On an ideal pairwise contact potential of proteins

Sanzo Miyazawa

Graduate School of Engineering, Gunma University, Japan

presented at Telluride

in July 5 - July 9 of 2010.

1. INTRODUCTION

1. Go's native potential has been used as an ideal interaction potential of proteins to study the folding processes of proteins and the conformational changes of large protein complexes. In the elastic network model a harmonic potential is assumed only between atoms or residues in contact in native structures.

   Why can such a native potential reflect characteristic features of real proteins?

2. Pairwise contact potentials, which depend only on the type of amino acids, are often used to discriminate a native fold from non-native ones.

   Can such a pairwise contact potential essentially make a specific fold the unique and lowest energy conformation?

3. There are studies of reconstructing three dimensional structures from one dimensional information such as contact numbers and the principal eigenvector of a contact matrix.

   Why do the principal eigenvector of a contact matrix and a contact number vector contain significant information on the native conformations of proteins?

To discuss these questions, we consider what properties are induced for the contact matrix of the lowest-energy conformation by pairwise contact interactions.

2. THEORY

Basic assumptions

The total conformational energy: the sum of pairwise contact interactions

$$E^{\mathrm{c}}(C,S) = \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}\mathcal{E}_{ij}(S)\Delta_{ij}(C) \tag{1}$$

$$= \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}(\delta\mathcal{E}_{ij}(S)+\varepsilon_0)\Delta_{ij}(C) = \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}\delta\mathcal{E}_{ij}(S)\Delta_{ij}(C) + \varepsilon_0 N_c(C) \tag{2}$$

where i and j indicate ith and jth atom or residue, N means their total number, and

$\mathcal{E}_{ij}(S) \equiv \delta\mathcal{E}_{ij}(S)+\varepsilon_0$	Sequence (S-) dependent contact energy (E matrix) between i and j; $\exists\mathcal{E}_{ij}(S)<0$ is assumed.
$\delta\mathcal{E}_{ij}(S) \equiv \mathcal{E}_{ij}(S)-\varepsilon_0$	ε_0 subtracted from contact energy
$0 \le \Delta_{ij}(C) \le 1$	Conformation (C-) dependent factor representing the degree of contact between i and j
$n_i(C) = \sum_j^N \Delta_{ij}(C)$	Contact number of the ith atom or residue
$N_{\mathrm{c}}(C) \equiv \frac{1}{2}\sum_i n_i(C)$	The total number of contacts; volume exlusion is taken into account only by restricting the total number of contacts to N_c.

Lower bounds of the total contact energy

1. $$E^{\mathrm{c}}(C,S) \geq \frac{1}{2}\sum_i \sum_j \delta\mathcal{E}_{ij}(S)\Delta_{ij}(C_{\mathsf{min}}) + \varepsilon_0(C_{\mathsf{min}})N_{\mathrm{c}}(C_{\mathsf{min}}) \qquad \text{for given } N_c \tag{3}$$

   $$\text{The equality is achieved iff } \Delta_{ij}(C_{\mathsf{min}}) = \begin{cases} 1 & \text{if } \delta\mathcal{E}_{ij}(S) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

   ε_0 corresponds to a threshold of contact energy for a residue/atom pair to be in contact. (5)

2. $$E^{\mathrm{c}}(C,S) \geq \min_{\varepsilon_0} [\, -\frac{1}{2}\|\delta\vec{\mathcal{E}}(S)\|\|\vec{\Delta}(C)\| + \varepsilon_0 N_{\mathrm{c}}(C)\,] \qquad \text{for given } N_c \tag{6}$$

   $$\text{The equality is achieved iff } \delta\mathcal{E}_{ij}(S) = \varepsilon\Delta_{ij}(C) \text{ with } \varepsilon < 0 \text{ and } \Delta_{ij}(C) = 1 \text{ or } 0. \tag{7}$$

   This lower bound is a special case of Eq. 3.

