Title
Biophysical and phylogenetic studies of Clathrin

Author
Wakeham, Diane Elizabeth

Publication Date
2003

Peer reviewed|Thesis/dissertation

Biophysical and Phylogenetic Studies of Clathrin

by

Diane Elizabeth Wakeham

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication

In memory of William Bailey,

who believed in me

before I believed in myself.

Acknowledgements

I am grateful to many people and groups, without whose scientific advice or emotional support and encouragement this thesis would never have been completed.

I thank my advisor Frances Brodsky for her unfailing optimism and insight, years of patience and support, and for teaching me to keep my eye on the big picture. I thank my thesis committee members Robert Fletterick and Charles Craik, for their support and inspiration just when I needed it most. To Wendell Lim and Keith Mostov for serving on my oral qualifying exam committee.

To Betsy Bennett, for late nights full of helpful advice on surviving the oral qualifying exam and for bringing so much fun into going to work. To Larry Lem, for mixing late night science with philosophical discussions on the meaning of life and for his friendship.

To Shu-Hui Liu, for her time and patience in guiding and advising me in Frances' absentia when I was learning the ropes. To Diana Sloweijko, Andy Wilde, and Barrie Greene for teaching a small molecule chemist the techniques of molecular biology and protein chemistry. To Peter Hwang, Joel Ybe, Manish Butte, and Michael Reese for support, encouragement, and occasional inspirational discussions on structural issues. To Chih-Ying (Jean) Chen, for her friendship and support and for her thoughtfulness in making sure that I always had something to eat when I was too focused on the tasks before me. To Tina

Tran, for her help in protein preparation and for sharing my love of poetry and spirituality.

To Christine Knuehl, Angela Stoddart, Victoria Crotzer, Mhairi Towler, Sherri Newmyer, Noriko Arase, David Riethof, Alejandra Solache-Diaz, Jyh-I Lu, Allan Mabardy, Venus Manalo, and all members of the Brodsky lab past and present, for tolerating long group meetings and offering support over many years.

To Peter Hwang, Giselle Knudsen, Dyche Mullins for helpful suggestions on AUC data collection and analysis, and especially to Giselle for her friendship, support, and mass spectrometry help. To Mei Li Wong, for her patience and advice in training me on the electron microscope. To Laurent Abi-Rached, Libby Guethlien and Erin Adams from Peter Parham's laboratory for introducing me to bioinformatics and providing useful advice.

To Brooke Yool, for being a fantastic roommate and for her example in how to be a fantastic teaching assistant. To Cath Gilchrist, for her support, encouragement, and for sharing her molecular biology expertise.

To Chris Olsen, Nancy Mutnick, Edna Rodas, Grace Stauffer, Diana Trail, Lisa Louie, Lois Greene, Nancy Quintrell, Nora Hom-Booher, and Liza Crosse for administrative support.

To my husband Ken Ricci, for his patience, understanding, devotion, and support and for his advice on how to survive graduate school. To my best friend Michele Harms, for being my counselor and providing Ben and Jerry's ice cream when I needed it most.

To my parents Peter and Eileen Wakeham, my sisters Lorna Roberts and Suzanne Wakeham, and all my family for having more confidence in me that I did in myself. To my in-laws Liz and Rich Ricci, for sending little notes of encouragement along the way.

To my many friends whose advice and encouragement along the way made such a difference: John Anderson, Keith Burdick, Paul Canavese, Bruce Caruthers, Stephen and Lucy Cho, Suzanne Frasca, Lauri Gilbilterra, Theresa Hannon, Andrew Jeung, Mark and Debbie Kaplan, Sherry LaPorte, Nancy Mar, Berford Moncriffe, Dustin and Diana Putnam, Mark and Lesley Ruzon, Jen and Marc Schaub, Linda Valenti, Amanda Vu, and to Mark Roberts, wherever he may be in life.

To the Catholic Community at Stanford, especially to the Chapel Choir under the direction of Teresa Pleins and PIA (Peninsula Interfaith Action), for reminding me that there was more to life than graduate school and for being so patient when I kept forgetting.

To William Draper, Donald Wijekoon, John Johnston, and all my former colleagues from the California Department of Health Services in Berkeley, for giving me my first glimpse of life as a research scientist.

To Kathleen Roskos, whose advice on applying for and surviving graduate school was invaluable, and to Andrea Rourke, Keith Washco, George Trager, Richard Lee, Richard Maskiewicz, Richard Jones, and all my former colleagues at Matrix Pharmaceutical, Inc. for their encouragement.

To my funding sources: this work was supported by N.I.H. grants GM55143 and GM38093 to the Brodsky Laboratory, a Graduate Research Opportunity Fellowship from the University of California in 1996-1997, and an ARCS (Achievement Rewards for College Scientists) Fellowship in 1997-1998.

Finally, to God, whose protein engineering of clathrin is a brilliant inspiration, and who is the source of both my strength and my passion for science. "Give thanks to the Lord... Sing to God a new song... The Lord loves justice and right and fills the earth with goodness" (Psalm 33).

Involvement of Coauthors- Scientific Contributions

The research presented in this dissertation resulted in two manuscripts. One is under revision after favorable review by EMBO Journal (Chapter 3) and the second is in preparation (Chapter 4). Significant contributions were made to three additional review publications (Appendix). Some of this research and much unpublished work (Chapter 2) involved the work of collaborators. The chapters and publications are listed below with the contributions of the various co-authors.

Chapter Two

Dr. Joel Ybe created the Ybe Model, on which my creation of Figure 2.1 was based. Ernest Chen created the plasmids for the mutants to test the Ybe Hypothesis. Dr. Joel Ybe and Dr. Peter Hwang solved the crystal structure of the proximal leg and created Figure 2.3A. Dr. Alan Roseman (MRC, Cambridge) and Dr. Peter Hwang (MSG, UCSF) created the Roseman Model of the docked proximal leg in the cryo-EM map and Figure 2.4, and provided the PDB file for the docked structure from which I generated Figures 2.8 and 2.9 and Table 2.6. Dr. Peter Hwang, Jeremy Wilbur, and Dr. Alan Roseman created the Wilbur Model and Figure 2.12, and provided the PDB file from which I generated Figure 2.13. Dr. Barrie Greene and Dr. Shu-Hui Liu provided guidance, and Dr. Frances Brodsky supervised and guided this work.

Chapter Three

Wakeham, D. E., Chen, C.Y., Greene, B., Hwang, P.K., and Brodsky, F.M. 2003. Clathrin self-assembly involves coordinated weak interactions favorable for cellular regulation. *Reviewed favorably by EMBO Journal.*

This publication is reproduced in Chapter 3, in a slightly modified form. Most of the work for this paper was done by me. The coauthors contributed as follows: the raw data for the FPLC work in Figure 3.2B is the work of Dr. Barrie Greene, the yeast-two-hybrid data in Table 3.1 is the work of Dr. Chih-Ying Chen, and the SPR result in Figure 3.5 is the work of Dr. Peter K. Hwang. Dr. Chih-Ying Chen also provided editorial comments. Dr. Frances Brodsky guided and supervised this work and provided intellectual and editorial input.

Chapter Four

Wakeham, D. E., Towler, M., Newmyer, S., Brodsky, F.M. 2003. Clathrin Specialization Explained by Phylogenetic Analysis. *Manuscript in preparation.*

This manuscript is based on the work presented in Chapter 4. The vast majority of the work was done by me. The coauthors contributed as follows: Dr. Mhairi Towler searched the database and located many sequences for clathrin heavy chains and contributed to the map-based paralogue analysis on the heavy chain chromosomes, and Dr. Sherri

Newmyer searched the database and located many sequences for clathrin light chains and contributed to the map-based paralogue analysis on the light chain chromosomes. Dr. Frances Brodsky guided and supervised this work.

Appendix

Brodsky, F.M.; Chen, C.Y.; Knuehl, C.; Towler, M.; Wakeham, D.E. Biological Basketweaving: Clathrin-Coated Vesicle Function, Structure, and Regulation. *Ann. Rev. Cell Dev. Biol.* 2001 17:517.

This publication is reprinted in the appendix, with permission, from the *Annual Reviews of Cell and Developmental Biology*, Vol. 17, ©2001 by Annual Reviews www.annualreviews.org. The text of the manuscript is the work of Dr. Frances Brodsky. I provided editorial input for the sections on the structure of the clathrin terminal domain and accessory proteins epsin, Eps15, dynamin, and AP180/ CALM. I created all figures. Figures 4 was an original effort and Figure 3 was a modification and addendum to previous original efforts from the second appendix paper below, while Figures 1 and 2 were based on revisions of previously existent unpublished figures belonging to Dr. Frances Brodsky. Figure 5 was a modification of the original cryo-EM images with permission of the authors (Smith et al., 1998) and publisher.

Wakeham, D.E.; Ybe, J.A.; Brodsky, F.M.; Hwang, P.K. Molecular Structures of Proteins Involved in Vesicle Coat Formation. *Traffic* 2000 1: 393-398.

This publication is reproduced in the appendix with persmission of Blackwell Publishing Ltd. All coauthors contributed substantially to this manuscript. I coordinated the efforts of the authors, provided editorial input, and created or revised figures. The text and figures from the clathrin terminal domain structure, arrestin, Hsc70, and Nef portions of the manuscript is my work. The text of the amphiphysin, dynamin, and Eps15 portions of the manuscript is the work of Dr. Frances Brodsky, with the figures for these sections created by me. Dr. Joel Ybe and Dr. Peter Hwang wrote the text and provided the figures for the ARF related structures, and Dr. Peter Hwang write the text and provided the figures for the adaptin structures and the clathrin proximal leg.

Ybe, J.A.; Wakeham, D.E.; Brodsky, F.M.; Hwang, P.K. Molecular Structures of Proteins Involved in Vesicle Fusion. *Traffic* 2000 1: 474-479.

This publication is reproduced in the appendix with permission of Blackwell Publishing Ltd. All coauthors contributed substantially to this manuscript. Dr. Joel Ybe and Dr. Frances Brodsky coordinated the efforts of the authors and provided editorial input. I also provided editorial input for the entire manuscript. The text and figures from the synaptotagmin portion of the manuscript is my work. Dr. Joel Ybe wrote the text and provided the figures for the SNAP25/SNARE, syntaxin, and NSF structures, and Dr. Peter Hwang wrote the text and provided the figures for the Rab/ Rabphilin and Rab-GDI sections.

Abstract

Biophysical and Phylogenetic Studies of Clathrin

Diane Elizabeth Wakeham

Clathrin has a unique triskelion shape that allows it to self-assemble into extended lattices along the cellular membranes. These clathrin cages surround membrane buds, which pinch inward to create vesicles within the cell during endocytosis and other steps of intracellular membrane traffic. While clathrins have been researched for decades, the nature of the interactions between clathrin heavy chains during clathrin self-assembly has been unclear. Site-directed mutagenesis and modeling were used to explore electrostatic interactions between clathrin leg segments during self-assembly and suggested they are repulsive in nature. Chimeric proteins were created which demonstrated that self-assembly is a cooperative, entropy-driven polymerization driven primarily by multiple weak hydrophobic interactions.

This work was followed by a bioinformatic study of clathrin genes demonstrating that the genes for both clathrin light chain and clathrin heavy chain subunits were duplicated during the time frame of evolution of chordates into vertebrates. Preliminary evidence suggests that the clathrin heavy chain gene may have duplicated in a large-scale genomic duplication event, in order to support increased neuronal and muscular sophistication in primitive vertebrates.

Table of Contents

DEDICATION III

ACKNOWLEDGEMENTS IV

INVOLVEMENT OF COAUTHORS- SCIENTIFIC CONTRIBUTIONS VIII

ABSTRACT XII

TABLE OF CONTENTS XIII

LIST OF FIGURES XVI

LIST OF TABLES XVI

CHAPTER ONE: INTRODUCTION 1

ENDOCYTOSIS: CELLULAR INTERNALIZATION 2

CLATHRIN-DEPENDENT INTRACELLULAR TRAFFICKING PATHWAYS 7

CLATHRIN POLYMERIZATION: SELF-ASSEMBLY 9

CLATHRIN MOLECULES AND STRUCTURES 13

REGULATORY CONSIDERATIONS: LATTICE INDUCTION AND MODIFICATION BY OTHER PROTEINS 17

CHAPTER TWO: EXPLORING THE MOLECULAR BASIS FOR CLATHRIN SELF-ASSEMBLY 19

ABSTRACT 19

INTRODUCTION 20

THE YBE HYPOTHESIS 21

Mutagenesis: Testing the Ybe Hypothesis *26*

THE CRYSTAL STRUCTURE AND THE ROSEMAN MODEL 31

The Role of Histidines *37*

The Role of Divalent Metal Ions *41*

Minimization of the Interface *46*

Mutagenesis: Testing the Roseman Model *50*

THE WILBUR MODEL 59

DISCUSSION 62

CHAPTER THREE: CLATHRIN SELF-ASSEMBLY INVOLVES COORDINATED WEAK INTERACTIONS FAVORABLE FOR CELLULAR REGULATION 69

ABSTRACT 69

INTRODUCTION 70

RESULTS 75

Individual Leg Segments Are Monomeric Under Assembly Conditions *75*

Chimeric Constructs Segment Interactions Assemble Using Cooperative Leg Interactions *85*

DISCUSSION 89

Thermodynamic Properties of Clathrin Assembly *90*

Cellular Regulation of Clathrin Assembly and Curvature *92*

CHAPTER FOUR: CLATHRIN SPECIALIZATION EXPLAINED BY PHYLOGENETIC ANALYSIS 97

ABSTRACT 97

INTRODUCTION 98

The Clathrin Sequences *99*

The Genomic Duplication Theory *102*

Primitive Chordates and Vertebrates and the Evolution of Complex Systems *105*

Single or Local Gene Duplication: Alternatives to Genome Duplication Theory *109*

RESULTS 110

The Chromosomal Loci of Clathrins *111*

Phylogenetic Analysis: A Tree Based Approach *115*

Paralogue Analysis: A Map Based Approach *120*

DISCUSSION 127

CHAPTER FIVE: MATERIALS AND METHODS 170

CHAPTER SIX: CONCLUSIONS AND PERSPECTIVES 203

REFERENCES 209

APPENDIX 246

BIOLOGICAL BASKETWEAVING: CLATHRIN-COATED VESICLE FUNCTION, STRUCTURE, AND REGULATION 247

MOLECULAR STRUCTURES OF PROTEINS INVOLVED IN VESICLE COAT FORMATION 305

MOLECULAR STRUCTURES OF PROTEINS INVOLVED IN VESICLE FUSION 311

List of Figures

FIGURE 1.1: FORMATION OF CLATHRIN-COATED PITS AND ENDOCYTOSIS 4

FIGURE 1.2: CLATHRIN'S ROLE IN MEMBRANE TRAFFIC IN THE EUKARYOTIC CELL 8

FIGURE 1.3: THE CLATHRIN TRISKELION: MORPHOLOGY AND DOMAINS 10

FIGURE 1.4: THE ASSEMBLED CLATHRIN CAGE 11

FIGURE 2.1: THE YBE STRONG SALT BRIDGE HYPOTHESIS OF CLATHRIN ASSEMBLY 23

FIGURE 2.2: ASSEMBLY OF MUTANT HUB MOLECULES TESTING THE YBE HYPOTHESIS 30

FIGURE 2.3: CRYSTAL STRUCTURES OF CLATHRIN 32

FIGURE 2.4: THE ROSEMAN MODEL FOR THE ASSEMBLED PROXIMAL LEG INTERFACE 35

FIGURE 2.5: DISTRIBUTION OF SURFACE- ACCESSIBLE HISTIDINES ON CLATHRIN PROXIMAL LEG SEGMENT 36

FIGURE 2.6: PREDICTED DISTRIBUTION OF HISTIDINES ALONG THE ENTIRE CLATHRIN LEG 39

FIGURE 2.7: THE EFFECT OF DIVALENT METAL IONS ON CLATHRIN HUB ASSEMBLY 44

FIGURE 2.8: CHCR6 IN THE ROSEMAN MODEL OF THE PROXIMAL LEG SEGMENT INTERFACE 49

FIGURE 2.9: POTENTIAL SALT BRIDGES IN THE ROSEMAN MODEL INTERFACE BETWEEN PROXIMAL LEG SEGMENTS 52

FIGURE 2.10: ASSEMBLY OF MUTANT HUBS TO TEST THE ROSEMAN MODEL 56

FIGURE 2.11: MUTANT HUB ASSEMBLY IN THE PRESENCE OF SALT: ELECTROSTATIC EFFECTS 57

FIGURE 2.12: THE WILBUR MODEL FOR THE ASSEMBLED PROXIMAL LEG BINDING 60

FIGURE 2.13: LOCATION OF PREVIOUSLY MUTATED RESIDUES WITHIN THE WILBUR MODEL 61

FIGURE 2.14: ELECTROSTATIC REPULSION AND PH VARIANCE OF OBSERVED MOLECULAR SIZE FOR PROXIMAL LEGS IN EQUILIBRIUM ULTRACENTRIFUGATION 65

FIGURE 2.15: SURFACE POTENTIAL- LACK OF COMPLEMENTARY ELECTROSTATIC REGIONS 67

FIGURE 3.1: CLATHRIN POLYMERIZATION 72

FIGURE 3.2: INDIVIDUAL LEG SEGMENTS ARE MONOMERIC BY FPLC 77

FIGURE 3.3: SEDIMENTATION VELOCITY ASSAYS. 79

FIGURE 3.4: PROXIMAL LEG SEGMENT IS A MONOMER AT PH 6.2 IN EQUILIBRIUM SEDIMENTATION ASSAY 80
FIGURE 3.5: SURFACE PLASMON RESONANCE FAILS TO DETECT INTERACTION BETWEEN PROXIMAL LEG SEGMENTS 83
FIGURE 3.6: CHIMERIC TRIMERIZED CLATHRIN PROXIMAL OR DISTAL LEG SEGMENTS ASSEMBLE 87
FIGURE 4.1: CLATHRIN DOMAINS AND SEQUENCE CONSERVATION 100
FIGURE 4.2: A TIMELINE OF VERTEBRATE EVOLUTION AND THE HYPOTHESIZED GENOME DUPLICATION 103
FIGURE 4.3: BAYESIAN PHYLOGENETIC TREE OF CLATHRIN HEAVY CHAINS 116
FIGURE 4.4: BAYESIAN PHYLOGENETIC TREE OF CLATHRIN LIGHT CHAINS 117
FIGURE 5.1: PLASMID CONSTRUCTS CREATED BY PREVIOUS LAB MEMBERS 171
FIGURE 5.2: CHCR6 CONSTRUCTS FOR PROTEIN EXPRESSION 173
FIGURE 5.3: CHIMERIC CLATHRIN HUB CONSTRUCTS 175

List of Tables

TABLE 2.1: SUMMARY OF INTERACTIONS IN THE YBE HYPOTHESIS 25

TABLE 2.2: SUMMARY OF MUTATIONS TESTING THE YBE HYPOTHESIS 27

TABLE 2.3: CONSERVATION OF HISTIDINES IN PROXIMAL LEG SEGMENT INTERFACE 40

TABLE 2.4: INTERACTIONS BETWEEN PROTEINS WITH CLATHRIN HEAVY CHAIN REPEATS (CHCRS) 47

TABLE 2.5: CONSERVATION OF POTENTIAL SALT BRIDGE RESIDUES 51

TABLE 2.6: DISTANCES BETWEEN POTENTIAL SALT BRIDGE PAIRS IN THE ROSEMAN MODEL 53

TABLE 2.7: SUMMARY OF MUTATIONS TESTING THE ROSEMAN MODEL 54

TABLE 2.8: ACIDITY OF RECOMBINANTLY EXPRESSED CLATHRIN FRAGMENTS 63

TABLE 3.1: PROXIMAL LEG SEGMENTS DO NOT INTERACT IN A YEAST-TWO-HYBRID ASSAY 82

TABLE 4.1: POTENTIAL PARALOGOUS GENES BETWEEN CHROMOSOMAL SEGMENTS 5Q35 AND 9P13. 121

TABLE 4.2: POTENTIAL PARALOGOUS GENES BETWEEN CHROMOSOMAL SEGMENTS 17Q23 AND 22Q11. 124

TABLE 4.3: ALIGNMENT OF CLATHRIN HEAVY CHAIN SEQUENCES AND SEQUENCE FRAGMENTS 130

TABLE 4.4: ALIGNMENT OF CLATHRIN LIGHT CHAIN SEQUENCES AND SEQUENCE FRAGMENTS 161

TABLE 4.5: DATABANK ACCESSION CODES FOR CLATHRIN HEAVY CHAIN SEQUENCES USED IN ALIGNMENT 167

TABLE 4.6: DATABANK ACCESSION CODES FOR CLATHRIN LIGHT CHAIN SEQUENCES USED IN THE ALIGNMENT 169

Chapter One: Introduction

Clathrin is a trimeric protein with a unique triskelion shape that has a crucial role in intracellular membrane traffic. It is composed of three heavy chains joined at the C-terminus. The globular N-terminal domains attach to rod-like distal and proximal leg domains, which radiate outward from the C-terminal trimerization domain like spokes of a wheel. Each heavy chain is bound by a smaller regulatory subunit, the clathrin light chain. The unusual pinwheel-like morphology allows clathrin to form extended lattices along the membrane surface and spherical baskets around vesicles. Clathrin lattices at the membrane function to sequester receptors and their bound ligands into clathrin-coated pits on the membrane surface, and to induce or stabilize curvature of these clathrin-coated pits as the membrane buds off and is moved about by the cell as a clathrin-coated vesicle. Clathrin is then removed from the surface of the vesicle and recycled for use in another round of vesicle formation, and the vesicle is tethered to the cytoskeletal elements in the cell and motored to its intended subcellular location.

In this dissertation I consider the mechanism of clathrin self-assembly and the duplication of the clathrin genes in the human genome. I begin by introducing the cellular role of clathrin (Chapter One). More detailed introductions to clathrin function (Brodsky et al., 2001) and structure (Wakeham et al., 2000; Ybe et al., 2000) to which I have contributed can be found in the Appendix of this dissertation. I describe my biochemical and biophysical

exploration of the molecular contacts between clathrin heavy chains during self-assembly, and the mechanism by which clathrins polymerize into a lattice (Chapter Two). I continue by detailing studies revealing the cooperative nature of clathrin self-assembly and discuss its implications in lattice stability and rearrangement (Chapter Three). Bioinformatic methods were then used to investigate the duplication of clathrin heavy chain and clathrin light chain genes in the human genome, to estimate when and how the duplicate copy evolved and to consider its functional implications (Chapter Four). The methodology used in all these experiments is described in detail later in the thesis (Chapter Five). Finally I consider future directions for research (Chapter Six).

Endocytosis: Cellular Internalization

All living cells are bounded by an external membrane made of a lipid bilayer. This plasma membrane maintains the integrity of the cell by isolating it from its surroundings, allowing it to collect and maintain a higher concentration of certain proteins and ions than that found in the extracellular medium. Translocation of even small molecules and ions into the cell through the membrane is difficult, owing to the hydrophobicity of the oily lipids and their natural impermeability to hydrophilic molecules found in the aqueous cytosolic fluid inside the cell. However, cells do need to exchange molecules with the extracellular fluid, and they have evolved several highly specialized systems to allow this to occur.

For some needs, transport proteins that span the membrane act to facilitate the transport of specific molecules. Carrier proteins, or transporters, bind to their specialized cargo and undergo a series of conformational changes to open successive gates and allow the cargo passage through to the interior of the cell. Carrier proteins are important for absorption of sugars, nucleotides, amino acids, small drug molecules, and some ions. Channel proteins, on the other hand, are narrow corridors through the membrane, where the specific size and charge of the ion determine its ability to pass through the channel. Most cells maintain electrochemical gradients using channel proteins (Alberts et al., 2002).

Cells also need to bring in molecules and complexes that are too large to fit through a narrow transporter gate or ion channel. Instead, the cells use endocytosis, a process in which the cellular membrane forms a local invagination coated with a clathrin lattice (also called a clathrin-coated pit), and the membrane buds and eventually pinches off inside the cell, forming a separate small, membrane-bound clathrin-coated vesicle inside the cell (Figure 1.1). Once the vesicle is severed from the membrane, the clathrin coat is removed from the outer surface. This clathrin-coated vesicle (CCV) is used as a shuttle to move the vesicle contents, which were formerly outside the cell, to a specific membrane-bound organelle inside the cell. Such organelles include the trans-Golgi network, endoplasmic reticulum, the nucleus, or lysosomes. The vesicle fuses with its target organelle to deliver its cargo (Alberts et al., 2002). There are additional endocytic pathways that do not depend on clathrin (Figure

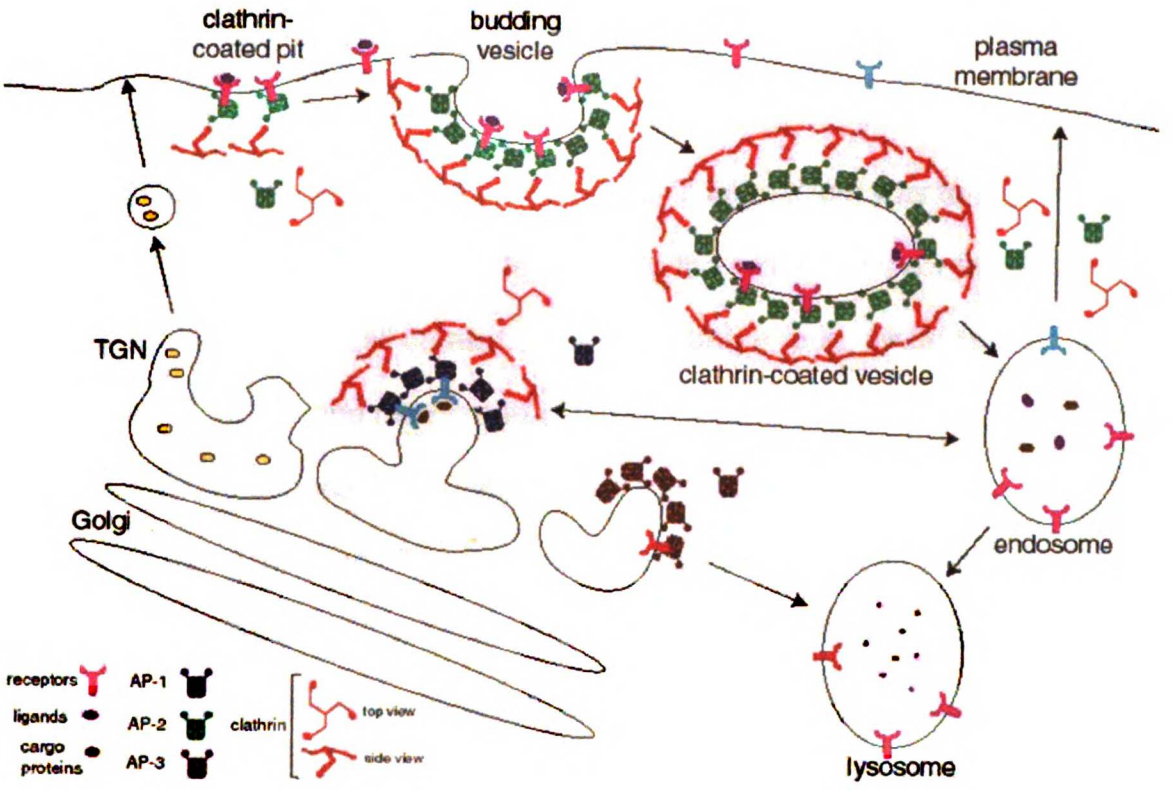


Figure 1.1: Formation of Clathrin-Coated Pits and Endocytosis
Transmembrane receptors bind to adaptor complexes, which bring them to docking sites for membrane traffic. The AP1 complex functions in TGN to endosome traffic, while the AP2 complex is involved in traffic from the plasma membrane to the endosomes. AP3 and AP4-coated vesicles in the cell appear to not involve clathrin coats. The clathrin assembles at the membrane into a clathrin-coated pit. Rearrangements in the clathrin lattice coincide with membrane deformation into a constricted bud containing the receptors and their bound ligands. The bud is then severed from the membrane surface, and the clathrin-coated vesicles are uncoated before the vesicle travels to its subcellular destination. The increasingly acidic pH in endosomes and particularly lysosomes results in dissociation of ligands from receptors or in protein degradation.

1.2A)(Brodsky et al., 2001), such as caveolae-mediated internalization of lipid raft domains (Anderson, 1998), but receptor-mediated endocytosis via clathrin-coated pits is the primary mechanism ascribed to the uptake of most larger molecules into the cell.

Receptor-mediated endocytosis (RME) is an efficient way to regulate transportation of specific macromolecules into the cell and deliver them to their desired locations. The cell creates a transmembrane receptor protein, which is embedded in the plasma membrane of the cell with a binding site for its cargo on the extracellular surface of the membrane and a signaling tail on the inside surface of the membrane. The number of receptors on the cell membrane is regulated to allow the cell to control the amount of cargo taken in by the cell. When an extracellular cargo molecule binds to the receptor, the cargo is incorporated with its receptor into a nascent vesicle. Specific amino acid sequences in the receptor tail, called cellular localization signals, direct the cargo to the appropriate location within the cell for its delivery (Blobel, 2000).

There are two types of receptor-mediated endocytosis (Brodsky et al., 2001). The first type involves signaling receptors that are internalized only when bound to cargo, for example G-protein coupled receptors (GPCRs) or growth factor receptors and other receptor tyrosine kinases (RTKs). Ligand binding triggers cascades that phosphorylate the intracellular receptor tail, which in turn allows it to be both mono-ubiquitinated and recognized by the adaptor complex AP2, a component of the clathrin internalization pathway. AP2 binding allows recruitment of the receptor into a clathrin-coated pit and thus incorporation into a

budding vesicle. Mono-ubiquitination allows Hrs binding, a second adaptor for clathrin coated vesicles moving from endosomes to lysosomes (Clague, 2002). These allow the duration of intracellular signaling to be a direct consequence of its activation, because internalization can lead to digestion or inactivation of the receptors and thus to termination of signaling and downregulation of the number of receptors on the cell surface (Tsao and von Zastrow, 2000). The second type of receptor-mediated endocytosis, constitutive endocytosis, occurs continually at a regular rate whether the ligand is bound to the receptor or not, because AP2 binding is not dependent on a phosphorylation event. One example of a receptor that uses constitutive internalization is transferrin receptor, the cellular mechanism for internalization of iron.

Clathrin thus mediates cellular internalization for a wide variety of ligands, such as cholesterol (Anderson et al., 1976), hormones (van Kerkhof et al., 2000), certain viruses (Helenius et al., 1980), liposomes (Straubinger et al., 1983), extracellular matrix proteins (Uekita et al., 2001), immune complexes (Willingham et al., 1979), insulin (Maxfield et al., 1978), and growth factors (Sorkin and Waters, 1993). The entire process of endocytosis, from coated pit formation through vesicle uncoating, can occur in under one minute (Kirchhausen, 2000a), or perhaps even faster in neuronal cells. A normal macrophage is estimated to endocytose and replenish by exocytosis rapidly enough that it internalizes 3% of its plasma membrane every minute, suggesting complete turnover every half hour (Brodsky, 1988; Robinson, 1987). Coated vesicles range in size from 50 to

250 nm in diameter (Pearse and Bretscher, 1981), perhaps allowing larger cargo molecules to tailor their clathrin coats for a custom fit.

Clathrin-Dependent Intracellular Trafficking Pathways

Cells contain many lipid membrane-bound organelles with vesicles trafficking between them, and clathrin is not limited to facilitation of membrane traffic at the plasma membrane (Brodsky et al., 2001). Newly synthesized proteins that have finished maturation in the Golgi bud from the trans-Golgi network (TGN) via clathrin-coated vesicles. In contrast to the plasma membrane vesicles, which use the tetrameric AP2 adaptor complex, the TGN vesicles use the AP1 adaptor complex. Both plasma membrane- and TGN- derived vesicles are motored to the endosomes, where they are sorted by the cell and sent either to the plasma membrane (for exocytosis or export of cytosolic proteins, or to recycle a receptor for another round of ligand binding) or to the lysosome (for its degradation). In addition, some secretory vesicles moving directly from the TGN to the plasma membrane appear initially associated with a partial clathrin coat, to remove unwanted receptors from secretory granules (Kuliawat et al., 1997; Molinete et al., 2001).

Additional tetrameric adaptor complexes homologous to AP1 and AP2 have been identified that form a coat around budding membranes but appear not to require a clathrin coat for their intracellular trafficking. AP3 appears to be involved in traffic from TGN or endosomes to lysosomes, and AP4 in traffic from endosomes to TGN (Robinson and Bonifacino, 2001). GGAs, which have

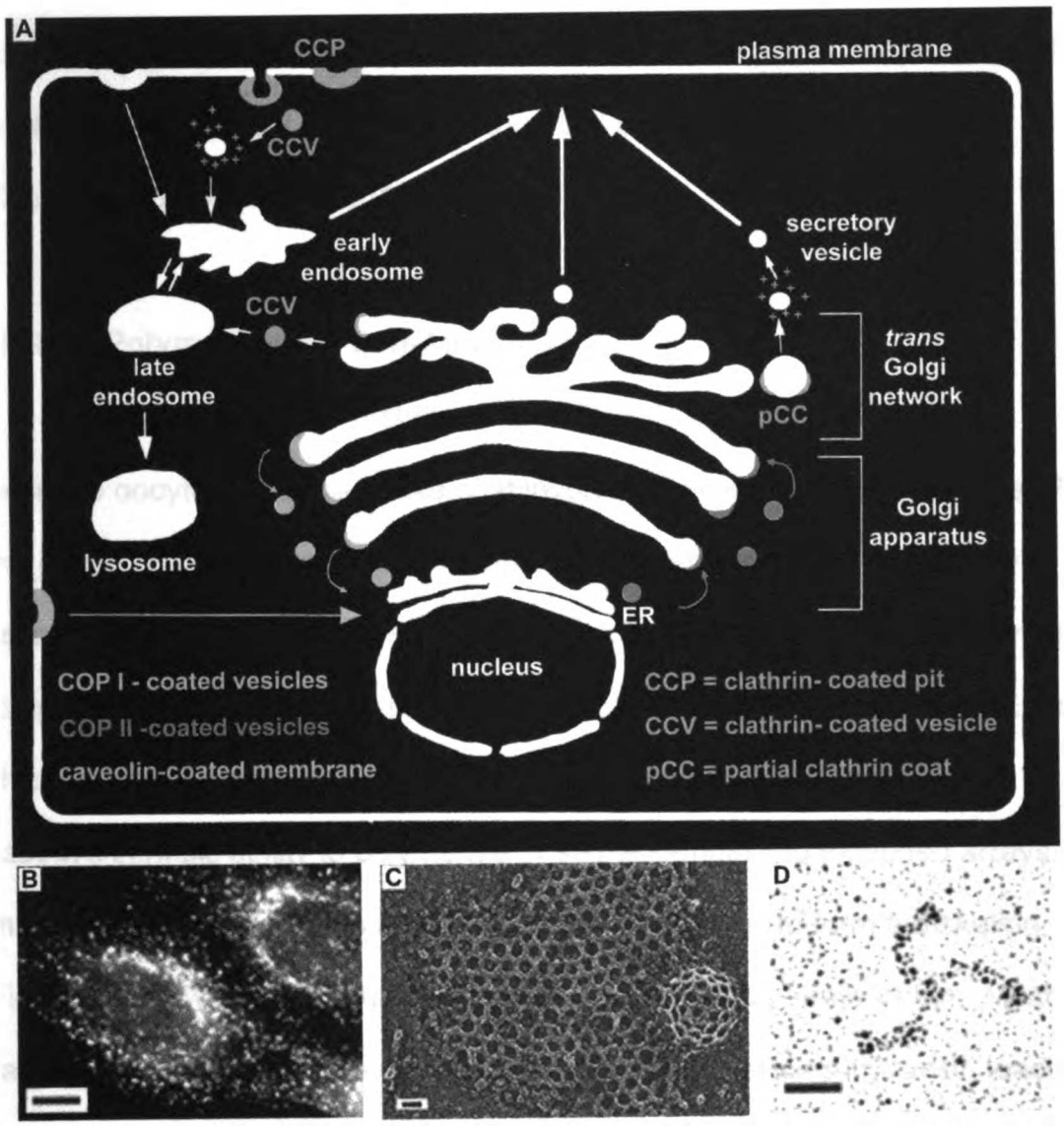


Figure 1.2: Clathrin's Role in Membrane Traffic in the Eukaryotic Cell
A. Clathrin-mediated cellular traffic is highlighted in blue to contrast with traffic mediated by other coating proteins. **B.** Distribution of clathrin-coated vesicles in a HeLa cell labeled with the X22 anti-clathrin heavy chain monoclonal antibody and fluorescent anti-immunoglobin viewed by microscopy (Liu et al., 2001a), reproduced with permission from Oxford University Press. The bar indicates 5 micrometers. **C.** An electron micrograph of a membrane-associated clathrin lattice and emerging clathrin-coated pit (Heuser et al., 1987), reproduced with copyright permission from Rockefeller University Press. The bar indicates 33 nanometers. **D.** A clathrin triskelion purified from bovine brain clathrin-coated vesicles and visualized by platinum shadowing (Liu et al., 2001a), reproduced with permission from Oxford University Press. The bar indicates 20 nanometers. This figure in its entirety has appeared in a review article (Brodsky et al., 2001) and is reproduced here with permission from Annual Reviews.

sequences homologous to adaptor subunits, appear to act as adaptors for ARF-dependent recruitment of clathrin to the TGN during AP1-vesicle budding (Boehm and Bonifacino, 2001).

Clathrin Polymerization: Self-Assembly

Clathrin coats were first discovered in 1964 by Roth and Porter in mosquito oocytes as a bristle-like coat involved in yolk protein endocytosis (Roth and Porter, 1964). These vesicles were isolated from pig brain in 1975 by Pearse and colleagues (Pearse, 1975), and clathrin was identified as the major component of this coat. The name clathrin comes from the Greek root klEithron, which led to the Latin word *clathratus*, meaning lattice-like, and it refers to clathrin's intrinsic ability to polymerize or self-assemble into extended arrays. Three clathrin heavy chains are joined at the C-terminus to form a triskelion (Figure 1.2D), with one clathrin light chain bound to each heavy chain. Most clathrin triskelia in a cell are cytosolic or surrounding vesicles in-transit, leading to a punctate distribution in the cell as visualized by immunofluorescence (Figure 1.2B). However, clathrin triskelia have the remarkable ability to self-assemble into a hexagonal lattice onto localized areas of the plasma membrane to form clathrin-coated pits (Figure 1.2C). Moreover, as pentagons are incorporated into this hexagonal lattice, a more puckered lattice is formed, which eventually allows the edges of the lattice to bud off sections of the membrane (Figure 1.2C). These constricted buds are then severed from the membrane to become free clathrin-coated vesicles (CCVs) inside the cell. Clathrin is able to self-assemble

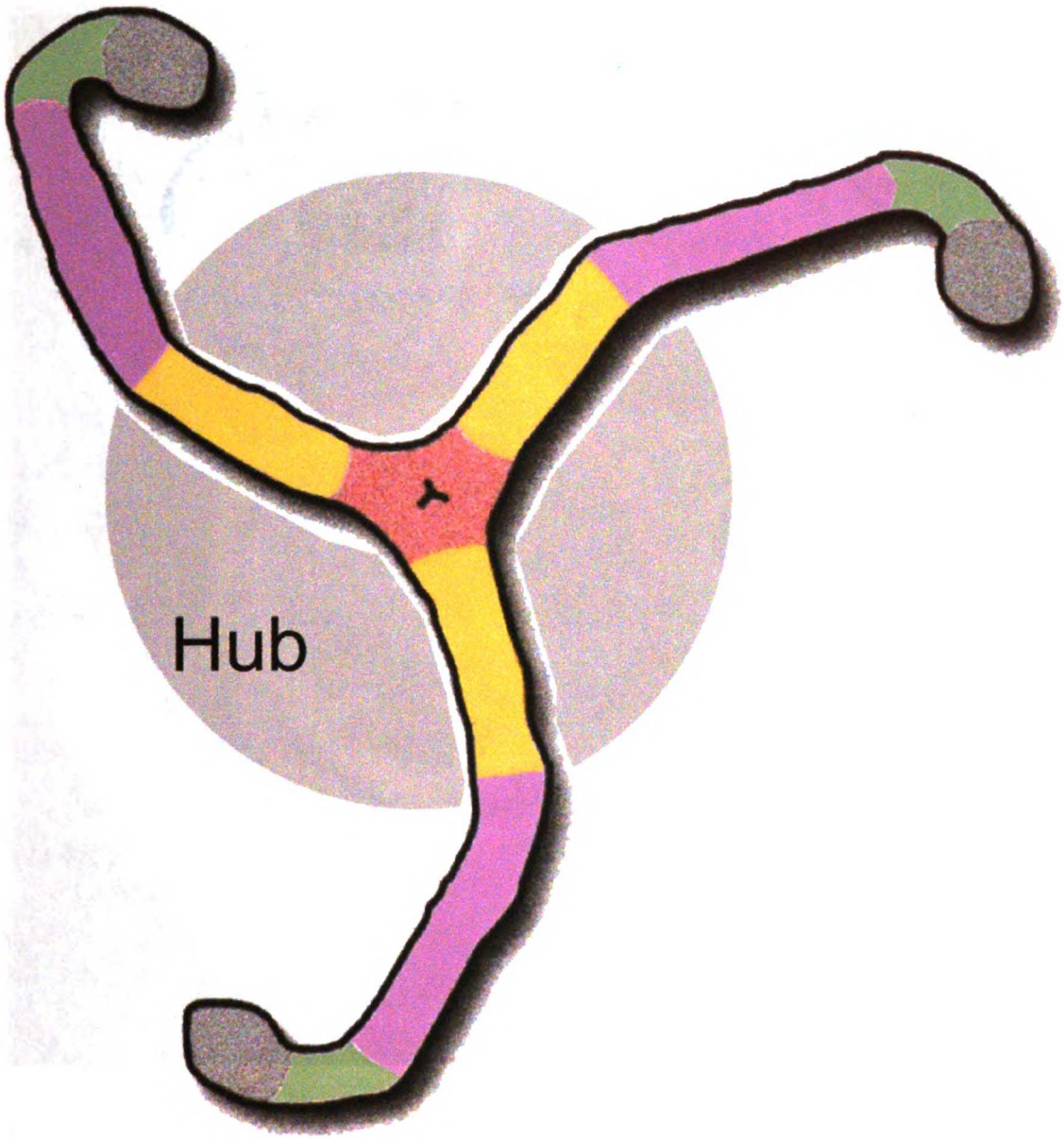


Figure 1.3: The Clathrin Triskelion: Morphology and Domains
Three clathrin heavy chains are joined at the C-terminal trimerization domain (1523-1675, orange) in a slightly puckered 120-degree angle. The proximal leg segment (1074-1522, yellow) is adjacent to the trimerization domain, and beyond a knee-like bend extends the distal leg segment (494-1074, lavender). The globular N-terminal trimerization (1-330, gray) domains connect to the distal legs via a flexible linker region (330-494, green). A clathrin light chain is bound to each heavy chain along the proximal leg domain. The Hub portion of the molecule is the C-terminal third, comprising proximal leg and trimerization domain (1074-1675). This image was modeled after a similar previously published image (Ybe et al., 1999).

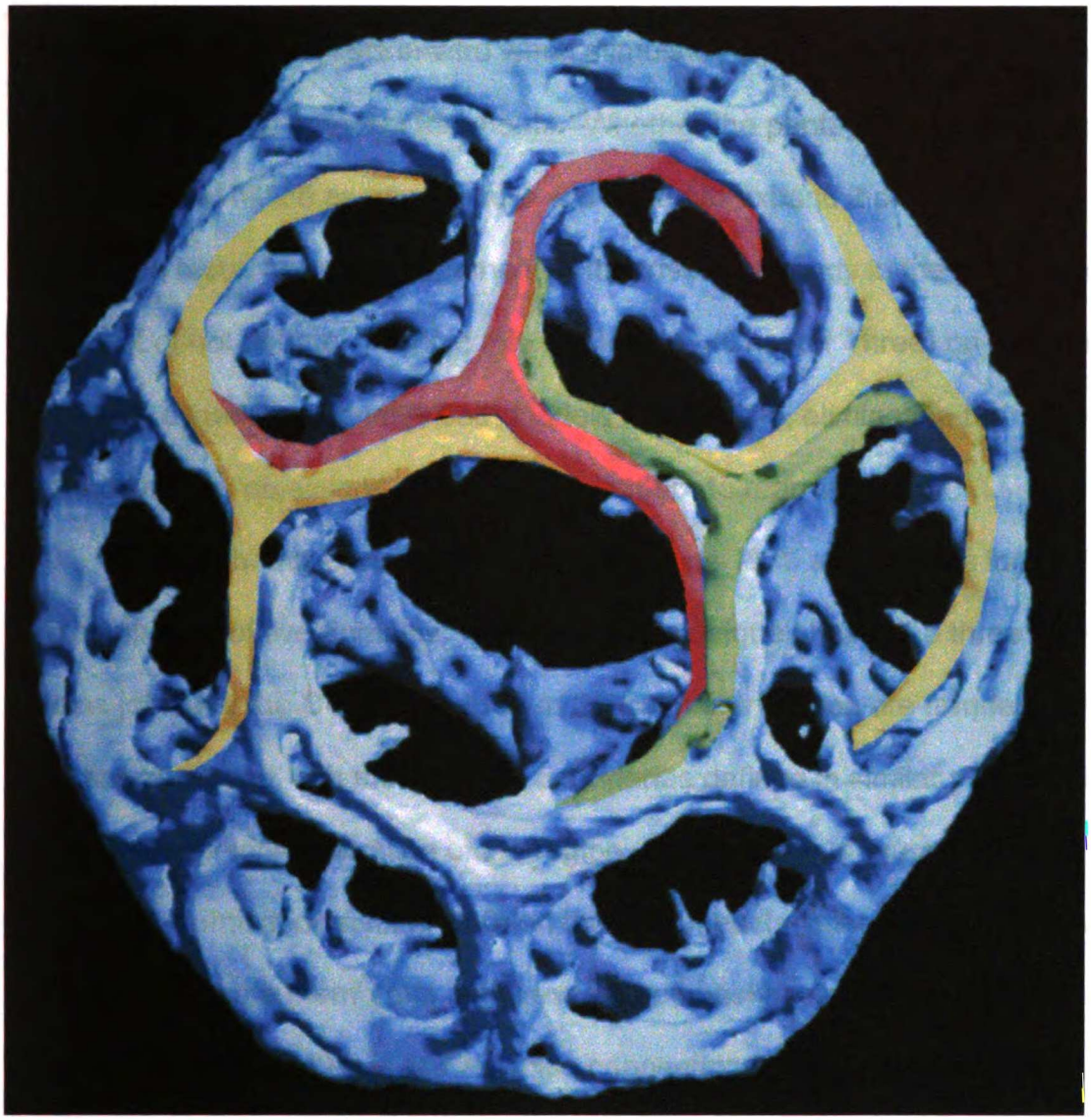


Figure 1.4: The Assembled Clathrin Cage
This 21 angstrom cryo-electron microscopy image of the whole clathrin basket (Smith et al., 1998), reproduced with permission from the authors and with copyright permission from Oxford University Press, has been colored to show the location of the individual triskelia. One side of a hexagon or pentagon in the triskelion lattice is composed of segments of clathrin legs from four different triskelia. Two proximal legs from the green and red triskelia make up the top layer of the lattice. The yellow distal legs of two triskelia centered at adjacent vertices in the lattice can be seen curving underneath the proximal legs to form a deeper layer of the lattice. Note that the terminal domains of the triskelia are not included in this image. The terminal domains curve into the center of the polyhedron, under the vertices (Musacchio et al., 1999). This figure has appeared in a review article (Brodsky et al., 2001) and is reproduced here with permission from Annual Reviews.

into similar closed baskets in-vitro when stimulated by increased acidity or the addition of adaptor complex fragments. Whether the pentagons are incorporated into the in-vivo lattice as new triskelia are added to expand it or whether the pentagons result from rearrangement of the triskelia already in the lattice remains a subject of hot debate. Work in this dissertation (Chapter Three) on clathrin assembly dynamics suggests both models are feasible, eliminating the thermodynamic objections (Kirchhausen, 2000b) to rearrangement.

Clathrin heavy chains are 192 kDa proteins 1675 amino acids long. They are composed of five domains. Because of the unique shape, the three heavy chains joined at the C-terminus to form the clathrin triskelion are frequently referred to as legs, and their domain boundaries defined in similar physiological terms to assist in visualization of the region on a triskelion (Figure 1.3). The globular N-terminal domain (1-330) on the foot of each clathrin leg points inward toward the plasma membrane in the assembled basket structure. A flexible linker domain (330- 494) is like an ankle connecting the pendulous terminal domain to the extended legs. The distal leg segment (494-1074), so called because it is the further of two leg segments from the trimerizing domain, forms the underlayer of clathrin leg segments strengthening the assembled basket. On the other side of the knee-like bend, the proximal leg segment (1074-1522), adjacent to the trimerization domain, forms the upper layer of clathrin leg segments of the hexagonal lattice side in the assembled clathrin cage. Finally, the trimerization domain (1522- 1675) at the C-terminus holds together the three heavy chains at the vertex (See also Figure 4.1).

Cryo-electron microscopy imaging of clathrin cages several years ago yielded a 21 angstrom resolution map of the smallest possible clathrin cage, a remarkable view of the clathrin legs in the extended structure (Smith et al., 1998)(Figure 1.4). The basket or cage has both pentagons and hexagons in a regular array surrounding adaptors and the plasma membrane. At each vertex of the clathrin cage is a trimerization domain. Each polygonal side in the lattice is composed of two layers: two antiparallel proximal legs on the top layer, and two antiparallel distal legs directly underneath them. The three proximal legs from that vertex contribute the upper layer to the three polygonal sides adjacent to this vertex, with three proximal legs from the three adjacent vertexes completing the upper layer of each polygonal side. The knee-like bend curves around beneath the trimerization domain of the adjacent vertex, such that the distal leg contributes to the polygonal segment over one vertex away from its own trimerization domain. The N-terminal domain would lie directly underneath the next vertex/ knee region (See also Figure 3.1). For clarity, the terminal domains, adaptors, and membrane structures in the center of the cage have been omitted in this image. Thus the assembled form of clathrin involves intertwined leg segments, with four separate segments (two proximal and two distal) comprising each polygonal side.

Clathrin Molecules and Structures

Clathrin is found in high abundance, comprising up to 1% of total protein in postmitochondrial extracts of many tissues and cells (Pearse, 1975; Pearse,

1976). About 2% of the plasma membrane is composed of clathrin-coated pits (Goldstein et al., 1979), and clathrin accounts for about 70% of the protein weight in coated vesicles (Pearse and Bretscher, 1981).

Clathrin heavy chains in fungi, invertebrates, and primitive chordates are bound by a single light chain, while those in vertebrates have two light chains LCa and LCb. Three light chains have been found in the plant species *Arabidopsis thaliana* (Scheele and Holstein, 2002). The two vertebrate light chains appear to have redundant function- both are expressed ubiquitously and bind to heavy chain with equal affinity. Similarly, yeasts, invertebrates, and primitive chordates have a single clathrin heavy chain, while vertebrates have two. The first heavy chain is found on chromosome 17 and is known as clathrin heavy chain or CHC17. It functions as an orthologue of the single clathrin heavy chain found in invertebrates, to regulate membrane traffic. It is expressed ubiquitously. The second clathrin heavy chain, found on chromosome 22, is known as CHC22. CHC22 does not bind either vertebrate light chain and appears not to share the invertebrate clathrin heavy chain and CHC17 function in membrane traffic. For this reason, CHC17 will be referred to simply as clathrin heavy chain throughout this dissertation, in agreement with literature conventions. CHC22 may have a role in myogenesis, the development of muscle tissue (Towler et al., 2002). Differences between the two forms of clathrin heavy chain and clathrin light chain are discussed more extensively later in this dissertation (Chapter Four).

The light chains are negatively charged 25-29 kDa subunits. Full length human LCa is 248 amino acids long and full length LCb is 228 amino acids long. Light chains have several important domains, enumerated here. Residues 162-192 in LCa and 155-172 in LCb of the full-length sequences span an alternatively spliced neuron-specific sequence, which is not present in the ubiquitous forms. The center of the light chains, residues 100-162 in LCa and 93-155 in LCb, is the heavy chain-binding region. At the N-terminus, the region from residues 28-50 of LCa and 20-42 of LCb is known as the consensus or conserved domain because it is absolutely conserved in mammalian light chains. Included in this region is 23EED, a sequence critical for light chain regulation of clathrin assembly but not for binding to heavy chain (Ybe et al., 1998). Residues 89-100 of LCa and 82-93 of LCb form a calcium-binding domain. Finally, the C-terminal domain is a calmodulin binding region from 192-248 of LCa and 173 to 228 of LCb (See also Figure 4.1). The light chains appear to associate tightly with the heavy chains shortly after their expression, and do not appear to release from the heavy chain under normal physiological conditions, although they may undergo conformational changes while bound. The light chain subunit has an undefined random structure when alone in solution, but evidence suggests that it acquires a helical conformation upon binding to heavy chain (Chen et al., 2002).

During the course of this work, two fragments of clathrin heavy chain were crystallized, and their structure solutions have led to predictions about the structure of the entire extended heavy chain and the potential interface between leg segments (Discussed in Chapter Two). The N-terminal domain is a seven-

bladed beta-propeller (ter Haar et al., 1998), where peptides from interacting proteins are able to bind in narrow slots between the propeller blades (ter Haar et al., 2000). The proximal leg segment is composed of tandem repeats of a novel structural domain called a clathrin heavy chain repeat (CHCR), a superhelix of ten alpha helices joined in a helix-turn-helix-loop configuration (Ybe et al., 1999). The degenerate sequence motif for CHCR thus identified led to the prediction that the extended clathrin distal and proximal leg segments are entirely composed of seven tandem CHCRs (Ybe et al., 1999). The structures of clathrin and other membrane trafficking proteins have been reviewed (Wakeham et al., 2000; Ybe et al., 2000) and included in the Appendix.

When clathrin is absent, the knockout is lethal or extensive defects are seen (Brodsky et al., 2001). The first organism found to survive without clathrin is yeast, whose internalization system has some fundamental differences from mammalian endocytosis (Baggett and Wendland, 2001; Geli and Riezman, 1998). Of the two yeast strains lacking clathrin, one is lethal, while the other survives but grows slowly and has vacuolar irregularities (Munn et al., 1991; Payne, 1990; Payne et al., 1987). Disruption of clathrin light chain in yeast also results in a slow growth phenotype (Lemmon and Jones, 1987; Silveira et al., 1990). Similarly in mammalian cells, expression of a dominant negative clathrin Hub fragment which binds up the exogenous light chain results in inhibition of endocytosis (Liu et al., 1998).

Regulatory Considerations: Lattice Induction and Modification by Other Proteins

Clathrin works with a number of other proteins and molecules to accomplish the budding of membranes during intracellular membrane traffic. Clathrin polymerization is highly regulated, to allow stimulated self-assembly into lattices only at well-defined clathrin coated pit locations (Gaidarov et al., 1999). Likewise, the full reversibility of clathrin assembly allows vesicles to uncoat before they need to tether and dock with their intended target organelles. The clathrin light chains act as regulatory subunits to inhibit assembly at physiological pH and thus prevent unproductive polymerization in the cytosol (Liu et al., 1995). This inhibition is overcome by the addition of the adaptor complexes, by unknown mechanism (Greene et al., 2000). The adaptor complexes bind both the cytosolic receptor tails and the clathrin terminal domain, recruiting these receptors to the location of a clathrin-coated pit (Robinson and Bonifacino, 2001).

Numerous other proteins have a role in induction of clathrin assembly and curvature of the membrane, most notably the monomeric clathrin adaptor AP180 (Ford et al., 2001; McMahon, 1999) and epsin (Ford et al., 2002). Both induce clathrin assembly and bind lipids, but epsin appears to have a more direct role in initiating membrane curvature. The dynamin GTPase is another important clathrin partner. It functions in vesicle scission, to constrict the neck of a budding vesicle and sever its connection to the plasma membrane (Hinshaw, 2000). The chaperone ATPase Hsc70 and auxilin are also notable for their role in uncoating clathrin and adaptors from the clathrin-coated vesicles (Lemmon, 2001;

Newmyer and Schmid, 2001). These and other cellular factors influencing clathrin assembly have been discussed extensively elsewhere (Brodsky et al., 2001)(Appendix).

Many of the cellular factors discussed appear to have functions similar to those described for clathrin. Adaptor complexes also self-assemble into coats around budding vesicles, and in the case of AP3 and AP4 they do not require clathrin. Some forms of endocytosis do not appear to require clathrin. Membrane curvature can be induced by epsin or by modifications in the lipid content on the outer face of the bilayer. However, the function of clathrin is distinct from all these co-factors despite the seeming overlap in abilities. Only clathrin has the intrinsic ability to reversibly polymerize into an extended regular lattice. Without clathrin, receptors bound by adaptor complexes would lack a location for their recruitment. Clathrin's ability to spontaneously form curved baskets not only assists in initiation of membrane curvature, but also stabilizes curvature induced by epsin or AP2. Multiple proteins work in a coordinated fashion to accomplish each of these goals because this redundancy ensures that these critical functions are maintained, that rapid membrane traffic is unimpeded by any roadblocks to allow cell survival. Clathrin's main functions are to self-assemble reversibly on cellular membranes, to ensnare receptors and their cargo in clathrin-coated pits for membrane transport, and to initiate and stabilize curvature in the budding membrane during the initial stages of vesicle formation.

Chapter Two: Exploring the Molecular Basis for Clathrin Self-Assembly

Abstract

Clathrin proximal legs lie adjacent and anti-parallel on each side of the assembled clathrin cage. The proximal legs are believed to contribute most of the affinity that drives clathrin self-assembly. In 1998, the Ybe Hypothesis of clathrin assembly was published. In this model, salt bridges between adjacent proximal legs are inhibited by competing salt bridges from regulatory clathrin light chain subunits. The specific, conserved charged residues proposed to be involved in the proximal leg to proximal leg interaction were altered by site-directed mutagenesis, but the extent of self-assembly of mutated clathrin Hubs remained unchanged. Meanwhile, the crystal structure of the proximal leg was solved, enabling the first glimpse at the structure of clathrin on a molecular level and the discovery of the seven tandem CHCR domains, three of which were in the proximal leg. Efforts to locate the critical interacting region between proximal legs by expressing a single recombinant CHCR as a first step in minimization of the interface were unsuccessful. The possibility that histidines and divalent metal ions might regulate clathrin self-assembly was briefly explored. Finally, using the Roseman Model of the proximal leg crystal structure docked into a high resolution cryo-EM map as a dimer, new salt bridge predictions were made. A new round of site-directed mutagenesis was undertaken. These mutations also

failed to show any inhibition of Hub assembly, though they did show unusual salt effects that suggested electrostatic repulsion rather than attraction was the dominant influence between leg segments. With the postulated structure of the light chain bound to the heavy chain, a new cryo-EM docked dimeric model, the Wilbur Model, is available to show the new hypothesized interface between proximal leg segments and the salt bridges possibly involved.

Introduction

Clathrin-coated vesicles sequester the cargo of intracellular membrane trafficking processes. It has long been known that individual clathrin triskelia polymerize into a lattice or basket, a process that can be stimulated in-vitro by adaptor complexes, or by a drop to pH below 6.5 (Pearse, 1975). Early papers in the field showed that both electrostatic and hydrophobic interactions are involved in clathrin assembly (Nandi and Edelhoch, 1984). However, the molecular determinants behind clathrin assembly remained elusive. In 1998, the Ybe Hypothesis that light chain regulation of clathrin assembly occurred via salt bridges was published (Ybe et al., 1998). In the studies described here, I sought to test the current hypothesis and ultimately to identify regions and residues of the clathrin proximal leg that are critical for clathrin self-assembly using molecular biology and biochemistry.

Clathrin Hub (1074-1675), a fragment comprised of the C-terminal third of the molecule, retained the ability to trimerize, to assemble into flat lattices, and to be regulated by clathrin light chain (Liu et al., 1995). This truncated recombinant

clathrin Hub construct (1074-1675) contains only the proximal leg domain (1074-1522) and trimerization domain (1523-1675), confirming earlier experiments that demonstrated that Interactions between adjacent antiparallel proximal legs the main driving force for clathrin assembly (Blank and Brodsky, 1986; Nathke et al., 1992). Whole clathrin assembly into spherical baskets rather than flat lattices involves further interactions between antiparallel distal legs beneath the exterior proximal leg shell, as well as distal-proximal interactions, which are absent in the Hub construct (Greene et al., 2000).

Both Hub and clathrin assembly occurs spontaneously in-vitro below pH 6.5, but at physiological pH assembly is inhibited when the clathrin light chain is bound, as it is in the cytoplasm of a cell. In whole clathrin this inhibition is overcome by the interacting adaptor complexes using an unknown mechanism.

The models for clathrin assembly from 1998- 2003 and my efforts to test each model are presented here chronologically. The Ybe Hypothesis was displaced by the Roseman Model and eventually by the Wilbur Model. However, in the course of studying each model, much was discerned about the nature of clathrin heavy chain interactions involved in the self-assembly reaction.

The Ybe Hypothesis

The first model of the molecular interactions potentially involved in clathrin assembly was put forward by Ybe and colleagues in 1998 (Ybe et al., 1998). Ybe found acidic residues on the light chain which were critical for light chain regulation of heavy chain assembly, but which did not affect light chain binding to

the heavy chain. Based on titration of assembly through addition of salt, Ybe proposed that this 22 EED sequence on the light chain interacted directly through a strong salt bridge with basic residues on the heavy chain to inhibit assembly at physiological pH. In the absence of light chains, then, these same basic residues would be free to interact with acidic residues on the adjacent heavy chain proximal leg segment without competition from the light chain salt bridges, and the resulting interactions would lead to assembly at additional acidity levels. This agreed with early suggestions that amino-carboxylate salt bridges were necessary in the assembled clathrin cages (Keen et al., 1979).

In order to fully develop this hypothesis, Ybe searched for a suitable conserved basic site on the heavy chain that could serve as the salt bridge mediator in this protein. The structure of the proximal leg segment of clathrin was at this time unknown. The best structural model of the clathrin heavy chain (Nathke et al., 1992) was based on mapping antibody mapping sites. The primary sequence of clathrin heavy chain was believed to span the Hub portion of the molecule lengthwise three times. Because the primary sequence was predicted to be helical, this was modeled as an extended three-helix bundle the length of the proximal leg segment. The structure of the light chain remains unknown, but it was likewise believed to be helical, and antibody mapping studies suggested that the N and C termini both were adjacent to the knee region between the distal and proximal leg segments of the heavy chain, with a turn in the light chain somewhere near the vertex of the heavy chain (Nathke et al., 1992). Thus the light chain would contribute two more helices to the proximal leg

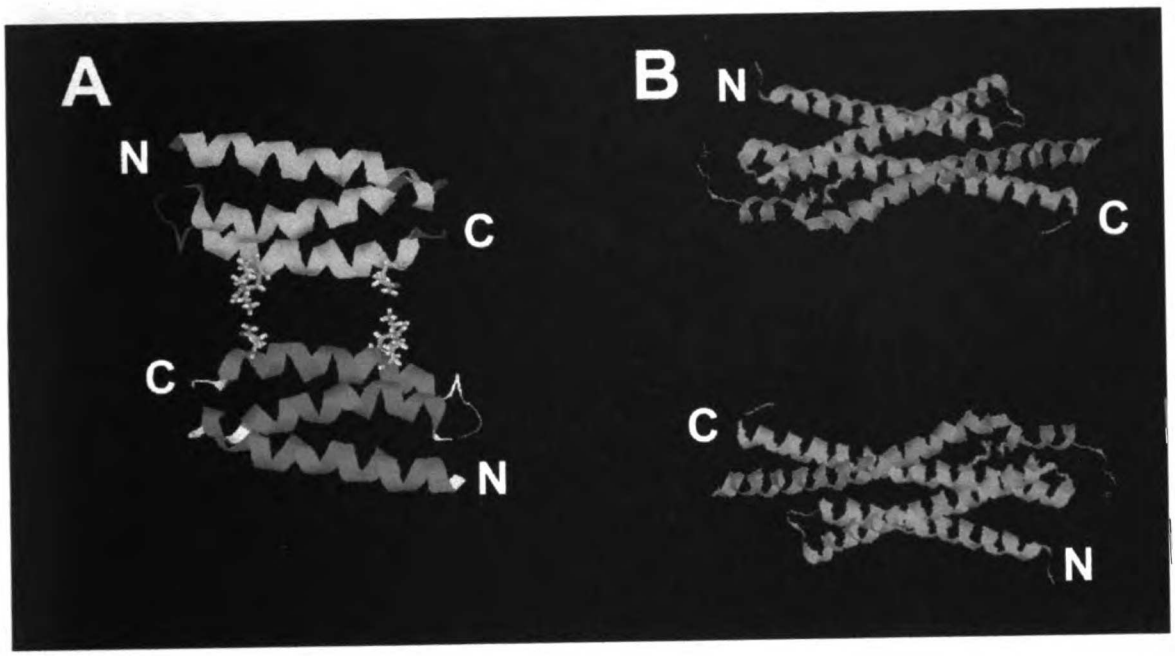


Figure 2.1: The Ybe Strong Salt Bridge Hypothesis of Clathrin Assembly

A conceptual model representing the interaction between proximal leg segments according to the Ybe Hypothesis. RasMol was used to create these images from the PDB coordinates for other unrelated proteins. **A.** Two proximal leg segments (cyan, red) in the absence of light chain are modeled as three-helix bundles, where each helix spans the length from the knee region to the trimerization domain. Basic cluster 1161 RKKAR is represented by the larger wireframe on the left of the cyan structure and the right of the magenta structure. An interacting acidic cluster is wireframed on the opposing proximal leg. PDB ID 2A3D was used for this illustration (Walsh et al., 1999). **B.** In the presence of light chain (green), the basic cluster (wireframed) on the proximal leg segments (magenta) is bound to 23 EED on the light chain (wireframed) and is not available for stabilizing heavy chain assembly. Thus in the presence of light chain clathrin assembly is inhibited at physiological pH and dependent on other interactions at low pH. PDB ID 1DN1 was used for this illustration (Misura et al., 2000).

segment, which would become a bundle of five helices. Ybe identified a basic region predicted to be near the knee of the proximal leg that was conserved in eight diverse species of heavy chain, and proposed this 1161 RKKAR as the main group of residues involved in driving clathrin assembly and interacting with the regulatory light chain sequence 23 EED. Two conserved acidic groups were found in a region of the heavy chain sequence that was proposed to lie opposite the 1161 RKKAR patch, 1132 DD and 1151 EE, and these were proposed to be the binding partners in the absence of light chain (Ybe et al., 1998). This strong salt bridge was purported to be active at all pH levels, in the absence of light chain (Figure 2.1) (Table 2.1).

The Ybe Hypothesis also predicted a second class of salt bridges of weaker affinity, whose affinity was considerably weakened at pH 6.7 compared with pH 6.2. This class of salt bridges would be able to assemble even in the presence of a competing light chain interaction. Histidines, with a pKa of 6.0-8.5, were implicated in this pH sensitivity; and three histidines at 1279, 1313, and 1335 were well conserved and were proposed to mediate this second salt bridge to acidic residues on the opposite strand. This weak salt bridge would contribute to assembly at pH 6.2, but its contribution would be absent at pH 6.7 where histidines are deprotonated (Table 2.1).

This salt bridge model was hailed as "appealingly simple" (Pishvaee and Payne, 1998) and it remained the accepted molecular model for clathrin assembly when the structure of the proximal leg was solved (Ybe et al., 1999). The mechanism by which the light chain inhibition was overcome during the .

	pH 6.2	pH 6.7 and physiological pH
With light chain	Weak histidine salt bridges	No interaction unless light chain inhibition is overcome by adaptor complex
Without light chain	Strong RKKAR salt bridge Weak histidine salt bridges	Strong RKKAR salt bridge

Table 2.1: Summary of Interactions in the Ybe Hypothesis

The proposed interactions of the Ybe Hypothesis and the conditions under which each is active are summarized here. Histidine salt bridges would be active only at low pH where histidine is protonated. The strong salt bridge interactions would be present only in the absence of competing salt bridges from the light chain. Note that under physiological conditions, neither the histidine interaction nor the strong salt bridge is active, because light chain inhibition prevents random Hub assembly in the cytosol. The interaction of the adaptor complex is required to overcome light chain inhibition, by unknown mechanism.

