

# VU Research Portal

## Preface to Introduction to Structural Bioinformatics

Feenstra, K. Anton; Abeln, Sanne

### **published in**

Introduction to Structural Bioinformatics  
2018

### **document version**

Publisher's PDF, also known as Version of record

### **document license**

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Feenstra, K. A., & Abeln, S. (2018). Preface to Introduction to Structural Bioinformatics. In *Introduction to Structural Bioinformatics* (pp. 5-9). Centre for Integrative Bioinformatics (IBIVU) and Department of Computer Science, Vrije Universiteit. <https://www.semanticscholar.org/paper/Introduction-to-Structural-Bioinformatics-Feenstra-Abeln/6943b5e364a9fcb82f868e3184bb234bf733cdc4>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Introduction to Structural Bioinformatics

K. Anton Feenstra      Sanne Abeln

Centre for Integrative Bioinformatics (IBIVU), and  
Department of Computer Science,  
Vrije Universiteit, De Boelelaan 1081A, 1081 HV Amsterdam, Netherlands

January 30, 2018

## **Abstract**

While many good textbooks are available on Protein Structure, Molecular Simulations, Thermodynamics and Bioinformatics methods in general, there is no good introductory level book for the field of Structural Bioinformatics. This book aims to give an introduction into Structural Bioinformatics, which is where the previous topics meet to explore three dimensional protein structures through computational analysis. We provide an overview of existing computational techniques, to validate, simulate, predict and analyse protein structures. More importantly, it will aim to provide practical knowledge about how and when to use such techniques. We will consider proteins from three major vantage points: Protein structure quantification, Protein structure prediction, and Protein simulation & dynamics.

# Contents

Contents	3
0 Preface to Introduction to Structural Bioinformatics	4

# **Chapter 0**

## **Preface to Introduction to Structural Bioinformatics**

**K. Anton Feenstra   Sanne Abeln**

Centre for Integrative Bioinformatics (IBIVU) and  
Department of Computer Science,  
Vrije Universiteit, De Boelelaan 1081A, 1081 HV Amsterdam, Netherlands

---

Why did we write this book? Firstly, because all the authors involved are very enthusiastic about protein structures and scientific computation, and secondly because we have been setting up a successful MSc level Bioinformatics in Structural Bioinformatics course that is in immediate need for a good text book.

While many good textbooks are available on Protein Structure, Molecular Simulations, Thermodynamics and Bioinformatics methods in general, there is no good introductory level book for the field of Structural Bioinformatics. This book aims to give an introduction into Structural Bioinformatics, which is where the previous topics meet to explore three dimensional protein structures through computational analysis.

This book will provide you an overview of existing computational techniques, to validate, simulate, predict and analyse protein structures. More importantly, it will aim to provide practical knowledge about how and when to use such techniques. We will consider proteins from three major vantage points, as illustrated in [Figure 1](#):

- I) Protein structure quantification;
- II) Protein structure prediction; and
- III) Protein simulation & dynamics.

Part I – *Structure Quantification* deals with comparing one protein structure to another. This can either be in a general sense: a protein structure is compared to a large set of reference structures to validate the experimental reliability of the structure. Or the comparison may be to one or more specific protein structures, with the question how similar the protein folds are.

Part II – *Structure prediction* deals with the question, how to predict the structure given a protein sequence. We start with a graceful introduction to protein structure basics [[Abeln et al., 2017a](#)]. We will see that there is a wide range of methods available, and that the reliability of such method varies strongly. It is important to understand how and when structure prediction will give you trustworthy outcomes [[Abeln et al., 2017b](#)]. Then we reach perhaps the most salient question: from an available structure, what can we say about the possible function this protein could perform.

Part III – *Simulation & dynamics* considers the dynamics or the ensemble of possible structures a protein may take up. Generally speaking however, it is not yet possible to allow proteins to fold within a simulation, nevertheless simulations are vital to understand functional mechanisms of structural ensembles of proteins, specifically in context of small metabolites, the cell membrane or other proteins.