An ideal pairwise potential of proteins

- Define $\Delta_{ij}(C)$ as $\Delta_{ij}(C_{\text{native}}) = 1$ or 0.

$$\mathcal{E}_{ij}(S) = \delta\mathcal{E}_{ij}(S) + \varepsilon_0 \tag{8}$$

$$\delta\mathcal{E}_{ij} + \varepsilon_0 < 0 \text{ and } \delta\mathcal{E}_{ij} < 0 \quad \text{for } (i,j) \text{ such that } \Delta_{ij}(C_{\text{native}}) = 1 \tag{9}$$

$$\delta\mathcal{E}_{ij} \geq 0 \quad \text{for } (i,j) \text{ such that } \Delta_{ij}(C_{\text{native}}) = 0 \tag{10}$$

	$\delta\mathcal{E}_{ij} + \varepsilon_0$ for non-native contacts		
	> 0	$= 0$	< 0
$\exists C \vert N_c(C) > N_c(C_{\text{native}})$	the lowest and unique	the lowest but may not be unique	may not be the lowest
$\forall C, N_c(C) \leq N_c(C_{\text{native}})$	the lowest and unique		

Thus, if $\delta\mathcal{E}_{ij} + \varepsilon_0 < 0$ for non-native pairs. *short-range interactions will be required* to make the native conformation the unique and lowest-energy conformation.

- Go's native potential: a special case of the above; $\delta\mathcal{E}_{ij} = \begin{cases} \varepsilon & \text{for native contacts} \\ 0 & \text{for non-native contacts} \end{cases}$

$$\mathcal{E}_{ij}(S) = \varepsilon\Delta_{ij}(C_{\text{native}}) + \varepsilon_0 \tag{11}$$

$$\varepsilon < 0 \text{ and } \varepsilon + \varepsilon_0 < 0 \tag{12}$$

Singular-spectral relationship between C and E matrices

Singular value decompositions: $\lambda_\mu(C)$ and $\varepsilon_\nu(S)$ are eigenvalues.

$$\Delta_{ij}(C) = \sum_\mu |\lambda_\mu(C)| L_{i\mu}(C) R_{j\mu}(C) \qquad \mathcal{E}_{ij}(S) = \sum_\nu |\varepsilon_\nu(S)| U_{i\nu}(S) V_{j\nu}(S) + \varepsilon_0 \tag{13}$$

$$|\lambda_1(C)| \geq \ldots \geq |\lambda_N(C)| \geq 0 \qquad |\varepsilon_1(S)| \geq \ldots \geq |\varepsilon_N(S)| \geq 0 \tag{14}$$

The total contact energy:

$$E^{\mathrm{c}}(C,S) \;=\; \frac{1}{2}\sum_\mu \sum_\nu |\lambda_\mu(C)||\varepsilon_\nu(S)|\omega_{\mu\nu}(C,S) + \varepsilon_0 N_{\mathrm{c}}(C) \tag{15}$$

where

$$1 \geq \;\; \omega_{\mu\nu}(C,S) \;=\; {}^t\boldsymbol{L}_\mu(C)\boldsymbol{U}_\nu(S)\,{}^t\boldsymbol{R}_\mu(C)\boldsymbol{V}_\nu(S) \;\; \geq -1 \tag{16}$$

Lower bounds :

$$E^{\mathrm{c}}(C,S) \;\geq\; \min_{\varepsilon_0}\Big[-\frac{1}{2}\sum_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} |\lambda_\xi(C)\varepsilon_\xi(S)| + \varepsilon_0 N_{\mathrm{c}}(C)\Big] \qquad \text{for given } N_c \tag{17}$$

$$\text{The equality is achieved iff } \omega_{\mu\nu} = -\delta_{\mu\nu} \quad \text{for } \{\mu|\lambda_\mu\varepsilon_\mu \neq 0\} \tag{18}$$

$$\geq\; \min_{\varepsilon_0}\Big[-\frac{1}{2}\|\vec{\lambda}(C)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}}\|\vec{\varepsilon}(S)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} + \varepsilon_0 N_{\mathrm{c}}(C)\Big] \qquad \text{for given } N_c \tag{19}$$

$$=\; \min_{\varepsilon_0}\Big[-\frac{1}{2}\|\delta\vec{\mathcal{E}}(S)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}}\|\vec{\Delta}(C)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} + \varepsilon_0 N_{\mathrm{c}}(C)\Big] \qquad \text{for given } N_c \tag{20}$$

$$\text{The equality is achieved iff } \varepsilon_\xi(S) = \varepsilon\lambda_\xi(C) \text{ with } \varepsilon < 0 \text{ for } \{\xi|\lambda_\xi\varepsilon_\xi \neq 0\} \tag{21}$$

Is a pairwise residue-residue potential sufficient to make native structures the unique and lowest-energy conformations?

