DTIC FILE COPY

(2)

(U)

SEC

AD-A209 574

...ENTATION PAGE

Form Approved
OMB No. 0704-0188

1a.	1b. RESTRICTIVE MARKINGS NA
2a. NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Distribution Unlimited
2b. DECLASSIFICATION / DOWNGR...NG SCHEDULE NA	
4. PERFORMING ORGANIZATION ...PO... NUMBER(S) NA	5. MONITORING ORGANIZATION REPORT NUMBER(S) NA
6a. NAME OF PERFORMING ORGANIZATION University of Rochester School of Medicine & Dentistry	6b. OFFICE SYMBOL (If applicable) NA
6c. ADDRESS (City, State, and ZIP Code) Dept. of Biophysics 601 Elmwood Avenue Rochester, NY 14642	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research 7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable) ONR
8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0322 10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N	RR04108		

11. TITLE (Include Security Classification)

(U) Quantitative Classification of Known Protein Structures

12. PERSONAL AUTHOR(S)
Shalom R. Rackovsky

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 7/88 TO 6/89	6/30/89	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
08			Protein structures, structure comparison, structure classification

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Mathematical methods have been developed for comparing the structures of proteins of arbitrary different molecular weight. A data base of structures representative of the entire Brookhaven Protein Data Bank has been assembled, containing 123 protein structures. This data base has been analyzed, using the methods developed, to determine the large-scale and local structure of the proteinome. Evidence exists for the existence of two distinct types of β-sheet/barrel proteins, and several types of mainly α-helical structures. Studies as a function of structural length scale and resolution are possible with the methods developed herein. Keywords: Molecular structure. (KT)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	(U)

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Michael T. Marron	(202) 696-4038	ONR

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

R&T CODE: 4415805

DATE: 1 JULY 1989

ANNUAL REPORT ON CONTRACT N00014-88-K-0322

PRINCIPAL INVESTIGATOR: Shalom R. Rackovsky

CONTRACTOR: University of Rochester

CONTRACT TITLE: Quantitative Classification of Known Protein Structures

START DATE: 1 July, 1988

RESEARCH OBJECTIVE: To develop mathematical methods for comparing and classifying the structures of proteins of arbitrary molecular weight; to develop a data base of protein x-ray structures representative of the entire set of known protein structures; to apply the methods developed to the data base, in order to understand the structural relationships between the known protein structures; to develop new insight into aspects of protein folding and evolutionary relationships based on the results of the comparison and classification studies.

PROGRESS (Year 1): Since ONR funding began on July 1, rapid progress has been made on several of the stated research objectives. A set of mathematical tools has been developed, using methods of graph, matrix and classification theories, which make it possible to analyze very large structural data bases rapidly, accurately and easily. These have been implemented computationally. A data base of 123 protein x-ray structures has been assembled from the Brookhaven Protein Data Bank, spanning the entire range of known protein structures. This set of proteins has been studied on the 4-, 5- and 6-alpha-carbon length scales, with respect to both global organization of the proteinome (i.e. the relationships between large groups of proteins) and local organization (i.e. definition of structures which are similar to any given protein). Observations on this data base confirm some of our earlier findings, made on a much smaller data base using less sophisticated methods of analysis, on structural relationships as a function of length scale and on convergent evolution of folding mechanisms. Some unsuspected structural relationships have come to light. For example, it has been demonstrated that there are two distinct classes of β-sheet/barrel proteins- one with mainly flat extended structures, and one with mainly twisted (right- or left-handed) extended structures. Evidence is also beginning to emerge for the existence of three classes of mainly helical proteins. It has been shown that the representational methods developed are able not only to demonstrate the similarity of related proteins, but to detect anomalies arising from the presence in the data base of proteins with less accurately determined struc-

tures. It has further been shown that these anomalous relationships disappear when the analysis is carried out at lower resolution, as befits low-quality structures.

WORK PLAN (Year 2): It is planned to extend the analysis to longer length scales, where one expects that structural relationships will become even more strongly defined. In addition, it is planned to study the effect of resolution of the representation, as this affects the results of the clustering studies which define the structure of the proteinome. It is also hoped to carry out studies of non-sequential backbone structures, i.e. those defined by virtual bonds which do not connect successive alpha carbons. Some of these, such as the crossover connection, have been shown to be highly characteristic features of protein structure. The methods we have developed make it possible to carry out a complete, quantitative census of such structures for the first time, and this can yield potentially useful, previously inaccessible data. It is anticipated that a start will be made this year in connecting the masses of data resulting from our analyses with the fundamental folding and evolutionary questions of interest. It is also planned to develop methods, based on our mathematical tools, for structural alignment of proteins and for the rapid search of the complete data base for fragments similar to a chosen structure.

INVENTIONS: No inventions have resulted from this work.

PUBLICATIONS AND REPORTS: A paper is currently in preparation which will detail the results of our studies to date. In addition, a Progress Report was submitted to the ONR Distribution List, as required, by June 1, 1989.

TRAINING ACTIVITIES: No students have been associated with this project this year. A graduate student will be joining the project on October 1, 1989.

AWARDS/FELLOWSHIPS: None.