Herein, we introduce relevant concepts such as structure basics, structure prediction, homology modeling, statistical thermodynamics, simulation. From here, the reader may be able to acquire further information through self study and further references. In particular, we will treat the protein not as a rigid structural entity, but as dynamic ensembles of structures which

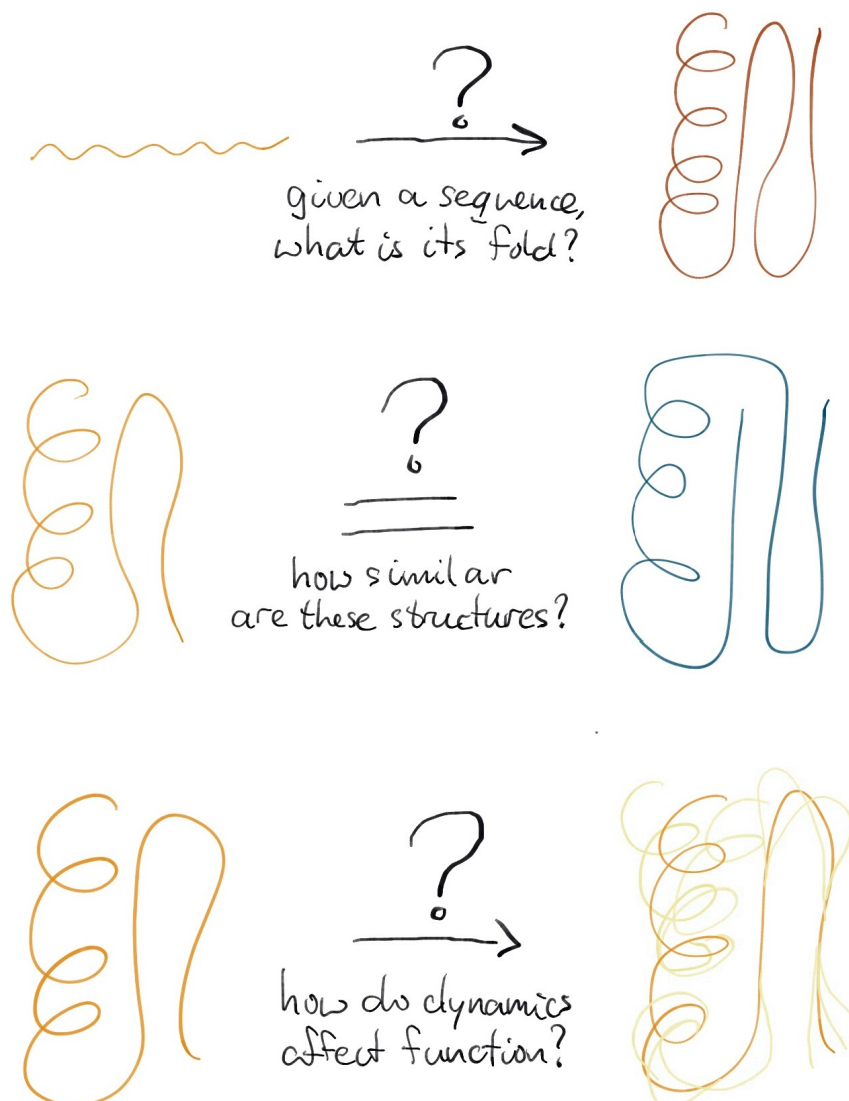


Figure 1: Within the field of Structural Bioinformatics three dimensional protein structures are investigated through computational analysis. Important problems that may be addressed computationally are shown in the form of cartoons. Firstly, how does the genomic sequence of a gene translate into the folded, functional protein structure? Secondly, when considering two proteins, how similar are their structures? And, last but not least, since we know proteins are not static entities, how do flexibility and dynamics play a role in the function of the protein?

make the protein able to perform its function or functions.

Hence this book aims to provide a framework for all important concepts within the field of Structural Bioinformatics. However, we do not aim to give extensive reviews of all the methods available in each of these subjects; for these we will refer, where appropriate, to other books.

The primary audience of this book are master students in the area of bioinformatics, or a related discipline like biotechnology. Basic background knowledge is assumed on protein structure, bioinformatics, sequence analysis, dynamic programming algorithms, calculus and basic chemistry. A few good books we would like to mention here, upfront, as they may be important reference material for readers of this book, depending on their background:

- Protein Structure – [Branden and Tooze \[1998\]](#)
- Understanding Bioinformatics – [Zvelebil and Baum \[2008\]](#)
- Essential Bioinformatics – [Xiong \[2006\]](#) (very basic)
- An introduction to Thermal Physics – [Schroeder \[1999\]](#)
- Introduction to Python – [Lutz \[2013\]](#)

And some more advanced books, that go further on particular topics than this book:

- Structural Bioinformatics – [Gu and Bourne \[2009\]](#)
- Sequence Analysis – [Durbin \[1998\]](#)
- Molecular Simulation – [Frenkel and Smit \[2002\]](#)
- Molecular Modelling – [Leach \[2001\]](#)

We hope to provide the reader with knowledge about the type of problems that are scientifically feasible to solve. For example, predicting a protein structure through homology modeling will generally give high quality and reliable results. In contrast, we also would like to point out which problems are still unlikely to yield good results, such as predicting protein-protein interactions, and which are still out of scope altogether for existing methods, such as correct prediction of protein folding from sequence information alone. Most importantly, we will introduce the fundamental concepts on which the method are based, and what assumptions there are for using such methods. Therefore it should become apparent which inherent limitations the techniques have and what the techniques can successfully be used for.