---

assembly process at physiological pH (when neither the strong nor weak salt bridge would be active, in this model) was presumed to be the effect of adaptor interactions, which might form additional competing salt bridges.

An alternative hypothesis (Pishvaee et al., 1997) involved light chains extending from the trimerization domain to form a molecular hinge, where the light chain is part of a four-helix bundle with the three-helix heavy chain proximal leg segment. In this model, light chains bridge between two heavy chains and assist the assembly mechanism by altering the triskelion pucker, based on experiments that relied heavily on the effects of mutations in the trimerization domain. This model did not speculate on the nature or location of the contacts

between proximal leg segments in the absence of light chain, which is physiologically omnipresent.

Mutagenesis: Testing the Ybe Hypothesis

Site-directed mutagenesis was chosen as the best way to test the Ybe Hypothesis to determine whether the conserved residues identified in the model were in fact critical for clathrin assembly. Ernest Chen created several mutant constructs of the clathrin Hub (1074-1675). I confirmed successful creation of these constructs by sequencing, and tested the effects of each mutation on the self-assembly of the recombinant mutant proteins and their regulation by clathrin light chain. The mutations chosen were divided into three groups: mutations of 1161 RKKAR, mutations of 1151 EE, and mutations of histidines (Table 2.2).

According to the model, mutation of the strong salt bridge would lead to a loss of the ability to assemble when light chain is absent under conditions where the histidine salt bridge is also absent, that is, above pH 6.5. The simplest mutation of the proposed heavy chain basic salt bridge site 1161 RKKAR was an alanine scan: mutation to 1161 AAAAA, a mutant known as Ala5. However, introduction of such a large hydrophobic patch might have other conformational consequences, so two less radical mutations were also pursued. Given the helical prediction for the protein, mutation of only one face of the helix by changing nonadjacent amino acids 3-4 residues apart would likely be sufficient. Thus M3, or Mutant 3 was created with 1161 REKAE. This charge reversal was designed to present an acidic, repulsive force on one face of the helix while

Mutant Construct Name	Residues Targeted	Mutation
Ala5	1161RKKAR	1161AAAAA
M2	1151EE	1151QQ
M3	K1162, R1165	E1161, E1165
M4	1161RKKAR	1161AAKAA
M5	H1313	Y1313
M6	H1335	Y1335
M56	H1313, H1335	Y1313, Y1335

Table 2.2: Summary of Mutations Testing the Ybe Hypothesis

The site-directed mutants were created to test the Ybe Hypothesis of the strong salt bridges involved in clathrin assembly are listed. All mutations were made in the Hub (1074-1675) recombinant clathrin fragment.

---

keeping a hydrophilic patch of residues exposed on the surface of the molecule. Finally, we created a gain of function mutation to complement the Ala5 loss of function mutation. M4, or Mutant 4, had an 1161 AAKAA mutation to restore minimal basic character and thus the ability to self-assemble to the sequence and prove that any loss of function from the Ala5 mutation was significant and not merely a structural disruption.

There was only one mutation targeted to the acidic partners of this basic site. This mutation was expected to have the same consequences as the mutation of its basic binding partner, loss of assembly when light chain is absent above pH 6.5. M2, or Mutant 2, had the targeted switch from 1151 EE to QQ. The change from glutamic acid to glutamine retains the size and hydrophobic

character of the residues but eliminates the overall charge and thus the potential to form salt bridges. Efforts to create a mutation for 1132 DD were unsuccessful and ultimately not pursued.

Two of the three identified conserved histidine residues were also targeted for mutations. M5, or Mutant 5, changed 1313 H to Y. Likewise M6, or Mutant 6 changed 1335 H to Y. Tyrosine was chosen for its similarity to histidine in both volume and polarity. If these mutations successfully disrupted the second salt bridge, Hub assembly without light chain at pH 6.7 would be expected to be unaffected, while the ability to assemble at pH 6.2 would be decreased in the absence of the histidine contribution. In the presence of light chain, when the strong salt bridges are competed out, clathrin assembly would rely entirely on this second salt bridge, so we would expect the ability to assemble at pH 6.2 to be abrogated by this mutation. Additionally, it was recognized that for a weak interaction, the effect of mutating one position may not be sufficient to abrogate the second salt bridge, while both together would have an additive effect. Thus M56, or Mutant 56, was created which incorporated both mutations 1313, 1335 HH to YY.

All mutant constructs proved expressible and were readily purified for assembly assays. Self-assembly was tested at both pH 6.2 and 6.7 in the presence and absence of light chains and compared with assembly of Hub. Ultimately, all mutant constructs were found to assemble to a similar extent to clathrin Hub in the presence and absence of light chain (Figure 2.2). Mutants were also tested in the presence or absence of calcium in the assembly buffer

(Data not shown). The presence of calcium led to higher extent of assembly, but the results were less reproducible and assembly in the absence of calcium was thus considered the standard assay.

This complete lack of significant effect on the ability of clathrin to self-assemble was attributed to several possible causes. The residues identified could have been completely uninvolved in the direct interactions between proximal leg segments during self-assembly, and could be instead a conserved binding site for regulatory or accessory proteins. Alternatively, the site could be one of several involved in self-assembly, with the mutation of this one conserved basic cluster or merely two histidines being insufficient to affect the overall binding interface enough to disrupt self-assembly to an observable extent, especially given the cooperative nature of a network of interactions in self-assembly polymerizations (Chapter Three). However, ultimately the crystal structure of the clathrin proximal leg segment was solved (Ybe et al., 1999) and found to be significantly different from the modeled structure. The 1161 RKKAR patch lies in a loop at the knee, too far from the adjacent proximal leg to mediate binding. The histidines, while near the center of the proximal leg, were partially buried in the structure. Thus the assembly results are explained by the location of mutated residues outside of the potential self-assembly interface.

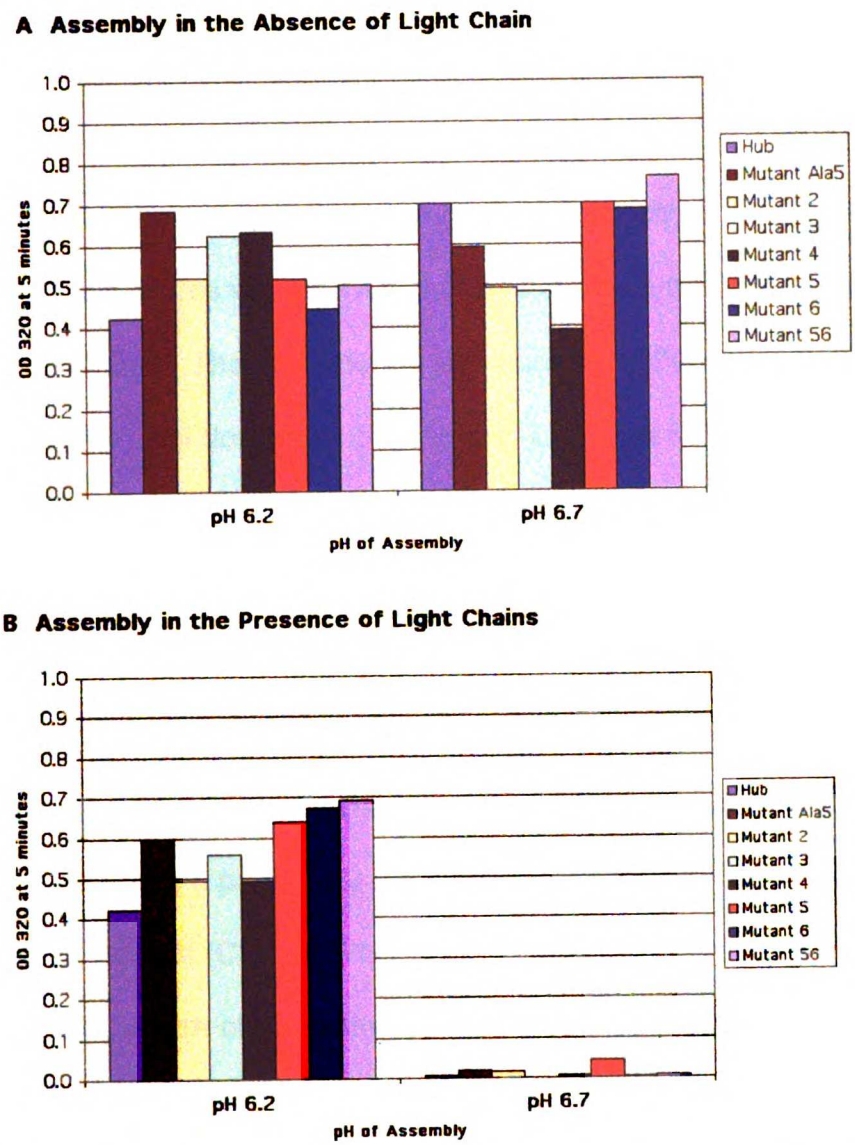


Figure 2.2: Assembly of Mutant Hub Molecules Testing the Ybe Hypothesis

Recombinant mutant clathrin Hubs were expressed and purified on nickel resin using the polyhistidine tag. The purified proteins at pH 7.9 were monitored spectroscopically at 320 nm while an aliquot of MES at pH 6.2 (left) or 6.7 (right) was added to induce self-assembly, causing measurable turbidity. After 5 minutes, Tris pH 9 was added to induce collapse of the assembled lattice, with full reversibility. The extent of assembly (plotted on the y-axis) was the change in absorbance at 320 nm after 5 minutes of assembly. **A.** Mutant Hubs were able to assemble comparably well to native Hub in the absence of light chain at pH 6.2 (left) and pH 6.7 (right). **B.** In the presence of light chain, mutant Hubs were able to assemble comparably well with native Hub at pH 6.2 (left). At pH 6.7 (right) and higher, light chains inhibited the assembly of mutant Hubs and native Hub alike. No significant effects of mutagenesis were evident in this assembly behavior.

The Crystal Structure and the Roseman Model

Just as the mutagenesis studies described above were drawing to a close, two crystal structures were solved that offered the first glimpses of clathrin at atomic resolution. The first was the structure from the Harvard group of the globular N-terminal domain of clathrin and its helical linker (1-494) (ter Haar et al., 1998) (Figure 2.3B), which was found to be a seven-bladed beta- propeller domain with a disordered alpha-helical zigzag linker. The second was the structure from the Brodsky and Fletterick laboratories here at UCSF of the proximal leg segment of clathrin (1210-1516) (Figure 2.3A) and the discovery that the entire clathrin heavy chain leg (423-1522) was composed of seven tandem repeats of a previously unknown structural motif dubbed the clathrin heavy chain repeat (CHCR) (Ybe et al., 1999) (Figure 2.3A,C).

The structure of the clathrin proximal leg segment (Figure 2.3A) was found to be a superhelix of helices: a box composed of two major faces composed of parallel helices and two minor looped faces joining them. The proximal leg (1074-1522) is projected to contain three tandem CHCR motifs, where each CHCR has 10 helices (5 helix hairpins in this helix-turn-helix-loop arrangement) in approximately 145 residues, where each helix is 28 angstroms long and each loop about 24 angstroms long.

Combining these crystallography results with a high-resolution cryo-electron microscopy map (Smith et al., 1998) allowed visualization of the orientation of the different leg segments in the assembled baskets and much

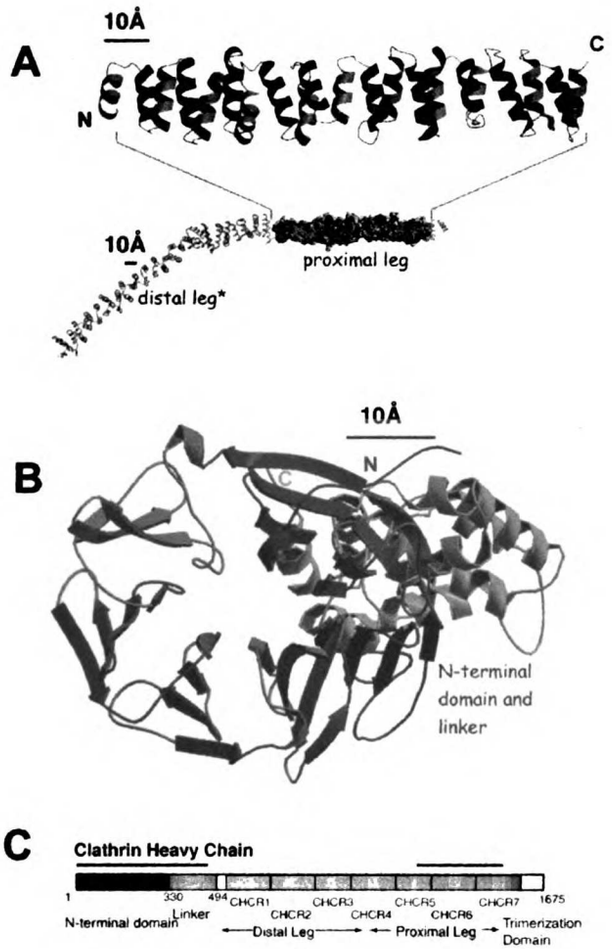


Figure 2.3: Crystal Structures of Clathrin
Molscript and Raster3D were used to generate these images from clathrin heavy chain PDB coordinates. **A.** Clathrin proximal leg segment (1210-1516) mediates lattice assembly. The structure is an elongated rod made up of an extended alpha/ alpha superhelix. Ten helices blocks form tandem structural domains called CHCR (clathrin heavy chain repeat). Alignments indicate that seven tandem CHCR domains form one clathrin leg (lower inset). PDB ID: 1B89 (Ybe et al., 1999). **B.** The globular N-terminal domain (1-494) projects toward the vesicle membrane to interact with the adaptor complex and other accessory proteins. This domain is a seven-bladed beta-propeller with a helical flexible linker (gray, right). Each blade of the propeller is a slightly twisted antiparallel beta sheet. PDB ID: 1BPO (ter Haar et al., 1998). **C.** Clathrin has an N-terminal domain, a helical linker, distal and proximal leg segments, and a C-terminal trimerization domain. The black lines above the bar representation indicate the portion of the protein included in the molecular structure solutions above. These images appeared in a previous publication (Wakeham et al., 2000) and are reproduced here with permission from Blackswell Munksgaard Publishing.

closer estimation for the location of the residues involved in the previous mutagenesis study. In the assembled form, the two proximal leg segments cross at a 20 degree angle (Smith et al., 1998) (Musacchio et al., 1999) concentrating the interacting surface in the center of the proximal leg segments and leaving the knee and vertex region free from self-interactions.

The 1161 RKKAR sequence targeted in the first round of mutagenesis above and its potential acidic binding partners were too close to the knee region in the assembled structure to reach close proximity with the adjacent proximal leg. However, the knee region does have a close approach with both the trimerization domain vertex and the linker region near the globular N-terminal domain. So while the crystal structure rules out a strong interaction between proximal leg segments involving the basic patch at 1161, it is quite possible for other self-interactions or binding interactions with other proteins to involve this region. The histidines targeted in the mutagenesis study, by contrast, were buried (H1335) or partially buried (H1313) residues that were not sufficiently surface accessible to allow their participation in the proposed salt bridges. The crystal structure therefore incontrovertibly changed the identity of the proposed contacts between proximal leg segments, but did not alter the underlying salt bridge theory put forth by Ybe.

While the crystal structure of the trimerization domain is unknown, the CHCR motif extends into the trimerization domain and the solved structure allowed estimation of the location of the mutations that were critical in the Pishvaee model (Pishvaee et al., 1997). Most of the mutations are likely to occur

on buried surfaces within the superhelical trimerization domain structure. While this does not rule out the concept that the light chain binding site extends into the trimerization domain, it does cast doubt on the specificity of the trimerization domain – light chain contacts found and the mutagenesis effects.

The crystal structure of the N-terminal domain was docked into the cryo-EM map to enable a visualization of the specific location of residues within the assembled structure (Musacchio et al., 1999). Unfortunately the rod-like proximal leg segment, with no prominent bulges or asymmetrical features, was easily docked into the cryo-EM map in any orientation and required more sophisticated algorithms to dock the crystal structure into the cryo-EM grid. The fit was especially problematic because the light chain density was included in the cryo-EM image, but the crystal structure of the proximal leg with light chain bound was unavailable. A collaboration was undertaken by Alan Roseman of MRC Cambridge and Peter Hwang of UCSF to dock the UCSF crystal structure into the Pearse EM using Roseman's algorithm (Roseman, 2000) (Figure 2.4). While a perfect alignment was elusive, the docked structure appeared to be accurate to within the twenty angstrom resolution limit of the cryo-EM and appeared to have one clear preferred orientation, in which the closest point of approach between two proximal legs was in residues 1331- 1337, with extensive contacts in CHCR6. GRASP was used to calculate the molecular surface area of the interface created in the docked structure, and it was found to encompass an area larger than 1500 square angstroms, a large area typical of many extended binding sites.

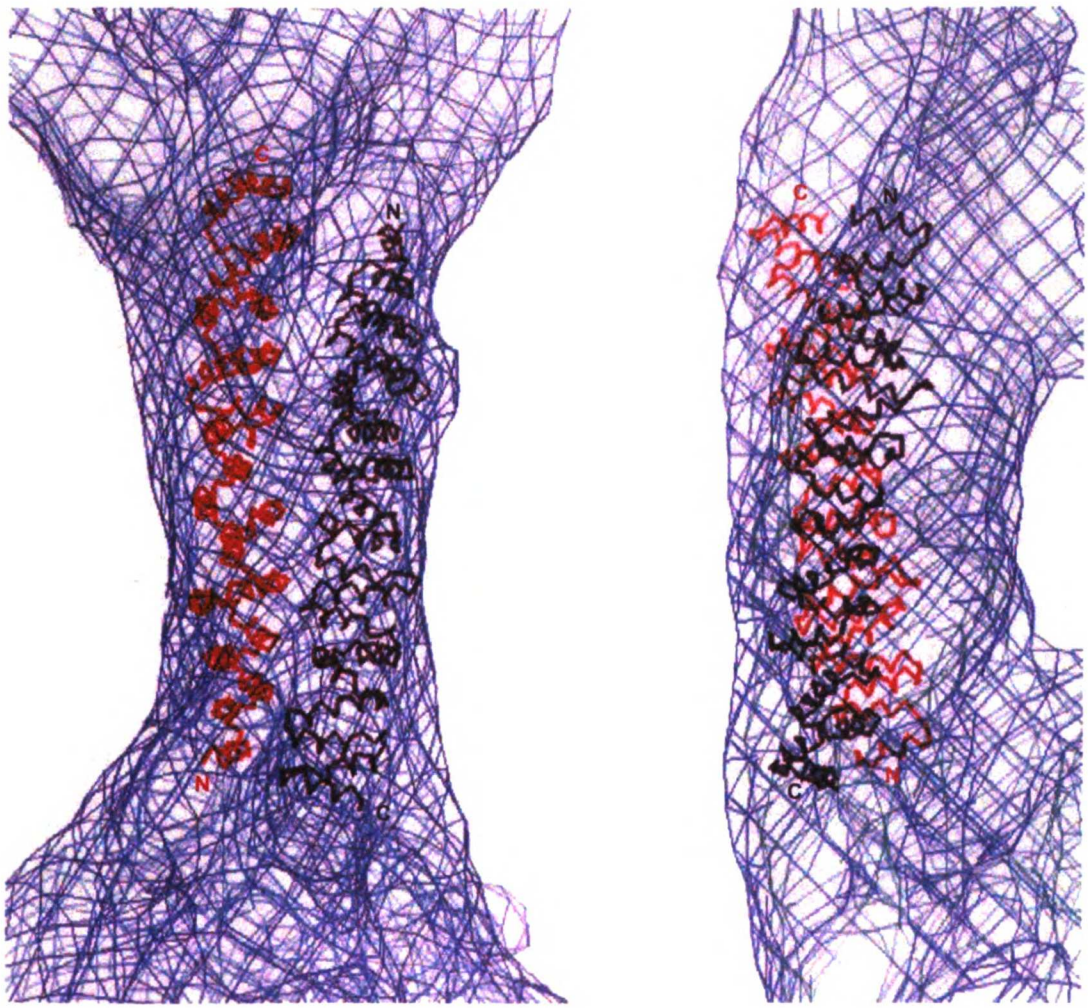


Figure 2.4: The Roseman Model for the Assembled Proximal Leg Interface

Two proximal leg crystal structures (red, navy) were docked into the cryo –electron microscopy image (cyan contours) of the assembled clathrin cage in order to model the interface between adjacent proximal leg segments in the assembled form. From the outside of the basket (left), the proximal legs are adjacent and antiparallel. However, a ninety degree rotation to view a lateral cross-section along one side (right) shows that the proximal legs cross at a 20 degree angle, with most contact area between the adjacent leg segments confined to the center of the interface.

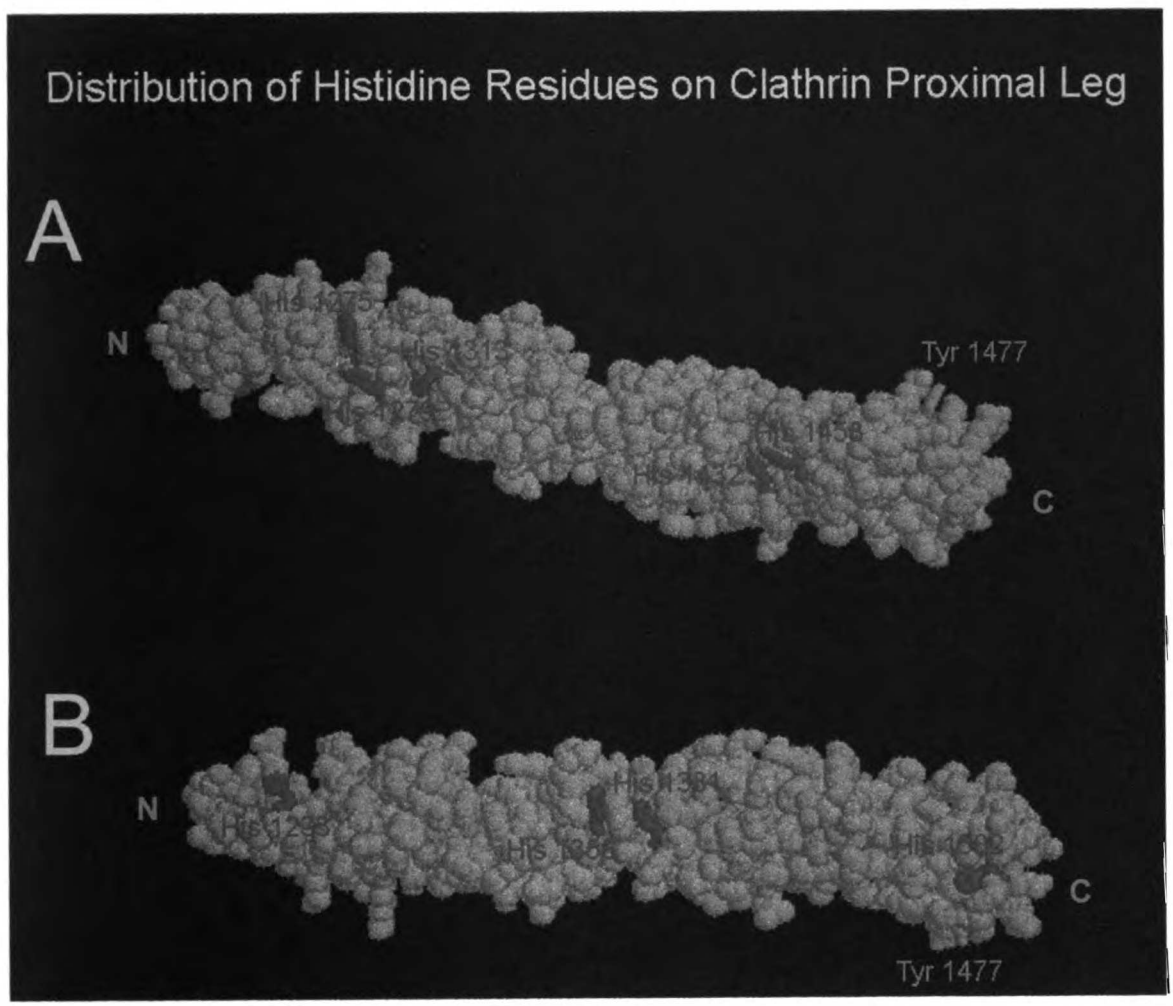


Figure 2.5: Distribution of Surface- Accessible Histidines on Clathrin Proximal Leg Segment
Histidines (blue) on the surface of the proximal leg segment (gray) as visualized by RasMol show an unexpected alignment along the two sides of the molecule. **A.** Along the inside surface of the interface, (assuming the tyrosine phosphorylation site Y1477 (green) points toward the cytosol), histidines are in two symmetrically spaced clusters. **B.** On the opposite looped face of the helix, forming the outside surface of each pentagonal or hexagonal leg segment, histidines are located in the center and at the ends of the side in a linear pattern. Proximal leg crystal coordinates are from PDB ID: 1B89 (Ybe et al., 1999).

The crystal structure of the proximal leg segment and the Roseman Model offered a plethora of new approaches to studying the possible mechanism of clathrin assembly. Three approaches are described in this dissertation. First, the role of histidines, first proposed in the Ybe Hypothesis, was investigated to explore whether histidines location in the crystal structure allowed a possible major role in self-assembly, and to determine how histidines might contribute to leg segment affinity. Next, efforts to locate the critical residues involved in self-assembly through minimization of interacting clathrin leg fragments is described. Finally, specific residues predicted to be crucial for self-assembly in the Roseman Model were targeted by site-directed mutagenesis.

The Role of Histidines

Hub, like clathrin, spontaneously assembles in-vitro in the presence of light chain below pH 6.5. A pH dependence in this range suggests the involvement of histidine residues (pKa 6.0-8.5) (Christianson, 1991; Ybe et al., 1998). Histidines have mediated pH switches by conformation changes or other mechanisms in many proteins (Borza et al., 1996; Hanakam et al., 1996; Schnizer et al., 1996). In glutamine synthetase, self-assembly involves metal ion-controlled supra-assembly of identical subunits, mediated by histidines (Dabrowski et al., 1994).

The crystal structure of the proximal domain of clathrin heavy chain revealed conserved surface-accessible histidines in near-linear alignment down two of the four faces (Figure 2.5) of the rod-like structure. Histidines 1233, 1356,

1381, and 1502 line up on one side of the molecule, with 1502 conserved between species. Similarly, Histidines 1275, 1279, 1313, 1432, and 1458 run down the opposite side of the molecule, with 1279 and 1313 conserved between species (Table 2.3). While the overall histidine content of the protein at 2.1 mole percent is not unusually high, the unusually high distribution of histidines on the surface (Meirovitch and Scheraga, 1980) and their alignment along exposed loop region between helical pairs is notable. Moreover, using the CHCR motif to predict the position of histidines on the distal leg suggests a similar distribution of histidines at the surface, primarily on only one face of the distal leg segment (Figure 2.6). Many of these histidines are found in close proximity with carboxylic acid residues, often indicative of a metal binding site (Christianson and Alexander, 1989), which also implicates a role in assembly.

The histidines within clathrin's proximal leg segment would be expected to be protonated during in-vitro assembly experiments, but deprotonated within the cell during physiological assembly, although local environments can cause significant pKa changes. Thus, histidine may control the pH dependence of clathrin assembly, through the proposed salt bridges when protonated, or through metal binding when unprotonated, or both. Assembly may be regulated by variations in histidine pKa due to factors in the local environment, or possibly by changes in the local concentrations of regulatory metal ions in the cell.

The Ybe Hypothesis suggested that in the presence and absence of light chain, charged histidine interactions provide a crucial interaction driving clathrin assembly. To explore the role of histidines in clathrin self-assembly, I attempted

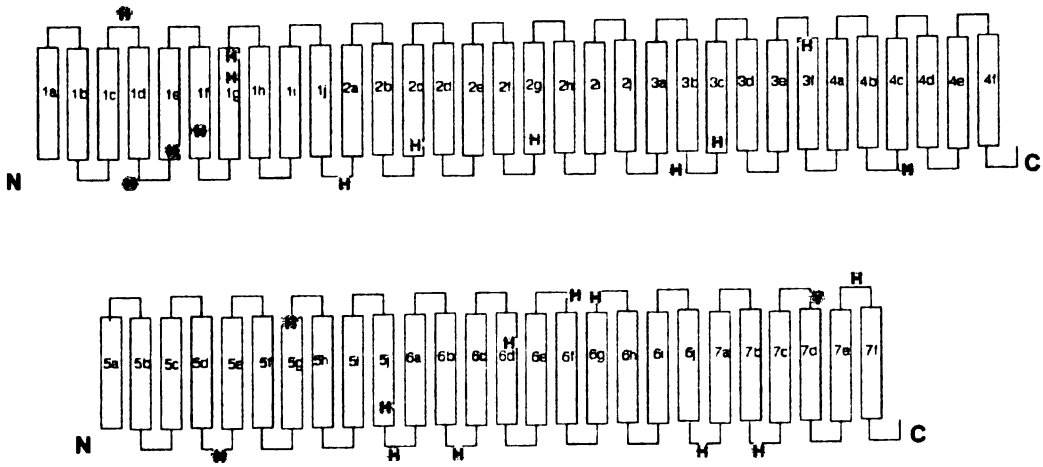


Figure 2.6: Predicted Distribution of Histidines Along the Entire Clathrin Leg

Based on the CHCR motif, the approximate location of each loop and helix along the entire length of the clathrin leg is known, despite the lack of a crystal structure for the distal leg segment. Helix numbers (1-7) refer to the tandem CHCR domains, and letters refer to which helix within that CHCR (a-j). The location of histidines within these helices and loops is denoted by an H in an orange circle. The phosphorylation site Y1477 is included as a reference point and denoted as a Y in a green circle. The distal leg segment (upper) contains a few histidines buried within CHCR1, but all other histidines are likely to localize to the surface-accessible loops or the very end of a nearby helix. One face of the rod-like distal leg segment is likely to have a linear alignment of histidines, possibly for functional use in interacting with the proximal leg segment that lies directly above it in the assembled basket. The opposite looped face of the distal leg has single symmetrically distributed histidines at either end. The proximal leg segment (lower) has only one buried histidine within CHCR5. The other histidines form a more or less symmetrical pattern on both looped faces around a center of helices 5f/ 5g.

Histidine	1233	1275	1279	1313	1356	1381	1432	1458	1502
human	H	H	H	H	H	H	H	H	H
bovine	H	H	H	H	H	H	H	H	H
rat	H	H	H	H	H	H	H	H	H
fly	Y	H	H	H	H	H	H	L	H
worm	R	H	H	H	H	H	H	L	H
slime	K	N	H	H	Q	H	H	Q	H
s. cer	Y	N	H	H	H	K	I	K	H
schpo	Y	N	H	H	H	Q	H	L	H
CHC22	H	H	H	H	H	H	H	H	H

Table 2.3: Conservation of Histidines in Proximal Leg Segment Interface

An alignment of clathrin heavy chain protein sequences from all species known in 1999 was created and used to evaluate conservation of surface-accessible histidines in the proposed interface between proximal legs according to the Roseman Model. Note that Histidine 1313 was previously mutated during testing of the Ybe model (Mutant 5).

a chemical modification of histidines on clathrin Hub in the absence of clathrin light chain. Caution must be taken in the interpretation of these results, since the recombinant fragment used had an N-terminal polyhistidine tag. Chemical modification of the accessible histidine residues with diethylpyrocarbonate (DEPC) in a clathrin Hub was expected to disrupt histidine salt bridges or interactions with metal. Failure to assemble the modified protein fragments, followed by assembly when the modifications are removed, would have confirmed a role for histidines in clathrin assembly.

DEPC chemical modification of clathrin Hub without light chain resulted in similar extent of assembly as untreated Hub. However, formation of the histidine-DEPC conjugate could not be quantitated. The DEPC reaction with buffers gave minimal absorption at 320 nm that could interfere with detection of assembly behavior. However, the assembly reaction itself appears to contribute most or all of the rise in absorption at 237 nm, where the conjugate formation would be observed. Also, DEPC did cause massive interference to observation of the assembly reaction at when combined with Tris buffer. The pH of the assembly reaction may not be favorable for histidine chemical modification under these conditions. A better approach may be mutagenesis of conserved surface-accessible histidine residues. It should be noted, however, that the conserved H1313 has been previously mutated during testing of the Ybe model (Mutant 5). Therefore mutation of multiple histidines on that side of the molecule, or beginning with mutation of one or more histidines on the opposite side of the proximal leg, would be appropriate. The experiments described here were inconclusive on the role of histidines in clathrin self- assembly.

The Role of Divalent Metal Ions

Metal ions are commonly involved in regulatory activity in the cell, and often play a role in assembly processes, particularly those involving cysteine or histidine residues. Metal binding sites are frequently engineered into protein interfaces to modulate protein-protein interactions. This strategy relies on mimicry of the naturally occurring metal sites at the interface of many protein

complexes such as insulin multimers, human growth hormone dimers, and glutamine synthetase (Matthews, 1995). Typical metal-ligand bond distances are on the order of 2 angstroms. Zinc, copper, and nickel have similar radii and properties such as a moderate hardness or polarizability which easily ligands to nitrogen, oxygen, or sulfur in proteins. Zinc in particular tends to bind to negatively charged residues such as carboxylates and thiolates, as well as neutral carbonyls and imidazoles (Christianson, 1991). Charged proteins can have appreciable long-range interactions with metal ions, so a physiological effect is quite possible.

The pH dependence of clathrin assembly led to the Ybe Hypothesis that clathrin assembly is controlled by two salt bridges between adjacent proximal legs. A titration of clathrin Hub assembly with sodium chloride suggested a low-affinity histidine salt bridge that is very sensitive to pH, with a midpoint of 50 mM NaCl at pH 6.2 and a midpoint of 175 mM NaCl at pH 6.7 (Ybe et al., 1998). Additional experiments suggested that much lower levels of zinc (2-5 mM) inhibited clathrin Hub assembly (Ybe et al., 1999; Ybe et al., 1998), suggesting a specific divalent metal interaction beyond the documented ionic strength effects (Nandi and Edelhoch, 1984; Van Jaarsveld et al., 1981). While the surface of clathrin proximal leg segment does not contain free cysteines, numerous potential metal interaction sites with histidines, glutamate, and aspartate exist. It seems possible that divalent cations could have a physiological regulatory role in clathrin assembly.

To assess this possibility that divalent ions controlled assembly, zinc titrations of clathrin Hub assembly without light chain were repeated in metal-free EPPS buffer to identify any pH dependence (Figure 2.7). Very low levels (5-100 μM) of zinc (on the order expected for specific interactions) did not appear to inhibit assembly below pH 6.5 in the absence of light chain. Note that typical zinc concentrations intracellularly vary from estimates of 0.4 femtomoles/cell in a fibroblast (Suhy et al., 1999) to 2.4×10^{-11} molar in a red blood cell (Simons, 1991). Zinc levels above 150 micromolar are toxic in Hep2 cells (Rudolf et al., 2003).

When zinc is added at higher concentrations (1-5 mM) to clathrin Hub, massive aggregation occurs. This result appears similar to the effect of eluting clathrin Hub from a nickel resin column without adding EDTA before overnight dialysis, which causes the protein to precipitate, presumably due to the presence of excess nickel leached from the column. Aggregation of Hub under these circumstances may be mostly reversed by addition of 1:1 volume ratios of 500 mM EDTA or ethanolamine, but reversibility is not readily achieved. (Preparations of proximal leg do not show any propensity to aggregate when eluted from nickel columns, and no EDTA is required before the proximal leg segment dialysis.) Additionally, addition of 3 mM calcium to the assembly reaction has been observed to give higher extents of assembly while decreasing reproducibility and full reversibility (Liu et al., 1995). Zinc also did not appear to significantly enhance assembly above pH 6.5 at low levels, which suggested that metal ion interactions at higher concentrations were nonspecific or that high

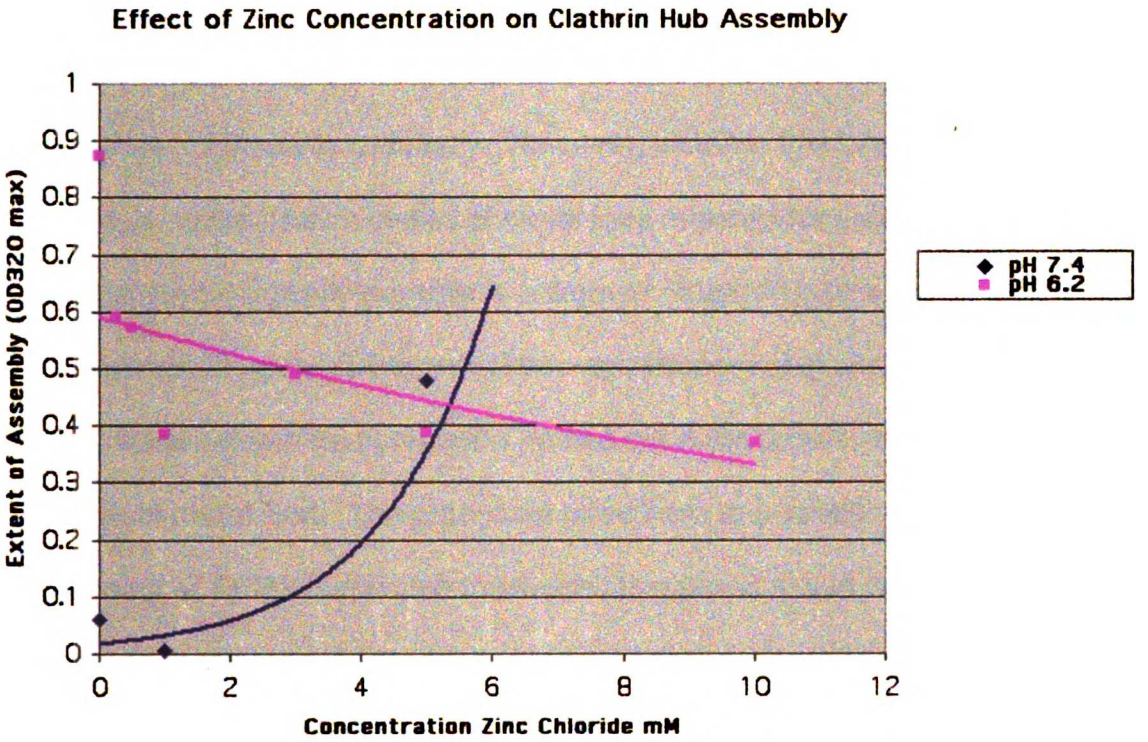


Figure 2.7: The Effect of Divalent Metal Ions on Clathrin Hub Assembly
At pH 6.2, addition of <10 mM of zinc to an assembly assay appears to inhibit the extent of clathrin Hub self- assembly in the absence of light chain slightly. However, at higher pH, zinc causes massive aggregation, as seen by the exponentially rising absorbance, which was irreversible under normal assembly assay conditions.

affinity metal binding sites were already fully occupied.

Given the overall negative charge of the proximal leg, excess metal ions could electrostatically shield the repulsion between proximal legs to cause aggregation or possibly assembly. However, whether this interaction is specific or whether concentration spikes of metal ions required for such an interaction would be physiologically possible is unknown. Clathrin preparation would need to be attempted under strictly metal-free conditions, using a mechanism other than a nickel column for purification, to find out if the protein is stable in the absence of metal ions. The procedure used here to prepare metal-free buffer (incubation of EPPS buffer with fresh chelating resin) would not have removed metal ions that were already tightly bound to clathrin from the purification protocol. The subject merits further exploration.

One possible approach for further study would be treatment of Hub with zinc chelator 4-(2-pyridylazo) resorcinol (PAR), which forms a colored complex with zinc that may be observed at 492 nm. Alternatively, purification of clathrin fragments in the presence of radiolabelled 65Zn would allow detection of the radioactive tightly bound metal. Copper binding proteins can be distinguished spectroscopically from other proteins by their intense absorption at 625 nm and irregularities in Electron Paramagnetic Resonance (EPR) spectra. Finally, EXAFS or XANES could be used to determine the local environment around a metal ion such as zinc or copper, with examination of the effects of different pH conditions (Hansen and Garner, 1987). These experiments were inconclusive on the role of divalent metal ions on clathrin assembly.

Minimization of the Interface

Sequence analysis indicates CHCRs comprise the entire distal and proximal legs of clathrin as well as part of the C-terminal trimerization domain (Ybe et al., 1999). Thus the interactions of clathrin heavy chain assembly, such as proximal-proximal and proximal-distal interactions, clearly are interactions between adjacent CHCR modules, which suggested that the function of CHCR is to mediate protein-protein interactions.

CHCR domains are found in other proteins also, present once or twice per protein, predominantly in proteins with unknown functions in trafficking pathways (Conibear and Stevens, 1998; Odorizzi et al., 1998; Ybe et al., 1999). In particular, Vps41p has been proposed to also function as a coat protein in TGN to lysosomal traffic (Conibear and Stevens, 1998) and alpha-COP, a coat protein for Golgi to ER retrograde traffic, may also contain a CHCR. Preliminary searches for CHCR-containing proteins using Pfam analysis yielded several human proteins of unknown function. CHCRs are present in KIAA0590, KIAA0770 (Vps39 homologue), KIAA0796, and KIAA0804 (Vps8 homologue) from the HUGE protein database of Kazusa DNA Research Institute. A CHCR is also found in human Vps41 homologue (SWISSPROT p49754)..

Interestingly, each of five confirmed yeast proteins with CHCR domains had known direct or indirect interactions with one or more other members of the set. These interactions range from genetic interactions to physical binding (Table 2.4).

	Clathrin	Vps41	Vps8	Pep5
Clathrin	Physical interaction in all species (Brodsky et al., 2001)			
Vps39		Possible physical interaction in yeast (Nakamura et al., 1997)		
Pep3		genetic interaction in fruit fly (Warner et al., 1998)		physical interaction in yeast (Rieder and Emr, 1997)
Pep5			genetic interaction in yeast (Woolford et al., 1990)	

Table 2.4: Interactions between proteins with Clathrin Heavy Chain Repeats (CHCRs)

Proteins with CHCRs were identified (Ybe et al., 1999) and I carried out literature searches to discover the functional interactions of the proteins. Interestingly all of the proteins with known functions had a role in vesicle traffic, and literature included reports of genetic or physical interactions between them. We hypothesized that these interactions were mediated by the CHCR domain.

We hypothesized that the CHCR domain was a scaffold upon which the specific determinants for different protein-protein interactions might be arranged. Possible binding interactions included self-assembly, protein interaction between CHCRs, or CHCR binding to one of the other domains present in trafficking

proteins (Ybe et al., 1999). Based on the biochemical data presented above, I hypothesized that CHCRs were the oligomerization domains mediating clathrin assembly and possibly also the assembly of other protein complexes with a function in vesicular trafficking.

The primary determinants for clathrin assembly are within the Hub, which contains CHCRs 5, 6, and 7. Most of the interface between adjacent proximal legs in the assembled basket, however, lies within CHCR6 (Figure 2.8). I attempted to determine whether a single CHCR module is sufficient for clathrin interaction. Recombinant purified CHCR6 from bovine clathrin was created with the intention to assay its self-association and binding to clathrin Hub. However, the CHCR6- 6XHis proved almost insoluble. Initial results suggesting solubility were eventually determined by Giselle Knudsen using mass spectrometry to be the presence of contaminant protein beta-lactamase. In order to increase solubility, fusion proteins were created with CHCR6 in other plasmids. A Thioredoxin-CHCR6 product also proved insoluble, and an MBP-CHCR6 (Maltose binding protein) construct yielded only soluble MBP. Ultimately, expression and purification of CHCR6 was not successful. However, proximal leg segment, spanning all of CHCRs 5 and 6 and part of CHCR7, was easily purified. Studies to minimize the interface between Hubs using proximal legs rather than CHCR6 proved fruitful (Chapter Three).

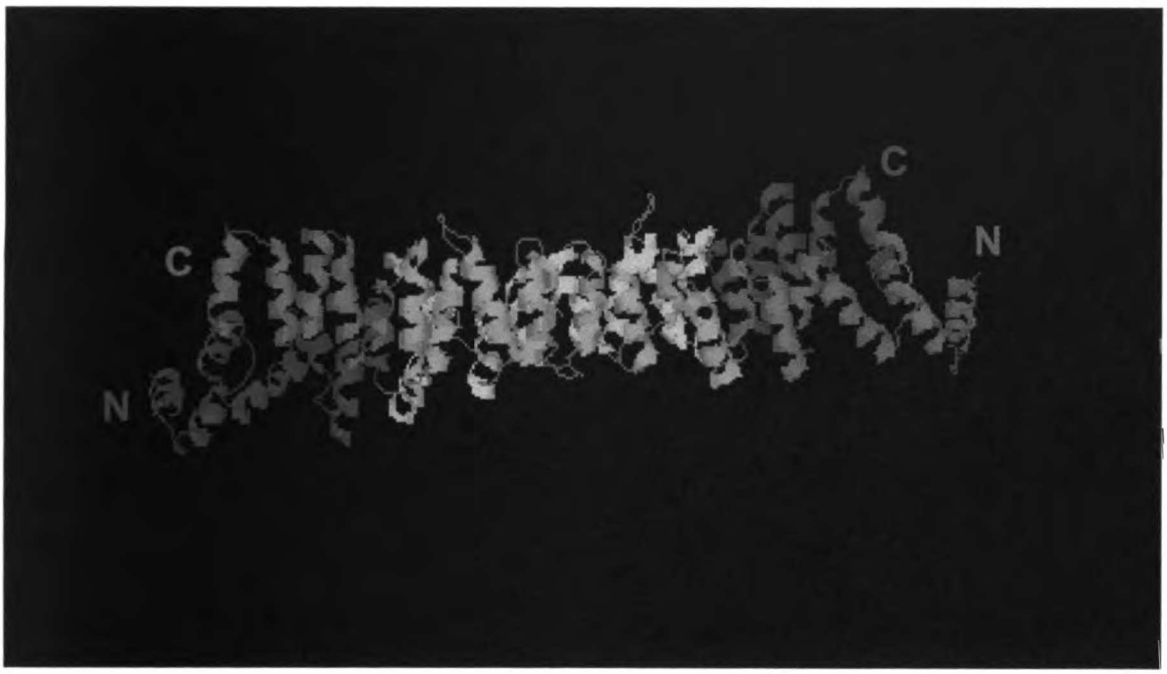


Figure 2.8: CHCR6 in the Roseman Model of the Proximal Leg Segment Interface
The docked proximal leg segments in the Roseman Model (shown in reddish hues and bluish hues) each contain portions of CHCRs 5 and 7 and the complete CHCR6. CHCR6 has been highlighted in each (cyan and magenta, center). Because the leg segments cross each other at an angle, more than half of the surface area of the interface between leg segments is from contacts between adjacent CHCR6 domains.

Mutagenesis: Testing the Roseman Model

An alignment of the clathrin heavy chain in twelve diverse species was used to determine the conservation of acidic and basic residues within the face-to-face contact area proposed by the Roseman Model (Table 2.5). Only seven bases were found within reach of the partner proximal leg in this arrangement, and only two had absolutely conserved basic character, R1267 and K1331. By contrast, fifteen acidic residues were found in the proposed interface, and four had absolutely conserved acidic character, D1242, D1298. E1304, and E1334.

I plotted the charged residues in the interface on the structural model of the dimeric interface (Figure 2.9), and scrutinized the distances between potential salt bridge pairs (Table 2.6). In the structural model of the dimeric interface, R1267 aligned most closely (within 3- 10 angstroms) with D1421, which was a conserved aspartatic acid in every species known at the time except malaria. K1331 was more promising: its presence and the existence of R1333 nearby in the central helix and closest point of approach put both within 10 angstroms of E1334 and E1337, well within range of binding. R1333 was conserved in insects and higher organisms, but was widely divergent in lower eukaryotes. By contrast, E1337 was conserved acidic in insects and higher organisms, but retained a basic charge in lower eukaryotes. Keith Burdick did an analysis to look for convergently evolved salt bridge pairs, but was unable to find a match for E1337 or any new nonconserved pairs. Hence, the best conserved pair and the point of closest approach of the helices was K1331- E1334.

BASES

SPECIES	R1267	K1328	K1331	R1333	K1387	K1449	R1453
CHC17 Ho. Sapiens	R	K	K	R	K	K	R
Bo. taurus	R	K	K	R	K	K	R
Ra. norvegicus	R	K	K	R	K	K	R
Ca. elegans	R	K	K	R	R	K	R
Dr. melanogaster	R	K	K	R	R	K	R
Ae. aegyptii	R	K	K	R	R	K	R
Di. dicdi	R	K	K	M	D	K	V
Ar. thaliana	R	R	K	M	E	K	V
Gl. max	R	R	K	M	D	K	V
Pl. falciparum	K	K	K	M	T	Q	E
Sa. cerevisiae	K	E	K	F	D	K	I
Sc. pombe	R	K	R	M	D	L	V
CHC22 Ho. sapiens	R	K	K	L	K	K	R

Table 2.5: Conservation of Potential Salt Bridge Residues

An alignment of clathrin heavy chains from all species known in 1999 was created using GCG software and used to evaluate the conservation of surface-accessible charged residues on the predicted interacting face of the proximal leg segment. This data was combined with distances between pairs (Table 2.6) to select targets for mutagenesis.

ACIDS

SPECIES	D 1212	E 1236	D 1242	E 1265	D 1284	E 1297	D 1298	E 1304	E 1334	E 1337	E 1360	D 1366	D 1384	E 1388	D 1393	D 1421
CHC17 Ho. sapiens	D	E	D	E	E	E	E	E	E	E	E	D	D	E	D	D
Bo. taurus	D	E	D	E	E	E	E	E	E	E	E	D	D	E	D	D
Ra. norvegicus	D	E	D	E	E	E	E	E	E	E	E	D	D	E	D	D
Ca. elegans	D	E	D	E	E	E	E	E	E	E	E	D	E	E	E	D
Dr. melanogaster	D	E	D	E	E	D	E	E	E	E	E	D	E	E	D	D
Ae. aegyptii	E	E	D	E	E	E	E	E	E	E	E	D	E	E	D	D
Di. dicdi	E	Q	D	E	E	N	E	E	E	K	Q	I	E	H	E	D
Ar. thaliana	E	Q	D	E	E	N	E	E	E	K	E	I	E	H	D	D
Gl. max	E	Q	D	E	E	N	E	E	E	K	E	I	E	H	D	D
Pl. falciparum	E	E	E	Q	D	N	E	E	E	R	E	I	T	A	Q	N
Sa. cerevisiae	K	D	D	E	D	E	E	E	E	K	E	A	K	H	E	D
Sc. pombe	E	E	D	E	P	E	E	E	E	K	E	V	E	H	D	D
CHC22 Ho. Sapiens	E	E	D	E	E	E	E	E	E	E	E	D	E	E	D	D

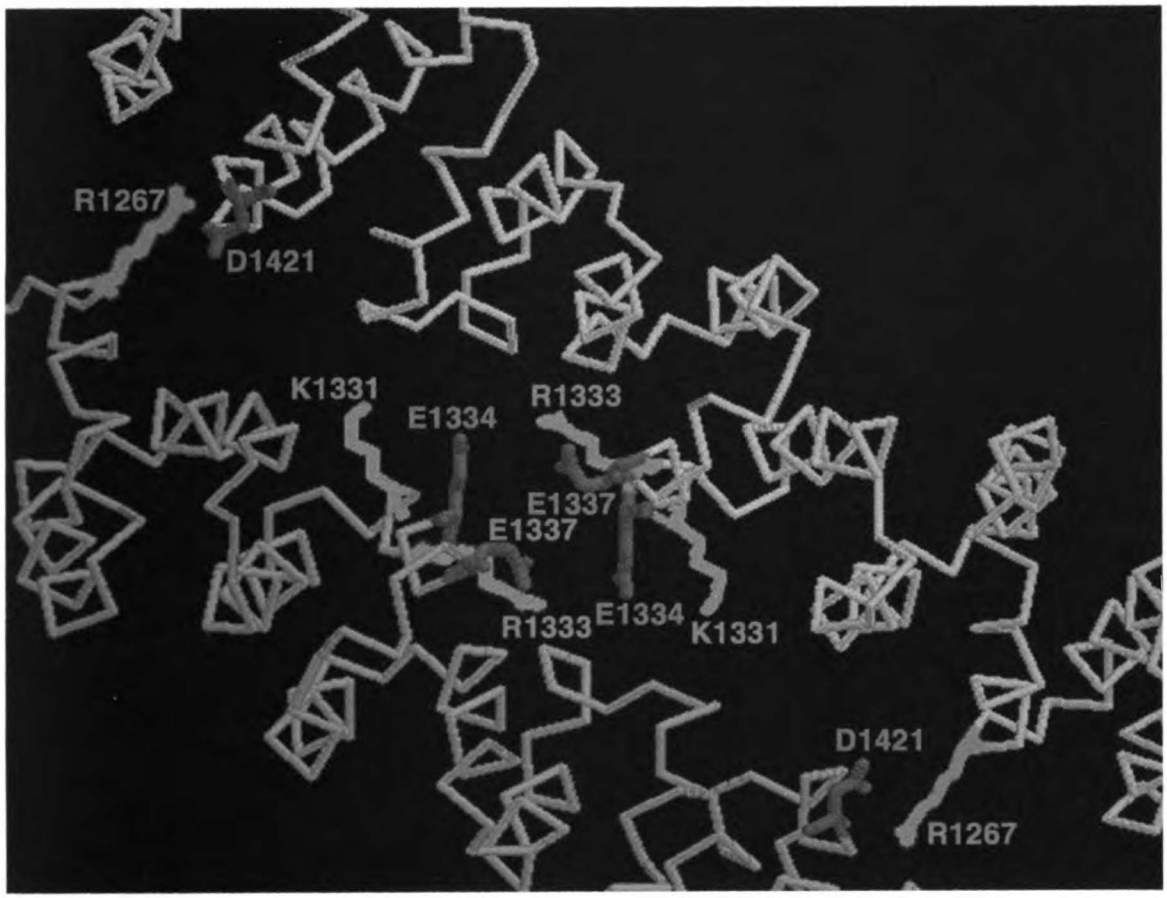


Figure 2.9: Potential Salt Bridges in the Roseman Model Interface Between Proximal Leg Segments
Enlargement of the Roseman Model docked interface between two proximal leg segments (cyan and magenta) in RasMol shows possible salt bridge interactions wireframed. K1331, R1333, E1334, and E1337 are at the closest point of approach between the two leg segments. Though farther from the center, R1267 and D1421 are another promising pair.

		Dock 1 (Older)		Dock 2 (Newer)				Intramolecular Bridges to Non- Conserved Residues		
Basic Ion	Partner	Chain A Dist. Å	Chain B Dist. Å	Chain A Dist. Å	Chain B Dist. Å	Avg Dist. Å	St. dev	Basic Ion	Partner	Dist. Å
R1267	D1421	11.1	3.8	9.2	3.3	6.9	3.9	R1267		
	E1388	15.7	10.3	10.1	3.7	10.0	4.9			
	D1393	14.0	14.6	4.2	7.4	10.1	5.1			
R1333	E1304	15.6	15.9	8.8	7.5	12.0	4.4	R1333	E1337	3.3
	E1334	6.0	8.0	5.1	6.1	6.3	1.2		E1360	4.2
	E1297	22.7	18.1	14.8	10.8	16.6	5.0		D1366	8.3
	E1337	8.9	11.2	13.3	14.2	11.9	2.4		D1384	10.1
	E1360	14.6	11.3	15.4	14.1	13.9	1.8			
K1331	E1334	19.6	21.3	10.9	12.1	16.0	5.2	K1331	E1334	6.7
	D1393	7.0	14.7	12.5	15.7	12.5	3.9		E1304	4.7
	E1360	17.0	15.8	8.1	7.9	12.2	4.9		E1297	8.9
	E1388	9.3	6.7	14.5	14.0	11.1	3.8			
	D1366	12.1	17.8	10.5	14.6	13.8	3.2			
	E1337	12.7	16.6	4.5	8.8	10 7	5.2			
	D1384	19.5	12.3	13.1	9.9	13.7	4.1			
K1387	E1334	2.3	12.7	10.2	14.5	9.9	5.4	K1387	D1393	13.5
	E1298	21.1	20.8	12.8	10.9	16.4	5.3		E1360	9.3
	E1304	9.8	15.4	6.6	8.1	10.0	3.8		D1384	8.3
	E1337	10.5	17.3	18.1	20.1	16.5	4.2		D1388	10.7
	E1297	15.1	13.2	7.8	3.8	10.0	5.1		E1337	11.1
R1453	D1212	15.8	19.3	15.2	8.0	14.6	4.7	R1453	E1474	4.8
	D1242	9.0	18.5	16.3	16.1	15.0	4.1		D1476	4.7
	E1236	9.8	15.5	12.1	7.8	11.3	3.3			
	E1265	9.4	9.6	16.3	12.8	12.0	3.2			
K1449	E1236	12.0	19.8	13.3	11.4	14.1	3.9	K1449	E1494	2.5
	E1265	7.6	11.9	14.6	12.2	11.6	2.9		D1476	6
									E1474	7.5
K1328	E1334	22.8	26.6	15.2	17.6	20.6	5.1	K1328	E1297	2.6
	E1337	17.3	23.0	11.0	15.1	16.6	5.0		E1304	6.6
	E1388	5.7	10.3	10.8	13.4	10.1	3.2		E1334	12.9
	D1393	10.6	20.5	14.3	19.4	16.2	4.6		E1360	15.3
	E1360	19.8	21.5	10.8	13.3	16.4	5.1		E1298	10.8
	D1421	11.3	16.0	18.4	21.1	16.7	4.2			
	D1384	19.4	16.2	10.8	10.1	14.1	4.4			

Table 2.6: Distances between Potential Salt Bridge Pairs in the Roseman Model

The distances between salt bridges projecting into the interface between proximal leg segments in the Roseman Model were measured using RasMol and compiled into this table. Two versions of the Roseman Model, considered equivalently likely fits and only slightly different in position from one another, were used. The distances between each proximal leg segment in the model for each of the two models were averaged when considering the distance to nearby conserved (left) and non-conserved (right) partners. Used together with information on the conservation of the charged residues (Table 2.5), those with consistently low distances to a neighboring partner were considered the most likely salt bridges and were chosen as targets for mutagenesis.

Site-directed mutagenesis was used to create two mutant Hub molecules, each with two point mutations (Table 2.7). Double was a mutant engineered to remove both basic charges from the center of the interface with the mutations K1331E and R1333E. Double was expected be a loss of function mutant, where the ability to assemble would be diminished. Switch, a charge reversal mutation, contained the changes K1331E and E1334K. Switch was intended to be a gain of function mutant, to restore the broken salt bridges by reversing the partnered residues. While Switch retains the overall electrostatic character of Hub, Double has an increased negative charge due to the replacement of two positively charged residues with acids. To control for electrostatic factors, M3 from the first round of mutations above, which contains the mutation 1161 RKKAR -> REKAE was used, because it similarly had two basic residues replaced with acidic residues, but the changes were known to be far from the likely interface being studied.

Mutant Construct Name	**Residues Targeted**	**Mutation**
Switch	K1331, R1333	E1331, E1333
Double	K1331, E1334	E1331, K1334
M3	K1162, R1165	E1161, E1165

Table 2.7: Summary of Mutations Testing the Roseman Model

The site-directed mutants Double and Switch created to test the Roseman Model of the proximal leg interface are listed, along with the M3 mutant, which was used as a control. All mutations were made in the Hub (1074-1675) recombinant clathrin fragment.

Switch and Double Hub Mutants were readily expressed and purified, and their assembly was compared with Hub and M3. In the presence or absence of light chain, Switch and Double were both able to assemble to a similar rate and extent as Hub in the presence or absence of light chain (Figure 2.10).

Previous studies have indicated an ionic strength effect on clathrin assembly, which is similar for sodium, lithium, cesium, and potassium (Van Jaarsveld et al., 1981). In the presence of salt, the assembly of clathrin Hub and mutant Hubs may similarly be titrated down to a very low level (Figure 2.11). Statistically, Double and M3, which have a lower net charge appear to assemble to a slightly greater extent than Hub in the absence of light chain. Switch behaved similarly to Hub in the presence of sodium chloride, with a maximum at 100- 200 mM NaCl, declining to no assembly at 800 mM. However, in the presence of salt Double and M3 both showed significantly enhanced assembly in the presence of 100- 200 mM NaCl, and assembly was above normal even when 400 mM salt was added.

Adding small amounts of salt appears to shield long-range repulsion between negatively charged clathrin Hubs, especially important for the more densely charged Double and M3. This shielding allows the Hubs and Mutant Hubs to approach each other in close enough proximity to enhance the overall assembly. In high salt concentrations, any specificity determining interactions that depend on electrostatics are shielded from one another, so assembly does not proceed. Because both Double and M3 have similar assembly behaviors, the difference between Hub and the mutants is due to overall electrostatic properties

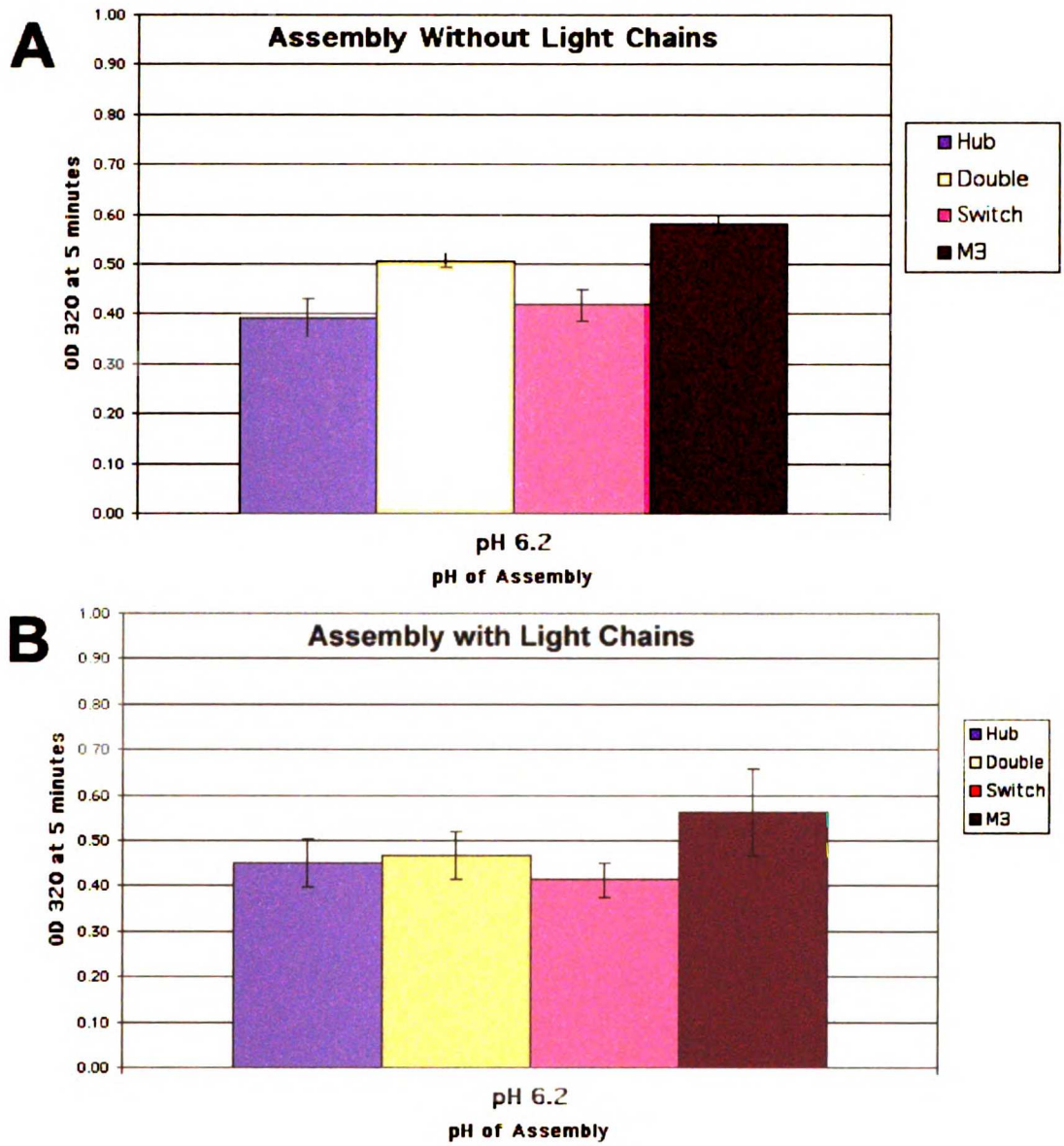


Figure 2.10: Assembly of Mutant Hubs to Test the Roseman Model
Recombinant mutant Hub molecules were expressed and purified, and tested spectroscopically for self-assembly. **A.** All mutants retained the ability to assemble at pH 6.2. Extent of assembly was slightly higher for Mutants Double and M3 than for native Hub or Switch. **B.** Assembly of all mutants was fully regulated by light chain.

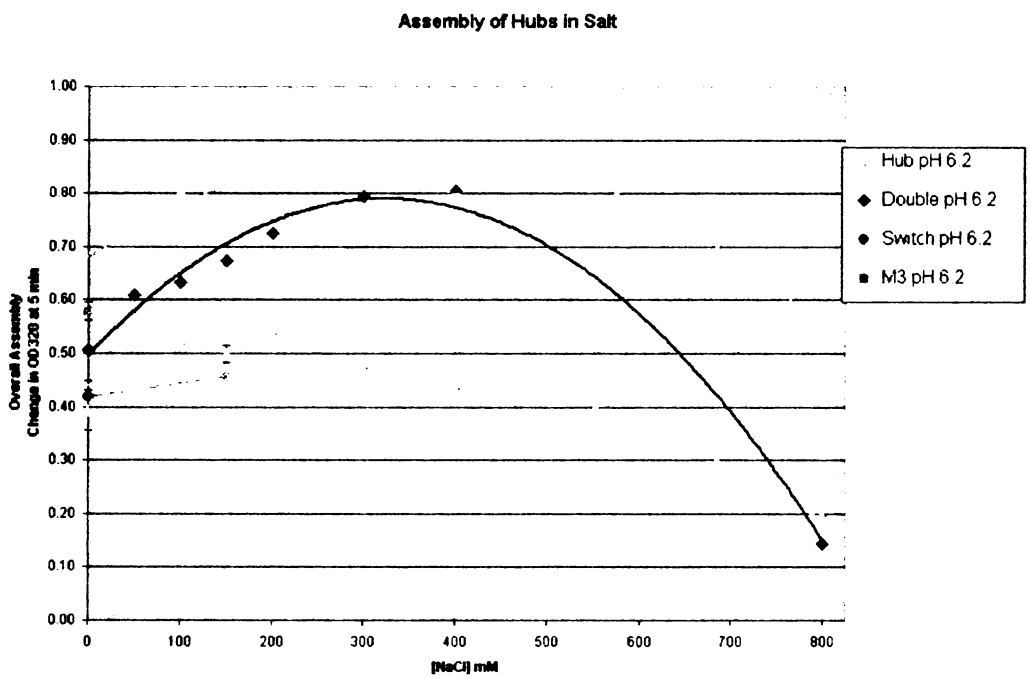


Figure 2.11: Mutant Hub Assembly in the Presence of Salt: Electrostatic Effects

Recombinant mutant Hubs were expressed and purified, and assembly in the absence of light chain was monitored spectroscopically. Extent of assembly is plotted as a function of sodium chloride concentration in the solution during the assembly reaction. Hub and Switch have a similar ability to assemble in the presence of salt, with maximum assembly seen close to physiological levels and a titration of the ability to assemble as salt levels rise beyond that point. Double and M3 mutants, with a more negative overall molecular charge, show a significantly higher extent of assembly as salt is added and electrostatic repulsion is shielded, before titration of ability to assemble.

of clathrin rather than an effect of the specific residues selected for mutation in Double, Switch, or M3.

There can be several explanations for the lack of effect of the mutations chosen on the overall self-assembly of clathrin Hub. It has been suggested that Hub assembly may not be fully representative of clathrin assembly. It is possible that Hub assembly, lacking a defined knee region, may allow some amount of sliding up and down the proximal legs into multiple "bound" configurations for a much looser and less structured sort of lattice. However, EM results (Liu et al., 1995) and the ability of isolated Hubs and terminal/distal fragments to combine to form whole baskets (Greene et al., 2000) suggests there is a well-defined specificity inherent in the Hub interactions. It is possible that R1267 or other residues may have acted instead of or in conjunction with the 1331-1334 location to bind clathrin proximal legs, where mutation of only one binding site is insufficient to cause a result. This is quite likely, given the cooperative nature of proximal leg interactions (Chapter Three). Finally, the docked cryo-EM/ crystal model may have been inaccurate, and the residues selected may not be located within the interface at all. Well-conserved basic residues not on the original proposed interface include R1245, K1246, K1254, R1293, R1311, R1342, K1347, R1350, K1397, R1429, R1434, and K1441. Any further mutagenesis in the absence of an improved model was considered unlikely to converge on the specific residues involved in a timely manner. However, the improved Wilbur Model has recently been created (see below) and further mutagenesis studies are likely.

