Contents

1 Introduction *1*
1.1 Background and Motivation *1*
Bibliography *12*

2 Fundamental of Theoretical Calculations on Protein Systems *15*
2.1 Strategies for Protein Calculations and Predictions *15*
2.2 Force Field-based Calculation on Protein Systems *21*
2.3 Quantum Mechanical Calculation on Protein Systems *28*
2.4 ML on Protein Systems *34*
Bibliography *39*

3 Protein Structure Prediction by Artificial Intelligence *43*
3.1 AlphaFold from Google *43*
3.1.1 Overview *43*
3.1.2 Previous Work *43*
3.1.3 What Is AlphaFold Expected to Do *44*
3.1.4 Architecture of AlphaFold *44*
3.1.4.1 Overall Architecture *44*
3.1.4.2 Evoformer Block Architecture *45*
3.1.4.3 Structure Prediction Model *47*
3.1.5 Training Methods: Using Both Labeled and Unlabeled Data *48*
3.1.6 Interpretation of the Network *48*
3.1.7 Conclusion and Discussion *48*
3.2 ESM and ESMFold from Meta AI *48*
3.2.1 Meta's ESM2 *49*
3.2.1.1 Background *49*
3.2.1.2 Methods *49*
3.2.1.3 Evaluation and Results *49*
3.2.1.4 Conclusion *51*
3.2.2 Meta's ESMFold *52*
3.2.2.1 Related Work *52*

3.2.2.2 Method *52*
3.2.2.3 Results and Discussions *54*
Bibliography *54*

4 Methods and Tools for Predicting Protein Folding Free Energy Change upon Mutation *57*
4.1 Introduction *57*
4.1.1 Background *57*
4.1.2 Experimental Methods *58*
4.1.3 Non-machine Learning Methods *58*
4.1.4 ML-based Methods *59*
4.2 Clustered Tree Regression *60*
4.2.1 Overview *61*
4.2.2 Data Distribution and Cleaning *62*
4.2.3 Feature Engineering *63*
4.2.4 Model Training *65*
4.2.5 Performance of CTR *65*
4.3 Materials and Methods *68*
4.3.1 Data Preparation *68*
4.3.2 Feature Extraction *68*
4.3.3 Experimental Setup *69*
4.4 Conclusion *69*
Bibliography *70*

5 Deep Neural Network-assisted Full-system Quantum Mechanical (FQM) Calculations of Proteins *75*
5.1 Introduction *75*
5.1.1 Background *75*
5.1.2 Advances in DNNs for Molecular Calculations *81*
5.2 DNN-based GFCC Method *84*
5.2.1 Methodology Overview *84*
5.2.2 Methods *87*
5.2.3 Results and Discussion *90*
5.2.4 Conclusion *93*
5.3 DNN-based Two-body Molecular Fractionation with Conjugate Caps *95*
5.3.1 Methodology Overview *95*
5.3.2 Neural Network Training and Energy Reconstruction *98*
5.3.3 Discussion and Implications *101*
5.4 Conclusion *106*
Bibliography *107*

6 Transfer Learning-assisted FQM Calculations of Proteins *111*
6.1 Introduction *111*
6.1.1 Background *111*

"ftoc_PrintPDF" — 2025/9/15 — 19:13 — page vii — #3

6.1.2 Bridging the Gap: from Classical Methods to TL in Protein FQM Calculations *112*
6.2 Inductive Transfer Learning Force Field *113*
6.2.1 Introduction *113*
6.2.2 Methods *114*
6.2.3 Discussion and Implications *116*
6.2.4 Conclusion and Outlook *118*
6.3 Transfer-learning-based Deep Learning Protocol for FQM Calculations *120*
6.3.1 Introduction *120*
6.3.2 Methods *121*
6.3.3 Discussion and Implications *124*
6.3.4 Conclusion and Outlook *126*
Bibliography *127*

7 Protein Interaction Prediction with Artificial Intelligence *129*
7.1 Background *129*
7.2 Methods and Technical Framework *131*
7.2.1 Sequence-based Methods *131*
7.2.2 Structure-based Methods *133*
7.2.3 Network-based Approaches *135*
7.2.4 Text Mining and Literature-based Prediction *136*
7.2.5 Hybrid Approaches *139*
7.3 Results *141*
7.4 Summary and Outlook *143*
7.4.1 Summary of Achievements *143*
7.4.2 Outlook and Future Directions *144*
Bibliography *146*

8 Protein Function Annotation with Machine Learning *149*
8.1 Background *149*
8.2 Methods and Technical Framework *152*
8.3 Protein Function Annotation *154*
8.3.1 Homology-based Annotation *156*
8.3.2 Structural Annotation *156*
8.3.3 Ontology-based Annotation *158*
8.3.4 Network-based Annotation *159*
8.3.5 ML- and AI-based Methods *160*
8.3.6 High-throughput Experimental Techniques *161*
8.3.7 Hybrid Methods *162*
Bibliography *163*

9 Machine Learning-driven Ab Initio Protein Design *165*
9.1 Background *165*
9.2 Advances and Applications in ML-driven Ab Initio Protein Design *167*
9.3 Graph Neural Networks for Flexible Protein Design *171*

9.3.1 Energy Function of Neural Networks for Protein Design *173*
9.3.2 ML Models and Technologies *176*
9.3.3 ML and AI-based Methods *177*
9.3.4 Conclusion and Outlook *179*
Bibliography *183*

10 Large Language Model of Protein Systems *185*
10.1 Background *185*
10.2 Methodology and Training of Protein Language Models *189*
10.2.1 Methodological Framework for Protein Language Models *189*
10.2.2 Training Strategies for PLMs *191*
10.3 Applications of PLMs *192*
10.3.1 Case Studies in the Application of PLMs *192*
10.4 Case Introduction: Decoding Protein Interactions Using ESM *195*
10.4.1 Technical Foundations *195*
10.4.2 Applications in Protein Design *196*
10.4.3 Future Directions *198*
10.5 Conclusion and Outlook *201*
Bibliography *203*

11 Outlook *207*
11.1 Background *207*
11.2 AI's Transformative Role *210*
11.3 Applications Across Protein Science *214*
11.4 AI-Augmented Experimental Techniques in Modern Protein Research *217*
Bibliography *220*

Index *223*