXA 2PSPA 2PU3A 2PV4A 2PVBA 2PWWA 2Q12A 2Q3TA 2Q3VA 2Q4MA 2Q6KA 2Q82A 2Q8PA 2Q9KA 2Q9RA 2QDJA 2QF4A 2QFEA 2QGUA 2QHQA 2QJLA 2QJZA 2QK1A 2QKVA 2QL8A 2QLTA 2QMLA 2QMQA 2QNGA 2QNIA 2QNKA 2QNLA 2QRUA 2QSBA 2QSKA 2QSWA 2QT1A 2QTDA 2QU1A 2QUDA 2QUOA 2QUPA 2QYWA 2QZ0A 2QZQA 2R01A 2R0XA 2R16A 2R2YA 2R31A 2R4GA 2R4QA 2R9FA 2RA9A 2RBKA 2RCIA 2RDCA 2RDQA 2RE2A 2RFAA 2RFRA 2RG8A 2RH2A 2RH3A 2RHFA 2RIQA 2RJ2A 2RJIA 2RK3A 2RKHA 2RKLA 2RKQA 2SAKA 2TGIA 2UU8A 2UURA 2UV4A 2UY2A 2UYOA 2UZ8A 2V03A 2V05A 2V1MA 2V1TA 2V33A 2V3GA 2V3SA 2V75A 2V79A 2V7FA 2V7SA 2V89A 2V9BA 2V9VA 2VB1A 2VC8A 2VDFA 2VDJA 2VEZA 2VGAA 2VH3A 2VHKA 2VJWA 2VK2A 2VLAA 2VMHA 2VOVA 2VPAA 2VPBA 2VQ2A 2VVWA 2VXGA 2VXNA 2VXZA 2VY8A 2VZCA 2W0GA 2W15A 2W1RA 2W2RA 2W31A 2W39A 2W3GA 2W3QA 2W47A 2W56A 2W7AA 2W9YA 2WAOA 2WBNA 2WCRA 2WDSA 2WF7A 2WFBA 2WFIA 2WFOA 2WH7A 2WJ5A 2WJRA 2WK1A 2WL1A 2WLTA 2WLVA 2WNFA 2WP7A 2WQFA 2WQPA 2WTAA 2WTPA 2WURA 2WVIA 2WWEA 2WY4A 2WZOA 2X2UA 2X3GA 2X3MA 2X46A 2X49A 2X4LA 2X55A 2X5CA 2X5GA 2X5NA 2X5RA 2X5XA 2X6UA 2X8WA 2X9ZA 2XBGA 2XCBA 2XETA 2XF7A 2XFVA 2XGRA 2XHFA 2XIOA 2XJ4A 2XLGA 2XM5A 2XODA 2XOMA 2XPWA 2XSEA 2XSKA 2XTPA 2XU3A 2XU8A 2XVSA 2XVYA 2XY4A 2XZ2A 2XZ7A 2Y0GA 2Y0OA 2Y1BA 2Y2ZA 2Y39A 2Y43A 2Y4XA 2Y5PA 2Y6CA 2Y6XA 2Y78A 2Y8GA 2Y8YA 2Y9UA 2YG2A 2YGOA 2YH9A 2YHCA 2YJMA 2YK4A 2YMVA 2YN0A 2YV4A 2YVQA 2YVTA 2YWJA 2YWWA 2YXFA 2YYOA 2YZTA 2YZYA 2Z0XA 2Z14A 2Z1CA 2Z2BA 2Z2NA 2Z51A 2Z5BA 2Z5EA 2Z5WA 2Z72A 2Z84A 2Z8LA 2Z8PA 2Z98A 2ZAYA 2ZCAA 2ZCUA 2ZCWA 2ZFGA 2ZGLA 2ZHJA 2ZOUA 2ZOVA 2ZPMA 2ZQOA 2ZVCA 2ZZJA 3A02A 3A0NA 3A0SA 3A0YA 3A1GA 3A27A 3A2ZA 3A4CA 3A4JA 3A57A 3A9FA 3AAFA 3ACHA 3ADOA 3ADYA 3AEIA 3AGNA 3AJ4A 3AJDA 3AKSA 3B09A 3B1VA 3B33A 3B42A 3B49A 3B4QA 3B6EA 3B79A