The Wilbur Model

A structural model for the light chain bound to proximal leg was recently published by our laboratory (Chen et al., 2002) and the density was re-docked into the cryo-EM map to generate a new model of the interface between proximal leg segments (Wilbur et al., 2002). In this new model, the proximal legs are tilted somewhat relative to their previous orientation, changing the location of the interface between proximal leg segments (Figure 2.12). Interestingly, in the Wilbur model the looped faces of the proximal legs interact extensively. This places one face of collinear histidines directly in the interface. Many of the residues mutated to test the Ybe and Roseman models can be located within this new dimeric model (Figure 2.13). While the amino acids mutated to test the Roseman Model still point inward toward the opposite proximal leg, they lie too far above the new proposed interface to mediate specificity. Hydrophobic interactions have now been strongly implicated as the primary driving force for clathrin assembly (Chapter Three), and these interactions are almost certainly supplemented by hydrogen bonds and a few salt bridges, with electrostatic and steric factors determining the specificity of the interaction. The Wilbur Model can thus be used to predict interactions, including possible salt bridges, and undergo another round of site-directed mutagenesis to determine the accuracy of the modeled interface. The leading candidates for salt bridge interaction sites based on the Wilbur alignment are D1448- K1246 and R1429- E1310. My hope is that others will succeed in these efforts and the molecular determinants for interaction will be located.

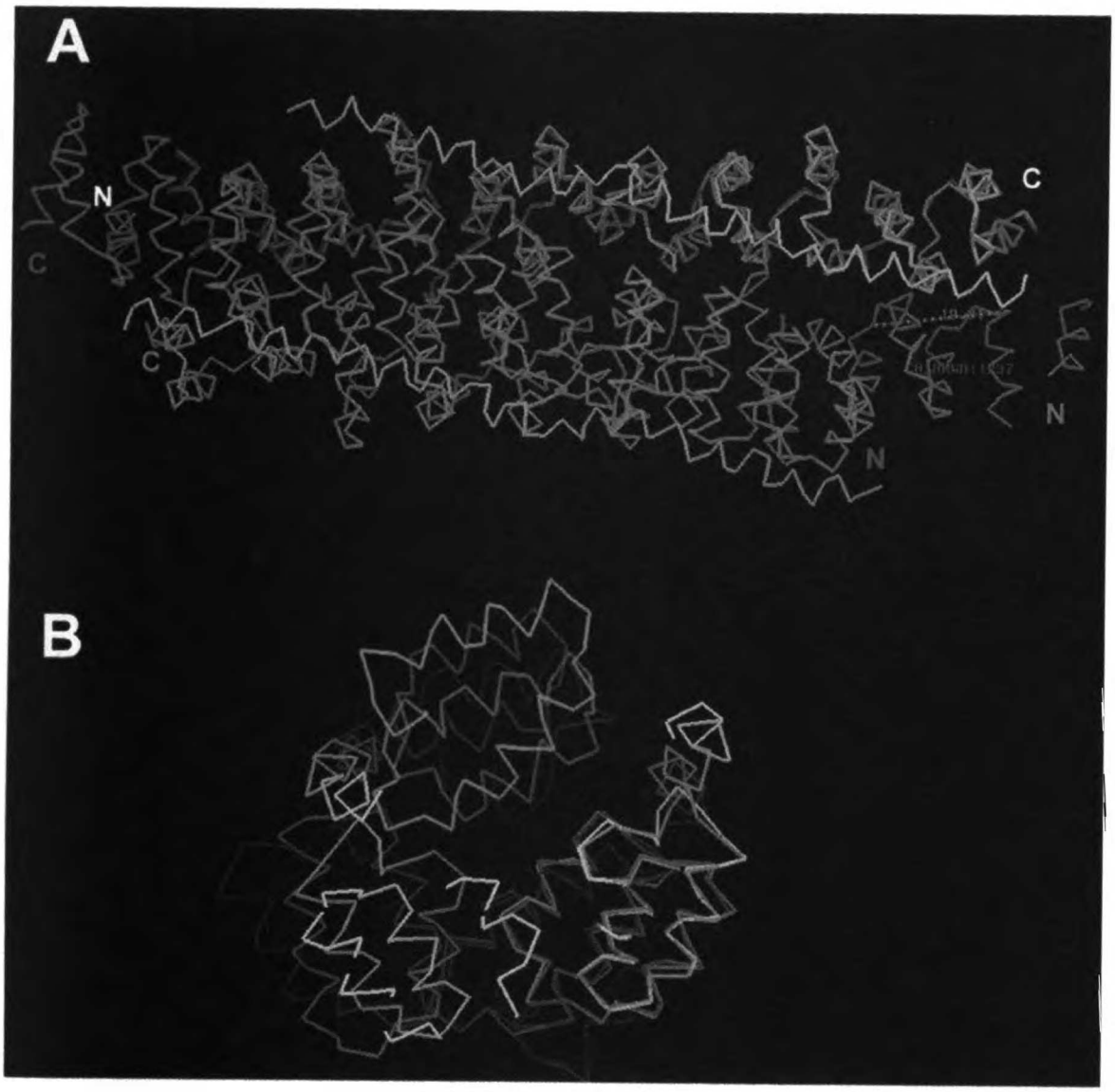


Figure 2.12: The Wilbur Model for the Assembled Proximal Leg Binding

The Wilbur model of two adjacent proximal legs with light chains bound (green) is superimposed on the Roseman Model for two proximal legs (blue) shows a rotation of the docking angle of the proximal legs, resulting in a new interface prediction. **A.** A lengthwise view of the models, where the light chain is visible as a thin extended green helix down the side of the superhelical proximal leg segment. **B.** An end-on view of the models makes the movement of the interface clear. The light chains in the Wilbur Model (green) are located above and to the outside of the interface between proximal legs. The heavy chains interact via a looped face and part of a helical face. The Roseman Model's interface (between the two blue proximal legs) of two flat helical interacting faces is rotated outward compared to the Wilbur interface (between green proximal legs).

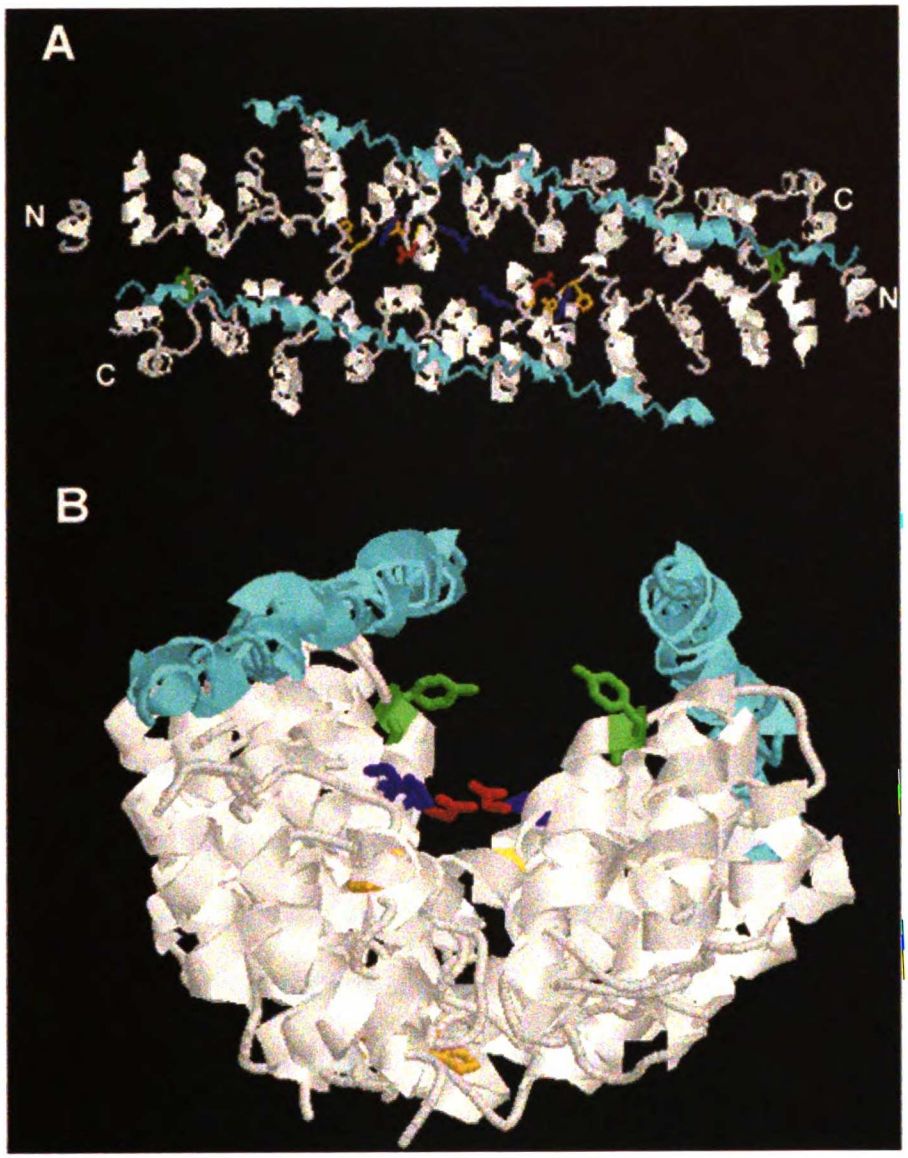


Figure 2.13: Location of Previously Mutated Residues Within the Wilbur Model

Two proximal leg domains (white) with bound light chains (cyan) interact in the Wilbur Model of the proximal leg interface between assembled clathrins, viewed from top (upper panel) and end (lower panel) orientations. The location of previously mutated residues is denoted in color using RasMol. The phosphorylation site Y1477 (green) is included as a point of reference. Histidines 1313 and 1335 mutated to test the Ybe Model are noted in yellow. H1335 is buried in the crystal structure, and 1313 is partially buried but located near the center of the interface. Basic residues K1331 and R1333, mutated to test the Roseman Model, are colored blue. While these residues still point into the interface, they are somewhat rotated out above the interface and are no longer at the point of closest approach. Moreover, they no longer align with E1334 (red).

Discussion

Clathrin self-assembly has long been known to involve a balance of electrostatic and hydrophobic interactions (Nandi and Edelhoch, 1984). The discovery of a charged patch on light chain that is critical for regulation of assembly (Ybe et al., 1998) led to the elegant Ybe salt bridge hypothesis that ultimately proved too simplistic to explain the assembly behavior of clathrin.

At physiological pH, the inhibition of assembly by light chains is somehow overcome by the presence of the adaptor complex. However, it seems unlikely that the adaptors are large enough to directly interact with the outer layers of the basket where proximal leg and light chains interact in order to overcome this inhibition by any direct mechanism. More likely, the adaptors gather and orient clathrin terminal domains to strengthen molecular interactions already existent between proximal and distal leg segments (Greene et al., 2000)(Chapter Three). Since adaptors and adaptor fragments exert their effect in-vitro in the absence of other proteins, any light chain-based salt bridge inhibition is unlikely to be directly competed out by any other protein to provide a mechanism for assembly in the presence of light chain. In other words, if a strong salt bridge were inactivated in the presence of light chain it would remain inactive physiologically in the assembled form. Thus, under physiological conditions in the Ybe Hypothesis, there are no active salt bridges in this model that could function under physiological conditions. Here I reconsider some of the more complex forces that could be involved in clathrin assembly based on early biochemical studies and my data.

Clathrin and all of its individual fragments- proximal leg, distal leg, Hub, and light chain- all have a significant net acidic charge (Table 2.8) that would hinder the initial approach of clathrin leg segments during the initiation phase of the assembly polymerization reaction. Consequently succinylation, which adds negative charge to lysines on the surface, prevents assembly; and polybasic amines, which attract the negative charge, stabilize assembled structures (Nandi and Edelhoch, 1984).

Fragment	**Isoelectric Point**	**Charge**
Clathrin Hub (1074- 1675)	pI 5.23	-19
Clathrin proximal leg (1074- 1522)	pI 6.14	-9
Clathrin light chain (1- 228, full length)	pI 4.44	-16

Table 2.8: Acidity of Recombinantly Expressed Clathrin Fragments

Clathrin heavy chain and its regulatory light chain subunit have a net negative charge and a low acidity, as calculated by GCG and Vector NTI software. Polyhistidine tags on recombinant constructs were not included in this calculation and would raise the isoelectric point slightly. Clathrin Hub assembles maximally at pH 6.2 where the proximal leg segments approach their pI and electrostatic repulsion is minimized. Binding of the light chain would heighten the repulsion between leg segments and lower the isoelectric point of Hub, altering the pH dependence of its assembly behavior. Note that distal leg is also acidic (data not shown).

Paradoxically, the even more electrostatically repelled Double and M3 mutants assembled better than Hub in the presence of salt. Once a close approach between these more charged leg segments was allowed, van der Waals interactions pulled the segments into tight binding and the electrostatic repulsion interaction term dropped off at this very close distance. The significant electrostatic repulsion thatwould be encountered if casual disassembly were to occur and the molecules were separated by a short distance once again was a barrier holding the lattices together more tightly than native Hub. Other evidence of this repulsion between leg segments is the abnormally small molecular size I observed during equilibrium ultracentrifugation experiments of proximal leg. Where pH was significantly different from the pI of the leg segments, the size appeared smaller than its actual size as determined by mass spectrometry, while close to the pI of the protein the size appeared normally (Figure 2.14). Because this is an experiment at equilibrium, this result is almost entirely independent of shape or size effects and is an indication of repulsion between leg segments in solution. Electrostatic effects should be considered along with shape effects in considering abnormal migrations from FPLC columns or any similar analysis. While the overall long-range electrostatic interaction is unfavorable for self-assembly, individual sites of complementary charges may exist and contribute short-range attraction and thus specificity through salt bridges. Plotting the surface of the proximal leg crystal structure in GRASP (Figure 2.15) failed to reveal any self-aligning oppositely charged patches of surface potential except in an orientation considered unlikely for binding, and the light chain is so acidic that

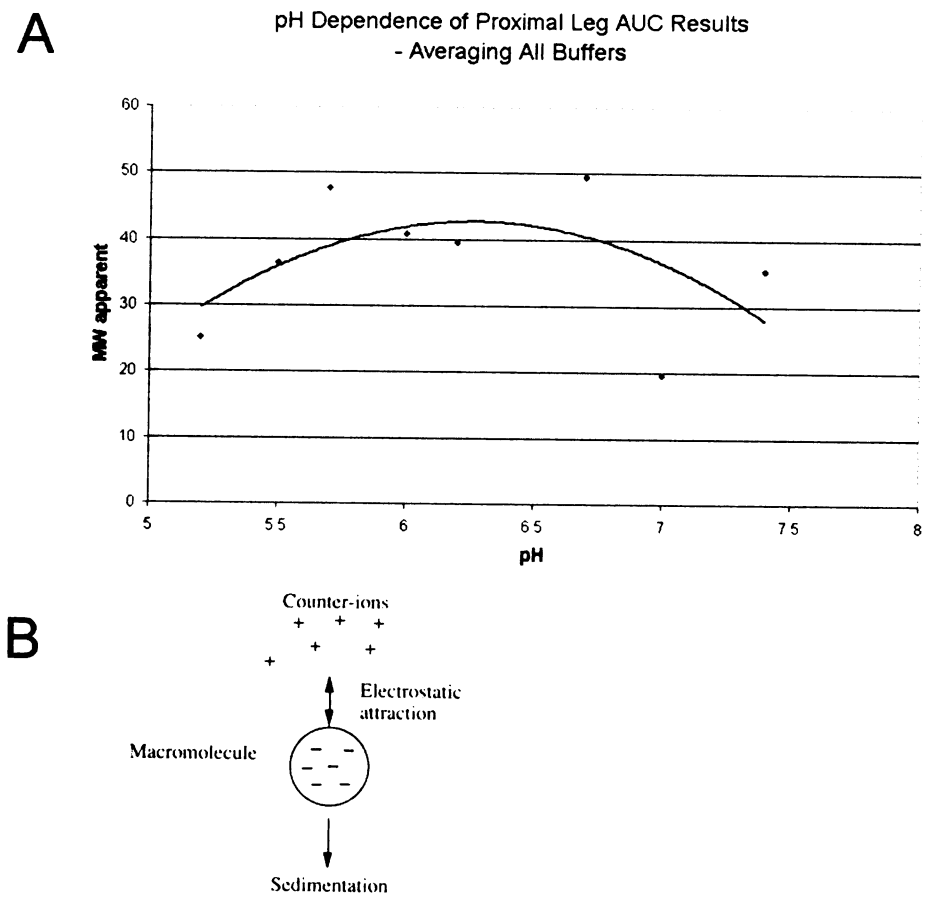


Figure 2.14: Electrostatic Repulsion and pH Variance of Observed Molecular Size for Proximal Legs in Equilibrium Ultracentrifugation **A.** Proximal leg segments were expressed, purified, and assayed in a Beckman analytical ultracentrifuge in an equilibrium sedimentation assay with similar buffer content to that used in an assembly assay. A variety of buffers were used to cover a large range of pH values: Tris, EPPS, MES, phosphate, and citrate. The pH dependence of the trend reflects the decreased electrostatic repulsion as the protein approaches its isoelectric point at pH 6.2. **B.** Electrostatic effects in a sedimentation velocity ultracentrifuge assay are illustrated. Negatively charged clathrin macromolecules move slower than similarly sized and shaped counterparts in a sedimentation assay because of their net attraction to the small counterions which migrate more slowly still. Electrostatic effects and shape would also affect elution time in a size-exclusion column for similar reasons. In an equilibrium sedimentation ultracentrifugation assay, the results are mostly independent of shape, but the net repulsion between leg segments keeps the macromolecules farther apart, broadening the gradient and making them appear to be smaller.

it is unlikely for light chain binding to create the necessary basic patch to allow electrostatic attraction. It seems to me that the electrostatic interactions between proximal leg segments are repulsive rather than attractive in nature, and any salt bridges are localized over a small area of the molecular surface.

The Ybe Hypothesis proposed that the protonation of histidine residues acted as a pH switch to allow salt-bridge driven assembly at acidic pH (Ybe et al., 1998). The unusual arrangement of histidines on the surface of the proximal leg suggests that this may be partly correct. At physiological pH, histidines are deprotonated and have two likely mechanisms of interaction. Histidines may partner with acidic residues through uncharged hydrogen bonds, or histidines on each leg may sandwich a divalent metal ion. Initial experiments to address these issues proved complicated by interferences and background interactions, but further exploration is merited. The specific site-directed mutations attempted failed to show any effect on clathrin assembly due to flaws in the models used for the interaction between clathrin proximal leg segments.

Ultimately, the existence of a dimeric crystal structure would be the ideal way to approach site-directed mutagenesis that attempts to block binding. Since such a dimeric structure of proximal leg segments has proven impossible (Chapter Three) and a crystal structure of assembled Hub has proven elusive (Ybe et al., 1999-2003), a higher resolution cryo-EM map would be exceedingly useful. During docking, the crystal structure must be reduced to the resolution of the cryo-EM before the algorithm attempts to locate it in the map. Thus, all of the surface contours and bumps in the crystal structure cannot be applied to help the

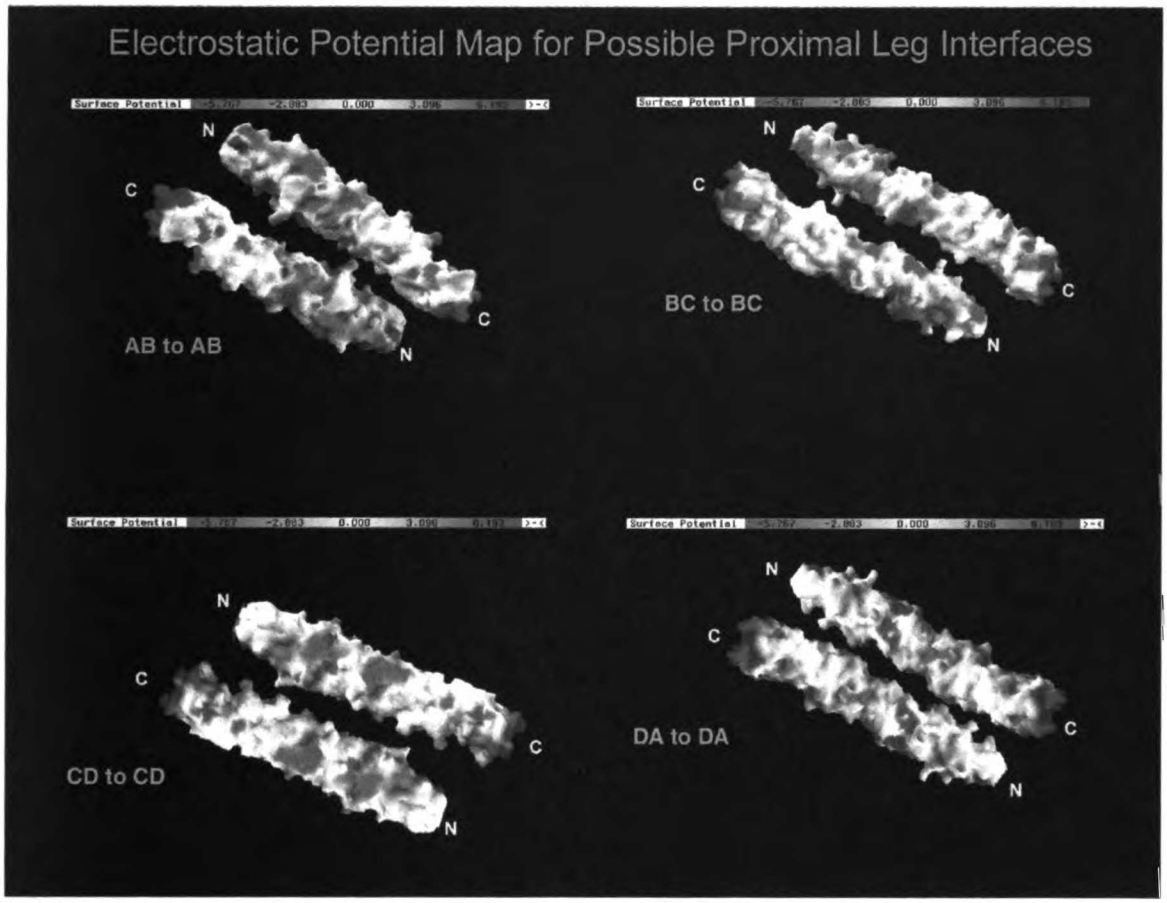


Figure 2.15: Surface Potential- Lack of Complementary Electrostatic Regions

The proximal leg segment crystal structure (PDB ID: 1B89) (Ybe et al., 1999) was used in GRASP to generate a map of electrostatic charge on the surface of the molecule. Equivalent charge density for bases or acids was given equivalent color density on the scale chosen (see scale bars). Two proximal leg molecules are aligned antiparallel side-by-side like an open sandwich, where the faces shown would be in nearly full contact in the Roseman Model proposed assembled form. The rod-like molecule was considered to have four faces, two helical (AB and CD) and two looped (BC and DA). The Roseman Model predicted an AB-to-AB interface. The Wilbur Model has the two faces slid to a different vertical alignment, with the same faces in contact. Charge complementarity in the center of the rod would be expected to be more important than at the ends, because the leg segments cross in the center in the assembled form. Significant electrostatic repulsion would be expected between the centers of the leg segments in every orientation except DA-to-DA. Note that sliding the alignment slightly would reduce repulsion for faces AB, BC, and CD, but there is no complementary basic patch available. This bolsters the concept that hydrophobic interactions provide most of the affinity between proximal leg segments.

algorithm find the correct orientation; instead the algorithm works from a blurred rod-like shape. A higher resolution cryo-EM map would retain more of the molecular details of the structure and enable a superior fit.

Interestingly, the new proposed interface in the Wilbur Model involves a looped face rather than a helical face, and aligns well with the collinear histidines identified earlier along the looped faces of the proximal leg structure. Only one of the looped faces had a visible basic patch near the center, offering the possibility of electrostatic complementarity between faces (Figure 2.15). All of these point toward a promising new model for the assembly interface.

However, mutagenesis may not be the optimal way to approach such a cooperative interaction (Chapter Three). While bulky residues might hinder the approach between leg segments, salt bridge mutations may ultimately prove fruitless if the hydrophobic residues surrounding a salt bridge maintain the affinity between leg segments. If specificity relies on a network of salt bridges and hydrogen bonds, disruption of only a few in a cooperative system is unlikely to have an effect on overall assembly, and it is difficult in such a system to measure individual binding affinities between segments. Crosslinking or FRET may be a more sensitive alternative if future rounds of mutagenesis continue to show no effect. Alternatively, mutagenesis could be used to target bulky surface-exposed hydrophobic residues that contribute more significant amounts of affinity to the interaction (discussed further in Chapter Six).

Chapter Three: Clathrin Self-Assembly Involves Coordinated Weak Interactions Favorable for Cellular Regulation

Abstract

The clathrin triskelion self-assembles into a polyhedral coat surrounding membrane vesicles that sort receptor cargo to the endocytic pathway. A triskelion contains three clathrin heavy chains joined at their C-termini, each extending from this vertex into proximal and distal leg segments linked to a globular N-terminal domain. In the clathrin coat, leg segments are entwined into parallel and anti-parallel interactions. Here we define the contributions of segmental interactions to the overall clathrin assembly reaction. Proximal and distal leg segments were found to lack sufficient affinity to form stable homo- or heterodimers under clathrin assembly conditions. However, chimeric constructs of proximal or distal leg segments, trimerized by replacement of the clathrin trimerization domain with that of the invariant chain protein, were able to self-assemble in reversible reactions. Thus clathrin assembly occurs because weak leg segment affinities are coordinated through trimerization, sharing properties with biopolymers whose polymerization depends of multiple weak hydrophobic interactions. Such polymerization is sensitive to small environmental changes

and is therefore compatible with cellular regulation of assembly, disassembly and curvature during formation of clathrin-coated vesicles.

Introduction

Clathrin self-assembly creates a basket-like polyhedral protein lattice that coats intracellular membranes involved in receptor-mediated endocytosis, organelle biogenesis from the trans-Golgi network and lysosomal targeting (Brodsky et al., 2001). The clathrin lattice organizes associated adaptor molecules to sequester receptors for selective sorting into clathrin-coated membrane domains or transport vesicles. The self-assembling unit is a clathrin triskelion comprising three heavy chains, non-covalently trimerized at their C-termini, each bound by a light chain subunit that associates irreversibly during biosynthesis. While purified clathrin can spontaneously and reversibly form closed polyhedral baskets, polymerization in the cell is highly regulated by accessory proteins. Clathrin light chains prevent random heavy chain assembly, and association with adaptor proteins and/or assembly co-factors overcomes this inhibition, directing polymerization onto membranes (Brodsky et al., 2001; Mishra et al., 2002). Adaptors and their partners, including the recently described epsin (Ford et al., 2002), contribute to the transformation of a hexagonal clathrin lattice into a curved lattice, ultimately containing the twelve pentagons needed for polyhedron closure. The heat shock family protein Hsc70 and its co-factor auxilin promote disassembly (Lemmon, 2001). To resolve the apparent thermodynamic paradox of how spontaneous clathrin polymerization reaction can be regulated

with sensitivity and reversibility in the cell, we undertook an analysis of the nature of the interactions between domains of the clathrin triskelion legs. In this study we define their binding properties and establish how these properties facilitate cellular regulation of clathrin self-assembly.

The structure and morphology of assembled clathrin is well characterized (Brodsky et al., 2001). Clathrin heavy chains have five domains: an N-terminal globular domain (1-390), a helical linker (390-494), a distal leg segment (494-1074), a proximal leg segment (1074-1522), and a C-terminal trimerization/vertex domain (1522-1675). In the assembled clathrin basket, each polygonal side is composed of two anti-parallel proximal leg segments, one from each of the vertices at either end of the segment, and two anti-parallel distal leg segments that lie below and parallel to the proximal leg segments, contributed from triskelia with adjacent vertices (Figure 3.1A, B) (Smith et al., 1998). Thus each side is composed of four leg segments from four separate triskelia. The clathrin heavy chain proximal leg segment is a super-helix of α-helices about 160 Å long with an oval cross section of 24 x 28 Å and comprises three clathrin heavy chain repeat (CHCR) domains (Ybe et al., 1999). The clathrin leg was found by sequence homology to share additional copies of the CHCR structural motif such that the rod-like super-helix structure is predicted to extend through the distal leg, ending at the linker region (Ybe et al., 1999). Analysis of high resolution cryo-electron microscopy images of assembled clathrin baskets (Smith et al., 1998) revealed that adjacent proximal leg segments have a center-to-center distance of about 25 Å, adjacent distal leg segments are separated by 25 Å, and the proximal-distal

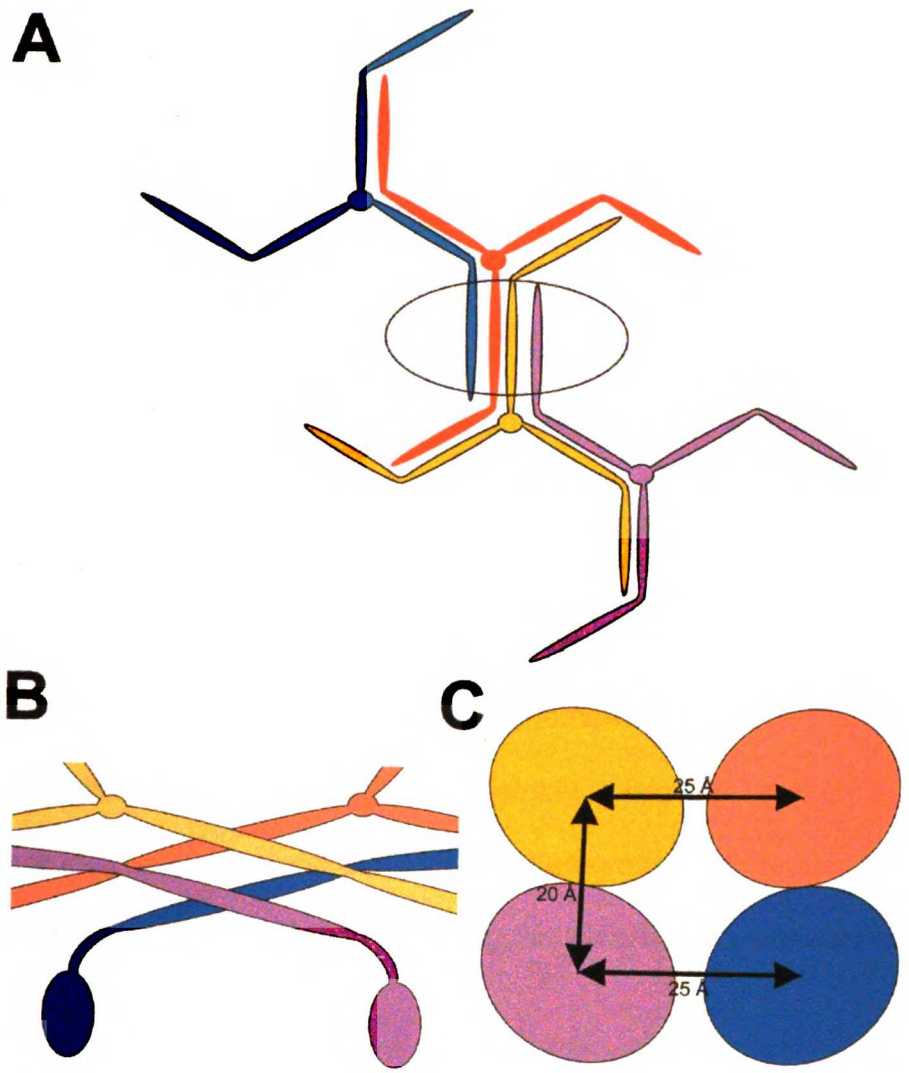


Figure 3.1: Clathrin Polymerization
A. Four triskelia together form one edge of a clathrin polyhedron. The completed edge (encircled) is composed of four leg segments: two anti-parallel distal leg segments (lavender and cyan) and two anti-parallel proximal leg segments (yellow and orange). **B.** This predicted arrangement of triskelia in one edge of the clathrin polyhedron, viewed from the side, shows two adjacent proximal leg segments (yellow, orange) coming from a vertex downward and crossing at an angle on the outer surface of the basket. Two crossing distal leg segments (lavender, cyan) lie underneath, with the globular N-terminal domain protruding down toward the inside of the basket. Based on the 21Å structure of a clathrin basket (Smith et al., 1998). **C.** A cross-section view of one edge of the clathrin polyhedron with two proximal leg segments on top (yellow, orange) and two distal leg segments beneath (lavender and cyan) indicates the center-to-center distance separation between leg segments. Based on the 21Å structure of a clathrin basket (Smith et al., 1998).

separation between layers of the basket is 20 Å (Musacchio et al., 1999) (Figure 3.1C). This close intertwining of the clathrin proximal and distal leg segments within the assembled basket suggests that interactions between leg segments are the driving force for clathrin assembly. Biochemical evidence points to a critical role for proximal leg segments in assembly. Monoclonal antibodies against the proximal leg segment block assembly (Blank and Brodsky, 1986), and recombinant "Hub" fragments (1074-1675), which contains three proximal leg segments and the trimerization domain, reversibly assemble into lattice-like structures (Liu et al., 1995). In addition, recombinant Hub and terminal-distal (1-1074) fragments of clathrin reconstitute assembly of baskets when stimulated with adaptor fragments (Greene et al., 2000), demonstrating that distal leg segments have a significant role in orienting the triskelia into a closed basket shape.

While the contacts between specific leg segments have been identified and their individual contributions characterized, the coordination and affinities of these binding interactions during clathrin polymerization needs to be established to understand how assembly can be regulated. In a typical protein polymerization reaction, the enormous entropic cost of gathering molecules into a supra-molecular assembly is overcome primarily by the entropy gain from the exclusion of water from hydrophobic subunit interfaces in the assembled structure (Cantor and Schimmel, 1980b). Entropy-driven reactions such as the polymerization of actin or tobacco mosaic virus are typically maintained over a

large surface area by extensive weak hydrophobic van der Waals interactions supplemented by a few electrostatic hydrogen bonds, with shape complementarity and electrostatic interactions providing more specificity than affinity to the interaction (Fersht et al., 1985; Prevelige, 1998). Experiments focusing on individual aspects of clathrin assembly have suggested that while electrostatic enthalpic interactions are involved (Ybe et al., 1998), entropic hydrophobic effects also make important contributions (Nandi and Edelhoch, 1984). In addition, theoretical models of the energetics of clathrin assembly (Nossal, 2001) have suggested that the binding energy between assembled triskelia is quite low, on the order of thermal energetic fluctuations, suggesting assembly may indeed be driven by weak van der Waals interactions. Resolution of this latter issue is critical for establishing whether curvature of a clathrin lattice can occur by rearrangement of a hexagonal lattice, which would involve local disassembly, or is limited to addition of new triskelia at the lattice edge.

To establish the relative contributions of the interaction of individual clathrin leg segments to the overall assembly reaction and the binding characteristics of the segmental interactions, we studied the dimerization of recombinant clathrin proximal and distal leg segments and the assembly of chimeric trimers of leg segments. Our data demonstrate that proximal-proximal, proximal-distal, and distal-distal interactions all contribute affinity to the clathrin polymerization process, but individually their dimerization interactions are too weak to detect. Thus clathrin assembly is consistent with the classical model of protein polymerization driven by multiple weak hydrophobic interactions. These

data account for the intrinsic ability for clathrin assembly to be regulated in the cell with respect to curvature changes and disassembly.

Results

Two complementary approaches were taken to address which segments of clathrin mediate the affinity for self-interaction. In the first approach, we sought to localize possible strong interactions between leg segments. To accomplish this, isolated leg segments were assayed for dimerizing interactions under conditions where whole clathrin assembles into a basket. In the second approach, we sought to identify cooperative interactions between leg segments. To that end, protein chimeras in which leg segments were trimerized using the trimerization domain from the invariant chain protein were engineered and tested.

Individual Leg Segments Are Monomeric Under Assembly Conditions

In order to localize the key contacts between clathrin leg domains, isolated leg fragments were tested to find the smallest possible fragment with sufficient interaction affinity to self-adhere. To accomplish this, clathrin leg segments were tested by a variety of methods to detect the presence of dimers under conditions where clathrin Hubs (1074-1675) assemble into lattices, in an approach similar to that taken with other proteins (Bellingham et al., 2001; Flood et al., 1997; Fujita et al., 1998).

Size-exclusion chromatography was used to examine the potential interactions of proximal (1074-1522) and terminal-distal (1-1074) leg domains

that were expressed as recombinant polyhistidine-tagged protein fragments and purified by nickel affinity column chromatography (Greene et al., 2000; Ybe et al., 1999). When the proximal leg fragment segment was applied to a calibrated FPLC high resolution size-exclusion column, it eluted as a monomer when chromatographed at pH 6.2 (Figure 3.2A), conditions where clathrin Hubs form a lattice (Liu et al., 1995). In a separate experiment using a lower resolution size-exclusion column, recombinant terminal-distal domain (TDD) eluted as a monomer significantly smaller than Hub, at pH 6.7, also recapitulating conditions favorable for Hub assembly (Figure 3.2B). Mixtures of TDD and proximal leg segments contained the two component peaks, but no new peak in the eluate corresponding to a complex (Figure 3.2B). Under similar conditions, a mixture of the proximal leg with clathrin light chain LCb at pH 6.2 eluted as a monomeric complex of the two subunits (Figure 3.2C), indicating that LC binding did not promote proximal leg interactions to detectable levels. The presence of both subunits in the eluate was confirmed by gel analysis (data not shown). However, the stable interaction between the proximal leg and LCb in this experiment reveals that the folding of the proximal leg segment was representative of the native triskelion structure, in which the clathrin heavy and light chains form a stable complex and co-elute from a sizing column (Ungewickell and Branton, 1981). Overall, this result suggests that any affinity between proximal or distal leg segments in the presence or absence of light chain is too weak to detect by this assay, and is thus significantly weaker than the $K_d \sim 10^{-9}$ heavy chain-light chain interaction (Winkler and Stanley, 1983).

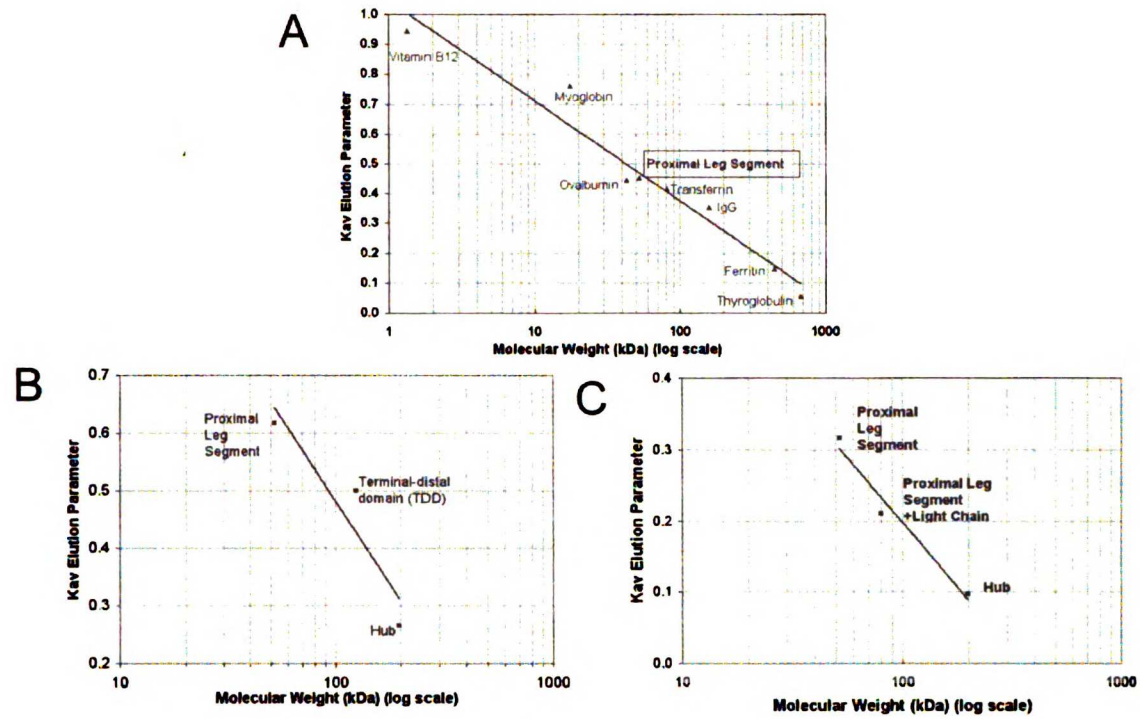


Figure 3.2: Individual Leg Segments Are Monomeric by FPLC. Fragments representing the clathrin proximal leg and terminal-distal domain were expressed as polyhistidine-tagged recombinant proteins, purified by a nickel-affinity column and analyzed by gel permeation chromatography. Kav versus molecular weight for sample and standards were plotted. **A.** Proximal leg segments are monomeric on a calibrated FPLC column. Proximal leg fragment and calibration standards were independently chromatographed on a Superdex 200 HR 10/30 column in Bis-tris pH 6.2 and elution times were noted. Proximal leg segment Kav was calculated as 0.454, indicating a size for clathrin proximal leg segment consistent with a monomeric 52kD species. **B.** Isolated proximal and distal leg segments are monomeric and do not form a complex when combined. Proximal and terminal-distal domains were chromatographed on a Superose 6 prep grade column in MES/Tris pH 6.7 and elution times were compared with elution of Hub or leg segments at pH 7.3. Kav was calculated. Mixtures of proximal and distal domain segments eluted as two separate populations; no larger peak corresponding to a complex was observed. **C.** Proximal leg segments bind light chains normally but do not dimerize in the presence of light chain. Proximal leg segments with and without purified light chains and Hub were chromatographed on a Superose 6 prep grade column in Bis-tris pH 6.2 and elution times were noted for calculation of Kav. While the mixture of proximal leg with light chain eluted as a complex, indicating proper folding of the recombinant protein fragments, the Kav of the complex is consistent with each complex containing only one light chain and one proximal leg. Thus the presence of light chain did not induce dimerization of the proximal leg.

More sensitive methods were chosen to pursue detection of a weak dimer between proximal leg segments. Analytical ultracentrifugation by sedimentation velocity yielded data fitting a single population of proximal leg segment species in solution at pH 6.2 (Figure 3.3A) or pH 6.7 (data not shown) with an apparent sedimentation coefficient of S=2.3. This coefficient for the proximal leg segment corresponds to a larger protein than typically expected for a sedimentation coefficient of this size, as evidenced by the frictional ratio of 1.8. The frictional ratio is proportional to the axial ratio of the protein, where an ideal spherical protein has F=1.0, while real data for globular proteins typically have F=1.2 (Cantor and Schimmel, 1980a). Clathrin proximal leg segment is a prolate ellipsoid (elongated rod) and Hub is an oblate ellipsoid (flattened disk), so each would be expected to have a high frictional ratio. In order to confirm that this species represented a monomer, the proximal leg segment was further analyzed by equilibrium sedimentation. Because this assay relies on the distribution of the species after equilibrium has been reached rather than during migration down the cell, it is insensitive to the non-spherical shape of clathrin proximal leg segments. Equilibrium sedimentation of proximal leg segments at pH 6.2, tested at multiple speeds and concentrations, yielded data that were fit simultaneously to a single set of parameters consistent with the proximal leg segment behaving as a monomer (Figure 3.4). The corresponding K_d for interaction of proximal leg segments determined from data that was fitted artificially to dimer behavior is at best on the order of 115 μM. This value is weaker than the nanomolar range

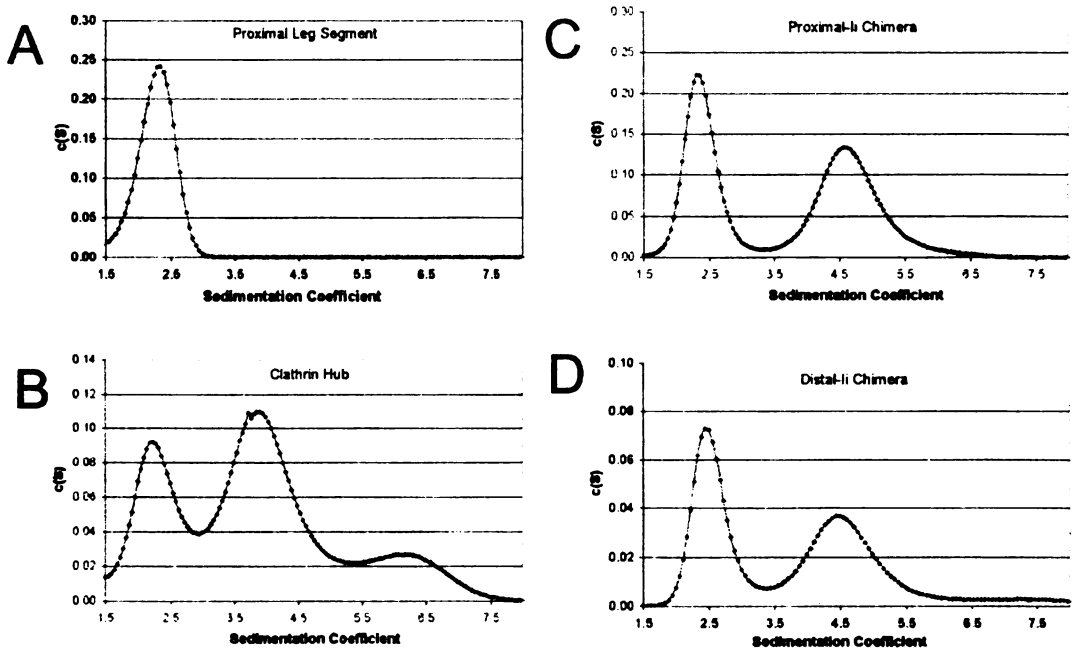


Figure 3.3: Sedimentation Velocity Assays.
Polyhistidine-tagged constructs were expressed and purified and their interaction assessed using sedimentation velocity by analytical ultracentrifugation in a Beckman XL-I at 40K rpm, 4 degrees, at pH 6.2. SedFit was used for a continuous c(S) distribution analysis of the data. For each fit, the root mean square deviation (rmsd) for the fit of the curves to the data, the number of data points (n), sum of squared residuals (SSR) and frictional coefficient (F, indicating shape deviations from a sphere) were noted. The confidence level (F-ratio) was 0.95. **A.** Proximal leg segment at pH 6.2 is composed of a single species with sedimentation coefficient S= 2.3 (rmsd at pH 6.2 = 0.006503, n=8372, SSR=0.354059, F=1.8). This species was confirmed as a monomer by equilibrium sedimentation (Figure 5), and components of trimeric constructs with S=2.3 were thus assigned as monomeric species in plots B-D. **B.** Hub at pH 7.9 is composed of three species in solution: the first two are at S=2.3 and S=3.9 and correspond to monomer (35.2%) and trimer (64.8%), respectively, and the third is a minor larger species at S=6.1 (rmsd = 0.007666, n=14198, SSR=0.834386, F=1.6). **C.** Proximal-Ii chimera at pH 7.9 is composed of two species in solution, at S=2.4 and S=4.7 (rmsd = 0.013468, n=16365, SSR=2.968339, F=1.3), which correspond to monomer (46.8%) and trimer (53.2%), respectively. **D.** Distal-Ii chimera at pH 7.9 is composed of two species in solution, at S=2.5 and S=4.6 (rmsd = 0.0.008542, n=18212, SSR=1.328910, F=1.2), which correspond to monomer (49.5%) and trimer (50.5%), respectively.

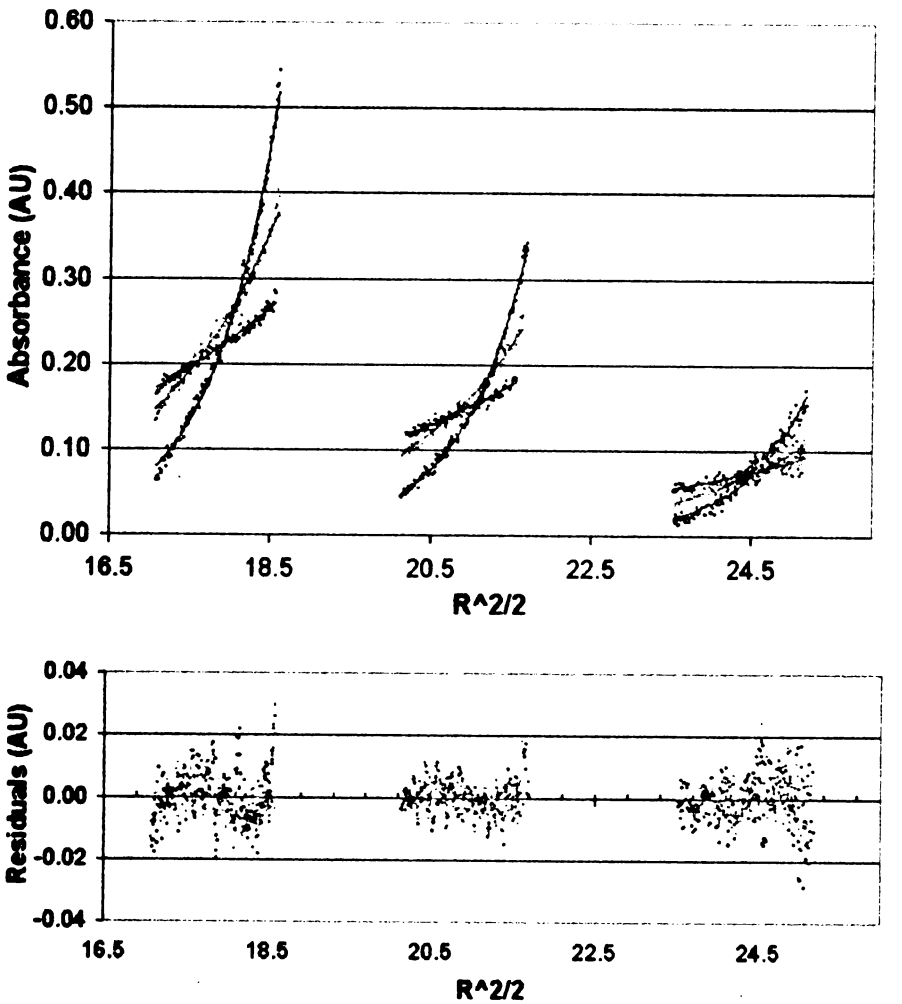


Figure 3.4: Proximal Leg Segment is a Monomer at pH 6.2 in Equilibrium Sedimentation Assay

Proximal leg was expressed and purified and its self-interaction was assessed by equilibrium sedimentation by analytical ultracentrifugation in a Beckman XL-I at 40K rpm, 4 degrees, pH 6.2. WinNonLin was used to simultaneously fit data points from samples of 0.4 (left), 0.2 (center), and 0.1 mg/mL (right) at 7K (green), 10K (orange), and 14K (purple) to the lines shown in black. (Radial distance) squared/2 is plotted against A280 (AU). A plot of the residuals (lower panel) indicates the systematic error in the fit between the lines and the raw data points. The fit gave values for parameter σ = 0.623 (0.547-0.696) where for the expected monomer σ = 0.667. The fit is consistent with a proximal leg segment fit to an N-mer of N=0.94, which represents monomer. Variance of fit was 4.6×10^{-5} with 1234 degrees of freedom.

affinity of enthalpy-driven self-assembly reactions (Mejean et al., 2001; Parker et al., 2001) and is more consistent with the order of magnitude expected for an entropy-driven network of weak hydrophobic subunit interactions (Cantor and Schimmel, 1980b). When complexes of proximal leg with LCb were tested under similar conditions, the equilibrium sedimentation data were also characteristic of behavior as a monomeric complex of the two subunits, consistent with the chromatography experiments above (data not shown).

A yeast-two-hybrid assay was then employed to examine further the possibility of weak interactions under more physiological conditions. Clathrin proximal leg segments or other heavy chain fragments were cloned into two-hybrid bait and prey vectors to express them as fusion proteins attached to GAL4 DNA binding or activation domains (Table 3.1). As a positive control, a strong interaction between the proximal leg segment and clathrin light chain LCb, cloned into the same prey, was demonstrated. Low levels of Hub assembly occur at physiological pH (in the absence of bound clathrin light chains) (Liu et al., 1995) and correspondingly, a weak interaction between Hub constructs (1073-1675) was detected. However, the proximal leg segment and its fragments showed no detectable interaction with one another or with Hub fragments (which extend into the trimerization/vertex domain). To rule out the possibility that the anti-parallel orientation of the interaction between leg segments and their elongated shape forced too great a separation between the GAL4 domains and allowed false negative interactions, the proximal leg segment was alsoconstructed upstream to the GAL4 activation domain in the bait vector. In this configuration the GAL4

domains would be adjacent during an anti-parallel interaction, but this construction also failed to show any interaction of the proximal leg segment with itself or with Hub (Chen and Brodsky, 2001). Thus the in-vivo evidence supported the in-vitro data, demonstrating a lack of detectable interaction between leg segments.

If proximal leg segments interacted strongly but only transiently to provide a nucleation center for clathrin assembly, the transient interaction would not be detectable in an equilibrium assay such as analytical centrifugation. Therefore, surface plasmon resonance was used to assay transient proximal leg segment interactions. In this assay a chip matrix was coated with immobilized proximal leg segment or clathrin light chain LCb and a solution of proximal leg segments was flowed over the chip surface in the chamber of an IASys optical biosensor

BD AD	1073-1675 Hub	1073-1522 Proximal Leg Segment	1204-1522	1271-1414
1073-1675 Hub	4			
1073-1522 Proximal Leg Segment	--	--	--	--
1204-1522		--	--	
1271-1414		--	--	--
LCb (Control)	65	65	200	

Table 3.1: Proximal Leg Segments Do Not Interact in a Yeast-Two-Hybrid Assay

[a]Interactions between binding domain (BD) and activation domain (AD) are noted in beta-galactosidase units.
-- indicates interaction was tested but could not be detected.

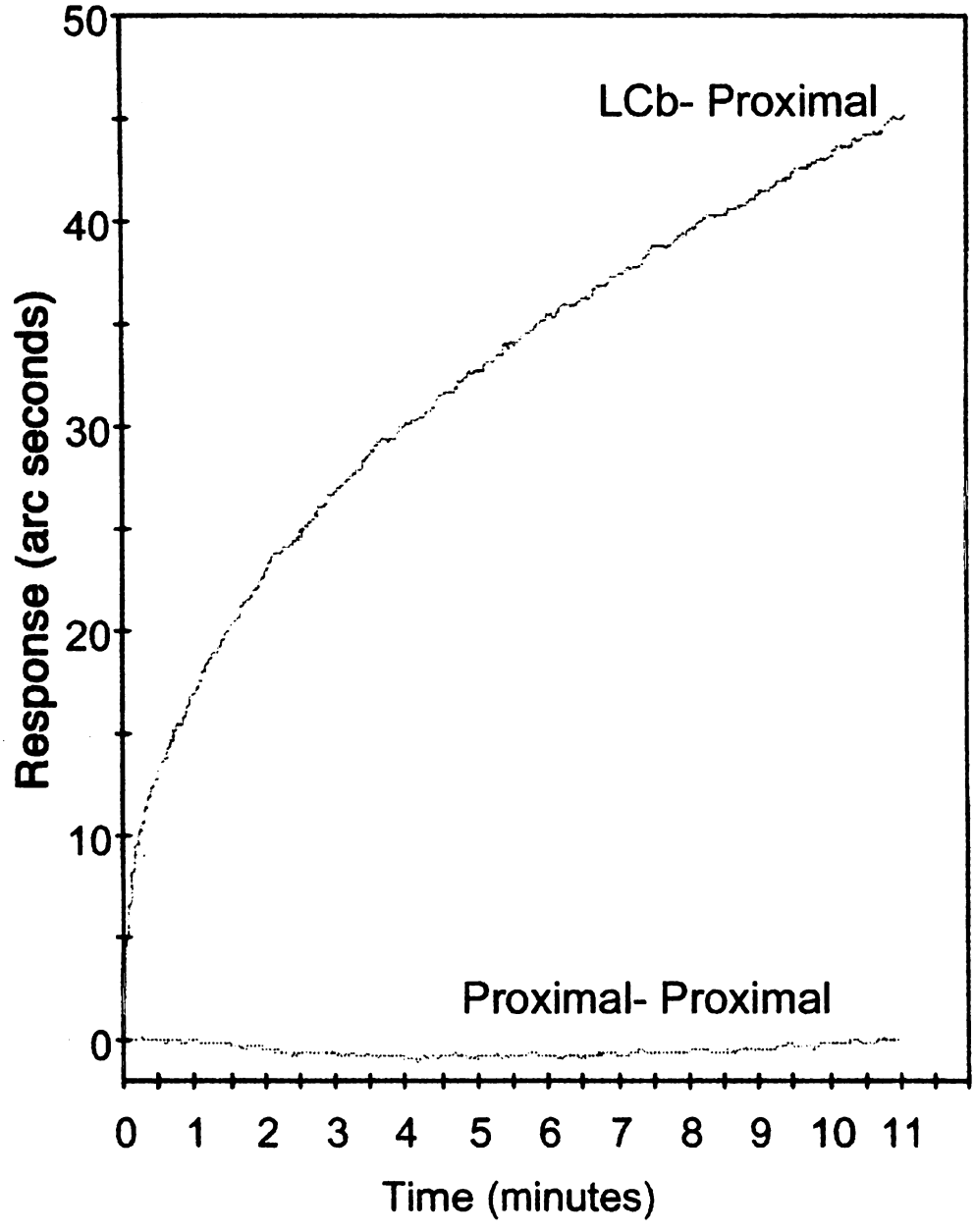


Figure 3.5: Surface Plasmon Resonance Fails to Detect Interaction Between Proximal Leg segments
Proximal leg segments were expressed and purified and then tested using an IAsys Plus biosensor. Proximal leg segments or clathrin light chain LCb were immobilized to a carboxylate cuvette, using EDC/NHS coupling chemistry. A solution of proximal leg segment in PBS was flowed over the surface and monitored for 10 minutes. Proximal leg segment failed to show any interaction with immobilized proximal leg segments (lower curve), but had a strong binding response to the control of immobilized clathrin light chain (upper curve).

(Figure 3.5). Measurement of changes in the refractive index of the immobilized protein on the chip reflects interaction with the protein in the flow chamber and, even for transient interactions, the kinetics of binding and dissociation can be determined. These measurements revealed that the proximal leg segments interacted strongly with immobilized light chain, but failed to show any interaction with immobilized proximal leg segments, at a pH favorable for clathrin assembly. Similar data were obtained with a longer proximal leg segment construct extending 1074-1590 (Ybe et al.). Thus, even transient dimer formation was not detectable between proximal leg segments, which were capable of binding to light chain subunits.

The failure to detect any interaction between proximal leg segments using multiple highly sensitive assays strongly indicates that isolated clathrin leg segment interactions have extremely weak affinity. This observation explains our previously published finding that an excess of proximal leg segments did not act as dominant negative inhibitors of clathrin assembly (Greene et al., 2000). The data reported here are consistent with a lack of binding of proximal leg segments to the whole triskelia tested in the assembly assay. In addition, we further demonstrate that binding clathrin light chain to proximal legs had no effect on their interaction. Together these data indicate that clathrin assembly relies on multiple interactions too weak to detect when individually isolated, such as would be found in a coordinated entropy-driven self-assembly reaction.

Chimeric Constructs Segment Interactions Assemble Using Cooperative Leg Interactions

The lack of detectable interaction between leg segments could be accounted for if the interactions between segments are extremely weak and rely on cooperative interactions, or if the leg segments do not mediate the interaction between clathrin molecules. If the latter possibility is correct, then assembly of Hub fragments (Liu et al., 1995), which are comprised of only three proximal leg segments and their trimerization domains, could only be accounted for by binding interactions between the trimerization domain and "knees" between proximal and distal leg segments, which are localized beneath the vertices in the assembled clathrin lattice and are present in the Hub constructs. In this case, the assembly inhibition activity of antibodies to the proximal leg segment would have to be ascribed to steric hindrance of the close approach of these domains. While a critical role for interactions apart from leg-leg binding during clathrin assembly is theoretically possible, the effects of the proximal leg-specific antibodies on assembly are also consistent with interference with weak binding between leg segments that is only sufficient to drive assembly when the segments are connected in a trimeric conformation to coordinate interactions.

We therefore sought to determine whether trimerization of leg segments with a non-clathrin trimerization domain would allow cooperative assembly between leg segments without the involvement of the trimerization domain itself. To this end, chimeric constructs were created with proximal or distal leg

segments attached to the C-terminal trimerization domain taken from the invariant chain (Ii) protein (Figure 3.6A).

In an antigen-presenting cell, trimerized Ii proteins occupy the peptide binding sites of three nascent MHC Class II complexes to stabilize their structure until they encounter antigen in the peptide loading compartment (Cresswell, 1996). The structure of the Ii trimerization domain has been determined (Jasanoff et al., 1998) and is composed of helix-and-loop segments not unlike the clathrin proximal leg segment CHCR motif (Ybe et al., 1999) that is predicted to continue into the clathrin trimerization domain. Like the clathrin trimerization domain, the three segments enter the C-terminal trimerization domain at a slightly puckered 120-degree angle, similar to the pucker of a clathrin triskelion. Ii trimerization domain constructs were engineered in which the Ii trimerization domain was joined to either the triskelion proximal leg (proximal-Ii) or distal leg (distal-Ii) segment by connecting loop segments between helices, creating a non-clathrin trimerization domain for the clathrin leg segments.

In order to confirm that the non-clathrin trimerization domain was able to produce a trimeric protein, sedimentation velocity experiments were performed to test the number of species in solution at pH 7.9, at which clathrin and Hub are disassembled. The results indicated that the purified proximal-Ii or distal-Ii chimeras each had a sedimentation coefficient distribution very similar to Hub (Figure 3.3B-D).

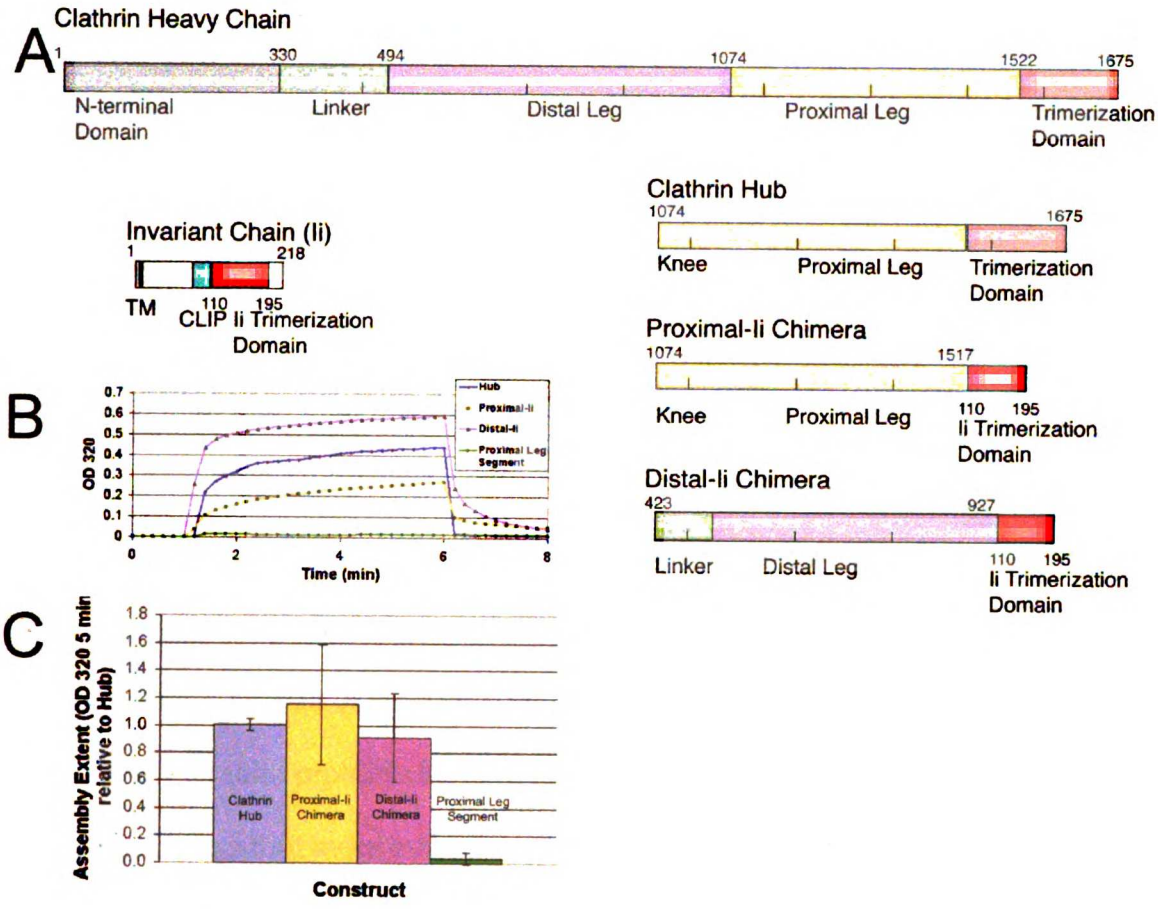


Figure 3.6: Chimeric Trimerized Clathrin Proximal or Distal Leg Segments Assemble

A. Chimeric constructs Distal-Ii and Proximal-Ii were created with clathrin distal or proximal leg segments, respectively, attached to the trimerization domain of the invariant chain protein. The bar diagrams indicate the protein domains of each donor protein from which chimeras were constructed and the domains in the resulting chimeras. Numbers indicate the residues flanking protein domains and the junctions where fragments were recombined. **B.** Purified chimeric proteins at pH 7.9 were induced to assemble by a drop in pH 6.7 through the addition of 1/25 volume of 1 molar MES pH 6.7, and turbidity (change in OD at 320 nm) was monitored for 5 minutes. Disassembly was induced by increasing the pH with an addition of 1/25 volume of 1M Tris pH 9 and monitoring OD 320 for 2 additional minutes as turbidity returned to baseline levels, demonstrating reversibility of the assembly process. Proximal leg segment, which is not trimeric and cannot assemble, served as a control. **C.** The chimeras had a similar overall extent of assembly to purified clathrin Hub in the pH 6.7 turbidity assay. Samples were induced to assemble as above and the extent of assembly (OD320 after 5 minutes) was recorded. Extent of assembly was normalized to average hub assembly over several preparations and averaged for this plot.

Each sample was composed of two species, the first of which at S~2.4 had the same apparent sedimentation coefficient as monomeric proximal leg (Figure 3.3A) and thus represents a monomeric species. For both hubs and chimeras, the distribution indicated a predominance of trimer (S=4-5) over monomer. The Hub and two chimeras are oblate ellipsoids and the data fit to an apparent frictional ratio of 1.6 for Hub (Figure 3.3B), indicating perhaps a slightly flatter disc shape than the two chimeras with apparent frictional ratios of 1.2 and 1.3 (Figures 3.3C,D). Thus, for both engineered constructs, the invariant chain trimerization domain was successful in trimerizing a significant proportion of chimeric leg segments in the expressed protein.

