補足

アミノ酸座位間における共進化に基づく残基間コンタクト予測：タンパク質立体構造予測にむけて

宮澤 三造
sanzo.miyazawa@gmail.com

03/11/2014

図 1. 共進化座位対と残基間コンタクト.

以下のタンパク質に関し, CES [10]とNMFI-DI [2]法で 真の残基間コンタクト数の1/3に等しい数の残基対を残基間コンタクトとして予測した結果を, α, β, $\alpha+\beta$とα/βフォールド の各タンパク質毎に, 各々左下図と右上図に, 真陽性(TP)を赤で偽陽性(FP)は藍色で示す. 図中, 灰色で示された残基対は残基間コンタクトで, 残基間での重原子の最小距離が 5 Å以下の 近接残基対でかつ 6残基以上隔たった残基対 ($|i-j|>5$)である. NMFI-DI法では, conservation filter [2]のみを用いた. 文献 [10]の図を改変.

Pfam ID / PDB ID	N^*	#contacts /#sites**	TP† +FP	PPV†† C_{ij}^{s}‡	PPV†† $\mathcal{C}_{ij}^{\mathrm{s}}$‡‡	PPV†† ρ_{ij}§	PPV†† DI§§	Remarks
α proteins								
Trans_reg_C / 1ODD-A:156-232	7720	111/76	37	0.189	$\ll$ 0.541	**0.622**	0.432	Transcriptional regulatory protein, C terminal
CH / 1BKR-A:5-107	2960	172/101	57	0.053	$\ll$ 0.439	**0.491**	0.439	Calponin homology domain
7tm_1 / 1GZM-A:54-306	6302	372/248	124	0.008	$\ll$ 0.290	**0.306**	0.169	Rhodopsin-like receptors
β proteins								
SH3_1 / 2HDA-A:97-144	4160	89/48	29	0.241	$\ll$ 0.621	**0.655**	0.552	SRC Homology 3 (SH3) domain
Cadherin / 2O72-A:113-212	7617	220/91	73	0.274	$\ll$ 0.726	**0.767**	0.753	Cadherin
Trypsin / 3TGI-E:16-238	6688	636/212	212	0.344	$\ll$ 0.575	**0.613**	0.533	Trypsin
$\alpha+\beta$ proteins								
Kunitz_BPTI / 5PTI-A:4-56	2130	111/53	37	0.216	$\ll$ 0.514	0.486	**0.541**	Kunitz domain
KH_1 / 1WVN-A:7-69	5114	90/57	30	0.367	$\ll$ 0.600	**0.700**	0.533	K Homology (KH) domain
RRM_1 / 1G2E-A:41-111	7684	133/70	44	0.295	$\ll$ 0.795	**0.795**	0.705	RNA recognition motif
FKBP_C / 1R9H-A:26-118	5695	200/92	66	0.197	$\ll$ 0.667	**0.727**	0.697	FK506 binding protein (FKBP)
Lectin_C / 1SL5-A:273-379	4479	246/103	82	0.171	$\ll$ 0.585	0.646	**0.671**	C-type lectin
α/β proteins								
Thioredoxin / 1RQM-A:1-103	7483	188/99	62	0.177	$\ll$ 0.581	**0.645**	0.565	Thioredoxin
Response_reg / 1E6K-A:8-121	7613	202/110	67	0.015	$\ll$ 0.657	**0.687**	0.642	Response regulator receiver domain
RNase_H / 1F21-A:3-142	4782	273/128	91	0.132	$\ll$ 0.407	0.407	**0.549**	RNase H
Ras / 5P21-A:5-165	6390	335/159	111	0.207	$\ll$ 0.640	**0.685**	0.631	Ras subfamily

文献 [10]の表を改変.

* 使用した相同配列の数. Pfam におけるFull alignmentから, Pfam系統樹において一定の閾値以下の短い枝長で分岐している配列は削除.

** 残基間コンタクト数/座位数. 残基間コンタクトの定義は残基間での重原子の最小距離が 5 Å以下でかつ 6残基以上隔たった残基対($|i-j|>5$)である.

† 予測した残基間コンタクトの数(真の残基間コンタクト数の1/3).

†† Positive predictive values (PPV = TP/(TP + FP))の値.