Continued on next page

3B7CA 3B7HA 3BA3A 3BBYA 3BCYA 3BDIA 3BEDA 3BFMA 3BHNA 3BI7A 3BJNA 3BK5A 3BLEA 3BN0A
3BOEA 3BPJA 3BQAA 3BS4A 3BUTA 3BV4A 3BV8A 3BWZA 3BY8A 3C0SA 3C1QA 3C26A 3C2EA 3C4BA
3C5VA 3C6AA 3C70A 3C7XA 3C8IA 3C8LA 3C8MA 3C8XA 3C9PA 3CANA 3CB9A 3CBNA 3CBZA 3CCDA
3CD1A 3CECA 3CEXA 3CG6A 3CHJA 3CHMA 3CHVA 3CI3A 3CIMA 3CJEA 3CKKA 3CKMA 3CLAA 3CP0A
3CP3A 3CQBA 3CT5A 3CT6A 3CZ6A 3D06A 3D0FA 3D0JA 3D1PA 3D2YA 3D32A 3D33A 3D3BA 3D3MA 3D4EA
3D79A 3D7IA 3D9NA 3DA5A 3DASA 3DCZA 3DD6A 3DD7A 3DDJA 3DDTA 3DEFA 3DEWA 3DF7A 3DF8A
3DFFA 3DFGA 3DGTA 3DKMA 3DKRA 3DLCA 3DMNA 3DNJA 3DNPA 3DQYA 3DSMA 3DT5A 3DXEA 3DXYA
3DZ1A 3E0HA 3E3XA 3E8TA 3E99A 3E9AA 3E9VA 3EA6A 3EDHA 3EEAA 3EERA 3EF8A 3EGAA 3EHGA
3EIPA 3EJCA 3EJFA 3EJVA 3ELFA 3EMFA 3EMIA 3ENUA 3EO6A 3EOIA 3ER7A 3ESMA 3EUNA 3EURA
3EUSA 3EVFA 3EVPA 3EXMA 3EYEA 3EZHA 3F14A 3F2EA 3F2ZA 3F43A 3F4MA 3F5BA 3F5HA 3F5RA
3F6GA 3F7EA 3F9SA 3FAJA 3FB9A 3FBLA 3FBUA 3FCNA 3FDQA 3FF0A 3FF2A 3FGHA 3FGYA 3FHGA
3FILA 3FKEA 3FL2A 3FLBA 3FLJA 3FMYA 3FN2A 3FNCA 3FNDA 3FPNA 3FPRA 3FPWA 3FRHA 3FRRA
3FSOA 3FSSA 3FT7A 3FTDA 3FTJA 3FX7A 3FXHA 3FYMA 3FYNA 3FYRA 3FZEA 3G16A 3G1JA 3G2BA
3G36A 3G3TA 3G46A 3G5TA 3G7PA 3G85A 3G91A 3G98A 3G9RA 3GA4A 3GA8A 3GBWA 3GBYA 3GDWA
3GGMA 3GHAA 3GI7A 3GJYA 3GK6A 3GKJA 3GMGA 3GMXA 3GOEA 3GOHA 3GONA 3GP6A 3GPIA 3GPVA
3GQJA 3GRDA 3GS2A 3GS9A 3GT0A 3GVAA 3GWIA 3GXHA 3GZRA 3H0NA 3H1ZA 3H20A 3H36A 3H4OA
3H4XA 3H51A 3H5JA 3H6QA 3H74A 3H75A 3H79A 3H7IA 3HA9A 3HBMA 3HC7A 3HFTA 3HJZA 3HM4A
3HMSA 3HP7A 3HPDA 3HRGA 3HSHA 3HTVA 3HVWA 3HWUA 3HYNA 3HZ7A 3HZ8A 3HZPA 3I0WA 3I10A
3I6IA 3I7MA 3IB5A 3IB7A 3IBWA 3ICVA 3IDUA 3IE4A 3IEEA 3IEZA 3IG9A 3IHTA 3II2A 3IIDA 3ILSA 3IM1A
3IMKA 3INOA 3IO3A 3IP0A 3IPFA 3IPJA 3IQUA 3IR4A 3IRBA 3ISXA 3ITFA 3IUOA 3IUWA 3IV3A 3IV4A
3IVVA 3IWFA 3IX3A 3IXLA 3JSRA 3JTZA 3JU3A 3JUDA 3JX9A 3JXOA 3JYOA 3JZ9A 3K0ZA 3K1HA 3K1UA
3K29A 3K3VA 3K5JA 3K67A 3K69A 3K8UA 3KA5A 3KBGA 3KBQA 3KBYA 3KDWA 3KE7A 3KEVA 3KF6A
3KFFA 3KFOA 3KGKA 3KHIA 3KJHA 3KKBA 3KKGA 3KLQA 3KOGA 3KOJA 3KSNA 3KUVA 3KVHA 3KWRA
3KXTA 3KYZA 3L00A 3L0FA 3L1NA 3L23A 3L2CA 3L32A 3L39A 3L51A 3L60A 3L6BA 3L7YA 3L81A 3L8WA
