Advances in Experimental Medicine and Biology 747

Jacqueline M. Matthews *Editor*

Protein Dimerization and Oligomerization in Biology

LANDES BIOSCIENCE

Springer

Protein Dimerization and Oligomerization in Biology

ADVANCES IN EXPERIMENTAL MEDICINE AND BIOLOGY

Recent Volumes in this Series

Volume 739
SENSING IN NATURE
Carlos López-Larrea

Volume 740
CALCIUM SIGNALING
Md. Shahidul Islam

Volume 741
STEM CELL TRANSPLANTATION
Carlos López-Larrea, Antonio López Vázquez and Beatriz Suárez Álvarez

Volume 742
ADVANCES IN MITROCHONDRIAL MEDICINE
Roberto Scatena

Volume 743
HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) AND BREASTFEEDING
Athena Kourtis and Marc Bulterys

Volume 744
RAMPs
William S. Spielman and Narayanan Parameswaran

Volume 745
NEW TECHNOLOGIES FOR TOXICITY TESTING
Michael Balls, Robert D. Combes and Nirmala Bhogal

Volume 746
GLIOMA: IMMUNOTHERAPEUTIC APPROACHES
Ryuya Yamanaka

Volume 747
PROTEIN DIMERIZATION AND OLIGOMERIZATION IN BIOLOGY
Jacqueline M. Matthews

Protein Dimerization and Oligomerization in Biology

Edited by

Jacqueline M. Matthews, PhD
School of Molecular Bioscience, University of Sydney, Sydney, Australia

Springer Science+Business Media, LLC

Landes Bioscience

Springer Science+Business Media, LLC
Landes Bioscience

Springer Science+Business Media, LLC, 233 Spring Street, New York, New York 10013, USA
http://www.springer.com

Please address all inquiries to the publishers:
Landes Bioscience, 1806 Rio Grande, Austin, Texas 78701, USA
Phone: 512/ 637 6050; FAX: 512/ 637 6079
http://www.landesbioscience.com

The chapters in this book are available in the Madame Curie Bioscience Database.
http://www.landesbioscience.com/curie

Protein Dimerization and Oligomerization in Biology, edited by Jacqueline M. Matthews. Landes Bioscience / Springer Science+Business Media, LLC dual imprint / Springer series: Advances in Experimental Medicine and Biology.

ISBN: 978-1-4614-3228-9

Library of Congress Cataloging-in-Publication Data

Protein dimerization and oligomerization in biology / edited by Jacqueline M. Matthews.
p. ; cm. -- (Advances in experimental medicine and biology ; v. 747)
Includes bibliographical references and index.
Summary: "This volume has a strong focus on homo-oligomerization, which is surprisingly common. However, protein function is so often linked to both homo- and hetero-oligomerization and many heterologous interactions likely evolved from homologous interaction, so this volume also covers many aspects of hetero-oligomerization"--Provided by publisher.
ISBN 978-1-4614-3228-9 (hardback)
I. Matthews, Jacqueline M. II. Series: Advances in experimental medicine and biology ; v. 747. 0065-2598
[DNLM: 1. Protein Multimerization. 2. Protein Folding. W1 AD559 v.747 2012 / QU 55.9]
LC classification not assigned
572'.633--dc23

2012005555

PREFACE

Proteins do not act in isolation. They interact with lipids, nucleic acids, carbohydrates, small molecules and ions. And of course they interact with proteins—either like proteins (self-association/homo-oligomerization) or different proteins (heterologous association/ hetero-oligomerization). Protein-protein interactions lie at the heart of essentially all biological processes and large-scale efforts to map and characterize protein-protein interaction networks have formed a major research focus in the post-genomic era. This volume has a strong focus on homo-oligomerization, which is surprisingly common. However, protein function is so often linked to both homo- and hetero-oligomerization and many heterologous interactions likely evolved from homologous interaction, so this volume also covers many aspects of hetero-oligomerization.

Chapter 1, by Matthews and Sunde, is a fairly general overview of protein dimerization and oligomerization, covering the prevalence of homodimers and higher-order oligomers of well characterized proteins, possible origins of self-association, and some of the many functional advantages conferred by homodimers and higher order oligomers.

