Properties of contact matrices

induced by pairwise interactions in proteins

Sanzo Miyazawa[1] and Akira R. Kinjo[2]

[1] miyazawa@smlab.sci.gunma-u.ac.jp

Graduate School of Engineering, Gunma University, Japan

[2]akinjo@protein.osaka-u.ac.jp

Institute for Protein Reseach, Osaka University, Suita, Japan

presented at

2009 annual meeting of Biophysical Society in Boston

(February 28 - March 4, 2009)

ABSTRACT

The properties of contact matrices (C matrices) needed for native proteins to be the lowest-energy conformations are considered in relation to a contact energy matrix (E matrix). *The total conformational energy is assumed to consist of pairwise interaction energies* between atoms or residues, each of which is expressed as a product of a conformation-dependent function (an element of a C *matrix*) and a sequence-dependent energy parameter (an element of a E *matrix*). *Such pairwise interactions in proteins force native C matrices to be in a relationship as if the interactions are a Go-like potential* [N. Go, Annu. Rev. Biophys. Bioeng. **12**. 183 (1983)] for the native C matrix, because the lowest bound of the total energy function is equal to the total energy of the native conformation interacting in a Go-like pairwise potential. This relationship between C and E matrices corresponds to (a) *a parallel relationship between the eigenvectors of the C and E matrices* and a linear relationship between their eigenvalues, and (b) a parallel relationship between a contact number vector and the principal eigenvectors of the C and E matrices; *the E matrix is eigen-decomposed with an additional constant term,* which corresponds to a threshold of contact energy that approximately separates native contacts from non-native ones. These relationships are confirmed in 182 representatives from each family of the SCOP database by examining inner products between the principal eigenvector of the C matrix, that of the E matrix evaluated with a statistical contact potential, and a contact number vector. In addition, the spectral representation of C and E matrices reveals that *pairwise residue-residue interactions*, which depends only on the types of interacting amino acids but not on other residues in a protein, *are insufficient and other interactions including residue connectivities and steric hindrance are needed to make native structures unique lowest-energy conformations.*

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

1. INTRODUCTION

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

Why the principal eigenvector of a contact matrix and a contact number vector contain significant information on protein structures ?

To answer this question, we consider what properties of a contact matrix are induced by pairwise contact interactions for native structures to be the lowest-energy conformations.

2. THEORY

Basic assumptions

The total conformational energy: the sum of pairwise interactions

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

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

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

$\mathcal{E}_{ij}(S)$		Sequence (S-) dependent contact energy (E matrix) between i and j
$\delta\mathcal{E}_{ij}(S)$	$\equiv \mathcal{E}_{ij}(S) - \varepsilon_0$	Contact energy relative to ε_0
$0 \le \Delta_{ij}(C) \le 1$		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

Lower bounds of the total contact energy are achieved by Go-like potentials.

1. $$E^{\mathrm{c}}(C,S) \geq -\frac{1}{2}\|\delta\vec{\mathcal{E}}(S)\|\|\vec{\Delta}(C)\| + \varepsilon_0 N_{\mathrm{c}}(C) \tag{3}$$

   The equality is achieved iff $\delta\mathcal{E}_{ij}(S) = \varepsilon\Delta_{ij}(C)$ with $\varepsilon < 0$ and $\Delta_{ij}(C) = 1$ or 0. (4)

2. $$E^{\mathrm{c}}(C,S) \geq \frac{1}{2}\sum_i\sum_j \delta\mathcal{E}_{ij}(S)\Delta_{ij}(C_{\mathrm{min}}) + \varepsilon_0 N_{\mathrm{c}}(C_{\mathrm{min}}) \tag{5}$$

   The equality is achieved iff $\Delta_{ij}(C_{\mathrm{min}}) = \begin{cases} 1 & \text{if } \delta\mathcal{E}_{ij}(S)) < 0 \\ 0 & \text{otherwise.} \end{cases}$ (6)

   ε_0 corresponds to a threshold of contact energy for a residue pair to be in contact in native structures.