3LAGA 3LAXA 3LAZA 3LB2A 3LD7A 3LFRA 3LHCA 3LHEA 3LHIA 3LHNA 3LI9A 3LLOA 3LLUA 3LMZA
3LO8A 3LPZA 3LQ9A 3LQNA 3LRUA 3LTIA 3LUYA 3LW3A 3LWAA 3LWXA 3LX3A 3LX7A 3LXYA 3LYDA
3LYEA 3LYGA 3LYHA 3LYWA 3LYYA 3LZWA 3M0BA 3M1XA 3M3PA 3M4IA 3M66A 3M7AA 3M7KA 3M86A
3M8JA 3M9QA 3MABA 3MAGA 3MAOA 3MC3A 3MCBA 3MCQA 3MDQA 3ME7A 3MHXA 3ML3A 3MMHA
3MQQA 3MQZA 3MR0A 3MSWA 3MTQA 3MTXA 3MVCA 3MVNA 3MVSA 3MW8A 3MWZA 3MX7A 3MXNA
3MXZA 3MZ0A 3N01A 3N08A 3N0RA 3N17A 3N1EA 3N2TA 3N4JA 3N6YA 3N72A 3NA6A 3NBCA 3NBMA
3NDQA 3NE0A 3NFTA 3NH4A 3NJCA 3NJNA 3NKGA 3NKLA 3NL9A 3NO2A 3NO7A 3NOJA 3NR5A 3NRFA
3NRLA 3NRWA 3NSOA 3NTKA 3NUFA 3NUQA 3NYMA 3O0LA 3O0PA 3O12A 3O15A 3O48A 3O4PA 3O6CA
3O7BA 3O7WA 3OBLA 3OCJA 3OCUA 3OEBA 3OHGA 3OIOA 3OIZA 3OJ0A 3OKXA 3OL3A 3OMDA 3OMYA
3ON9A 3ONHA 3ONJA 3OOSA 3OOUA 3ORUA 3OSDA 3OSEA 3OSTA 3OV5A 3OXPA 3P02A 3P0KA 3P2TA
3P4EA 3P4HA 3PE6A 3PESA 3PF6A 3PFGA 3PFTA 3PI7A 3PIWA 3PLUA 3PMSA 3POJA 3POWA 3PP2A
3PSHA 3PT3A 3PT5A 3PUCA 3PVHA 3PVVA 3PYFA 3PYWA 3Q13A 3Q1NA 3Q1XA 3Q20A 3Q2UA 3Q46A
3Q6BA 3Q7RA 3QAGA 3QC0A 3QC7A 3QDHA 3QH6A 3QHPA 3QL9A 3QLEA 3QM9A 3QP4A 3QPAA 3QR7A
3QWLA 3QX1A 3QY7A 3QZBA 3QZMA 3QZXA 3R15A 3R2QA 3R5TA 3R6DA 3R87A 3RAYA 3RC1A 3RFNA
3RJVA 3RKCA 3RKGA 3RL5A 3RLGA 3RLKA 3RLOA 3RLSA 3RM3A 3RMHA 3RMQA 3RN4A 3RNLA 3RNVA
3RO3A 3ROFA 3RPJA 3RPZA 3RQ4A 3RR6A 3RVCA 3RX9A 3S25A 3S2RA 3S4EA 3S6EA 3S6FA 3S83A 3S8IA
3S8SA 3S9XA 3SBMA 3SD2A 3SIBA 3SIGA 3SJMA 3SK7A 3SMVA 3SMZA 3SNOA 3SNZA 3SO6A 3SOJA 3SOVA
3SU6A 3SXMA 3SXUA 3SY1A 3T0HA 3T47A 3T4RA 3T7AA 3T7LA 3T7ZA 3T8JA 3T92A 3T9YA 3TBDA 3TBNA
3TE8A 3TEEA 3TEUA 3TG2A 3TJ8A 3TJMA 3TJYA 3TNTA 3TOWA 3TR0A 3TS9A 3TT9A 3TU8A 3TYSA
3TZ6A 3U0VA 3U1DA 3U1UA 3U25A 3U2AA 3U4GA 3U4VA 3U5SA 3U5VA 3U6GA 3U7ZA 3U81A 3U8VA 3U97A
3U9JA 3UAWA 3UC9A 3UEJA 3UFEA 3UFZA 3UI4A 3UJCA 3ULBA 3ULJA 3ULTA 3UMHA 3UMZA 3UP3A
3URGA 3URRA 3USHA 3UUEA 3UV0A 3UX2A 3V0EA 3V1EA 3V46A 3V68A 3V7BA 3VBCA 3VDJA 3VEJA
3VGIA 3VGPA 3VHJA 3VMVA 3VN5A 3VNEA 3VORA 3VP5A 3VPZA 3VQJA 3VTTA 3VUBA 3VURA 3VVVA
3VWCA 3VXVA 3VZHA 3W06A 3W07A 3W0EA 3W1OA 3W2ZA 3W43A 3WCQA 3WDCA 3WDNA 3WGXA