If the contact potential depends only on the type of residues in contact, then the eigenvalues that are equal to zero must exist for longer proteins than 20 amino acids.

$$\exists \varepsilon_\xi = 0 \ \text{ because } \ \text{rank}(\mathcal{E}_{ij}) \leq 20. \tag{22}$$

Therefore, it is highly possible that the multiple lowest-energy conformations may exist, because

$$E^c(C,S) = \frac{1}{2}\sum_\nu |\varepsilon_\nu| ({}^tU\Delta(C)V)_{\nu\nu} \tag{23}$$

and

$$\Delta(C_{\text{lowest}}) \text{ and } \Delta(C) \text{ may exist, such that } \varepsilon_\xi ({}^tU(\Delta(C_{\text{lowest}}) - \Delta(C))V)_{\xi\xi} = 0 \text{ for } \forall\xi \tag{24}$$

Relationships to be satisfied in the lower bound of conformational energy

1. All the left ($\boldsymbol{L}_\mu$ and $\boldsymbol{U}_\mu$) and the corresponding right ($\boldsymbol{R}_\mu$ and $\boldsymbol{V}_\mu$) singular vectors of C- and E- matrices are parallel or anti-parallel to each other, that is, Eq. 12; $\omega_{\mu\nu} = -\delta_{\mu\nu}$.

   Bastolla etl al. (2005) pointed out such a relation in which the term of ε_0 was ignored, that is, $\varepsilon_0 = 0$.

2. The principal eigenvector $\boldsymbol{R}_1$ of the C matrix and the contact number vector $\boldsymbol{n}$ tend to be parallel to make the principal eigenvalue larger, that is, to make the total energy lower.

$$\lambda_\mu(C) \;=\; \frac{{}^t\boldsymbol{R}_\mu(C)\boldsymbol{n}(C)}{{}^t\boldsymbol{R}_\mu(C)\mathbf{1}} = \langle n_\bullet^2\rangle^{1/2}\ {}^t\boldsymbol{R}_\mu\boldsymbol{n}\|\mathbf{1}\|/({}^t\boldsymbol{R}_\mu\mathbf{1}\|\boldsymbol{n}\|) \tag{25}$$

3. The contact number vector $\boldsymbol{n}$ and the vector of mean relative contact energy $\delta\vec{\mathcal{E}}_\bullet$ tends to be anti-parallel.

$$E^{\mathrm{c}}(C,S) \;\approx\; \frac{1}{2}\sum_i\sum_j[\frac{1}{N}\sum_k \delta\mathcal{E}_{ik}(S)]\Delta_{ij}(C) + \varepsilon_0 N_{\mathrm{c}}(C) \tag{26}$$

$$\geq\; -\frac{1}{2}\|\delta\vec{\mathcal{E}}_\bullet(S)\|\|\boldsymbol{n}(C)\| + \varepsilon_0 N_{\mathrm{c}}(C) \tag{27}$$

$$\delta\vec{\mathcal{E}}_\bullet(S) \;\equiv\; {}^t(\ldots,\ \frac{1}{N}\sum_k \delta\mathcal{E}_{ik}(S),\ \ldots) \tag{28}$$

Native conformations approach the C-matrix as closely as possible in which the relationships above are satisfied.

3. DATA ANALYSES

The relationships between E and C matrices indicated for lower energy conformations are examined by crudely evaluating pairwise interactions in native structures with a statistical contact potential.

Pairwise contact potential used:

A statistical estimate of contact energies with a correction for the Bethe approximation (Miyazawa & Jernigan, Proteins 34, 49, 1999); the contact energy between amino acids a and b is evaluated as

$$e_{ab} = e_{rr} + \alpha'[\Delta e_{ar}^{\text{Bethe}} + \Delta e_{rb}^{\text{Bethe}} + \frac{\beta'}{\alpha'}\delta e_{ab}^{\text{Bethe}}] \tag{29}$$

where $\frac{\beta'}{\alpha'} = 2.2$, and the subscript r represents the mean effects from an amino acid.