Spectral relationship between C and E matrices

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

$$\Delta_{ij}(C) = \sum_\mu |\lambda_\mu(C)| L_{i\mu}(C) R_{j\mu}(C) \qquad \mathcal{E}_{ij}(S) = \sum_\nu |\varepsilon_\nu(S)| U_{i\nu}(S) V_{j\nu}(S) + \varepsilon_0 \tag{7}$$

$$|\lambda_1(C)| \geq \ldots \geq |\lambda_N(C)| \geq 0 \qquad |\varepsilon_1(S)| \geq \ldots \geq |\varepsilon_N(S)| \geq 0 \tag{8}$$

The total contact energy:

$$E^{\mathrm{c}}(C,S) \;=\; \frac{1}{2}\sum_\mu \sum_\nu |\lambda_\mu(C)||\varepsilon_\nu(S)|\omega_{\mu\nu}(C,S) + \varepsilon_0 N_{\mathrm{c}}(C) \tag{9}$$

where

$$\omega_{\mu\nu}(C,S) \;=\; {}^t\boldsymbol{L}_\mu(C)\boldsymbol{U}_\nu(S)\,{}^t\boldsymbol{R}_\mu(C)\boldsymbol{V}_\nu(S) \tag{10}$$

The lower bounds are achieved by Go-like potentials:

$$E^{\mathrm{c}}(C,S) \;\geq\; -\frac{1}{2}\sum_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} |\lambda_\xi(C)\varepsilon_\xi(S)| + \varepsilon_0 N_{\mathrm{c}}(C) \tag{11}$$

$$\text{The equality is achieved iff } \omega_{\mu\nu} = -\delta_{\mu\nu} \quad \text{for } \{\mu|\lambda_\mu\varepsilon_\mu \neq 0\} \tag{12}$$

$$\geq\; -\frac{1}{2}\|\vec{\lambda}(C)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}}\|\vec{\varepsilon}(S)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} + \varepsilon_0 N_{\mathrm{c}}(C) \tag{13}$$

$$=\; -\frac{1}{2}\|\delta\vec{\mathcal{E}}(S)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}}\|\vec{\Delta}(C)\|_{\{\xi|\lambda_\xi\varepsilon_\xi\neq 0\}} + \varepsilon_0 N_{\mathrm{c}}(C) \tag{14}$$

$$\text{The equality is achieved iff } \varepsilon_\xi(S) = \varepsilon\lambda_\xi(C) \text{ with } \varepsilon < 0 \text{ for } \{\xi|\lambda_\xi\varepsilon_\xi \neq 0\} \tag{15}$$

Relationships to be satisfied in the lower bounds of conformational energy

1. All the corresponding left ($\boldsymbol{L}_\mu$ and $\boldsymbol{U}_\mu$) and right ($\boldsymbol{R}_\mu$ and $\boldsymbol{V}_\mu$) singular vectors of the C- and E- matrices are parallel or anti-parallel to each other, that is, Eq. 12; $\omega_{\mu\nu} = -\delta_{\mu\nu}$.
2. The principal eigenvector $\boldsymbol{R}_1$ of the C matrix and contact number vector $\boldsymbol{n}$ tends to be in parallel to make the principal eigenvalue larger.

$$\lambda_\mu(C) \;=\; \frac{{}^t\boldsymbol{R}_\mu(C)\boldsymbol{n}(C)}{{}^t\boldsymbol{R}_\mu(C)\mathbf{1}} = \langle n_\bullet^2\rangle^{1/2}\,{}^t\boldsymbol{R}_\mu\boldsymbol{n}\|\mathbf{1}\|/({}^t\boldsymbol{R}_\mu\mathbf{1}\|\boldsymbol{n}\|) \tag{16}$$

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.

$$E^{\mathrm{c}}(C,S) \;\approx\; \frac{1}{2}\sum_i\sum_j[\frac{1}{N}\sum_k \delta\mathcal{E}_{ik}(S)]\Delta_{ij}(C) + \varepsilon_0 N_{\mathrm{c}}(C) \tag{17}$$

$$\geq\; -\frac{1}{2}\|\delta\vec{\mathcal{E}}_\bullet(S)\|\|\boldsymbol{n}(C)\| + \varepsilon_0 N_{\mathrm{c}}(C) \tag{18}$$

$$\delta\vec{\mathcal{E}}_\bullet(S) \;\equiv\; {}^t(\ldots,\; \frac{1}{N}\sum_k \delta\mathcal{E}_{ik}(S),\; \ldots) \tag{19}$$

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

It is highly possible that multiple lowest energy conformations may exist, because

1. $$E^c(C,S) = \frac{1}{2}\sum_{\nu} |\varepsilon_\nu|({}^tU\Delta(C)V)_{\nu\nu} + \varepsilon_0 N_c(C) \tag{20}$$
2. $$\exists \varepsilon_\xi = 0 \ \text{ because } \ \mathrm{rank}(\delta\mathcal{E}_{ij}) \leq 20. \tag{21}$$

