

Introduction

More than three decades ago, the hope emerged that protein engineering would be able to predict protein and enzyme function on the basis of X-ray crystal structures. The expectations were that we should be able to create goal-oriented functions in the enzyme of interest. A large effort was made to obtain the structures of enzymes of great importance for understanding biological processes and enzymes of general commercial interest in many industries. A large variety of structures of enzymes from many biological pathways, as well as enzymes of commercial interest, have been solved, including carbohydrate-acting enzymes, proteolytic enzymes, and lipolytic enzymes, and have helped tremendously in understanding the structure-function relationships. They have also revealed how much we still need to learn in order to manipulate genes to make enzymes react in a desired way.

Today, there are at least two major focuses on gaining benefit from and knowledge about enzyme function: (1) data analysis and (2) a more detailed understanding. Much learning cannot be said to be statistically feasible, but I hope the scientific society will still accept a few examples as feasible hypotheses to investigate further. With the increasing knowledge on enzyme function, with input from atomistic mobility and hydrogen bonding, the shifting electrostatics situation due to mobility and changes in relative coordinated atoms and macroscopic dependencies on enzyme environment changes leaves us with a very complex multidimensional space for how enzymes work. This makes it nearly experimentally unfeasible to have enough statistics on all the possible impact characteristics, as theoretically needed, making it difficult to draw sound, comprehensive, and significant conclusions. Commonly, even very large data sets will reveal single conclusions but are incorrectly

drawn since the number of data sets for each parameter alone is too few to make findings statistically significant. The data analysis will definitely add to a more detailed understanding and to suggestions for function. Some chapters touch upon data-driven discovery, but most of the chapters are focused on hypothesis-driven research testing one specific enzyme in a specific environment and with few parameters, giving exciting insights into the complexity of enzyme nature.

During my work in developing enzymes for technical use and work on the enzyme-substrate interaction, it has been tempting to combine the information from quantum mechanical calculations of the energetics in the catalytic reaction, and the overall molecular mobility using standard force fields, as well as electrostatics calculations and docking in order to inform on three important topics of enzyme function, namely (1) the initial substrate binding to the enzyme, (2) the important local fitting to accommodate the correct spatial state that can contain the reactive state as seen by molecular dynamics mobility and hydrogen bonding patterns, and (3) the reactive state energetics as measured by quantum mechanical calculations. This overall reaction could be stated in a formula as shown below:

$$\text{Enzyme function} = f(\text{overall binding}) \\ + f(\text{local fluctuations and interactions}) + f(\text{reactive energy})$$

Or in other words, enzyme function is a function of three major key factors: (1) the overall fitting of the substrate for binding with the correct orientation for the more detailed local interactions in the nearer active site surroundings, (2) the necessary hydrogen bonding and electrostatic interactions to secure the correct arrangements for the catalysis reaction to take place, and (3) the quantum mechanical energy in the catalysis reaction. Seen from molecular dynamics simulations some hydrogen bonds are only present at a certain time during the simulation, indicating that activity only occurs when the structure is in a certain subdomain structure containing the important hydrogen bonds. If certain hydrogen bonds are in place at the same time the reaction can occur. If one of the three stated factors is not fulfilled at the same time, then no reaction occurs. Examples of important hydrogen bonds are presented in Chapter 15.

In Chapter 10 on sequences and design the combination of sequence alignment information, docking, and molecular simulation of variant molecules to extract more combinatorial information is discussed.

This book focuses on the current understanding obtained in the past 10–15 years to the present. In the 1980s focus was on making 3D structures and understanding and analyzing proteins. In the 1990s focus was on diversity methods and screening methods, whereas in the 2000s the focus has been on bioinformatics and simulation methods and statistical methods, as well as ultrahigh-throughput methods with revised views on proteins. Today we hope the analysis of large data will help find the desired results. Many new technologies have brought new insights into enzyme function, with emphasis on single-molecule behavior and molecular mobility and electrostatics, as well as enzymes working on large substrates and complex substrates. Focus on the mobility impact on substrate interaction can be found in Chapter 2.