3WH1A 3WH2A 3WHJA 3WI9A 3WITA 3WJPA 3WJTA 3WMVA 3WS7A 3WURA 3WVAA 3WW9A 3WX4A
3WZ3A 3WZSA 3X0FA 3X0IA 3X0TA 3X2MA 3X34A 3ZBDA 3ZDBA 3ZFIA 3ZGHA 3ZH5A 3ZHIA 3ZIHA 3ZJAA
3ZN3A 3ZN4A 3ZNVA 3ZRGA 3ZRXA 3ZSJA 3ZSUA 3ZUCA 3ZUDA 3ZUIA 3ZXCA 3ZXNA 3ZY7A 3ZYHA
3ZYPA 3ZZLA 3ZZOA 3ZZPA 3ZZYA 4A02A 4A29A 4A2VA 4A3PA 4A3ZA 4A4JA 4A56A 4A5UA 4A61A 4A6QA
4A7UA 4A8TA 4ABLA 4ACJA 4ADZA 4AE7A 4AEQA 4AFFA 4AFMA 4AIVA 4AIWA 4AL5A 4ALZA 4AQOA
4ATEA 4ATGA 4ATMA 4AU1A 4AVAA 4AXOA 4AXQA 4B1LA 4B21A 4B2FA 4B4CA 4B4DA 4B5OA 4B89A
4B8VA 4B8XA 4B9GA 4BFCA 4BFOA 4BGCA 4BGPA 4BHRA 4BI3A 4BJ0A 4BJAA 4BJIA 4BK0A 4BN4A
4BOQA 4BOUA 4BPFA 4BPSA 4BQYA 4BT7A 4BTBA 4BWOA 4BYZA 4BZAA 4BZPA 4C1WA 4C5PA 4C6AA
4C6SA 4C81A 4C84A 4C98A 4CBEA 4CC2A 4CCVA 4CCWA 4CD8A 4CFIA 4CG1A 4CHEA 4CICA 4CMFA
4CO8A 4CV7A 4CZ5A 4CZGA 4D01A 4D0QA 4D53A 4D5RA 4D6QA 4D8BA 4DB5A 4DB6A 4DDPA 4DE9A
4DI9A 4DJGA 4DK2A 4DLMA 4DNYA 4DOLA 4DQ9A 4DQJA 4DQNA 4DT5A 4DVCA 4DYQA 4DZOA 4E0AA
4E14A 4E1PA 4E1SA 4E40A 4E4RA 4EA9A 4EBGA 4EFOA 4EG9A 4EHCA 4EHSA 4EHXA 4EICA 4EKFA
4EL6A 4EO0A 4EP4A 4EPZA 4EQPA 4ERCA 4ERNA 4ERRA 4ERYA 4ES1A 4ES7A 4ESMA 4ESQA 4ETXA
4EULA 4EUNA 4EV1A 4EVUA 4EW5A 4EW7A 4EXOA 4EXRA 4EYCA 4EYSA 4EZ8A 4F0WA 4F2EA 4F54A
4F55A 4F67A 4F80A 4F87A 4F98A 4FBSA 4FD9A 4FDBA 4FE3A 4FF5A 4FGOA 4FH3A 4FK9A 4FLBA 4FTFA
4FUUA 4FVGA 4FXIA 4FZPA 4G08A 4G0XA 4G29A 4G3NA 4G3VA 4G4KA 4G54A 4G5AA 4G75A 4G78A
4G7XA 4G9QA 4G9SA 4GA2A 4GB5A 4GBMA 4GC3A 4GCNA 4GDZA 4GEIA 4GIOA 4GJZA 4GMQA 4GOFA
4GS3A 4GT8A 4GT9A 4GUCA 4GWBA 4GZCA 4GZJA 4H08A 4H14A 4H3UA 4H4NA 4H7WA 4H86A 4H8EA
4HCJA 4HCSA 4HDDA 4HE6A 4HEIA 4HEOA 4HFVA 4HHXA 4HI8A 4HIKA 4HJIA 4HKGA 4HLSA 4HLYA
4HNOA 4HP4A 4HROA 4HRVA 4HS1A 4HS2A 4HTGA 4HTLA 4HTUA 4HU2A 4HVYA 4HWFA 4HWMA 4HY4A
4HYLA 4HYQA 4HZOA 4I1FA 4I1KA 4I4OA 4I66A 4I6RA 4I6XA 4I71A 4I8HA 4I8IA 4I90A 4I95A 4I9OA 4IAUA
4IC3A 4IC9A 4IEJA 4IFAA 4IGIA 4II9A 4IIKA 4IILA 4IKCA 4IKGA 4IL7A 4IN0A 4IPVA 4IRFA 4IT6A 4IUSA
4IX7A 4J32A 4J37A 4J39A 4J42A 4J4RA 4J5RA 4J8AA 4J8CA 4J8SA 4JB3A 4JBDA 4JCCA 4JCYA 4JDEA
4JDUA 4JEJA 4JF5A 4JF8A 4JG2A 4JGLA 4JHTA 4JJ2A 4JMPA 4JNFA 4JNUA 4JP6A 4JQFA 4JTMA 4JXHA