Traditionally, "dimerization" refers to the coming together of two similar subunits, but is often used more loosely to refer to any type of protein association—often because the stoichiometry of association is unknown. In Chapter 2, Gell, Grant and Mackay outline many of the key experimental approaches that can be used to detect protein-protein interactions and characterise the nature of protein dimerization and oligomerization. In Chapter 3, Jones describes what is known about protein association from analysis of structures, and how this information can be harnessed to predict and further analyze protein dimers and oligomers.

Enzymes form one of the best characterised class of proteins, and one in which homo-oligomerization is particularly prevalent. In Chapter 4, Mackenzie and Clarke describe the caspase system, which provides many examples of the different ways in which enzyme activity can be regulated by protein oligomerization. In Chapter 5, Griffin and Gerrard focus on the relationships between oligomeric state and enzyme function, including engineering approaches in which manipulation of oligomeric state has been used to regulate function.

Interactions between proteins and nucleic acids are essential to many aspects of cell function. In Chapter 6, Wilce, Vivian and Wilce provide a comprehensive overview of the contributions of protein dimer and oligomer formation to nucleic acid binding, while in Chapter 7, Funnell and Crossley focus on the roles that protein homo- and hetero-oligomers play in the regulation of transcription.

Many membrane channel proteins form oligomers, and in Chapter 8, Clarke and Gulbis describe, using potassium channels as an example, the intimate relationships between oligomerization and ion channel function.

One interesting mode of protein oligomerization is domain swapping—the exchange of elements of structure between like subunits. In Chapter 9, Rousseau, Schymkowitz and Itzhaki explain the implications of domain swapping in for protein folding and function, and how the same phenomenon may be involved in misfolding events.

Finally, in Chapter 10, Itzhaki and Lowe provide an overview of repeat proteins, pseudo-multimeric proteins that keep their subunits firmly in place by effectively positioning subunits on the same polypeptide chain.

Jacqueline M. Matthews, PhD

ABOUT THE EDITOR...

JACQUELINE (JACQUI) M. MATTHEWS is currently a Senior Research Fellow of the National Health and Medical Research Council of Australia, and the Professor of Protein Chemistry at the University of Sydney. Her research focuses on regulatory proteins involved in development and disease, and in particular on protein-protein and protein-DNA interactions within transcription factor complexes. Dr. Matthews received her undergraduate training (BSc Hons) in chemistry and biochemistry at the University of New South Wales in Sydney, Australia, and a PhD in biological chemistry (for work on protein folding under the supervision of Sir Professor Alan Fersht) from the University of Cambridge, UK. She is currently President of the Sydney Protein Group, a member of the Executive Council of the Protein Society and a member of the Australian Society for Biochemistry and Molecular Biology, and Australian Biophysical Society.

PARTICIPANTS

A. Clay Clark
Department of Molecular
and Structural Biochemistry
North Carolina State University
Raleigh, North Carolina
USA

Oliver B. Clarke
The Walter and Eliza Hall Institute
of Medical Research
and
Department of Medical Biology
The University of Melbourne
Melbourne, Victoria
Australia

Merlin Crossley
School of Biotechnology
and Biomolecular Sciences
University of New South Wales
and
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
Australia

Alister P. W. Funnell
School of Biotechnology
and Biomolecular Sciences
University of New South Wales
and
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
Australia

David A. Gell
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
and
Menzies Research Institute
University of Tasmania
Hobart, Tasmania
Australia

Juliet A. Gerrard
Biomolecular Interaction Centre (BIC)
School of Biological Sciences
University of Canterbury
Christchurch
New Zealand

Richard P. Grant
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
Australia

Michael D.W. Griffin
Bio21 Institute of Molecular Science
and Biotechnology
University of Melbourne
Melbourne, Victoria
Australia

Jacqueline M. Gulbis
The Walter and Eliza Hall Institute
of Medical Research
and
Department of Medical Biology
The University of Melbourne
Melbourne, Victoria
Australia