The book is divided into four major sections: enzyme function (Chapters 1–8), enzyme design (Chapters 9–15), enzyme diversity (Chapters 16–20), and enzyme screening and analysis (Chapters 21–25). The enzyme function part addresses the enzyme kinetics on simple substrates in Chapter 1, as well as the more complex interaction on larger substrates in Chapters 4, 7, and 8. Also structural aspects are addressed in Chapter 6, NMR structures in Chapter 5, and further dynamic aspects in Chapters 2 and 3. The enzyme design part is focused on the sequence-derived design methods in Chapters 9, 10, and 11, as well as in Chapter 20, and 3D structural methods. The 3D structural design/understanding is mainly discussed in Chapters 12–15. The design area is also covered partly under enzyme diversity, especially in Chapter 16, which has a review of both diversity methods and some design ideas. Further under enzyme diversity are handled metagenomics, circular permutations, and ancestral reconstruction in Chapters 18, 19, and 20, respectively, as well as the number issues in directed evolution in Chapter 17. The enzyme screening and analysis part includes both *in silico* screening in Chapters 23 and 24 and wet chemistry screening methods in Chapters 21 and 22, as well as an example of analysis of enzyme variants in Chapter 25.

The computer simulations reveal great insight into the function of enzymes and can help in designing new functionalities and activities. The predictive power is still not precise, but we can use the simulations to screen for potential variants of interest, which then need testing for the desired function. Decades ago, one specific predicted variant was selected for testing—today it is commonly understood that a certain number of the, say, top 10 or 100 candidates could potentially be of interest. The speed of computers today allows for this kind of suggestions and sometimes also a reasonable simplification is used for making the screening possible. Chapters 23 and 24 address these possibilities. Also Chapter 16 touches upon the *in silico* design possibilities.

It is now more than a decade ago that enzyme promiscuity became a major field of interest. The versatility of enzymes and their activities are more open today than ever and the general EC classification system is seldom fully explanatory today. A few chapters touch upon the promiscuity—not from a specificity issue but rather a reaction mechanistic view; see Chapters 15 and 23.

Other screening methods in the wet chemistry part are being developed, and while screening has come out of the first decade in protein engineering, the limitations are getting more visible and the possibilities better utilized. A few chapters address the methodologies (Chapters 16, 17, 21, and 22)—micronanotechnology has gone into the screening area and possibilities for very high numbers have become a reality. Smart techniques to secure the picking of hits are important and an interesting method is mentioned in Chapter 22.

In an earlier book I edited, *Enzyme Engineering: Function, Design, Variant Generation and Screening*, the focus was more on the variant generation and screening part and less on the function and design part. In this book the main focus is on enzyme function and design and less on variant generation and screening methods. This reflects the fact that many new insights into the more complex enzyme function have emerged during the past many years. Massive quantities of information on variants of enzymes and the multiple states of the structures as well as single-molecule insight have added to the colligative understanding of enzyme function.

The production of many mutations has, besides a lot of data, also resulted in the realization of how little we still understand about

enzyme function. Therefore, this has been emphasized in the first eight chapters with examples from the versatility of factors influencing enzyme activity and enzyme-substrate interaction. Around 20 years ago the main enzyme understanding was based on simple kinetics and soluble substrate interactions. In industry, we are aware that the main enzyme function often occurs under conditions other than the simple substrate-enzyme interaction theory, very well described with mathematical equations. Chapter 3 (on single-enzyme function) and Chapter 2 (on enzyme motions) emphasize the rather complicated behavior of the enzymatic function, which continues to open new depths of understanding. Examples of these complicated behaviors are presented in Chapter 4 on surface-active enzymes and Chapter 7 on the carbohydrate-hydrolyzing enzyme family.

During the work on writing the book chapters representing important directions in enzyme research on enzyme function, design, engineering, and analysis, recent aspects have been published, including enzymes' use of the energy coming from the catalyzed chemical reaction itself, which adds to the chapters on mobility of the enzymes. Also the importance of electrostatics and the impact on enzyme function has not been directly addressed in the chapters but is clearly a major part of some of the added chapters and has been established as an important factor in enzyme function and catalysis. Clearly, more combinations of these factors mentioned in the chapters and above are needed in the future to further understand the full functional space of enzymes and thus understand how to address improvements by protein engineering.