4K0DA 4K0NA 4K4OA 4K7BA 4K82A 4K8WA 4KA9A 4KAGA 4KDDA 4KDRA 4KDWA 4KEFA 4KEXA 4KH8A
4KIIA 4KM6A 4KRUA 4KSNA 4KSYA 4KT3A 4KTWA 4KUIA 4KV2A 4KW4A 4KZKA 4L0JA 4L1IA 4L3UA
4L8AA 4L9EA 4L9NA 4LBAA 4LD1A 4LEBA 4LF0A 4LGJA 4LHSA 4LJ6A 4LLDA 4LMYA 4LPQA 4LQTA
4LQZA 4LRUA 4LTTA 4LUAA 4LUPA 4LWSA 4LWUA 4LZHA 4M0NA 4M5BA 4M7TA 4M8AA 4M91A 4M9KA
4MAIA 4MAKA 4MDAA 4MDWA 4MDYA 4ME2A 4MF5A 4MFKA 4MGQA 4MGSA 4MI7A 4MISA 4MJDA 4MJFA
4MLOA 4MNNA 4MNOA 4MP8A 4MT2A 4MT8A 4MTMA 4MTUA 4MUMA 4MVKA 4MXTA 4MYKA 4MZCA
4MZJA 4N1IA 4N30A 4N67A 4N6QA 4N77A 4N7CA 4N7FA 4NC6A 4NC7A 4NDJA 4NDSA 4NE3A 4NGDA 4NI6A
4NL9A 4NMIA 4NNOA 4NOAA 4NOHA 4NS5A 4NTDA 4NWBA 4NYQA 4O0AA 4O0KA 4O1TA 4O65A 4O66A
4O6GA 4O6UA 4O7JA 4O7KA 4OA3A 4OD6A 4OE9A 4OFAA 4OIEA 4OKEA 4ONRA 4OOXA 4OQ8A 4OQPA
4OSNA 4OTMA 4OTNA 4OUNA 4OUSA 4OVYA 4OY3A 4P0MA 4P1MA 4P3AA 4P3VA 4P5NA 4P7XA 4P98A
4P9IA 4PAUA 4PBDA 4PBOA 4PDNA 4PF3A 4PGRA 4PH2A 4PH8A 4PHJA 4PI8A 4PKMA 4PLZA 4PNOA
4PQDA 4PQQA 4PS2A 4PS6A 4PSFA 4PSYA 4PUIA 4PWWA 4PWYA 4PXVA 4PZ0A 4PZ1A 4PZ3A 4Q0PA
4Q2LA 4Q2SA 4Q34A 4Q53A 4Q63A 4Q6VA 4Q7OA 4QA8A 4QB0A 4QEKA 4QHJA 4QM6A 4QMAA 4QMHA
4QNDA 4QP5A 4QPNA 4QPTA 4QRKA 4QTQA 4QUKA 4QXLA 4QY7A 4R03A 4R1BA 4R1JA 4R2HA 4R3QA
4R5RA 4R7QA 4R9PA 4RAXA 4RAYA 4RBRA 4RD7A 4RDBA 4REOA 4RGDA 4RGIA 4RK4A 4RMLA 4RNAA
4RO3A 4RP3A 4RPTA 4RT1A 4RTIA 4RU3A 4RUVA 4RVQA 4RWUA 4RXIA 4RXVA 4RYOA 4RZ9A 4S1PA
4S24A 4TJVA 4TKBA 4TMDA 4TQ1A 4TQ2A 4TQRA 4TTWA 4TXRA 4TYZA 4U12A 4U3VA 4U6NA 4U6OA
4U89A 4U9BA 4UDSA 4UE8A 4UHHA 4UHQA 4UHTA 4UIOA 4UJ7A 4UMGA 4UNUA 4UOBA 4UONA 4UOSA
4UQWA 4UQXA 4UU3A 4UVQA 4UY5A 4UYIA 4UYRA 4V0KA 4V0SA 4V17A 4V1GA 4V1JA 4V1KA 4V23A
4V3IA 4W5XA 4W6TA 4W79A 4W7WA 4W8HA 4W8PA 4W8QA 4W97A 4W9ZA 4WBJA 4WCJA 4WDCA 4WE2A
4WEEA 4WH9A 4WHEA 4WHIA 4WIQA 4WJIA 4WJQA 4WJTA 4WN5A 4WNBA 4WOLA 4WPGA 4WPKA
4WPYA 4WRIA 4WSFA 4WT3A 4WTPA 4WUIA 4WUTA 4WV4A 4WWBA 4WWFA 4WY4A 4WY9A 4WYHA
4WZ0A 4WZXA 4X2RA 4X3IA 4X5MA 4X5PA 4X7GA 4X8QA 4X8YA 4X9RA 4X9XA 4X9ZA 4XABA 4XALA
4XB4A 4XCVA 4XDUA 4XDXA 4XEHA 4XEPA 4XEZA 4XFSA 4XINA 4XKBA 4XKZA 4XO1A 4XPCA 4XPLA