Laura S. Itzhaki
University of Cambridge
Department of Chemistry
Cambridge
UK

Susan Jones
Department of Biochemistry
School of Life Sciences
University of Sussex
Brighton
UK

Alan R. Lowe
Department of Physics
University of California-Berkeley
Berkeley, California
USA

Joel P. Mackay
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
Australia

Sarah H. MacKenzie
Department of Molecular
and Structural Biochemistry
North Carolina State University
Raleigh, North Carolina
USA

Jacqueline M. Matthews
School of Molecular Bioscience
University of Sydney
Sydney, New South Wales
Australia

Frederic Rousseau
VIB Switch Laboratory
Department of Molecular Cell Biology
Katholieke Universiteit Leuven
Leuven
Belgium

Joost Schymkowitz
VIB Switch Laboratory
Department of Molecular Cell Biology
Katholieke Universiteit Leuven
Leuven
Belgium

Margaret Sunde
School of Molecular Bioscience
and
Discipline of Pharmacology
University of Sydney
Sydney, New South Wales
Australia

Julian Vivian
Department of Biochemistry
and Molecular Biology
and
School of Biomedical Sciences
Monash Univerity
Melbourne, Victoria
Australia

Jackie Wilce
Department of Biochemistry
and Molecular Biology
and
School of Biomedical Sciences
Monash Univerity
Melbourne, Victoria
Australia

Matthew Wilce
Department of Biochemistry
and Molecular Biology
and
School of Biomedical Sciences
Monash Univerity
Melbourne, Victoria
Australia

CONTENTS

CHAPTER 1

DIMERS, OLIGOMERS, EVERYWHERE

Jacqueline M. Matthews*,[1] and Margaret Sunde[1,2]

[1]*School of Molecular Bioscience, University of Sydney, Sydney, Australia;* [2]*Discipline of Pharmacology, University of Sydney, Sydney, Australia*

**Corresponding Author: Jacqueline M. Matthews—Email: jacqui.matthews@sydney.edu.au*

Abstract: The specific self-association of proteins to form homodimers and higher order oligomers is an extremely common event in biological systems. In this chapter we review the prevalence of protein oligomerization and discuss the likely origins of this phenomenon. We also outline many of the functional advantages conferred by the dimerization or oligomerization of a wide range of different proteins and in a variety of biological roles, that are likely to have placed a selective pressure on biological systems to evolve and maintain homodimerization/oligomerization interfaces.

INTRODUCTION

Proteins rarely act alone. They commonly bind other biomolecules, including other proteins, to generate a biological response. A large percentage of proteins appear to self-associate to form dimers or higher-order oligomers. Dimerization and oligomerization can confer several different structural and functional advantages to proteins, including improved stability, regulation of activity and increased complexity.[1] Here, we consider how the phenomenon of specific self-association of proteins may have arisen and why homodimerization and homo-oligomerization of proteins remain a common feature of biological systems.

THE PREVALENCE OF PROTEIN HOMODIMERS AND HOMO-OLIGOMERS

Many proteins self-associate to form homodimers or higher order homo-oligomers (e.g., Fig. 1).[1] Anecdotally, self-association appears to be very common but it is often

Protein Dimerization and Oligomerization in Biology, edited by Jacqueline M. Matthews.