The chimeric proteins were then assessed for their ability to assemble under conditions that induce Hub assembly (Liu et al., 1995), using a light scattering assay. In this assay, self-assembly is distinguishable from protein aggregation in that it requires proper folding of the protein and is reversible (Kentsis et al., 2002). Spectrophotometric data indicated that both the proximal-Ii and distal-Ii constructs were able to assemble at either pH 6.2 or 6.7 (Figure 3.6B). Assembly of the chimeric constructs proceeded at a similar rate compared to that of Hub assembly (data not shown) and to a similar extent, as revealed by the endpoint spectrophotometric measurements (Figure 3.6C). The assembly reactions of the chimeras were reversible by raising the pH to 7.9, which causes assembled Hub to disassemble, thus demonstrating specificity and reversibility of their assembly (Figure 3.6B). Analysis of assembled chimeric constructs by electron microscopy revealed pseudo-lattices similar to those

previously observed for assembled Hub (Greene et al., 2000; Liu et al., 1995) in which assemble proceeds in multiple planes, so that lattice morphology is poorly stained (data not shown). In control reactions, the proximal leg segment did not produce a spectrophotometric signal upon lowering the pH to 6.2 or 6.7, indicating no formation of larger aggregates under these conditions, consistent with their lack of dimerization in other assays and providing a specificity control for assembly of the chimeras. Chimera assembly showing proximal-proximal and distal-distal interactions in a trimeric configuration complements previous data regarding the role of distal leg segments in clathrin assembly (Greene et al., 2000). These segments on their own were unable to interact effectively with Hubs during assembly, but when extended to include the terminal domain and multimerized by adaptors into a counter-Hub, these segments could influence Hub assembly to form a curved structure. Thus, clathrin assembly depends on coordination of weak proximal-proximal, proximal-distal and distal-distal leg interactions. Taken as a whole, our analysis of leg dimerization combined with the behavior of the trimerized leg segment constructs indicates that trimerization allows cooperativity between multiple simultaneous weak interactions of leg segments resulting in clathrin assembly.

Discussion

Here we define the nature of domain interactions contributing to clathrin assembly. We demonstrate that individual leg segment affinities are weak and rely on trimerization to allow multiple coordinated contributions to the formation of

a clathrin lattice. In addition, we confirm that clathrin light chain binding is not required for trimerized proximal leg segment interactions and we demonstrate that light chain binding does not promote interaction of individual proximal leg segments. Finally, we show that distal-distal leg segment interaction contributes to clathrin assembly, in addition to proximal-proximal and proximal-distal interactions that have been previously implicated (Blank and Brodsky, 1986; Greene et al., 2000; Smith et al., 1998). We cannot rule out additional roles for vertex-knee interactions in assembly from the data reported here, but the chimeric constructs demonstrate that leg segment interactions when trimerized are sufficient to sustain assembly without these interactions. From this analysis of clathrin assembly, we can infer important features of the thermodynamics of clathrin assembly that provide insight into the regulation of clathrin assembly and curvature in the cell.

Thermodynamic Properties of Clathrin Assembly

Multiple weak, simultaneous interactions have been proposed as the mechanism for the assembly of clathrin terminal domains with adaptor complexes and beta-arrestin (Kirchhausen, 2000b). Here we demonstrate that similar principles govern clathrin self-assembly. Thus, as hypothesized over twenty years ago (Crowther and Pearse, 1981), we now have evidence that the process of clathrin coat formation joins actin, amyloid, tubulin, and others in the large and diverse group of entropy-driven condensation reactions, for which the tobacco mosaic virus rod-forming polymerization process is a model (Lauffer,

1975). Based on what is known about these reactions, we must consider that the primary driving force in clathrin assembly is hydrophobic, rather than electrostatic. Typically, electrostatically mediated assembly would involve early formation of weak long-range interactions, followed by the docking of proteins into a higher affinity configuration requiring strong hydrophobic interactions (Schreiber and Fersht, 1996). The predicted electrostatic interactions between clathrin domains (Ybe et al., 1998) would be unlikely to act attractively at long range due to electrostatic repulsion between charged, predominantly acidic leg segments, especially if acidic clathrin light chains are bound (Chapter Two). Furthermore, the absence of strong binding affinity between isolated leg segments suggests that strong short-range salt bridges are also absent in this interface However, electrostatic repulsion can function as a specificity determinant to block misalignment of leg segments, as has been seen for other self-assembling systems (Coleman et al., 1999). Thus, the weakness of the interaction between leg segments demonstrates a possible role for repulsive electrostatic specificity-determining interactions which are balanced out by hydrophobic attractive interactions, as previously suggested for clathrin assembly (Nandi and Edelhoch, 1984). In fact, the pH and salt sensitivity observed for clathrin assembly is common among entropy-driven polymerizations (Lauffer, 1975) and may result in part from a simple reduction of the electrostatic repulsion between acidic heavy chains near the isoelectric point of the leg segments (Chapter Two). The possibility has been raised that the ionization state of surface-accessible histidines is crucial to the pH-dependence of the assembly

process (Ybe et al., 1998). Such histidines may participate in salt bridge interactions, stabilize the heavy chain structure (Sali et al., 1988) or mediate regulation of the dimerization with zinc or other metal ions (Christianson and Alexander, 1989) (Chapter Two). However, given the evidence for electrostatic interaction, the most likely explanation is the existence of pH-dependent electrostatic hydrogen bonds via histidines, as has been found in other proteins (Mejean et al., 2001).

The interface between two proximal leg segments in an assembled basket is predicted to have a buried surface area on the order of 1500 $Å^2$ (Chapter Two), and would be expected to have affinity contributed by extensive van der Waals associations, several charged and uncharged hydrogen bonds, and possibly a small contribution from one or more salt bridges (Prevelige, 1998). Taking all of these properties into consideration, along with the date reported here, we propose that clathrin self-assembly is driven primarily by the entropic gain of the hydrophobic effect when the assembled leg segments exclude water from their interface.

Cellular Regulation of Clathrin Assembly and Curvature

The unstable and weak interaction between isolated leg segments raises the question of how nucleation of clathrin assembly occurs in the cell. Clathrin polymerization appears to proceed through an intermediate structure with 4-7 triskelia (Schwarz et al., 1989). With such weak interactions between isolated leg segments as demonstrated here, we would expect that intermediates smaller

than this would be unstable. In the cell, adaptors and accessory co-factors AP180 (Ford et al., 2001) and epsin (Ford et al., 2002) are therefore required to play a key role in initiating assembly through recruitment and stabilization of clathrin polymerization on the membrane. Based on the thermodynamic considerations above, assembly could also be readily stimulated by locally lowered pH due to proton pump activity (Heuser, 1989) or recruitment to acidic membrane lipids, which may have a similar effect on clathrin assembly (Maezawa et al., 1989) as they do on laminin assembly (Freire and Coelho-Sampaio, 2000). Clathrin propagation, when recruited to the membrane in the vicinity of a stable clathrin-coated pit locus (Gaidarov et al., 1999) and stimulated to undergo a coordinated assembly, would occur more readily than nucleation because it would by nature involve multiple segment interactions to join an existing lattice.

A second issue relevant to our data is how curvature can be introduced into a clathrin lattice. There are three factors that contribute to lattice curvature. The first is the geometric requirement for introduction of twelve pentagons into a hexagonal lattice in order to achieve a closed sphere or vesicle coat (Kaneseki and Kadota, 1969). The second factor is the shape of the triskelion itself and its inherent flexibility. In an assembled clathrin basket, a triskelion has increased pucker at the vertex as well as greater curvature of the legs than it displays in solution (Musacchio et al., 1999). It has been suggested that the knee region between proximal and distal leg segments, which has little intrinsic curvature in solution (Yoshimura et al., 1991) and no inherent additional elasticity compared

with the proximal or distal leg segments (Jin and Nossal, 2000; Kocsis et al., 1991), is responsible for most of the curvature in the assembled basket (Kirchhausen et al., 1986; Musacchio et al., 1999). We have further shown that ultimate flexibility at the knee by dissociation of the distal from the proximal leg domain is compatible with introduction of curvature into a clathrin lattice (Greene et al., 2000). In addition, theoretical analyses have indicated that triskelia may form pentagons as well as hexagons with some torsional strain but little overall free energy cost (Jin and Nossal, 1993), particularly in the presence of tetrameric adaptors (Nossal, 2001). Thus changes in triskelion conformation can readily occur and contribute to lattice curvature without posing a significant energy barrier. The third factor regulating curvature of a clathrin lattice at the cellular membrane is the participation of adaptors and accessory proteins in this process. Tetrameric adaptor proteins AP1 (at the TGN) and AP2 (at the plasma membrane) work with monomeric adaptors such as GGAs (Robinson and Bonifacino, 2001) and Dab2 (Mishra et al., 2002) to induce invaginations in the clathrin lattice. Epsin, which binds lipid, tetrameric adaptors and clathrin, is able to accomplish this curvature even without the aid of tetrameric adaptors (Ford et al., 2002; Nossal and Zimmerberg, 2002). In this case, epsin's mechanism of action relies on deformation of the membrane by insertion of its hydrophobic domain and its ability to manipulate clathrin to form pentagons through binding to clathrin. Whether epsin and other adaptors can perform clathrin manipulations within a preformed hexagonal array or only at the edge of such a lattice by

interaction with newly recruited triskelia depends on the energetics of the interaction of individual clathrin leg domains.

Here we show that the energetics of clathrin self-assembly are compatible with lattice-modifying proteins being able to produce a bud from the interior of an existing clathrin lattice, essentially to regulate rearrangement of a lattice. Evidence presented here for weak interactions between clathrin leg segments supports earlier theoretical free energy calculations for the assembly process (Nossal, 2001), which estimated that the free energy gain on the addition of an individual triskelion to a partially formed lattice is quite small, on the order of thermal fluctuations. Such a low free energy threshold permits rearrangement in the center of a lattice by cytosolic exchange of triskelia (Brown and Petersen, 1999; Wu et al., 2001) and is compatible with removal of a triskelion for rearrangement of a hexagon into a pentagon (Kaneseki and Kadota, 1969; Nossal, 2001). While our data support such a model for lattice rearrangement, they do not absolutely contradict the suggestion that such pentagonal arrangements are limited to the edges of the lattice (Harrison and Kirchhausen, 1983; Kirchhausen, 2000b). In either case, we would suggest that the presence of lattice-modifying proteins in a budding segment of membrane allows rearranged or new triskelia to be oriented so that they interact with more than one leg segment of the partially or fully completed lattice at the same time, stabilizing the interaction when compared with isolated leg segment interactions. Thus, a delicate balance between individually weak, repulsive and attractive electrostatic, hydrogen bonding, and hydrophobic interactions can be exploited

by regulatory proteins. This allows clathrin self-assembly to be spontaneous, readily reversible and finely regulated in the cell.

Chapter Four: Clathrin Specialization Explained by Phylogenetic Analysis

Abstract

It has long been known that vertebrates contain two clathrin light chains LCa and LCb, but the functional difference between them remains unclear. In recent years a second clathrin heavy chain has been identified in humans on chromosome 22, which has a distinct function from the conventional clathrin heavy chain on chromosome 17. Here we construct phylogenetic trees for clathrin heavy chain and clathrin light chain. The position of the radiation nodes for the second form of both clathrin heavy and light chains suggests each gene was duplicated between 750 and 420 million years ago. Efforts are ongoing to identify additional paralogous genes surrounding clathrin genes to determine whether these duplications arose as part of a large-scale (possibly genomic) duplication during chordate evolution. Such a duplication event would have allowed rapid specialization to support increasingly sophisticated neuronal and muscular functions. Further investigation of clathrin duplications may yield clues to the functional differences between the two vertebrate forms of clathrin heavy chain and clathrin light chain.

Introduction

Conventional clathrin heavy chain CHC17 has fascinated researchers ever since it was identified in insect oocytes endocytosing yolk protein in 1964 (Roth and Porter, 1964). It was subsequently cloned and sequenced in rat (Kirchhausen et al., 1987a) and mapped to chromosome 17 in humans (Dodge et al., 1991). No evidence was found of additional clathrin heavy chain proteins in humans by biochemical studies or Southern blotting. It was therefore entirely unexpected that a novel form of clathrin heavy chain would be identified on chromosome 22 in the course of sequencing the human genome, now called CHC22 (Brodsky, 1997; Gong et al., 1996; Holmes et al., 1997; Kedra et al., 1996; Long et al., 1996; Sirotkin et al., 1996). The function of CHC22 is still not fully defined. By contrast, the presence of two forms of clathrin light chain was shown in 1983 (Brodsky and Parham, 1983), and the sequence of each vertebrate clathrin light chain was documented in 1987 (Jackson and Parham, 1988; Jackson et al., 1987; Kirchhausen et al., 1987b). Interestingly, both LCa and LCb appear to bind CHC17 in-vitro but not CHC22 (Liu et al., 2001b), and they compete with one another for binding sites on the heavy chain in an apparently random distribution with redundant function (Brodsky et al., 1991). The investigation of the evolutionary divergence between the two forms of clathrin heavy and light chains described here may shed light on the function of CHC22, any specialized differences between light chains LCa and LCb, and whether regulatory light chain subunits exist for CHC22. In addition, determination of the circumstances under which the gene duplications occurred

may lend support to the theory of en-bloc genome duplication during chordate evolution, and may possibly help to identify an extended locus of genes involved in membrane traffic.

The Clathrin Sequences

Efforts to determine the function of the novel CHC22 form of clathrin in humans were undertaken by our laboratory upon its discovery (Liu et al., 2001b). CHC22 was shown to be expressed at high levels in cardiac and skeletal muscle and in the testis, leading some researchers to call it muscle- specific clathrin (Brodsky, 1997). CHC22 has 90% protein sequence similarity and 85% protein sequence identity with CHC17. The homology is strong throughout the sequence (Figure 4.1A) with only the C-terminal trimerization having below 80% protein sequence identity. The DNA shows identical intron/ exon boundaries (data not shown), and there is 74% nucleotide sequence identity between CHC17 and CHC22 genes. CHC22 is 1640 amino acids long compared with 1675 for CHC17. While the function of CHC22 is still undetermined, it does not appear to function in endocytosis or conventional membrane trafficking pathways alongside traditional CHC17. Studies suggest a structural or cytoskeletal role for CHC22 (Liu et al., 2001b; Towler et al., 2002), possibly in muscle cell differentiation. CHC22 orthologous sequences have now been found in chicken and fish in the course of this study, suggesting its presence in all vertebrate genomes. Parasites, plants, and invertebrates each appear to have a single heavy chain,

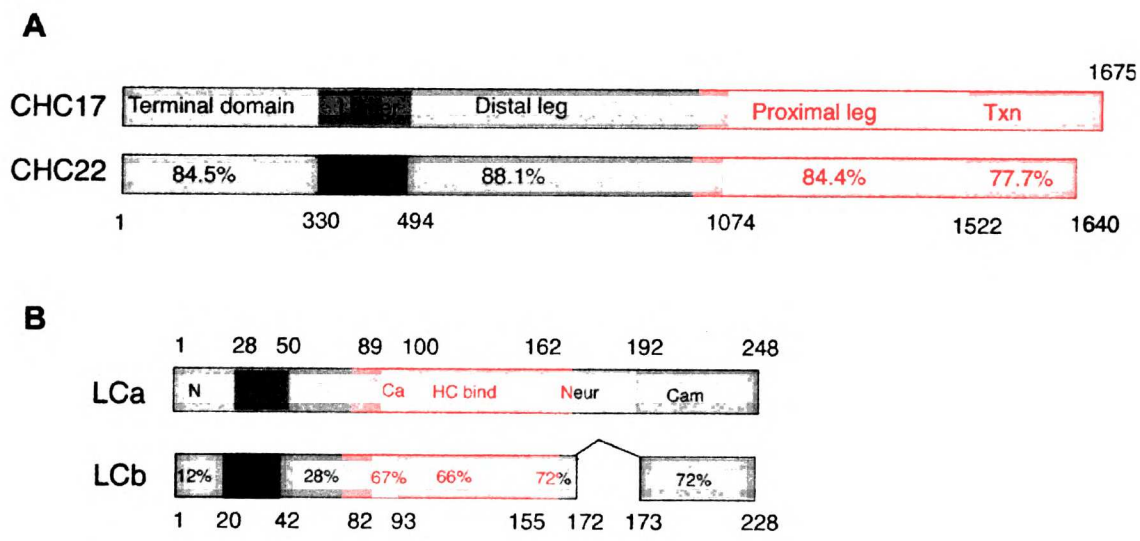


Figure 4.1: Clathrin Domains and Sequence Conservation
The two human forms of clathrins are compared side by side, with percent identity noted in each domain. **A.** Clathrin heavy chain CHC17 shows high conservation to CHC22 throughout the length of the sequence. Domains include the globular N-terminal domain (blue) and linker (green), distal leg segment (purple), proximal leg segment (yellow) and trimerization domain (orange). Homology is weakest in the C-terminal trimerization domain, which is somewhat shorter in CHC22. **B.** Clathrin light chains LCa and LCb show variable conservation between domains, with the N-terminal domain (N, blue)) having lowest homology and the adjacent alpha-consensus domain (Con, green) being absolutely conserved. Other domains include the calcium binding domain (Ca, yellow), the heavy chain binding domain (HC bind, orange), the neuron-specific insert (Neur, magenta) and the calmodulin binding domain (Cam, cyan).

where plant and some of the other sequences share the same short length as CHC22.

While invertebrates and yeasts have a single clathrin light chain and plants appear to have multiple light chains, vertebrates have two light chains. Each of the mammalian light chains is encoded on a different chromosome and they share 60% sequence identity. Both light chains are expressed in every cell type and they show no marked preferential ability to bind heavy chain relative to one another (Brodsky et al., 1987), but bind in accordance with the ratio of the protein expression level between the two forms. Within each cell there are therefore populations of CHC17 triskelia with the three binding sites occupied by three LCa molecules, three LCb molecules, two LCa and one LCb, or two LCb and one LCa. All light chain binding sites are fully occupied in the cell, and light chains do not appear to dissociate from heavy chains once bound (Brodsky et al., 1991). In humans, the sequence homology between the two light chain forms varies widely between the domains of the primary sequence (Figure 4.1B), with the conserved consensus domain (28-50 LCa, 20-42 LCb) being absolutely conserved and the neuronal insert/ calmodulin binding site (162-248 LCa, 155-228 LCb) having 72% homology. The least conserved region is the far N-terminus (1-28 LCa, 1-20 LCb).

The apparent functional redundance of the two vertebrate light chains in tissue culture cells has led to studies to determine any specialized functions or differences between the two. The potential exists for differential regulatory control of the light chains by phosphorylation, binding of Hsc70, or disulfide bond

formation. LCb has a casein kinase II serine phosphorylation site near the N-terminus which is absent in LCa (Chu et al., 1999; Hill et al., 1988). In contrast, LCa has a higher affinity binding site at positions 47-71 for Hsc70 (DeLuca-Flaherty et al., 1990), the heat shock protein involved in uncoating vesicles that have budded from the membrane (Newmyer and Schmid, 2001; Rothman and Schmid, 1986). The C-terminus of each light chain has a cysteine conserved at position 228 which may be involved in transient disulfide bond interactions. However, LCb has an additional disulfide bond (position 237), as does the neuronal form of LCa (position 195) (Parham et al., 1989). Both forms of clathrin light chain bind calcium (Nathke et al., 1990) and calmodulin (Pley et al., 1995), but the former interaction may require high levels of calcium that are not physiologically relevant (Nathke et al., 1990). However, the only observed functional difference between the two vertebrate forms was the observation that LCb is expressed at a higher ratio in cells with a role in regulated secretion (Acton and Brodsky, 1990).

The Genomic Duplication Theory

Susumu Ohno proposed that during chordate evolution two polyploidizations, or en-bloc duplications of the entire genome from N to 4N, allowed rapid specialization of proteins to support increasingly complex organisms (Ohno, 1970). The first duplication in this 2R hypothesis occurred after the split between primitive chordates represented by *amphioxus* (lancelets) and Cephalochordates (jawless fishes, such as the hagfish or lamprey), some

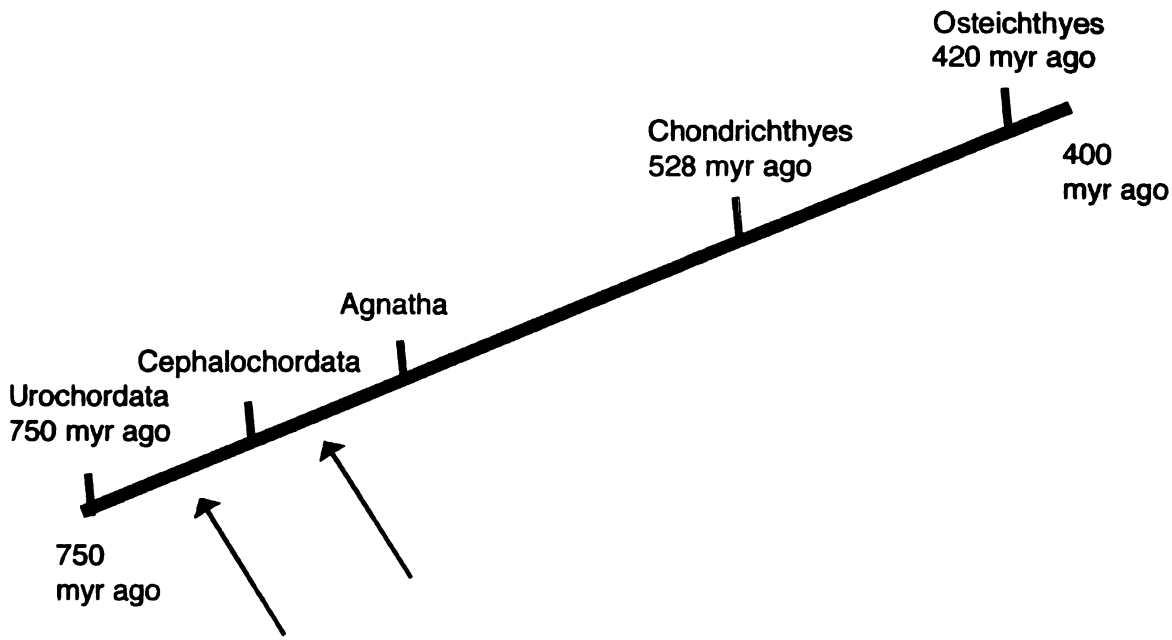


Figure 4.2: A Timeline of Vertebrate Evolution and the Hypothesized Genome Duplication
Radiation of the branches leading to modern day chordate and vertebrate species is noted along an evolutionary timeline. Arrows denote the timing of the theorized genome duplications during chordate evolution.

750 million years ago. The second duplication may have occurred as recently as 528 million years ago, before the split between Chondrichthyes (cartilaginous fishes, such as sharks) and Osteichthyes (bony fishes, such as pufferfish or zebrafish) (Abi-Rached et al., 2002) (Figure 4.2). The duplicate chromosomes arising from such an event, containing duplicated genes called "paralogues" of one another, would then evolve over time. On each chromosome copy, many of the genes would mutate to lose function, whereas other genes would be retained with their mutations, leading each copy to a specialized or unique function within the same gene family (Wolfe, 2001). This duplication would then allow rapid diversification of protein families to support increasingly complicated systems in the developing chordate and vertebrate lineage. Eventually, sequence divergence between the duplicated chromosomes would return the polyploid genome to the diploid state found in the human genome today, a process called paleopolyploidy.

While this genomic duplication theory has gained acceptance in recent years (Spring, 2002), skeptics remain who believe that while polyploidization is evident in some genomes such as zebrafish (Taylor et al., 2001) and pufferfish (Williams et al., 2002), the as-yet incompletely annotated human genome does not provide strong support for en-bloc genome duplications. Specifically, if a genome duplication occurred, it may have resulted from one rather than two duplication events (Wolfe, 2001).

The prototypical examples in support of genome duplication are the HOX gene clusters (Stellwag, 1999). Hox genes are crucial during embryogenesis for

bilateral symmetry and overall morphology. The 38 HOX genes found in humans, arranged in four clusters on separate chromosomes, appear to have descended directly from a single cluster in the primitive chordate ancestor of amphioxus through chromosomal duplications during the time frame proposed for genome duplication, a perfect example of four modern gene clusters resulting from two duplications of a single ancestral gene locus. Moreover, the clustered organization of Hox genes suggested that entire blocks of genes like this would be found functioning together at a single locus, having duplicated and evolved in tandem. A similar clustered locus of genes was found for the the MHC (Major histocompatibility complex) (Kasahara et al., 1999), with further investigation showing evidence of genome duplication in MHC vertebrate evolution (Abi-Rached et al., 2002; Kasahara et al., 1999). Phylogenetic studies of other membrane trafficking proteins have generated support for the 2R hypothesis, suggesting that COPI subunits (involved in Golgi to ER membrane traffic) evolved away from a common ancestor with the clathrin-associated AP adaptor subunits after a genome-wide duplication (Schledzewski et al., 1999). In an ideal scenario, a chromosomal locus for genes involved in membrane traffic might be located near one of the clathrin genes, waiting for discovery.

Primitive Chordates and Vertebrates and the Evolution of Complex Systems

Here I consider the early vertebrate families that evolved from our common chordate ancestor during the development of vertebrate complexity.

This sheds light on which functions our vertebrate ancestors developed in the time frame when genomic duplication is proposed to occur. In particular, the evolutionary acquisition of specialized functions for muscular and neuronal tissue is the potential driving force leading to clathrin diversification. The modern day descendants of chordates and early vertebrates and the novel features distinguishing each from its ancestors on the chordate evolutionary tree are detailed in this section.

The tunicates, or Urochordata, are perhaps the most primitive chordates. These acraniate vertebrates include ascidians such as *ciona*, the sea squirt. The adult sea squirt, a siphoning creature commonly attached to submerged rocks, hardly resembles a close ancestor of humans. However, the tadpole larva has a dorsal notochord and nerve chord, firmly establishing it as a chordate (Holland, 2002; Meinertzhagen and Okamura, 2001). The *ciona* genome is 17 times smaller than the human genome and its radiation from the common chordate ancestor occurred about 750 million years ago, before the first proposed round of gene duplication, making ciona a good starting point for considering the development of vertebrates.

The next group to radiate from the ancestral chordates was the Cephalochordata, also known as *amphioxus* or more commonly as lancelets. The lancelets are small fish-like marine chordates with no scales and a simple dorsal notochord and nerve chord that persist throughout the animal's life. Lancelets show a far more sophisticated musculature than ciona, with defined blocks of muscles to allow efficient swimming. They have a cilia-based plankton

feeding system. They have no distinguishable head and no eyes, nose, or ears, though they do have primitive light receptors (Raven and Johnson, 1986). *Amphioxus* appears to have a simple genome, which has not undergone duplication or rearrangement in millions of years. While Ohno proposed the first round of genome duplication before *amphioxus* radiation, most modern studies estimate the first duplication happened just after *amphioxus* radiation (Furlong and Holland, 2002).

The first primitive vertebrates that evolved had a distinct skull (cranium) and brain, and the first traces of a vertebral column. This tremendous leap in neuronal complexity is evident today in the Agnatha (jawless fishes, such as lampreys and hagfish). Unlike lancelets, hagfish lack cilia for feeding on plankton and instead force water through their pharynx using muscle contraction. Thus both muscular and neuronal development evolved rapidly between Cephalochordates and Agnatha (Klein et al., 2000; Raven and Johnson, 1986).

Chondrichthyes, which are cartilaginous vertebrate fishes such as the ancestral shark, skate, ray, and dogfish shark, were the first Gnathostomes (jawed vertebrates) that evolved. Their jaws, fins, and light skeletons allowed them to become efficient swimmers and predators. The neuronal development required for predator sensory perception and behavior is clear (Kusunoki and Amemiya, 1995). The liver contributes a significant portion of the body weight, providing buoyancy and possibly representing a more sophisticated metabolism appropriate for the diet of a predator. Constant swimming is required to push water past the gills to provide oxygen, revealing further muscular development

(Raven and Johnson, 1986). Cartilaginous fishes are thought to have evolved just after the second round of genomic duplication about 528 myr (million years) ago (Abi-Rached et al., 2002; Furlong and Holland, 2002).

For almost 100 myr, cartilaginous fishes were the only vertebrates. Bony fishes, or Osteichthyes, evolved about 420 myr ago, possibly in fresh water. A bony skeleton and vertebral column that was present in the primordial ancestor of the cartilaginous fish is retained in modern bony fishes such as pufferfish or zebrafish. The swim bladder, used to regulate buoyancy, frees bony fishes from the need to continually swim and thus allows heavier skeletal structure. Most modern fish are derived from this class of animals (Grunwald and Eisen, 2002; Raven and Johnson, 1986).

Thus the muscular, neuronal, respiratory, and metabolic systems were undergoing significant development during the time frame in which the vertebrates evolved and genome duplication took place. Perhaps it is unsurprising then that the strongest support for the 2R genome hypothesis comes from researchers studying evolutionary development (Postlethwait et al., 1998), who maintain that the increase in sophistication required for chordates to develop these pathways rapidly was best met by duplication of the entire genome to specialize existing proteins rather than evolution of an entire set of proteins *de novo*.

Single or Local Gene Duplication: Alternatives to Genome Duplication Theory

Alternative explanations for the rapid evolution of gene families have been put forward by skeptics of Ohno's theory (Wolfe, 2001). The chances that a second or third gene evolved *de novo* and became highly homologous to a gene already in the genome are considered remote, by any standard. However, there are several possibilities that fall short of whole genome duplication that should be considered. Aneuploidy, or the presence of extra or fewer copies of a given chromosome, is one possibility. The paralogous chromosomal regions identified to date (Spring, 2002) could be isolated examples of chromosome duplication rather than extended duplication of the entire genome.

Also, continuous small-scale duplications, such as tandem or segmental duplications, have been shown to have a significant role in the same time frame as genome duplication (Gu et al., 2002). A molecular clock for such small duplications was estimated by considering the rate at which gene duplications occurred in time frames *not* associated with genome-wide duplication, such as before 750 million years ago or after 400 million years ago. Thus the overall number of gene duplication events can be expressed as mR + C, where m is the number of genome wide duplications (likely 1 or 2), R is the genome-wide duplication, and C is the continuous small-scale duplication component.

Over time, tandem genes on a single chromosome resulting from such small-scale duplications could be transposed to different genomic loci, away from the originally amplified "founder gene" and from other paralogous genes.

Tandem arrays of coding or noncoding sequences are considered by nature dynamic and subject to ongoing gene deletions and expansions due to unequal crossing-over or replication slippage.

While some have interpreted the duplication of many proteins in the same time frame as evidence of genome-wide duplication, others point out that proteins that interact must co-evolve to maintain those interactions (Goh and Cohen, 2002). In some cases this may involve duplication to mimic the new regulatory or other features of the novel interacting gene copy. However, in the case of clathrin, the light chain duplicates both remained associated with a single heavy chain gene (Liu et al., 2001b), suggesting light chain and heavy chain duplication, if it occurred in the same time frame, did not result from co-evolutionary pressures to allow each light chain to specialize to one of the heavy chain variants.

Results

Here I investigate the phylogeny and evolution of clathrins. Three approaches were taken in these studies. First, through literature searches and searches of genomic databases, the genomic loci surrounding the clathrin genes were identified and characterized. Clathrin sequences from all organisms were then gathered from the public databases and used to create an alignment and subsequently a phylogenetic tree, to determine when in evolutionary time clathrin gene duplication occurred. Finally, a map-based approach to finding paralogons in the chromosomal segments surrounding the clathrin genes was undertaken to

determine whether clathrin duplication was a result of a small or large scale gene duplication event.

The Chromosomal Loci of Clathrins

Conventional clathrin heavy chain CHC17 is located on chromosomal segment 17q23.2 (Dodge et al., 1991). The 17q23 segment has been shown to be amplified in poor-prognosis breast cancer (Barlund et al., 2000; Monni et al., 2001) and in other tumors (Anderson et al., 2002), implicating several genes in this region with the progression of the disease. Cancer is the likely outcome of altered kinetics of endocytosis and recycling of growth factor receptors, such as might result from amplification of clathrin (Floyd and De Camilli, 1998). However, what is more relevant to this study is the propensity for amplification in this segment of DNA. A study of the bovine chromosome containing the orthologous CHC17 gene suggests that clathrin CHC17, MTMR4 (myotubularin related protein 4), and MMD (monocyte-to-macrophage differentiation-associated) are part of a single chromosomal block produced by radiation-induced breakage during radiation mapping. This radiation hybrid linkage group has a less stable centromeric location in cattle (Kirkpatrick et al., 2002), as compared with its mid-chromosomal location in humans (Dodge et al., 1991). Similarly, the porcine gene corresponding to the human CHC17 (Thomsen et al., 1998) maps to porcine chromosome 2, as opposed to the remainder of the human chromosome 17 which maps to porcine chromosome 12 (Shi et al., 2001). These studies suggest the CHC17 gene lies in a chromosomal region that was once prone to

rearrangement or duplication, and may still tend toward instability under adverse conditions.

Moreover, the genomic region 22q11.21 around CHC22 is even more unstable and is commonly deleted in various tumors and congenital disorders (Edelmann et al., 1999; Shaikh et al., 2001), such as cat-eye syndrome, der(22) syndrome, and velo-cardio-facial syndrome/ DiGeorge syndrome (VFCS/DGS), a fact which led to the initial discovery of the CHC22 gene. VCFS/DGS is the most common microdeletion disorder in humans with symptoms of cleft palate, heart defects, learning disabilities, and behavioral disorders (Ryan et al., 1997). Studies have revealed that the 22q11 locus is prone to deletions (Shaikh et al., 2000) resulting from meiotic inter- and intra- chromosomal homologous recombination events between two segmental duplication breakpoints (also called low copy repeats or LCRs) on either side of CHC22 and scattered throughout 22q11(Edelmann et al., 1999). Further, most single nucleotide polymorphisms (SNPs) in LCR segments, including those in 22q11, are in fact paralogous sequence variants (PSVs) resulting from segmental duplications of the human genome in the progeny of subjects with a predisposition for certain specific genetic rearrangements (Estivill et al., 2002). Interestingly, the pericentromeric region of chromosomes, including the loci near CHC22, is identified as a preferential site for integration of transposed duplicated DNA (Bridgland et al., 2003; Horvath et al., 2000). This suggests not only that 22q11 is in an unstable portion of the genome, but also that CHC22 may have integrated at that site quite independently of the duplication event that generated

its existence. Thus, the absence of a mouse CHC22 gene in the orthologous mouse chromosome 16A region (Botta et al., 1997; Lund et al., 2000), confirmed during the course of this work by the discovery of a mouse pseudogene at that chromosomal location (mouse chr16: 17421-17396k, contig_53192, contig_53193, and contig_224459) does not rule out the existence of CHC22 elsewhere in the mouse genome, particularly since CHC22 orthologues appear to exist in dogs, rabbits, and primates (Holmes et al., 1997) and have been identified in this work in the genomes of chicken and pufferfish (Table 4.5). CHC22 has differentially spliced variants (Kedra et al., 1996; Long et al., 1996), of unknown function. Perhaps the most unexpected discovery is the close homology of CHC22 to CHC17, given the dynamic and unstable region of the genome in which it is found. Presumably the selective pressure to maintain features of its sequence is quite strong, which points toward a crucial, if yet unspecified, role in the cell.

In contrast to the instability of the genomic sequences surrounding the heavy chain genes, the light chain genes are located in chromosomal regions of relative stability. The sequencing of the human genome has revealed that the initially reported genomic location for the clathrin light chains was incorrect (Jackson and Parham, 1988). The gene for light chain A (LCa) is found at 9p13.3. LCa is surrounded by genes whose single nucleotide polymorphisms lead to disease. Its downstream neighbor genes include GNE, which is associated with hereditary inclusion gene myopathy (IBM2) (Eisenberg et al., 2001), and PAX5, which is fused with other genes in leukemia (Cazzaniga et al.,

2001). Meanwhile, just upstream lies the tumor suppressor RECK (Eisenberg et al., 2002b). Light chain B (LCb) is located at a chromosome end, 5q35.2. Deletion of this 5q terminus is associated with central nervous system structural defects (Angle et al., 2003). Each light chain gene has differentially spliced variants (Jackson et al., 1987; Kirchhausen et al., 1987b). The longer forms of light chain have a neuronal insert believed to function in the coordinated and rapid speed of endocytic uptake of neurotransmitter required in brain tissue. Interestingly, two clathrin light chain pseudogenes, here called LCps8 and LCps12, are also found in the human genome. LCps8 is found at 8p22 and LCps12 at 12p13.31. Each lacks introns, making them likely products of random mRNA re-incorporation into the genome. Such processed pseudogenes are common; it has been estimated that one of every five genes has spawned a processed pseudogene in the human genome (Harrison and Gerstein, 2002). Ps8 would code for a full-length protein, except for several single nucleotide deletions and one single nucleotide insertion that make proper translation impossible. The closest coding region to Ps8 is the putative N33 prostate cancer tumor suppressor. Ps12 is missing the first 45 amino acids and has one single nucleotide deletion and two stop codons making proper translation impossible. The closest coding region is for predicted protein NT_024397, of unknown function.

Phylogenetic Analysis: A Tree Based Approach

A preliminary phylogenetic analysis suggested that CHC22 and CHC17 evolved from a common vertebrate ancestor (Sirotkin et al., 1996). We have extended that analysis by searching for all known clathrin heavy chain sequences or sequence fragments, including protists, parasites, plants, fungi, and sea squirt (see Table 4.5 at end of chapter for accession codes). This represents a significant leap from the sequences currently published or aligned and allows a fuller view of the clathrin family and its conserved sequences. In addition, efforts have been made to specifically locate CHC22 sequences, leading to the identification of sequence fragments for chicken and pufferfish CHC22. An alignment was created of all heavy chain sequences (Table 4.3, at end of chapter), and used to construct a phylogenetic tree (Figure 4.3).

The results indicate that CHC22 and CHC17 diverged sometime after the radiation of the tunicate sea squirt (*Ciona intestinalis*) but before the radiation of the modern osteichthyes Japanese pufferfish (*Takifugu rubripes*). This narrows down the time frame for clathrin heavy chain gene duplication to between 750 and 420 million years ago, squarely in the time frame in which the theorized en-bloc genome duplication was occurring. CHC22 is somewhat closer in sequence to its invertebrate ancestors than CHC17. A Bayesian algorithm was then used to infer the ancestral sequence of clathrin at the node for CHC22 radiation (Table 4.3, listed as ANCESTRAL CHC). The program infers an amino acid for each position in the alignment, including those where most species have a gap--

Figure 4.3: Bayesian Phylogenetic Tree of Clathrin Heavy Chains
A phylogenetic tree for clathrin heavy chain proteins was constructed using MrBayes. Clades for CHC22, mammalian and bird CHC17, fish CHC17, and heavy chains in insects, roundworms, fungi, and plants are well defined. Full species names and accession codes for sequences are noted in Table 4.5.

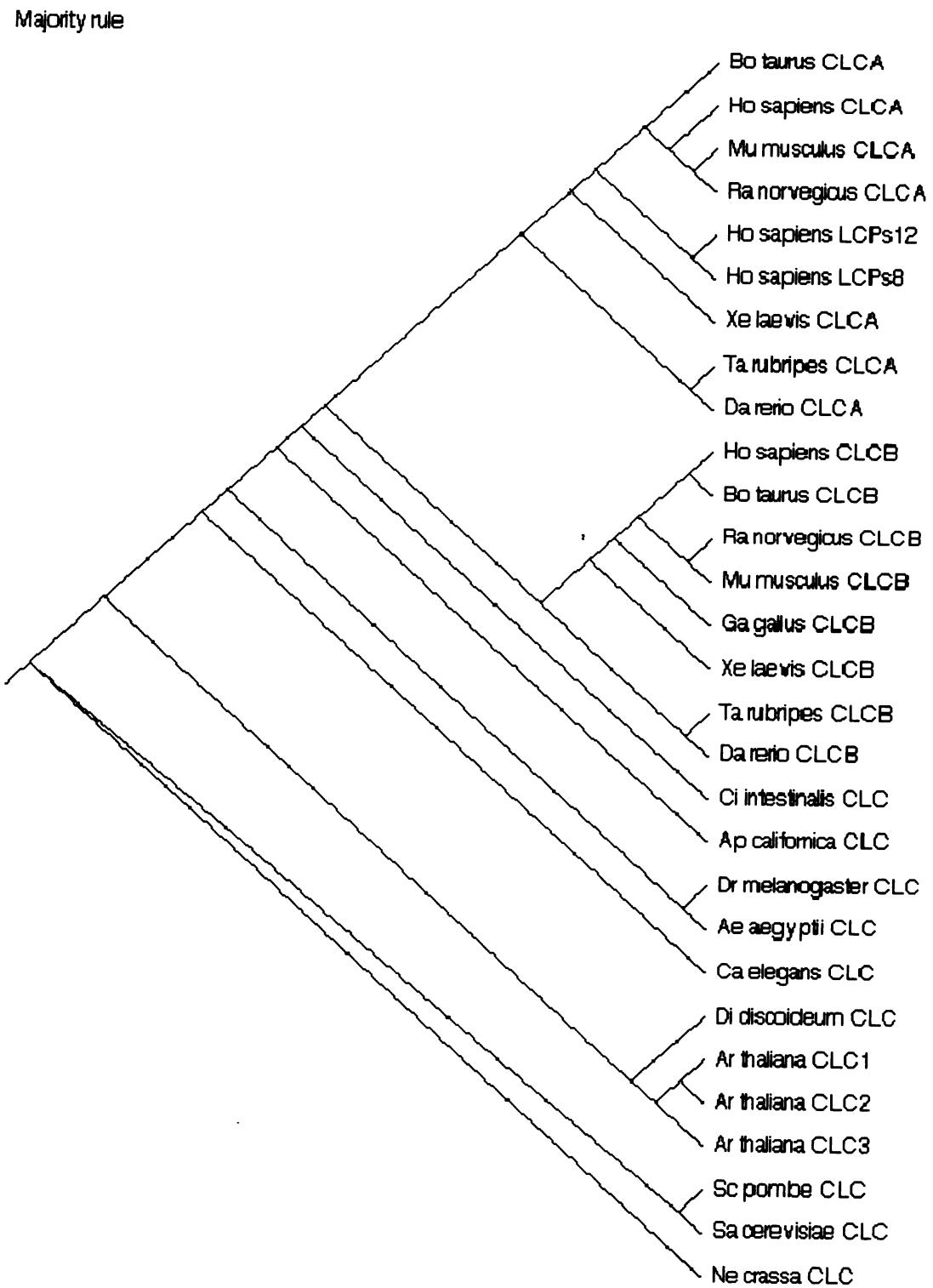


Figure 4.4: Bayesian Phylogenetic Tree of Clathrin Light Chains
A phylogenetic tree for clathrin light chain proteins and theorized sequences from translation of pseudogenes (ignoring stop codons and single nucleotide deletions) was constructed using MrBayes. Clades for LCa in mammals and fish, LCb in mammals and fish, and light chains in fungi, and plants are well defined. Full species names and accession codes for sequences are noted in Table 4.6.

an assumption requiring cautious interpretation. The resulting sequence strongly resembles both ciona and the CHC22 orthologue in pufferfish, the two closest sequences to the node. I anticipate that this sequence will be useful as a query in BLAST searches for locating the CHC22 orthologue in other species, such as zebrafish.

Likewise, we searched the databases for all light chain sequences and sequence fragments and located the sequence for species ranging from the fungi *neurospora* to the sea squirt (see Table 4.6 at end of chapter for accession codes). Fortunately, a few fungi and plant light chains, which are quite divergent in sequence from mammalian light chain, have been discovered and confirmed biochemically (Scheele and Holstein, 2002; Silveira et al., 1990). This facilitated identification of additional fungi, plant, and protist orthologues. Both LCa and LCb sequences are known for many vertebrates including fish, chicken, and rodents. An alignment of all light chain proteins was constructed (Table 4.4, at end of chapter) and a phylogenetic tree was generated (Figure 4.4).

The light chain phylogenetic tree indicates that light chains also duplicated after the radiation of the primitive chordate sea squirt (*Ciona intestinalis*) and before the radiation of the modern bony fishes zebrafish (*Danio rerio*) and pufferfish (*Takifugu rubripes*), between 750 to 420 million years ago. LCb appears to be somewhat closer in sequence to its ancestors than LCa. Both light chain pseudogenes LCps8 and LCps12 are located within the LCa clade in the tree, and appear to have been incorporated into the genome sometime after the radiation of the amphibian frog (*Xenopus laevis*) and before the radiation of the

rodents Norway rat (*Rattus norvegicus*) and Mouse (*Mus musculus*), relatively recently in evolutionary time. Thus, the pseudogenes are not candidates for CHC22-binding light chains that were lost over time due to mutations, but LCa mRNA insertions of into an advanced vertebrate genome.

The more rapid rate of mutation of the clathrin light chain, as opposed to the heavy chain, is evident in the lower homology between the LCa and LCb domains (Figure 4.1) and in the differences in sequence between the incorporated pseudogenes and the coding genes. The ancestral sequence at the nodes where LCb radiation occurred from a chordate ancestor and where LCa radiation occurred from an LCb-like ancestral sequence were also inferred using Bayesian algorithms (Table 4.4, bottom of table ANCESTRAL LCa and ANCESTRAL LCb). Again, the algorithm specification to infer a residue at every position results in assignment of residues even in positions where most sequences have a gap. The true ancestral sequence would likely lack most of the residues in the gaps and at the end of the protein. The resulting sequence draws heavily both from vertebrate LCb sequences and from the invertebrate sequences such as insect or roundworm clathrin light chains. The domain of residues near the N-terminus site conserved between mammalian LCa and LCb (28-50) is also absolutely conserved in the inferred ancestral LCb sequence. This inferred ancestral sequence should be a useful query tool to find additional light chain sequences.

Paralogue Analysis: A Map Based Approach

According to the 2R genome duplication theory, vertebrates could have up to 4 gene family members for every gene derived from an invertebrate ancestor, but many would have mutated and been lost (Spring, 2002). However, a recent study shows a statistically significant number of chromosomal regions are retained in modern day organisms which contain more genes with closely related sequences and functions than would be located there by random chance (McLysaght et al., 2002). Such segments were typically about 2 Mb long and are called paralogous chromosomal regions. Two genes with similar function within these domains are called paralogons. In order to determine whether clathrin duplications were the result of a large event such as a genomic duplication or a localized event, efforts are ongoing to locate additional paralogons between chromosomal regions 17q23 and 22q11 (Table 4.2), and likewise between chromosomal regions 5q35 and 9p13 (Table 4.1).

Potential paralogues were identified by examining the function, domains, and similar proteins in the database for each coding gene near the clathrin genes. All genes located near the clathrin genes in the current version of the human genome were recorded. For each, a literature search and sequence analysis yielded a list of protein domains and any potential functional role. Genes with similarities of domain or function to those on the potential paralogous region were identified as possible paralogons and investigated more deeply. A pairwise BLAST alignment was created between these sequences (data not shown) and any known phylogenetic relationship between the sequences was

Segment 5p35 Genes		
Gene	Name	Location
CLTA	Clathrin light chain LCa	36360431
RNF38	Ring finger 38 FLJ21343	36505839
MELK	Mammal embryonic leucine zipper kinase	36742315

Segment 9p13 Genes		
Gene	Name	Location
CLTB	Clathrin light chain LCb	176638272
KIAA1100	RING finger protein KIAA1100	176772487
GPRK6	G-protein coupled receptor kinase 6	177696412

Table 4.1: Potential Paralogous Genes Between Chromosomal Segments 5q35 and 9p13.

researched in the literature. If necessary, this analysis will be extended to align multiple family members and determine if the sequences are closely related on a phylogenetic tree generated de novo. In order to satisfy the statistical criteria for paralogous chromosomal regions put forth by Wolfe and colleagues (McLysaght et al., 2002), we would need to locate six paralogons within a ~2 MB sequence near the clathrin genes.

Chromosomes 5 and 9

For the chromosomal regions 5q35 and 9p13, five potential paralogon pairs were identified in addition to clathrin light chains. The first potential pair of paralogons was the RNA helicases ABS and VCP. Interestingly, VCP has been shown to interact with clathrin heavy chain (Pleasure et al., 1993), implicating any paralogon as a likely clathrin binding partner. Another pair of possible paralogons was LIM-motif containing proteins ENIGMA and TESK. A third potential pair of paralogons were Rab24 and Rab1b-like protein, which are interesting because Rabs have a critical function in membrane traffic (Alberts et al., 2002). However, a pairwise BLAST alignment showed no significant similarity with a BLAST E-value of 10 for these three pairs, eliminating them from consideration. In contrast, a typical BLAST hit for any closely related gene has an E-value (expected frequency of such a hit being the result of random chance) far below one. The Rabs in question are from completely different branches of the Rab tree (Pereira-Leal and Seabra, 2001), and their short length combined with diversity within the Rab family makes them divergent enough to show no significant similarity.

GPRK6, a GPCR kinase, and MELK, mammalian embryonic leucine zipper kinase, showed similarity over half their sequences and merited further investigation.

RING finger protein genes KIAA1100 and RNF38 near the light chain genes were evaluated as potential paralogons. These proteins were identified as possible paralogues in an analysis by McLysaught. While a phylogenetic tree

could not be found which included them because of their recent cloning (Eisenberg et al., 2002a; Kikuno et al., 1999), they are the top blast hit for one another and are likely paralogons. RNF38 is expected to be an E3 ubiquitin ligase with RING-H2 and KIL domains. A phylogenetic tree of these latter two potential paralogous gene pairs is in progress.

My preliminary analysis failed to generate a statistically significant number of paralogons on these chromosomal regions. The presence of one or two paralogons in a chromosomal region sequence of that size is a criteria met by 90% of the human genome (McLysaght et al., 2002). Loosening the stringency of these statistical criteria may yield further evidence of genome duplication and aid in the search for other possible light chains in the human genome. However, unless additional paralogue candidates can be identified, I must conclude that the light chains duplicated by a local duplication, possibly containing one or two adjacent genes, or that this small number of paralogous sequences were clustered together by random chance.

Chromosomes 17 and 22

Six pairs of proteins were identified as potential paralogons on chromosome segments 17q23 and 22q11 near the clathrin heavy chain. Two of these, PNUTL1 with PNUTL2 and TBX1 with TBX 2 and 4, had been previously identified in the automated analysis for paralogous chromosomal regions by McLysaght (McLysaght et al., 2002), and the remainder were located by mapping

Segment 17q23 Genes			Segment 22q11 Genes		
Gene	Name	Location	Gene	Name	Location
FLJ20315	Hypothetical RING finger	58912512	TUBAL2	Tubulin, alpha-2-like	15533812
MTMR4	Myotubularin related protein FYVE-DSP2	59050488	CLTCL1	Clathrin heavy chain CHC22	16107229
PNUTL2	Septin4	59079091	PNUTL1	Septin5	16642252
CLTC	Clathrin heavy chain CHC17	60146892	TBX1	T-box1 isoform B	16684453
TUBD1	Delta Tubulin	60387345	KIAA0371	MTMR3 FYVE-DSP1c	27170470
TBX2 TBX4	T-box2 T-box4	61963306 62019593	KIAA1133	KIAA1133 RING finger	26075986

Table 4.2: Potential Paralogous Genes Between Chromosomal Segments 17q23 and 22q11.

---

the loci. One pair of ubiquitin proteases, NY-REN-60 and USP18, failed to show any significant similarity in the Pairwise BLAST, eliminating it from consideration.

Delta tubulin on chromosome 17 and alpha-2-like tubulin on chromosome 22 were evaluated and found to have significant sequence similarity. However, a literature search of phylogenetic trees places alpha and delta tubulins in distant

clades (Jenkins et al., 2002). It remains to be seen if this alpha-like tubulin actually groups in the alpha clade in further studies.

Two hypothetical proteins, FLJ20315 on chromosome 17 and KIAA1133 on chromosome 22, also showed some similarity and were tested as possible paralogues. A partial alignment across two-thirds of the sequence was seen, with a reasonable E-value of 3×10^{-50}. Further investigation suggested these hypothetical proteins are RING finger proteins of unknown function. While KIAA1133 is more distant from CHC22 than most other pairs under consideration, this region of chromosome 22 appears to have made tandem duplications of itself in this pericentrosomic region. Therefore, I believe the broader region should also be considered and this pair merits further consideration.

T box transcriptional factor genes on chromosomes 17 and 22 were also analyzed. TBX2 and TBX4 on chromosome 22 lie adjacent, the likely result of a tandem duplication. Each shows partial alignment with TBX1 from chromosome 17 by pairwise BLAST with good homology. A literature search for phylogenetic trees located an analysis which places TBX2 and TBX4 in the same subclade, near a TBX1 clade (Agulnik et al., 1998). This is a promising potential paralogue.

Septin genes PNUTL1 (Septin 5) on chromosome 22 and PNUTL2 (Septin 4) on chromosome 17 were also investigated. A pairwise BLAST search showed good alignment. This was an exciting find because in addition to their well documented role in cytokinesis, septins may have a role in vesicle targeting and fusion, particularly during exocytosis (Kartmann and Roth, 2000), suggesting a

possible functional grouping with clathrin. However, despite their common categorization as Group IIIa septins in one literature report (Kartmann and Roth, 2000), a phylogenetic tree from another group places them in distant clades (Momany et al., 2001). Further investigation is merited.

Finally, myotubularin related proteins MTMR4 on chromosome 17 and KIAA0371 (MTMR3) on chromosome 22 were evaluated as possible paralogues. Both were found to be FYVE-DSP (dual specificity) phosphatases, acting on both phosphotyrosine and phosphoserines. Mutations in the related gene MTMR1 causes a severe congenital muscular disorder, X-linked myotubular myopathy. The protein sequences show good alignment in a pairwise BLAST search, and literature reports suggest that MTMR4 is more closely related to MTMR3 than other family members (Laporte et al., 1998). This is intriguing given the association of CHC22 with muscle function (Towler et al., 2002) and the muscular disorder VFCS/DGS (Holmes et al., 1997). However, this pair was also further beyond CHC22 than most genes considered in this analysis, and is only considered because of the likely tandem duplication along the labile 22q11 segment.

My preliminary analysis has located two likely paralogues in addition to clathrin, and three other possible paralogue pairs. Efforts are ongoing to identify additional possible paralogous pairs. If all six of these are confirmed as paralogues by further investigation, it would be statistically unlikely for these to be located here by random chance (McLysaght et al., 2002). Then we would conclude that the heavy chain gene probably duplicated as a result of a large-

scale duplication event such as a segmental or chromosomal duplication, in accordance with the genome duplication theory.

Discussion

The duplication of both the clathrin heavy chain and light chain genes occurred during the course of chordate evolution, between 750 to 420 million years ago, when chordate and early vertebrate ancestors were developing sophisticated muscular and neuronal systems. The mechanism of the duplication, however, may differ significantly between heavy chains and light chains.

Preliminary data suggest that the light chain duplication was the result of a localized gene duplication event, affecting only a few genes at most. A tandem copy of the light chain on the same chromosome most likely over evolutionary time was moved to a different chromosome where we find it today. The light chains evolved rapidly following their duplication, retaining the ability to bind the more ubiquitous CHC17 but seemingly not CHC22. Pseudogenes in the genome do not represent ancient CHC22-binding light chains, but merely genomic error in reincorporating processed mRNA transcripts into the DNA genome. The existence of alternatively spliced neuronal light chain variants suggests a probable primary function of light chain in regulating the increasingly rapid and complex endocytic events at the synapse. However, the functional difference between LCa and LCb remains elusive.

In contrast, the preliminary data suggests that the heavy chain duplication may have been the result of a large-scale duplication event such as the duplication of an entire genome, chromosome, or large chromosomal segment. Despite the unstable and dynamic nature of the genomic loci of the heavy chain genes, they remain highly conserved. This is surprising given the resiliency of the structure and assembly mechanism when mutation is deliberately introduced (Chapter Two). The selective pressures to resist mutation on clathrin heavy chains are probably the result of binding interactions with other components of the endocytic machinery. Whether the paralogous proteins near the clathrin genes represent a functional unit of proteins that may interact with clathrin is inconclusive, though the presence of myotubularins and septins is intriguing.

Clathrin duplication allowed divergence of a novel form of clathrin that may have a role in muscle function (CHC22). Surprisingly, CHC22 is evolutionarily closer to the ancestral invertebrate light chain, yet it is the form with less ubiquitous function and it has lost its ability to bind light chains in-vitro. However, a recent yeast-two-hybrid screen in our laboratory for proteins interacting with CHC22 yielded the more ancestral LCb but not LCa as a weakly interacting protein. In muscle tissue, where CHC22 is most strongly expressed, the relative expression levels for the light chains LCa and LCb have not been defined. It is possible that the expression of LCb is low in muscle tissue, allowing other proteins to out-compete LCb for binding to CHC22. This may have allowed CHC22 to evolve in the absence of bound light chain and thus weakened the interaction with ancestral light chain over evolutionary time. Thus clathrin CHC17

has co-evolved with both light chains, but may have particularly co-evolved with LCa, to allow an expanded neuronal role as the CNS developed in early chordates.

Further analysis of the phylogenetic and evolutionary relationship between clathrins may lead to a deeper understanding of the functional differences between the two vertebrate clathrin heavy chains and two vertebrate clathrin light chains.

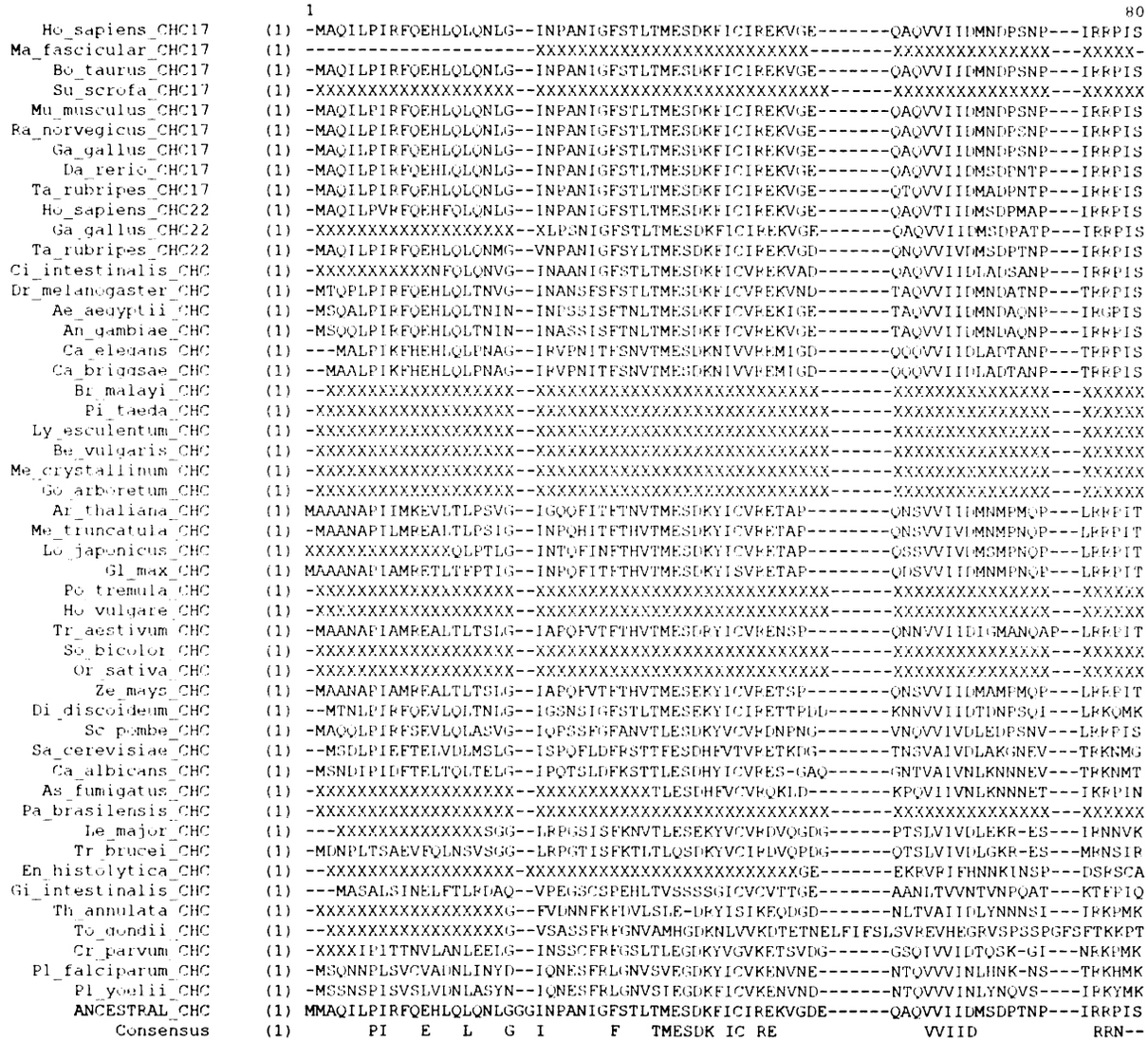


Table 4.3: Alignment of Clathrin Heavy Chain Sequences and Sequence Fragments

		81 160
Ho_sapiens_CHC17	(68)	-ADSAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGE--
Ma_fascicular_CHC17	(47)	-XXXXXXXXXXX-XXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Bo_taurus_CHC17	(68)	-ADSAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGE--
Su_scrofa_CHC17	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Mu_musculus_CHC17	(68)	-ADSAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGE--
Ra_norvegicus_CHC17	(68)	-ADSAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGE--
Ga_gallus_CHC17	(68)	-ADSAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGE--
Da_rerio_CHC17	(68)	-ADSAIMNPASK-VIALK-GK------TLQIFNIEMKSKMKAHTMTEDV--SFWKWISLNTIALVTDNAVYHWSMEGD--
Ta_rubripes_CHC17	(68)	-ADSAIMNPASK-VIALKAAK------TLQIFNIEMKSKMKAHTMTDDV--TFWKWISLNTVALVTDNAVYHWSMEGD--
Ho_sapiens_CHC22	(68)	-AESAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMAEEV--IFWKWVSVNTVALVTETAVYHWSMEGD--
Ga_gallus_CHC22	(68)	-AESAIMNPASK-VIALKAGK------TLQIFNIEMKSKMKAHTMAEEV--IFWKWISVNTVALVTETAVYHWSMERE--
Ta_rubripes_CHC22	(68)	-ADSAIMNPASK-VIALKAAK------TLQIFNIEMKSKLKAHTMTEDV--MFWKWISVNTVALVTDSAVYHWSMEGD--
Ci_intestinalis_CHC	(68)	-AESAIMNPASK-VIALKAGR------TLQIFNMEMKSKLKAHTMTEDV--VFWKWISTNMVGLVTETSVYHWSMEGNS-
Dr_melanogaster_CHC	(68)	-ADSAIMNPASK-VIALKAQK------TLQIFNIEMKSKMKAHTMNEDV--VFWKWISLNTLALVTETSVFHWSMEGD--
Ae_aegyptii_CHC	(68)	-ADSAIMNPASK-VIALKAQK------TLQIFNIEMKSKMKAHAMTEEV--VFWKWITLNTLSLVTETSVYHWSMEGD--
An_gambiae_CHC	(68)	-ADSAIMNPASK-VIALKAQK------TLQIFNIEMKSKMKAHTMTEEV--VFWKWITLNTLSLVTETSVYHWSMEGD--
Ca_elegans_CHC	(66)	-ADSVIMHPTAK-IILALKSGK------TLQIFNIELKAKVKAHQNVEDV--VYWKWISEKTIALVSDTAVYHWSIEGD--
Ca_briggsae_CHC	(67)	-ADSVIMHPTAK-IILALKSGK------TLQIFNIELKAKVKAHQNVEDV--VYWKWISDKLIALVSDTAVYHWSIEGD--
Br_malayi_CHC	(67)	-XXXXXXXXXXX-XXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Pi_taeda_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Ly_esculentum_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Be_vulgaris_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Me_crystallinum_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Go_arboretum_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Ar_thaliana_Chc	(68)	-ADSALMNPNSR-IILALKAQVPGTTQDHLQIFNIEAKAKLKSHQMPEQV--AFWKWITPKMLGLVTQTSVYHWSIEGD--
Me_truncatula_CHC	(67)	-ADSALMNPNSR-IILALKAQLQGTTQDHLQIFNIELKAKMKSYQMPEQV--VFWKWIXPXLXXXXXXXXXXXXXXXXX--
Lo_japonicus_CHC	(68)	-ADSALMNPNSR-IILALKXXX------XLQIFNIELKAKMKSYQMPEQV--VFWKWISPKTLGLVTLTSVYHWSIEGD--
Gl_max_CHC	(68)	-ADSALMNPNSP-SILALKTQLQGTTQDHLQIFNIEFKSKMKSYQIPEQV--SFWKWITPKLLGIVTQTSVYHWSIEGD--
Po_tremula_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Ho_vulgare_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Tr_aestivum_CHC	(68)	-ADSALMNPNTR-IILALKAQIAGTTQDHLQIFNIEAKTKVKSHQMPEQV--VFWKWITPKLLGLVTQTSVYHWSIEGD--
So_bicolor_CHC	(69)	-XXXXXXXXXXX-XXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Or_sativa_CHC	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Ze_mays_CHC	(67)	-ADSALMNPNAR-IILALKAQIPGTTQDHLQIFNIEAKTKIKSHQMPEQV--VFWKWITPKLLGLVTQTSVYHWSIEGD--
Di_discoideum_CHC	(68)	-TDAAIMNPKEP-IILALKIGQ------VLQLISIEQKMQLKSCQMQEPL--EFWKWISPNTLALVTATSVHHWTKEGN--
Sc_pombe_CHC	(68)	-ADSVILHPKKK-IIALKAQR------QLQVFDLEAKAKINSYVMNQDV--VYWTWISDSVIGMVTDTSVHHWTVSG---
Sa_cerevisiae_CHC	(67)	-GDSAIMHPSQM-VISVRANG-----TIVQIFNLETKSKLKSFTLDEPV--IFWRWLSETTLGFVTARSILTSNVFDGN-
Ca_albicans_CHC	(68)	-ADNAIMHPKEF-VISLRAGT------TLQIFNLGSKQRLKAFTMDQPV--IFWKWLDDTYLGLVTQSSIYYWNVFDG--
As_fumigatus_CHC	(67)	-ADSAIMHWSRN-IIALRAGR------TIQIFDLSAKQKLKSAVMNEDV--VYWKWFSETSLGLVTDASVYHWDVFDP--
Pa_brasilensis_chc	(69)	XXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX--
Le_major_CHC	(66)	DAESCIMNPKSK-IILALRSGR------NLQVFDVDASRRLKATLFHEDV--AYWRWIDDRTLGIVTAKAVYHWSLDTAT-
Tr_brucei_CHC	(68)	DAESAIMNPMAK-IILALRSGR------NLQIFDVDAANRLKVVVFNEDV--VYWCWIDARTVGIVSNTAVYHWSLDTAAD
En_histolytica_CHC	(68)	-ADFAIMHPTKQ-IMAVAAGT------TVQMFDVQSKSKVADITLQTPL--VFWKWANEHILAIVTSTSVFHFDLNSP--
Gi_intestinalis_CHC	(66)	-AQHAIQHRKNPNIVVLRTPV------LLQTLDVSTMTPYGRFDFPNPIGVDYWGWATDEKLALVCPDAVYFWNCMTG--
Th_annulata_CHC	(67)	-AEAAIMNPNKP-IIALRAXX------XXXVFNLENKEKLGYYQFDQRI--IYWKWLNNMELVIITETLVYHWMIGSN--
To_gondii_CHC	(77)	QAEAAIMHPSEK-VVALRAKTEGSAGHMIQVLNLETKVRLGTAQMNEPV--VYWKWVAPTLLALVTDRAVYHWTVGENGE
Cr_parvum_CHC	(68)	-AESALIHPIEN-ILVVRGPYEDN-GCTVQIFNLDSKEKLGAFLFPESV--VFWPWLTPRILAIVGDKGIYHWTIESGN-
Pl_falciparum_CHC	(67)	-AESVIIHPNDP-IILALKGTIKNMNTIFLQVFNIETKEKICSLNLNEYM--NYWKWINNDTIAIVCEKNVYHWNIDIHNT
Pl_yoelii_CHC	(67)	-ADSVIIHPNDP-IILALRGSIKNANTIFLQVFNIETKEKICSLNLNEHI--NYWKWINNDTIAIVCEKNVYHWNIDMYNT
ANCESTRAL_CHC	(72)	DADSAIMNPASKNVIALKAGKPGTTQDTLQIFNIEMKSKMKAHTMTEDVNVMFWKWISVNTVALVTDTAVYHWSMEGDSQ
Consensus	(81)	ADSAIMNP ALKA LQIFNIE K K K M E V FWKWI LVT VYHW IDG