Continued on next page

4XPXA 4XT6A 4XTBA 4XU4A 4XUOA 4XUWA 4XVVA 4XXXA 4XY5A 4XZFA 4Y2FA 4Y6WA 4Y7SA 4Y88A 4Y9IA 4Y9WA 4YAAA 4YBGA 4YE7A 4YG0A 4YI8A 4YMYA 4YNHA 4YNXA 4YQDA 4YSIA 4YTKA 4YTLA 4YTVA 4YU8A 4YUCA 4YWZA 4YX1A 4YZ6A 4Z2NA 4Z3HA 4Z47A 4Z4AA 4Z85A 4Z8WA 4ZAVA 4ZBHA 4ZC3A 4ZCEA 4ZDSA 4ZF7A 4ZGFA 4ZGMA 4ZHWA 4ZJ9A 4ZJHA 4ZJUA 4ZLDA 4ZMKA 4ZNKA 4ZOTA 4ZQXA 4ZSIA 4ZV5A 4ZVAA 4ZVCA 4ZVFA 4ZX2A 4ZY7A 4ZY9A 4ZYAA 4ZZ1A 5A0NA 5A1QA 5A3AA 5A3DA 5A4AA 5A67A 5A6WA 5A8CA 5A99A 5A9AA 5A9TA 5AE0A 5AG8A 5AGIA 5AGRA 5AIGA 5AIMA 5AIZA 5AJGA 5AJJA 5AMHA 5AMTA 5AOTA 5AOYA 5AOZA 5AQ0A 5AZBA 5AZWA 5AZYA 5B1NA 5B1RA 5B3KA 5B3PA 5B42A 5B4ZA 5BOIA 5BOWA 5BP9A 5BPXA 5BTYA 5BVLA 5BX1A 5BXGA 5BY4A 5BY5A 5BY8A 5BYKA 5C12A 5C17A 5C1EA 5C2MA 5C30A 5C5GA 5C5ZA 5C6SA 5CDKA 5CE7A 5CHPA 5CKLA 5CL8A 5CNWA 5COFA 5COTA 5COWA 5CR9A 5CTVA 5CUMA 5CVWA 5CWGA 5CXUA 5CYVA 5D22A 5D2FA 5D3KA 5D3XA 5D7UA 5D8MA 5D8VA 5DAWA 5DBLA 5DFSA 5DGJA 5DHDA 5DJEA 5DJOA 5DLBA 5DLOA 5DM2A 5DMAA 5DMDA 5DMMA 5DOMA 5DP2A 5DPGA 5DPOA 5DRFA 5DTCA 5DTXA 5DXLA 5DZ9A 5DZEA 5E0LA 5E10A 5E1WA 5E2CA 5E3QA 5E50A 5E5YA 5E6XA 5E9PA 5EC6A 5ECDA 5EDFA 5EDLA 5EH1A 5EIPA 5EIUA 5EJUA 5EKYA 5EMIA 5EOHA 5EP2A 5EPEA 5EQ0A 5EQVA 5EQZA 5EUBA 5EV5A 5EVFA 5EWOA 5EYNA 5EYRA 5EYSA 5EZQA 5EZUA 5F18A 5F47A 5F4CA 5F5LA 5F6RA 5F7GA 5F8ZA 5FA8A 5FAFA 5FAZA 5FB2A 5FBFA 5FCEA 5FD9A 5FDBA 5FENA 5FF3A 5FFHA 5FFSA 5FHVA 5FIDA 5FMTA 5FOTA 5FPZA 5FQIA 5FS4A 5FSVA 5FU5A 5FUBA 5FZSA 5G2HA 5G38A 5G3QA 5HB7A 5HBDA 5HBPA 5HBQA 5HD9A 5HDWA 5HE9A 5HGZA 5HI8A 5HJ1A 5HJ9A 5HKQA 5HQHA 5HQTA 5HRPA 5HRRA 5HRSA 5HTLA 5HUSA 5HWAA 5HWIA 5HWVA 5HXFA 5HYAA 5HZ6A 5HZ8A 5I0YA 5I1SA 5I45A 5I5CA 5I5NA 5I77A 5I8JA 5I9PA 5IA7A 5IBWA 5ICUA 5IDBA 5IDVA 5IEOA 5IG0A 5IGCA 5IGEA 5IGFA 5IHFA 5IHWA 5II6A 5II8A 5IIFA 5IMAA 5IMUA 5IO9A 5IOCA 5IODA 5IQNA 5IT3A 5ITMA 5IUCA 5IWBA 5IWHA 5IXBA 5IXHA 5J1ZA 5J2OA 5J3QA 5J4FA 5J4LA 5J4RA 5J5LA 5J8EA 5JA5A 5JBRA 5JDKA 5JE5A 5JEDA 5JEEA 5JELA 5JGJA 5JH8A 5JICA 5JIGA 5JJ2A 5JJOA 5JJXA 5JLBA 5NULA 7A3HA