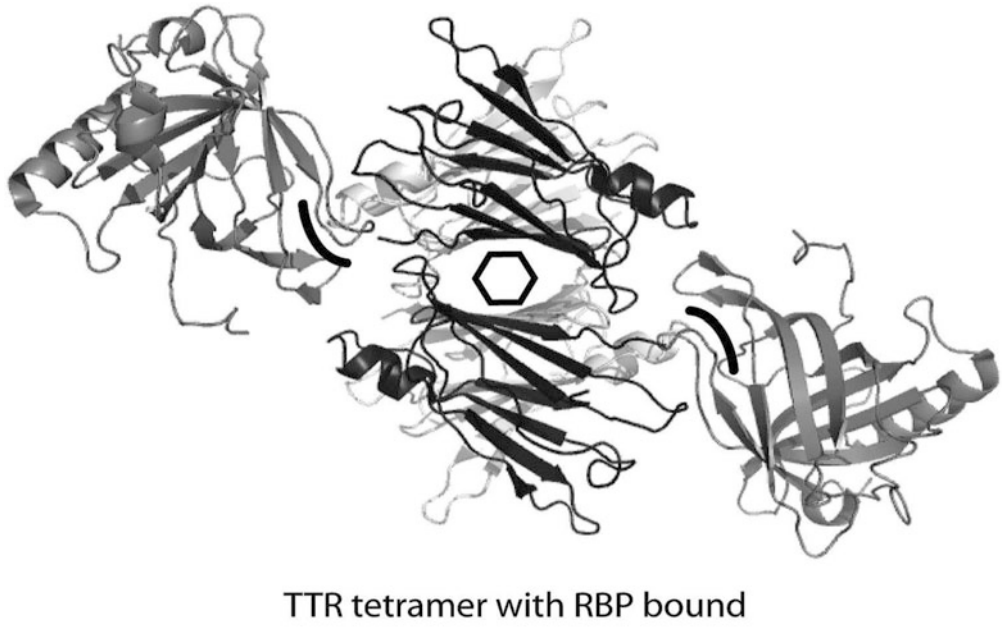

Figure 1. The Transthyretin (TTR):retinol binding protein (RBP) complex consists of a homotetramer and two heterodimers. The association of transthyretin monomers (coloured black and white) generates a homotetramer with a large central channel. The thyroid hormone thyroxine, represented as a black hexamer, binds within this channel. In addition, the tranthyretin tetramer forms a heterocomplex with two molecules of retinol binding protein (coloured in grey), which bind on either side of the tetramer. The binding site for retinol in each RBP molecule is indicated with a black line. (PDB 2WQA).

hard to quantify, partly because the stoichiometry of self-association for many proteins has not been characterized by robust biophysical methods, such as those outlined in the accompanying chapter by Gell, Grant and Mackay. Fortunately, however, several databases exist that are highly populated by very well characterized proteins, allowing us to gauge the prevalence of homodimers and higher order homo-oligomers.

The Brenda enzyme database (http://www.brenda.uni-koeln.de/) contains entries for tens of thousands of enzymes originating from all domains of life, about a third of which (~11,000 entries in July 2011) report a defined subunit composition. For this subset of enzymes, homodimers and higher order homo-oligomers by far outnumber monomers (Table 1). Not all entries have the same type of annotation so the absolute numbers vary according to how the search is carried out, but the proportions of monomer to dimer (and higher order oligomers) are all similar. Overall, monomers comprise about a quarter to a third of enzymes with a defined subunit composition and dimers plus higher order oligomers are at least twice as prevalent as monomers. Of these oligomers, dimers are most prevalent (36–38%) followed by tetramers (19%). Numbers of entries decrease rapidly as the oligomeric subunit number increases and enzymes with an odd number of subunits are less prevalent than those with an even number. Note that hetero-oligomers are a relatively poorly represented group. The database appears to be dominated by enzymes from bacterial species, but the proportions are similar for human enzymes, with the minor exception that dimers appear to be

Table 1. Subunit composition of enzymes

	Enzymes From All Species		Human Enzymes	
	Subunit Name[1]	Number of Like Subunits[2]	Subunit Name[1]	Number of Like Subunits[2]
Monomer	4847 (25%)	3641 (33%)	419 (23%)	270 (35%)
Dimer	7553 (38%)	4000 (36%)	864 (47%)	351 (45%)
Trimer	880 (4%)	382 (3%)	95 (5%)	24 (3%)
Tetramer	3719 (19%)	2075 (19%)	290 (16%)	105 (13%)
Pentamer	88 (0.4%)	42 (0.4%)	2 (0.1%)	2 (0.3%)
Hexamer	845 (4%)	449 (4%)	50 (3%)	19 (2%)
Higher order oligomers	1162 (6%)[3]	415 (4%)[4]	55 (3%)[3]	10 (1%)[4]
Hetero	620 (3%)		67 (4%)	
Total	19714	11004	1842	781

These numbers were generated through the "Search Subunits" module of the Brenda database in July 2011.
1. Subunit names as indicated were entered as the main search function. Note that this will capture both homo and heterooligomers, but heterooligomers appear to comprise a small proportion of entries.
2. The "Number of like subunits" is the output after entering "N *" (where N = 1 for monomer, 2 for dimer etc) in the Commentary window, which for many enzymes lists the number of copies of subunits. This search should exclude most hetero-oligomers, but will also exclude entries for which the subunit composition is not specified in this format (or at all in the Commentary window).
3. Subunit name: heptamer through to tetraeicosamer and poly.
4. Commentary window "N *" where N = 7–24,30,36,48,60.

particularly highly represented, apparently at the expense of higher order oligomers from tetramers upwards.

