



Understanding Enzymes

Function, Design, Engineering, and Analysis

edited by

Allan Svendsen



Understanding Enzymes

Function, Design, Engineering, and Analysis

edited by

Allan Svendsen



PAN STANFORD  PUBLISHING

Published by

Pan Stanford Publishing Pte. Ltd.
Penthouse Level, Suntec Tower 3
8 Temasek Boulevard
Singapore 038988

Email: editorial@panstanford.com
Web: www.panstanford.com

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

**Understanding Enzymes: Function, Design, Engineering,
and Analysis**

Copyright © 2016 Pan Stanford Publishing Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-4669-32-0 (Hardcover)
ISBN 978-981-4669-33-7 (eBook)

Printed in the USA

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

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.

Contents

Introduction

xix

PART I ENZYME FUNCTION

1 A Short Practical Guide to the Quantitative Analysis of Engineered Enzymes	3
<i>Christopher D. Bayer and Florian Hollfelder</i>	
1.1 Introduction	3
1.2 Quantifying Reaction Progress	4
1.3 Typical Saturation Plots Give Michaelis–Menten Parameters	5
1.4 What Can Go Wrong?	8
1.5 Dealing with Multiphasic and Pre-Steady-State Kinetics	12
1.6 Evaluating Enzymes	16
 2 Protein Conformational Motions: Enzyme Catalysis	 21
<i>Xinyi Huang, C. Tony Liu, and Stephen J. Benkovic</i>	
2.1 Introduction	21
2.2 Multidimensional Protein Landscape and the Timescales of Motions	22
2.3 Conformational Changes in Enzyme–Substrate Interactions	26
2.4 Conformational Changes in Catalysis	28
2.4.1 Protein Dynamics of DHFR in the Catalytic Cycle	30
2.4.2 Temporally Overlap: Correlation Does Not Mean Causation	32
2.4.3 Fast Timescale Conformational Fluctuations	34

2.4.4	Effect of Conformational Changes on the Electrostatic Environment	36
2.5	Conservation of Protein Motions in Evolution	38
2.6	Designing Protein Dynamics	39
2.7	Concluding Remarks	40
3	Enzymology Meets Nanotechnology: Single-Molecule Methods for Observing Enzyme Kinetics in Real Time	47
	<i>Kerstin G. Blank, Anna A. Wasielec, and Alan E. Rowan</i>	
3.1	Introduction	48
3.2	Single-Turnover Detection	53
3.2.1	Fluorescent Reporter Systems	53
3.2.2	Measurement Setup	56
3.2.3	Data Analysis	57
3.3	Single-Enzyme Kinetics	60
3.3.1	<i>Candida antarctica</i> Lipase B	63
3.3.2	<i>Thermomyces lanuginosus</i> Lipase	67
3.3.3	α -Chymotrypsin	73
3.3.4	Nitrite Reductase	78
3.3.5	Summary	84
3.4	New Developments Facilitated by Nanotechnology	88
3.4.1	Nano-optical Approaches	89
3.4.2	Nano-electronic Approaches	96
3.4.3	Nanomechanical Approaches	103
3.4.4	Summary	108
3.5	Conclusion	110
4	Interfacial Enzyme Function Visualized Using Neutron, X-Ray, and Light-Scattering Methods	125
	<i>Hanna Wacklin and Tommy Nylander</i>	
4.1	Phospholipase A ₂ : An Interfacially Activated Enzyme	126
4.1.1	Neutron Reflection	129
4.1.2	Ellipsometry	130
4.1.3	Activity of <i>Naja mossaambica mossaambica</i> PLA ₂	130
4.1.4	Fate of the Reaction Products	133
4.1.5	The Lag Phase and Activation of Pancreatic PLA ₂	135
4.1.6	Distribution of Products during the Lag Phase	138

