
Introduction to Protein Structure

Carl Branden

Uppsala Biomedical Center
Swedish University of Agricultural Sciences
Uppsala, Sweden

John Tooze

European Molecular Biology Organization
Heidelberg, Germany

Garland Publishing, Inc.
New York and London, 1991

THE COVER

Front: The background photograph of the cover is of a Laue x-ray diffraction pattern produced by a crystal of the plant enzyme ribulose biphosphate carboxylase. This technique is described in Chapter 17. Information derived from such x-ray patterns, together with a knowledge of the amino acid sequence, enabled the three-dimensional arrangement of atoms in the protein to be determined. A simplified representation of this protein structure is shown in color, superimposed on the diffraction pattern. The enzyme, which is involved in the fixation of carbon dioxide, is a member of the large class of α/β barrel protein structures. This class of structures is discussed in detail in Chapter 4.

Back: Tomato bushy stunt virus is a spherical virus made from 180 protein subunits. Arms extending from sixty of these subunits contribute to an internal framework that determines the size of the correctly assembled virus particle. The interdigitated arms from three subunits meet at each of the twenty icosahedral threefold axes of the virus. One such axis is shown here with the β strands from three subunits shown in different shades of green. Virus structure is described in more detail in Chapter 11.

© 1991 Carl Branden and John Tooze

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Branden, Carl.

Introduction to protein structure / Carl Branden, John Tooze.

p. cm.

Includes index.

ISBN 0-8153-0344-0 — ISBN 0-8153-0270-3 (pbk.)

1. Proteins—Structure. I. Tooze, John. II. Title.

QP551.B7635 1991

574.19'245—dc20

91-11788

CIP

Published by Garland Publishing, Inc.

136 Madison Ave., New York, New York, 10016

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4

E 5 a / 83



94/308

Contents

Protein Structure			
Part 1 Basic Structural Principles	1		
1. The Building Blocks	3		
Proteins are polypeptide chains	4		
The genetic code specifies 20 different amino acid side chains	4		
Cysteines can form disulfide bridges	5		
Peptide units are building blocks of protein structures	8		
Glycine residues can adopt many different conformations	8		
Conclusion	9		
Selected Readings	10		
2. Motifs of Protein Structure	11		
Few general principles emerged from the first protein structure	11		
The interior of proteins is hydrophobic	12		
The alpha (α) helix is an important element of secondary structure	12		
The α helix has a dipole moment	13		
Some amino acids are preferred in α helices	13		
Beta (β) sheets usually have their beta (β) strands either parallel or antiparallel	15		
Loop regions are at the surface of protein molecules	18		
Schematic pictures of proteins highlight secondary structure	19		
Topology diagrams are useful for classification of protein structures	20		
Secondary structure elements are connected into simple motifs	21		
The hairpin β motif occurs frequently in protein structures	23		
The Greek key motif is found in antiparallel β sheets	24		
The β - α - β motif contains two parallel β strands	24		
Protein molecules are organized in a structural hierarchy		26	
Large polypeptide chains fold into several domains		26	
Domains are built from structural motifs		27	
Simple motifs combine to form complex motifs		28	
Protein structures can be divided into three main classes		28	
Conclusion		29	
Selected Readings		30	
3. Alpha-Domain Structures	33		
Two adjacent α helices are usually antiparallel		33	
The four-helix bundle is a common domain structure in α proteins		33	
The active site is between the α helices in four-helix bundle structures		35	
The globin fold is present in myoglobin and hemoglobin		35	
Geometric aspects determine α -helix packing		36	
Ridges of one α helix fit into grooves of an adjacent helix		36	
The globin fold has been preserved during evolution		37	
The hydrophobic interior is preserved		38	
Helix movements accommodate interior side-chain mutations		38	
Sickle-cell hemoglobin confers resistance to malaria		39	
Conclusion		40	
Selected Readings		41	
4. Alpha/Beta Structures	43		
Parallel β strands are arranged in barrels or sheets		43	
α/β barrels occur in many different enzymes		44	
Branched hydrophobic side chains dominate the core of α/β barrels		45	