Table S4: PDB code and chain identification for proteins in the validation set

1H97A 1J5XA 1N12A 1NFPA 1ODMA 1OZ9A 1QJPA 1SENA 1TQGA 1U6TA 1WLJA 1XYIA 1ZI8A 2CAYA 2HE7A 2HNGA 2HQQA 2IIHA 2NQ3A 2OA2A 2OO3A 2OSOA 2P58A 2P5DA 2PSPA 2RH2A 2RIQA 2W3GA 2WY4A 2YGOA 2Z8PA 2ZHJA 3A4CA 3D06A 3D1PA 3DT5A 3EJVA 3FILA 3FMYA 3G5TA 3GPIA 3H5JA 3H7IA 3HRGA 3INOA 3JU3A 3LO8A 3LZWA 3MC3A 3MMHA 3N6YA 3QPAA 3U4VA 3U8VA 3VGPA 3ZRXA 3ZZLA 4A02A 4A4JA 4BOQA 4CCVA 4ETXA 4EUNA 4F2EA 4HDDA 4I71A 4I95A 4JG2A 4K0NA 4KEXA 4L3UA 4N7CA 4NDSA 4ONRA 4Q2SA 4Q7OA 4RO3A 4TKBA 4V17A 4V3IA 4WHIA 4WJQA 4WNBA 4XKZA 4XVVA 4XXXA 4YBGA 4YE7A 4YTLA 4ZX2A 5A9AA 5A9TA 5AE0A 5C17A 5E0LA 5E50A 5FENA 5HGZA 5IOCA 5J2OA