The Protein Data Bank (PDB) provides a compilation of highly characterized proteins from a much wider variety of different classes, although we note that these data are heavily biased towards soluble proteins and homomeric samples (e.g., note the relatively low number of protein hetero-oligomers; Table 2). Deposited structures are highly represented by monomers, but as at least as many proteins form dimers or higher order oligomers. Note that assignment of the biologically relevant oligomeric state from crystal structures is not trivial. It has been estimated that for 20% of dimers in the PDB the chance of misrepresentation is as high as 50%.[2] Examination of the nature and size of interfaces in crystallized complexes will reflect only the enthalpic component of complex formation and not the entropy loss on formation of the complex. Although weak interactions may be manifest in highly concentrated crystallization conditions, they may also be displaced by crystal packing contacts that result in a more favorable global energy. Several automated analysis procedures have been developed to analyze the complexes observed in crystals but complementary noncrystallographic studies should always be used to support identification of biologically significant macromolecular complexes.

In addition to the databases that report highly characterized oligomers, high-throughput studies of protein-protein interaction networks from eukaryotic organisms indicate a statistical bias towards homo-oligomeric interactions; 25–200 times more homomeric interactions were identified than could be expected if homodimers and higher order homo-oligomers randomly appeared in the course of the evolution.[3]

Table 2. Subunit composition in protein structures. Searches specified only structures that contained proteins [Macomolecule Type: Contains Protein—Yes; other options—Ignore]; and queried the [Number of Chains (Biological Assembly)] option such that monomer refers to 1, dimer to 2 etc. A 95% sequence identity cutoff was used to reduce the numbers of mutant proteins.

	One Protein Entity Only[1]	Any Number of Protein Entities[2]
Monomer	14636	14689
Dimer or larger	16780	21098
Heterooligomers[3]	-	3569
Break down of nonmonomeric structures		
Dimer	9219	10728
Trimer	2052	2561
Tetramer	3274	4151
Pentamer	133	266
Hexamer	1014	1339
Higher order oligomers[4]	1088	2053

1. [Number of Entities: Entity type—Protein; *between* 1 *and* 1 (column 1)].
2. [Number of Entities: Entity type—Protein; *between* 1 *and* 10^6 (column 2)].
3. For heterodimers [Number of Entities: Entity type—Protein; *between* 2 *and* 10^6]/The [Number of Chains (Biological Assembly): *between* 2 *and* 10^6 chains].
4. [Number of chains (Biological Entity): *between* 7 *and* 10^6 chains].

THE EVOLUTION OF PROTEIN DIMERS

Origins of Protein Self-Association

The tendency of many proteins to self-associate is a property well known to structural biologists. Indeed, for many proteins self-association is a major problem at the concentrations required for NMR spectroscopy, X-ray crystallography and techniques such as small angle scattering methodologies (e.g., ref. 4). Modeling of protein-like surfaces show they have a statistically higher affinity for self attraction compared with the propensity for attraction between *different* proteins.[5] These statistical propensities are likely to produce self-self or similar interfaces of very low affinity, but it is reasonable to assume that any such interfaces that confer a functional advantage to an organism could evolve into higher affinity interfaces that mediate specific oligomer formation. Indeed, dimer interfaces have a high degree of conservation in evolutionarily related proteins.[6]

From Simple Homo-Oligomers to Complex Systems

In prokaryotes multi-protein complexes tend to have a simpler composition than in eukayotes. For example, the catalytic core units of proteasomes are made up of two rings of alpha and two rings of beta subunits, with each ring containing seven subunits. In bacteria and archaea there is a single type of alpha and a single type of beta subunit, but in eukaryotes there are seven different types each of alpha or beta subunits (Fig. 2).[7,8]

