(別紙様式13)

ABSTRACTS OF RESEARCH PROJECT, GRANT-IN-AID FOR SCIENTIFIC RESEARCH (2001)

1. RESEARCH INSTITUTION NUMBER : 12301

2. RESEARCH INSTITUTION : Gunma University

3. CATEGORY : Grant-in-Aid for Scientific Research (C) (2)

4. TERM OF PROJECT (2000 ~ 2001)

5. PROJECT NUMBER : 12680651

6. TITLE OF PROJECT : Protein sequence–structure alignments for predicting protein's structures and functions from their sequences

		REGISTERED NUMBER	NAME	INSTITUTION, DEPARTMENT, TITLE OF POSITION
7. HEAD INVESTIGATOR		60190774	Miyazawa, Sanzo	Gunma University, Faculty of Technology Associate Professor
8. INVESTIGATORS	(1)	REGISTERED NUMBER	NAME	INSTITUTION, DEPARTMENT, TITLE OF POSITION
	(2)	〃	〃	〃
	(3)	〃	〃	〃
	(4)	〃	〃	〃
	(5)	〃	〃	〃

9. SUMMARY OF RESEARCH RESULTS

We examine how effectively simple potential functions previously developed can identify compatibilities between sequences and structures of proteins for database searches. The potential function consists of pairwise contact energies, repulsive packing potentials of residues for overly dense arrangement and short-range potentials for secondary structures, all of which were estimated from statistical preferences observed in known protein structures. Each potential energy term was modified to represent compatibilities between sequences and structures for globular proteins. Pairwise contact interactions in a sequence-structure alignment are evaluated in a mean field approximation on the basis of probabilities of site pairs to be aligned. Gap penalties are assumed to be proportional to the number of contacts at each residue position, and as a result gaps will be more frequently placed on protein surfaces than in cores. In addition to minimum energy alignments, we use probability alignments made by successively aligning site pairs in order by pairwise alignment probabilities. Results show that the present energy function and alignment method can detect well both folds compatible with a given sequence and, inversely, sequences compatible with a given fold, and yield mostly similar alignments for these two types of sequence and structure pairs. Probability alignments consisting of most reliable site pairs only can yield extremely small root mean square deviations, and including less reliable pairs increases the deviations. Also it is observed that secondary structure potentials are usefully complementary to yield improved alignments with this method. Remarkably, by this method some individual sequence-structure pairs are detected having only 5-20 % sequence identity.

10. KEY WORDS

(1) Statistical potentials (2) Protein inverse folding (3) Protein fold recognition
(4) Protein sequence-structure alignments (5) Protein sequence alignments (6) Protein homology search
(7) Protein structure prediction (8) Structural genomics

(CONTINUE TO NEXT PAGE)

11. REFERENCES

AUTHORS , TITLE OF ARTICLE	JOURNAL, VOLUME-NUMBER, PAGES CONCERNED, YEAR
Miyazawa, S. Identifying sequence-structure pairs undetected by sequence alignments.	*Protein Engineering*, **13**, 459-475, 2000.
Miyazawa, S. Protein sequence-structure alignment based on site-alignment probabilities.	*Genome Informatics*, **11**, 141-150, 2000.