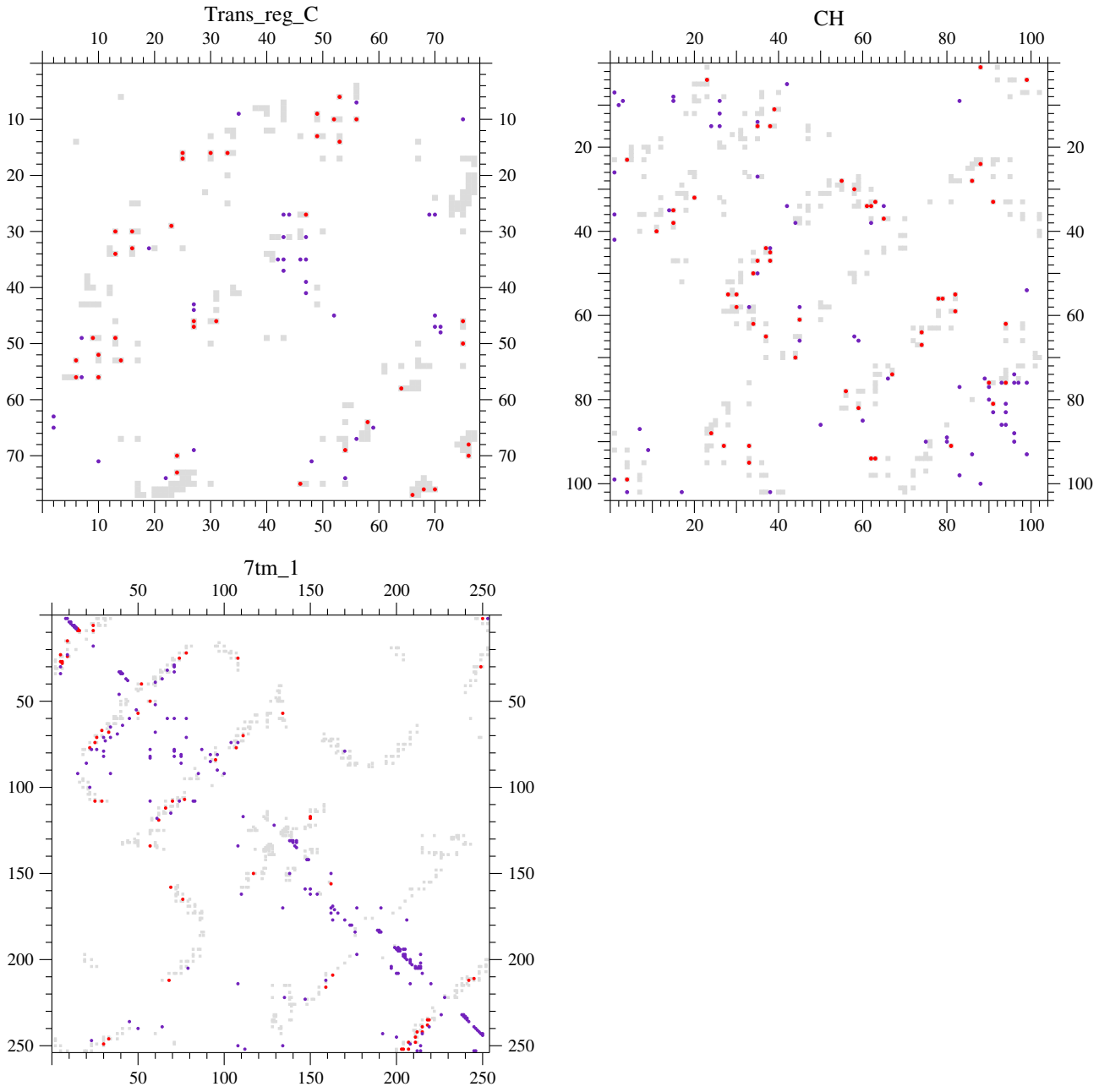
‡ 置換確率ベクトル間の相関係数による予測. 両末端座位は除外.

‡‡ 置換確率ベクトル間の偏相関係数による予測. 両末端座位は除外.

