

# Contents

|          |                                                                                                                          |           |
|----------|--------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Formal Model of 3D Protein Structures for Functional Genomics, Comparative Bioinformatics, and Molecular Modeling</b> | <b>1</b>  |
| 1.1      | Introduction                                                                                                             | 1         |
| 1.2      | General Definition of Protein Spatial Structure                                                                          | 2         |
| 1.3      | A Reference to Representation Levels                                                                                     | 4         |
| 1.3.1    | Primary Structure                                                                                                        | 4         |
| 1.3.2    | Secondary Structure                                                                                                      | 6         |
| 1.3.3    | Tertiary Structure                                                                                                       | 8         |
| 1.3.4    | Quaternary Structure                                                                                                     | 11        |
| 1.4      | Relative Coordinates of Protein Structures                                                                               | 13        |
| 1.5      | Energy Properties of Protein Structures                                                                                  | 17        |
| 1.6      | Summary                                                                                                                  | 19        |
|          | References                                                                                                               | 20        |
| <b>2</b> | <b>Multithreaded PSS-SQL for Searching Databases of Secondary Structures</b>                                             | <b>25</b> |
| 2.1      | Introduction                                                                                                             | 25        |
| 2.2      | Storing and Processing Secondary Structures in a Relational Database                                                     | 28        |
| 2.2.1    | Data Preparation and Storing                                                                                             | 28        |
| 2.2.2    | Indexing of Secondary Structures                                                                                         | 30        |
| 2.2.3    | Alignment Algorithm                                                                                                      | 30        |
| 2.2.4    | Multithreaded Implementation                                                                                             | 33        |
| 2.3      | SQL as the Interface Between User and the Database                                                                       | 36        |
| 2.3.1    | Pattern Representation in PSS-SQL Queries                                                                                | 37        |
| 2.3.2    | Sample Queries in PSS-SQL                                                                                                | 38        |
| 2.4      | Efficiency of the PSS-SQL                                                                                                | 41        |
| 2.5      | Discussion                                                                                                               | 43        |
| 2.6      | Summary                                                                                                                  | 45        |
|          | References                                                                                                               | 45        |

|          |                                                                                |     |
|----------|--------------------------------------------------------------------------------|-----|
| <b>3</b> | <b>Parallel CUDA-Based Protein 3D Structure</b>                                |     |
|          | <b>Similarity Searching</b> . . . . .                                          | 49  |
| 3.1      | Introduction . . . . .                                                         | 49  |
| 3.1.1    | What Makes the Problem . . . . .                                               | 50  |
| 3.1.2    | CUDA Architecture and Construction<br>of GPU Devices . . . . .                 | 51  |
| 3.1.3    | CUDA-Enabled GPUs in Processing Biological Data . . .                          | 52  |
| 3.2      | CASSERT for Protein Structures Similarity Searching . . . . .                  | 53  |
| 3.2.1    | General Course of the Matching Method . . . . .                                | 55  |
| 3.2.2    | First Phase: Low-Resolution Alignment . . . . .                                | 56  |
| 3.2.3    | Second Phase: High-Resolution Alignment . . . . .                              | 58  |
| 3.2.4    | Third Phase: Structural Superposition<br>and Alignment Visualization . . . . . | 59  |
| 3.3      | GPU-Based Implementation of the CASSERT . . . . .                              | 60  |
| 3.3.1    | Data Preparation . . . . .                                                     | 60  |
| 3.3.2    | Implementation of Two-Phase Structural Alignment<br>in a GPU . . . . .         | 63  |
| 3.3.3    | First Phase of Structural Alignment in the GPU . . . . .                       | 64  |
| 3.3.4    | Second Phase of Structural Alignment in the GPU . . . . .                      | 68  |
| 3.4      | GPU-CASSERT Efficiency Tests . . . . .                                         | 70  |
| 3.5      | Discussion . . . . .                                                           | 74  |
| 3.6      | Summary . . . . .                                                              | 76  |
|          | References . . . . .                                                           | 77  |
| <b>4</b> | <b>Cloud Computing for 3D Protein Structure Alignment</b> . . . . .            | 81  |
| 4.1      | Introduction . . . . .                                                         | 81  |
| 4.1.1    | Cloud Computing in Bioinformatics<br>and Life Sciences . . . . .               | 82  |
| 4.1.2    | Cloud Deployment and Service Models . . . . .                                  | 83  |
| 4.1.3    | Microsoft Azure . . . . .                                                      | 84  |
| 4.2      | Cloud4Psi for 3D Protein Structure Alignment . . . . .                         | 86  |
| 4.2.1    | Use Case: Interaction with Cloud4Psi . . . . .                                 | 87  |
| 4.2.2    | Architecture and Model of the Cloud4Psi . . . . .                              | 89  |
| 4.2.3    | Algorithms for Protein Structure Similarity Searching . . .                    | 92  |
| 4.2.4    | Implementation of Similarity Searching<br>in Azure Cloud . . . . .             | 92  |
| 4.3      | Efficiency of the Cloud4Psi . . . . .                                          | 98  |
| 4.4      | Discussion . . . . .                                                           | 100 |
| 4.5      | Summary . . . . .                                                              | 100 |
|          | References . . . . .                                                           | 101 |

|                                                    |      |
|----------------------------------------------------|------|
| Contents                                           | xvii |
| <b>5 General Discussion and Concluding Remarks</b> | 103  |
| 5.1 General Discussion                             | 103  |
| 5.2 Concluding Remarks                             | 105  |
| References                                         | 106  |
| <b>Index</b>                                       | 107  |

High-Performance Computational Solutions in Protein  
Bioinformatics

Mrozek, D.

2014, XIX, 109 p. 56 illus., 21 illus. in color., Softcover

ISBN: 978-3-319-06970-8