The setup of this books follows the same guiding principle that we apply throughout our masters' programme. We introduce a **focus** on open and challenging research problems to solve, and the technical skills to solve them; a **balance** between basic skills, such as biology, mathematics (modelling) and computational tools (programming); then, building on this focus and balance, **translate** between different adjoining disciplines, and between tools and methods on the one hand, and application and problems on the other. This is summarized in [Figure 2](#).

Lastly, we would like to thank all our students who have followed the MSc course Structural Bioinformatics at the VU University in Amsterdam



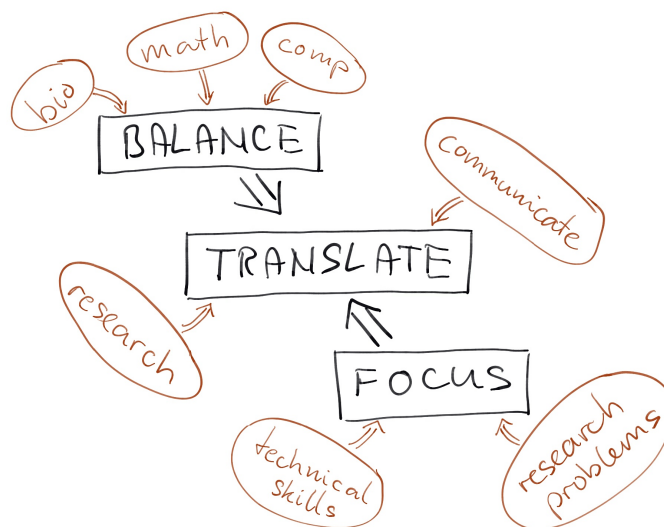


Figure 2: Conceptual organisation of our bioinformatics education programme along three key elements: Translate, Balance, and Focus.

for their enthusiasm in pointing out mistakes in our lectures and asking important additional questions; without this vital input it would have been absolutely impossible to write this book. In particular, we thank Nicola Bonzanni, Kamil K. Belau, Ashley Gallagher, Jochem Bijlard, and Reza Haydarlou for insightful discussions and critical proofreading of early versions.

# Bibliography

- Sanne Abeln, Jaap Heringa, and K. Anton Feenstra. Introduction to Protein Structure Prediction. *arXiv*, 1712.00407, dec 2017a. URL <http://arxiv.org/abs/1712.00407>.
- Sanne Abeln, Jaap Heringa, and K. Anton Feenstra. Strategies for protein structure model generation. *arXiv*, 1712.00425, dec 2017b. URL <http://arxiv.org/abs/1712.00425>.
- C Branden and J Tooze. *Introduction to protein structure*. garland publishing, New York, 1998. ISBN 9780815323044.
- Richard Durbin. *Biological sequence analysis : probabilistic models of proteins and nucleic acids*. 1998. ISBN 9780521629713.
- Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Pr, San Diego, second edition, 2002. ISBN 0122673514.
- Jenny. Gu and Philip E. Bourne. *Structural bioinformatics*. Wiley-Blackwell, 2009. ISBN 0470181052.
- Andrew Leach. *Molecular Modelling: Principles and Applications*. Pearson, 2001. ISBN 9780582382107.
- Mark Lutz. *Learning Python*. O'Reilly, 2013. ISBN 1449355730.
- Daniel V. Schroeder. *An Introduction to Thermal Physics*. Addison-Wesley Publishing Company, San Francisco, CA, 1999. ISBN 0201380277.
- Jin Xiong. *Essential Bioinformatics*. Cambridge University Press, 2006. ISBN 9780511806087. doi: 10.1017/CBO9780511806087.
- M. Zvelebil and J.O. Baum. *Understanding Bioinformatics*. Garland Science, Taylor & Francis Group, New York – London, nov 2008. ISBN 9780815340249.