Table 4.3 continued

```
                                 161                                                                            240
        Ho_sapiens_CHC17   (136) --------------------------SQPVKMFDRHSSLAG----CQIINYRTDAKQKWLLLTGISA--QQ----NRVVG
   Ma_fascicular_CHC17   (115) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXX--XX----XXXXX
        Bo_taurus_CHC17   (136) --------------------------SQPVKMFDRHSSLAG----CQIINYRTDAKQKWLLLTGISA--QQ----NRVVG
        Su_scrofa_CHC17   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
      Mu_musculus_CHC17   (136) --------------------------SQPVKMFDRHSSLAG----CQIINYRTDAKQKWLLLTGISA--QQ----NRVVG
   Ra_norvegicus_CHC17   (136) --------------------------SQPVKMFDRHSSLAG----CQIINYRTDAKQKWLLLTGISA--QQ----NRVVG
        Ga_gallus_CHC17   (136) --------------------------SQPVKMFDRHSSLAG----CQIINYRTDAKQKWLLLTGISA--QQ----NRVVG
         Da_rerio_CHC17   (135) --------------------------SQPIKVFDRXXXXXX----XXXXXXRTDAKQKWLLLIGISAQ-VQ----NRVVG
     Ta_rubripes_CHC17   (136) --------------------------SQPVKVFDRHSSLAG----CQIINYRTDAKQKWLLLIGISAQVQQ----NRVVG
      Ho_sapiens_CHC22   (136) --------------------------SQPMKMFDRHTSLVG----CQVIHYRTDEYQKWLLLVGISA--QQ----NRVVG
        Ga_gallus_CHC22   (136) --------------------------SQPQKXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXX--XX----XXXXX
     Ta_rubripes_CHC22   (136) --------------------------FHPAKVFDRHASFAG----CQIINYRTDQQQKWLLLIGISA--QQ----NRVVG
   Ci_intestinalis_CHC   (137) --------------------------QQPVKMFDKHPSLAK----CQIINYRTDAKQKWLLLIGISA--QA----NRVVG
   Dr_melanogaster_CHC   (136) --------------------------SMPQKMFDRHSSLNG----CQIINYRCNASQQWLLLVGISA--LP----SRVAG
        Ae_aegyptii_CHC   (136) --------------------------STPVKMFERHSSLNG----CQIINYRTDPKQAWLLLVGISA--QQ----NRVIG
        An_gambiae_CHC   (136) --------------------------STPIKMFERHSSLNG----CQIINYRTDPKQAWLLLVGISA--QQ----NRVIG
        Ca_elegans_CHC   (134) --------------------------AAPVKMFDRHQSLAG----TQIINYRADAENKWLVLIGISA--KD----SRVVG
       Ca_briggsae_CHC   (135) --------------------------AAPVKMFDRHQSLAG----TQIINYRADAESKWLVLIGISAK-D-----SRVVG
         Br_malayi_CHC   (135) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXX--XX----XXXXX
          Pi_taeda_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
    Ly_esculentum_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
        Be_vulgaris_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
  Me_crystallinum_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
      Go_arboretum_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
       Ar_thaliana_Chc   (142) --------------------------SEPVKMFDPTANLAN----NQIINYKCSPNEKWLVLIGIAPGSPERP--QLVKG
     Me_truncatula_CHC   (141) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
      Lo_japonicus_CHC   (136) --------------------------SEPIKMFERTANLAN----NQIINYRSDPTEKWLVLIGIAX--XX----XXVKG
            Gl_max_CHC   (142) --------------------------SEPVKMFERTANLPN----NQIINYRCDPTEKWLVLIGILHGSPERP--QLVKG
        Po_tremula_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
        Ho_vulgare_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXX
       Tr_aestivum_CHC   (142) --------------------------SEPTKMFDKTANLGN----NQIINYRCDPAXKWLXXXXXXXXXXX----XXXXX
        So_bicolor_CHC   (137) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXX--XX----XXXXX
         Or_sativa_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXXXXXXXX--XXXXX
           Ze_mays_CHC   (141) --------------------------SEPTKMFDPTANLAN----NQIINYCRDPAEKWLVLIGIAPGAPERP--QLVKG
    Di_discoideum_CHC   (136) --------------------------SDPVKVFDRHPDLQN----TEIINYRSDSTQNWLVLVAIHQ--RD----GRVVG
          Sc_pombe_CHC   (135) --------------------------SDPVKMFDRHSSLNG----TQIISYKSNYNEEWFTLIGISS--RD----NRIAG
     Sa_cerevisiae_CHC   (137) ------------------------VNAKPQLLTLRHANLNN----TQIINFVANKNLDWFAVVGILQ--EN----GRIAG
       Ca_albicans_CHC   (136) ------------------------TNNGPTKLTDRHHTLNN----CQIINFVAEPDLNWFAVTGIAQE-D-----GRIAG
      As_fumigatus_CHC   (135) ------------------------TQNQPLKIFDRLPNLSVS---CQIINYRVNDDEKWMVVVGISQ--QQ----GRVVG
   Pa_brasilensis_CHC   (139) --------------------------XXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXXX--XX----XXXXX
          Le_major_CHC   (136) --------------------------DAPQHIFDRSADYDSS---VQILSYRSDEQKKWLVVTGVARD-PS----GTMVG
         Tr_brucei_CHC   (139) --------------------------APPQLVFSPAPEFDAS---EQVLSYQTDEQKKWLMLCGVMR--TA----EGMVG
    En_histolytica_CHC   (136) --------------------------SAPKKIVDRHSELAAS----QIIGYKIDPTEKWVAILALLQ--KN----NEIVG
   Gi_intestinalis_CHC   (137) ---------------------------EFSPGFARDVSLTSTYSNSRITDLAVSSDADWVAITALGT--DQA---NHTQG
       Th_annulata_CHC   (135) ----------------------------PKYMFELTGKLLDSN--TKIVGYSSDITNKWCLVFGIYSN-DQG---VSIDG
         To_gondii_CHC   (154) ------------------------TTAEPEKVCSREGRLAD---AVQIISYAVDKDLKWCILTGIXXXXXXXXXXXXXXX
         Cr_parvum_CHC   (142) ------------------------SNSIPVRIFERAGKLAEQS--TQIVGYQTDSGQRWCMLMGLCPVTNESTGSISVKG
     Pl_falciparum_CHC   (143) KKNKDNNDNNNNNNNNNNNNNNNNYFNTLTKVFEKAQIFIDNN--SQILYYSTDKDMKWCILCGIST--DQG---KSIDG
          Pl_yoelii_CHC   (143) KKNK--------------------ENNVLTKVFEKAQIFIDNN--SQILYYGTDKEMKWCILCGISTQ-DQG---KSIDG
       ANCESTRAL_CHC   (152) NKNKDNNDNNNNNNNNNNNNNNNNYFSQPAKMFDRHASLAGNNSNCQIINYRTDQQQKWLLLIGISAQQQQRP--NRVVG
             Consensus   (161)                             P K F R   L       QII Y  D   KW  L GI                 G
```

Table 4.3 continued

```
                              241                                                                              320
      Ho_sapiens_CHC17  (180) AMQLYSVDRKVSQPIEGHAASFAQFKMEGN-AEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
   Ma_fascicular_CHC17  (157) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXX--------------
       Bo_taurus_CHC17  (180) AMQLYSVDRKVSQPIEGHAASFAQFKMEGN-AEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
       Su_scrofa_CHC17  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
     Mu_musculus_CHC17  (180) AMQLYSVDRKVSQPIEGHAASFAQFKMEGN-AEESTLFCFAVRGQAG----GKVHIIEVGTPPTGN--------------
  Ra_norvegicus_CHC17   (180) AMQLYSVDRKVSQPIEGHAASFAQFKMEGN-AEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
       Ga_gallus_CHC17  (180) AMQLYSVDRKVSQPIEGHAASFAQFKMEGN-AEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
        Da_rerio_CHC17  (180) AMQLYSVDRKVSQPIEGHAAGFAQFKMEGN-AEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
     Ta_rubripes_CHC17  (182) AMQLYSVDRKVSQPIEGHAAGFAQFKMEGN-SEESTLFCFAVRGQAG----GKLHIIEVGTPPTGN--------------
      Ho_sapiens_CHC22  (180) AMQLYSVDRKVSQPIEGHAAAFAEFKMEGN-AKPATLFCFAVRNPTG----GKLHIIEVGQPAAGN--------------
       Ga_gallus_CHC22  (178) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXX--------------
     Ta_rubripes_CHC22  (180) AMQLYSVDRKVSQPIEGHAAAFGDFKVEGN-TKPSTLFCFAVRSQAG----GKLHIIEVGQPAAGN--------------
   Ci_intestinalis_CHC  (181) AMQLYSVERRVSQPIEGHAASFAQIKLEGN-XXXSTLFCFAVRGPQG----GKLHIIEVGSPAGN---------------
   Dr_melanogaster_CHC  (180) AMQLYSVERKVSQAIEGHAASFATFKIDAN-KEPTTLFCFAVRTATG----GKLHIIEVGAPPNGN--------------
       Ae_aegyptii_CHC  (180) AMQLYSVERKVSQAIEGHAASFATFKMEEN-KEPSTLFCFAVRSANA----AKLHIIEVGAPPTGN--------------
       An_gambiae_CHC   (180) AMQLYSVERKVSQAIEGHAASFATFKMEEN-KELSTLFCFAVRSQTA----AKLHIIEVGTPPAGN--------------
        Ca_elegans_CHC  (178) SMQLYSTERKVSQPIEGHAACFVRFKVDGN-QNPSNLFCFSVKTDNG----GKLHVIEVGTPAAGN--------------
       Ca_briggsae_CHC  (179) SMQLYSTERKVSQPIEGHAASFVRFKVDGN-QQPSNLFCFSVKTETG----GKLHIIEVGTPATGN--------------
         Br_malayi_CHC  (179) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXX----XQLHVIEIGSPPAGN--------------
          Pi_taeda_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
     Ly_esculentum_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
       Be_vulgaris_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
  Me_crystallinum_CHC   (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
      Go_arboretum_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
       Ar_thaliana_CHC  (190) NMQLFSVDQQRSQALEAHAASFAQFKVPGN-ENPSILISFASKSFNAG---QITSKLHVIELGAQPGK------------
    Me_truncatula_CHC   (187) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
      Lo_japonicus_CHC  (180) NMQLFSVDQQRSQALEAHAASFAQYKVRSLMSYDSVFIFVMXXXXXX-----XXXXXXXXXXXAGK--------------
            Gl_max_CHC  (190) RMQLFSVEQQRSQALEAHAASFAQFKVPGN-ENPSTLFLFATNPNAG----QIITKLHVIELGAQPG-------------
        Po_tremula_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
        Ho_vulgare_CHC  (185) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
       Tr_aestivum_CHC  (188) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
        So_bicolor_CHC  (181) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXX--------------
         Or_sativa_CHC  (187) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXX--------------
           Ze_mays_CHC  (189) NMQLFSVDQQRSQALEAHAASFATFKVVGN-ENPSTLICFASKTTNAG---QITSKLHVIELGAQPGK------------
     Di_discoideum_CHC  (180) RIQLYSVEKQISQSIEGHAACFANYIVPGA-TRPSTLFAISSRTQNA----SKILVLEVSKGDG---------------
          Sc_pombe_CHC  (179) NLQLYSKKRKVSQPLESHASAFAVIQPEGV-DHEVQVLALASRLPTG----SKLSIVEVDRNPN-N--------------
    Sa_cerevisiae_CHC   (183) RIQLFSKQRNISQAIDGHVAIFTNILLEGNGSTPVQVFVTGNRNATTG--AGELRIIEIDHDASLP-------------
       Ca_albicans_CHC  (182) HIQLYSKSRNVSQAIEGHVCKFASISLSGG-VQPTKVFCVGNKNAQGQ---GNMHIIEIDHVDG-N-------------
      As_fumigatus_CHC  (182) SMQLYSKDRGISQFIEGHAAAFASIRVEGS-PLEHKLFTFAVRTQTG----AKLQIAEIDHQEPN--------------
   Pa_brasilensis_CHC   (181) XXXXXXXXNSLVIVVDGHAGTFISTNTPTE-PRACNIMCLAWNTPHEG---GKVQLMELPTSPKMD-------------
          Le_major_CHC  (182) KALLYSVENRSSRVLDGHACCFISTSTPSE-PRKCNVMCLAWNSPQQG---GQVMIMELPTGSKTD-------------
         Tr_brucei_CHC  (184) KTQLFSVENNSGRVLDGHSGTFISTNTPTD-PRSCNLMCLAWNNPSTG---GKILIMELPTGSKTD-------------
   En_histolytica_CHC   (180) KIQLFSIEKNASQIIDGYAASFYNYKFDGA-SQGISYVCLCTKXXXXX---XXXXXXXXXXXXXXX-------------
  Gi_intestinalis_CHC   (185) MIQLYCFSRAYTRMLSGQCVTFVDWQLDAT-RPKTTLLAYAHKPTGSQ--EITINIMEPAPQVGADGASI----------
       Th_annulata_CHC  (181) VIQLYSVDKRQQQLFEGYAGTFTQLRTNNQTLKKSNLLIFCEHKKNTHN--IKLHLMDIGSLGQSN----TSNGVSSGVG
          To_gondii_CHC (207) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXX
         Cr_parvum_CHC  (196) QLQLFSVEKRQQQLIEGFSGNFGELIVDDFVHSPASIVCFVEKKQEHSN--ARLHVMDISGQRAGSGE-----FGSSLSP
    Pl_falciparum_CHC   (216) YMQLYSCEKKFDQTIEGFIGCFGSFVFDNL--DMKPLFCFVEKKKNS--TISRLHLMDIYTNKTET-------------
          Pl_yoelii_CHC (197) HMQLYSCEKKLHQIIEGFIGCFGSFIFENW--DTKPLFCFIEKKKNSS--MSRLHLMDIYNSKTEG-------------
       ANCESTRAL_CHC    (230) AMQLYSVDRKVSQPIEGHAAAFAEFKMEGNMAKPSTLFCFAVRSQAGG---QITGKLHIIEVGQPAAGNRES--------
          Consensus     (241)   QLYS      SQ  EGHA  F                L C                E
Table 4.3 continued
```

```
                              321                                                                            400
     Ho_sapiens_CHC17   (241) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQISEKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
   Ma_fascicular_CHC17  (218) -----XXXXXXXXXXXXXXXXX-------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
      Bo_taurus_CHC17   (241) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQISEKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
      Su_scrofa_CHC17   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-
    Mu_musculus_CHC17   (241) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQVSEKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
  Ra_norvegicus_CHC17   (241) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQISEKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
      Ga_gallus_CHC17   (241) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQISDKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
       Da_rerio_CHC17   (241) -----QAFPKKAVDVWFPPEAQ-------NDFPVALQISSKHDVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
    Ta_rubripes_CHC17   (243) -----QPFPKKAVDVFFPPEAQ-------NDFPVAMQVSAKHNVVFLITKYGYIHLYDLETGTCIYM-NRISGETIFVT-
     Ho_sapiens_CHC22   (241) -----QPFVKKAVDVFFPPEAQ-------NDFPVAMQIGAKHGVIYLITKYGYLHLYDLESGVCICM-NRISADTIFVT-
      Ga_gallus_CHC22   (239) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
    Ta_rubripes_CHC22   (241) -----QPFAKKAVDVFFPPEAQ-------TDFPVAMQIGSKHGVIYLITKYGYIHLYDLESGVCIYM-NRISAETIFVT-
  Ci_intestinalis_CHC   (241) -----QPFTKKAVDVFFPPEAA-------NDFPVAMKVSPNQNVVFVVTKYGYIHMYDLETAVCIFM-NRISADTIFVT-
  Dr_melanogaster_CHC   (241) -----QPFAKKAVDVFFPPEAQ-------NDFPVAMQVSAKYDTIYLITKYGYIHLYDMETATCIYM-NRISADTIFVT-
      Ae_aegyptii_CHC   (241) -----TAFTKKAVDVFFPPEAQ-------SDFPVAMQVSPKYDVIYLITKYGYIHMYDIETGTCITM-NRISADTIFVT-
      An_gambiae_CHC    (241) -----VAFTKKAVDVFFPPEAQ-------SDFPVAMQVSPRYDVIYLITKYGYIHMYDIETATCIYM-NRISGDTIFVT-
      Ca_elegans_CHC    (239) -----TPFQKKNVDVPYTADTA-------GDFPVSMQVSAKQGIIYLVTKQGYVHLYDVESGTRIYS-NRISTDTVFVT-
      Ca_briggsae_CHC   (240) -----SPFQKKNVDVPYTADTA-------GDFPVSMQVSTKQGIIYLVTKQGYVHLYDVESGLRIYS-NRISTDTVFVT-
       Br_malayi_CHC    (240) -----QPFQKKQVEVYYPAEAA-------TDFPVAMQVSSKHGIIYLVTKYGYVHLYDIETAVCIYM-NRISSETIFVT-
        Pi_taeda_CHC    (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
    Ly_esculentum_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
      Be_vulgaris_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
  Me_crystallinum_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-
     Go_arboretum_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
      Ar_thaliana_CHC   (254) -----PSFTKKQADLFFPPDFA-------DDFPVAMQVSHKFNLIYVITKLGLLFVYDLETASAIYR-NRISPDPIFLT-
    Me_truncatula_CHC   (249) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
     Lo_japonicus_CHC   (241) -----PSFTKKQADLFFPPDFA-------DDFPVAMQISHKYSLIYVITKLGLLFVYDLETATAVYR-NRISPDPIFLT-
          Gl_max_CHC    (252) -----NHHLPRNKQIFSSPQIFS------DDFPVAMQISHKYNLIYVITKLGLLFVYDLETATAVYR-NRISPDPIFLT-
       Po_tremula_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
       Ho_vulgare_CHC   (248) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-
      Tr_aestivum_CHC   (250) --XXXXXXXXXXXXXXXXXXXFQ-------DDFPVAMQISQKYGLVYVITKLGLLFVYDLETAAAVYR-NRISPDPIFLT-
        So_bicolor_CHC  (242) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXX-
         Or_sativa_CHC  (250) -----XXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-
          Ze_mays_CHC   (253) -----PGFSKKQADLFFPPDFQ-------DDFPVAMQVSQKYGLIYVITKLGLLFVYDLETAAAVYR-NRISPDPIFLT-
    Di_discoideum_CHC   (239) -----PNFQKRASDVFYPPEIGA------SDFPVAMQVSEKYEVIYMVTKLGYIHLFDLGTANLIYR-NRISNENIFVT-
         Sc_pombe_CHC   (239) -----PAFATKTVDLFFPPEAV-------NDFPIAIEIGSTYNVAYVVTKYGFIHVYDLETAKCIYM-NRVSGESIFVT-
    Sa_cerevisiae_CHC   (247) -----SQYQKETTDIFFPPDAT-------NDFPIAVQVSEKYGIIYLLTKYGFIHLYELETGTNLFV-NRITAESVFTA-
       Ca_albicans_CHC  (243) -----PQFQKKVVDIFFPPDAS-------NDFPISLQASNKYGIVYVLTKYGFIHLYDMETGANLFV-NRITADPVFTA-
      As_fumigatus_CHC  (242) -----PRFQKKAVEVYFPQEAV-------NDFPVAMQVSRKYDIVYLVTKYGFIHLYDLETGTCIFM-NRISSETIFTT-
   Pa_brasilensis_CHC   (243) -----ITMPRRVVDVAFFP----------GDFPVALTVSPRHKLLTIMTSRGTAILMDIFTGTIIQS-QQVTSNIVFCG-
         Le_major_CHC   (244) -----LSVLRPVYSVHIQP----------GDFPVAMNVSDRHKLLTVATSRGSFVLMDIFTGTILVE-QQFSSSVVFCG-
        Tr_brucei_CHC   (246) -----LTVQRPMIDVPFA---Q-------GDFPVATHVSPRHKLLTVVTSRGSVVLMDLFTGVVIKT-HQLPNNTIFCG-
   En_histolytica_CHC   (242) -----XXXXXXXXXXXXPQQQQ------ADFMVGLVVSVKYETILLVSRMGYIYGFYSGDGTLFFQ-GRVIDDIVFLT-
  Gi_intestinalis_CHC   (252) -----SGFTTKAVQATVSAALVG------VDHPLFLRGKEKYGFLYMFSKQGNVTLVDIYSGLLIYQ-GPISGEMVFAIG
      Th_annulata_CHC   (255) TSVGTTGTTTGGTDILKINCIMERNEEFPNDFPISIHVFDSNGIILILTKNGFAHFYFSNTLTHLFTXXXXXXXXXXX-
        To_gondii_CHC   (286) XXXXXXXXXXXXXXXXXXXXXAP-------TDFPLSIHISQKFGVVYIVTKGGYLILLDALTGXXXXXXHRILQDAVFLA-
        Cr_parvum_CHC   (269) PTPFKAVIELPTVDEGSAGGAHS------FDFPIYTYVSSYFGVIFIITRGGILYIVEPTSNTLLYC-NKVCQDSVFLG-
    Pl_falciparum_CHC   (278) -----TIPYKIVKEINLINENL-------NDFPIYISLNTLQGVIYIVTKCSYVYVFDEGTLSLIVK-EKISEDNIFIC-
        Pl_yoelii_CHC   (259) -----STPYKIVKEINLINDTL-------NDFPIYISLNTLQGVIYVITKCSYVYIFDESTLTEIVR-EKISDDNIFIC-
      ANCESTRAL_CHC     (299) ----IQPFAKKAVDVFFPPEAGQ------NDFPVAMQIGAKHGVIYLITKYGYIHLYDLESGVCIYM-NRISAETIFVTG
          Consensus     (321)        F K   D    P             DFPVA Q S K        TK G     YD ET    I    NRIS    IF T
Table 4.3 continued
```

```
                                 401                                                                          480
     Ho_sapiens_CHC17    (307) -----APHEATAGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRMAVRNNLAGAEELFARK-------FNA
   Ma_fascicular_CHC17   (284) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
      Bo_taurus_CHC17    (307) -----APHEATAGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRMAVRNNLAGAEELFARK-------FNA
      Su_scrofa_CHC17    (316) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
    Mu_musculus_CHC17    (307) -----APHEATAGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRMAVRNNLAGAEELFARK-------FNA
   Ra_norvegicus_CHC17   (307) -----APHEATAGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRMAVRNNLAGAEELFARK-------FNA
      Ga_gallus_CHC17    (307) -----AQHEATAGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRMAVRNNLAGAEELFARK-------FNA
       Da_rerio_CHC17    (307) -----APHEATSGIIGVNRKGQVLSVRVEEENIIPYITNVLQNPD----LALRMAVRINLAGAEELFARK-------FNN
    Ta_rubripes_CHC17    (309) -----APHEATSGIIGVNRKGQVLSVCVEEENIIPYITNVLQNPD----LALRLAVRNNLAGAEELFARK-------FNN
     Ho_sapiens_CHC22    (307) -----APHKPTSGIIGVNKKGQVLSVCVEEDNIVNYATNVLQNPD----LGLRLAVRSNLAGAEKLFVRK-------FNT
      Ga_gallus_CHC22    (306) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
    Ta_rubripes_CHC22    (307) -----APHEATSGIIGVNKKGQVLSVCVEEENIVNYATNVLQNPD----LALRIALRSNLAGAEEIFARK-------FNT
   Ci_intestinalis_CHC   (307) -----APHEATVGLIGVNRKGQVLSVSVDEENVVPYITNTLQNPD----LALRFAVRNNLGGADELFSRK-------FNT
   Dr_melanogaster_CHC   (307) -----APHEASGGIIGVNRKGQVLSVTVDEEQIIPYINTVLQNPD----LALRMAVRNNLAGAEDLFVRK-------FNK
      Ae_aegyptii_CHC    (307) -----APHESSGGIIGVNRKQQVLSVTVDEEQIIPYINTVLQNPD----LALRMAVRNNLSGAEDLFVRK-------FNQ
       An_gambiae_CHC    (307) -----APHESSGGIIGVNRKGQVLSVTVDEEQIIPYINTVLQNPD----LALRMAVRNNLSGAEDLFVRK-------FNQ
       Ca_elegans_CHC    (305) -----CEYTATGGIMGINRKGQVLSVSIDEANLVPFVTNQLQNPD----LALKLAVRCDLPGAEELFVRK-------FNL
      Ca_briggsae_CHC    (306) -----CEYTATGGIMGINRKGQVLSVSIDEANLVPFVTNQLQNPD----LALKLAVRCDLPGAEELFVRK-------FNL
        Br_malayi_CHC    (306) -----AEYTATEGIIGVNRKGQVLSVSIDEQNMIPYVTQTLQNPE----LALKLAVRCDLPGAEELFVRK-------FNL
         Pi_taeda_CHC    (315) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
     Ly_esculentum_CHC   (315) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
       Be_vulgaris_CHC   (315) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
   Me_crystallinum_CHC   (316) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
      Go_arboretum_CHC   (315) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
      Ar_thaliana_CHC    (320) -----SEASSVGGFYAINRRGQVLLATVNEATIIPFISGQLNNLE----LAVNLAKRGNLPGAENLVVQR-------FQE
     Me_truncatula_CHC   (316) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
     Lo_japonicus_CHC    (307) -----SEATTVGGFYAINRRGQVLLATVNEQTIVNFVSXXXXXXX----XXXXXXXXXXXXXXXXXXXXR-------FHE
           Gl_max_CHC    (319) -----SEATSVGGFYAINRRGQVLLATVNEQTIVNFVSGQLNNLE----LAVNLAKRGNLPGAEKLVVER-------FHE
       Po_tremula_CHC    (315) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
       Ho_vulgare_CHC    (316) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
      Tr_aestivum_CHC    (319) -----AESSTTGGFYAINRRGQVLHATVNDATVVPFVSGQLNNLE----LAVNLAKRANLPGAENLVVQR-------FQE
       So_bicolor_CHC    (309) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
        Or_sativa_CHC    (318) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
          Ze_mays_CHC    (319) -----AESSSTXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------XXX
     Di_discoideum_CHC   (306) -----AFEESTNGIIAVNRKGQVLSVSIDDKNIIPYICNTLNNLE----LAISMACKNNLPGAEGLLTTQ-------FER
         Sc_pombe_CHC    (305) -----TAHKSVNGLMAINRKGQVLSVSINPETIIPYILSNLNDPG----LAVRMASHANLPGADNLYMQQ-------FQQ
    Sa_cerevisiae_CHC    (313) -----APYNHENGIACINKKGQVLAVEISTSQIVPYILNKLSNVA----LALIVATRGGLPGADDLFQKQ-------FES
      Ca_albicans_CHC    (309) -----SSYNDGTGLITINKAGQVLSVEVSRDKIIPYVLDKLSNVP----LALALSSRGGFPGAENLFQQQ-------FQT
      As_fumigatus_CHC   (308) -----APDSESTGLVGVNRKGQVLSVSVDEGTIIQYLMENPAMSG----LAVKLASKAGLPGADHLYQQQ-------FDN
   Pa_brasilensis_CHC    (306) -----TLYTKTGGLLCVNNQGSVFHVSPNDNGIIPFIKDQMRNPE----LAVRIASSANLGGVDDLYRMQ-------LEN
         Le_major_CHC    (307) -----AADNKVGGVLCVNNQGAVMRIGPNNDGIIDFVKNRMHNAE----LALRIASTANLGGVDDLFKVQ-------LDN
        Tr_brucei_CHC    (309) -----TPYTKTGGILCVNNAGSVFHVSPNDNTIVPFVKNQLQNAD----LALRIAGSANLGGVDDLYRVQ-------LDN
    En_histolytica_CHC   (309) -----AYNQSTDGVTVVGKKGVVTNIYIEGDKLVPFICNGLKNVG----LGCKIAARCNLPGADDIFVER-------FNA
   Gi_intestinalis_CHC   (320) TCYEGGTFSDTHDFIGITTDGRCYGIGVDKINLVNYVMTKRGQPD----VAYAIALRTGIRGADSLFQQK-------LDG
      Th_annulata_CHC    (334) -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXXXXXXXXXXXXX-------XXX
        To_gondii_CHC    (358) -----TDSPQTGGILTVNKRGLVCLCNINLQALIPYINQALVYVPNRQQIATSLAKRYGLPGAEEXXXXX-------XXX
        Cr_parvum_CHC    (341) ------SPSRFHGICMANKKGLVLHITLNSSSVLTYIQSNPELSLN--SNLLRWTQRYGYQGTDEFSIRM-------FNE
    Pl_falciparum_CHC    (344) -----CDSKNGEGIYAVNKKGIIFYITINYFHLINHLKLSNLEFKDK--IIKNLCVKYGYPGCDYISAYKKCINDMDFKK
        Pl_yoelii_CHC    (325) -----CDNKNGEGIIAVNKKGKIYYITLNYISLINSLKSSNIDVKDK--IIRNLSLKYGYPGCDYISVYKKCISEMDFKK
       ANCESTRAL_CHC     (368) TCYEGAPHEATSGIIGVNKKGQVLSVCVEEENIVNYATNVLQNPDDK--LALRLAVRSNLAGAEELFARKKCINDMDFNT
           Consensus     (401)             G    N KGQVL                         L N     LA   A R  L GA  L           F
Table 4.3 continued
```

```
                              481                                                                            560
Ho_sapiens_CHC17      (371) LFAQGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPAQ--PGQTSP-LLQYFGILLD-QGQL
Ma_fascicular_CHC17   (348) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXX-XXXXXXXXXX-XXXX
Bo_taurus_CHC17       (371) LFAQGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPAQ--PGQTSP-LLQYFGILLD-QGQL
Su_scrofa_CHC17       (380) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXX
Mu_musculus_CHC17     (371) LFAQGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPAQ--PGQTSP-LLQYFGILLD-QGQL
Ra_norvegicus_CHC17   (371) LFAQGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPAQ--PGQTSP-LLQYFGILLD-QGQL
Ga_gallus_CHC17       (371) LFAQGNYSEAAKVAANAPKGI-------------------LRTPDTIPRFQSVPAQ--PGQTSP-LLQYLRILLD-QGQL
Da_rerio_CHC17        (371) LFAGGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPAQ--PGQTSP-LLQYFGILLD-QGQL
Ta_rubripes_CHC17     (373) LFAAGNYSEAAKVAANAPKGI-------------------LRTPDTIRRFQSVPTQ--PGQTSP-LLQYFGILLD-QGQL
Ho_sapiens_CHC22      (371) LFAQGSYAEAAKVAASAPKGI-------------------LRTRETVQKFQSIPAQ--SGQASP-LLQYFGILLD-QGQL
Ga_gallus_CHC22       (370) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXFQSVPAQ--PGMRLP--CSSTWNMLD-QGQL
Ta_rubripes_CHC22     (371) LFSQGNYSEAAKVAASAPKGI-------------------LRTAETIRKFQSVPAQ--PGQASP-LLQYFGILLD-QGQL
Ci_intestinalis_CHC   (371) LFAQGNYTEAAKVAANAPKGI-------------------LRTSQTIQKFLSVPAQ--PGQNSPPLLQYFGILLD-QGKL
Dr_melanogaster_CHC   (371) LFTAGQYAEAAKVAALAPKAI-------------------LRTPQTIQRFQQVQTP--AGSTTPPLLQYFGILLD-QGKL
Ae_aegyptii_CHC       (371) LFQNGQYAEAAKVAAIAPRGI-------------------LRTPQTIQKFQQVPAQ--PGTNSPPLLQYFGILLD-QGKL
An_gambiae_CHC        (371) LFQNGQFAEAAKVAAIAPKGI-------------------LRTPQTIQKFQQVPAQ--PGTNSPPLLQYFGILLD-QGKL
Ca_elegans_CHC        (369) LFSNGQFGESAKVAASAPQGI-------------------LRTPATIQKFQQCPST--GPGPSP-LLQYFGILLD-QGKL
Ca_briggsae_CHC       (370) LFSNGQFGEAAKVAASAPQGI-------------------LRTPATIQKFQQCPST--GSGPSP-LLQFFGILLD-QGKL
Br_malayi_CHC         (370) LFGNGNYAEAAKVAATAPQGI-------------------LRTPQTIQKFQQCPHT--GGGPSP-LLQYFGILLD-QGQL
Pi_taeda_CHC          (379) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Ly_esculentum_CHC     (379) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Be_vulgaris_CHC       (379) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXX
Me_crystallinum_CHC   (380) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXX
Go_arboretum_CHC      (379) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Ar_thaliana_CHC       (384) LFAQTKYKEAAELAAESPQGI-------------------LRTPDTVAKFQSVPVQ--AGQTPP-LLQYFGTLLT-RGKL
Me_truncatula_CHC     (380) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Lo_japonicus_CHC      (371) LFAQTKYKEAAELAAESPQGI-------------------LRTPDTVAKFQSVPVQ--SGQTPP-LLQYFGTLLT-RGKL
Gl_max_CHC            (383) LFAQTKYKEAAELAAESPQGI-------------------LRTPDTVAKFQSVPVQ--AGQTPP-LLQYFGTLLT-RGKL
Po_tremula_CHC        (379) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Ho_vulgare_CHC        (380) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXX
Tr_aestivum_CHC       (383) LFAQTKYKEAAELAAESPQGL-------------------LRTPETVAKFQSVPVQ--AGQTPP-LLQYFGTLLT-RGKL
So_bicolor_CHC        (373) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Or_sativa_CHC         (382) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXQSVPVQ--AGQTPP-LLQYFGTLLT-RGKL
Ze_mays_CHC           (383) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX---XXXXX-XXXXXXXXXX-XXXX
Di_discoideum_CHC     (370) YFQQGQYKEAAKVAADSPGSI-------------------LRNLQTIQKFQSIPPI--PDQPSA-LLQYFGMLLE-KGKL
Sc_pombe_CHC          (369) LMAQGNYSEAAKVAASSPRGI-------------------LRTSQVIDQFKLIQAA--PGQIAP-ILQYFGTLLD-KGPL
Sa_cerevisiae_CHC     (377) LLLQNDYQNAAKVAASSTS---------------------LRNQNTINRLKNIQAP--PGAISP-ILLYFSTLLD-KGKL
Ca_albicans_CHC       (373) YLNQGDYSNAAKVAASSEQ---------------------LRTQDTINKLKNITPQ--PGQISP-ILQYFSTLLD-RGTL
As_fumigatus_CHC      (372) LLAQGNYSEAAKIAANSPRGF-------------------LRTPETINRFKNAPQT--GQQMSV-ILQYFGMLLD-KGSL
Pa_brasilensis_CHC    (370) GLRVGNIEEAIRACLRAPNNA-------------------LRVPEVLQRFVHQIDS--RGRHAPXXXXXXXXXX-XXXX
Le_major_CHC          (371) FLRNGNVEEAVRTCLRAPGNS-------------------LRTREILMRFMQMPQQ--PGQ-PPATSTYFKIALA-ETSL
Tr_brucei_CHC         (373) QLRAGNIEEAVRTCLRAPNNA-------------------LRGPDILSRFQLMPPI--PGQQPA-ISTYFKVAVA-ETTL
En_histolytica_CHC    (373) LLNAGQIREAALIAANSPGEL-------------------IRNANTINKLSQMQAQ--PGQN-PPLFLYFATILE-SSSL
Gi_intestinalis_CHC   (389) YLGQYQWSEAAELVKIAPAGV-------------------LRTKATIDRFLAGAAFTAPGAPTP-IMIYLTRMLENDGKL
Th_annulata_CHC       (399) XXXXXXXXXXXXXXXXXXXXX------------------XXXXXXXXXXXXXXXXXXXXXXXXXALSYYFSILLE-VDKL
To_gondii_CHC         (426) XXXXXXXXXXXXXXXXXXXXX-------------------XXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXX-XXXX
Cr_parvum_CHC         (406) SIKRQDYQNACRVVSLNKNGS-------------------LRTPATLNHFKMVDS------SNKLLFQYFTTVEK-FHTL
Pl_falciparum_CHC     (417) ASKIICLMKNTKIFEEHSNNIAESILIMNRKNIDIRIIPPLRTQQVLNSFKSFKNT--SGQLSP-LLLYFSVLLE-YDKL
Pl_yoelii_CHC         (398) ASKIICLLKNPKIFEKYSNSISEAITIMNKKNMDIRIVSPLRTQQVLNSFKTFRNS--QGQLSP-LLLYFSVLLE-LDKL
ANCESTRAL_CHC         (446) LFAQGNYSEAAKVAASAPKGIAESILIMNRKNIDIRIIPILRTAETIRKFQSVPAFTQPGQASPPLLQYFGILLDNQGQL
Consensus             (481) L   G   EAA  AA  P                      LRT  T    F   P     GQ  P LLQYF  LL    G L
```

Table 4.3 continued

```
                                561
Ho_sapiens_CHC17      (428) NKYE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKSVDPT-LALSVYLRAN--VPNKVIQ
Ma_fascicular_CHC17   (405) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Bo_taurus_CHC17       (428) NKYE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKSVDPT-LALSVYLRAN--VPNKVIQ
Su_scrofa_CHC17       (439) XXXXXXXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX--XXXXXXX
Mu_musculus_CHC17     (428) NKYE-SLELCRPVLQ------------QGRXXXXXXXXXXXXXXCSEELGDLVKSVDPT-LALSVYLRAN--VPNKVIQ
Ra_norvegicus_CHC17   (428) NKYE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKSVDPT-LALSVYLRAN--VPNKVIQ
Ga_gallus_CHC17       (428) NKYE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLVCWDELGDLVKSVDPT-LALSVYLRAN--VPNKVIQ
Da_rerio_CHC17        (428) NKFE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKSVDPT-LALSVYLRAS--VPIKVIQ
Ta_rubripes_CHC17     (430) NKFE-SLELCRPVLQ------------QGRQ--AAARKMVEQLECSEELGDLVKAVDPT-LALSVYLRAN--VPNKVIQ
Ho_sapiens_CHC22      (428) NKLE-SLELCHLVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKTTDPM-LALSVYLRAN--VPSKVIQ
Ga_gallus_CHC22       (426) NKFE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKTADPT-LALSVYLRAN--VPNKVIQ
Ta_rubripes_CHC22     (428) NKFE-SLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKASDPT-LALSVYLRAN--VPNKVIQ
Ci_intestinalis_CHC   (429) NKFE-SLELCRPVLQ------------QGRKHLMEKWLKDEKLECSEELGDLVKPVDST-LGLSIYLRAT--IPNKVVM
Dr_melanogaster_CHC   (429) NKFE-SLELCRPVLL------------QGKKQLCEKWLKEEKLECSEELGDLVKASDLT-LALSIYLRAN--VPNKVIQ
Ae_aegyptii_CHC       (429) NKYE-SLELCRPVLA------------QGRKQLCEKWLKEEKLECSEELGDLVKPSDHM-LALSIYLRAN--VPNKVIQ
An_gambiae_CHC        (429) NKYE-SLELCRPVLA------------QGRKQLCEKWLKEEKLECSEELGDLVKPSDPT-LALSIYLRSN--VPNKVIQ
Ca_elegans_CHC        (426) NKYE-TLELCRPVLA------------QGRKELITKWLNDQKLECCEELGDLIKPHDVN-TALSVYLRGN--VPHKVVQ
Ca_briggsae_CHC       (427) NKYE-TLELCRPVLA------------QGRKELITKWLNEQKLECCEELGDLIKPHDVN-TALSVYLRGN--VPHKVVQ
Br_malayi_CHC         (427) NKYE-TLELCRPVLA------------QGRKQLLEKWLTEGKLECTEELGDLVRPHDVN-VALSVYLRGN--VPHKVVQ
Pi_taeda_CHC          (437) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Ly_esculentum_CHC     (437) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Be_vulgaris_CHC       (438) XXXXXXXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX--ATPKVVA
Me_crystallinum_CHC   (439) XXXXXXLELSPLVVN------------QNKKNLLENWLADDKLECSEELGDLVKTDNDL--ALKIYIKAR--ATPKVVA
Go_arboretum_CHC      (437) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Ar_thaliana_CHC       (441) NSYE-SLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKTVDND-LALKIYIKAR--ATPKVVA
Me_truncatula_CHC     (438) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXTVDND-LALKIYIKAR--ATPKVVA
Lo_japonicus_CHC      (428) NAFE-SLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKXXDND-LALKIYIKAR--ATPKVVA
Gl_max_CHC            (440) NALE-SLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKTVDND-LALKIYIKAR--ATPKVVA
Po_tremula_CHC        (437) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Ho_vulgare_CHC        (439) XXXXXXLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKTVDND-LALKIYIKAR--ATPKVVA
Tr_aestivum_CHC       (440) NAYE-SLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKTVDND-LALKIYIKAR--ATPKVVA
So_bicolor_CHC        (431) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Or_sativa_CHC         (439) NAYE-SLELSPLVVN------------QNKKNLLENWLAEDKLECSEELGDLVKTVDND-LALKIYIKAR--ATPKVVA
Ze_mays_CHC           (439) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXDLALKIYIKAR--ATPKVVA
Di_discoideum_CHC     (427) NKVE-SLELVRPVLA------------QGKKPILEKWLTEDKLECSEQLGDEVRPHDRK-LALSIYYRAN--ASDKVIT
Sc_pombe_CHC          (426) NEHE-TIELARPVLA------------QNRIQLLEKWYGENKLACTEALGDLVKPYNTP-FALKIYETAN--VPNKVVM
Sa_cerevisiae_CHC     (442) NKEE-TIELARPVLQ------------QDRKQLFEKWLKEDKLECSEELGDIVKPFDTT-LALACYLRAG--AHAKVIS
Ca_albicans_CHC       (428) NKFE-SIELAKPVLQ------------QDRKPLFEKWLKEDKLTSSEELGDIVKSYDTT-LALAVYIRAN--VNIKVVS
As_fumigatus_CHC      (429) NKYE-SVELVRPVLQ------------QNRKHLLEKWMRENKLEASEELGDIVRPYLMN-MALAIYLQAN--VPHKVIA
Pa_brasilensis_CHC    (428) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXX--XXXXXXX
Le_major_CHC          (428) NECE-SVELARAVVP------------KGGIGYVKQQFDEGKLTASEELGDLVQQADST-MALKIFHQGN--AHAKVLN
Tr_brucei_CHC         (430) NAHE-SAELARAVIP------------KGGVDYVKQQYAADKLTASEELGDLMSAADPE-LAIKIYHKAE--AHAKVVG
En_histolytica_CHC    (430) NALE-GFTLIQLLIP------------QNKLPAVEAWIVNDKIECSEKAGDLVKPYIQR-LAIMIYSKH---ATGKVVQ
Gi_intestinalis_CHC   (449) AIEEESMELAKFLVQ------------TSNVSTIETYIKEKRFFATEGLGDYLASVGLEQPSFTTYANAG--CHEKAIA
Th_annulata_CHC       (460) NEIE-SIELIKPVIL------------QNRKELIKKWLENNKLTESETLGDLIYEIDYT-IAFKIYNKLH--LYQKAIN
To_gondii_CHC         (484) XXXX-XXXXXXXXXX------------XXXXXXXXXXXXXXXXXXXXELGDVVRTLDAQ-LAVRVYREAK--AGAKVLQ
Cr_parvum_CHC         (460) NQFE-STEFCKLLLSPNSTSLISEFTSITQPLAFLQVLINEEKLAFSEELGDTLLQNGEKKLALKIFKKCTPQNPTKILQ
Pl_falciparum_CHC     (493) NTYE-SIELVKPVVL------------QKKKEYLEKWIKDDKLTCSEELGDLVKVLDLR-LALNIYLRCS--SHNKILS
Pl_yoelii_CHC         (474) NTYE-SIELVKPVVL------------QKKKEYLEKWIKDDKLACSEELGDLVKSLDLR-LSLIIYLRCG--AHNKIIS
ANCESTRAL_CHC         (526) NKFEESLELCRPVLQ------------QGRKQLLEKWLKEDKLECSEELGDLVKASDPTQLALSVYLRAN--VPNKVIQ
Consensus             (561) N  E SLEL RPV               Q  K L E WL E KLECSEELGDLVK  D   LAL  Y  A      KV
```

Table 4.3 continued

```
                                641                                                                                 720
   Ho_sapiens_CHC17   (491) CFAETGQVQKIVLYAKKV--GYTPDWIFLLRNVMR-IS--PDQGQQ--------------------------------FA
Ma_fascicular_CHC17   (468) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXX-XX--XXXXXX--------------------------------XX
    Bo_taurus_CHC17   (491) CFAETGQVQKIVLYAKKV--GYTPDWIFLLRNVMR-IS--PDQGQQ--------------------------------FA
    Su_scrofa_CHC17   (504) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
  Mu_musculus_CHC17   (491) CFAETGQVQKIVLYAKKV--GYTPDWIFLLRNVMR-IS--PDQGQQ--------------------------------FA
Ra_norvegicus_CHC17   (491) CFAETGQVQKIVLYAKKV--GYTPDWIFLLRNVMR-IS--PDQGQQ--------------------------------FA
    Ga_gallus_CHC17   (491) CFPETGQVQKIVCYAMKV--GYTPDWIFFVRNVMR-IS--PDQGQQ--------------------------------FA
     Da_rerio_CHC17   (491) CFAETGQFQKIVLYAKKV--GYTPDWIFLLRNVMR-IS--PEQGLQ--------------------------------FS
  Ta_rubripes_CHC17   (491) CFAETGQFPKIVLYAKKV--GYTPDWIFLLRNVMR-IN--PEQGLQ--------------------------------FA
    Ho_sapiens_CHC22  (491) CFAETGQFQKIVLYAKKV--GYTPDWIFLLRGVMK-IS--PEQGLQ--------------------------------FS
    Ga_gallus_CHC22   (489) CFAETGQFQKIVLYAKKV--GYTPDWIFLLRSVMR-VS--PEQGLQ--------------------------------FS
  Ta_rubripes_CHC22   (491) CFAETGQFQKIVLYAKKV--GYTPDWVFLLRNVMR-VN--PDQGLQ--------------------------------FA
Ci_intestinalis_CHC   (492) CFAETGQFQKIVMYAKKV--GYTPDYVFLLRNVMR-VN--PETGKQ--------------------------------FA
Dr_melanogaster_CHC   (492) CFAETGQFQKIVLYAKKV--NYTPDYVFLLRSVMR-SN--PEQGAG--------------------------------FA
    Ae_aegyptii_CHC   (492) CFAETGQFQKIVLYAKKV--NYSPDYIFLLRSVMR-TN--PEQGAG--------------------------------FA
     An_gambiae_CHC   (492) CFAETGQFQKIVLYAKKV--NYSPDYVFLLRSVMR-TN--PEQGSG--------------------------------FA
     Ca_elegans_CHC   (489) SFAETGQFDKIVMYAKRV--GFQPDYLFQLRQILRNSN--PDHGAK--------------------------------FA
    Ca_briggsae_CHC   (490) SFAETGQFDKIVMYAKRV--GFQPDYLFQLRQILRNSN--PDNGAK--------------------------------FA
      Br_malayi_CHC   (490) CFAETGQFDKIILYAKKV--NFEPDYLFQLRQVLR-SN--PEMGAK--------------------------------FA
       Pi_taeda_CHC   (500) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
  Ly_esculentum_CHC   (500) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
     Be_vulgaris_CHC  (503) AFAERREFDKILIYSKQV--GYTPDYLFLLQTILR-TD--PQGAVN--------------------------------FA
Me_crystallinum_CHC   (502) AFAERREFDKILVYSKQV--GYTPDYLFLLQTILR-TD--PQGAVN--------------------------------FA
    Go_arboretum_CHC  (500) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
     Ar_thaliana_CHC  (504) AFAERREFDKILIYSKQV--GYTPDYMFLLQTILR-TD--PQGAVN--------------------------------FA
  Me_truncatula_CHC   (501) AFAERREFDKILVYSKQVX-XXXXXXXXXXXXXXX-XX--XXXXXX--------------------------------XX
    Lo_japonicus_CHC  (491) AFAERREFDKILIYSKQV--XXXXXXXXXXXXXXX-XX--XXXXXXX-------------------------------XX
          Gl_max_CHC  (503) AFAERREFDKILIYSKQV--GYTPDYLFLLQTILR-TD--PQGAVN--------------------------------FA
       Po_tremula_CHC (500) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
      Ho_vulgare_CHC  (503) AFAERREFDKILIYSKQV--GYTPDYLFLLQTILR-TD--PQGAVN--------------------------------FA
     Tr_aestivum_CHC  (503) AFAERREFDKILIYSKQV--GXXXXXXXXXXXLR-TD--PQGAVN--------------------------------FA
      Sc_bicolor_CHC  (494) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXX-XX--XXXXXX--------------------------------XX
       Or_sativa_CHC  (502) AFAERREFDKILIYSKQV---YLFXXXXXXXXXXX-XX--XXXXXX--------------------------------XX
        Ze_mays_CHC   (503) AFAERREFAKILIYSKQV--CYXXXXXXXXXXXXX-XX--XXXXXX--------------------------------XX
  Di_discoideum_CHC   (490) LFAETGEFDKIIAYCKKF--NYKPDFMFLLQRMAN-AN--PMGAAD--------------------------------FA
       Sc_pombe_CHC   (489) CLSELGDFGKLATYTSQQ--NITPDYVSLLQNLVR-VN--PDQAAE--------------------------------FA
  Sa_cerevisiae_CHC   (495) CLAELQQFEKIIPYCQKV--GYQPNFLVLISSLIR-SS--PDRASE--------------------------------FA
     Ca_albicans_CHC  (491) CLAELGQFDKILPYCQKV--GYNPDYTNLIQNLVR-VN--PDKASE--------------------------------FA
    As_fumigatus_CHC  (492) GFAETGQFDKILTYSKQV--GYQPDYTQLLQHIVR-VN--PEKGAE--------------------------------FA
  Pa_brasilensis_CHC  (491) XXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXX--XXXXXX--------------------------------XX
       Le_major_CHC   (491) LLIQRNETQKAVEYCKRA--SFSPDWRVIMNNFIR-VA--PQDAVN--------------------------------LG
       Tr_brucei_CHC  (493) VLLQRNETQKAVEYCKRA--GFTPNWRVILNNAIH-VN--PQGAVG--------------------------------LA
  En_histolytica_CHC  (492) SLAEMGQSDKIVAYCERA--GFSPDYLQILQXXXX-XX-----XXX---------------------------------
 Gi_intestinalis_CHC  (514) WLIAHNQYAIIPQYAQKFP-EYTPDYPRIISQALKQAEQRPDGASTIES-------------------------IYQFS
     Th_annulata_CHC  (523) CLLXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXX-XX--XXXXXXXXXXX-------------------------XX
       To_gondii_CHC  (547) TLTELGNFDEIIAFAKET--KLEADYAGLLRNLVN-VH--PENAVK--------------------------------FA
       Cr_parvum_CHC  (539) TLIEIGNFSHVSEFIREQRASNWPTITTDIRSLLNSLLIANNIDATVEFVKTVILPPAP------------------SSQ
  Pl_falciparum_CHC   (556) TYCLLNMFNNVLSYINNFK-QINFDYVGIFITIVNYEF--PQGNADNKDSAFVGDFNSGSKNETMDFFN--------DGA
       Pl_yoelii_CHC  (537) IYCLLNMFNNVMTYINNYKHITNFDFVNIFITIINYEF--PTTSGGGSFNSGNTNGNDSGSSLDLHHTEKNESLDFFNDN
       ANCESTRAL_CHC  (591) CFAETGQFQKIVLYAKKVKHGYTPDWVFLLRNVMRNVNTDPEQGLQNKDSAFVGDFNDGSKNLFLYFTNKNEMDFFYDFA
             Consensus  (641)  FAE G F KI  Y K V  GY PD   LL    R     P                                    FA
```

Table 4.3 continued

```
                        721                                                                            800
     Ho_sapiens_CHC17  (534) QMLVQDEE--------------------------PLADITQIVDVFMEYNLIQQCTAFLLDALK-NNRPS-EGPLQTRLL
   Ma_fascicular_CHC17 (511) XXXXXXXX--------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXX-XXXXXXXXX
      Bo_taurus_CHC17  (534) QMLVQDEE--------------------------PLADITQIVDVFMEYNLIQQCTAFLLDALK-NNRPS-EGPLQTRLL
      Su_scrofa_CHC17  (548) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Mu_musculus_CHC17  (534) QMLVQDEE--------------------------PLADITQVMDVFMEYNLIQQCTAFLLDALK-NNRPS-EGPLQTRLL
  Ra_norvegicus_CHC17  (534) QMLVQDEE--------------------------PLADITQIVDVFMEYNLIQQCTAFLLDALK-NNRPS-EGPLQTRLL
      Ga_gallus_CHC17  (534) QMLFQDEE--------------------------PLADITQIVDVSMEYNLIQQCTAFLLDALK-NNPPS-EGPLQTRLL
        Da_rerio_CHC17 (534) QMLVQDEE--------------------------PLADITQIVDVFMEYNLIQQCTSFLLDALK-NNRPT-EGPLQTRLL
    Ta_rubripes_CHC17  (534) QMLVQDEE--------------------------PLADITQIVDVFMEYNLIQQCTSFLLDALK-NNRPS-EGPLQTRLL
     Ho_sapiens_CHC22  (534) RMLVQDEE--------------------------PLANISQIVDIFMENSLIQQCTSFLLDALK-NNRPA-EGLLQTWLL
      Ga_gallus_CHC22  (532) QMLVQDEE--------------------------PLANINQIVDVFMENSLIQQCTSFLLDALK-NNRPA-EGHLQTRLL
    Ta_rubripes_CHC22  (534) QMLVQDEE--------------------------PLANINQIVDVFMEGNLIQQCTSFLLDALK-NNRPA-EGHLQTRLL
  Ci_intestinalis_CHC  (535) MMLVQDDE--------------------------PLADVAQVXDVFMEQNLIQLCTAFLLDALK-NNRPA-EGALQTRLL
  Dr_melanogaster_CHC  (535) SMLVAEEE--------------------------PLADINQIVDIFMEHSMVQQCTAFLLDALK-HNRPA-EGALQTRLL
      Ae_aegypti_CHC   (535) SMLVADEE--------------------------PLADINQIVDIFMEQNMVQQCTAFLLDALK-NNRPS-EGALQTRLL
     An_gambiae_CHC    (535) SMLVADEE--------------------------PLADINQIVDIFMEQNMVQQCTAFLLDALK-NNRPA-EGALQTRLL
      Ca_elegans_CHC   (533) QLLVSESENGE-----------------------PLADLSQIIDCFMEVQAVQPCTSFLLEVLK-GDKPE-EGHLQTRLL
     Ca_briggsae_CHC   (534) QLLVSESENGE-----------------------PLADLSQIIDCFMEVQAVQPCTSFLLEVLK-GDKPE-EGHLQTRLL
       Br_malayi_CHC   (533) QMLVSEGENX------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXX
        Pi_taeda_CHC   (544) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXX
  Ly_esculentum_CHC    (544) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXX
     Be_vulgaris_CHC   (546) LMMSQMEGG-------------------------CPLDYNTITDLFLQRNLIREATAFLLDVLK-PNLPE-HGYLQTKVL
 Me_crystallinum_CHC   (545) LMMSQMEGG-------------------------CPMDYNTITDLFLQRNLIREATAFLLDVLK-PNLPE-HGYLQTKVL
   Go_arboretum_CHC    (544) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXX
     Ar_thaliana_CHC   (547) LMMSQMEGG-------------------------SPVDYNTITDLFLQRNLIREATSFLLDVLK-PNLPE-HAFLQTKVL
   Me_truncatula_CHC   (545) XXXXXXXX--------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXX--XXXXX-XXXXXXXXX
    Lo_japonicus_CHC   (535) XXXXXXX---------------------------XXXXXXXXXXXXXXXNLIREATAFLLDVLK-PNLPE-HGFLQTKVL
          Gl_max_CHC   (546) LMMSQMEGG-------------------------CPVDYNTITDLFLQRNLIREATAFLLDVLK-PNLPE-HGYLQTKVL
      Po_tremula_CHC   (544) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXX
      Ho_vulgare_CHC   (546) LMMSQMEGG-------------------------CPLDYNTITDLFLQRNMIREATAFLLDVXX-XXXXX-XXXXXXXXX
     Tr_aestivum_CHC   (546) LMMSQMEGG-------------------------CPLDYNTITDLFLQRNMIREATAFLLDVLK-PNLEE-HAFLQTKVL
      So_bicolor_CHC   (537) XXXAQMEGG-------------------------CPVDYNTITDLFLQRNMIREATAFLLDVLK-PNLPE-HAFLQTKVL
       Or_sativa_CHC   (544) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXX-XXXXXXXXX
         Ze_mays_CHC   (546) XXXXXXXXX-------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXVL
  Di_discoideum_CHC    (533) VKLVKEEGG-------------------------PYIDANQVVELFSARNMIQETSNFLFAILD-GDRPQ-DANLQTKLL
        Sc_pombe_CHC   (532) TQMFNSN---------------------------PSINLEKIVDIFMSQNLVQQATAFLLDALK-DDNPE-HSHLQTRLL
  Sa_cerevisiae_CHC    (538) VSLLQNPETA------------------------SQIDIEKIADLFFSQNHIQQGTSLLLDALK-GDTPD-QGHLQTRVL
     Ca_albicans_CHC   (534) TSLLSSPD--------------------------ANLNVEQIADLFFSQNYIQQGTAFLLDALK-NDTPA-EGHLQTKVL
    As_fumigatus_CHC   (535) AQLANEESG-------------------------ALIDLDRVVDVFLSQNMIQQATSFLLDALK-DNRPE-HGHLQTRLL
  Pa_brasilensis_CHC   (535) XXXXXXXX--------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
        Le_major_CHC   (534) LMLYREMGDR------------------------PVLTADEVVDMFVSAQQIQQATEFLLEILR-NHNDESTMELQTKLL
       Tr_brucei_CHC   (536) QVLHRDMGDA------------------------PVVDPIEVVDMFVTAQHIQQATEFVLEVLR-DNTGENTKDLQTKLL
  En_histolytica_CHC   (530) ----------------------------------XXIMISIIDLFARYRYLKEITALILDTLD-GDNEQ-YANIQTKVL
 Gi_intestinalis_CHC   (567) LSITQALDATT-----------------------DGDKILDIADCFADIGRVNEATTILLEVCKRYNYSDDTVIYQSMVI
     Th_annulata_CHC   (571) XXXXXXXXXXXX----------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
       To_gondii_CHC   (590) QQLLSSEP--------------------------PLADVTQVSEVLLQQHKYQEFTSLMLDFLK-GNKPE-QGPLQTRLL
       Cr_parvum_CHC   (601) TDLANSTSAESLD---------------------LNIDKTSVVEIFVSHSRYKEITSILLDHLK-ANKPE-DSALQTKLL
   Pl_falciparum_CHC   (625) ASISNDQNNSINKLNKGHINEVAIQYIKFLCENNISFDINKIIDYLLSKKKLQEATSILLDYLK-ENKAE-HKNLQTKLF
       Pl_yoelii_CHC   (615) NSVSNSKDLKNGNNHRNTNIEIATQYIKLLSDNNIQLDIGKIVDYLLNNNKLQEATSILLDYLK-DNKPE-HKHLQTKIF
       ANCESTRAL_CHC   (671) QMLVQDEEPTEFCRLLLTINEIAIQYIKLLCDGEPLANINQIVDVFMEGNLIQQCTSFLLDALKRNNRPADEGHLQTRLL
           Consensus   (721)  L                                  D   I D F   N IQ  T FLLD LK  N P    G LQT LL
Table 4.3 continued
```

```
                                801                                                                                 880
    Ho_sapiens_CHC17   (586) EMNLMHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HFTDLYDIKRAVVHT-HLLNP-----
 Ma_fascicular_CHC17   (561) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXX-----
    Bo_taurus_CHC17    (586) EMNLMHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HFTDLYDIKRAVVHT-HLLNP-----
    Su_scrofa_CHC17    (603) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
  Mu_musculus_CHC17    (586) EMNLMHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HFTDLYDIKRAVVHT-HLLNP-----
 Ra_norvegicus_CHC17   (586) EMNLMHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HFTDLYDIKRAVVHT-HLLNP-----
    Ga_gallus_CHC17    (586) EMNPMQA-PQVADAILGKPIFIHFDRAHIAQLCEKAGLLQGALE----------HFQDLYDIKRAVVHT-HLLNP-----
     Da_rerio_CHC17    (586) EMNLMHA-PQVANAILGNQMFTHYDRAHVAQLCEKAGLLQRALE----------HYTDLYDIKRAGVHT-HLLNP-----
   Ta_rubripes_CHC17   (586) EMNLMHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-HLLNP-----
    Ho_sapiens_CHC22   (586) EMNLVHA-PQVADAILGNKMFTHYDRAHIAQLCEKAGLLQQALE----------HYTDLYDIKRAVVHT-HLLNP-----
    Ga_gallus_CHC22    (584) EMNLIHA-PQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-HLLNP-----
   Ta_rubripes_CHC22   (586) EMNLIHA-PQVADAILGNQMFTHYDRAHVAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-HLLNP-----
 Ci_intestinalis_CHC   (587) EMNLMHA-PQVADAILGNQMFTHYDRAHVAQLCEKAGLLQRALE----------HFVDLYDIKRAIIHT-HLLNP-----
 Dr_melanogaster_CHC   (587) EMNLMSA-PQVADAILGNAMFTHYDRAHIAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-HMLNA-----
     Ae aegyptii_CHC   (587) EMNLMSA-PQVADAILGNAMFTHYDRAHIAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-HLLNG-----
     An_gambiae_CHC    (587) EMNLMSA-PQVADAILGNAMFTHYDRAHIAQLCEKAGLLQRALE----------HYTDLYDIKRAVVHT-QLLNG-----
     Ca_elegans_CHC    (588) EMNLLAA-PAVADAILANKMFSHYDRAAIGQLCEKAGLLQRALE----------HFTDLYDIKRTVVHT-HLLKP-----
     Ca_briggsae_CHC   (589) EMNLLAA-PAVADAILANKMFSHYDRAAIGQLCEKAGLLQRALE----------HFTDLYDIKRTVVHT-HLLKP-----
       Br_malayi_CHC   (588) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXX-----
        Pi_taeda_CHC   (598) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
   Ly_esculentum_CHC   (598) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
     Be_vulgaris_CHC   (599) EINLVTF-PNVADAILANGMFSHYDRPRIAQLCEKGGLYLQALK----------HYTELPDVKRVIVNT-HAIEP-----
  Me_crystallinum_CHC  (598) EINLVTF-PNVADAILANGMFSHYDRSRNVQLCEKGGLYLQLLT----------LY-GLPDXXXXXXX-XXXXX-----
    Go_arboretum_CHC   (598) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
     Ar_thaliana_CHC   (600) EINLVTF-PNVADAILANGMFTHYDRPRIAQLCEKAGLYIQSLK----------HYSELPDIKRVIVNT-HAIEP-----
   Me_truncatula_CHC   (596) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXX-----
     Lo_japonicus_CHC  (586) EINLVTF-PNVADAILANGMFSHYDRPRIAQLCEKAGLYVRALQ----------HYTELPDIKRVIVNT-HAIEP-----
          Gl_max_CHC   (599) EINLVTF-PNVADAILANGMFSHYDRPRIAQLCEKAGLYVRALQ----------HYTELPDIKRVIVNT-HAIEP-----
       Po_tremula_CHC  (598) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
       Ho_vulgare_CHC  (599) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXX-----
     Tr_aestivum_CHC   (599) EINLVTY-PNVADAILANGMFSHYDRPRIAQLCEKAGLYLRALQ----------HYAELPDIKRVIVNT-HAIEP-----
       Sb_bicolor_CHC  (590) EINLVTY-PNVADAILANGMFSHYDRPRIAQLCEKAGLYLRALQ----------HYSELPDIKRVIVNT-HAIEP-----
        Or_sativa_CHC  (597) XXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXX-----
         Ze_mays_CHC   (600) EINLVTY-PNVADAILANGMFSHYDRPRIAQLCEKAGLYLRALQ----------HYSELPDIKRVIVNT-HAIEP-----
   Di_discoideum_CHC   (586) EINLLHA-PQNADAIMGGQKFTHYNPLRIGGLCEKAGLYQRALE----------HYTDLADIKRVLSHAGHMVNQ-----
       Sc_pombe_CHC    (583) EINLINA-PQVADAILGNQMFTHFDRAVIASLCERAGLVQRALE----------LYDKPADIKRVIVHS-NLLNP-----
    Sa_cerevisiae_CHC  (592) EVNLLHA-PQVADAILGNNIFSHYDKPTIASLSEKAGLYQRALE----------NYTDIKDIKRCVVHT-NALPI-----
     Ca_albicans_CHC   (586) EINLLHA-PQVADAILGNQMFSHYDKPTIGKLCEKSGLFQRALE----------HYDDLKDIKRVIVHT-NVLPS-----
     As_fumigatus_CHC  (588) EMNLVNA-PQVADAILGNEIFTHYDRPRISQLCENAGLIQRALE----------NTDDPAVIKRNIVRT-DKLNP-----
  Pa_brasilensis_CHC   (589) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
        Le_major_CHC   (589) EMNLKYSHPSVADKIFVRKICQCYDGMRLAPLCERAGLHQHAIDCYIMAQQQDSDLNNLASIRRCLQQL-QNFNP-----
       Tr_brucei_CHC   (591) EINLKHSHPSVAEKIFARGVCVHFDAMLLAPLCERASLPQPAIECYVMAQRMDPGIDNLANIRRCFSNA-QVLNP-----
   En histolytica_CHC  (573) EMNLIGA-PQIADTIFENDMLKHFDHNRIGKLCEQAGLFKKALQ----------IFEQFDDIRRVLAHA-SAIPP-----
   Gi_intestinalis_CHC (624) ELNRGVN-QAVVDAIQAKGVLARINPAEITKRSEESGLYQRAFE----------FCDSDAECRRIASTYAAKLGKSIFS-
      Th_annulata_CHC  (629) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXX-----
        To_gondii_CHC  (642) EVNLLHS-PQVAETIFQMEMLTHFDRAKIAALCEKVGLSQRALE----------LYTDIADIKRVMLQSGGKISQ-----
        Cr_parvum_CHC  (658) EVNLLHA-PQVAEALFQMDLFTHYDKHAIAALCEKAGLYERALE----------NFSDMRDIRRILGVACGSLNT-----
   Pl_falciparum_CHC   (703) EFNLYNN-VQVAETLFQMDIFTYYDKNKIAYLCEEKGLYQRALE----------NYTNINDIKRVITKSTCFQKGGNNNN
       Pl_yoelii_CHC   (693) EFNLHHN-VQIAETIFQMDIFTYYDKNRIAYLCEEKGLYQRALE----------NYTNINDIKRVITKSACFQKNGN---
       ANCESTRAL_CHC   (751) EMNLIHAHPQVADAILGNQMFTHYDRAHIAQLCEKAGLLQRALENYVMAQRMDPHYTDLYDIKRAVVHTGHLLNPGNNNN
            Consensus  (801) E NL    PQVADAIL N MF HYDR  IAQLCEKAGL QRALE          H   L DIKR  V T     P
```