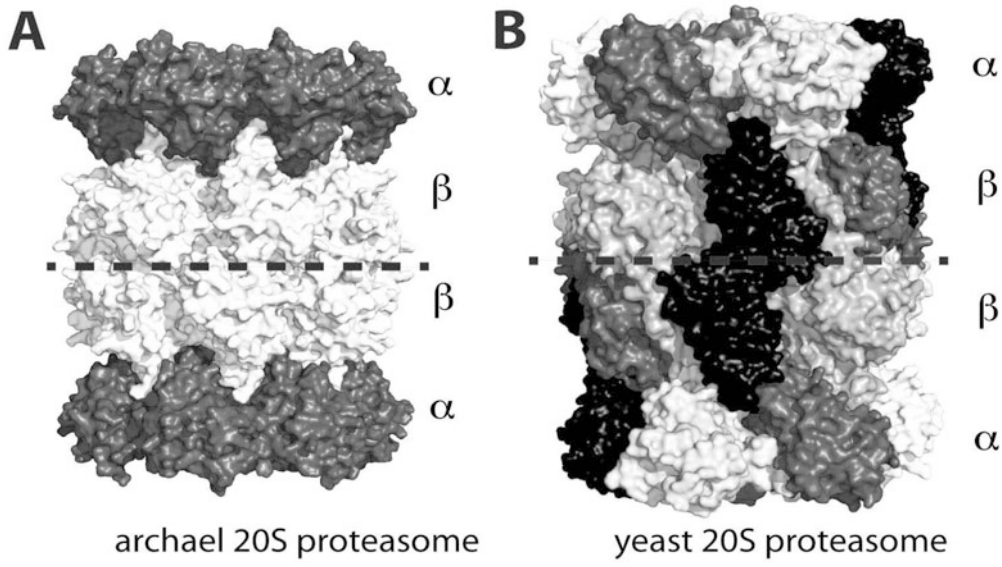

Figure 2. Comparison of the archael (*Thermoplasma acidophilum*) and yeast (*Saccharomyces cerevisiae*) 20S proteosome structures. A) The core units of the archaeal proteosome consist of two rings of alpha (α) and two rings of beta (β) subunits, with each ring containing seven subunits (alpha and beta subunits coloured white and black, respectively). B) In eukaryotes there are seven different types of alpha and seven different types of beta subunits (coloured in different shades of grey). (PDB 3IPM and 3NZJ). Black dashed lines demarcate the two β rings.

Similarly, proteasome-associated AAA ATPases tend to be homo-hexamers in bacteria and archaea and hetero-hexamers in eukaryotes. Proteins that regulate gene expression in prokaryotes are often homodimers or oligomers, but in eukaryotes, processes that regulate gene expression appear to rely heavily on the formation of multiprotein complexes.[9] The expanded sizes of eukaryotic genomes compared to prokaryotic genomes appear to have been caused in part by genome duplication events. On an evolutionary timescale identical copies of genes gradually diverge in sequence and function to form paralogs and through additional genome duplication events become families of related proteins.[10] Thus, homodimeric proteins could evolve into so called "superfamily heterodimers", families of related proteins that can form homomeric and/or heteromeric interactions with other family members. The homo- and heterodimerizing superfamilies include receptors, enzyme complexes, transcription factors and ion channels and are often functionally very important. Indeed, there is a positive correlation between the number of protein partners and importance to the viability of an organism,[11] and large scale protein-protein interaction screening studies show that proteins that can form homo-oligomers are more likely to have an increased number of binding partners.[3]

Different combinations and permutations of subunits in complexes tend to have different activities, such as transcription factor complexes targeting different DNA sequences or recruiting different cofactors (see accompanying chapters on nucleic acid binding proteins by Wilce, Vivian and Wilce and transcription factors by Funnell and Crossley). The exchange of a single component can transform a transcription complex from one that activates to one that represses transcription. This ability to use transcription factors and other regulatory