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

$$e_{ab} = e_{rr} + \alpha'[\Delta e_{ar}^{\text{Bethe}} + \Delta e_{rb}^{\text{Bethe}} + \frac{\beta'}{\alpha'}\delta e_{ab}^{\text{Bethe}}] \quad (22)$$

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

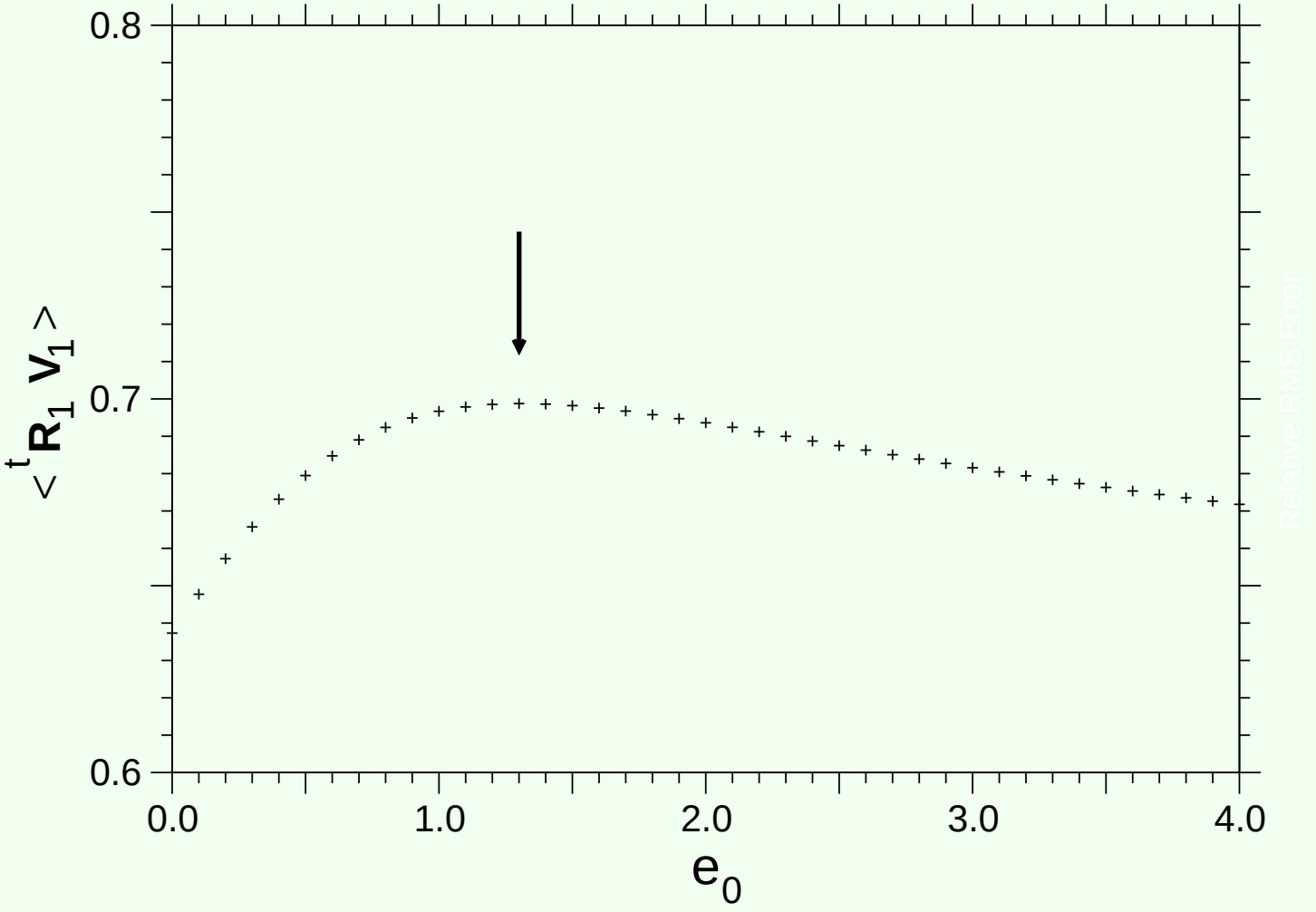
Protein structures analyzed:

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

An optimum value of ε_0 in Eqs. 3, 5, 11, and 13,

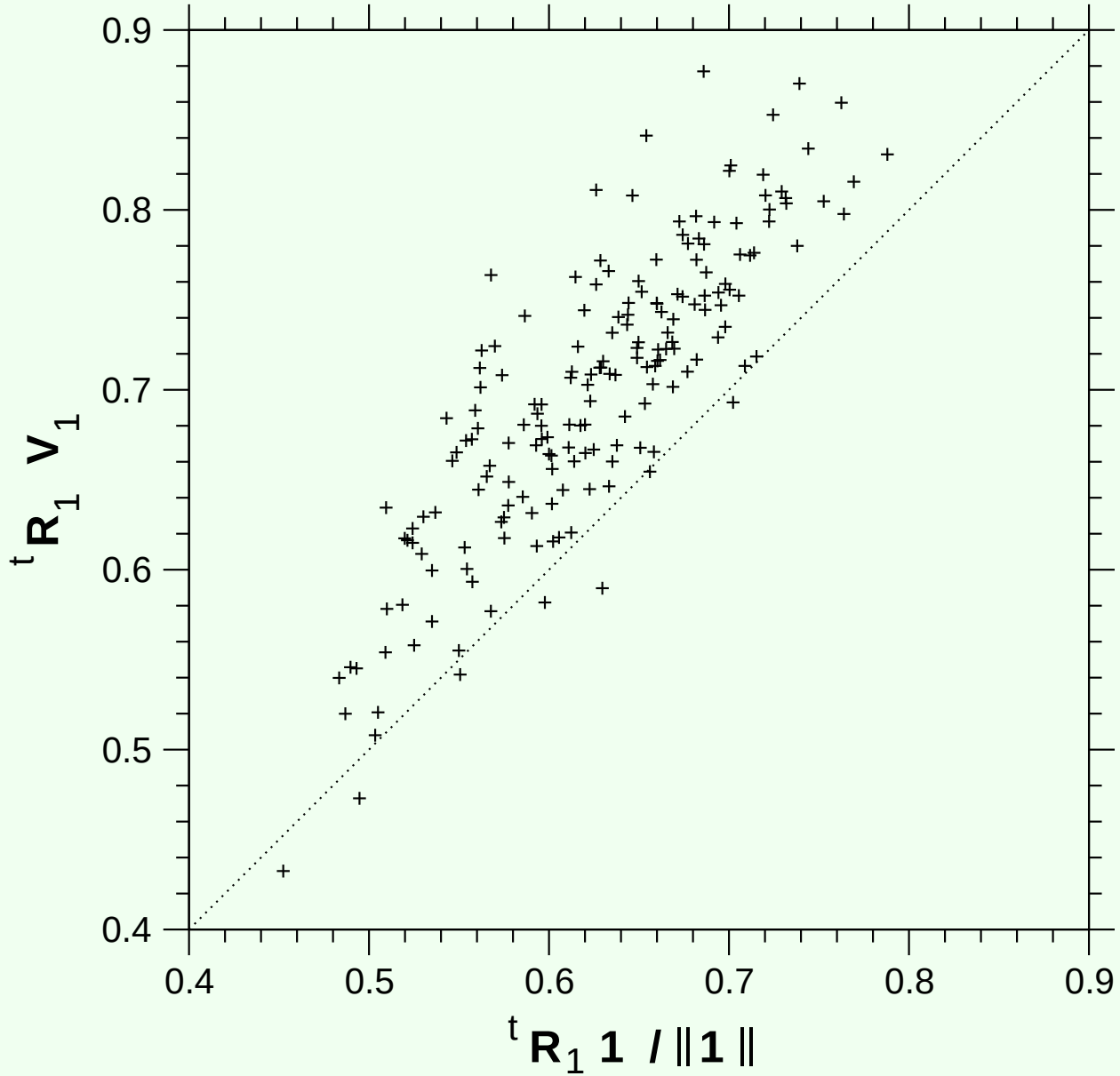
where the average of ${}^tR_1V_1$ over 182 proteins has a maximum;