Table 4.3 continued

```
                                881                                                                            960
      Ho_sapiens_CHC17    (649) -----------------------------EWLVNYFGSLS--VEDSLECLRAMLSANIRQNLQICVQVA---SKYHEQLS
    Ma_fascicular_CHC17   (624) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
        Bo_taurus_CHC17   (649) -----------------------------EWLVNYFGSLS--VEDSLECLRAMLSANIRQNLQICVQVA---SKYHEQLS
        Su_scrofa_CHC17   (668) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
      Mu_musculus_CHC17   (649) -----------------------------EWLVNYFGSLS--VEDSLECLRAMLSANIRQNLQICVQVA---SKYHEQLS
    Ra_norvegicus_CHC17   (649) -----------------------------EWLVNYFGSLS--VEDSLECLRAMLSANIRQNLQIWVQVA---SKYHEQLS
        Ga_gallus_CHC17   (649) -----------------------------ECLVNYFGSLS--VENSLEWLSAILSANIRQNLQICVQVA---SKYHEQLS
          Da_rerio_CHC17  (649) -----------------------------EWLVNFFGSLS--VEDSMECLRAMLSANIRQNLQICVQVA---SKYHEQLS
      Ta_rubripes_CHC17   (649) -----------------------------EWLVNFFGSLS--VEDSLECLRAMLSANIRQNLQICVQVA---SKYHEQLS
       Ho_sapiens_CHC22   (649) -----------------------------EWLVNFFGSLS--VEDSVECLHAMLSANIRQNLQLCVQVA---SKYHEQLG
        Ga_gallus_CHC22   (647) -----------------------------EWLVNFFGSLS--VEDSVECLRAMLSANIRQNLQLCVQVA---SKYHEQLG
      Ta_rubripes_CHC22   (649) -----------------------------EWLVNFFGSLS--VEDSLECLRAMLSANIRQNLQLCVQVA---SKYHEQLG
    Ci_intestinalis_CHC   (650) -----------------------------EWLVNYFGSLS--VEDSLECLKAMLTANIRQNLQICVQVA---TKYHEQLT
    Dr_melanogaster_CHC   (650) -----------------------------EWLVSFFGTLS--VEDSLECLKAMLTANLRQNLQICVQIA---TKYHEQLT
        Ae_aegyptii_CHC   (650) -----------------------------DWLVGFFGTLS--VEDSLECLKAMLTANIRQNLQICVQIA---TKYHEQLT
        An_gambiae_CHC    (650) -----------------------------DWLVGFFGTLS--VEDSLECLKAMLTANIRQNLQICVQIA---TKYHEQLT
        Ca_elegans_CHC    (651) -----------------------------DWLVGYFGSLS--VEDSVECLKAMLTQNIRQNLQVVVQIA---SKYHEQLG
       Ca_briggsae_CHC    (652) -----------------------------DWLVGYFGSLS--VEDSLECLKAMLTQNIRQNLQVVVQIA---SKYHEQLG
          Br_malayi_CHC   (651) -----------------------------XXXXXXXXXXX--XXXXXXXXXAMLQTNMRQNLQIVVQIA---TKYHEQLT
           Pi_taeda_CHC   (662) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
     Ly_esculentum_CHC    (662) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
        Be_vulgaris_CHC   (662) -----------------------------QALVEFFGTLS--KEWALXCMKDLLLVNLRGNLQIIVQTA---KEYSDQLG
   Me_crystallinum_CHC    (660) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
      Go_arboretum_CHC    (662) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
        Ar_thaliana_CHC   (663) -----------------------------QALVEFFGTLS--SEWAMECMKDLLLVNLRGNLQIIVQAC---KEYCEQLG
     Me_truncatula_CHC    (659) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
       Lo_japonicus_CHC   (649) ----------------------------QVCLVEFFGTLS--QEWALECMKDLLLANLRGNLQIIVQVA---KEYCEQLG
             Gl_max_CHC   (662) -----------------------------QSLVEFFGTLS--REWALECMKDLLLVNLRGNLQIIVQVA---KEYCEQLG
         Po_tremula_CHC   (662) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
         Ho_vulgare_CHC   (662) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
       Tr_aestivum_CHC    (662) -----------------------------QALVEFFGTLS--KEWALECMKDLLLVNLRGNLQIVVQAA---KEYCEQLG
         So_bicolor_CHC   (653) -----------------------------QALVEFFGTLS--REWALECMKDLLLVNLRGNLQIVVQAA---KEYSEQLG
          Or_sativa_CHC   (660) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
            Ze_mays_CHC   (663) -----------------------------QALVEFFGTLS--REWALECMKDLLLVNLRGNLQIVVQAA---KEYSEQLG
     Di_discoideum_CHC    (650) -----------------------------EFLVSYFGSLN--PEDRMECMRDFLRTNPRQNLQLVVAIA---VSYSDQIT
           Sc_pombe_CHC   (646) -----------------------------EWLMNYFSRFS--PDEVYDYLREMLRSNLRQNLQIVVQIA---TRYSDLVG
    Sa_cerevisiae_CHC     (655) -----------------------------DWLVGYFGKLN--VEQSLACLKALMDNNIQANIQTVVQVA---TKFSDLIG
        Ca_albicans_CHC   (649) -----------------------------DWLVSYFGQLN--VDQSVACIKELLSNNMQQNLQVVIQVA---TKYSDLIG
      As_fumigatus_CHC    (651) -----------------------------EWLMNYFGRLS--VEQTLDCMDTMLEVNIRQNLQAVVQIA---TKFSDLLG
    Pa_brasilensis_CHC    (654) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
           Le_major_CHC   (663) -----------------------------EWIVDFFGKLN--KQDSLKCLEDLCTDS-RQNFKVIVQVA---TKYSDALG
          Tr_brucei_CHC   (665) -----------------------------DWVIEFFGKLS--PGDSMKCLKDLLANH-HQNFKIIVQVA---TKYNEALG
    En_histolytica_CHC    (636) -----------------------------ELIIESFRKLE--PEQAILVLQDLLRTNPRGNLQIIIKIL---LEFHTGLG
    Gi_intestinalis_CHC   (692) ---------------------------SPRWYGAYIKRASRFSNPTDDVIKSLFAFQ-QQMLEITIGGE---RIYADHVA
        Th_annulata_CHC   (694) -----------------------------XXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXX---XXXXXXXX
          To_gondii_CHC   (706) -----------------------------EFTQQFFGNLP--PDASLEILTDMLRSS-SQNLQAVVAVA---IKFHGQIG
          Cr_parvum_CHC   (722) -----------------------------DWLANYLSKLS--PRTRFDCLKELLXXXXXXXLQSVIQVC---IKNVDNIG
      Pl_falciparum_CHC   (772) TTTTTTTTSSNMGDGHFDMNNVSKGKISIEWIKNYFSTLS--DSVCEELLFDFMKGS-KINMEVVISIC---VQYYDKIG
          Pl_yoelii_CHC   (759) -------ISNNLNSEGGSGNMSLHGGISLEWIKNYFSTLS--DSVCQDLLFDFMKGN-KVNIEIIISIC---VQYYNKIG
          ANCESTRAL_CHC   (831) TTTTTTTISSNLGDGGFDMNNVLKGKISQEWLVNFFGSLSRFVEDSLECLRAMLSANIRQNLQLCVQVASACSKYHEQLG
              Consensus   (881)                              WLV  FG LS    E   ECL    L  N RQNLQ  VQ A     KY EQL
```

Table 4.3 continued

```
                                961                                                                                1040
   Ho_sapiens_CHC17   (695) T---------------QSLIELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
Ma_fascicular_CHC17   (670) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
     Bo_taurus_CHC17   (695) T---------------QSLIELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
     Su_scrofa_CHC17   (714) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
   Mu_musculus_CHC17   (695) T---------------QSLIELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
Ra_norvegicus_CHC17   (695) T---------------QSLIELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
     Ga_gallus_CHC17   (695) T---------------QSLIELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
      Da_rerio_CHC17   (695) T---------------QSLTELFESFK--SFEGGLFYFLGSIVN----------------------------FSQD----
   Ta_rubripes_CHC17   (695) T---------------QALTELFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
     Ho_sapiens_CHC22  (695) T---------------QALVELFESFK--SYK-GLFYFLGSIVN----------------------------FSQD----
     Ga_gallus_CHC22   (693) T---------------QSLVXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
   Ta_rubripes_CHC22   (695) T---------------QSLVELFESFK--SYE-GLFYFLGSIVN----------------------------FSQE----
Ci_intestinalis_CHC    (696) T---------------EKLIEVFESFK--SFE-GLFYFLGSIVN----------------------------FSQD----
Dr_melanogaster_CHC    (696) N---------------KALIDLFEGFK--SYD-GLFYFLSSIVN----------------------------FSQD----
     Ae_aegyptii_CHC   (696) T---------------KALIDLFESFK--SYE-GLFYFLGSIVN----------------------------FSQD----
     An_gambiae_CHC    (696) T---------------KALIDLFESFK--SYE-GLFYFLGSIVN----------------------------FSQD----
     Ca_elegans_CHC    (697) A---------------DKLIEMFENHK--SYE-GLFYFLGSIVN----------------------------FSQD----
     Ca_briggsae_CHC   (698) A---------------DKLIEMFETHK--SYE-GLFYFLGSIVN----------------------------FSQD----
       Br_malayi_CHC   (697) T---------------HALIDLFESFK--SYE-GLFYFLGSIVN----------------------------FSQD----
       Pi_taeda_CHC    (708) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
  Ly_esculentum_CHC    (708) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
     Be_vulgaris_CHC   (708) V---------------DSCIKLFEQFK--SYE-GLYFFLGSYLS----------------------------SSED----
Me_crystallinum_CHC    (706) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
    Go_arboretum_CHC   (708) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
     Ar_thaliana_CHC   (709) V---------------DACIKLFEQFK--SYE-GLYFFLGSYLS----------------------------MSED----
   Me_truncatula_CHC   (705) X---------------XXXXKIFEQFR--SYE-GLYFFLGSYLS----------------------------SSED----
     Lo_japonicus_CHC  (696) V---------------DACIKLFEQFR--SYE-GLYFFLGSYLS----------------------------DSED----
           Gl_max_CHC  (708) V---------------DACIKLFEQFR--SYE-GLYFFLGSYLS----------------------------SSED----
       Po_tremula_CHC  (708) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
       Ho_vulgare_CHC  (708) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
     Tr_aestivum_CHC   (708) V---------------DACIKLFEQFK--SYE-GLYFFLGSYLS----------------------------SSED----
       So_bicolor_CHC  (699) V---------------DACIKLFEQFK--SYE-GLYFFLGSYLS----------------------------SSED----
       Or_sativa_CHC   (706) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
         Ze_mays_CHC   (709) V---------------DACIKLFEQFK--SYE-XXXXXXXXXXX----------------------------XXXX----
   Di_discoideum_CHC   (696) P---------------EAIIAMFESFR--LYE-GLYLYLTQVVV----------------------------TSTS----
       Sc_pombe_CHC    (692) A---------------QRIIEMFEKFK--TFE-GLYYYLGSIVN----------------------------ITED----
   Sa_cerevisiae_CHC   (701) P---------------STLIKLFEDYN--ATE-GLYYYLASLVN----------------------------LTED----
     Ca_albicans_CHC   (695) A---------------AKLIKIFEEYK--CTE-GLYYYLSSIVN----------------------------LTQD----
    As_fumigatus_CHC   (697) P---------------GQLISLFEKYR--TAE-GLYYYLGSIVN----------------------------LSED----
  Pa_brasilensis_CHC   (700) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
       Le_major_CHC    (708) A---------------ADLINLFLEHS--LYD-VLYYYLGAVVP----------------------------YTRD----
       Tr_brucei_CHC   (710) A---------------DKLINVFLRARN-CLD-ILYYYLGAVVP----------------------------YTRD----
  En_histolytica_CHC   (682) D---------------DKVIDLLQKFD--CWE-GLFYFLSQIVN----------------------------TTND----
Gi_intestinalis_CHC    (741) MGFALSPIVVSNVSLCVTVVKVFEKYGMGNSE-ALYLFLSMIAP----------------------------TCLD----
     Th_annulata_CHC   (740) X---------------XXXXXXXXXXX--XXX-XXXXXXXXXXX----------------------------XXXX----
       To_gondii_CHC   (751) T---------------TKLVEMFEKFS--SYE-GVFYFLGSILA----------------------------FSSD----
       Cr_parvum_CHC   (768) I---------------ENIITLFEQQG--IWE-GIYYVVGSXXX----------------------------XXXX----
   Pl_falciparum_CHC   (846) I---------------KKIINKFEENK--NYE-GIFYFVSNILNDLPNLIVKGNNNDN-------------LSNNSLST
       Pl_yoelii_CHC   (826) I---------------KKIVNKFEENK--NYE-GIFYFISSILNELQNMGTVGNANDIGKYNDNNSISSNNN1NNNLSSS
      ANCESTRAL_CHC    (911) TGFALSPIVVSNVSLCQALVELFESFKNGSYE-GLFYFLGSIVNDLPNLIVKGNNNDIGKYNDNNSISSNDNFSQDLLSS
          Consensus    (961)                    I LFE FK  S E GL YFLGSIV                                 S D
```

Table 4.3 continued

```
                                1041                                                                1120
   Ho_sapiens_CHC17    (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
Ma_fascicular_CHC17    (700) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
     Bo_taurus_CHC17   (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
     Su_scrofa_CHC17   (744) ----------------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   Mu_musculus_CHC17   (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
Ra_norvegicus_CHC17    (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
     Ga_gallus_CHC17   (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
      Da_rerio_CHC17   (726) ----------------------------------PEVHFKYIQAAWKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
   Ta_rubripes_CHC17   (725) ----------------------------------PEVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
    Ho_sapiens_CHC22   (725) ----------------------------------PDVHLKYIQAACKTG-QIKEVERICRES-SCYNPERVKNFLKEAK
     Ga_gallus_CHC22   (723) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXX
   Ta_rubripes_CHC22   (725) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDPERVKNFLKEAK
Ci_intestinalis_CHC    (726) ----------------------------------PDVHFKYIQAACKTG-QIKEVERICRES-NCYDADKVKNFLKEAK
Dr_melanogaster_CHC    (726) ----------------------------------PEVHFKYIQAACKTN-QIKEVERICRES-NCYNPERVKNFLKEAK
     Ae_aegyptii_CHC   (726) ----------------------------------PEVHFKYIQAACKTN-QIKEVERICRES-NCYNAERVKNFLKEAK
      An_gambiae_CHC   (726) ----------------------------------PEVHFKYIQAACKTN-QIKEVERICRES-NCYNAERVKNFLKEAK
      Ca_elegans_CHC   (727) ----------------------------------PEVHFKYIQAATRTG-QIKEVERICRES-QCYDAERVKNFLKEAK
     Ca_briggsae_CHC   (728) ----------------------------------PEVHFKYIQAATRTG-QIKEVERICRES-QCYDAERVKNFLKEAK
       Br_malayi_CHC   (727) ----------------------------------PEVHFKYIQAATRTG-QIKEVERICRES-NCYEAERVKNYLKEAK
        Pi_taeda_CHC   (738) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
  Ly_esculentum_CHC    (738) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
    Be_vulgaris_CHC    (738) ----------------------------------PDIHFRYIEAAAKTG-QLKEVERVTRES-NFYDAEKTKNFLMEAK
Me_crystallinum_CHC    (736) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
    Go_arboretum_CHC   (738) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
     Ar_thaliana_CHC   (739) ----------------------------------PEIHFKYIEAAAKTG-QIKEVERVTRES-NFYDAEKTKNFLMEAK
   Me_truncatula_CHC   (735) ----------------------------------PDIHFKYIEAAAKTG-QIKEVERVYRES-SFYDPEKTKNFLMEAK
    Lo_japonicus_CHC   (726) ----------------------------------PDIHFKYIEAAAKTG-QIKEVERVTRES-NFYDAEKAKNFLMEAK
          Gl_max_CHC   (738) ----------------------------------PDIHFKYIEAAAKTG-QIKEVERVTRES-SFYDPEKTKNFLMEAK
      Po_tremula_CHC   (738) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
      Ho_vulgare_CHC   (738) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
     Tr_aestivum_CHC   (738) ----------------------------------PDIHFKYIESAARTG-QIKEVERVTRES-NFYDAEKTKNFLMEAK
      So_bicolor_CHC   (729) ----------------------------------PDIHFKYIEAAARTG-QIKEVERVTRES-NFYDAEKTKNFLMEAK
       Or_sativa_CHC   (736) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
        Ze_mays_CHC    (739) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXX-XXXXXXXXXXXXXXXX
  Di_discoideum_CHC    (726) ----------------------------------PEVHFKYIEAAAKIN-QFKEVERMCRDS-NYYDPEKTRDFLKEAK
       Sc_pombe_CHC    (722) ----------------------------------PEVVYKYIQAACLMN-QFTEVERICRDN-NVYNPEKVKNLLKEAK
   Sa_cerevisiae_CHC   (731) ----------------------------------KDVVYKYIEAAARMK-QYREIERIVKDN-NVYDPERVKNFLKDAN
     Ca_albicans_CHC   (725) ----------------------------------PDVVFKYIQAAARMN-QTKEIERVVRDN-NVYNGEKVKNFLKEFK
    As_fumigatus_CHC   (727) ----------------------------------PEVHFKYIEAATAMG-QITEVERICRES-NYYNPEKVKNFLKEAK
  Pa_brasilensis_CHC   (730) ----------------------------------XXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXX
        Le_major_CHC   (738) ----------------------------------PEVHFRYIEAAAEMG-QMQELERMTRES-PCYDAERTKNYLKSKQ
       Tr_brucei_CHC   (741) ----------------------------------PEVHFRYIEAAAEVG-QVQELERMTRES-PCYDPERTKNYLKNKK
  En_histolytica_CHC   (712) ----------------------------------KMVVFKYIEACARSG-QFQELARVVQEN-DVYDPKEVKEFLKNAK
 Gi_intestinalis_CHC   (788) ----------------------------------SSIHNFLLRTILQLDSKTLELTHIVAED-QYLDPNECFTIMTTTT
     Th_annulata_CHC   (770) ----------------------------------XEIIIPYMKLCIEYN-DIIELENILLYN-NNFNLFDAKELIKNSQ
       To_gondii_CHC   (781) ----------------------------------PEVHFKYIEAAAKLN-HTQEVERVCRES-KCYEPQRVKEFLKVKE
       Cr_parvum_CHC   (798) ----------------------------------XXXHFKYIEASVHLG-QIQEAERICRDFPQSYEPEQVIEYFKSIK
   Pl_falciparum_CHC   (894) YNNMNDEDASILLTSDICSTPQYSYETANLKLEDVHYIMFKYIEACVKIN-NIQELDRICKDKNAKYNPEQIKNFLKDCK
        Pl_yoelii_CHC  (888) RSNNDDTSSILINSDIGIGTQFPYENNSVLTIEDLHHIMFKYIEACVKIN-NIQELDRICKDPNAKYNPEQIKNFLKECK
        ANCESTRAL_CHC  (990) YNNMDDEDAILLLSDDIISTPFYYYENANLKLEDLPDVHFKYIQAACKTGSQIKEVERICRESNNCYDPERVKNFLKEAK
            Consensus  (1041)                                  P VHFKYI AA   G QIKEVERI RES N YD E  KNFLKEAK
```

Table 4.3 continued

```
Bo_taurus_CHC17      (768) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------XXXXXXXXXXXXXXX-XXXXXXXXXX---------
Su_scrofa_CHC17      (789) XXXXXXX----XXXXXXXXXXXXXXXXXXXXX-------------NLQKYIEIYVQKVNP-SRLPVVIGGL---------
Mu_musculus_CHC17    (768) LTDQLP-----IIVCDRFDFVHDLVLYLYRN-------------XXXXXXXXXXXXXXX-XXXXXXXXXX---------
Ra_norvegicus_CHC17  (768) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NLQKYIEIYVQKVNP-SRLPVVIGGL---------
Ga_gallus_CHC17      (768) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------SLQKYIEIYVQKVNP-SRLPVVIGGL---------
Da_rerio_CHC17       (769) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NLQKYIEIYVQKVNP-SRLPVVIGGL---------
Ta_rubripes_CHC17    (768) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NLQKYIEIYVQKVNP-SRLPVVVGGL---------
Ho_sapiens_CHC22     (768) LTDQLPL----IIVCDRFGFVHDLVLYLYRN-------------NLQRYIEIYVQKVNP-SRTPAVIGGL---------
Ga_gallus_CHC22      (767) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Ta_rubripes_CHC22    (768) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NLQKYIEIYVQKVNP-SRLPVVIGGL---------
Ci_intestinalis_CHC  (769) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NIPKYIEIYVQKVNP-SRLPVVVGGL---------
Dr_melanogaster_CHC  (769) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------NLQKYIEIYVQKVNP-SRLPVVVGGL---------
Ae_aegyptii_CHC      (769) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------SLQKYIEIYVQKVNP-SRLPVVVGGL---------
An_gambiae_CHC       (769) LTDQLPL----IIVCDRFDFVHDLVLYLYRN-------------SLQKYIEIYVQKVNP-SRLPVVVGGL---------
Ca_elegans_CHC       (770) LNDQLPL----IIVCDRHNMVHDLVLYLYRN-------------QLQKYIEVFVQKVNA-ARLPIVVGAL---------
Ca_briggsae_CHC      (771) LNDQLPL----IIVCDRHNMVHDLVLYLYRN-------------QLQKYIEVFVQKVNA-ARLPIVVGAL---------
Br_malayi_CHC        (770) LTDQLPL----IIVCDRHDMVHDLVLYLYRN-------------NLQKYIEVFVQKVNP-ARLPIVVGGL---------
Pi_taeda_CHC         (781) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Ly_esculentum_CHC    (781) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Be_vulgaris_CHC      (781) LPDARPL----INVCDRFGYVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-ANAPLVVGQL---------
Me_crystallinum_CHC  (779) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Go_arboretum_CHC     (781) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Ar_thaliana_CHC      (782) LPDARPL----INVCDRFSFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Me_truncatula_CHC    (778) LPDARPL----INVCDRFGFVPDLTHYLYTS-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Lo_japonicus_CHC     (769) LPDARPL----INVCDRFGFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Gl_max_CHC           (781) LPDARPL----INVCDRFGFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Po_tremula_CHC       (781) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Ho_vulgare_CHC       (781) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Tr_aestivum_CHC      (781) LPDARPL----INVCDRFGFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
So_bicolor_CHC       (772) LPDARPL----INVCDRFGFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Or_sativa_CHC        (779) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Ze_mays_CHC          (782) XPDARPL----INVCDRFGFVPDLTHYLYTN-------------NMLPYIEGYVQKVNP-GNAPLVVGQL---------
Di_discoideum_CHC    (769) LPDQLPL----IIVCDRYEFISDLTNYLYKN-------------NLSKYIEAYVQKINP-VNTPLVVGAL---------
Sc_pombe_CHC         (765) LADQLPL----ILVCDRYDFVNDLVFYLFRN-------------NMFQFIEIYVQRINP-SKTPQVVGAL---------
Sa_cerevisiae_CHC    (774) LEDQLPL----VIVCDRFDFVHEMILYLYKS-------------QNLKFIETYVQQVNP-SKTAQVVGAL---------
Ca_albicans_CHC      (768) LDDQLPL----IIVCDRFNFVHDLILYLYKN-------------QYFKFIEVYVQSVNP-ANTPQVVAGL---------
As_fumigatus_CHC     (770) LTEQLPL----IIVCDRFNFIHDLVLYLYQN-------------QQYKSIEVYVQRVNP-SRTPAVVGGL---------
Pa_brasilensis_CHC   (774) XXXXXXX----XXXXXXXXXXXXXXXXXXX-------------XXXXXXXXXXXXXXX-XXXXXXXXX----------
Le_major_CHC         (781) LTDMWPF----INVCDQHNMVNEMVHYLVET-------------GNESFIEQYVTRRNP-GKTPQVVQAL---------
Tr_brucei_CHC        (784) MTNLWPL----INVCDQHNFVDELIRYLIDT-------------NNEALIEQYVQRRNP-LKTPAVVGAL---------
En_histolytica_CHC   (755) IADQIPF----IIVCDKHGYVEEMTQYLYQN-------------GNHRFIEAYVKKINP-MNTPKVVGAL---------
Gi_intestinalis_CHC  (832) PASPDAFLQAFLTFCERFHLGKKLGLFLTDSITNKVLGKDTFFTLNAATIVEAFIGGHGV-AQTHDLLFGLFEG------
Th_annulata_CHC      (813) LSTTKPF----IIICYRLNNIQELLLFLFNYYDTVLAQPDTN---STNNLTTSTIGTVVPGIKDTTTVGASTVTGTVKEE
Tg_gondii_CHC        (824) LPDPRPL----IYVCDLHGYVTELAEYLFKN-------------SLLKYIEVYVSRVNS-ANAPLVIGTL---------
Cr_parvum_CHC        (842) MSDLRPL----IWVCDLHHRVEELISYLYHM-------------SLYKYIQVYTLKINP-SQTPLVIGTL---------
Pl_falciparum_CHC    (973) LSDPRPL----IYVCDIHNYIEELAEYLYKN-------------SLLKYIEVYVIKVNP-NNAHKVIGVL---------
Pl_yoelii_CHC        (967) LSDPRPL----IYVCDIHNFIEELAEYLYKN-------------SLLKYIEVYVIKVNP-HNAHKVIGVL---------
ANCESTRAL_CHC       (1070) LTDQLPLLQAFIIVCDRFDFVHDLVLYLYRNITNKVLGKDTFFTLNLQKYIEIYVQKVNP-SRLPVVIGGLFEGGVLAVP
Consensus           (1121) L D  PL    I VCDRF FV DL  YLY N                  YIE YVQKVNP    P VVG L
```

Table 4.3 continued

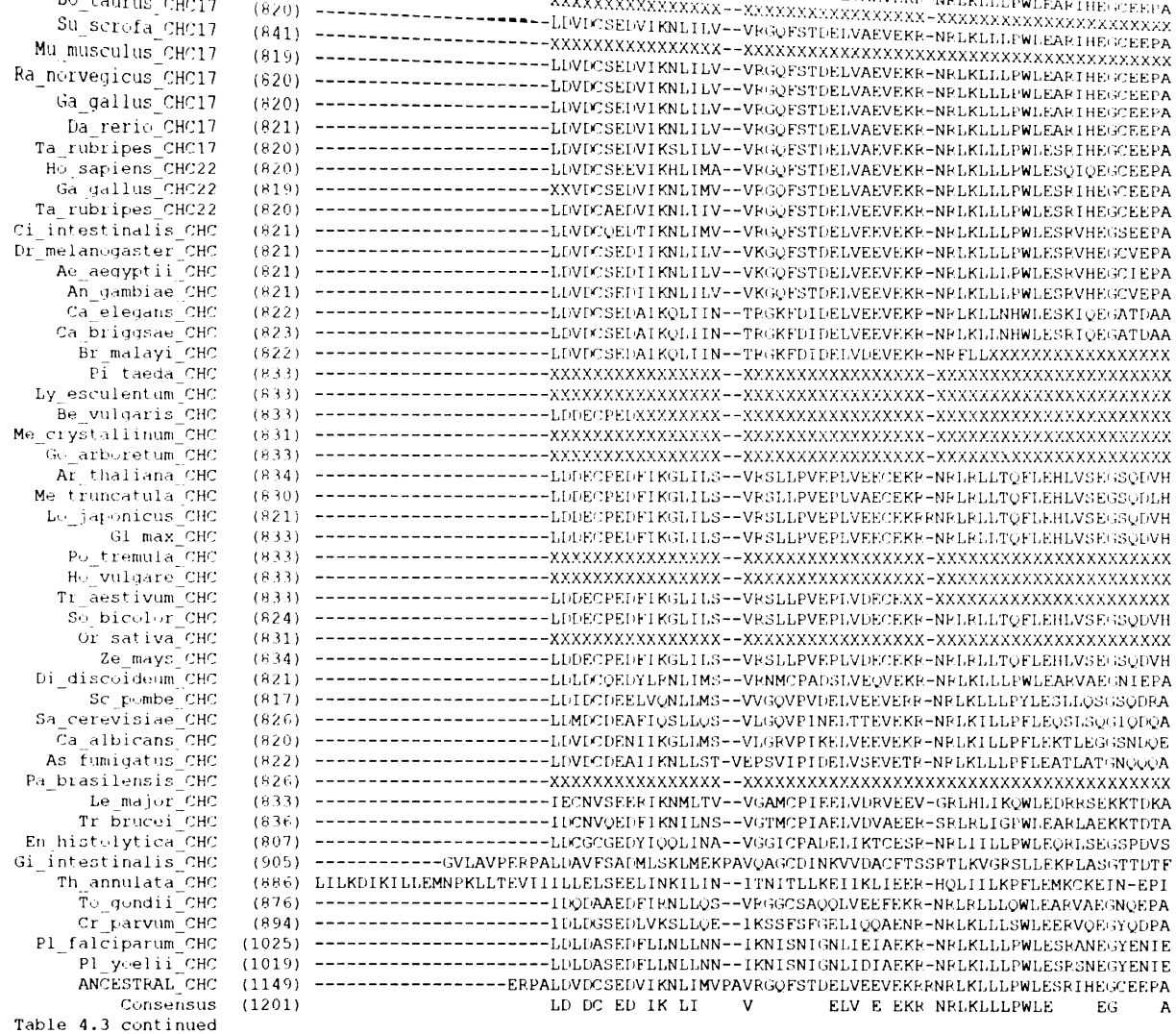


Table 4.3 continued

```
          Bo_taurus_CHC17  (875) THNALAKIYIDS-NNNPERFLRENPYYDSRVVGKY------XXXXXXXXXXXXXX----------GQCD
          Su_scrofa_CHC17  (897) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----CEKRDPHLACVAYER----------XXXX
        Mu_musculus_CHC17  (874) THNALAKIYIDS-NNNPERFLRENPYYDSRVVGKY-----CEKRDPHLACVAYER----------XXXX
      Ra_norvegicus_CHC17  (875) THNALAKIYIDS-NNNPERFLRENPYYDSRVVGKY-----CEKRDPHLACVAYER----------GQCD
         Ga_gallus_CHC17   (875) THNALAKKYIDT-NNNPERFLRENPYYDSRVVGKY-----CEKRDPHLACVAYER----------GQCD
          Da_rerio_CHC17   (876) THNALAKIYIDS-NNNPERFLRENPYYDSRVVGKY-----CEKRDPHLACVAYER----------GQCD
      Ta_rubripes_CHC17    (875) THNALAKIYIDS-NNNPERFLRENPFYDSRVVGKY-----CEKRDPHLSCVAYER----------GQCD
        Ho_sapiens_CHC22   (875) THNALAKIYIDS-NNSPECFLRENAYYDSSVVGRY-----CEKRDPHLACVAYER----------GQCD
         Ga_gallus_CHC22   (874) THNALAKIYIDS-NNNPERFLRENPYYDSRVVGKY-----CEKRDPHLACVAYER----------GQCD
      Ta_rubripes_CHC22    (875) THNALAKIYIDS-NNTPERFLKENPFYDSAVVGKY-----CEKRDPHLACVAYER----------GQCD
    Ci_intestinalis_CHC    (876) THNALAKIYIDS-NNNPERFLRENPFYDSRVVGKY-----CEKRDPHMACVCYER----------GQCD
    Dr_melanogaster_CHC    (876) THNALAKIYIDS-NNNPERYLKENQYYDSRVVGRY-----CEKRDPHLACVAYER----------GLCD
        Ae_aegyptii_CHC    (876) THNALAKIYIDS-NNNPERFLKENQFYDSCVVGRY-----CEKRDPHLACVAYER----------GHCD
         An_gambiae_CHC    (876) THNALAKIYIDS-NNNPERFLKENQFYDSRVVGRY-----CEKRDPHLACVAYER----------GQCD
          Ca_elegans_CHC   (877) THNAMAKIYIDS-NNNPERFLKENPYYDSKVVGKY-----CEKRDPHYAFLSYER----------GQCD
        Ca_briggsae_CHC    (878) THNAMAKIYIDS-NNNPERFLKENPYYDSKVVGKY-----CEKRDPHYAFLSYER----------GQCD
          Br_malayi_CHC    (877) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
           Pi_taeda_CHC    (888) XXXXXXXXXXXX-XXXXEHFLTTNPYYDSYVVGKH-----CEKRNPTLAVVLYWR----------GQCD
        Ly_esculentum_CHC  (888) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
          Be_vulgaris_CHC  (888) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
      Me_crystallinum_CHC  (886) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXX----------XXXX
         Go_arboretum_CHC  (888) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
          Ar_thaliana_CHC  (889) VHNALGKIIIDS-NNNPEHFLTTNPYYDSKVVGKY-----CEKRDPTLAVVAYRR----------GQCD
        Me_truncatula_CHC  (885) VCNTPGKIIIDI-NNNLEHFVTTNAYYDSPXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXX
          Lo_japonicus_CHC (877) VHNALGKIIIES-NNNPEHFLTTNPYYDSRVVGKY-----CEKRDPTLAVVAYRR----------GQCD
               Gl_max_CHC  (888) VHNALGKIIIDS-NNNPEHFLTTNPYYDSRVVGKY-----CEKRDPTLAVVAYRR----------GLCD
            Po_tremula_CHC (888) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
            Ho_vulgare_CHC (888) XXXXXXXXXXXX-XXXPEHFLTTNPFYDSRVVGKY-----CEKRDPTLAVVAYRR----------GQCD
           Tr_aestivum_CHC (888) XXXXXXXXXXXX-XXXXEHFLTTNPFYDSRVVGKY-----REKRDPTLAVVAYRR----------GQCD
            So_bicolor_CHC (879) VHNALGKIIIDS-NNNPEHFLTTNPFYDSRVVGKY-----CEKRDPTLAVVAYRR----------GQCD
              Or_sativa_CHC (886) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
                Ze_mays_CHC (889) VHNALGKIIIDS-NNNPEHFLTTNPFYDSRVVGKY-----CEKRDPTLAVVAYRR----------GQCD
          Di_discoideum_CHC (876) IHNALAKIYIDS-NKNPEAFLLHNQFYDSKVVGKY-----CEKRDPYLSFVAYKR----------GLCD
               Sc_pombe_CHC (872) IYDALAKIYIDS-NNNPEVFLKENNFYDTLTVGKY-----CEKRDPYLAFIAYEK----------GGND
        Sa_cerevisiae_CHC   (881) VYNALAKIYIDS-NNSPEKFLKENDQYDTLDVGHY-----CEKRDPYLAYIAYEK----------GQND
          Ca_albicans_CHC   (875) VYNTLAKIYIDS-NNSPEKFLQENNNYDTLVVGKY-----CEKRDPYLAYISYSK----------GGND
         As_fumigatus_CHC   (878) VYNALAKIYIDS-NNNPEKFLKENDMYDTLTVGKY-----CEKRDPNLAYIAYRK----------GQND
       Pa_brasilensis_CHC   (882) XXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXX-----XXXXXXXXXXXXXXX----------XXXX
            Le_major_CHC    (888) LYNALAKIYVDI-GQSPEQFLSENEYYDAQVVGKY-----CETRDPNLAYIIYCC----------GHCS
          Tr_brucei_CHC     (891) LHNAIAKLLVIS-GNLPEKFIEENDYYDPVVMGKY-----CEDRDPNLSYLVYRK----------AKMS
       En_histolytica_CHC   (862) LHTALAKVYVDQ-GKDAEKFLKYDMYYDSLIVGKY-----CENRDPYLSICYKRK----------NLD
    Gi_intestinalis_CHC     (973) VHTAFAKLIMEEPNGNAKRLLMEDPYIDHALIGTFILESPTHFTDDFLVDLAMTSLTIGAFGPTTAGADFSDHIINGECA
          Th_annulata_CHC   (962) LHTALMKIYITL-GENPEDYLINNKYYLKEEIANY-----CQDIDLQLSLLIYKH----------NSMY
            To_gondii_CHC   (931) VHNALAKIYIXX-XXXXESFLKTNAYYDSLVVGKY-----CEERDPHLAFTAYKR----------AWGACD
            Cr_parvum_CHC   (949) LHNALAKIYIDM-NKDSENFLKTNPHYDAKTIGDY-----CEDRDPYLAYIAYRK----------GQCD
         Pl_falciparum_CHC (1080) LHNALAKIYIDL-NKDPENFLKNNNFYDKKMIGKY-----CEDLDPHLAYTAYERS----------NGECD
            Pl_yoelii_CHC  (1074) LHNALAKIYIDL-NKDPETFLKNNNFYDKKMIGKY-----CEDLDPHLAYTAYERS----------NGQCD
            ANCESTRAL_CHC  (1211) THNALAKIYIDSKNNTPERFLRENPYYDSAVVGKYILESPCEKRDPHLACVAYERSTIGAFGPTTAGADFSDHIINGQCD
                Consensus  (1281)  HNALAKIYIDS NNNPE FL  N  YDS VVGKY     CEKRDP LA  AY R                GQCD
```

Table 4.3 continued

```
                Bo_taurus_CHC17  (928) LELINVCNENSLFKSLSRYLVRRKDPELWGSVLL----------ESNPYRRPLIDQVVQ-TALSETQDP
                Su_scrofa_CHC17  (951) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------ESNPYRRPLIDQVVQ-TALSETQDP
              Mu_musculus_CHC17  (927) LELINVCNENSLFKSLSRYLVRRKDPELWGSVLL----------ESNPYRRPLIDQVVQ-TALSETQDP
            Ra_norvegicus_CHC17  (928) LELINVCNENSLFKSLSRYLVRRKDPELWGSVLL----------ESNPYRRPLIDQVVQ-TALSETQDP
                Ga_gallus_CHC17  (928) LELINVCNENSLFKSLSRYLVRRKDPELWASVLL----------ESNPYRRPLIDQVVQ-TALSETQDP
                 Da_rerio_CHC17  (929) QELIHVCNENSLFKSLSRYFVRRKDPELWASVLL----------EVNPYRRPLIDQVVQ-TALSETQDP
              Ta_rubripes_CHC17  (928) QELINVCNENSLFKSLSRYLVRRKDPELWASVLL----------ESNPFRRPLIDQVVQ-TALSETQDP
                Ho_sapiens_CHC22 (928) LELIKVCNENSLFKSEARYLVCRKDPELWAHVLE----------ETNPSRRQLIDQVVQ-TALSETRDP
                Ga_gallus_CHC22  (927) LELIKVCNENSLFKSEARYLVRRKDPELWANVLE----------ENNPFRRQLIDQVVQ-TALSETQDP
              Ta_rubripes_CHC22  (928) LDLIKVCNENSLFKSEARYLVHRKDPELWANVLE----------ENNPFRRQLIDQVVQ-TALSETQDP
            Ci_intestinalis_CHC  (929) DELIRVCNENSLFKSEARYLVRRKDVELWGKVLE----------DTNRYRRQLIDQVNVQTALGEAQDP
            Dr_melanogaster_CHC  (929) RELIAVCNENSLFKSEARYLVGRRDAELWAEVLS----------ESNPYKRQLIDQVVQ-TALSETQDP
                Ae_aegyptii_CHC  (929) RELIAVCNENSLFKSEARYLVRRRDAELWADVLS----------ESNPYKRQLIDQVVQ-TALSETQDP
                An_gambiae_CHC   (929) RELIAVCNENSLFKSEARYLVRRRDAELWAEVLS----------EANPYKRQLIDQVVQ-TALSETQDP
                Ca_elegans_CHC   (930) AELINVCNENSLFKNLARYLVKRRDFTLWEQVLN----------EENVHRRQLIDQVVQ-TALSETQDP
                Ca_briggsae_CHC  (931) AELINVCNENSLFKNLARYLVKRRDFTLWEQVLN----------EENVYRRQLIDQVVQ-TALSETQDP
                  Br_malayi_CHC  (930) XXXXXVCNENSLFKNLARYLVKRRDFPLWAQVLA----------EDNQHRRQLIDQVVQ-TALSETQDP
                   Pi_taeda_CHC  (941) DELIDVSNKNCLFELQASYVVERADPQLWKKVLA----------LENPLRRKLIDQVVS-TTLPESKNL
              Ly_esculentum_CHC  (941) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
                Be_vulgaris_CHC  (941) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
            Me_crystallinum_CHC  (938) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
               Go_arboretum_CHC  (941) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
                Ar_thaliana_CHC  (942) EELINVTNKNSLFKLQARYVVERMDGDLWEKVLT----------EENEYRRQLIDQVVS-TALPESKSP
              Me_truncatula_CHC  (948) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXLIDQVVS-TAFPESKXX
                Lo_japonicus_CHC (930) DELINVTSKNSLFKLQARYVVERMDGDLWAKVLT----------PDNEYRRQLIDQVVS-TALPESSSP
                     Gl_max_CHC  (941) DELINVTNKNSLFKLQARYVVERMDGDLWEKVLN----------PDNAYRRQLIDQVVS-TALPESKSP
                  Po_tremula_CHC (941) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
                  Ho_vulgare_CHC (941) EELINVTNKNSLFKLQARYVVERMDGDLWDKVLQ----------PDENRRQFIDQVVS-TALPESKSP
                 Tr_aestivum_CHC (941) EELINVTNKNSLFKLQARYVVERMDGDLWDKVLQ----------PDNEYRRQFIDQVVS-TALPESKSP
                  Se_bicolor_CHC (932) DELINVTNKNSLFKLQARYVVERMDGDLWDKVLQ----------PENEYRRQLIDQVVS-TALPESKSP
                   Or_sativa_CHC (939) XXXXXXXXXXXXXXXXARYVVERMDGDLWDKVLQ----------PENEYRRQLIDQVVS-TALPESKSP
                     Ze_mays_CHC (942) DELINVTNXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXXXXXXXXXXX
              Di_discoideum_CHC  (929) YELIEVTNKNTLFKNQARYLVERQDPDLWAYVLS----------DQNEYKRSLIDQVVQ-TALPESTNA
                  Sc_pombe_CHC   (925) TEIINLCNENSMFKQLARYLLKRSDSNLWSEVLQ----------DS-AYRRPLLDQVIA-TAVPESSDP
               Sa_cerevisiae_CHC (934) DDLIRITNENSMYKYQARYLLERSDLDLWNKVLN----------QENIHRRQLIDSVIS-VGIPELTDP
                 Ca_albicans_CHC (928) DQLISITNENKMYKYQARYLLSKSDIELWNKVLS----------SDNIHRRQLVDQVIS-TGIPELDDP
               As_fumigatus_CHC  (931) LELINITNENAMYPAQARYLVERADPEIWSFVLS----------ENNLHRRPLVIQVIA-TAVPESTEP
             Pa_brasilensis_CHC  (935) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXXXXXXXX-XXXXXXXXX
                  Le_major_CHC  (941) EMMIDLCHRNGMYKQLARYLVKEQSLDLWASVLG----------QNSAERQQLIESVQQ-IALPESESS
                Tr_brucei_CHC   (944) TELVEITSKNGMWKQLARYLVKQQDPALWASVLT----------ESSINRDRLVEAVQQ-TALPESEVT
             En_histolytica_CHC (914) DELIDVTNRHNLFKSKQIILLKKXXXXXXXXVLT----------EENEHRKSVVEQVIH-TALPECTDP
            Gi_intestinalis_CHC (1053) RMLVLVAYAGSAYKELAERLLLHKCSELWQLALVGVMNEDPASDVIKAITQGGMRRPMSMHRDMLLQVLAHSPPIPLDTLS
                Th_annulata_CHC (1015) EHFIKFSLKHQLYKSLYQFLLKESNXXXXXXXXX----------XXXXXXXXXXX-XXXXXXXXX
                  To_gondii_CHC  (986) EQLVXLTNKNGLFRLQARYLVERQAPELWAYVLR----------EENPHRRQAIDQVVS-SALPESSSA
                  Cr_parvum_CHC (1002) DEIIQVTYKNDLYRLLARYLVERQDLDLWNKVLGNSSNS----------IEPVVNEQCRQAIIDQVTS-SILPEFYNK
             Pl_falciparum_CHC (1135) EELIYITSKNGLFKLEARYLISRQSIELWKVVLD----------ETNKYRKNVIIDQVIG-STLIESNNA
                  Pl_yoelii_CHC (1129) EELINITSKNGLFKLQAKYLVSRQSMGLWSMVLD----------ESNKYRKNVIIDQVIG-STLIESNNA
                ANCESTRAL_CHC   (1291) LELIKVCNENSLFKSEARYLVHRKDPELWANVLEGVMNEDPASDVIKAITQGGMRENNPFRRQLIDQVVQQTALSETQDP
                    Consensus   (1361)  ELI V N NSLFK  ARYLV R D  LW  VL                          N  RR LIDQVV  TAL E   P
```

Table 4.3 continued

```
Ma_fascicular_CHC17  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYINRLDN-YDAPDIA
    Bo_taurus_CHC17  (960) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXX-XXXXXXX
    Su_scrofa_CHC17  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYINRLDN-YDAPDIA
  Mu_musculus_CHC17 (1010) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Ra_norvegicus_CHC17  (985) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYINRLDN-YDAPDIA
    Ga_gallus_CHC17  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYINRLDN-YDAPDIA
     Da_rerio_CHC17  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRYLQELLILTGIKG-DRTRVMGFINRLDN-YDAPDIA
     Da_rerio_CHC17  (987) EE----VSVTVKAF--MTADLPNELIELLEKIVLDH-SVFSEHRNLQNLLILKAIKA-DRTPVLEYINRLDN-YDAPDIA
  Ta_rubripes_CHC17  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYISRLDN-YDAPDIA
    Ho_sapiens_CHC22 (986) EE----ISVTVKAF--MTADLPNELIELLEKIILDS-SVFSDHRNLQNLLILTAIKA-DRTRVMDYINRLEN-YDAPDIA
    Ga_gallus_CHC22  (985) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTGIKA-DRTRVMEYINRLDN-YDAPDIA
  Ta_rubripes_CHC22  (986) EE----VSVTVKAF--MTADLPNELIELLEKIVLDN-SVFSEHRNLQNLLILTAIKA-DRTRVMEYINRLDN-YDAPDIA
Ci_intestinalis_CHC  (988) DE----ISVAVKAF--MAADLPNELIELLEKIVLDN-SMFSDHRNLQNLLILTAIKA-DSSRVMDYITQLDN-YDAPDIA
Dr_melanogaster_CHC  (987) DD----ISVTVKAF--MTADLPNELIELLEKIILDS-SVFSDHRNLQNLLILTAIKA-DRTRVMDYINRLEN-YDAPDIA
    Ae_aeuyptii_CHC  (987) DD----ISVTVKAF--MTADLPNELIELLEKIVLDS-SVFSDHRNKQNLLILTAIKA-DRSRVMDYINRLDN-YDAPDIA
     An_gambiae_CHC  (987) DD----ISVTVKAF--MTADLPNELIELLEKIVLDS-SVFSDHRNLQNLLILTAIKA-DRSRVMDYINRLDN-YDAPDIA
     Ca_elegans_CHC  (988) ED----ISVTVKAF--MAADLPNELIELLEKIVLDN-SAFSEHRNLQNLLILTAMRA-DRTRVMEYIQKLDN-YDAPDIA
    Ca_briggsae_CHC  (989) ED----ISVTVKAF--MAADLPNELIELLEKIVLDN-SAFSEHRNLQNLLILTAMRA-DRTRVMEYIQKLDN-YDAPDIA
      Br_malayi_CHC  (988) ED----ISATVKAF--MAADLPNELIELLEKIVLDN-SVFSEHRXXXXXXXXXXXXX-XXXXXXXXXXXXX-XDISDIT
       Pi_taeda_CHC  (999) HQ----IYATVKAF--MDADRPHELIELLEKIVLQN-STLSGSANLQYLLILXXXXX-XXXXXXXXXXXXX-XXXXXXX
  Ly_esculentum_CHC  (999) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
    Be_vulgaris_CHC  (999) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXLDN-LDGPAVG
Me_crystallinum_CHC  (996) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
   Go_arboretum_CHC  (999) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXMDTINRLDN-FDGPAVG
     Ar_thaliana_CHC (1000) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYINRLDN-FDGPAVG
  Me_truncatula_CHC (1009) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
    Lo_japonicus_CHC  (988) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYINRLDN-FDGPAVG
          Gl_max_CHC  (999) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYINRLDN-FDGPAVG
      Po_tremula_CHC  (999) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXLDN-FDGPAIG
      Ho_vulgare_CHC  (998) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DSSRVMDYVNRLDN-FDGPAVG
     Tr_aestivum_CHC  (999) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYVNRLDN-FDGPAVG
     So_bicolor_CHC  (990) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYVNRLDN-FDGPAVG
      Or_sativa_CHC  (997) EQ----VSAAVKAF--MTADLPHELIELLEKIVLQN-SAFSGNFNLQNLLILTAIKA-DPSRVMDYVNRLDN-FDGPAVG
       Ze_mays_CHC (1001) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
  Di_discoideum_CHC  (987) TE----VSATVKAF--MDANLPNELIELLEKIVLEG-KEFKTAKELQNLLILTAIRA-DKSRVTDYINRLDN-FDGSKLA
     Sc_pombe_CHC  (982) EA----VSIVVKAL--MEVDLPSQLIELLEKIVLQP-SSFSENANLQNLLFLTAIKA-DKSRVMEYIDKLDK-YDVDEIA
   Sa_cerevisiae_CHC  (992) EP----VSLTVQAF--MTNGLKLELIELLEKIILEP-SPFNENVALQGLLLLSAIKY-EPTKVSSYIEKLDN-YDADEIA
    Ca_albicans_CHC  (986) EP----ISITVKAF--MENDLPEEIMELLEKIILEP-SPFSENASLQGIMILTAIKA-DSSKVSSYIEKLDK-FDPNEIA
    As_fumigatus_CHC  (989) DK----VSVAVKAF--LDADLPGELIELLEKIILEP-SPFSDNGSLQNLIMLTAAKA-DKGRLMEYIHQLNE-FSADEIA
  Pa_brasilensis_CHC  (993) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
      Le_major_CHC  (999) EE----VSTTVRAF--MNANLTEELTSLLGQIVVHG--RFRKNRFLENLLIMSAVRA-RKDKVMEYVTTLED-YDAKDIA
     Tr_brucei_CHC (1002) EE----VSTTVKAF--MDAELTEELTSILDQIVVRG--RFRKNRYLENLLIMSAVRA-SKAKVMEYVTTLDS-YDAKEIA
  En_histolytica_CHC  (972) NQ----ILKTVQAF--LQNDLPNDLIDLLDKLVSQN-PEYSTNPTLQKLLIQTSIKA-APNRVMDYIRKLND-YDPKEIG
  Gi_intestinalis_CHC (1133) DAQIFVLSSSILAIKNESDAIPRYLQAILEKIITSPGSVHASNPGMQNLLIVTALNTKDYPKVSALIRSAIMCYDQDVIA
     Th_annulata_CHC (1070) XX----XXXXXXXX--XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXX-XXXXXXX
       To_gondii_CHC (1044) DE----VSAAVNAF--INAQLPHELIELLEKIVLHN-SDFSNNKNLQNLLILTAMKA-DSSRIIDYVNRLHN-YDGAAIA
      Cr_parvum_CHC (1069) SEE---ISCVIRAF--INAEVPNSLLEVLEKIIVNN-TEFSQNKNLQNLLLLTSIKI-DVRRLDDYVLRLDN-YDAKEVA
  Pl_falciparum_CHC (1193) DE----ITVTVKAF--IEKKLSSELIELLEKIVLHN-SEFSDNKNLQNLLILTAIKS-DSKKVMEYINRLDN-FSGPQIA
       Pl_yoelii_CHC (1187) DE----ITVTVKAF--IEKKLSSELIELLEKIVLHN-SEFRDNKNLQNLLILTAIKS-DSKKVMEYINRLDN-FSGPQIA
       ANCESTRAL_CHC (1371) EEQIFVVSVTVKAFKNMTADLPNELIELLEKIVLDNGSVFSEHRNLQNLLILTAIKAKDRTRVMEYINRLDNCYDAPDIA
           Consensus (1441) E      S TVKAF  M ADLP ELIELLEKIVL N S FS   NLQNLLILTAIKA D  RVM YI RLDN YD P IA
```

Table 4.3 continued

```
Md_fascicular_CHC17 (1057) NIAISNE-------LFEEAFAIFRKFDVNTSAV---QVLIEHIGN-----LDRA----------------------------
          Bo_taurus_CHC17 (1031) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
        Su_scrofa_CHC17 (1057) NIAISNE-------LFEEAFAIFRKFDVNTSAV---QVLIEHIGN-----LDRA---------------------------- 1600
      Mu_musculus_CHC17 (1084) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
     Ra_norvegicus_CHC17 (1056) NIAISNE-------LFEEAFAIFRKFDVNTSAV---QVLIEHIGN-----LDRA----------------------------
        Ga_gallus_CHC17 (1057) NIAISNE-------LFEEAFAIFRKFDVNTSAV---QVLIEHIGN-----LDRA----------------------------
         Da_rerio_CHC17 (1057) NIAIRNE-------LFDEAFAIFRKFDVNTSPV---QVLIEHIGN-----LDRA----------------------------
      Ta_rubripes_CHC17 (1058) NIAISNE-------LFEEAFAIFRKFDVNTSAVQ-VQVLIEHIGN-----LDRA----------------------------
      Ta_rubripes_CHC17 (1057) NIAISNE-------LFEEAFAIFKKFDVNTSAV---QVLIEHIGN-----LDRA----------------------------
       Ho_sapiens_CHC22 (1057) SIAVSSA-------LYEEAFTVFHKFDMNASAI---QVLIEHIGN-----LDRA----------------------------
        Ga_gallus_CHC22 (1056) NIAISNE-------LYEEALAIXXXXXXXXXXX----XXXXXXXX-----XXXX----------------------------
      Ta_rubripes_CHC22 (1057) NIAISNE-------LFEEAFAIFKKFDVNTSAI---QVLIEHIGN-----LDRA----------------------------
    Ci_intestinalis_CHC (1059) NICIENE-------LFEEAFAIFNKFDVNTSAI---QVLIEHVKN-----LDRA----------------------------
    Dr_melanogaster_CHC (1058) NIAISNQ-------LYEEAFAIFKKFDVNTSAI---QVLIDQVNN-----LERA----------------------------
       Ae_aegyptii_CHC (1058) NIAINNE-------KYEEAFAIFKKFDVNTSAI---QVLIEQVHN-----LERA----------------------------
        An_gambiae_CHC (1058) NIAINNE-------LYEEAFAIFKKFDVNTSAI---QVLIEQVNN-----LERA----------------------------
        Ca_elegans_CHC (1059) NIAITSE-------LYEEAFAIFKKFDVNSSAI---NVLIENVNN-----LDRA----------------------------
       Ca_briggsae_CHC (1060) NIAITSE-------LYEEAFAIFKKFDVNSSAI---NVLIENVNN-----LDRA----------------------------
         Br_malayi_CHC (1059) SYILIVK-------LVRTIFWIGRKFDINIAAM---QVLIDNVSN-----LDRA----------------------------
          Pi_taeda_CHC (1070) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
      Ly_esculentum_CHC (1070) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
        Be_vulgaris_CHC (1070) EVAVEAQ-------LFEEAFAIFKKFNLNVQAV---NVLLDNIQS-----IDRA----------------------------
    Me_crystallinum_CHC (1067) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
       Go_arboretum_CHC (1070) EVAVEAQ-------LYEEAFAIFKKFNLNVQAV---NVLLDNIRS-----IDRA----------------------------
       Ar_thaliana_CHC (1071) EVAVEAQ-------LYEEAFAIFKKFNLNVQAV---NVLLDNVRS-----IERA----------------------------
      Me_truncatula_CHC (1080) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
       Lo_japonicus_CHC (1059) EMAVEGQ-------LYEEAFAIFKKFNLNVQAV---DVLIENIHS-----IDRA----------------------------
            Gl_max_CHC (1070) EMAVEAQ-------LYEEAFAIFKKFNLNVQAV---NVLLDHIRS-----IDRA----------------------------
         Po_tremula_CHC (1070) EVAVEAQ-------LYEEAFAIFKKFNLNFSAV---NVLLDNIRS-----IDRA----------------------------
         Ho_vulgare_CHC (1069) EVAVEAQ-------LYEEAFAIFKKFNLNVQAV---DVLLDNIPS-----IERA----------------------------
       Tr_aestivum_CHC (1070) EVAVEAQ-------LYEEAFAIFKKFNLNVQAV---DVLLDNIRS-----IERA----------------------------
        So_bicolor_CHC (1061) EVAVEAQ-------LYEEAFAIFKKFNLNVQAV---NVLLDNIRS-----IERA----------------------------
          Or_sativa_CHC (1068) EVAVEAQ-------LFEEAFAIFKKFNLNVQAV---NVLLDNIRS-----IERA----------------------------
            Ze_mays_CHC (1072) XXXXXSQ-------LYEEAFAIFKKFNLNVQAV---DVLLDNIRS-----IERA----------------------------
      Di_discoideum_CHC (1058) PIAIESQ-------LYEEAFFMYKKFQFNVEAI---DVLITHIGS-----IERA----------------------------
          Sc_pombe_CHC (1053) EIAIENG-------LYEEAFRIYKIHNKHEQAM---KVLVEDIVS-----LDRA----------------------------
      Sa_cerevisiae_CHC (1063) PLCIEHD-------LKEEAFEIYDKHEMYGKAL---KVLIEDIMS-----LDRA----------------------------
       Ca_albicans_CHC (1057) PLCIDNG-------LNEEAFEVYDKFELRSEAM---KVLVEDIVS-----LDRA----------------------------
     As_fumigatus_CHC (1060) EMCISVG-------LYEEAFEIYKKVNNYIAAV---NVLVENIVS-----IDRA----------------------------
   Pa_brasilensis_CHC (1064) XXXXXXX-------XXXXXXXXXXXXXXXXXXX---XXXXXXXXX-----XXXX----------------------------
         Le_major_CHC (1069) HVASGAG-------MHEIAFVVYTRHNMQKEAM---TVLLQDMND-----ISRG----------------------------
        Tr_brucei_CHC (1072) GIATAAE-------LHEEAMTVYDKFEMRMEAA---TVLLPDLKD-----LPRG----------------------------
   En_histolytica_CHC (1043) ESCIADN-------LFEEAFEVYTKFDLQEEAM---DVLLQHIKD-----LTRA----------------------------
  Gi_intestinalis_CHC (1213) KACVETAKKNEYPPLYEDAFEVFHRAGRPIDAI---RVLLDNIS------IDRG----------------------------
      Th_annulata_CHC (1141) XXXXXXX-------XXXXXXXXXXXXXXXXXXX----XXXXXXXX-----XXXX----------------------------
        To_gondii_CHC (1115) QVAMEYG-------LREEAFTIYKKFGLFGEAA---DTLLKSAEEGSEADLERA----------------------------
        Cr_parvum_CHC (1141) KVAIDHG-------LYSHAFQIYKKFSFNNEAVET-LLMISKLEL-----LKNEGIEIDGQ--------------------
   Pl_falciparum_CHC (1264) AVAYEYK-------LREEAFVIYKKFNCYTSAI---SVLLDKILY-----NKNQKNKNFHEFSSHDSTVPYYSNNNNKDN
        Pl_yoelii_CHC (1258) EVAYEYK-------LREEAFVIYKKFNCNTSAI---SVLLDNILT-----FNKNRKGKMSANAPRKKIELALFNTIINDY
        ANCESTRAL_CHC (1451) NIAISNEKKNEYPPLREEAFAIFKKFDVNTSAISLQQVLIEHIGN-----LDRANKKKFSEFSRRDKIVLYYFNNIIKDY
             Consensus (1521)   A           L EEAFAIFKKF  N  A     VL   I         RA
```

Table 4.3 continued

```
                                1601
    Ho_sapiens_CHC17  (1096) ----------------------------------------------------------------------------------------------------
Ma_fascicular_CHC17   (1070) ----------------------------------------------------------------------------------------------------
     Bo_taurus_CHC17  (1096) ----------------------------------------------------------------------------------------------------
     Su_scrofa_CHC17  (1123) ----------------------------------------------------------------------------------------------------
   Mu_musculus_CHC17  (1095) ----------------------------------------------------------------------------------------------------
Ra_norvegicus_CHC17   (1096) ----------------------------------------------------------------------------------------------------
     Ga_gallus_CHC17  (1096) ----------------------------------------------------------------------------------------------------
      Da_rerio_CHC17  (1099) ----------------------------------------------------------------------------------------------------
   Ta_rubripes_CHC17  (1096) ----------------------------------------------------------------------------------------------------
    Ho_sapiens_CHC22  (1096) ----------------------------------------------------------------------------------------------------
     Ga_gallus_CHC22  (1094) ----------------------------------------------------------------------------------------------------
   Ta_rubripes_CHC22  (1096) ----------------------------------------------------------------------------------------------------
Ci_intestinalis_CHC   (1098) ----------------------------------------------------------------------------------------------------
Dr_melanogaster_CHC   (1097) ----------------------------------------------------------------------------------------------------
     Ae_aegyptii_CHC  (1097) ----------------------------------------------------------------------------------------------------
      An_gambiae_CHC  (1097) ----------------------------------------------------------------------------------------------------
      Ca_elegans_CHC  (1098) ----------------------------------------------------------------------------------------------------
     Ca_briggsae_CHC  (1099) ----------------------------------------------------------------------------------------------------
       Br_malayi_CHC  (1098) ----------------------------------------------------------------------------------------------------
        Pi_taeda_CHC  (1109) ----------------------------------------------------------------------------------------------------
   Ly_esculentum_CHC  (1109) ----------------------------------------------------------------------------------------------------
     Be_vulgaris_CHC  (1109) ----------------------------------------------------------------------------------------------------
Me_crystallinum_CHC   (1106) ----------------------------------------------------------------------------------------------------
    Gb_arboretum_CHC  (1109) ----------------------------------------------------------------------------------------------------
     Ar_thaliana_CHC  (1110) ----------------------------------------------------------------------------------------------------
   Me_truncatula_CHC  (1119) ----------------------------------------------------------------------------------------------------
    Lo_japonicus_CHC  (1098) ----------------------------------------------------------------------------------------------------
          Gl_max_CHC  (1109) ----------------------------------------------------------------------------------------------------
       Po_tremula_CHC  (1109) ----------------------------------------------------------------------------------------------------
       Ho_vulgare_CHC  (1108) ----------------------------------------------------------------------------------------------------
      Tr_aestivum_CHC  (1109) ----------------------------------------------------------------------------------------------------
       So_bicolor_CHC  (1100) ----------------------------------------------------------------------------------------------------
        Or_sativa_CHC  (1107) ----------------------------------------------------------------------------------------------------
          Ze_mays_CHC  (1111) ----------------------------------------------------------------------------------------------------
  Di_discoideum_CHC    (1097) ----------------------------------------------------------------------------------------------------
       Sc_pombe_CHC    (1092) ----------------------------------------------------------------------------------------------------
  Sa_cerevisiae_CHC    (1102) ----------------------------------------------------------------------------------------------------
     Ca_albicans_CHC   (1096) ----------------------------------------------------------------------------------------------------
    As_fumigatus_CHC   (1099) ----------------------------------------------------------------------------------------------------
 Pa_brasilensis_CHC    (1103) ----------------------------------------------------------------------------------------------------
        Le_major_CHC   (1108) ----------------------------------------------------------------------------------------------------
       Tr_brucei_CHC   (1111) ----------------------------------------------------------------------------------------------------
 En_histolytica_CHC    (1082) ----------------------------------------------------------------------------------------------------
Gi_intestinalis_CHC    (1258) ----------------------------------------------------------------------------------------------------
     Th_annulata_CHC   (1179) ----------------------------------------------------------------------------------------------------
       To_gondii_CHC   (1159) ----------------------------------------------------------------------------------------------------
       Cr_parvum_CHC   (1189) ----------------------------------------------------GS--E----------FMKNLDSMISIEKLAN
  Pl_falciparum_CHC    (1329) IFGNKSDNENDEYLKQFTQNNSYREKYGSYYNNNDISNSNNNNIHLHPEFNNEDKFSSDTLHNYEDIDSTESTCDLSNYE
       Pl_yoelii_CHC   (1323) SGNDSDGYLNVTNFSGALSSSYNEKPDGASSIYGSNEVFRNYGSGQLSEFEDEN-------NEEQEMEKDETHPNIENYE
       ANCESTRAL_CHC   (1526) IFGDKDDYLNDEYLKGFLSNNYYREKYGSYYIYNDIENFNNYNIGLLSEFNDEDKFSSDTLNNYEDIDKDESTCDLENYE
           Consensus   (1601)
```