4.1.7	Hydrolysis of DPPC by Pancreatic PLA ₂	139
4.1.8	Role of the Reaction Products in PLA ₂ Activation	141
4.1.9	Effect of pH and Activation by Me- β -cyclodextrin	144
4.2	Other Lipolytic Enzyme Reactions on Surfaces	150
4.2.1	Triacylglycerol Lipases and the Role of Lipid Liquid Crystalline Nanostructures	150
4.3	Cellulase Enzymes	154
4.4	Conclusion	158
5	Folding Dynamics and Structural Basis of the Enzyme Mechanism of Ubiquitin C-Terminal Hydrolyases	167
	<i>Shang-Te Danny Hsu</i>	
5.1	Introduction	169
5.1.1	UCH-L1	171
5.1.1.1	Genetic association between UCH-L1 and neurodegenerative diseases	171
5.1.1.2	UCH-L1 in oncogenesis	175
5.1.2	Molecular Insights into the Pathogenesis Associated with UCH-L1	175
5.1.3	UCHL3	177
5.1.4	UCHL5	178
5.1.5	BAP1	179
5.2	UCH Structures	180
5.3	Folding Dynamics and Kinetics	183
5.4	Substrate Recognition	184
5.5	Enzyme Mechanism	186
5.6	Conclusion	189
6	Stabilization of Enzymes by Metal Binding: Structures of Two Alkalophilic <i>Bacillus</i> Subtilases and Analysis of the Second Metal-Binding Site of the Subtilase Family	203
	<i>Jan Dohnalek, Katherine E. McAuley, Andrzej M. Brzozowski, Peter R. Østergaard, Allan Svendsen, and Keith S. Wilson</i>	
6.1	Introduction: Subtilases and Metal Binding	203
6.1.1	Calcium-Binding Sites in <i>Bacillus</i> : Proposal for a Standard Nomenclature	209

6.1.2	The Weak Metal-Binding Site	214
6.2	Two New Structures of Subtilases with Altered Calcium Sites	216
6.2.1	Proteinase SubTY	216
6.2.1.1	The overall fold	216
6.2.1.2	The active site	216
6.2.1.3	SubTY calcium and sodium sites	218
6.2.1.4	SubTY disulfide bridge	219
6.2.2	SubHal	220
6.2.2.1	The unliganded form of SubHal	220
6.2.2.2	The SubHal:CI2A complex	221
6.2.2.3	Termini, surface, and pH stability of SubHal	221
6.2.2.4	The two crystallographically independent SubHal:CI2A complexes	223
6.2.2.5	The calcium sites in SubHal	224
6.2.2.6	The active site of SubHal	226
6.2.3	Enzymatic Activity of SubTY and SubHal	228
6.2.4	Comparison of SubTY and SubHal with Other Subtilases	228
6.2.5	The SubHal C-domain Compared to the Eukaryotic PCs, Furin and Kexin	232
6.2.5.1	Active site comparison	233
6.2.5.2	The specificity pockets	234
6.2.5.3	Inhibitor CI2A binding	234
6.2.6	Activity Profiles	236
6.2.7	Comparison of Metal Binding at the Strong and Weak Sites in the S8 Family	236
6.2.8	The Ca-II and Na-II Metal-Binding Sites	237
6.3	Conclusion: Implications for Structural Studies of Enzymes	248
6.4	Materials and Methods	249
6.4.1	SubTY	249
6.4.1.1	Protein production and purification	249
6.4.1.2	Purification of the SubTY:CI2A (1:1) complex	250
6.4.1.3	Crystallization	250
6.4.1.4	Structure determination	251

6.4.2	SubHal	251
6.4.2.1	Protein production and purification	251
6.4.2.2	Purification of the SubHal:CI2A (1:1) complex	252
6.4.2.3	Crystallization	252
6.4.2.4	Structure determination	253
6.4.3	Protease Assays	256
6.4.4	pH Stability	257
6.4.5	Data Deposition	257
7	Structure and Functional Roles of Surface Binding Sites in Amylolytic Enzymes	267
	<i>Darrell Cockburn and Birte Svensson</i>	
7.1	Introduction	267
7.2	Identification of SBSs: X-Ray Crystallography	271
7.3	Bioinformatics of SBS Enzymes	273
7.4	Binding Site Isolation	275
7.5	Protection of Binding Sites from Chemical Labeling	277
7.6	Nuclear Magnetic Resonance	277
7.7	Binding Assays	278
7.8	Activity Assays	282
7.9	Future Prospects	283
7.10	Conclusion	286
8	Interfacial Enzymes and Their Interactions with Surfaces: Molecular Simulation Studies	297
	<i>Nathalie Willems, Mickaël Lelimousin, Heidi Koldsø, and Mark S. P. Sansom</i>	
8.1	Introduction	297
8.2	Enzyme Interactions at Interfaces	299
8.3	Molecular Dynamic Simulations of Biomolecular Systems	301
8.4	Lipases	303
8.4.1	Atomistic MD Studies of Lipase Interactions with Interfaces	304
8.4.2	The Role of Water in Lipase Catalysis at Interfaces	307

8.5	Coarse-Grained MD Studies of Interfacial Enzymes: Orientation and Interactions	309
8.5.1	Phospholipase A2	309
8.5.2	PTEN	310
8.6	Conclusions	311

