Documentation of NNCon 1.0

NNCon 1.0 is one of the most effective protein contact map predictors. It was ranked as one of the best residue-residue contact predictor in the Eighth Critical Assessment of Techniques for Protein Structure Prediction (CASP8), 2008 (for details, see Dr. Michael Tress's CASP8 assessment. Group id of NNCon (under the group name MULTICOM-CMFR) is RR069.).

NNCon 1.0 uses two steps to predict protein contact map. First, it uses a 2D-Recursive Neural Network to predict a general residue-residue contact map at 8 Angstrom threshold (Cheng et al., Nucleic Acids Research, 2005). Second, it uses the similar 2D-Recursive Neural Network to predict the special beta-sheet residue pairing map (Cheng and Baldi, Bioinformatics, 2005). Then the two maps are combined into a final map. The second step can improve contact prediction of proteins containing beta-sheets. In addition to good performance, NNCon is a fast predictor that usually can make predictions in just several minutes.

Input: target name, email address, and a plain protein sequence.

Output: in the body of a returned email, the predicted contact number and contact order are included. A list of predicted residue-residue contacts in the CASP format is also included. The probability threshold used to select contacts is set to 0.1. These predicted residue-residue contacts are formatted as line entries with columns meaning: index of residue i, index of residue j, starting range for prediction, ending range for prediction, probability residue i and residue j are within the specified distance threshold. Furthermore, a full contact probability matrix of all residue-residue pairs in 8 Angstrom and 12 Angstrom are attached. The format of the attached file is: name, predicted secondary structure, predicted relative solvent accessibility at 25% threshold, and a n * n contact probability matrix. n is the length of the protein. The contact map picture (.png file) is also attached with the email.

Example: the following is an example with the input sequence (CASP8 target T0387):
SMKPKLCRLAKGENGYGFHLNAIRGLPGSFIKEVQKGGPADLAGLEDEDVIIEVN GVNVL DEPYEKVVDRIQSSGKNVTLLVCGKKAQDTV

The output email that a user would receive is:

SMKPKLCRLAKGENGYGFHLNAIRGLPGSFIKEVQKGGPADLAGLEDEDVIIEVNGVN VLDEPYEKVVDRIQSSGKNVTLLVCGKKAQDTV

The predicted contact number is: 3.033.

The predicted contact order is: 0.914.

The attached files are:

--full predicted probabilities with 8 Angstrom (.con8a.comb file) and 12 Angstrom (.cm12a.comb file);

--the illustration of contact map (.png file).

Predicted residue-residue contacts at 8 Angstrom in CASP format are as follows. Predicted residue and residue contacts shown above are formatted as
line entries with columns meaning: index of residue i, index of residue j, starting range for prediction, ending range for prediction,
probability residue i and residue j are within the specified distance threshold. For more interpretations of outputs,
please go to http://casp.rnet.missouri.edu/nncon_help.html.

PFRMAT RR
*TARGET test_zheng_April_13_2009_short_*
AUTHOR NNcon
REMARK Sequence separation of predicted contacts >= 6.
METHOD neural network contact map prediction
MODEL 1
SMKPKLCRLAKGENGYGFHLNAIRGLPGSFIKEVQKGGPADLAGLEDEDV
IIEVNGVNVLDEPYEKVVDRIQSSGKNVTLLVCGKKAQDTV
* 4   22 0   8 0.132*
* 4   55 0   8 0.106*
* 4   80 0   8 0.137*
* 4   83 0   8 0.119*
.
.
.
* 83   89 0   8 0.179*
* 83   90 0   8 0.118*
* 83   91 0   8 0.191*
* 84   90 0   8 0.158*
* 84   91 0   8 0.125*
END

Interpretations of residue-residu contacts:
PFRMAT RR
TARGET temp

AUTHOR NNcon
REMARK Sequence separation of predicted contacts >= 6.
METHOD neural network contact map prediction
MODEL 1
HLEGSIGILLKKHEIVFDGC   # <- entire target sequence

HDFGRTYIWQMSD
1   9   0   8   0.70
1 10   0   8   0.70   # <- indices of residues: i and j (integers),
1 12   0   8   0.60   # <- the range of Cb-Cb distance predicted
1 14   0   8   0.20   #   for the residue pair: d1 and d2 (real),
1 15   0   8   0.10   # <- probability of the distance between
1 17   0   8   0.30   #   Cb atoms being within the specified
1 19   0   8   0.50   #   range: p (real)
2   8   0   8   0.90
3   7   0   8   0.70
END