Table S5: PDB code and chain identification for proteins in the benchmark set

1AHSC 1C2YD 1C9YA 1CCTA 1COZA 1DBRA 1DCHF 1EDIA 1EFDN 1F46B 1F68A 1FHIA 1FJRB 1FS0G 1G61A 1GJJA 1GLGA 1GPSA 1H68A 1I95E 1I97T 1IMBB 1IMXA 1IR1S 1IS9A 1JGPR 1JH0L 1K6LH 1KNVB 1KNYB 1KQPA 1LDIA 1LQKB 1M12A 1MB6A 1MFRP 1MR7A 1N2ZB 1N5BA 1N60C 1NQLB 1OAGA 1OTFF 1P3HE 1PCFA 1PDFE 1PS1A 1RD9D 1RH7C 1RL9A 1S3FB 1S68A 1SUDA 1SWXA 1SYHA 1TD4A 1TFKB 1TJLD 1UWZB 1VCRA 1VJNA 1VQZA 1W8AA 1W9GB 1WD5A 1WIGA 1WPVB 1X0PJ 1X48A 1X8HA 1X91A 1XBAA 1XQFA 1XS6A 1Y4HD 1Y60C 1YG6F 1YHQO 1YQFF 1YWSA 1Z7ME 1ZD7B 1ZJ0A 1ZWYC 2A84A 2A9KB 2AMCA 2AV5D 2B9NX 2BWEL 2C2OA 2CB6A 2CCCA 2CDMC 2CJRB 2CSMA 2D0PB 2D2CN 2DIOC 2E2AB 2EJNA 2F0RA 2FEEB 2FJCO 2GVIA 2H44A 2HGHA 2HI7B 2HJJA 2HL0A 2I9LI 2IA9E 2II9B 2J1KQ 2J3WA 2J8WB 2JOVA 2JYNA 2KYSA 2KZSA 2M0MB 2NQ2A 2NR9A 2OF5H 2OGFD 2OHCA 2OJ5C 2ONKC 2OPIA 2PAVP 2PLSF 2Q7RA 2QQDE 2QYFD 2RDOL 2RMRA 2RTBB 2VGRA 2VT8A 2WNKA 2WNYA 2XVTF 2Y9PB 2YADB 2YZOA 2ZITD 3A1JB 3ANZW 3AXGI 3B2UB 3B71B 3B7AA 3BLAB 3BP9B 3CPWT 3CVZC 3CXJC 3D2QD 3DBYF 3DKXB 3EB6B 3EW1A 3G74B 3GUVA 3GYVA 3GZFC 3H8DB 3H90A 3HPGL 3HTYJ 3I9OB 3IQZF 3K43B 3K8RB 3KZLA 3LW5L 3M71A 3MEZA 3N1GA 3NJSA 3O7JA 3OFEB 3OQIA 3P45J 3PC7B 3PJZA 3QE7A 3QNQA 3RBYB 3T3TB 3UD2B 3UWSA 3UYUB 3V3LB 3VHHB 3VX6A 3ZNUG 3ZUXA 4A5ZB 4AI3A 4ARDB 4AU0B 4DLHB 4E1YB 4E6FA 4F0DA 4HBRC 4IOSH 4IZJC 4J32B

Table S6: PDB code and chain identification for CASP12 targets

T0859 T0860 T0861 T0862 T0863 T0864 T0865 T0866 T0868 T0869 T0870 T0872 T0877 T0878 T0879 T0880 T0882 T0883 T0884 T0885 T0886 T0889 T0891 T0892 T0893 T0894 T0895 T0900 T0902 T0903 T0904 T0907 T0909 T0912 T0917 T0918 T0920 T0921 T0922 T0928 T0929 T0930 T0932 T0933 T0942 T0943 T0944 T0945 T0948

References

Baldassi, C., Zamparo, M., Feinauer, C., Procaccini, A., Zecchina, R., Weigt, M., and Pagnani, A. (2014a). Fast and accurate multivariate gaussian modeling of protein families: Predicting residue contacts and protein-interaction partners. *PLOS ONE*, **9**(3), 1–12.

Baldassi, C., Pagnani, A., Weigt, M., Feinauer, C., Procaccini, A., Zecchina, R., and Zamparo, M. (2014b). Gaussdca.jl - first release.

Cheng, H., Schaeffer, R., Liao, Y., Kinch, L., Pei, J., Shi, S., Kim, B., and Grishin, N. (2014). ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol*, **10**(12), e1003926.

Clevert, D., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). *CoRR*.

Dunn, S., Wahl, L., and Gloor, G. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, **24**(3), 333–340.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, **abs/1502.03167**.

Menéndez Hurtado, D., Uziela, K., and Elofsson, A. (2018). Deep transfer learning in the assessment of the quality of protein models. *ArXiv e-prints*.

Michel, M., Skwark, M. J., Menendez Hurtado, D., Ekeberg, M., and Elofsson, A. (2017). Predicting accurate contacts in thousands of pfam domain families using pconsc3. *Bioinformatics*, **33**(18), 2859–2866.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15**(1), 1929–1958.

Wang, G. and Dunbrack, Jr., R. L. (2003). PISCES: a protein sequence culling server. *Bioinformatics*, **19**(12), 1589–1591.