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



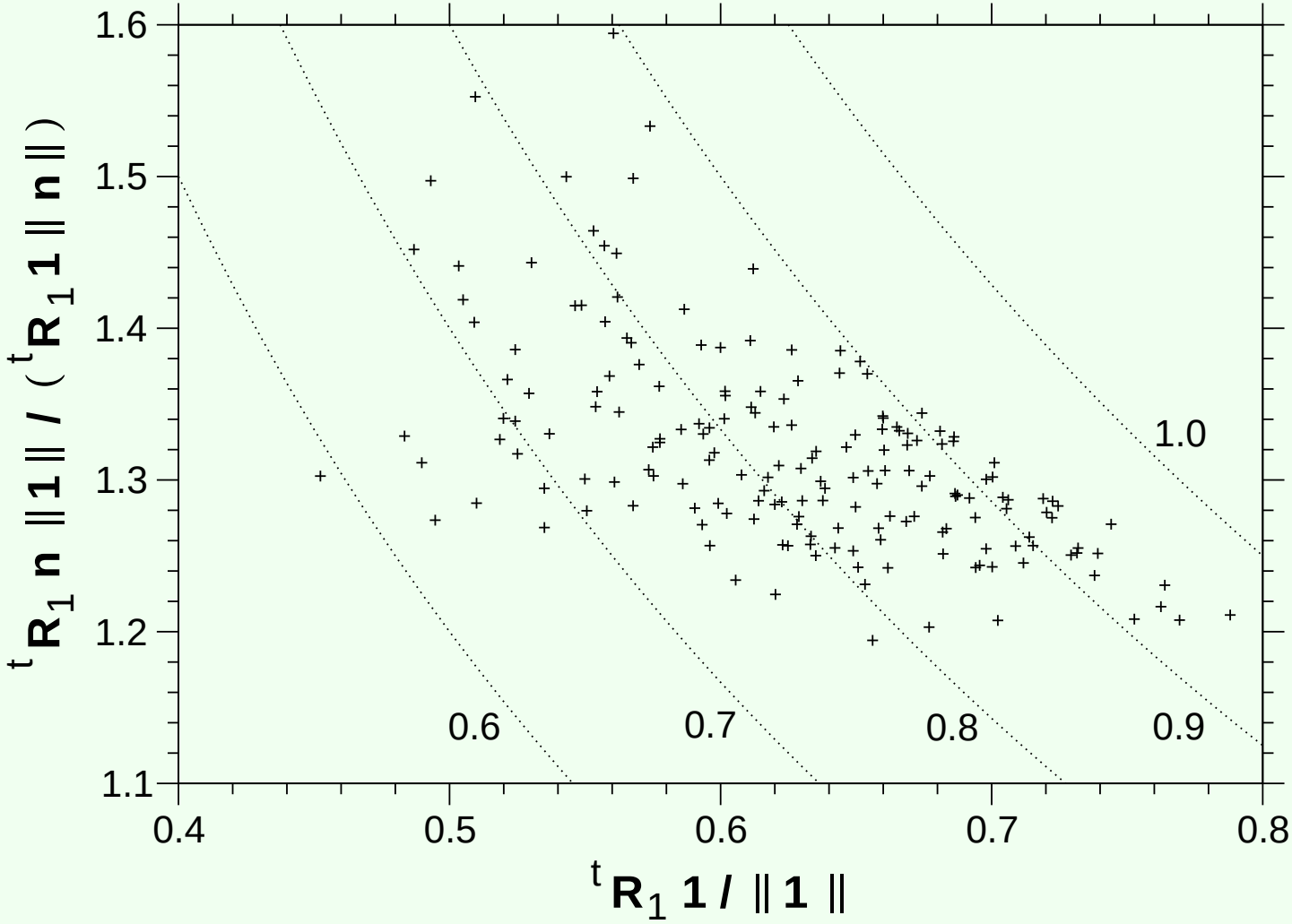
This optimum value of ε_0 is used for the singular decomposition of E matrices in the following analyses.