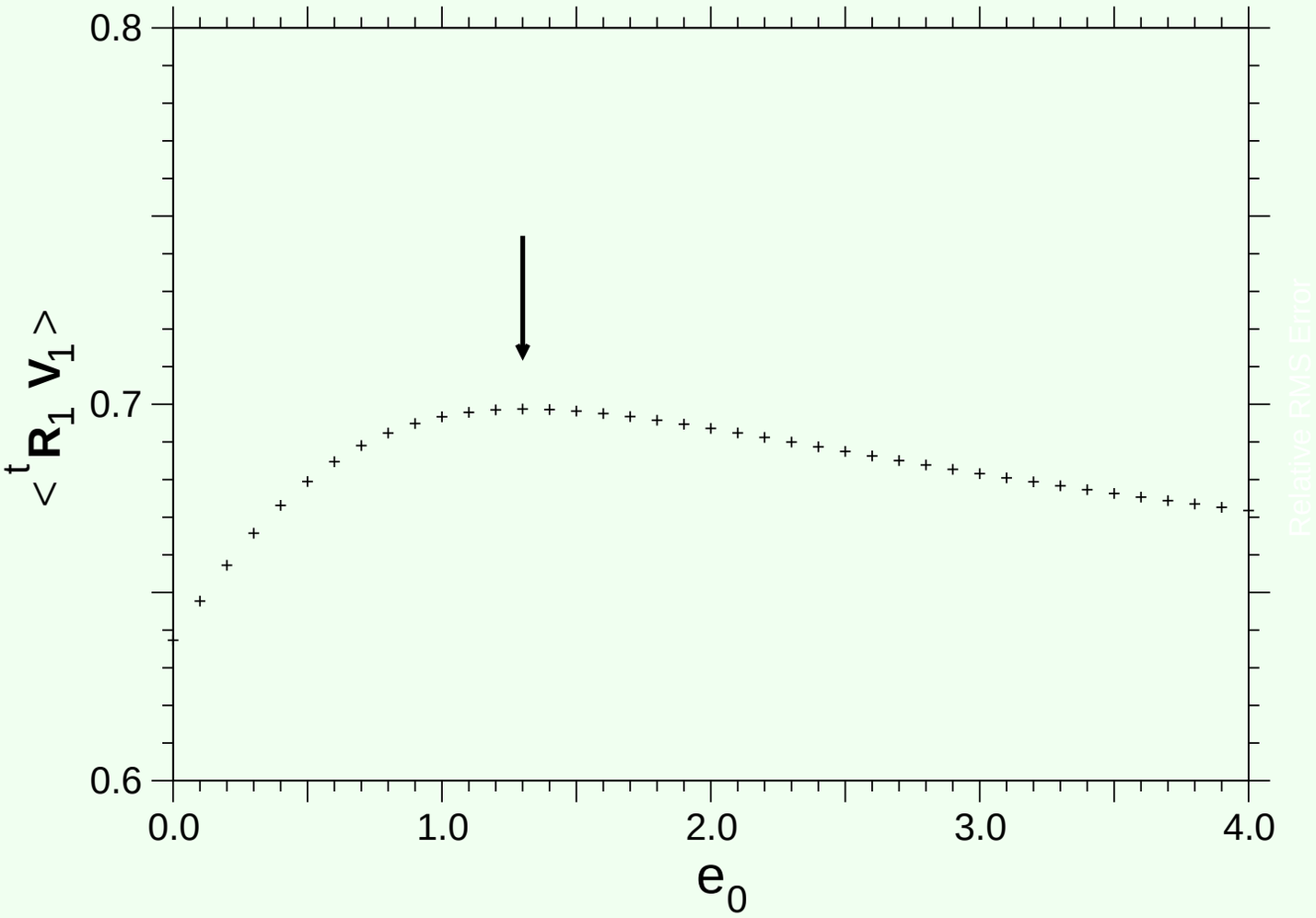
Protein structures analyzed:

182 proteins of representatives from each family of classes 1 – 4 in SCOP 1.69.