Table 4.3 continued

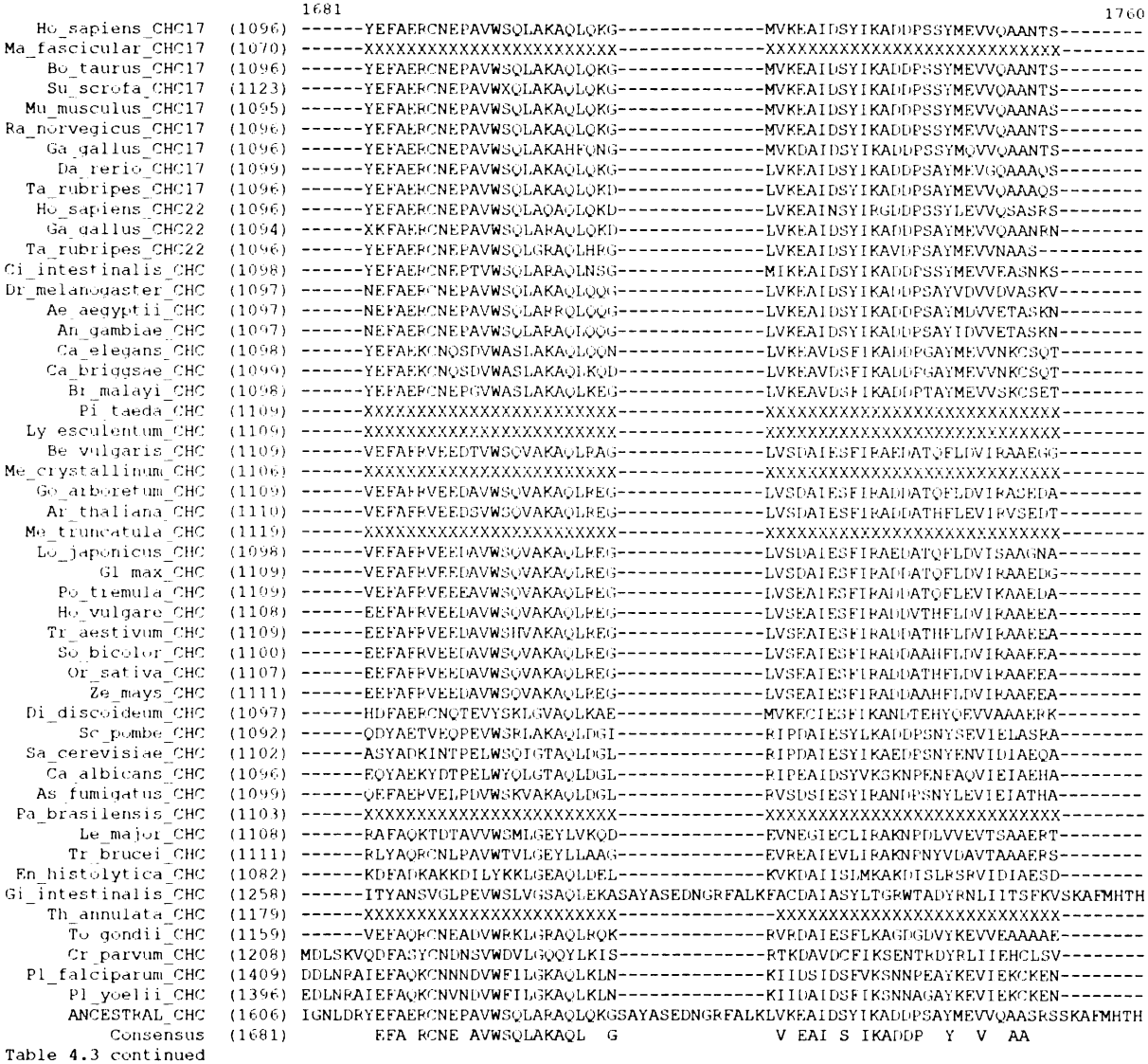


Table 4.3 continued

```
                                   1761                                                                                1840
      Ho_sapiens_CHC17   (1148) -----GNWEELVKYLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
  Ma_fascicular_CHC17    (1122) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
       Bo_taurus_CHC17   (1148) -----GNWEELVKYLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
      Su_scrofa_CHC17    (1175) -----GNWEELVKYLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
    Mu_musculus_CHC17    (1147) -----GNWEELVKYLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
  Ra_norvegicus_CHC17    (1148) -----GNWEELVKYLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
      Ga_gallus_CHC17    (1148) -----GNWEDFVKYFQMARKK--AR-----ESYVETELIFVLAKTNRLAELEEFING---------PNNAHIQQVG----
       Da_rerio_CHC17    (1151) -----GNWEDLVKFLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVGDRVG
    Ta_rubripes_CHC17    (1148) -----GNWEDLVKFLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFING---------PNNAHIQQVG----
     Ho_sapiens_CHC22    (1148) -----NNWEDLVKFLQMARKK--GR-----ESYIETELIFALAKTSRVSELEDFING---------PNNAHIQQVG----
      Ga_gallus_CHC22    (1146) -----DNWEDLVKFLQMARKK--AR-----ESYVETELIFAFAKTNRLSELEEFISG---------PNNAHIQQVG----
    Ta_rubripes_CHC22    (1146) ------NWEDLVKFLQMARKK--AR-----ESYVETELIFALAKTNRLAELEEFVSG---------PNNAHIQQVG----
  Ci_intestinalis_CHC    (1150) -----GNWEDLVRYLQMARKK--AR-----ESYVETELVFAFAKTNRLADLEEFISG---------PNNANIQQVG----
  Dr_melanogaster_CHC    (1149) -----ESWDDLVRYLQMARKK--AR-----ESYIESELIYAYARTGRLADLEEFISG---------PNHADIQKIG----
      Ae_aegyptii_CHC    (1149) -----ESWEDLVRYLQMAGKK--AR-----ESYIESELIYAYARTGRLADLEEFVSG---------PNHADIQKIG----
      An_gambiae_CHC     (1149) -----DSWEDLVRYLQMARKK--AR-----ESYIESELIYAYARTGRLADLEEFVSG---------PNHADIQKIG----
      Ca_elegans_CHC     (1150) -----EHWEDLVRYLQMARKK--SR-----ESYIETELVFALAKTGRLTELEEFIAG---------PNHAQIGQIG----
     Ca_briggsae_CHC     (1151) -----EHWEDLVRYLQMARKK--SR-----ESYIETELVYALAKTGRLTELEEFIAG---------PNHAQIGQIG----
       Br_malayi_CHC     (1150) -----XNWEDLVRYLQMARKK--SR-----ESYIETELVYAYAKTNRLADLEEFISG---------PNHAQIQQIG----
        Pi_taeda_CHC     (1161) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
   Ly_esculentum_CHC     (1161) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
      Be_vulgaris_CHC    (1161) -----DVYHDLVKYLLMVRQQ--TK-----EPRVDSELIYAYAKIDRLSDIEEFILM---------PNVANLQNVG----
 Me_crystallinum_CHC     (1158) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
     Go_arboretum_CHC    (1161) -----DVYPDLVRYLLMVRQK--VK-----EPKVDGELIYAYAKIDRLGEIEEFILM---------PNVANLQNVG----
     Ar_thaliana_CHC     (1162) -----DVYDDLVKYLLMVRQK--VK-----EPKVDSELIYAYAKIDRLGEIEEFILM---------PNVANLQHVG----
   Me_truncatula_CHC     (1171) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
    Lo_japonicus_CHC     (1150) -----NVYHDLVKYLLMVRQK--AK-----EPKVDSELIYAYAKIDRLSDIEEFILM---------PNVANLQNVG----
          Gl_max_CHC     (1161) -----NAYHDLVRYLLMVRHK--TK-----EPKVDSELIYAYAKIDRLSDIEEFILM---------PNVANLQNVG----
      Po_tremula_CHC     (1161) -----DVYHDLVRYLIMVRQK--SX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
      Ho_vulgare_CHC     (1160) -----NVYHDLVKYXXXXXXX--X------EPKVDSELIFAYAKIDRLNDIEEFILM---------PNVANLQNVG----
    Tr_aestivum_CHC      (1161) -----DVYHDLVKYLLMVRQK--AR-----EPKVDGEPIFAYAKIDRLSDIEEFILM---------PNVANLQNVG----
      So_bicolor_CHC     (1152) -----NVYDDLVKYLLMVRQK--AR-----EPKVDGELIFAYAKIDRLSDIEEFILM---------PNVANLQNVG----
       Or_sativa_CHC     (1159) -----NVYDDLVKYLLMVRQK--AR-----EPKVDGELIFAYAKIDRLSDIEEFILM---------PNVANLQNVG----
         Ze_mays_CHC     (1163) -----NVYNDLVKYLLMVRQK--AR-----EPKVDGELIFAYAKIDRLSDIEEFILM---------PNVANLQNVG----
   Di_discoideum_CHC     (1149) -----DEYEDLVKFLQMCRKK--TK-----EPAIESELIFAYAKVNKLAEMEDFINS---------PNSAHIQVVG----
        Sc_pombe_CHC     (1144) -----GKYEELIKYLLMARSK--MH-----EPDVDSALLIAYAKTNQLTEMETFLIG---------SNVADVKAVG----
  Sa_cerevisiae_CHC      (1154) -----GKYEELIPFLLMARKT--LK-----EPKIDGALIIAYAELNKIHEIENLLAG---------SNVANLDHVG----
     Ca_albicans_CHC     (1148) -----GKEEELITFLIMARET--LR-----EPVVDGALINAYATLDRLSDMEKFVGG---------SNVADLEAIG----
    As_fumigatus_CHC     (1151) -----GKDEELVKYLKMARKT--LR-----EPAIDTALAFCYARLDQLPELEDFLRT---------TNVADVETSG----
 Pa_brasilensis_CHC      (1155) -----XXXXXXXXXXXXXXX--XX-----XXXXXXXXXXXXXXXXXXXXXXXXXXX---------XXXXXXXXXX----
        Le_major_CHC     (1160) -----NQFGDLIKYLTMARQCSRAK-----DNKIDTALVITYAKTGRLEELEEFLKQ---------THNVKIGAIA----
       Tr_brucei_CHC     (1163) -----NQFGDLVKYLNMAPQESTSN-----DNKIESVLLLTYARTGRLSELEELLQN---------THNVQIQPVA----
  En_histolytica_CHC     (1134) -----GSYEDLINYLLMCKEE--TK-----DMMVETELLYCYAKLKKNDEIENFLKT---------ANCANLTSIA----
 Gi_intestinalis_CHC     (1332) SVEVRSELAALLEYLRMARVVVAQKGETSSQAEIDTAVMYCLARLDDYGALSEFLESVNPRDSTVIPNRGGIKEVG----
     Th_annulata_CHC     (1230) -----XXXXXXXXXXXXXXX---XX--------XXXXXXXXXXXXXXXXXXXXXXXXX----------XXXXXXXXX----
       To_gondii_CHC     (1211) -----DAYDALVDFLLMARKKITVK-----DQVIDSELVYAYAKTDRLEEMDAFLSG---------TNTANVQAVG----
       Cr_parvum_CHC     (1266) -----KAYRELLGYLQMVRR---SK-----DPIVDTELAYCMSKLELLQDLQSFLQG---------INTVQLQKIG----
   Pl_falciparum_CHC     (1467) -----NFYEHLITYLNTLREQNSLK-----DVLVDSELLYAYAKLKKTNEMNKFINT---------TNSANLQLIG----
        Pl_yoelii_CHC    (1454) -----NFYEQLITYLNTLREQNSLK-----DVLVDSELLYAYAKLKKTLEMTKFIAS---------TNLANMQLIG----
       ANCESTRAL_CHC     (1686) SVEVRNNWEDLVKFLQMARKKSSARGETSSESYVETELIFALAKTNRLAELEEFINGVNPRDSTVIPNNAHIQQVG----
           Consensus     (1761)         EDLVKYL MAR K   R     E  V  ELI AYAK  RL E EEFI              PN A  Q VG
Table 4.3 continued
```

```
                                1841                                                                                  1920
      Ho_sapiens_CHC17  (1203)  DRCYDEKMYDAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
     Ma_fascicular_CHC17 (1177) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
       Bo_taurus_CHC17  (1203)  DRCYDEKMYDAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
      Su_scrofa_CHC17   (1230)  DRCYXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
     Mu_musculus_CHC17  (1202)  DRCYDEKMYDAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
   Ra_norvegicus_CHC17  (1203)  DRCYDEKMYDAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
       Ga_gallus_CHC17  (1203)  DRCYDEKMYEAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
        Da_rerio_CHC17  (1210)  DRCYDEKMYEAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAHMCGLHIVVHA
     Ta_rubripes_CHC17  (1203)  DRCYDEKMYEAAKLLYNNVSNFGRLASTLVHLG-EYQAAVDGARKANSTRTWK-EVCFACVDGKEFRLAQMCGLHIVVHA
      Ho_sapiens_CHC22  (1203)  DRCYEEGMYEAAKLLYSNVSNFARLASTLVHLG-EYQAAVDNSRKASSTRTWK-EVCFACMDGQEFRFAQLCGLHIVIHA
       Ga_gallus_CHC22  (1201)  DRCYEEGMYEAAKLLYNNVSNFARLASTLVHLG-EYQAAVDSGRKANSTRTWK-EVCFACVDGREFRLAQICGLQXXXXX
     Ta_rubripes_CHC22  (1200)  DRCYEEGMYEAAKLLYNNVSNFARLASTLVHLG-EYQAAVDSARKANSTRTWK-EVCFACVDGEEFRLAQICGLHIVIHA
   Ci_intestinalis_CHC  (1205)  DRCYDQKMYEAAKVLFNNISNFARLASTVVHLEGEYQTAVDSARKANSTRTWK-EVCFACVDGEEFRLAQICGLHIVVHA
   Dr_melanogaster_CHC  (1204)  NRCFSDGMYDAAKLLYNNVSNFARLAITLVYLK-EFQGAVDSARKANSTRTWK-EVCFACVDAEEFRLAQMCGLHIVVHA
       Ae_aegyptii_CHC  (1204)  DRCFNDKMYEAAKLLYNNVSNFARLAITLVHLK-EFQGAVDGARKANSTRTWK-EVCFACVDAEEFRLAQMCGLHIVVHA
       An_gambiae_CHC   (1204)  DRCFNDRMYEAAKLLYNNVSNFARLAITLVHLR-EFQGAVDGARKANSTRTWK-EVCFACVDAEEFRLAQMCGLHIVVHA
       Ca_elegans_CHC   (1205)  DRCFDNGMFDSAKILFNNVSNFAKLSVTLVRLG-EYQGAVDAARKANSTKTWK-QVCFSCVENGEFRLAQMCGLHIVVHA
      Ca_briggsae_CHC   (1206)  DRCFDNGMFDAAKILFNNVSNFAKLSVTLVRLG-EYQGAVDAARKANSTKTWK-QVCFSCVENGEFRLAQMCGLHIVVHA
        Br_malayi_CHC   (1205)  DRCFDSGMYEAAKILHNNISNFAKLSVTLVRLN-EFQGAVDAARKANSTKTWK-QVCFACVDNEEFRLAQMCGLHIVVHA
         Pi_taeda_CHC   (1216)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
    Ly_esculentum_CHC   (1216)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
      Be_vulgaris_CHC   (1216)  DRLFDEALYEAAKIIYAFISNWAKLAVTLVRLK-QXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
  Me_crystallinum_CHC   (1213)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
     Go_arboretum_CHC   (1216)  DRLFDEALYEAAKIIFAFISNWAKLAVTLVRLK-QFQGAVDAARKANSAKTWX-XXXXXXXXXXXXXXXXXXXXXXXXXX
      Ar_thaliana_CHC   (1217)  DRLYDEALYEAAKIIYAFISNWGKLAVTLVKLQ-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIIQV
    Me_truncatula_CHC   (1226)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXANSAKTWK-EVCFACVDEEEFRLAQICGLNIIIQV
     Lo_japonicus_CHC   (1205)  DRLYDEALYEAAKIIFAFISNWAKLAITLVKLK-QFQGAVD--RKANSSKTWK-EVCFACVDEEEFRLAQICGLNIIIQV
          Gl_max_CHC    (1216)  DRLYDEELYEAAKIIFAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIIQV
       Po_tremula_CHC   (1216)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
       Ho_vulgare_CHC   (1214)  DRLYDEELYEAAKIIYAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIVQV
      Tr_aestivum_CHC   (1216)  DRLYDEELYEAAKIIYAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIIQV
       So_bicolor_CHC   (1207)  DRLYDEELYEAAKIIYAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIVQV
        Or_sativa_CHC   (1214)  DRLYDEELYEAAKIIYAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIVQV
          Ze_mays_CHC   (1218)  DRLYDEELYEAAKIIYAFISNWAKLAVTLVKLK-QFQGAVDAARKANSAKTWK-EVCFACVDAEEFRLAQICGLNIIVQV
    Di_discoideum_CHC   (1204)  DRCFENGLYEAAKVLYTNISNFSRLTSCLVKLG-QYQAAVDAARKANSTKTWK-EVSAACIDAKEFRLAQVCGINIIVHG
         Sc_pombe_CHC   (1199)  DRCFESKNYEAAKLMYSSISNWSMLATTLVYLG-EYQGAVDCARKANSIKVWK-QVGTACIDKREFRLAQICGLNLIVHA
    Sa_cerevisiae_CHC   (1209)  DKLFENKEYKAARLCYSAVSNYSKLASTLVYLG-DYQAAVDTARKASNIKVWK-LVNDACIEKKEFKLAQICGLNLIVHA
      Ca_albicans_CHC   (1203)  DKLFEAKNYKGAKILYSNVSKYSKLATTLVYLG-DYQGAVDCARKASNTQVWK-QVNSACIENKEFRLAQICGLNLIIDA
     As_fumigatus_CHC   (1206)  DKAYAEGYHQAAKIFYTSISNWAKLATTLVHLE-DYQAAVECARKANSVKVWK-EVNQACVDKKEFRLAQICGLNLIVHA
   Pa_brasilensis_CHC   (1210)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX
         Le_major_CHC   (1217)  DKCFQDKLYESARVLYTVANNYARVASTEVMIN-NLPAAVEAAKKAKSIHAYK-EANLACIEAGDLKLAGVCAVPVVLKA
        Tr_brucei_CHC   (1220)  DKCFEDGYYDSARLLYSMSMNFHKLALTLVRMN-NLAEAVDAAQKAQSRSTWD-AVNHACIEANDVRLAAICAVPLVLQV
   En_histolytica_CHC   (1189)  ERCYNEELYGAAKILYTSLNNYIKLASCLLKLK-DYAGAVEAAKKANSTRTWK-EVTFACIDAKEFTLAQETGINILIAG
   Gi_intestinalis_CHC  (1408)  EKCFAEEHYKAAKYFFEFVGDWSRLSLTLVKLR-CLKEAVEAATRAADPECWK-AVAAECLEIRDFELAKSVFLNLVLVE
      Th_annulata_CHC   (1280)  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXX-XXXXXXXXXXDNKLLNRVGIELLNYP
        To_gondii_CHC   (1268)  DRLFAEQRYKVRRLLYASLPNYAKLASCFVRLE-DFAASVDAARKAKNPKTWK-EVAFAALSKGELKCAHAAALSLIVHP
        Cr_parvum_CHC   (1320)  DRLMDEQDYRYSIIFYQAIPNYSRLTSCYIQLG-EYNNALETAKKANSPKTWK-ELLQICMQIGESELAHQAGLNIIVYP
    Pl_falciparum_CHC   (1524)  DRLYKEEEYEVAKILYSNIPNNRKLTACYLKLK-EYALAIEAAKKAKSLKTWEKEVNFICVKYKQLKYAHTAGLQLIMHA
        Pl_yoelii_CHC   (1511)  DRLYKEKEYDAAKILYSSIPNNQKLTFCHLKLK-EYSLAIEAAKKTKSLKTWR-EVNLVCVKYKQLKFAHIAGLQLIMHA
        ANCESTRAL_CHC   (1762)  DRCYEEGMYEAAKLLYNNVSNFARLASTLVHLG-EYQAAVDSARKANSTRTWKKEVCFACVDGEEFRLAQICGLHIVIHA
            Consensus   (1841)  DRCY E  YEAAK LY   SN   LA TLV L    Q AVD ARKANS   TWK EVCFACVD  EFRLAQ CGL I
```

Table 4.3 continued

```
                                1921                                                                                2000
   Ho_sapiens_CHC17  (1281) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
Ma_fascicular_CHC17  (1255) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
     Bo_taurus_CHC17  (1281) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
     Su_scrofa_CHC17  (1308) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
   Mu_musculus_CHC17  (1280) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
Ra_norvegicus_CHC17  (1281) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
     Ga_gallus_CHC17  (1281) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
      Da_rerio_CHC17  (1288) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
   Ta_rubripes_CHC17  (1281) DELEELINYYQDRGYFEELITMLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
     Ho_sapiens_CHC22  (1281) DELEELMCYYQDRGYFEELILLLEAALG------LERAHMGMFTELAILYSKFK-------PQKMLEHLE--LFWSR-VN
     Ga_gallus_CHC22  (1279) XXXXXXISYYQDRGYFEELMALLEAALG------LERAHMGMFTXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
   Ta_rubripes_CHC22  (1278) DELEDLISYYQDRGYFEELIALLEAALG------LERAHMGMFTELAILYSKFK-------PQKMREHLE--LFWSR-VN
Ci_intestinalis_CHC  (1284) DELEELITXXXXXXXXXXXXMLEAALG------LERAHMGMFTELAILYSKFK-------PGKMREHLE--LFWSR-VN
Dr_melanogaster_CHC  (1282) DELEDLINYYQNRGYFDELIALLESALG------LERAHMGMFTELAILYSKFK-------PSKMREHLE--LFWSR-VN
     Ae_aegyptii_CHC  (1282) DELEDLITYYQDRGHFEELIGLLEAALG------LERAHMGMFTELAILYSKYK-------PAKMREHLE--LFWSR-VN
     An_gambiae_CHC  (1282) DELEDLINYYQDRGYFEELIGLLEAALG------LERAHMGMFTELAILYSKYK-------PAKMREHLE--LFWSR-VN
     Ca_elegans_CHC  (1283) DELEELINFYQDRGHFEELIALLEAALG------LERAHMGMFTELAILYSKYK-------PEKMREHLE--LFWSR-VN
     Ca_briggsae_CHC  (1284) DELEELINFYQDRGHFEELIALLEAALG------LERAHMGMFTELAILYSKYK-------PEKMREHLE--LFWSR-VN
      Br_malayi_CHC  (1283) DELEELINYYQDRGYFEELIGLLEAALG------LERAHMGMFTELAILYSKYK-------AEKMREHLE--LFWSR-VN
       Pi_taeda_CHC  (1294) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
   Ly_esculentum_CHC  (1294) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
     Be_vulgaris_CHC  (1294) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
Me_crystallinum_CHC  (1291) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
    Go_arboretum_CHC  (1294) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
     Ar_thaliana_CHC  (1295) DDLEEVSEYYQNRGCFNELISLMESGLG------LERAHMGIFTELGVLYARYR-------YEKLMEHIK--LFSTR-LN
   Me_truncatula_CHC  (1304) DDLEEVSEYYQNRGCFNELISLMESGLG------LERAHMGIFTELGVLYARYR-------PEKLMEHIK--LFATR-LN
     Lo_japonicus_CHC  (1281) XXXXXXXXYYQNRGHFNELISLMESGLG------LERAHMGIFTELGVLFARYR-------PEKLMEHIK--LFSTR-LN
          Gl_max_CHC  (1294) DDLEEVSEYYQNRGCFNELISLMESGLG------LERAHMGIFTELGVLYARYR-------HEKLMEHIK--LFATR-PN
       Po_tremula_CHC  (1294) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
       Ho_vulgare_CHC  (1292) DDLEEVSEYYQNRGCFSELISLMESGLG------LERAHMGIFTELGVLYARYR-------PEKLMEHIK--LFSTR-LN
     Tr_aestivum_CHC  (1294) DDLEEVSEYYQNRGCFNELISLMESGLG------LERAHMGIFTELGVLYARYR-------PEKLMEHIK--LFSTR-LN
     So_bicolor_CHC  (1285) DDLEEVSEYYQNRGCFNELIALMESGLG------LERAHMGIFTELGVLYARYR-------PEKLMEHIK--LFSTR-LN
      Or_sativa_CHC  (1292) DDLEEVSEYYQNRGCFNELISLMESGLG------LERAHMGIFTELGVLYARYR-------XXXXXXXXX--XXXXX-XX
       Ze_mays_CHC  (1296) DDLEEVSEYYQNRGCFSELIALMESGLG------LERAHMGIFTELGVLYARYR-------SEKLMEHIK--LFSTR-LN
   Di_discoideum_CHC  (1282) DELEELIPQYEDRGYFNELISLLESGLA------SERAHVGMFTELAILYSKYK-------EEKLMEHLK--LFYSR-LN
       Sc_pombe_CHC  (1277) EELPGLIPLYEERGYFEEVISLMEAGLG------LERAHMAFYTELAILYAKYK-------PERMMEHLK--LFWGR-LN
   Sa_cerevisiae_CHC  (1287) EELDELVERYESNGYFEELISLFEAGLG------LERAHMGMFTELAILYSKYE-------PDKTFEHLK--LFWSR-IN
     Ca_albicans_CHC  (1281) EELPELVKTYEYNGYFNELIALFENGLS------LERAHMGMFTELAILYAKYS-------PEKVMEHLK--LFWSR-IN
     As_fumigatus_CHC  (1284) EELQDLVRQYEPNGYFDELISVLEAGLG------LERAHMGMFTELGIALSKYH-------PDRVMEHLK--LFWSR-IN
  Pa_brasilensis_CHC  (1288) XXXXXXXXXXXXXXXXXXXXXXXXXXXX------XXXXXXXXXXXXXXXXXXXX-------XXXXXXXXX--XXXXX-XX
        Le_major_CHC  (1295) EEVSGMCNRYESRGLWEELFSVLRNASS------HQGAHMSIFTEMGVLLAKYR-------PEKLMEHVI--MYAKK-IN
      Tr_brucei_CHC  (1298) ESLQDVVNRYEAYGLYDELFAVLKSAST------NSGAHMGIFTEMGLQLAKYR-------PEKLLEHVH--MYSKK-IN
  En_histolytica_CHC  (1267) DEITELVYYYEKNELYDQVIELLEAGLK------IENVHVSMFTELAILYSKYK-------EEKLYDYLK--QYVAK-IQ
Gi_intestinalis_CHC  (1486) SELPSIVQYYEKYGFIEELLEVLEAGAEQPQAATYTSMETNIFTILAILYCKYMWIVKKTDPQRLQTYIK--THGNK-IH
     Th_annulata_CHC  (1358) EFLVSVVASYESMGLFDDLIELLRNTTK----------TVATSTELAICIAKYH-------PEELMEHLRNVAFESNSLN
       To_gondii_CHC  (1346) DHLDSLIERYEQLCLFKELIELLEQGLQ------GEPTHVGLYTELGVLYATYE-------SSKLMDYIR--QHSGK-VN
       Cr_parvum_CHC  (1398) DYCELVVSEYEKKGLTAELLTLLEGAIQN-----TDRANGSLFTELGILYAKYT-------PEKLMDYCS--SYSGR-IN
   Pl_falciparum_CHC  (1603) DHLDEIIKIYEKKKYINELMNLLENGLN------NERAHVGIYTELGILYAKYK-------PEKLMEFIR--NYTNK-MN
        Pl_yoelii_CHC  (1589) DHLDEIIKIYEKKKYINELLSLLENGLN------SERAHVGIYTELGMLYAKYK-------PEKLMEFIR--SYTNK-MN
       ANCESTRAL_CHC  (1841) DELEELISYYQDRGYFEELIALLEAALGNPQAATLERAHMGMFTELAILYSKFKWIVRKTDPQKMREHLE--LFWSR-VN
           Consensus  (1921) DELEE   YYQ RG F ELI LLE  LG      LERAHMG FTEL ILY KY        P K  EH    LF  R  N
```

Table 4.3 continued

```
                                2001                                                                                                2080
   Ho_sapiens_CHC17   (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYRAIQFYLEFKPLLLNDLL
Ma_fascicular_CHC17   (1319) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   Bo_taurus_CHC17    (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYRAIQFYLEFKPLLLNDLL
   Su_scrofa_CHC17    (1372) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  Mu_musculus_CHC17   (1344) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYKAIQFYLEFKPLLLNDLL
Ra_norvegicus_CHC17   (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYKAIQFYLEFKPLLLNDLL
   Ga_gallus_CHC17    (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYKAVQFYLEFKPLLLNDLL
    Da_rerio_CHC17    (1352) IPKVLRAAEQAHLWGELVFLYDKYEEYDNAIITMMNHPADAWKESQFKDIITKVVANVELYYRAVQFYLEFKPLLLNDLL
  Ta_rubripes_CHC17   (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMMNHPTDAWKEGQFKDIITKVAN-VELYYKAVQFYLEFKPLLLNDLL
   Ho_sapiens_CHC22   (1345) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAVLTMMSHPTEAWKEGQFKDIITKVAN-VELCYRALQFYLDYKPLLINDLL
   Ga_gallus_CHC22    (1343) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  Ta_rubripes_CHC22   (1342) IPKVLRAAEQSHLWAELVFLYDKYEEYDNAVLTMMSHPTEAWKEGLFKDIIPKVAN-VELYYKSLSFYLDYKPLLLNDLL
Ci_intestinalis_CHC   (1348) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAIVTMMAHPTDAWKEGQYKDIITKVAN-IELYYKSLQFYLDYKPMLINDLL
Dr_melanogaster_CHC   (1346) IPKVLRAAESAHLWGELVFLYDKYEEYDNAVLAMMAHPTEAWREGHFKDIITKVAN-IELYYKAIEFYLDFKPLLLNDML
    Ae_aegyptii_CHC   (1346) IPKVLRAAEQAHLWSELVFLYDKYEEYDNAVLAMMAHPTEAWREGHFKDIITKVAN-IELYYKAIQFYLDYKPLLLNDML
    An_gambiae_CHC    (1346) IPKVLRAAEQAHLWSELVFLYDKYEEYDNAVLAMMAHPSEAWREGHFKDIITKVAN-IELYYKAIQFYLDYKPLLLNDML
    Ca_elegans_CHC    (1347) IPKVLRAAEQAHLWSELVFLYDKYEEYDNAALTMMQHPTESWREQHFKEVIAKVAN-VELYYKAMQFYLDYKPLLLNDLL
   Ca_briggsae_CHC    (1348) IPKVLRAAEQAHLWSELVFLYDKYEEYDNAALTMMQHPTESWREQHFKEVIAKVAN-VELYYKAMQFYLDFKPLLLNDLL
     Br_malayi_CHC    (1347) IPKVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     Pi_taeda_CHC     (1358) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  Ly_esculentum_CHC   (1358) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Be_vulgaris_CHC   (1358) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Me_crystallinum_CHC   (1355) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Go_arboretum_CHC  (1358) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Ar_thaliana_CHC   (1359) IPKLIRACDEQQHWQELTYLYIQYDEFDNAATTVMNHSPEAWEHMQFKDIVAKVAN-VELYYRAVQFYLEFKPLLLNDLL
  Me_truncatula_CHC   (1368) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMXXX--XWDHMQFKDVAVKVAN-VELYYKAVHFYLEEHPDLINDIL
    Lo_japonicus_CHC  (1345) IPKLIRACDEQQHWTELTYLYIQYDEFDNAATTIMNHSPEAWDHMQFKDVIVKVGN-VELYYKAVHFXXXXXXXXXXXXX
          Gl_max_CHC  (1358) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMNHSPEAWDHMQLKDVVVKVAN-VELYYKAVHFYLQEHPDLINDVL
     Po_tremula_CHC   (1358) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTIMNHSPEAWDHMQFKDVVVKVAN-VELYYKAVHFYLQEHPDLINDLL
     Ho_vulgare_CHC   (1356) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMNHSPDAWDHMQFKDVCVKVAN-VEIYYKAVHFYLQEHPDLINDLL
    Tr_aestivum_CHC   (1358) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMNHSPDAWDHMQFKDVAVKVAN-VEIYYKAVHFYLQEHPDLINDLL
    So_bicolor_CHC    (1349) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMNHSPDAWDHMQFKDVCVKVAN-VELYYKAVHFYLQEHPDLINDML
     Or_sativa_CHC    (1356) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
       Ze_mays_CHC    (1360) IPKLIRACDEQQHWKELTYLYIQYDEFDNAATTIMNHSPDAWDHMQFKDVCVKVAN-VELYYKAVHFYLQEHPDLINDML
  Di_discoideum_CHC   (1346) VPKVIKACQANQQWPQLTYLYIHYDFHDNAIQTMINHSIEAWDHVLFKETIPKVAK-LDLYYSAISFYLEEQPLLINDLL
      Sc_pombe_CHC    (1341) MAKVIRACDQMHLWNEAVFLYVHDQSYDNAAAVMMEQP-EAFDHQSFKDIIVHVAN-LELYYRALNFYLEQHPMLLTDLL
  Sa_cerevisiae_CHC   (1351) IPKVIRAVEQAHLWSELVFLYAHYDEWDNAALTLIEKSTKDLDHAYFKEVVVKVSN-LEIYYKAINFYVKFHPSLLVDLL
    Ca_albicans_CHC   (1345) IPKVLTACEDAHLYPELIFLYCHYEEWDNAALTMIEKSEVAFDHSSFKEIIVKAPN-LEIHYKAIQFYMNENPSLLVDLL
   As_fumigatus_CHC   (1348) IPKMIRACEEANLWPELVFLYCHYDEWDNAALAMMERAADAWEHHSFKDIIVKVAN-LEIYYRALNFYLQEQPLLLTDLL
  Pa_brasilensis_CHC  (1352) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      Le_major_CHC    (1359) THKMITVCEQYHHWVVLRVLHTNNEDWLAATNSMMQHHADAFDHEIFKDAVSHLGA-SELLYTAIGFYLKTHPDQLNDFL
     Tr_brucei_CHC    (1362) AHKLISVCEEYHHWLALRVLHVGNEDWLAAAKTMMCHFADAFDHDVFKDVASHLGA-SDFVYNAISFYVNTCPQNLCDFL
  En_histolytica_CHC  (1331) CQKVIPTVNMNQQWKELVFLYVQVDQXXXAIETMISYPDDCFDHQLMKELLVNVPR-IDMIYKAESYYLAEKPEKVNEML
 Gi_intestinalis_CHC  (1563) IPTLLHWTRETRLWGEYAYLLAASRDFDKAVVEMIAHPPSSFNHDVMKKVIGRVSN-PELVAQVVSFYIEYSPEYLCDFL
    Th_annulata_CHC   (1421) ISKTAPECSNLWLWREAVFLYT-IDDSDKAILSMILHP-ECFEEQLFFRTLANVSN-TEVIYKALYFYIQQYPTAVNKLL
     To_gondii_CHC    (1410) IPRLIRACERQSLWKEAVYLHMNYDEYEQAANCLIMHP-AAWSHELFVQVSXXXXX-XXXVYRAISFYSEYHPLQLCLLL
     Cr_parvum_CHC    (1463) IPKLTRICEQRQLWNEVVYLYLQYQEFDQAVLTVISHPKEAWKNDQFLSILQNVTN-VDILYKSMTFYLQEHPELLNSLL
  Pl_falciparum_CHC   (1667) TRKLIDVCENEYLLKEAVYLYISYDEYNLAVDTIIKHSPTAYTADTFMQVIHKVTN-SDIIHKVIDFYIEEHPLNLYNLL
     Pl_yoelii_CHC    (1653) TRKLIDVCHNEYLLKEAVYLYISYDEYNLAVDTIIKHSPIAYQPDIFMQVIHKVTN-SDIIHKVIDFYIEENPLNLYNLL
     ANCESTRAL_CHC    (1918) IPKVLRAAEQAHLWAELVFLYDKYEEYDNAVLTMMSHPTEAWKEGQFKDIITKVAN-VELYYKALQFYLDYKPLLLNDLL
           Consensus  (2001) IPK  RA     LW ELV LY  Y E DNA  TMM H   AW    FKD   KVAN  ELYYKA  FYL   P LLNDLL
```

Table 4.3 continued

```
                                  2081                                                                           2160
     Ho_sapiens_CHC17    (1424) MVLSP-------------RLDHTRAVNYFSKVKQLPLVKPYLRSVQNHNNKSVNESLNNLFITEEDYQALRTSIDAYDNF
  Ma_fascicular_CHC17    (1398) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      Bo_taurus_CHC17    (1424) MVLSP-------------RLDHTRAVNYFSKVKQLPLVKPYLRSVQNHNNKSVNESLNNLFITEEDYQALRTSIDAYDNF
      Su_scrofa_CHC17    (1451) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Mu_musculus_CHC17    (1423) MVLSP-------------RLDHTRAVNYFSKVKQLPLVKPYLRSVQNHNNKSVNESLNNLFITEEDYQALRTSIDAYDNF
  Ra_norvegicus_CHC17    (1424) MVLSP-------------RLAHTRAVNYFSKVKQLPLVKPYLRSVQNHNNKSVNESLNNLFITEEDYQALRTSIDAYDNF
      Ga_gallus_CHC17    (1424) MVLSP-------------RLDHTRAVTFFTKVKQLPLVKPYLRSVQNHNNKSVNESLNNLFIIEEDYQALRTSIDAYDNF
       Da_rerio_CHC17    (1432) IVLSP-------------RLDHTRAVNFFSKVKQLPLVKPYLRSVQNHNNKSVNEALNNLFITEEDFQALRTSIDAYDNF
    Ta_rubripes_CHC17    (1424) MVLSP-------------RLDHSRAVNFFSKVKQLPLVKPYLRSVQNHNNKSVNEALNNLFIIEEDYQALRTSIDAYDNF
     Ho_sapiens_CHC22    (1424) LVLSP-------------RLDHTWTVSFFSKAGQLPLVKPYLRSVQSHNNKSVNEALNHLLTEEEDYQGLRASIDAYDNF
      Ga_gallus_CHC22    (1422) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Ta_rubripes_CHC22    (1421) TILSP-------------RLDHSRAVSFFSKVNQLKLVKPYLRSVQNHNNKSVNEALNNLLTEEEDYQGLRASIDAYDNF
 Ci_intestinalis_CHC     (1427) NVLSP-------------RMDHTRAVAFFRKVQHLPLVKPYLRTVQTHNNKAINEALNDLLIEEEDYSGLRNSIDAYDNF
 Dr_melanogaster_CHC     (1425) LVLAP-------------RMDHTRAVSYFSKTGYLPLVKPYLRSVQSLNNKAINEALNGLLIDEEDYQGLRNSIDGFDNF
     Ae_aegyptii_CHC     (1425) LVLAP-------------RMDHTRSVNFFTKQGHLQLVKTYLRSVQSLNNKAINEALNALLIDEEDYQGLRTSIDAFDNF
      An_gambiae_CHC     (1425) LVLAP-------------RMDHTRAVSFFTKQGHLQLVKTYLRSVQSLNNKAINEALNGLLIDEEDYQGLRTSIDAFDNF
      Ca_elegans_CHC     (1426) TVLSP-------------RLDHSRTVLFFNKLKQIPLVKPYLRQVQNLNNKAINEALNQLLIDEEDHAGLRSSIEAQDNF
     Ca_briggsae_CHC     (1427) AVLSP-------------RLDHSRTVLFFNKLKQIPLVKPYLRQVQNLNNKAINEALNQLLIDEEDHAGLRSSIEAQDNF
       Br_malayi_CHC     (1426) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
        Pi_taeda_CHC     (1437) XXXXX-------------XXDHTRVVDIMRKAGHLHLVKPYMVAVQSNNVAAVNEALNAIYVEEEDYERLRESIDMHDNF
   Ly_esculentum_CHC     (1437) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXIELHDSF
     Be_vulgaris_CHC     (1437) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 Me_crystallinum_CHC     (1434) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    Go_arboretum_CHC     (1437) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     Ar_thaliana_CHC     (1438) IVLSP-------------RLDHTRAVDIMRKAGHLRLIKPYMIAVQSNNVSAVNEALNEIYVEEEDYDRLRESIDLHDSF
   Me_truncatula_CHC     (1445) NVLAL-------------RVDHTRVVDIMRKAGHLRLVKPYMVAVQSSNVSAVNEALNEIYVEEEDYDRL-ESIDLHDNF
     Lo_japonicus_CHC    (1424) XXXXX-------------XXXXXXXXXXXXXXXLRLVKPYMVAVQSNNVSAVNEALNEIYVEEEDYDRLRESIDLHDNF
          Gl_max_CHC     (1437) NVLAL-------------RVDQARVVDIMRKAGHLRLVKPYMVAVQSNNVSAVNEALNEIYVEEEDYDRLRESIDLHDNF
      Po_tremula_CHC     (1437) NVIAL-------------RVDHTRVVDIMRKAGQLRLVKPYMVAVQSNNVSAVNGALNGIYIEEEDYDRLRESIELHDNF
      Ho_vulgare_CHC     (1435) NVLAL-------------RLDHTRVVDIMRKAGQLHLVKPYMVAVQSNNVSAVNEALNELYVEEEDYERLRESVDMHDNF
     Tr_aestivum_CHC     (1437) NVLAL-------------RLDHTRVVDIMRKAGQLHLVKPYMVAVQSNNVSAVNEALNELYVEEEDYERLRESVDMHDNF
      So_bicolor_CHC     (1428) NVLAL-------------RLDHTRVVDIMRKAGQLHLVKPYMVAVQSNNVSAVNEALNELYVEEEDYERLRESVDMHDNF
       Or_sativa_CHC     (1435) XXXX--------------XXDHTRVVDIMRKAGQLHLVKPYMVAVQSNNVSAVNEALNELYVEEEDYERLRESVDMHDNF
         Ze_mays_CHC     (1439) NVLAL-------------RLDHTRVVDIMRKAGQLHLVKPYMVAVQSNNVSAVNEALNELYVEEEDYERLRESVDMHDNF
  Di_discoideum_CHC      (1425) SVLSP-------------RIDHTRAVTLIRSLGHLPLVKPYLVSAQDQNVAALNEALNELYVEEEDYESLRSSIDANSNF
        Sc_pombe_CHC     (1419) AALTP-------------RIDHPRVIRIFEKSENTPLILNFMVAIQHLNIQAVNHAYNDLLIEMEDYQSLQDSIENYDHF
  Sa_cerevisiae_CHC      (1430) TSLTP-------------RLDIPRTVKIFSKSDNLPLIKPFLINVLPKNNSVVNQAYHDLMIEEEDYKALQDAVDSYDKF
     Ca_albicans_CHC     (1424) KVLTP-------------KLDLPRVVRMFVQSDNLPMIKPFLISVLDKNNSVVNSAYHDLLIEEEDYKSLRSSIESNNRF
    As_fumigatus_CHC     (1427) QVLTP-------------RIDVNRVVRIFQSSDNIPLIKPFLLNVQSQNKRAVNDAINDLLIEEEDYKTLRDSVDNYDNF
 Pa_brasilensis_CHC      (1431) XXXXX-------------XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
        Le_major_CHC     (1438) SSIFK-------------RVDPERVMAETTKVAPIHVIRTYLEAAQERNAKKVNEALNNLYIEEDDFKALRHSVETYNNF
       Tr_brucei_CHC     (1441) TSMFK-------------VLDPDRVLREVKNVAPIHLILPYLESAQPRNSRLINDALNDLYVEEENFVALRNSVENYNNF
  En_histolytica_CHC     (1410) IAVAH-------------RCDHTQVISIARKEKDLKTIREYLLYCLDLNKDSVNEALIDLFIDDGDAKSLKSLIEKNTNF
 Gi_intestinalis_CHC     (1642) KACQLGVPNADGSPGIPAALDPIRVIAMVRSYDCVPMIKSYLELFVGAHNQEINETLIKIYLESYDDEGLRRLIAATTAY
     Th_annulata_CHC     (1498) SCVKF-------------KLDTGRVIKILRNNNCLQLAKEYFQ-QSDRNSQQINDTLYEIYVEEHDFESLEQDVSKYINY
       To_gondii_CHC     (1488) KSLDK-------------KLDHSRVVQHVRKAGHLAVVEKYLRETQHLNITAVNEAVNELLVEGEDVDGLRESILEYDNF
       Cr_parvum_CHC     (1542) MLTLK------------XXXXXXXXXXXXXXXXXISIIQEFLESISDENIQVVNEALINLYIEKVEVEKLMKLILTCDNY
   Pl_falciparum_CHC     (1746) KILEN-------------KIDNNRLVQTLKKSNNLPLIQKYLEDIQAQNITAVNETLNEIYLQNDDYISLRNSIDEYDNF
       Pl_yoelii_CHC     (1732) KILEN-------------KIDNNRLVQTMKKSNNLPLIQKYLEDIQTQNITSVNETLNEIYLENDDYISLRKSIDEYDNF
     ANCESTRAL_CHC       (1997) TVLSPGVPNADGSPGIPARLDHTRAVSFFSKVGQLPLVKPYLRSVQNHNNKSVNEALNNLLTEEEDYQGLRASIDAYDNF
           Consensus     (2081) VL                R DH R V    K   L LVKPYL  VQ  N    VNEALN L   EEEDY   LR SID  DNF
```

Table 4.3 continued

```
                                2161                                                                           2240
      Ho_sapiens_CHC17   (1491) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDSLYKDAMQYASESKDTELAEELL-----QWFL-QEEK
   Ma_fascicular_CHC17   (1465) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
       Bo_taurus_CHC17   (1491) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDSLYKDAMQYASESKDTELAEELL-----QWFL-QEEK
       Su_scrofa_CHC17   (1518) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
     Mu_musculus_CHC17   (1490) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDSLYKDAMQYASESKDTELAEELL-----QWFL-QEEK
   Ra_norvegicus_CHC17   (1491) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDSLYKDAMQYASESKDTELAEELL-----QWFL-QEEK
       Ga_gallus_CHC17   (1491) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDRLYKDAMQYASESKDTELAEELL-----QWFL-QENK
        Da_rerio_CHC17   (1499) DNISLAQS-LEKHELIEFRRIAAYLFKGNNRWKQSVELCKKDKLYKVDAMQYASESKDTELAEELL-----QWFL-EDDK
     Ta_rubripes_CHC17   (1491) DNISLAQR-LEKHELIEFRRIAAYLFKGNNRWKQSVELC-KKDKLYKDAMQYASESKDIELAEELL-----SWFL-QEDK
      Ho_sapiens_CHC22   (1491) DNISLAQQ-LEKHQLMEFRCIAAYLYKGNNWWAQSVELC-KKDHLYKDAMQHAAESRDAELAQKLL-----QWFL-EEGK
       Ga_gallus_CHC22   (1489) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
     Ta_rubripes_CHC22   (1488) DTIGLAQR-LEKHDLIEFRRIAAYLYKGNNRWRQSVELC-KKDKLYKDAMLYAAESKDAELAENLL-----QWFL-EEGR
    Ci_intestinalis_CHC  (1494) DTIALAQQRLEKHELIEFRRIAAYLYKGNNRWKQSVELC-KKDLLYKDAMTYAGESRDTELAEELL-----EWFL-ERQR
    Dr_melanogaster_CHC  (1492) DNIALAQK-LEKHELTEFRRIAAYLYKGNNRWKQSVELC-KKDKLYKDAMEYAAESCKQDIAEELL-----GWFL-ERDA
        Ae_aegyptii_CHC  (1492) DNIALAQK-LEKHELTEFRRIAAYLYKGNNRWKQSVELC-KKDRLFKDAMEYAAESHQGELAEELL-----GWFL-ERGA
        An_gambiae_CHC   (1492) DNIALAQK-LEKHELTEFRRIAAYLYKGNNRWKQSVELC-KKDRLFKDAMEYAAESHQGELAEELL-----GWFL-ERGA
        Ca_elegans_CHC   (1493) DNITLAQQ-LEKHPLVEFRRISAYLFKGNNRWKQSIELC-KKDKLYKDAMEYAAESRNGELAEELL-----SFFL-DEKL
       Ca_briggsae_CHC   (1494) DNISLAQQ-LEKHPLVEFRRISAYLFKGNNRWKQSIELC-KKDKLYKDAMEYAAESRNGELAEELL-----SFFL-DEKL
         Br_malayi_CHC   (1493) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXMEYAAESRQPEIAEELL-----SIFL-ENKL
          Pi_taeda_CHC   (1504) DQIGLAQK-VEKHELLEFRRIAAYIYKKAGRWRQSVALS-KKDNLYKDAMETSSQSGDRELAEELL-----VYFV-EQGK
     Ly_esculentum_CHC   (1504) DQIGLAQK-IEKHELLEMRRVAASIYKKAGRWKQSIALS-KKDNLYKDAMETASQSGDRELAEELL-----VYFI-EQGK
       Be_vulgaris_CHC   (1504) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
   Me_crystallinum_CHC   (1501) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
      Go_arboretum_CHC   (1504) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXL-----VYFI-EQGK
       Ar_thaliana_CHC   (1505) DQIGLAQK-IEKHELVEMRRVAAYIYKKAGRWKQSIALS-KKDNMYKDCMETASQSGEHELAEQLL-----VYFI-EQGK
     Me_truncatula_CHC   (1511) DQIGLAQK-IEKHELLEMRRVAAYIYKKAGRWKQSIALS-KKDNLYKDAMETASQSGERELAEELL-----VYFI-DQGK
      Lo_japonicus_CHC   (1489) DQIGLAQK-IEKHELLEMRRVAAYIYKKAGRWKQSIALS-KKDNLYKDAMETASQSGERELAEELL-----VYFL-MQGK
            Gl_max_CHC   (1504) DQIGLAQK-IETHELLEMRRVAAYIYKKAGRWKQSIELCHKKDTLYKDAMETVSQSGDRELVEELL-----VYFI-DQGK
        Po_tremula_CHC   (1504) DQIGLAQK-IEKHELLEMRRVAAYIYKKAGRWKQSIALS-KKDNLYKDCMETCSQSGERELSEELL-----VYFI-EQGK
       Ho_vulgare_CHC    (1502) DQIGLAQK-LEKHELLEMRRIAAYIYKKAGRWKQSIALS-KKDNMYKDCMETCSQSGDRELSEDLL-----VYFI-EQGK
       Tr_aestivum_CHC   (1504) DQIGLAQK-LEKHELLEMRRIAAYIYKKAGRWKQSIALS-KKDNMYKDCMETCSQSGDRELSEDLL-----VYFI-EQGK
        So_bicolor_CHC   (1495) DQIGLAQK-LEKHELLEMRRIAAYIYKKAGRWKQSIALS-KKDNMYKDCMETCSQSGDRELSEDLL-----VYFI-EQGK
         Or_sativa_CHC   (1501) DQIGLAQK-LEKHELLEMRRIAAYIYKKAGRWKQSIALS-KKDNMYKDCMETCSQSGDRELSEDLL-----VYFI-EQGK
           Ze_mays_CHC   (1506) DQICLAQK-LEKHELLEMRRIAAYIYKKAGRWKQSIALS-KKDNMYKDCMETCSQSGDRELSEDLL-----VYFI-EQGK
    Di_discoideum_CHC    (1492) GTIALAQK-LEKHELLEFRRIAAYLYKKNNRWAQSVELS-KKDKLYKDAIQSASDSKNPAIGEELL-----QYFV-DQQN
          Sc_pombe_CHC   (1486) DAIALARR-LEKHSLLEFRRIAAYIYRKNKRWTQSIELS-KQDRFYKDAIITARDSDQTTIAEDLM-----KYFV-EIGN
     Sa_cerevisiae_CHC   (1497) DQLGLASR-LESHKLIFFKKIGALLYRRNKKWAKSLSIL-KEEKLWKDAIETAAISQDPKVVEALL-----TYFV-ETGN
       Ca_albicans_CHC   (1491) NKLDLAER-LEKHDLIFFRQIAATLYTKEKKFNRAISIL-KTDKLWPDLLRTVAISKSKKIAHELL-----DYFV-ETGN
      As_fumigatus_CHC   (1494) DAVQLAQR-LEKHDLIFFRQIAANIYRNNKRWEKSISLS-KQDKLYKDAIETAAISGKSDVVEELL-----RYFV-DIGS
    Pa_brasilensis_CHC   (1498) XXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXX-XXXX
          Le_major_CHC   (1505) DSEELSSR-LEKIELFEFRKIALLLHRRNKRYSHAIAVA-KANELFEEAIDTAAESGDSELVNKLL-----DYFA-KDYP
         Tr_brucei_CHC   (1508) DSVELSAR-LEKMELFEFRKIALFLHRRHKAFDHALAVA-KENKLFQEAIDTAVESADPKVVEELL-----DFFA-VERP
   En_histolytica_CHC    (1477) NKTGLANR-LKVHEDVEFRKIAAYLYSSSDDYKSAIELC-KSEGFDDDAMEIAKKSKDEEKVTDLL-----NFFV-EQDK
   Gi_intestinalis_CHC   (1722) DAHALLDNLVSEDQTDDMRKLAVGLYGRLNRYDEGIKFG-LQNWFYEDAAECAADSGQIEHCEALLRLVCSSQFPTPKIR
       Th_annulata_CHC   (1564) DYMKLCTL-LEEHPLARMREIGAKILSRHNHYSRASIIYMKNNN-YIKAIDCARLSKSTQLVHETI-----NNLI-LKEE
         To_gondii_CHC   (1555) DQLALAQT-LENHPRVEMRRYAXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXXXRICFDLF-----RFFLADLED
         Cr_parvum_CHC   (1610) DQAKLSAR-LEKHPMNDLRKLAVKILDKNSNYQQALSIC-QKEMLIDEAILVVYKSGNVALIEELL-----EFLL-SNNK
     Pl_falciparum_CHC   (1813) NQTNLLNK-LENHKLPEMRRIAALLYKKNKKYKEAINLS-KKEKQYKDAIEIARVSKNNVYIEDLI-----NYFI-ESKN
         Pl_yoelii_CHC   (1799) NQTNLINK-LENHKLAEMRRIAALLYKKNKKFKEAINLS-KKEGQIKDAIDIARVSKNHLYVEDLI-----NYFI-QIKN
     ANCESTRAL_CHC       (2077) DNISLAQRRLEKHDLIEFRRIAAYLYKGNNRWRQSVELCHKKDKLYKDAMQYAAESKDAELAENLLNLVCSSQWFLEEGK
             Consensus   (2161) D I LAQ  LEKHEL EFRRIAAYLYK   RWKQS  L  KKD LYKDAM  A  S   EL EELL        F
Table 4.3 continued
```

```
                                2241                                                                                2320
   Ho_sapiens_CHC17  (1563) RECFGACLFTCYDLLRPDVVLETAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLLDASESLR--------KEEE-----
Ma_fascicular_CHC17  (1537) XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXX----XXXXXQVDKLLDASESLR--------KEEE-----
    Bo_taurus_CHC17  (1563) RECFGACLFTCYDLLRPDVVLETAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLLDASESLR--------KEEE-----
    Su_scrofa_CHC17  (1590) XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXX----XXXXXXXXXXXXXXXXXX-------XXXX-----
  Mu_musculus_CHC17  (1562) RECFGACLFTCYDLLRPDVVLETAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLLDASESLR--------KEEE-----
Ra_norvegicus_CHC17  (1563) RECFGACLFTCYDLLRPDVVLETAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLLDASESLR--------KEEE-----
    Ga_gallus_CHC17  (1563) KECFGACLFTCYDLLRPDVVLETAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLLDASESLR--------KEEE-----
     Da_rerio_CHC17  (1572) KECFAACLFTCYDLLRPDVVLETAWRHNL-MDFSMPYFIQVM----KEYLSKVDKLLDASESLR--------KEEE-----
  Ta_rubripes_CHC17  (1563) KECFAACLFSCYDLLRPDVVLETAWRHNI-MDFSMPYFIQVM----KEYLSKVDKLESSESVR--------KEEE-----
    Ho_sapiens_CHC22  (1563) RECFAACLFTCYDLLRPDMVLELAWRHNL-VDLAMPYFIQVM----KEYLSKVDKLLDALESLR--------KQEE-----
    Ga_gallus_CHC22  (1563) XXCFAACLFTCYDLLHPDVVLELAWRHNI-MDFAMPYFIQVM----KEYLTKVDNLLDASESLR--------KEEE-----
  Ta_rubripes_CHC22  (1560) KECFAACLFASYDLLHPDVVLELAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLEEAESQR--------KTEE-----
Ci_intestinalis_CHC  (1567) YECFSAGLYSCYDLLRPDVVLEQAWRHGI-SDYAMPYLIQVM----KEYTIRVDKLETSENVR--------KQEE-----
Dr_melanogaster_CHC  (1564) YDCFAACLYQCYDLLRPDVILELAWKHKI-VDFAMPYLIQVL----KEYTTKVDKLELNEAQR--------EKED-----
    Ae_aegyptii_CHC  (1564) HDCFAACLFQCYDLLR---RHILSWRHNI-MDFAMPYLIQVT----KEYTSKVDKLEVADAER--------QKEG-----
     An_gambiae_CHC  (1564) YDCFAACLFQCYDLLRPDVILELAWRHNI-MDFAMPYIIQVT----KEYTSKVDKLEASDAER--------QKEG-----
     Ca_elegans_CHC  (1565) YDCFAASLYHCYDLLHPDVIMELAWKHKI-MDYAMPYMIQVM----RDYQTRLEKLERSEHER--------KEEK-----
    Ca_briggsae_CHC  (1566) YDCFAASLYHCYDLLHPDVIMELAWKHKI-MDYAMPYMIQVM----RDYQTRLEKLERSEHER--------KEEK-----
      Br_malayi_CHC  (1566) HDCFAASLYQMYDLLHPDVILELAWKHKI-MDFAMPYMIQVM----RDYHSPVXXXXXXXXXX--------XXXX-----
       Pi_taeda_CHC  (1576) KECFASCLFTCYDLIRPDVALELAWFNNM-IDFSFEYLLQFI----KEYTTKVDELVKDKLEN-------LQEVK----S
  Ly_esculentum_CHC  (1576) KECFATCLFVCYDLIRADVALELAWMNNM-IDFAFPYLLQFI----KEYTGKVDELIKDKIEA-------QKEAK----A
     Be_vulgaris_CHC  (1576) XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXX----XXXXXXXXXXXXXXXXXX-------XXXX-----
Me_crystallinum_CHC  (1573) XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXX----XXXXXXXXXXXXXXXXXX--------XXXX-----
    Go_arboretum_CHC  (1577) KECFASCLFVCYDLIRADVALELAWINNM-IDFAFPYLLQFI----KEYTGKVDELIKDKIEA-------QKEVK----A
      Ar_thaliana_CHC  (1577) KECFATCLFVCYDLIRPDVALELAWINNM-IDFAFPYLLQFI----KEYSGKVDELIKDKLEA-------QKEVK----A
   Me_truncatula_CHC  (1583) KECFASCLFVCYDLIRADVVLELAWMHNM-IDFAFPYLLQFI----KEYTGKVDELVKHKIES-------QNEVK----A
    Lo_japonicus_CHC  (1561) KECFASCLFVCYDLIRADIALELAWMHNM-IDFAFPYLLQV-----------------------------------------
          Gl_max_CHC  (1577) KECFASCLFVCYDLIRADIVLELAWMNNM-IDFAFPYLLQFI----KEYTGKVDELVKDKIEA------QNQVKA-----
       Po_tremula_CHC  (1576) KECFAACLFVCYDLIRPDVALELAWMNNM-IDFAFPYLLQFI----KEYTSKVDELIKAKLEA-------LSEVK----A
       Ho_vulgare_CHC  (1574) KECFASCLFICYDLIRADVALELAWTNNM-LDFAFPYLLQFI----KEYTSKVDDLVKDRIES-------QKEEK----A
     Tr_aestivum_CHC  (1576) KECFASCLFICYDLIRADVALELAWTNNM-LDFAFPYLIQFI----KEYTSKVDDLVKDRIES-------QKEEK----A
      So_bicolor_CHC  (1567) KECFASCLFICYDLIRPDVALELAWMNNM-LDFAFPYLLQFI----KEYTSKVDDLVKDKIES-------QKEER----A
        Or_sativa_CHC  (1573) KECFASCLFICYDLIRADVALELAWMNNM-VDXCIPXXXXXX----XXXXXXXXXXXXXXXXX--------XXXX-----
          Ze_mays_CHC  (1578) KECFASCLFICYDLIRPDVALELAWMNNM-MDFAFPYLLQFI----KEYTSKVDDLVKDKIES-------QKEER----A
   Di_discoideum_CHC  (1564) NSAFAACLYTCYDFLKPDAVIELAWRNNI-LNYSFPYLIQYV----KEYTTKVDQLVDDFKAR-----QKKTEEE-----
       Sc_pombe_CHC  (1558) YECFAAILYTCYHLLRNDLVMEISWRKGL-QDYAYPYFINFQ----CEMFSKVLNLEKDLKDR-----QAVKSEE-----
  Sa_cerevisiae_CHC  (1569) RRGFVALLYAAYNLVFIEFVLEISWMNSL-EDYIKPFEISIK----KEQNDSIKKITEELAKK-------SGSNE-----
     Ca_albicans_CHC  (1563) HECFVALLYTSYEFIANDYVMELSWLHNL-SNFIKPYEISIAFENQKKLNEVYQDLQKPKEAD-------PKQEE-----
   As_fumigatus_CHC  (1566) RECYVGMLYACYDLIRPDVILEMSWRHGL-HDFTMPYMINFL----CEQTRTIEMLKKDNEER--------KAPE-----
  Pa_brasilensis_CHC  (1570) XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXX----XXXXXXXXXXXXXXXXX--------XXXX-----
       Le_major_CHC  (1577) -ECFAACLYACYDLLKPATVLQKAWLNQR-TEMAMPYMIQVL----QDYSSKIEKMEKLMMDAQTAAKDAARRAG---PV
       Tr_brucei_CHC  (1580) -DSFVSCLYACYDYLSPDVVLEKAWLNNR-INIAMPYLIQAI----HDFTQRVSRLEKGANDG-----MQPSKDG-----
  En_histolytica_CHC  (1549) KEAFGNCLNVCYDYVPTEVALELGYKNKL-MIQTMPFFMCKKM----KDTNDRIKKLEQAEKDR-----QNAKKELGYKNK
Gi_intestinalis_CHC  (1801) KEVFAVILIRCGHCIRPDVVMELAWVNRM-TDYMMPYMISTL----QRYGSAVRTLAGSIAEA--------R--------
    Th_annulata_CHC  (1636) LNYLLVALVINFHLLDLPMVLESVWLNNVNLDVIMPLII-----------------------------------------
      To_gondii_CHC  (1628) PAGFAACLYTCYPLVKPDVALELAWRHKC-MDYCMPFLIQ-----------------------------------------
      Cr_parvum_CHC  (1682) KENFVACLYTCYEFLRPDTVMEMAWKHNC-LDATMPFFIQSL----RDMTNTIDXXXXXXXXX--------XXXX-----
  Pl_falciparum_CHC  (1885) KEAFCACLIVCYDLLKPDYILEIVWTSGF-KDQAMLYFIQII----SDYTNQIEHMKKQIEDM-------EKEKK-----
      Pl_yoelii_CHC  (1871) KEALCACLIACYDILKPDYVLEIIWLSGF-KDHAMLYFIQII----SDYTQQIETMKKQIEDI-------EKEKK-----
      ANCESTRAL_CHC  (2157) KECFAACLFTCYDLLRPDVVLELAWRHNI-MDFAMPYFIQVM----KEYLTKVDKLEAAESLRQNAKKELGKQEE-----
          Consensus  (2241)  ECFAACLF CYDLLRPDV LELAW  N   DFAMPY IQ      REY  KVD L               K E
```

Table 4.3 continued

```
                                2321                                                                                2400
    Ho_sapiens_CHC17  (1625) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYTA-PPYGQP---------QPGFGYSM--
Ma_fascicular_CHC17   (1599) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYT-APPYGQP---------QPGFGYSM--
    Bo_taurus_CHC17   (1625) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYTA-PAYGQP---------QPGFGYSM--
    Su_scrofa_CHC17   (1653) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----------------XXXXXXX-XXXXXX---------XXXXXXXX--
  Mu_musculus_CHC17   (1624) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYTA-PPYGQP---------QPGFGYSM--
Ra_norvegicus_CHC17   (1625) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYTA-PPYGQP---------QPGFGYSM--
    Ga_gallus_CHC17   (1625) QATETQPIVYGQPQLMLTAGPSVAVPPQAP-----------------FGYGYTA-PPYGQP---------QPGFGYSM--
     Da_rerio_CHC17   (1634) QLTEAQPIVYGTPQLMLTAGPNVAVPPQQP-----------------YGYGYPTATGYTQP---------AQPGFGYGM--
  Ta_rubripes_CHC17   (1625) QATETQPIVYGTPQLMLTAGPSVPVAPQQA-----------------YGYGYQA-PAGYTQP------APQPGFTYGM--
    Ho_sapiens_CHC22  (1625) HVTEPAPLVFDFDGHE-------------------------------------------------------------------
    Ga_gallus_CHC22   (1625) QVTEPTPIVFXXXXXX-------------------------------------------------------------------
  Ta_rubripes_CHC22   (1622) EVTEPQPMVFGR---------------------------------------------------------------------
Ci_intestinalis_CHC   (1629) EQTENKPIVYG----------------------------------------------------------------------
Dr_melanogaster_CHC   (1626) DSTEHKNIIQMEPQLMITAGPAMGIPPQYA-----------------QNYPPGA-----ATVTAAGGKN-MGYPYL----
    Ae_aegyptii_CHC   (1623) ENSEHKSIILPEPQLMLTAGPGMGMP-QYA-----------------PQYAGGYVP--AGP-NMG-----PYPGYGGM--
     An_gambiae_CHC   (1626) ESTEHKSIIMPEPQLMLTAGPGIGMP-QYA-----------------PQYAGAYVP--PQP-NMP-----PYQYGGFN--
     Ca_elegans_CHC   (1627) AEQQQNNGMTMEPQLMLTYGAPAPQMTYPG-----------------TTGGYGGQPAYGQP------GQPGYNAPGFM--
    Ca_briggsae_CHC   (1628) AEQQQNNKMTMEPQLMLTYGAPAPQM--------------------GYPGAPA-GYGGQP----AYGQPGQPGFNAPGF
      Br_malayi_CHC   (1628) XXXXXXXXXXXXXXXXX-------------XXXXX----------------------------------------------
       Pi_taeda_CHC   (1640) KEKEEKDMVAQQNMYAQ-------------------------------------------------------------------
  Ly_esculentum_CHC   (1640) KENEEKDVMKQQNMYAQ-------------------------------------------------------------------
     Be_vulgaris_CHC  (1638) XXXXXXXXXXXXXX-----------------------------------------------------------------------
Me_crystallinum_CHC   (1635) XXXXXXXXXXXXXXXX---------------------------------------------------------------------
    Go_arboretum_CHC  (1641) KEQEEKEVIAQQNMYAQ-------------------------------------------------------------------
     Ar_thaliana_CHC  (1641) KEQEEKDVISQQNMYAQLLPLALPAPPMPGMGGGGYGPPPQMGG----------MPG--MSG--MPP-----MPPYGMPP
   Me_truncatula_CHC  (1647) KEQEEKEVIAQQNMYAQ-------------------------------------------------------------------
    Lo_japonicus_CHC  (1601) ---------------------------------------------------------------------------------
          Gl_max_CHC  (1641) KEQEEKEVIAQQNMYAQLLPLALPAPPMPGMG-GGFAPPPPMGG--------LGMPP---------------MPPFGMPP
       Po_tremula_CHC (1640) KEKEEKDMVSQQNMYAQ-------------------------------------------------------------------
       Ho_vulgare_CHC (1638) KDNEEKELVAQQNMYAQ-------------------------------------------------------------------
     Tr_aestivum_CHC  (1640) KEQEEKDVVAQQNMYAQLLPLALPAPP--GMG-----GPPPMGGMGMPPMGGMGMPP--MGGMGMPPMGPGPMPAYGMPP
      So_bicolor_CHC  (1631) KEKEEKDLVAQQNMYAQLLPLALPAPPMPGMG-----GPPPMGGMGMPPMGGMGMPP----------MGPGPMPAFGMPP
       Or_sativa_CHC  (1635) XXXXXXXXXX-----------------------------------------------------------------------
         Ze_mays_CHC  (1642) KEKEEKDLVAQQNMYAQ-------------------------------------------------------------------
  Di_discoideum_CHC   (1629) KEQQNIESSQYQPDLTNL------SYGYA--ATGGMLALPPAVG---YQQQQQP-QQMYNPNQ-MMGGFQQNYNQYGGF-
       Sc_pombe_CHC   (1623) ESASTIGAGILGNTLMLTQGPMANNNDQF-------DSFQ-----QASPMPRLGNF----------------------
  Sa_cerevisiae_CHC   (1632) EHKDGQPLMLMNSAMNVQPTGF-----------------------------------------------------------
    Ca_albicans_CHC   (1630) EPGVGQPLMNGF-------------------------------------------------------------------
     As_fumigatus_CHC (1628) ---------------------------------------------------------------------------------
  Pa_brasilensis_CHC  (1632) XXXXXXXXXXXXXXXXX---------------------------------------------------------------
       Le_major_CHC   (1648) QDPPSVPLMIEQGGGMP-----------------------------------------------------------------
      Tr_brucei_CHC   (1644) SRPGGVPGYAGGNDPLMIQAGPAQPMGVPM----------HNVNIHPQPGYGG----------VPGQ--GYAGGMGNPN
  En_histolytica_CHC  (1619) LMDQTMPFMCKKMKDTNDRIKKLEQAEKDR---QNAKKEP------QLAQQDTP-FGMLALPA----APGMMPMQQPMGG
Gi_intestinalis_CHC   (1860) --GMIPAQQAPQQH-----------------------------------------------------------------------
    Th_annulata_CHC   (1675) ---------------------------------------------------------------------------------
      To_gondii_CHC   (1667) ---------------------------------------------------------------------------------
      Cr_parvum_CHC   (1744) XXXXXXXXXXXXXXXXX-------------XXXXXXXXX-------------------------------------------
  Pl_falciparum_CHC   (1948) MNKSAPNDYSAMNNQFN-YSLNN-----------------NLSIMPPQNNF------MSSN---------SFDKYDMFN
       Pl_yoelii_CHC  (1934) MNKSAPNDYSANTISNQFTYSLNK----------------NLSIMPPQNNY-----IPNNSDFDKYDISYMTTCEFYFL
       ANCESTRAL_CHC  (2227) EVTEPQPMVFGRDGHELTLLGYGLA-----LAGPSVAVPPQAP--YGYGYQA--------PPGYGQPAPQPGFGYGMFMY
           Consensus  (2321)    E
```

Table 4.3 continued

		2401 2413
Ho_sapiens_CHC17	(1676)	----------
Ma_fascicular_CHC17	(1650)	----------
Bo_taurus_CHC17	(1676)	----------
Su_scrofa_CHC17	(1708)	----------
Mu_musculus_CHC17	(1675)	----------
Ra_norvegicus_CHC17	(1676)	----------
Ga_gallus_CHC17	(1676)	----------
Da_rerio_CHC17	(1687)	----------
Ta_rubripes_CHC17	(1679)	----------
Ho_sapiens_CHC22	(1641)	----------
Ga_gallus_CHC22	(1641)	----------
Ta_rubripes_CHC22	(1634)	----------
Ci_intestinalis_CHC	(1640)	----------
Dr_melanogaster_CHC	(1679)	----------
Ae_aegypti_CHC	(1675)	----------
An_gambiae_CHC	(1678)	----------
Ca_elegans_CHC	(1682)	----------
Ca_briggsae_CHC	(1682)	M----------
Br_malayi_CHC	(1651)	----------
Pi_tacda_CHC	(1657)	----------
Ly_esculentum_CHC	(1657)	----------
Be_vulgaris_CHC	(1652)	----------
Me_crystallinum_CHC	(1651)	----------
Gy_arboretum_CHC	(1658)	----------
At_thaliana_CHC	(1702)	MGGY---------
Me_truncatula_CHC	(1664)	----------
Lo_japonicus_CHC	(1681)	----------
Gl_max_CHC	(1697)	MGGY---------
Po_tremula_CHC	(1657)	----------
Ho_vulgare_CHC	(1655)	----------
Tr_aestivum_CHC	(1711)	MGNY---------
So_bicolor_CHC	(1696)	MGGY---------
Or_sativa_CHC	(1646)	----------
Ze_mays_CHC	(1659)	----------
Di_discoideum_CHC	(1695)	----------
Sc_pombe_CHC	(1667)	----------
Sa_cerevisiae_CHC	(1654)	----------
Ca_albicans_CHC	(1642)	----------
As_fumigatus_CHC	(1628)	----------
Pa_brasilensis_CHC	(1650)	----------
Le_major_CHC	(1665)	----------
Tr_brucei_CHC	(1701)	MMPY---------
En_histolytica_CHC	(1685)	MG----------
Gi_intestinalis_CHC	(1872)	----------
Th_annulata_CHC	(1675)	----------
To_gondii_CHC	(1667)	----------
Cr_parvum_CHC	(1771)	----------
Pl_falciparum_CHC	(1994)	NN----THF----
Pl_yoelii_CHC	(1992)	CFFYFPYLFALRK
ANCESTRAL_CHC	(2292)	GGF---YLFALRK
Consensus	(2401)	

