
Contents

1	Introduction	1
2	Proteins	3
2.1	Primary Structure of the Protein	4
2.1.1	Amino Acids	4
2.1.2	Polypeptide Composition	5
2.1.3	Primary Structure	6
2.2	Secondary Structure	6
2.2.1	Torsion Angles	7
2.2.2	Secondary Structure Elements	8
2.3	Tertiary and Quaternary Structure	8
2.4	Determination of Protein Structure	10
2.5	Protein Databases	11
2.5.1	Primary Databases	11
2.5.2	Secondary Databases	13
2.6	Comparison of the Protein Structure	14
2.7	Sequence Motifs	15
3	Cluster Analysis	19
3.1	Data Structures in Clustering	20
3.2	Partitioning Methods	21
3.2.1	The K-means Method	22
3.2.2	The K-medoids Method	24
3.3	Density-based Methods	27
3.3.1	Clustering Based on Density Distribution Functions ...	28
4	State of the Art	31
4.1	Sequence-Structure Relationships in Proteins	31
4.1.1	Sequence-Structure Fragments Based on Clustering of Sequence Space	32

VIII Contents

4.1.2	Sequence-Structure Fragments Based on Clustering of Structure Space	32
4.1.3	Protein Loops Motifs	34
4.1.4	Nonlocal Structure Motifs	35
4.1.5	Summarization	36
4.2	Clustering of Protein Fragments	36
4.3	Protein Structure Prediction	37
4.3.1	Ab Initio Modeling	38
4.3.2	Comparative Modeling	38
4.3.3	Protein Threading	38
4.3.4	Fragment Assembly	39
5	Motivation and Goals of this Monograph	41
6	Density-based Clustering of Protein Substructures	45
6.1	DENCLUE-PS Method	46
6.2	Parameters of the DENCLUE-PS Method	51
6.3	Input Data	52
6.4	Experimental Results	53
6.4.1	Experiment 1 - Verify the Existence of the Target Sequences	55
6.4.2	Experiment 2 - Confirm the Unity with the Requirement	55
6.4.3	Experiment 3 - Tabulate the Number of Sequence Fragments	56
6.4.4	Experiment 4 - Properties of the Method	59
6.5	Comparison of the Method	59
7	Application of DENCLUE-PS Method on Fragment Search	67
7.1	Sequence Fragment Selection	68
7.2	Seed Search	68
7.3	Refinement of the Seeds of the Sequence-structure Fragments	70
7.4	An Example of Sequence-Structure Fragment	73
8	Summary	83
References	85	
Input set for the Comparison	93	
AVIVL fragment	99	