The optimum value of ε_0 in Eqs. 6, 3, 17, and 19,

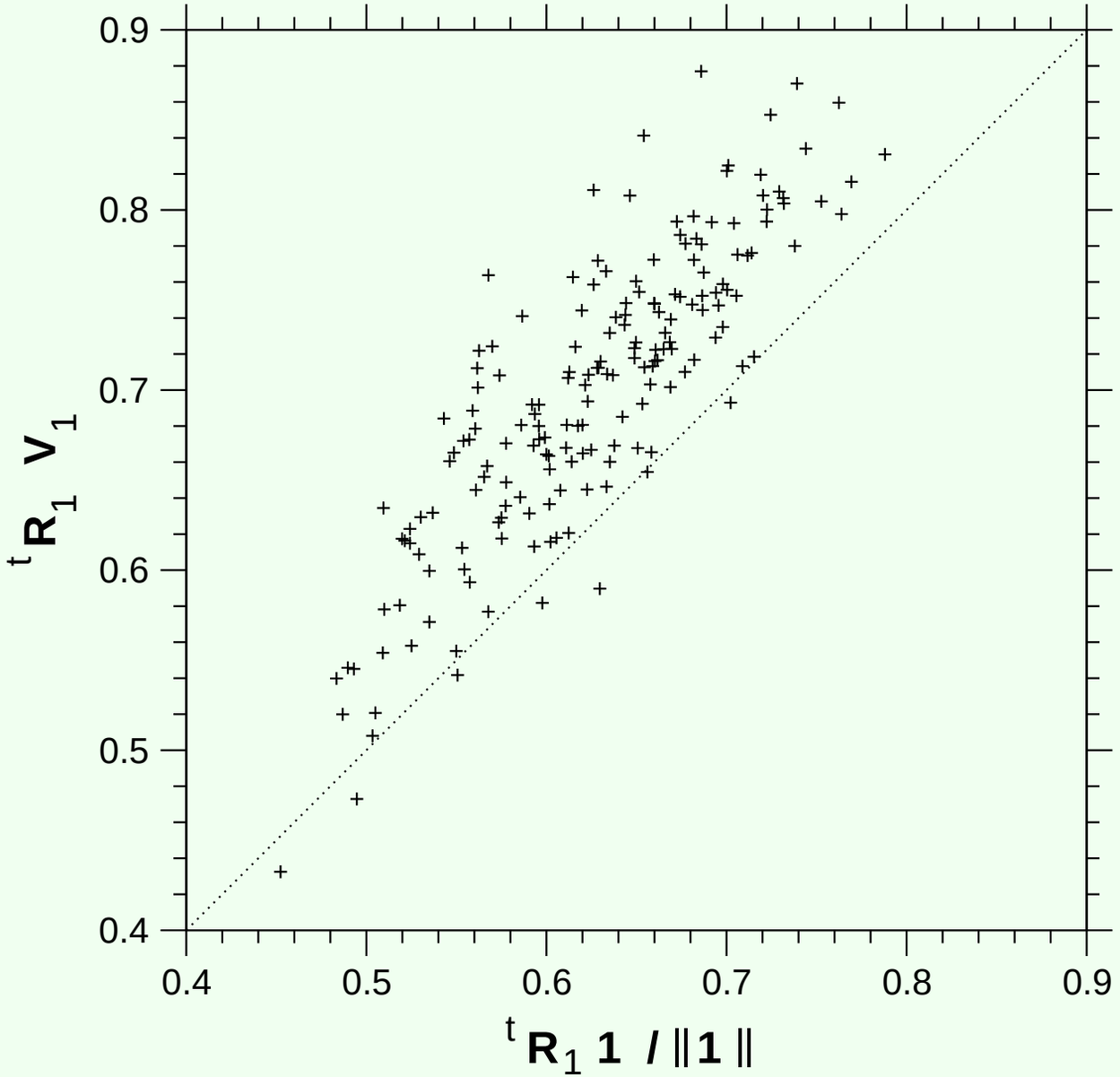
where the average of ${}^tR_1V_1$ over 182 proteins has a maximum;

ε_0 corresponds to a threshold of contact energy that separates native contacts from non-native ones.