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



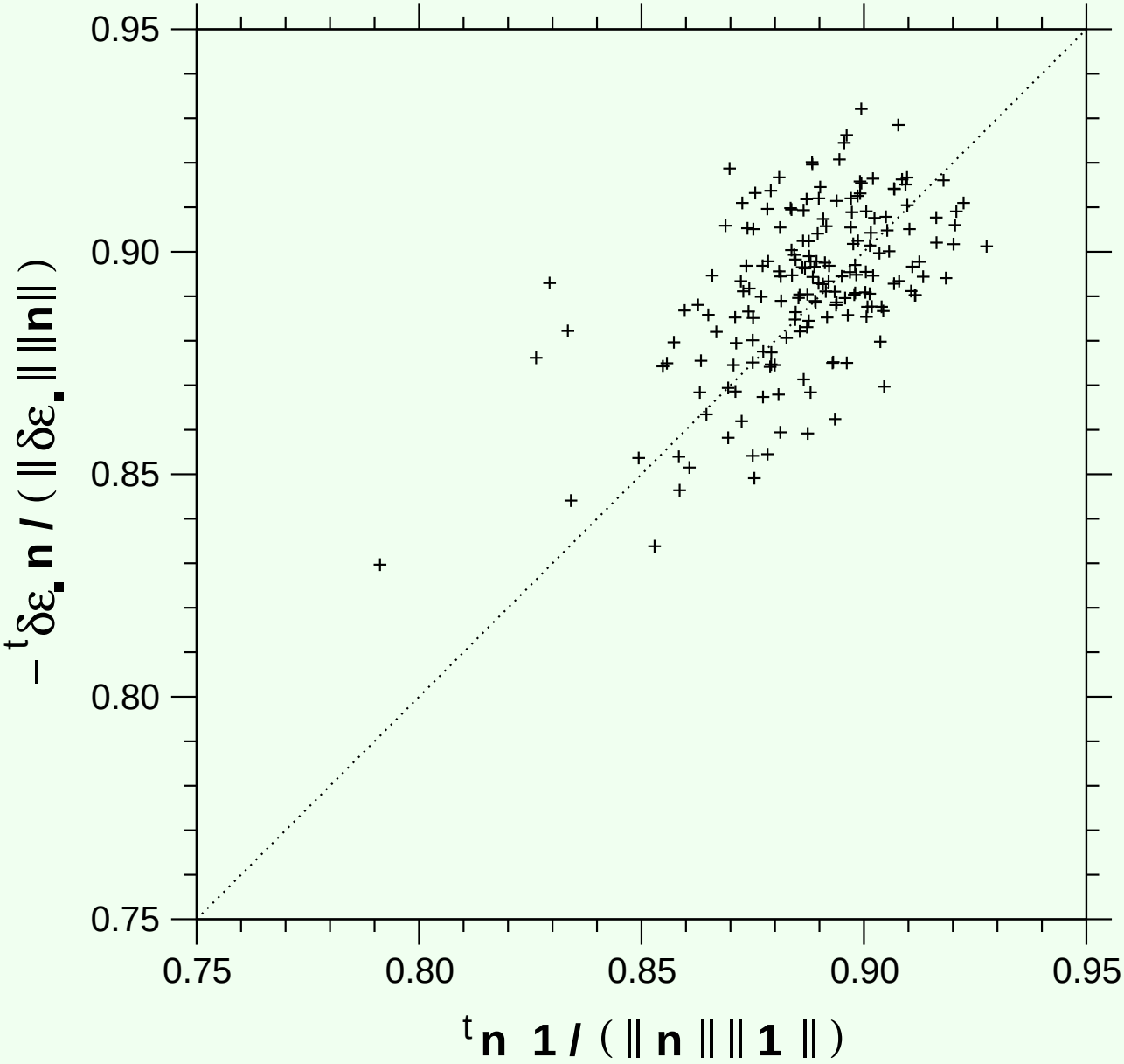
In the t-tests of the correlation coefficients between $\boldsymbol{R}_1$ and $\boldsymbol{V}_1$, the geometric mean of probabilities for a significance over 182 proteins is equal to $\exp(-18.4)$.

The primary 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.

The contact vector, n, tends to be parallel to the mean contact energy vector, $-\delta\vec{\mathcal{E}}_\bullet$, although not strongly.



In the t-tests of the correlation coefficients between n and $-\delta\vec{\mathcal{E}}_\bullet$, the geometric mean of probabilities for a significance over 182 proteins is equal to $\exp(-27.9)$.

4. CONCLUSIONS

Pairwise contact interactions in proteins force native C matrices to be in a relationship as if the interactions are a Go-like potential. As a result, the following relationships between E- and C- matrices for protein native structures are expected, and have been observed in 182 representative proteins;

1. a parallel relationship between the principal eigenvectors of the C- and E-matrices, and
2. a parallel relationship between a contact number vector and the principal eigenvectors of the C- and E-matrices,

provided that the *E-matrix is eigen-decomposed with an additional constant term* that corresponds to a threshold of contact energy that approximately separates native contacts from non-native ones.

The spectral representation of C- and E-matrices also reveals that *pairwise residue-residue interactions*, which depends only on the types of interacting amino acids, *are insufficient and other interactions including residue connectivities and steric hindrance are needed to make native structures unique lowest-energy conformations.*

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