Pyruvate kinase contains several domains, one of which is an α/β barrel	46	The receptor binding site is formed by the jelly roll domain	74
Double barrels have occurred by gene fusion	47	Hemagglutinin acts as a membrane fusogen	74
The active site is formed by loops at one end of the α/β barrel	47	Conclusion	74
Stability and function are separated	48	Selected Readings	76
Was there an ancestral barrel?	48	6. DNA Structures	79
α/β twisted open-sheet structures contain α helices on both sides of the β sheet	49	The DNA double helix is different in A- and B-DNA	79
Open β -sheet structures have different topologies	49	The DNA helix has major and minor grooves	80
The position of the active sites can be predicted in α/β structures	51	Z-DNA forms a zigzag pattern	81
Tyrosyl-tRNA synthetase has two different domains ($\alpha/\beta + \alpha$)	51	B-DNA is the preferred conformation <i>in vivo</i>	81
Carboxypeptidase is an α/β protein with mixed β sheet	53	Specific base sequences can be recognized in B-DNA	82
Arabinose-binding protein has two similar α/β domains	54	Conclusion	83
Conclusion	56	Selected Readings	84
Selected Readings	56	Part 2 Structure, Function, and Engineering	85
5. Antiparallel Beta Structures	59	7. DNA Recognition by Proteins with the Helix-Turn-Helix Motif	87
Up-and-down barrels have a simple topology	60	A molecular mechanism for gene control is emerging	87
Retinol-binding protein folds into an up-and-down β barrel	60	Repressor and Cro proteins operate a procaryotic genetic switch region	88
Retinol is bound inside the β barrel	61	The x-ray structure of the complete lambda Cro protein is known	89
Amino acid sequence reflects β structure	62	The x-ray structure of the DNA-binding domain of the lambda repressor is known	90
The retinol-binding protein belongs to a superfamily of protein structures	62	Both lambda Cro and repressor proteins have a specific DNA-binding motif	91
Retinol binding in humans and biliverdin binding in insects show evolutionary relationship	62	Model building predicts Cro-DNA interactions	92
Structure suggests function for β -lactoglobulin	63	Genetic studies agree with the structural model	93
Neuraminidase folds into up-and-down β sheets	64	The x-ray structure of DNA complexes with 434 Cro and repressor revealed novel features of protein DNA interactions	94
Folding motifs form a superbarrel in neuraminidase	64	The structures of 434 Cro and 434 repressor-binding domains are very similar	95
The active site is at one end of the superbarrel	65	The B-DNA conformation is distorted in the complexes	95
Greek key motifs occur frequently in the antiparallel β structures	66	Conformational changes of DNA are important for differential repressor binding	96
The γ -crystallin molecule has two domains	67	Sequence-specific protein-DNA interactions recognize operator regions	96
The domain structure has a simple topology	68	Nonspecific protein-DNA interactions determine DNA conformation	97
Two Greek key motifs form the domain	68	Local DNA structure modulates repressor binding	98
The two domains have identical topology	69	The essence of phage repressor and Cro	99
The two domains have similar structures	69	Amino acid sequence relations identify helix-turn-helix motifs	99
The Greek key motifs are evolutionarily related in γ crystallin	70	Phage Cro and repressor proteins are evolutionarily related	100
Intron positions separate the four Greek key motifs	70	Cro and repressor have homologous sequences but partly different structures	101
The Greek key motifs can form jelly roll barrels	70	Sequence comparison using strong stereochemical constraints identifies helix-turn-helix motifs	101
The jelly roll motif is wrapped around a barrel	71	Sequence comparisons without using stereochemical constraints do not identify helix-turn-helix motifs unambiguously	103
The jelly roll barrel is usually divided into two sheets	71	DNA binding is regulated by allosteric control	104
A folding scheme has been suggested for the jelly roll barrel structure	72		
The hemagglutinin polypeptide chain folds into a complex structure	72		
The subunit structure is divided into a stem and a tip	73		
The hemagglutinin molecule is trimeric	73		