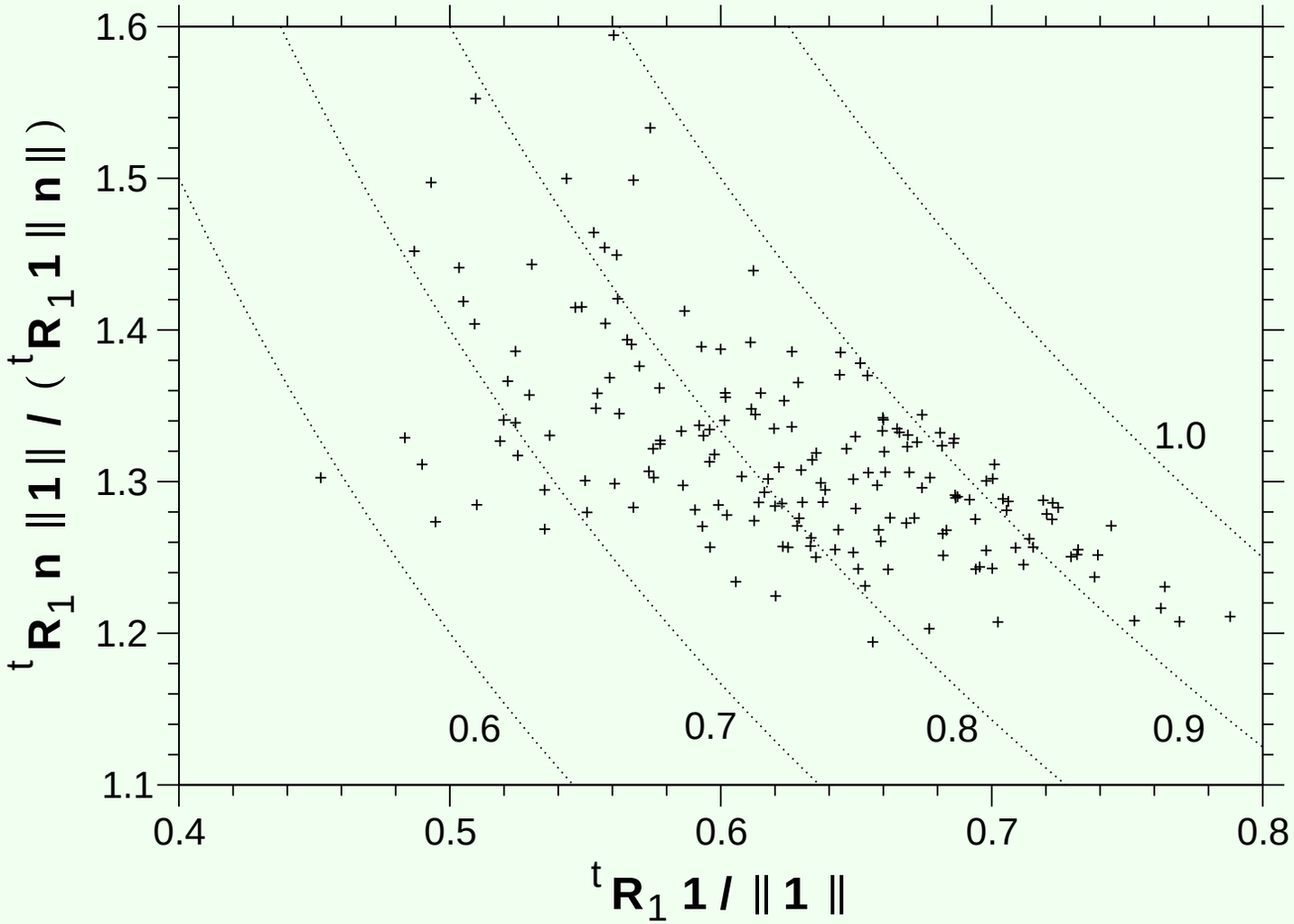
This optimum value of e_0, ($\varepsilon_0 \propto e_0 +$ constant), is used for the singular decomposition of E matrices in the following analyses.

The principal eigenvectors, $\boldsymbol{R}_1$ and $\boldsymbol{V}_1$, of C- and E- matrices tend to be parallel to each other.



In the t-tests of the correlation coefficients between $\boldsymbol{R}_1$ and $\boldsymbol{V}_1$, the geometric mean of probabilities for a significance over 182 proteins is equal to $\exp(-18.4)$.

The principal eigenvector, R_1, of C-matrix tends to be parallel to the contact number vector, n.



The dotted lines indicate the iso-value lines for ${}^t\boldsymbol{R}_1\boldsymbol{n}/\parallel \boldsymbol{n} \parallel$, whose values are shown in the figure.

4. CONCLUSIONS

1. The total energy of a native structure with the Go's native potential corresponds to a lower bound of the total energy of the protein interacting with a pairwise contact potential.
2. Native conformations must approach the one as closely as possible for which real interactions become an ideal potential. Thus, the following relationships between E- and C- matrices for protein native structures are expected, and have been observed in 182 representative proteins;
   (a) a parallel relationship between the principal eigenvectors of the C- and the E-matrices, and
   (b) a parallel relationship between a contact number vector and the principal eigenvectors of the C- and the E-matrices,

   provided that the *E-matrix is singular-decomposed with an additional constant term*; it corresponds to the threshold of contact energy that separates native contacts from non-native ones.
3. Any pairwise residue-residue contact potential, which depends only on the type of interacting amino acids, is not sufficient and other interactions including residue connectivities and steric hindrance are needed to make native structures the unique and lowest-energy conformations.

This work was done in collaboration with Dr. Akira R. Kinjo in Protein Research Institute of Osaka Univ.

Reference: Phys.Rev.E, **77**:051910, 2008.