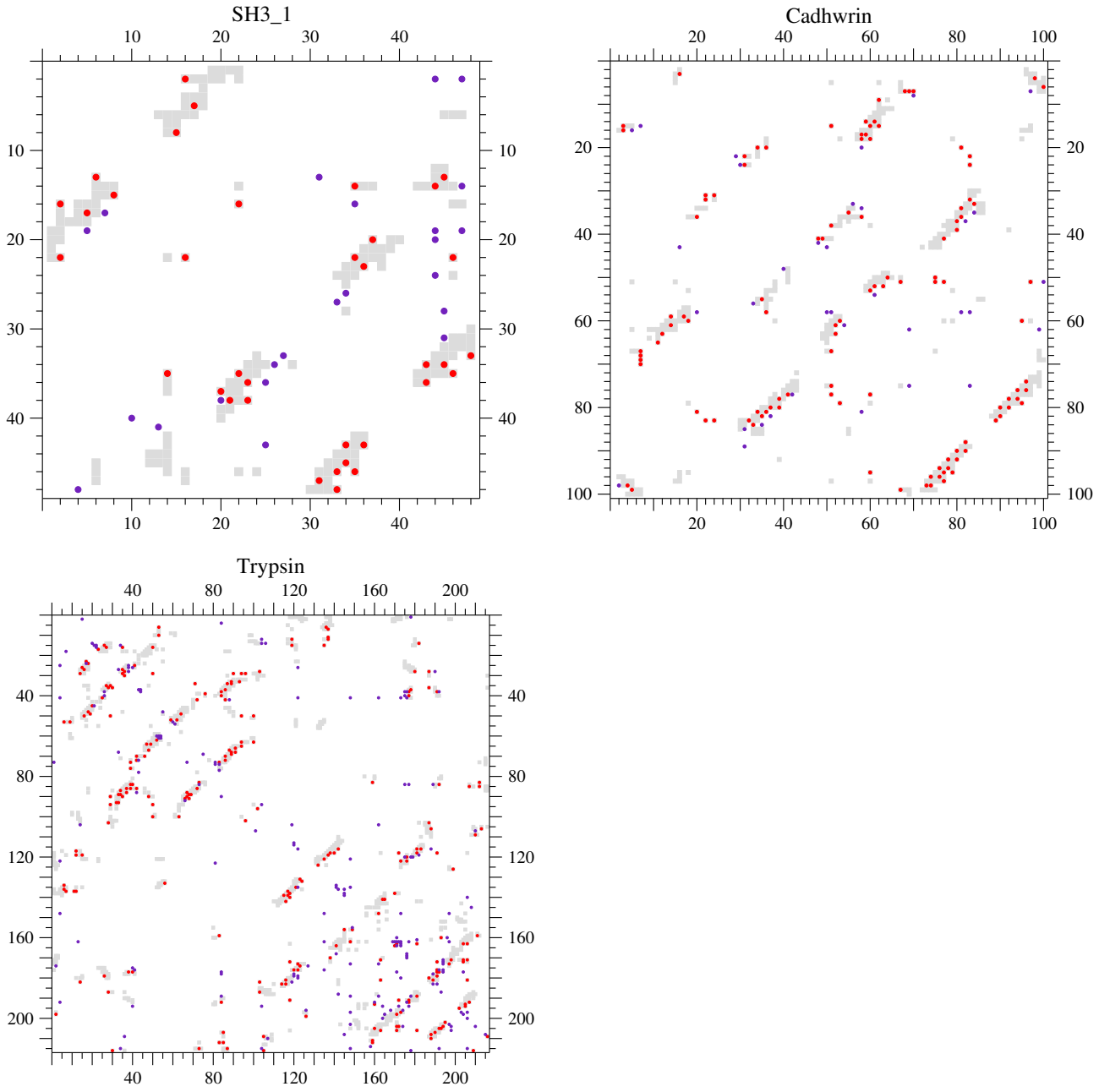
§ CES法による予測; 太字はDI法より高い精度を示す.

§§ NMFI-DI(DCA)法による予測; conservation filter [2]を用いた.