The <i>trp</i> repressor forms a helix-turn-helix motif	104	NAD binds in a similar way to each domain	146
A conformational change provides the molecular mechanism of the functional switch	105	Hydride transfer to NAD is stereospecific	147
CAP is a positive control element	106	Are the NAD-binding domains evolutionarily related?	148
The polypeptide chain of CAP folds into two domains	106	The NAD-binding motif can be predicted from amino acid sequence	148
A molecular mechanism has been suggested for the switch in CAP	107	FAD- and NAD-binding domains have essential similarities	151
<i>Met</i> and <i>arc</i> repressors belong to a family of β -sheet DNA-binding proteins	108	Gene fusion has occurred between an FMN-binding α/β barrel and a cytochrome	152
Conclusion	109	The cytochrome- b_2 domain is homologous to mammalian cytochrome b_5	153
Selected Readings	110	The FMN-binding domain is an α/β barrel	154
8. Structural Motifs of Eucaryotic Transcription Factors	113	Hexokinase validates the theory of induced fit	155
Transcription factors have two functionally different domains	114	Conclusion	157
Three different families of zinc fingers have been observed	115	Selected Readings	158
The classic zinc finger has two cysteine and two histidine ligands bound to zinc	116	11. The Structure of Spherical Viruses	161
Two zinc fingers in the glucocorticoid receptor form one DNA-binding domain	118	The protein shells of spherical viruses have icosahedral symmetry	162
Yeast transcription factor Gal 4 contains a binuclear zinc cluster in its DNA-binding domain	119	The icosahedron has high symmetry	163
Retroviral zinc fingers have one histidine and three cysteine residues bound to zinc	120	The simplest virus has a shell of 60 protein subunits	164
Monomers of homeodomains bind to DNA through a helix-turn-helix motif	121	Complex spherical viruses have more than one polypeptide chain in the asymmetric unit	165
Leucine zippers provide dimerization interactions for some eucaryotic DNA-binding proteins	124	Structural versatility gives quasi-equivalent packing in T = 3 plant viruses	166
Conclusion	126	The protein capsid of picorna viruses contains four polypeptide chains	167
Selected Readings	127	There are four different structural proteins in picorna viruses	168
9. DNA Polymerase Is a Multifunctional Enzyme	129	The arrangement of subunits in the shell of picorna viruses is similar to that of T = 3 plant viruses	168
The Klenow fragment of <i>E. coli</i> DNA polymerase I (Pol I) can be crystallized	131	The coat proteins of spherical plant and animal viruses have similar structure, the jelly roll barrel structure, indicating an evolutionary relationship	169
The Klenow fragment has two separate domains	131	Drugs against common cold may be designed from the structure of rhinovirus	171
The large domain has a large binding cleft	132	Bacteriophage MS2 has a different subunit structure	173
The large domain has polymerase activity	132	Both the core and the spikes of enveloped viruses have icosahedral symmetry	174
The small domain has an α/β structure	133	The subunits in polyoma virus have nonequivalent environments	174
The small domain has 3'-5' exonuclease activity	134	Conclusion	175
How the processivity and fidelity of DNA synthesis is achieved	135	Selected Readings	176
Comparison with other polymerases reveals evolutionary relationships	136	12. Recognition of Foreign Molecules by the Immune System	179
DNA-binding proteins are constructed from modules	137	The polypeptide chains of antibodies are divided into domains	181
Conclusion	138	Antibody diversity is generated by several different mechanisms	182
Selected Readings	139	All immunoglobulin domains have similar three-dimensional structure	183
10. Enzymes That Bind Nucleotides	141	The immunoglobulin fold is best described as two antiparallel β sheets packed tightly against each other	184
The structures of several NAD-dependent dehydrogenases are known	142		
The dehydrogenase polypeptide chains are modular	143		
The NAD-binding domains have similar structures	144		