PART II ENZYME DESIGN

9	Sequence, Structure, Function: What We Learn from Analyzing Protein Families	321
	<i>Michael Widmann and Jürgen Pleiss</i>	
9.1	Introduction	321
9.2	Detection of Inconsistencies Utilizing a Standard Numbering Scheme	323
9.3	Identification of Functionally Relevant Positions	327
9.4	The Modular Structure of Thiamine Diphosphate-Dependent Decarboxylases	330
9.5	Stereoselectivity-Determining Positions: The S-Pocket Concept in Thiamine Diphosphate-Dependent Decarboxylases	333
9.6	Regioselectivity-Determining Positions: Design of Smart Cytochrome P450 Monooxygenase Libraries	336
9.7	Substrate Specificity-Determining Positions: The GX/GGGX Motif in Lipases	340
9.8	Conclusion	341
10	Bioinformatic Analysis of Protein Families to Select Function-Related Variable Positions	351
	<i>Dmitry Suplatov, Evgeny Kirilin, and Vytas Švedas</i>	
10.1	Introduction	352
10.2	Bioinformatic Analysis of Evolutionary Information to Identify Function-Related Variable Positions	359
10.2.1	Problem Definition	359
10.2.2	Scoring Schemes in the Variable Position Selection: High-Entropy, Subfamily-Specific, and Co-Evolving Positions	361
10.2.3	Association of the Variable Positions with Functional Subfamilies	366

10.2.4	How to Select Functionally Important Positions as Hotspots for Further Evaluation: Implementation of Statistical Analysis	366
10.3	The Bioinformatic Analysis of Diverse Protein Superfamilies	369
10.3.1	Bioinformatic Challenges at Studying Enzymes	369
10.3.2	Zebra: A New Algorithm to Select Functionally Important Subfamily-Specific Positions from Sequence and Structural Data	370
10.4	Subfamily-Specific Positions as a Tool for Enzyme Engineering	375
10.5	Conclusion	377
11	Decoding Life Secrets in Sequences by Chemicals	387
	<i>Zizhang Zhang</i>	
11.1	Introduction	388
11.2	Linking an Enzyme's Activity to Its Sequence	389
11.3	Refining the Sequence Space to a Specific Function by Directed Evolution	395
11.4	Linking Chemistry to -Omics with High-Throughput Screening Methods	398
11.5	Finding Large Sequence Space of a Specific Function from Microbial Diversity	400
11.6	Linking Sequences to Substromes at the Molecular Level	404
11.6.1	Biocatalytic Study of EHs	405
11.6.2	Pharmacological Study of EHs	407
11.6.3	Mechanistic Study of EHs	407
11.6.4	What We Have Learned from the Studies of EH	410
11.6.5	Technologies with Potentials in Genochemistry Approach	410
11.7	Correlating with Computational Methods	410
11.8	Problems That Genochemistry Can Potentially Tackle	413
11.9	Conclusion	414

12	Role of Tunnels and Gates in Enzymatic Catalysis	421
	<i>Sérgio M. Marques, Jan Brezovsky, and Jiri Damborsky</i>	
12.1	Introduction	421
12.2	Protein Tunnels	423
12.2.1	Structural Basis and Function	423
12.2.2	Identification Methods	427
12.2.3	Molecular Engineering	429
12.3	Protein Gates	431
12.3.1	Structural Basis and Function	431
12.3.2	Identification Methods	437
12.3.3	Molecular Engineering	440
12.4	Conclusions	442
13	Molecular Descriptors for the Structural Analysis of Enzyme Active Sites	465
	<i>Valerio Ferrario, Lydia Siragusa, Cynthia Ebert, Gabriele Cruciani, and Lucia Gardossia</i>	
13.1	Introduction: Molecular Descriptors for Investigation of Enzyme Catalysis	465
13.2	Molecular Descriptors Based on Molecular Interaction Fields	467
13.3	Multivariate Statistical Analysis for Processing and Interpretation of Molecular Descriptors	472
13.4	Grind Descriptors for the Study of Substrate Specificity	475
13.5	VolSurf Descriptors for the Modeling of Substrate Specificity	477
13.6	Differential MIF _s Descriptors for the Study of Enantioselectivity	479
13.7	Hybrid MIF _s Descriptors for the Computation of Entropic Contribution to Enantiodiscrimination	481
13.8	Analysis of Enzyme Active Sites for Rational Enzyme Engineering	484
13.9	BioGPS Descriptors for in silico Rational Design and Screening of Enzymes	489
13.10	Conclusions	495