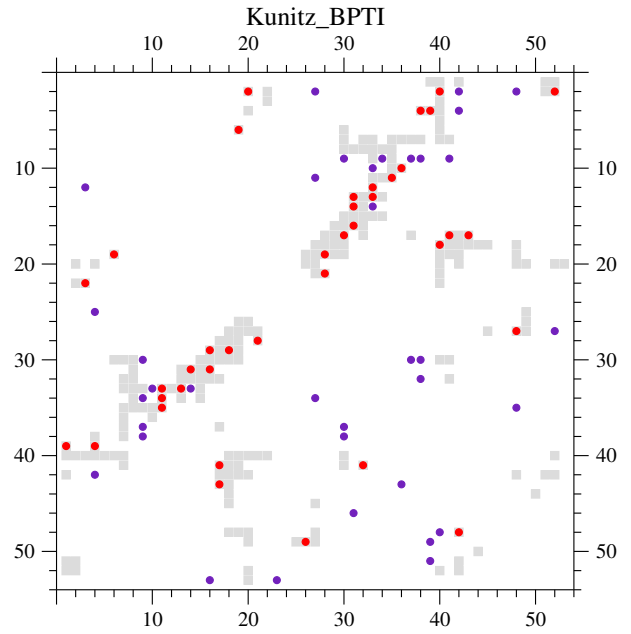
α proteins

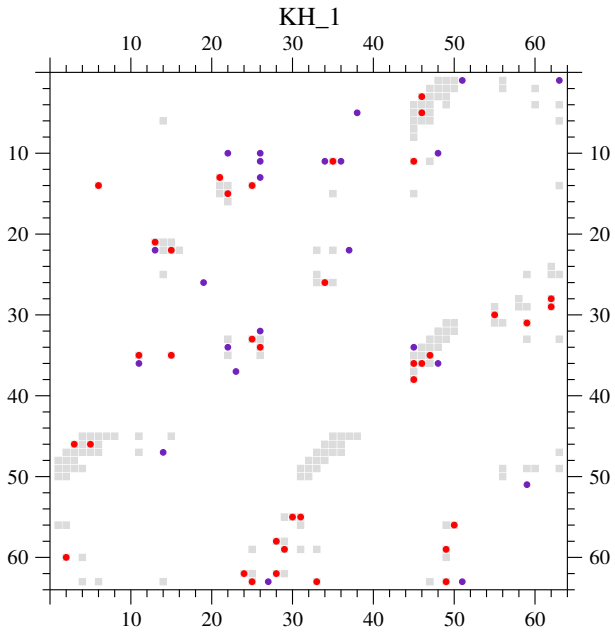


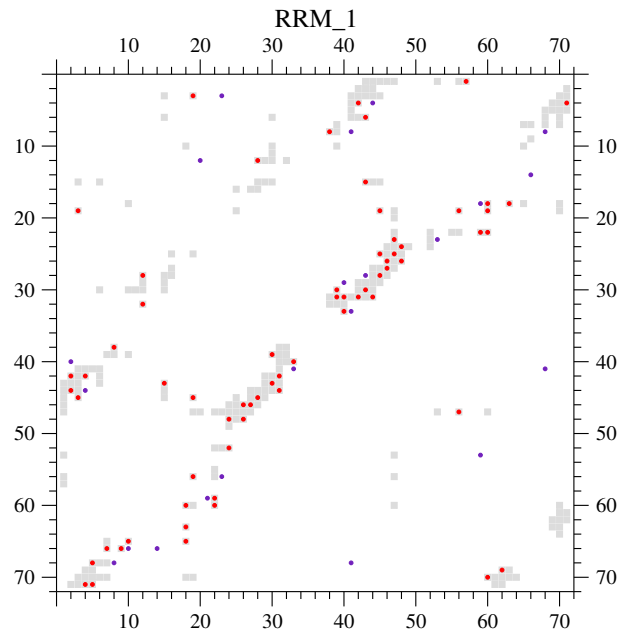
β proteins

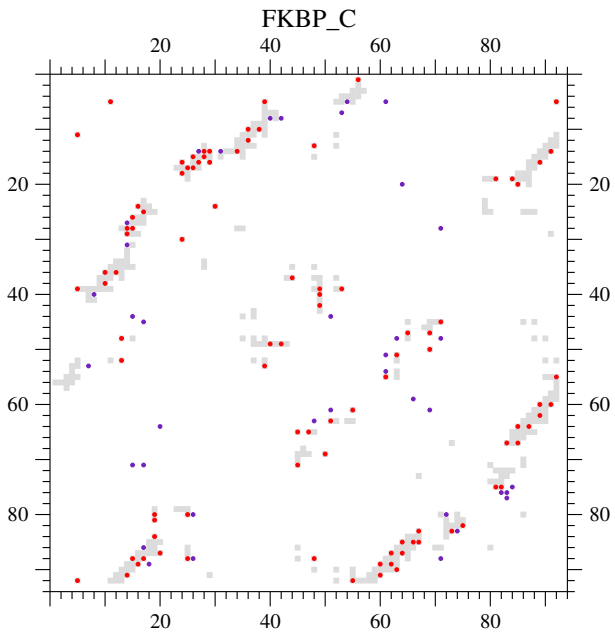


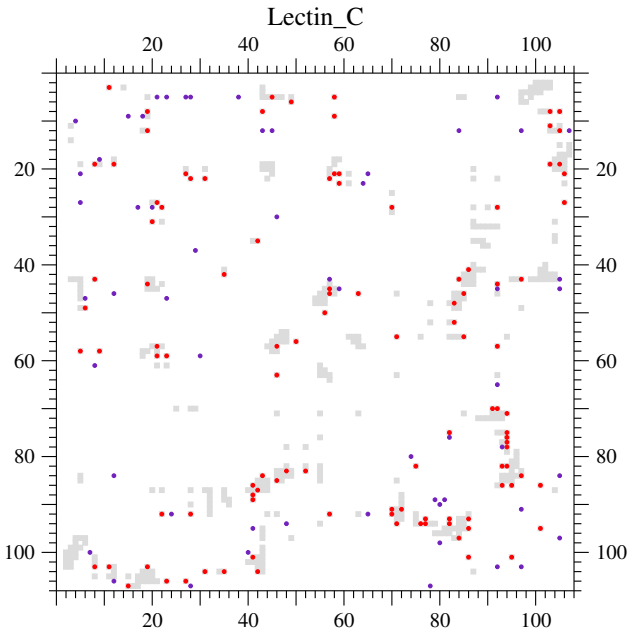
$\alpha + \beta$ proteins



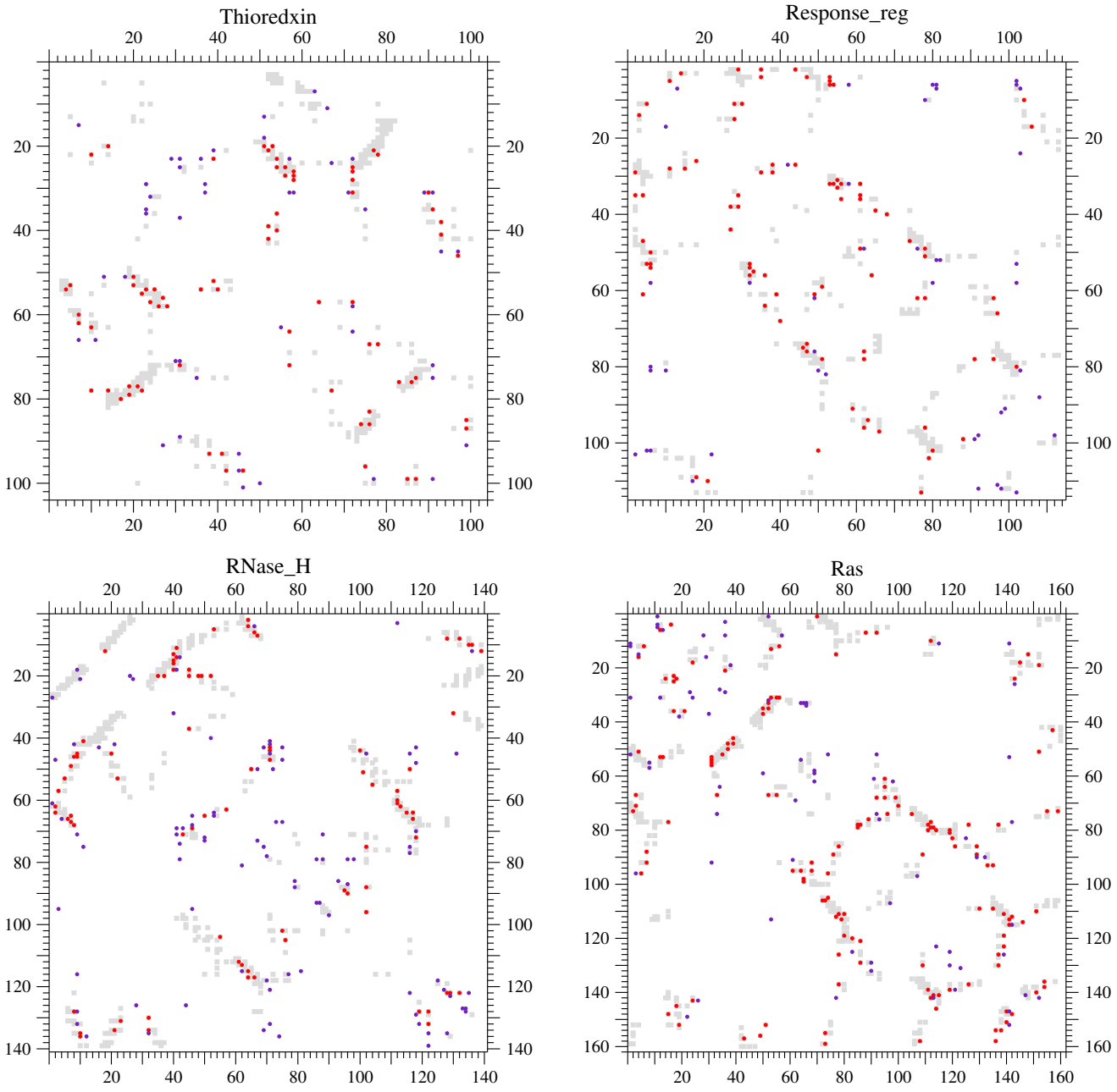








α/β proteins



文献

1) Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M.: Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc. Natl. Acad. Sci. USA **108**, 1293–1301 (2011). doi:10.1073/pnas.1111471108

2) Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., Sander, C.: Protein 3D structure computed from evolutionary sequence variation. PLoS ONE **6**(12), 28766 (2011). doi:10.1371/journal.pone.0028766

3) Taylor, W.R., Sadowski, M.I.: Structural constraints on the covariance matrix derived from multiple aligned protein sequences. PLoS ONE **6**(12), 28265 (2011). doi:10.1371/journal.pone.0028265

4) Hopf, T.A., Colwell, L.J., Sheridan, R., Rost, B., Sander, C., Marks, D.S.: Three-dimensional structures of membrane proteins from genomic sequencing. Cell **149**, 1607–1621 (2012). doi:10.1016/j.cell.2012.04.012