The hypervariable regions are clustered in loop regions at one end of the variable domain	185	The receptors for insulin and epidermal growth factor are evolutionarily related	220
The antigen binding site is formed by close association of the hypervariable regions from both heavy and light chains	186	The PDGF receptor is also a protein tyrosine kinase receptor	221
The antigen binding site binds haptens in crevices and protein antigens through large flat surfaces	188	Similar mechanisms are used for signal transduction across the membrane	221
The structure of an idiotype-anti-idiotype complex has been determined	192	G proteins are molecular amplifiers	222
An IgG molecule has several degrees of conformational flexibility	193	G proteins are homologous in sequence to elongation factor Tu and cH-ras p21	223
The structure of a human MHC molecule has provided insights into the molecular mechanism of T-cell activation	194	Point mutation generates transforming <i>ras</i> oncogenes	223
Recognition of antigen is different in MHC compared to immunoglobulins	194	A deletion mutant gene is used to produce truncated p21	224
Conclusion	197	The crystal structures of p21 and EF-Tu have the same fold of their polypeptide chains	224
Selected Readings	199	Regions of conserved amino acid sequence bind GTP	225
13. Membrane Proteins	201	Oncogenic activation is caused by mutations in the GTP-binding loops	226
Membrane proteins are difficult to crystallize	202	The molecular basis of autophosphorylation of viral p21	226
Bacteriorhodopsin contains seven transmembrane α helices	203	Protein-protein interactions can modulate the GTP-binding and hydrolysis properties of <i>ras</i> p21	227
The bacterial photosynthetic reaction center is built up from four different polypeptide chains and many pigments	203	Receptors that utilize G proteins contain seven transmembrane helices	227
The L, M, and H subunits have transmembrane α helices	205	Conclusion	228
The photosynthetic pigments are bound to the L and M subunits	206	Selected Readings	229
Reaction centers convert light energy into electrical energy by electron flow through the membrane	208	15. An Example of Enzyme Catalysis: Serine Proteinases	231
The reaction center is a quantum-mechanical tunneling device	209	Proteinases form four functional families	231
Transmembrane α helices can be predicted from the amino acid sequence	209	The catalytic properties of enzymes are reflected in K_m and k_{cat} values	232
Hydrophobicity scales measure the degree of hydrophobicity of different amino acid side chains	210	Enzymes decrease the activation energy of chemical reactions	232
Hydropathy plots identify transmembrane helices	210	Serine proteinases cleave peptide bonds by forming tetrahedral transition states	234
Reaction-center hydropathy plots agree with crystal structural data	211	Four essential structural features are required for the catalytic action of serine proteinases	235
Membrane lipids have no specific interaction with protein transmembrane α helices	212	Convergent evolution has produced two different serine proteinases with similar catalytic mechanisms	236
Structural rearrangements convert a water-soluble protein to a membrane-bound form	212	The chymotrypsin structure has two antiparallel β -barrel domains	236
Conclusion	213	The active site is formed by two loop regions from each domain	237
Selected Readings	214	Did the chymotrypsin molecule evolve by gene duplication?	238
14. Receptor Families	217	Different side chains in the substrate specificity pocket confer preferential cleavage	238
Tyrosine kinase growth factor receptors and G-protein linked receptors form two different receptor families	217	Engineered mutations in the substrate specificity pocket change the rate of catalysis	239
The epidermal growth factor (EGF) receptor folds into distinct domains	218	The Asp 189-Lys mutant in trypsin has unexpected changes in substrate specificity	241
The <i>v-erb B</i> oncogene is a coopted EGF receptor gene	219	The structure of the serine proteinase subtilisin is of the α/β type	241
		The active sites of subtilisin and chymotrypsin are similar	241

A structural anomaly in subtilisin has functional consequences	243	Hydrogen bonds give small energy contributions to ligand binding	261
Transition-state stabilization in subtilisin is dissected by protein engineering	243	Hydrogen bonds involving charged groups contribute more to specificity than those between uncharged groups	262
Catalysis occurs without a catalytic triad	243	Circularly permuted α/β barrels fold correctly	263
Substrate molecules provide catalytic groups in substrate-assisted catalysis	243	Protein structures can be designed from first principles	265
Conclusion	245	Conclusion	266
Selected Readings	245	Selected Readings	267
16. Prediction, Engineering, and Design of Protein Structures	247	17. Determination of Protein Structures	269
Prediction of protein structure from sequence is an unsolved problem	248	Several different techniques are used to study the structure of protein molecules	269
Many different amino acid sequences give similar three-dimensional structures	248	Protein crystals are difficult to grow	270
Amino acid sequence homology implies similarity in structure and functions	249	X-ray sources are either monochromatic or polychromatic	272
Homologous proteins have conserved structural cores and variable loop regions	249	X-ray data are recorded either on films or by electronic detectors	273
Knowledge of secondary structure is necessary for prediction of tertiary structure	251	The rules for diffraction are given by Bragg's law	273
Prediction methods for secondary structure have low accuracy	251	Phase determination is the major crystallographic problem	274
The tertiary fold of an enzyme has been successfully predicted	252	Building a model involves subjective interpretation of the data	276
Folded and unfolded proteins are almost equally stable	256	Errors in the initial model are removed by refinement	277
Proteins can be made more stable by engineering	257	Amino acid sequence is essential for x-ray structure determination	278
Disulfide bridges increase protein stability	257	Recent technological advances have greatly influenced protein crystallography	279
Glycine and proline have opposite effects on stability	259	NMR methods use the magnetic properties of atomic nuclei	280
Stabilizing the dipoles of α helices increases stability	259	Two-dimensional NMR spectra of proteins are interpreted by the method of sequential assignment	281
Mutants that fill cavities in hydrophobic cores do not stabilize T4 lysozyme	260	Distance constraints are used to derive possible structures of a protein molecule	283
Binding energy in molecular recognition and specificity has been analyzed by protein engineering	260	Biochemical studies and molecular structure give complementary functional information	284
		Conclusion	284
		Selected Readings	285