5) Marks, D.S., Hopf, T.A., Sander, C.: Protein structure prediction from sequence variation. Nature Biotech. **30**, 1072–1080 (2012). doi:10.1038/nbt.2419

6) Sułkowska, J.I., Morcos, F., Weigt, M., Hwa, T., Onuchic, J.N.: Genomics-aided structure prediction. Proc. Natl. Acad. Sci. USA **109**, 10340–10345 (2012). doi:10.1073/pnas.1207864109

7) Nugent, T., Jones, D.T.: Accurate de novo structure prediction of large transmembrane protein domains using fragmentassembly and correlated mutation analysis. Proc. Natl. Acad. Sci. USA **109**, 1540–1547 (2012). doi:10.1073/pnas.1120036109

8) Jones, D.T., Buchan, D.W.A., Cozzetto, D., Pontil, M.: Psicov: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics **28**, 184–190 (2012). doi:10.1093/bioinformatics/btr638

9) Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., Aurell, E.: Improved contact prediction in proteins: Using pseudolikelihoods to infer potts models. Phys. Rev. E **87**, 012707–116 (2013). doi:10.1103/PhysRevE.87.012707

10) Miyazawa, S.: Advantages of a mechanistic codon substitution model for evolutionary analysis of protein-coding sequences. PLoS ONE **6**(12), 28892 (2011). doi:10.1371/journal.pone.0028892

    Miyazawa, S.: Superiority of a mechanistic codon substitution model even for protein sequences in phylogenetic analysis. BMC Evol. Biol. **13**, 257 (2013). doi:10.1186/1471-2148-13-257

11) Schneidman, E., Berry II, M.J., Segev, R., Bialek, W.: Weak pairwise correlations imply strongly correlated network states in a neural population. Nature **440**, 1007–1012 (2006). doi:10.1038/nature04701

12) Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., Hwa, T.: Identification of direct residue contacts in protein-protein interaction by message passing. Proc. Natl. Acad. Sci. USA **106**, 67–72 (2009). doi:10.1073/pnas.0805923106

13) Kappen, H.J., Rodríguez, F.B.: Efficient learning in boltzmann machines using linear response theory. Neural Computation **10**, 1137–1156 (1998). doi:10.1162/089976698300017386