Table 4.3 continued

Table 4.4: Alignment of Clathrin Light Chain Sequences and Sequence Fragments

```
                              81                                                                             160
     BO TAURUS CLCA   (53) -------GA-PGSQPHGEPPGI-PDAVDGVTNGDYYQ-ESN----------GPTDSYAAISQ---VDRL-QSEPESIRKW
     HO SAPIENS CLCA  (58) -------GA-PGPQPHGEPPGG-PDAVDGVMNGEYYQ-ESN----------GPTDSYAAISQ---VDRL-QSEPESIRKW
    MU MUSCULUS CLCA  (58) -------GA-PG-RATRRAGGG-PDAVDGVMNGEYYQ-ESN----------GPTDSYAAISE---VDRL-QSEPESIRKW
 RA NORVEGICUS CLCA   (58) -------GA-PGPQAHGEPPGG-PDAVDGVMNGEYYQ-ESN----------GPTDSYAAISE---VDRL-QSEPESIRKW
    HO SAPIENS LCPS8  (59) -------VT-PRPQPQGEQLGX-PDAVDGVMXGDYYQ-ESN----------GPRDNYAAISQ---VDRF-PSXPENICRC
   HO SAPIENS LCPS12  (12) -------GA-PRPQAHGKL-----DAVDGVMNGKYYQ-ESN----------SPTDSCAAISQ---VDRL-QSEPESIRKW
       XE LAEVIS LCA  (47) -------GE-LPVSLQT----G-DGATDTVMNGDFYQ-EAN----------GPTDGYAAISH---ADRL-RAEPESIRKW
     TA RUBRIPES LCA  (49) -------GD-VPSSLT--------DSNGGCN---LP--ESN----------GPSDAYAAISN---ADRL-QAEPESLRKW
        DA RERIO LCA  (51) -------GD-VPSSLSQ-------DQDGGAMNGDLHG-ESN----------GPSDVYAAISS---VDRL-QAEPESLRKW
     HO SAPIENS LCB   (50) -------SH-AAPAQPGPTSGAGSEDMGTTVNGDVFQ-EAN----------GPADGYAAIAQ---ADRL-TQEPESIRKW
       BO TAURUS LCB  (50) -------SQ-GGLAQPGPASGA-SEDMGATVNGDVFQ-EAN----------GPADGYAAIAQ---ADRL-TQEPESIRKW
  RA NORVEGICUS LCB   (50) -------SQ-VASAQPGLASGGGSEDMGTTVNGDVFQ-EAN----------GPADGYAAIAQ---ADRL-TQEPESIRKW
     MU MUSCULUS LCB  (50) -------SQ-VASAQPGLASGAGSEDMSTTVNGDVFQKESN----------GPTDSYAAISE---VDRL-QSEPESIRKW
       GA GALLUS LCB  (47) -----DSAS-APAGQAAPPETAGFQNGGATVNGDVFQ-ESN----------GPTDAYAAIAK---ADRL-TQEPESIRKW
       XE LAEVIS LCB  (47) --------D-SPVEEAEMVDFN--DFGTIPMNGELYQ-ETS----------DYTDGYSAIAQ---ADRL-TQEPESLRKW
     TA RUBRIPES LCB  (41) -------AD-GQPQSAN---------YGKLSRLTSDY-ESN----------GPTDGYAAIAQ---ADVQ-RQEPESLRKW
        DA RERIO LCB   (1) --------------------------------------DTN----------GMTDSYSAIAQ---VDIQ-RQEPESLRKW
 CI INTESTINALIS LC   (55) -------IA-VADAR--------------------------------------------------LEEL-RTEPEKIRVW
  AP CALIFORNICA LC   (43) ----------QPEAAQGFDAFGSAEGTDAPSGDQQTN--------------GPSSMYEAISQ---HDTL-RAEPEKIKIW
 DR MELANOGASTER LC   (43) -------PAASTDEGLGELLGGTASEGDLLSAGGTGGLESSTGSFEVIG--GESNEPVGISG----PPPSREEPEKIRKW
      AN GAMBIAE LC   (52) -------GDPVPKAESTGLENELNGSFEMISNADAQDPPTTTTGDNNLNTEDDNDDFAGFT----VPKQVTEEPEKIRKW
      CA ELEGANS LC   (53) AGDEP--PPVVHPTDSGVDLDGLVDDNAAAPAIVVPAVEPMVNG----NHSASSGGSKGPSPIL--STVPRIEAEKIRLW
   DI DISCOIDEUM LC   (39) ----------ISFETTTFDNSNNNNNNNNHNNNSYNS--------------GFDGDLSSVDG----DMKPKETAPAMREY
    AR THALIANA LC1   (41) --------VEDSVDDVFAAPSSDYGAYSNGDGIFGSN--------------GDHDGPILPPP----SEMESDEGFALREW
    AR THALIANA LC2   (33) -----------KESDLPCGDSSPRPETQSPPSINSFD---------------DTNDSILPPP----SAMEKEEGFALREW
    AR THALIANA LC3   (49) --------DPASPNGYGFGASSPNHDFSSPFESSVNDANGNGGG------SGGDAIFASDGPILPDPNEMREEGFQRREW
         SC POMBE LC  (55) -------FENDSEAEQTRFEQNFPPIDAEMQASGTFS--------------APKAPYMGQAEVH-PPEDESGDPEPVRKW
   SA CEREVISIAE LC   (50) -----EASPAKDDDEIRDFEEQFPDINSANGAVSSDQNGSATVSSGNDNGEADDDFSTFEGANQSTESVKEDRSEVVDQW
        NE CRASSA LC  (63) -------GGLDSSAGAGAGLETSAAFESQFPDLSAGN-ESVAPGG---TITGAGPSVTYNSGYAPYAQE-EQEPEVIREW
      ANCESTRAL LCa   (81) AGDEPDSGDVVPPSLQGEPPGGGDDANGGVMNGDLYQKESNTGGDNNLNTEGPTDGYAAISQYFTVDRLPQAEPESIRKW
      ANCESTRAL LCb   (81) AGDEPDSADVVPPAQAGPPDGGGSDDGGAPMNGDLYQKESNTGGDNNLNTEGPTDGYAAIAQYLTVDRLPRQEPESIRKW
           Consensus  (81)                                NG   Q E N          GPTD YAAI      DRL   EPESIRKW
```

Table 4.4 continued

		161 240
BO TAURUS CLCA	(109)	REEQT-ERLEALDANSRKQEAEWKEKAIKELDEWYARQDEQLQKTKANNRVADEAFYKQPFADVIGYVTNINHPCYSLEQ
HO SAPIENS CLCA	(114)	REEQM-ERLEALDANSRKQEAEWKEKAIKELEEWYARQDEQLQKTKANNRVADEAFYKQPFADVIGYVTNINHPCYSLEQ
MU MUSCULUS CLCA	(113)	REEQT-ERLEALDANSRKQEAEWKEKAIKELEEWYARQDEQLQKTKANNRAAEEAFYKQPFADVIGYV------------
RA NORVEGICUS CLCA	(114)	REEQT-ERLEALDANSRKQEAEWKEKAVKELEEWYARQDEQLQKTKASNRVADEAFYKQPFADVIGYVTNINHPCYSLEQ
HO SAPIENS LCPS8	(115)	KEEKM-ECLAALAAYSPKQEAEWKEKAINELEEWRARQDEQLQKRKANYR------------------------------
HO SAPIENS LCPS12	(64)	REEQT-ERLEALDANSQKQEAEWKEKAIKELEVXYVRXDEQLQKTKANRA------------------------------
XE LAEVIS LCA	(99)	REEQR-SRLEMLDANSKKQESEWKEKASKELEEWYTRQDELLQKAKANNR------------------------------
TA RUBRIPES LCA	(94)	REEQQ-DRLEVLDKNSRKQEAEWKEKAKVELEEWNARQNEQLEKTKTNNRVLDEDFYKQPFSELIGYVYVAPTNILPPSS
DA RERIO LCA	(101)	REEQR-DRLEELDANSRKQEAEWKEKAKLELEEWHTRQNEQLEKTKVNNR------------------------------
HO SAPIENS LCB	(107)	REEQR-KRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRIADKAFYQQPDADIIGY------------V
BO TAURUS LCB	(106)	REEQR-KRLQELDAASKVMEQEWREKAKKDLEEWNQRQSEQVEKNKINNRIADKAFYQQPDADIIGY------------V
RA NORVEGICUS LCB	(107)	REEQK-KRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNRIADKAFYQQPDADTIGY------------V
MU MUSCULUS LCB	(108)	REEQQRKRLQELDAASKVTEQEWREKAKKDLEEWNQRQSEQVEKNKINNR------------------------------
GA GALLUS LCB	(106)	REEQK-KRLEELDAASKVTEQEWREKAKKDLEEWNLRQNEQMEKNRANNR------------------------------
XE LAEVIS LCB	(101)	REEQK-KRLEELDAASKVTEQEWREKAKKDLDEWYQRQSEQTEKNKVSNR------------------------------
TA RUBRIPES LCB	(89)	REEQK-QRLEELDLASKAKEEEWREKAKKELEDWHVHQSEQMEKNKANNRIADKAFYKQPSSDVIG------------LV
DA RERIO LCB	(29)	REEQK-TRLEELDSASKAAEAVWREKAKKELEDWHIHQTEQMEKNKVNNRIADKAFYKQPNSDVIG------------FV
CI INTESTINALIS LC	(76)	REENK--LLAEKDRESERKQQEWLAQARKELEDWDRNRLEQVEKTKESNRPFVQIGYTVNFYHACYSPNFT---------
AP CALIFORNICA LC	(95)	REEQK-ARLEKKDSEEDKKKLELKEVAKKELDDWYKHHTEQLEKTKENNR------------------------------
DR MELANOGASTER LC	(110)	REEQK-QRLEEKDIEEERKKEELRQQSKKELDDWLRQIGESISKTKLASR------------------------------
AN GAMBIAE LC	(121)	REDQK-ARLEEKDREEERKKDELREQARKELEDWYKHHEETISKTKSANR------------------------------
CA ELEGANS LC	(125)	KAQQE-QLLSKKDEAEEKKKIELRANAKKELEEWYKQREKTLQLSHDENLK--------------------------NEK
DI DISCOIDEUM LC	(91)	LEKHE-KEMQEKKKKSEEKRQKKIAEAKQSLDNFYSEREAKKKTALKNNR------------------------------
AR THALIANA LC1	(95)	R-RQNAIQLEEKEKREKELLKQIIEEADQYKEEFHKKIEVTCENNKAANR------------------------------
AR THALIANA LC2	(83)	RRLNA-LRLEEKEKEEKEMVQQILEAAEQYKAEFYSKRNVTIENNKKLNR------------------------------
AR THALIANA LC3	(115)	RRLNT-IHLEEKEKKEKEMRNQIITEAEDFKKAFYEKRDKTIETNKTDNR------------------------------
SA CEREVISIAE LC	(125)	KQRRA-VEIHEKDLKDEELKKELQDEAIKHIDDFYDSYNKKKEQQLEDAA-----------------------------K
SC POMBE LC	(113)	KEDQM-KRIQERDESSKKLRESNIEKARKAIDDFYENFNDKRDKVIAKSR-----------------------------K
NE CRASSA LC	(131)	REKRD-AKLAKRAEQFAAQRAETIAEAQKNIDEFYENYNNKKEKAIGQTR-----------------------------K
ANCESTRAL LCa	(161)	REEQRRDRLEELDANSKKQEAEWKEKAKKELEEWYARQNEQLEKTKVNNRVADEAFYKQPFSDVIGYVYNANHPILPPEQ
ANCESTRAL LCb	(161)	REEQKRQRLEELDAASKAKEQEWREKAKKELEEWYQRQSEQMEKNKVNNRIADKAFYKQPFSDVIGYVYNAPHPILPPEV
Consensus	(161)	REEQ RLEELDA SK E EW EKAKKELEEWY RQ EQ EK K NNR

Table 4.4 continued

```
                                 241                                                                            320
BO TAURUS CLCA        (188)  AAEEAFVNDIEES---SPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
HO SAPIENS CLCA       (193)  AAEEAFVNDIDES---SPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
MU MUSCULUS CLCA      (180)  AAEEAFVNDIDES---SPGTEWERVAPLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
RA NORVEGICUS CLCA    (193)  AAEEAFVNDIDES---SPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
HO SAPIENS LCPS8      (164)  -A-EDFVNDIDES---SPCTGWXQVARLFDCNPKXS------KQGKEVS--------------------CMHSVLNSLKQ
HO SAPIENS LCPS12     (113)  -AEEAFVNDIDQF---FPGTEWESVAWLCNFNPKSR------RQAKDIS--------------------QRCSVLISLKQ
XE LAEVIS LCA          (43)  AAEEAFVSDVEET---SPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
TA RUBRIPES LCA       (173)  AAEEAMISDLDDN---NPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
DA RERIO LCA          (150)  AAEEAMVSELDEN---SPGTEWERVARLCDFNPKSS------KQAKDVS--------------------RMRSVLISLKQ
HO SAPIENS LCB        (174)  ASEEAFVKESKEE---TPGTEWEKVAQLCDFNPKSS------KQCKDVS--------------------RLRSVLMSLKQ
BO TAURUS LCB         (173)  ASEEAFVKESKEE---TPGTEWEKVAQLCDFNPKSS------KQCKDVS--------------------RLRSVLMSLKQ
RA NORVEGICUS LCB     (174)  ASEEAFVKESKEE---TPGTEWEKVAQLCDFNPKSS------KQCKDVS--------------------RLRSVLMSLKQ
MU MUSCULUS LCB       (158)  ASEEAFVKESKEE---TPGTEWEKVAQLCDFNPKSS------KQCKDVS--------------------RLRSVLMSLKQ
GA GALLUS LCB         (155)  ASEEAFLKESKEE---TPGSEWEKVAQLCDFNPKSS------KQSKDVS--------------------RMRSVLISLKQ
XE LAEVIS LCB         (119)  ASDEALAVDTKEE---ETGTEWERVARLCDFNPKSS------KQSKDVS--------------------RL---------
TA RUBRIPES LCB       (156)  SSEEAFLAETDDS---CPGSEWERVARLCDFNPKTS------KQAKDVS--------------------RMRSVLISLKQ
DA RERIO LCB           (96)  ASEEAFLKECEDD---SPGTEWEKVARLCDFNPKTS------RQTKDVS--------------------GMRSVLISLKQ
CI INTESTINALIS LC    (145)  -AEEQFLKDRDQ----AAPSEWERVSKLCEFNPKST------KSTKDVT--------------------RMRSTLLHLKQ
AP CALIFORNICA LC     (144)  AAEDAFVKERDEK---VSGQAWEKITRLCEFNPKNS------KTTKDVG--------------------RFRGILLQLKQ
DR MELANOGASTER LC    (159)  NAEKQAATLENGT--IEPGTEWERIAKLCDFNPKVN------KAGKDVS--------------------RMRSIYLHLKQ
AN GAMBIAE LC         (170)  NAEKQFVAETDEI---EPGTEWERIAKLCDFNPKTN------KSNKDIS--------------------RMRSIILQLKQ
CA ELEGANS LC         (178)  SNQELFAKQQDG------DAQWETVNKLVDQ--QKS------KSGKDLS--------------------RLKTLLAGLKH
DI DISCOIDEUM LC      (140)  DHNKSLETDSTSG---NTTHTWESVVSMIDLQAKPN------PANKDTS--------------------RMREILIRLKN
AR THALIANA LC1       (144)  EKEKLYLENQEKFYAESSKNYWKAIAELVPKEVPTIEKRRGKKEQQDPKK-PTVSVIQGPKPGKPTDLTRMRQILVKLKH
AR THALIANA LC2       (132)  --EKE--KNQEKFYAEADKNNWKAIAELIPREVPVIEN----RGNKKKTATITVIQ--GPKPGKPTDLSRMRQVLTKLKH
AR THALIANA LC3       (164)  EKEKLYWANQEKFHKEVDKHYWKAIAELIPREVPNIEK---KRGKKDPDKKPSVNVIQGPKPGKPTDLGRMRQIFLKLKT
SC POMBE LC           (163)  EQEKLLEENESK--S-TGTTSWERILKLIDLSDKPE------AHGRSTE--------------------RFRELLISLAK
SA CEREVISIAE LC      (175)  EAEAFLKKR-DEFFG-QDNTTWDRALQLINQDDADI------IGGRDRS--------------------KLKEILLRLKG
NE CRASSA LC          (181)  EAEEFLAS--REDTT-SGGTSWERIAKLVDVSGKGA------KGGAAGSG-----------------KERFRELLISLKK
ANCESTRAL LCa         (241)  AAEEAFVSEIDESYAISPGTEWERVARLCDFNPKSSEKRRGKKQAKDVSKKPAVAVIQGPKPGKPADLERMRSVLISLKQ
ANCESTRAL LCb         (241)  ASEEAFVKETDESYAISPGTEWERVARLCDFNPKSSEKRRGKKQAKDVSKKPAVAVIQGPKPGKPADLERMRSVLISLKQ
Consensus             (241)  AAEEAF      E    PGTEWE VA LCDFNPK S      KQ KDVS                    RMRSVL SLKQ
```

Table 4.4 continued

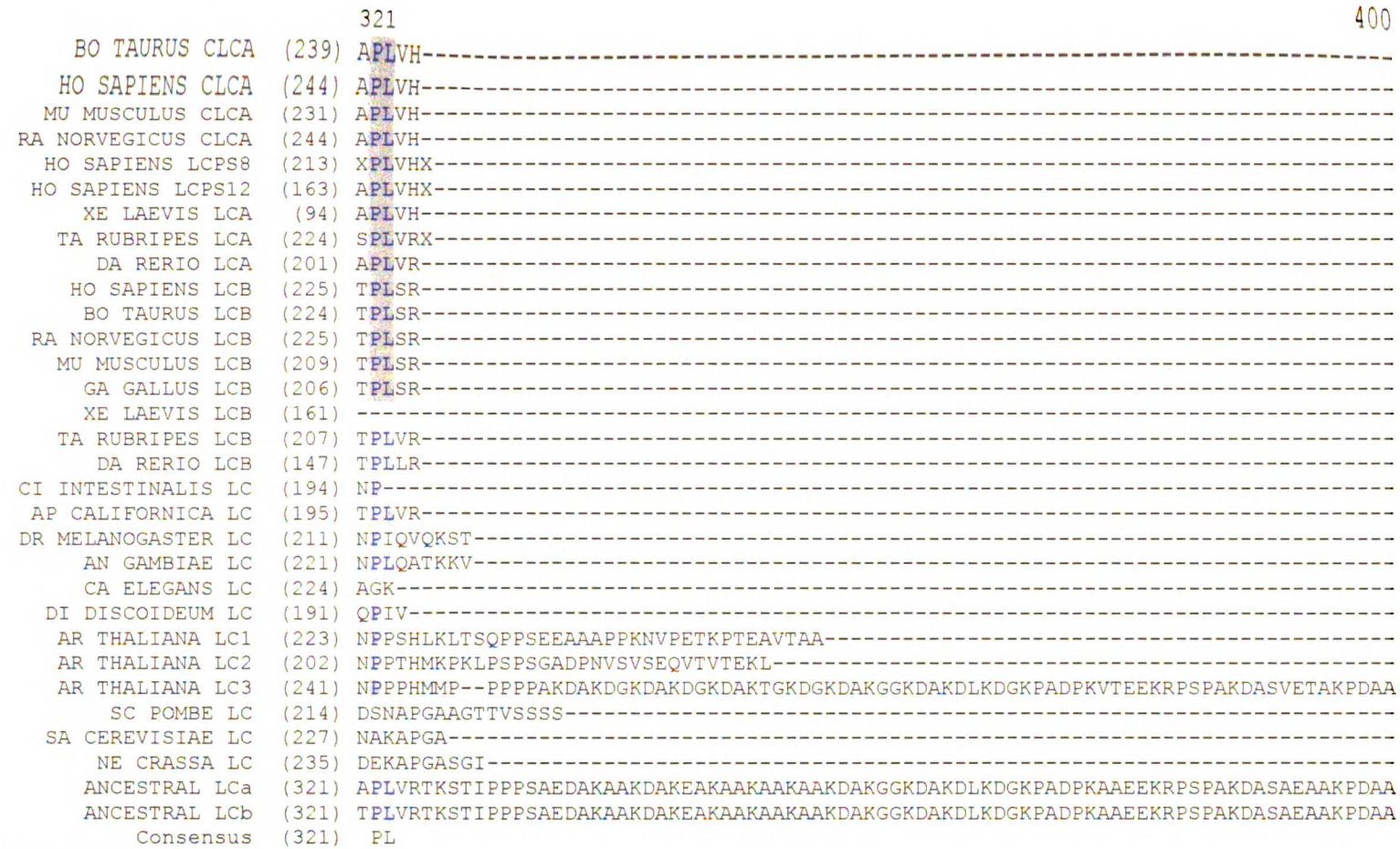


Table 4.4 continued

		401 420
BO TAURUS CLCA	(244)	----------
HO SAPIENS CLCA	(249)	----------
MU MUSCULUS CLCA	(236)	----------
RA NORVEGICUS CLCA	(249)	----------
HO SAPIENS LCPS8	(219)	----------
HO SAPIENS LCPS12	(169)	----------
XE LAEVIS LCA	(157)	----------
TA RUBRIPES LCA	(230)	----------
DA RERIO LCA	(206)	----------
HO SAPIENS LCB	(230)	----------
BO TAURUS LCB	(229)	----------
RA NORVEGICUS LCB	(230)	----------
MU MUSCULUS LCB	(214)	----------
GA GALLUS LCB	(211)	----------
XE LAEVIS LCB	(161)	----------
TA RUBRIPES LCB	(212)	----------
DA RERIO LCB	(152)	----------
CI INTESTINALIS LC	(196)	----------
AP CALIFORNICA LC	(200)	----------
DR MELANOGASTER LC	(220)	----------
AN GAMBIAE LC	(230)	----------
CA ELEGANS LC	(227)	----------
DI DISCOIDEUM LC	(195)	----------
AR THALIANA LC1	(259)	----------
AR THALIANA LC2	(234)	----------
AR THALIANA LC3	(319)	ASGEGEKPVAVTEAEGTKAE
SC POMBE LC	(230)	----------
SA CEREVISIAE LC	(234)	----------
NE CRASSA LC	(245)	----------
ANCESTRAL LCa	(401)	ASGEGEKPAAAAEAEGAKAE
ANCESTRAL LCb	(401)	ASGEGEKPAAAAEAEGAKAE
Consensus	(401)	

Table 4.4 continued

Common name	NCBI DNA Accession	NCBI Protein Accession	Kingdom	Class	Genus	Species	Ref
Human CHC17	G1705916	Q00610	Metazoa	Mammalia	Homo	sapiens	(Dodge et al., 1991; Nomura et al., 1994)
Macaque CHC17	AB050403	-	Metazoa	Mammalia	Macaca	fascicularis	(Osada et al., 2001)
Cow CHC17	G1705915	P44951	Metazoa	Mammalia	Bovine	taurus	(Liu et al., 1995)
Pig CHC17	Ensembl: F14522	-	Metazoa	Mammalia	Sus	scrofa	(Wintero et al., 1996)
Mouse CHC17	Embl: AL592222.9 NCBI contig NT_031414 and NW_000039	-	Metazoa	Mammalia	Mus	musculus	(Waterston et al., 2002)
Rat CHC17	G116514	P11442	Metazoa	Mammalia	Rattus	norvegicus	(Kirchhausen et al., 1987a; ter Haar et al., 1998)
Chicken CHC17	GGA427965	CAD20886	Metazoa	Aves	Gallus	gallus	(Wettey et al., 2000)
Zebrafish CHC17	BM777438 BQ284965 AW421256 AI558722 BG307402 BQ284322 Sanger supercontig z06s013284	-	Metazoa	Actinopterygii	Danio	rerio	(Sprague et al., 2001)
Japanese Pufferfish CHC17	FRUP00000157850 Scaffold 621 v 3	-	Metazoa	Actinopterygii	Takifugu	rubripes	(Aparicio et al., 2002; Elgar et al., 1999)
Human CHC22	G2506298	P53675	Metazoa	Mammalia	Homo	sapiens	(Kedra et al., 1996; Long et al., 1996; Sirotkin et al., 1996)
Chicken CHC22	603984283F1 603485646F1 603788828F1 603747580F1 047368.1 021350.1	-	Metazoa	Aves	Gallus	gallus	(Boardman et al., 2002)
Japanese Pufferfish CHC22	FRUP00000149307 Scaffold 385	-	Metazoa	Actinopterygii	Takifugu	rubripes	(Aparicio et al., 2002; Elgar et al., 1999)
Sea squirt CHC	Ci0100133226	-	Metazoa	Phlebobranchia	Ciona	intenstinalis	(Dehal et al., 2002)
Fruit Fly CHC	NM_057694.2	NP_477042	Metazoa	Insecta	Drosophila	melanogaster	(Adams et al., 2000)
Mosquito CHC	AY009101	AAK12626	Metazoa	Insecta	Aedes	aegyptii	(Kokoza et al., 1997)
African malaria mosquito CHC	AAAB01008859.1	EEA08110	Metazoa	Insecta	Anopheles	gambiae	(Celera, 2002)
Roundworm CHC	NM_066859.1	NP_499260	Metazoa	Secernentea	Caenorhabditis	elegans	(Gonczy et al., 2000)
Roundworm CHC	AC084685.1 ENSCBRP0000000804 1		Metazoa	Secernentea	Caenorhabditis	briggsae	(Harris et al., 2003)
Elephantiasis CHC	TIGR_6279 194346 189088 214282 189086 194672 201727		Metazoa	Secernentea	Brugia	malayi	(Williams et al., 2000)
Loblolly pine CHC	AW010935		Viridiplantae	Pinaceae	Pinus	taeda	(Whetten et al., 2001)
Tomato CHC	BG127598		Viridiplantae	Solanaceae	Lycopersicon	esculentum	(Quackenbush et al., 2000)
Sugar beet CHC	BI096310 BQ489421 BQ594159		Viridiplantae	Chenopodiaceae	Beta	vulgaris	(de los Reyes et al., 2001)
Iceplant CHC	BE037308		Viridiplantae	Aizoaceae	Mesembryanthemum	crystallinum	(Bohnert et al., 2000)

Table 4.5: Databank Accession Codes for Clathrin Heavy Chain Sequences Used in Alignment

Tree cotton CHC	AI731103 BG443262 CA991331		Viridiplantae	Malvoideae	Gossypium	arboreum	(Blewitt et al., 1999)
Thale cress CHC	NM_111688.1 NM_111950.1 AC012562.5 AY094397	NP_187466 NP_187724 AAG51341	Viridiplantae	Brassicaceae	Arabidopsis	thaliana	((AGI), 2000; Theologis et al., 2000)
Barrel medic CHC	BQ140972 BI265895 MTtuc02-12-15.11958 MTtuc02-12-15.9453 AW689178 CA991331 AG137552 BH739319		Viridiplantae	Papilionoideae	Medicago	truncatula	(Bell et al., 2001)
Japanese Lotus Flower CHC	AP004467		Viridiplantae	Papilionoideae	Lotus	japonicus	(Sato et al., 2001)
Soybean CHC	G7441347	T06779	Viridiplantae	Papilionoideae	Glycine	max	(Blackbourn et al., 1996)
European aspen CHC	BU894962 BU810505 BU895942		Viridiplantae	Salicaceae	Populus	tremula	(Sterky et al., 1998)
Wild barley CHC	CA030159 CA022613 HVtuc02-11-10.12031 HVtuc02-11-10.8717 BU994164		Viridiplantae	Pooideae	Hordeum	vulgare	(Matthews et al., 2003)
Bread wheat CHC	CA726748 CA734056 BE515818 BJ316154 BQ905469 TAtuc-02-12-22.22669 TAtuc-02-12-22.11721 TAtuc-02-12-22.9539		Viridiplantae	Pooideae	Triticum	aestivum	(Matthews et al., 2003)
Broomcorn CHC	SBtuc-02-10-21.5120		Viridiplantae	Panicoideae	Sorghum	biocolor	(Mullet et al., 2002)
Rice CHC	CA756890 CA757329 D22409 BX000501 BX000491 OStuc-02-10-21.5570		Viridiplantae	Ehrhartoideae	Oryza	sativa	(Bohnert et al., 2000; Yu et al., 2002)
Corn CHC	AY103752 AW053034 BM318825 BH882780 BZ413705 CA399753 BZ405349 PCO092520	-	Viridiplantae	Panicoideae	Zea	mays	(Coe et al. 2002)
Slime mold CHC	167687	P25870	Mycetozoa	Dictyostelium	Dictyostelium	discoideum	(O'Halloran and Anderson, 1992)
Fission Yeast CHC	NC_003424.1	NP_594148	Fungi	Taphrinomycetes	Schizosaccharomyces	pombe	(Wood et al., 2002)
Baker's Yeast CHC	G116515	P22137	Fungi	Endomycetes	Saccharomyces	cerevisiae	(Feuermann et al., 1997; Lemmon et al., 1991)
Thrush CHC	Contig6.2026 SDSTC_5476	-	Fungi	Endomycetes	Candida	albicans	(Tait et al., 1997)
Beano CHC	TIGR_5085.735	-	Fungi	Plectomycetes	Aspergillus	fumigatus	(Denning et al., 2002)
Fungus CHC	AF441250.1	AAL35337	Fungi	Eurotiomycetes	Paracoccoidioides	brasilensis	(Felipe et al., 2003)
Leishmaniasis CHC	Sanger_5664 LM36.1 contig7		Euglenozoa	Trypanosomatida e	Leishmania	major	(Myler et al., 2001)
Sleeping sickness CHC	TBR78858	CAC51440	Euglenozoa	Trypanosomatida e	Trypanosoma	brucei	(Morgan et al., 2001)
Amoebic dysentery CHC	Ehistolyt_ENTFZ89TF TIGR_5759	-	Entamoebid ae	Entamoeba	Entamoeba	histolytica	(Clark et al., 2003)
Giardia CHC	AY125725.1	AAM83403	Diplomonadi dae	Hexamitidae	Giardia	intestinalis	(Marti and Hehl, 2002)
Protozoan parasite CHC	Sanger_5874 Contig1542, contig3		Alveolata	Theileriidae	Theileria	annulata	(Tait and Hall, 2003)
Toxoplasma CHC	TIGR_5811 contig:33126	-	Alveolata	Eimeriida	Toxoplasma	gondii	(Kissinger et al., 2003)
Parasite CHC	CVMUMN_5807 cparvum_Contig1711	-	Alveolata	Eimeriida	Crytosporidium	parvum	(Strong and Nelson, 2000)
Malaria CHC	AAN36274.1	NP_701550.1	Alveolata	Plasmodium	Plasmodium	falciparum	(Gardner et al., 2002)
Rodent malaria CHC	AABL01000504	EAA21246.1	Alveolata	Plasmodium	Plasmodium	yoelii yoelii	(Carlton et al., 2002)

Table 4.5 continued

Common name	NCBI DNA Accession	NCBI Protein Accession	Kingdom	Class	Genus	Species	Ref
Human CLCA	M20471	P09496	M etazoa	Mammalia	Homo	sapiens	(Jackson and Parham, 1988)
Human LCps8 pseudogene	NT_015280		Metazoa	Mammalia	Homo	sapiens	(Lander et al., 2001)
Human LCps12 pseudogene	NT_024397		Metazoa	Mammalia	Homo	sapiens	(Lander et al., 2001)
Cow CLCA	X04849	P04973	Metazoa	Mammalia	Bovine	taurus	(Brodsky et al., 1987; Jackson et al., 1987)
Rat CLCA	M19260	P08081	Metazoa	Mammalia	Rattus	norvegicus	(Kirchhausen et al., 1987b)
Mouse CLCA	U91848	O08585	Metazoa	Mammalia	Mus	musculus	(Scott et al., 1996)
Frog CLCA	BG814584 BG885389 BG813947 BG611814	-	Metazoa	Amphibia	Xenopus	laevis	(Klein et al., 2002)
Japanese Pufferfish CLCA	CAAB01001172		Metazoa	Actinopterygii	Takifugu	rubripes	(Aparicio et al., 2002; Elgar et al., 1999)
Zebrafish CLCA	CA975665		Metazoa	Actinopterygii	Danio	rerio	(Sprague et al., 2001)
Human CLCB	NM_007097	P09497	Metazoa	Mammalia	Homo	sapiens	(Jackson and Parham, 1988; Ponnambalam et al., 1994)
Bovine CLCB	X04852	P04975	Metazoa	Mammalia	Bovine	taurus	(Brodsky et al., 1987; Jackson et al., 1987)
Rat CLCB	M15883	P08082	Metazoa	Mammalia	Rattus	norvegicus	(Kirchhausen et al., 1987b)
Mouse CLCB	AK009844.1	BAB26539.1	Metazoa	Mammalia	Mus	musculus	(Cerninci and Hayashizaki, 1999)
Chicken CLCB	BM490884	-	Metazoa	Aves	Gallus	gallus	(Boardman et al., 2002)
Frog CLCB	BF613534 BG515527 BE491314 AW767161	-	Metazoa	Amphibia	Xenopus	laevis	(Klein et al., 2002)
Japanese Pufferfish CLCB	CAAB01001207		Metazoa	Actinopterygii	Takifugu	rubripes	(Aparicio et al., 2002; Elgar et al., 1999)
Zebrafish CLCB	Zfish41356-385B12.p1c Z35724-a3423c10.p1c Zfish35934-350a11.p1c Zfish35935-271607.p1c		Metazoa	Actinopterygii	Danio	rerio	(Sprague et al., 2001)
Sea squirt CLC	C10100148440 Scaffold 140		Metazoa	Phlebobranchia	Ciona	intestinalis	(Dehal et al., 2002)
Sea Hare CLC	L14586	APLCLTLCA	Metazoa	Gastropoda	Aplysia	californica	(Hu et al., 1993)
Fruit Fly CLC	NM_079454	Q9VWA1	Metazoa	Insecta	Drosophila	melanogaster	(Adams et al., 2000; Vasyukevich and Bazinet, 1999)
Mosquito CLC	AAAB01008933.1	EAA09983	Metazoa	Insecta	Anopheles	gambiae	(Celera, 2002)
Roundworm CLC	AAB37069	T29460	Metazoa	Secernentea	Caenorhabditis	elegans	(Waterston and Consortium, 1998)
Slime mold CLC	AC117081	AAM43766	Mycetozoa	Dictyostelium	Dictyostelium	discoideum	(Glockner et al., 2002)
Thale cress CLC	NM_129564 AF049236 AC006234	NP_565921 AAC14416 AAD20919	Viridiplantae	Brassicaceae	Arabidopsis	thaliana	(Scheele and Holstein, 2002)
Fission yeast CLC	NC_003423	NP_595750	Fungi	Taphrinomycetes	Schizosaccharomyces	pombe	(Wood et al., 2002)
Baker's yeast CLC	NC_001139	NP_011683	Fungi	Endomycetes	Saccharomyces	cerevisiae	(Goffeau et al., 1996; Tettelin et al., 1997)
Fungus CLC	AL807366.1	CAD36975	Fungi	Pyrenomycetes	Neurospora	crassa	(Schulte et al., 2002)

Table 4.6: Databank Accession Codes for Clathrin Light Chain Sequences Used in the Alignment

Chapter Five: Materials and Methods

Construct creation

Production of the plasmids for the polyhistidine-tagged bovine proximal leg segment (1074-1522) and Hub (1074-1675) in pET15b by S. H. Liu (Liu et al., 1995) and for the for polyhistidine-tagged distal leg segment (1-1074) in pET 23d by B. Greene (Greene et al., 2000) have been previously described.

Briefly, for the Hub, SK991-1675 was digested with AccIII (also called BspEI) and HindIII. The resulting fragment contained a.a. 1074 (AccIII) to 3' UTR HindIII region. Both ends were Klenow filled-in, and the 2 kb fragment was isolated. The pET15b vector was digested with XhoI and BamHI and Klenow filled-in, using CIAP to dephosphorylate the vector. Then vector and insert were ligated together. Protein expression was checked to ensure the isolated clones had the correct orientation.

For the proximal leg, the Hub construct was used with Stratagene QuikChange Mutagenesis Kit in a PCR to insert a stop codon after 1522. The primer used was named 1522C: 5'CTT GCA CAG ATC TAC ACT CTA TTT and its complementary strand. Successful constructs were confirmed by sequencing.

The first round of clathrin mutants to test the Ybe Hypothesis (Chapter Two) were created by Ernest Chen and screened and sequenced by Diane Wakeham. The mutations were created using a Stratagene QuikChange

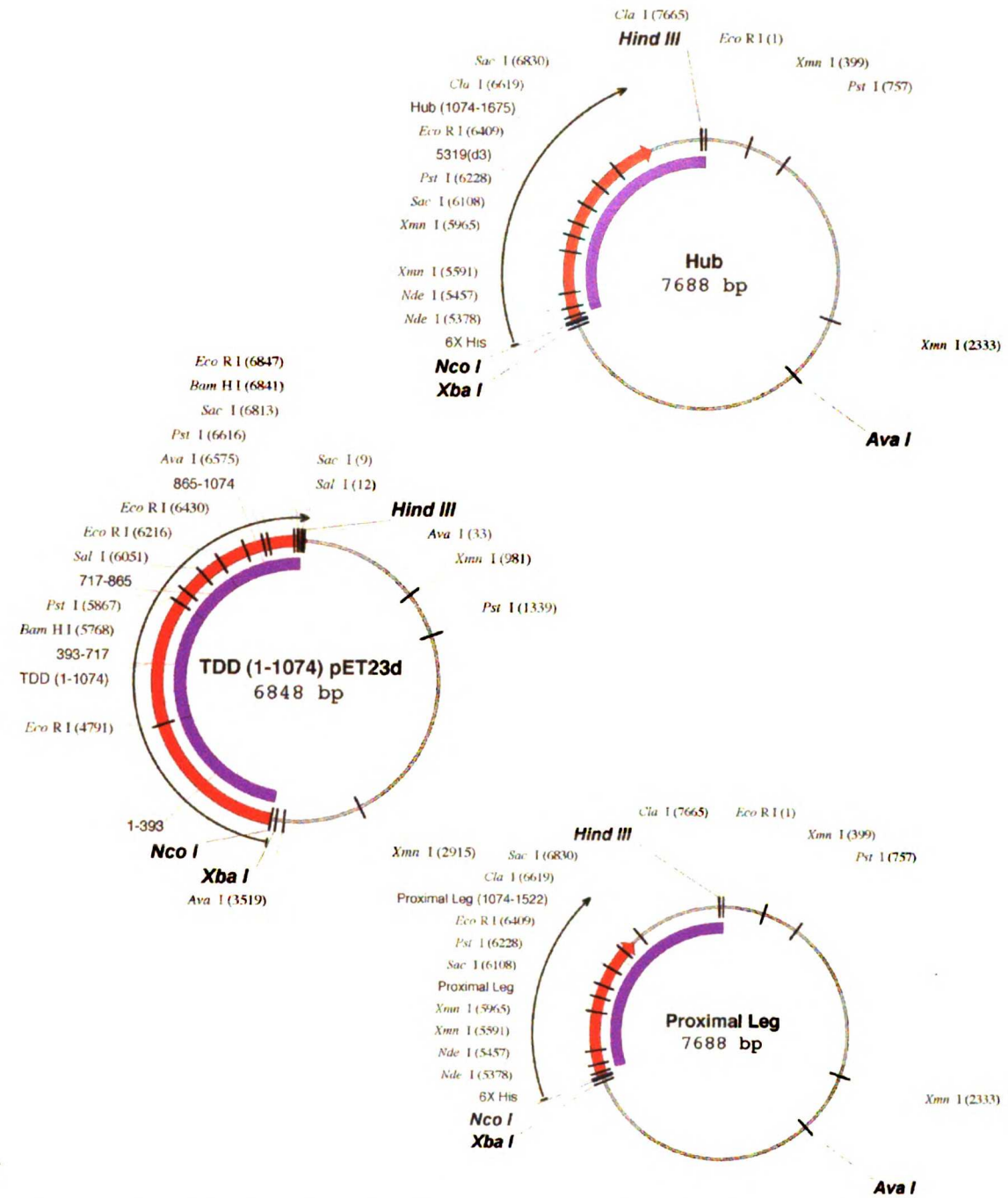


Figure 5.1: Plasmid constructs created by previous lab members
Clathrin heavy chain Hub and proximal leg segment in pET15b were created by Shu-Hui Liu (Liu et al., 1995) and was used as the basis for mutagenesis and chimera studies. Proteins expressed from these constructs were used in biophysical assays. Clathrin terminal domain and distal leg (TDD) in pET 23d was created by Barrie Greene (Greene et al., 2000) and used to create chimeric constructs.

Mutagenesis Kit and the primers noted here. For "Ala5" (1161RKKAR->AAAAA) the oligos were named Hub1171-Ala5 and HubR1171-Ala5. The sequence of the sense oligo was 5' TTG CAG ATG GCC GCC GCG GCC GCT GCT GAG TCC TAT GTG. For "Mutant 2" (1151EE->QQ) the sense oligo read 5' GTG GAA ACT GGC AAC AGC TGG TGA AGT AC. For "Mutant 3" (1162KKAR->EKAE) the sense oligo read 5' GAT GGC CCG CGA GAA GGC TGA GGA GTC CTA TGT GG. For "Mutant 5" (1313H->Y) the sense primer read 5' GACTTGAGCGAGCTTACATGGGGATG. For "Mutant 6" (1335 H->Y) the sense oligo read 5' GAA AAT GAG GGA GTA CCT GGA GCT GTT C.

Two additional mutants (Chapter Two) were created by E. Chen using as a template two of the mutants above. "Mutant 4" used "Ala5" as a template (1161 AAAAA->AAKAA) with the sense primer 5' GAT GGC CGC CGC CAA GGC TGC TGA GTC C. Also, "Mutant 56" was created identically to "Mutant 6", but it used "Mutant 5" as a template such that "Mutant 56" had two point mutations (1313, 1335 HH->YY).

The CHCR6 fragment of the proximal leg (1271-1419) (Chapter Two) was created in pET15b as follows: The clathrin Hub (1074-1675) in pET15b with an N-terminal polyhistidine tag was used as a template in a polymerase chain reaction with TOPO primers TOPO6N (5' CTC GCT CAT ATG TGT GGG CTT C) and TOPO6C (5' CAT CAG CGG ATC CTA TAA CAA CAG TAG C). This amplified fragment was cloned into a TOPO-TA vector, pCR 2.1-TOPO. The resulting pCR 2.1-CHCR6 clones were digested by BamHI and NdeI, to reveal the orientation of the cloning and isolate the desired insert. Likewise, the

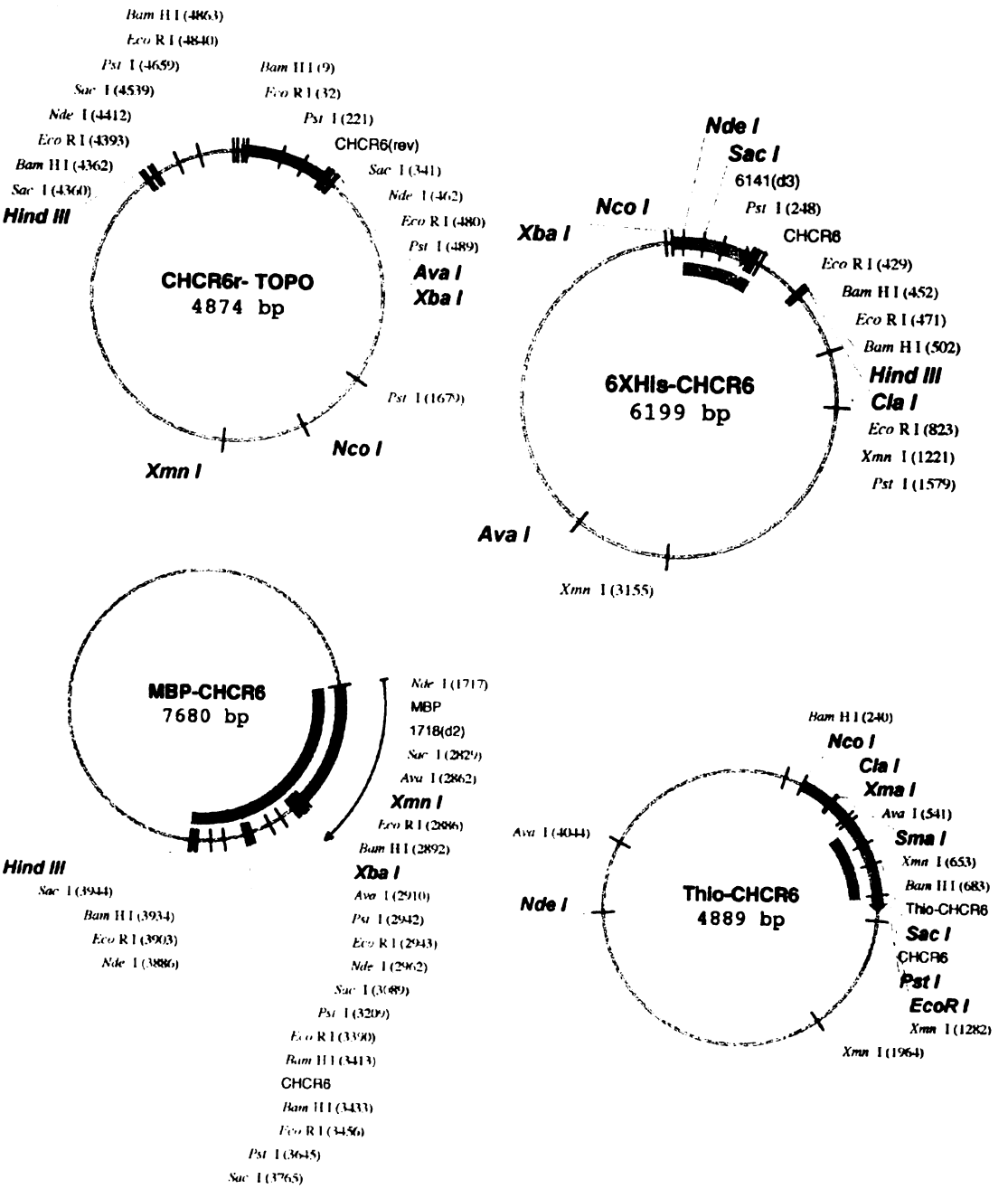


Figure 5.2: CHCR6 Constructs for Protein Expression
After amplification by PCR, CHCR6 was inserted into a TOPO vector. From there it was cut and pasted into pET15b for an N-terminal 6X Histidine tag, into pMAL-c2X for a fusion product with an N-terminal maltose binding protein, and into pBAD/Thio-TOPO for a fusion product with an N-terminal thioredoxin protein.

TDD/pET15b construct was digested by BamHI and NdeI to provide the vector fragment. The digested fragments were separated on a gel, extracted and purified from the gel, and combined for ligation using Promega Ligase T4. Successful construction was confirmed by digestion and sequencing.

CHCR6-MBP (Chapter Two) was created from the pCR2.1-CHCR6 intermediate construct, which was digested with XbaI and HindIII. The resulting 500 bp insert was separated on an agarose gel and extracted and purified from the gel slice. The pMAL-c2X vector was similarly cut with XbaI and HindIII and separated from the gel. Insert and vector were joined by T4 Ligase and the clones were sequenced to confirm successful creation.

ThioCHCR6 (Chapter Two) was created from Hub (1074-1675) through amplification with primers ThioNCHCR6 (5' GGG CTT CAT ATT GTA GTA CAT GC) and ThioCCHCR6 (5' CAG TGG CTT GAA TTC TAA GTA GAA TTG). The PCR product was cloned directly into a pBAD/Thio-TOPO plasmid (Invitrogen) and sequenced to confirm correct orientation and sequence of the clone.

For the second round of mutagenesis to test the Roseman Model (Chapter Two), Stratagene QuikChange Mutagenesis Kit was again used with Hub as a template. "Double" (1331 KMR->EME) was created using sense oligo 5' CT AAG TTT AAG CCA CAG GAA ATG GAG GAG CAC CTG GAG CTG, and "Switch" (1331 KMRE->EMRK) used 5' CT AAG TTT AAA CCA CAG GAA ATG AGG AAG CAC CTG GAG CTG TTC.

Chimeric constructs (Chapter Three) were created using the Seamless Cloning Kit (Stratagene Cat #214400). Seamless primers were used to amplify

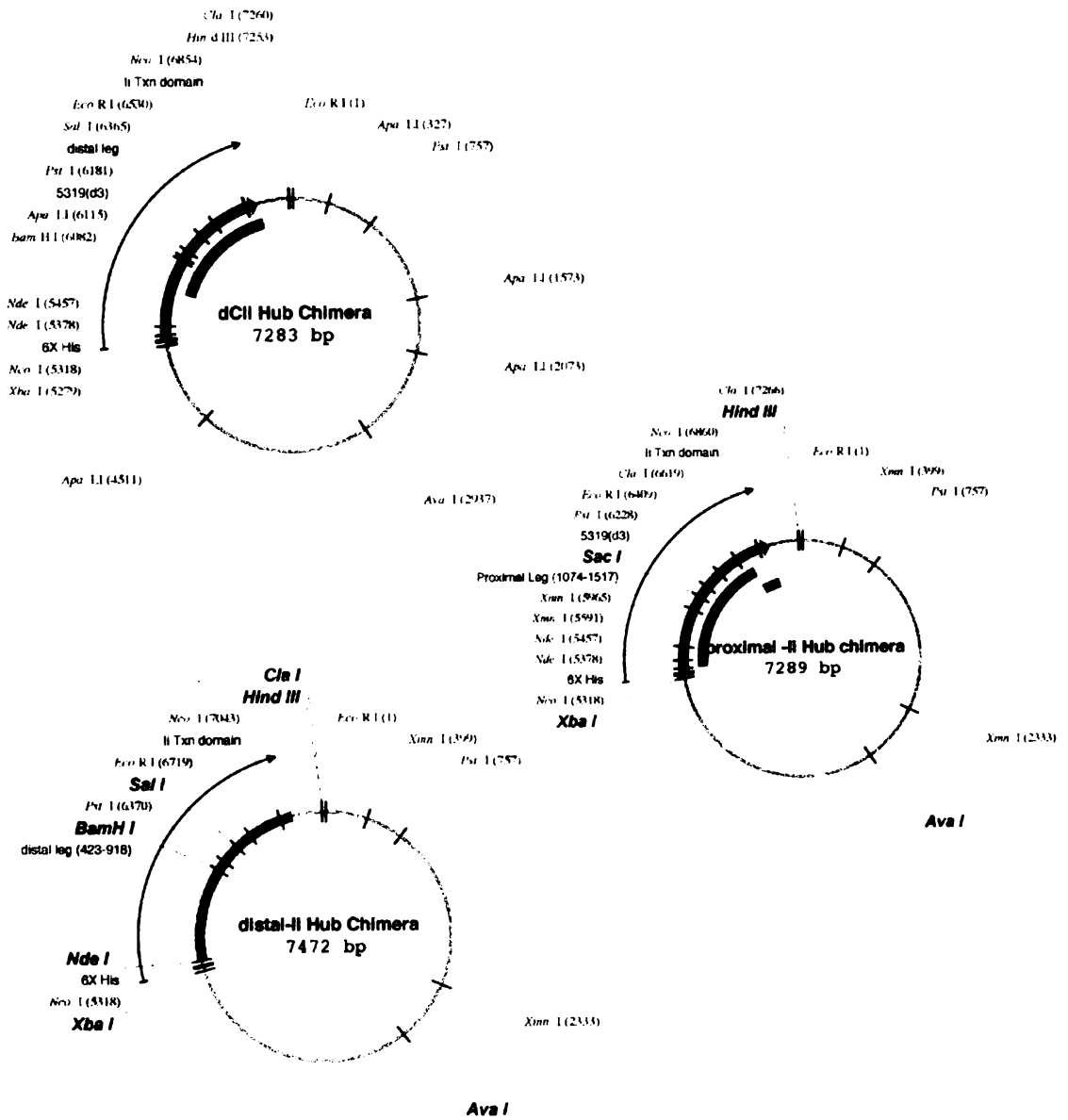


Figure 5.3: Chimeric Clathrin Hub Constructs
Clathrin Hub was expressed with a CHCR from distal leg substituted for the CHCR6 in dCHub (not described in text). In the proximal-Ii chimera, clathrin heavy chain proximal leg was fused to the trimerization domain from invariant chain. In the distal-Ii chimera, the distal leg was fused to the invariant chain trimerization domain.

cDNA for the invariant chain trimerization domain (110-195) from previously described plasmid MT ΔIP33 #7 (Liu et al., 1998) and ligate that together with the cDNA for bovine clathrin proximal leg segments (1074-1517) or distal leg segments (423-925) within pET15b according to Seamless Protocol. Purification protocol was identical to previously published Hub purification. XL-10 Gold cells were used for transformation, to give larger colonies. Primers used for Proximal-Ii from Hub were as follows: f-pET15b-A: 5' AGC TCT TCC GCC TGC TAA CAA AGC CCG AAA GG; r-Hub-B: 5' CCA GCG ACT CTT CCC TTT GAA GAG GTA AGC AGC; f-IiTxn-C: 5' GCC CTC TTC AAA GGG CCC ATG CAG AAT GCC ACC; R-IiTxn-D: 5' C CTC CTC TTC CGG CGA CTA TTT CGG TGG AGC GTC AGT GGG C. Primers used for Distal-Ii from Proximal-Ii were as follows: AfliHub: 5' TTC TCT TCG CGG ATG CAG AAT GCC ACC AAG TAT GGC; B-ryHub: 5' C ATC TCT TCT TCG GTC GAG CAT ATG GCT GCC GCG CGG; C-fd423: 5' GGC TCT TCT CGA GAC CAG GGC CAG CTA AAC AAA TAT G; D-CHCR3.5r: 5' GAC TCT TCC CCG ATC ACA CTG GCC ACG CTC GTA AGC.

Agarose Gel Analysis

Agarose gel was prepared by combining 0.5 grams of agarose with 50 mL of TAE buffer, and microwaving to fully dissolve the sugar. Then 0.5 µL of EtBr was added for visualization of the gel and the solution was poured into a mold and a comb added. I prepared 50 X TAE by combining 60.5 g Tris base, 14.3 mL glacial acetic acid, and 25 mL of 0.5 M EDTA into 250 mL total volume. I

diluted this stock to prepare 1X TAE solution. The solidified 1% gel was immersed in 1X TAE buffer and loaded with samples after 6X stop reaction had been added in the appropriate quantity to each sample, for a total volume of 10-12 µL per well, alongside a sizing reference.

An electric current of 60- 70 volts is applied across the gel for 1-2 hours, or until reasonable spread between the two dye bands at 500 bp and 4 kb was achieved. The gel was visualized on a UV light box using protective eyewear or beneath a shield. Photographs of the gel were taken. Bands were excised from the gel using a razor blade and a smaller, hand-held UV lamp to visualize the location of the band.

DNA Preparation and Handling

Promega Wizard mini-preps and Qiagen mini-preps were used to isolate DNA plasmids from DH5α cells (Gibco #18265-017). Qiagen QIAquick Gel Extraction kit was used to recover and purify DNA from agarose gels. Restriction enzymes were purchased primarily from NEB or Gibco and digests were done at 37° C, with typically a 10 µL digest using 1-3 µL DNA and 0.5-1 µL of enzyme. Custom primers were ordered from Gibco. Primers and plasmids were stored at -20° C in a non-defrosting freezer; cells were stored at -70 or -80. T4 DNA Ligase (Gibco) was used to join DNA fragments according to the protocol.

DNA Spectrophotometric Concentration Assay

A glass cuvette with sloped sides was used, holding a final volume of 1 mL. The spectrophotometer was blanked with water. A sample volume of 2 µL was added to the tube and diluted in 1 mL water, briefly vortexing to mix. Absorbance was measured at 260nm using 320 nm as a background correction. The concentration of DNA for double-stranded DNA was calculated as follows: 50 µg/mL X OD260 X 1 mL/2 µL = µg/mL ds-DNA. For single stranded DNA concentration was calculated as follows: 33 µg/mL X OD260 X 1 mL/2 µL = µg/mL ss-DNA.

DNA Sequencing

Outsourced sequencing was done by the UCSF Biomolecular Resource Center or by Elim Biopharmaceuticals. Sequencing was also done in-house according to the Thermo-Sequenase Kit procedure, involving PCR with ^{33}P-labelled dNTPs. Glass sequencing plates were kept scrupulously clean by washing with soap, water, and 70% ethanol. The outer glass plate was then coated lightly in the hood with Sigmacote to grease and allowed to dry for 30 minutes.

The sequencing apparatus was assembled according to package directions and the gel was mixed: 44.1 g urea, 12.6 mL Long-Ranger Gel Solution (50%), 5.25 mL water, 5.25 mL of 20% TTE Glycerol tolerant buffer, which had been dissolved over low heat and brought to a total volume of 100 mL.

The warm gel was filtered through a 0.45 micron filter and diluted to 105 mL. Then polymerization was induced through the addition of 375 µL 10% APS and 37.5 µL TEMED. The solution was inverted once to mix and withdrawn immediately into a 100 mL syringe, expelling bubbles and slowly pouring the gel over about 70 seconds to fill the space between the glass plates of the apparatus, tilting and tipping to expel bubbles as needed. A comb was inserted and clamped into place, and the gel was laid down horizontally and allowed to polymerize over 30 minutes, covered with damp paper towels and saran wrap for longer gellation times.

To run, the gel plates were set upright in a running base and upper tray filled with 350- 500 mL of 1X TTE buffer. The comb was removed; the wells were rinsed with a syringe full of buffer, and 2.5 µL of stop sequence was loaded on various columns at the edges or near bubbles to assess which lanes were usable. The gel was heated by applying electric current 80- 120 Watts. The temperature of the gel was monitored using a sticker on the glass plate. The sample solutions were warmed to 70 degrees, and the gel to 45- 50 degrees, before loading 3 µL sample into each sample well. For a primer 100 bp upstream, running until the light blue dye band is 7.5" from the bottom of the plate allowed optimal gel reading. The buffer was drained and the gel cooled before transfer to filter paper, and a gel dryer was used for 1-2 hours to remove residual moisture. The gel was exposed to film over one week before reading sequences.

Creation of Competent Cells

A 100 µL aliquot of an existing batch of competent cells were spread onto an Agar plate (no ampicillin added) and grown overnight at 37 degrees. A calcium chloride solution was mixed by combining 0.88 g $CaCl_2$ dihydrate, 11.90 mL glycerol, and 0.30 g PIPES, and adjusting to pH 7. This solution was autoclaved and cooled before use. A single colony was selected to grow in 2 mL of LB at 37 degrees for 3-4 hours or until OD ~0.5. This culture was diluted into 100 mL LB in a 500 mL flask and shaken at 37 degrees until OD ~0.125. The cells were transferred to 50 mL polypropylene tubes and cooled on ice 10 minutes.

The culture was then centrifuged at 4 degrees 10 minutes, 4000 rpm in GS3 or 2800 rpm in Sorvall RT6000B (1600G). I decanted the media away from pellets, inverting tubes one minute to drain all media. I then resuspended each pellet in 10 mL of calcium chloride solution and placed it on ice 30 minutes. The cells were recovered by centrifugation 10 min 4 degrees as before, decanting fluid as before. I resuspended each pellet in 2 mL ice-cold calcium chloride solution

I transferred 200 µL aliquots of the newly created competent cells to Eppendorf microfuge tubes, and froze them in a dry ice/ ethanol bath. One tube was retained on wet ice for efficiency testing. From this remaining tube, I aliquotted 3 tubes with 50 µL cells. I transformed the first (T3) with 5 µL pUC18 plasmid, the second (T2) with 0.5 µL pUC18 at 0.1 ng/ µL, and the third (T1) with

5 µL of T2, for a 100-fold dilution, bringing each up to a 100 µL plating volume with LB. These were grown at 37 degrees overnight.

I calculated the transformation efficiency as follows: T3: # colonies/100 µL plated X 1 mL LB cells/ 0.5 ng DNA X 1000 ng/1 µg X 1000 µL/ 1 mL = # colonies/ µg DNA. T2: # colonies/ 100 µL plated X 1 mL LB cells/ 0.05 ng DNA X 1000 ng/1 µg X 1000 µL/ 1 mL = # colonies/ µg DNA. T1: # colonies/ 100 µL plated X 1 mL LB cells/ 0.005 ng DNA X 1000 ng/1 µg X 1000 µL/ 1 mL = # colonies/ µg DNA.

Transformation of competent cells

I obtained or created competent cells (such as DH5α Gibco #18265-017, or BL21(DE3) from Stratagene) and stored them in a dry ice/ ethanol bath at -70° C or in a freezer under similar conditions until ready to use. I chilled empty 1.5 mL microfuge tubes and the desired plasmid on wet ice. The competent cells were thawed on wet ice. I gently mixed thawed cells, then aliquotted 50 µL competent cells into as many chilled tubes as needed. Unused competent cells were refrozen immediately in liquid nitrogen or a dry ice/ ethanol bath and returned to the freezer.

A 1-2 µL aliquot of miniprepped clean plasmid DNA was added directly to the chilled cells. I mixed them gently, and incubated cells on wet ice for 30 minutes. Heat shock was used to transform the cells by immersion at 37-42° C for 20- 30 seconds (without shaking). The cells were placed on ice for 2 minutes

following immersion. Then 0.95 mL of room temperature or pre-warmed LB broth was added, and the cells were placed on a rocker at 37° C for one hour to express the plasmid. To select the transformed cells, I spread 100 µL or more of the mixture on Amp+ LB plates, using a triangular glass rod sterilized in ethanol and flames, or by shaking sterile glass beads. The plates were incubated at 37 degrees overnight.

I isolated one colony from the plate using a toothpick and grew it in 2 mL LB/ Amp at 37 degrees in a shaker, for 2-6 hours. I withdraw two 100 µL aliquots to a microfuge tube, added 100 µL of 40% glycerol, and immediately froze them in a dry ice/ ethanol bath or in liquid nitrogen. These stocks were stored in the freezer and freeze/ thaw cycles were minimized for best longevity. The remainder of the sample broth was spun down in a microfuge tube and used for a miniprep, if needed.

Note that for some transformations (TOPO 10 cells, for example) beta-mercaptoethanol is required during the step where the plasmid is added to the cells. Certain cell types required special media, for example TOP10 cells required SOC rather than LB. Spreading 80 µL of X-gal 20 mg/mL onto the plates before the transformation with certain plasmids allowed visual identification of cells transformed by observing the presence of a blue color.

Construct expression and purification

For construct expression in pET15b, BL21 stock were streaked onto LB-Amp+ plates and grown overnight at 37° C. A single colony was selected and inoculated in 2 mL of LB with 2 µL of 100 mg/mL filter-sterilized Ampicillin and shaken 2-3 hours, until growth is evident. Then this culture is used to inoculate a much larger culture of LB, typically one liter, with final concentration of ampicillin of 100 µg/mL. When A600 reaches 0.4- 0.5 AU, cool at 30° C and induce with IPTG for a final concentration of 100 µg/mL. Express at 30° C for 3-5 hours. After centrifugation and discarding the supernatant, the cells were used fresh or frozen for protein purification.

All polyhistidine-tagged constructs were purified using nickel affinity chromatography similar to what has been previously described (Liu et al., 1995). Specifically, binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris, pH 7.9) and wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris pH 7.9) were prepared. On occasion HEPES was used in lieu of Tris as the buffer. Charging buffer (100 mM nickel chloride) and "Magic Elution Buffer" (250 mM imidazole, 0.5 M NaCl, 20 mM Tris pH 7.9) were also prepared. The pellet was resuspended in Binding buffer with protease inhibitors added- per 25 mL of culture, I added 0.143 mL of 17.4 mg/mL PMSF in isopropanol, 25 µL of 1 mg/mL aprotinin and leupeptin, and 25 mg of lysozyme. BugBuster was also found to successfully work if used in combination with Binding Buffer components. The chilled Binding Buffer plus protease inhibitors/ lysozyme mixture was added to the pellet and swished until the pellet was fully resuspended. Then it was left on

ice 15-30 minutes. The resulting lysate was sonicated in Rosette cooling cells on ice at 50% of microtip limit, for 30 second intervals, for 3 minutes total time. If the sonicated solution remained viscous, 5 µg/mL DNAse was added. A 200 µL aliquot of the lysed cells were spun in a TL-100.3 rotor for one hour at 55K at 4 degrees, to test for solubility. The bulk of the lysed cells were transferred to Ti 50.2 tubes and spun at 4 degrees under vacuum at 40 K for 1 hour (150,000 G). Supernatant was kept on ice, pellet was discarded.

Fresh chelating Sepharose resin was used in a column or in a 50 mL conical tube. Seven milliliters of resin was rinsed 3 times with 45 mL of water to remove residual ethanol content and then combined with 40 mL of nickel charging buffer. After a brief shake to mix the resin with its new buffer, the resin was pelleted in a desktop centrifuge and the supernatant discarded. The charged resin was then combined with 40 mL of binding buffer, which changed the color to blue. Finally the resin was combined with the sample and set to rock gently at room temperature for 30 minutes. Excessive clumping indicates good yield. The resulting resin was then transferred into a column and settled by gravity or a peristaltic pump. The resin was washed with 30 mL of binding buffer and 50 mL of wash buffer before elution with 15 mL of Magic Elution Buffer and collection in 1 mL tubes. Bradford was used to test for presence of eluted protein. Fractions kept were combined with 30 µL of EDTA per milliliter of eluate to remove excess metal ions from the column and thus to prevent protein aggregation, and the protein was dialyzed overnight into 10 mM Tris pH 7.9. Do not use DTT.

Chimera purification protocol was identical to Hub purification, though yields were somewhat lower.

Light chain LCb was expressed and purified as above along with Hub, and then isolated by boiling the Hub-LCb complex to release light chains. Fresh proximal legs were then incubated with excess light chain on ice to reoccupy light chain binding sites, as previously described (Liu et al., 1995).

For expression of CHCR6-MBP (Chapter Two), induction was similar to above but IPTG levels were 0.3 mM and purification was done using a maltose resin. The construct was soluble but expressed only MBP and not the full fusion protein.

For expression of ThioCHCR6 (Chapter Two), induction was done in TOP10 cells (LB with 1 µL/mL ampicillin) or LMG194 cells (RM with 1 µL/mL ampicillin and 20% glucose) and induction was with 0.00002- 0.2% arabinose. In both cell lines the construct was pelleted (insoluble) though induction was good at 0.2% arabinose for 4 hours.

Size-exclusion chromatography

For routine purification following nickel affinity column when yield was sufficient, the proteins were concentrated down to about 1 mL using a Centriprep with a 30 kD cutoff and then further purified by FPLC using a Sepharose 6 prep grade column XK16/40 at 1 mL/min in degassed 10 mM Tris pH 7.9 with 0.02%

sodium azide. The strip chart recorder monitored A280 and the fraction collector gathered 1 mL fractions for minutes 20-55 of the 110 minute method.

For calibration of the high resolution FPLC columns (Chapter Three), a Superdex HR200 10/30 column (Amersham Pharmacia #17-0188-01) in pH 6.2 50mM Bis-tris buffer was calibrated at 0.5 ml/min by injection of individual standards using Acetone (1%, EM Science #AX0120-14), Vitamin B12 (0.1 mg/mL, Sigma #V-2876), Myoglobin (1.5 mg/mL, Sigma #M-0630), Transferrin (3 mg/mL, Sigma #T-0178), Bovine IgG (3 mg/mL, Sigma #I-5506); Dextran Blue (0.1 mg/mL), Ferritin (0.5 mg/mL), Ovalbumin (3.7 mg/mL), and Thyroglobulin (3 mg/mL) (Amersham Gel Filtration Calibration Kits Cat#17-0441-01 and #17-0442-01). Kav= (Ve-Vo)/(Vt-Vo) of proximal leg segments and calibration standards were plotted against the log molecular weight to determine molecular size of proximal leg segments, according to Amersham Gel Filtration Kit Instruction Manual.

Additionally a 60 mL XK16/40 Superose 6 column (Amersham Pharmacia #17-0489-01) in pH 6.7 MES 40mM/ Tris 10 mM was used to determine the size of proximal leg segments, terminal/distal domain, or mixtures of both (Chapter Three). These were compared to the size of Hub and terminal/distal domain at pH 7.3 on the same column.

This 60 mL XK16/40 Superose 6 column was also used to look at light chain-proximal leg segment complexes, isolated proximal leg segments, and Hub (Chapter Three) at a flow rate of 1 mg/mL with a 10 mM Bis-tris pH 6.2 mobile phase. Hub at pH 6.2 assembled into a large aggregate that was centrifuged

and removed, leaving small amounts of unassembled Hub in the supernatant for this analysis.

SDS-PAGE Gel Analysis- Coomassie and Western Blots

SDS-PAGE gels were purchased commercially and used according to package directions, or poured manually according to the procedure below. Four stock solutions of lower acrylamide, lower Tris (4X), upper acrylamide, and upper Tris (4X) were created and stored at 4 degrees. The lower portion of a 10% gel was created by combining 4 mL lower acrylamide, 3 mL lower Tris, 5 mL water, 100 µL APS 10%, and 20 µL TEMED. 6 mL was needed per mini-gel, filling 2/3 of the space between glass plates that had been sealed with 1% agarose. The upper gel was created by combining 1.5 mL upper acrylamide, 2.5 mL upper Tris, 6 mL water, 100 µL APS 10% and 20 µL TEMED, and immediately inserting a comb and clamping in place. Samples are heated to boiling 10 minutes before loading on gel. The gels were run in running buffer at 100 volts to stack, then increased to 150 volts for separation.

Some gels were immediately stained in Coomassie to visualize total protein by soaking in Coomasie 30 minutes to one hour, then destaining in 10% acetic acid, 40% methanol.

Other gels instead were transferred onto nitrocellulose for a western blot by creating 1X Laemmli, 10X TBS (6.05 g Tris base, 45 g NaCl in 500 mL pH 7.4), and 3% Milk in TBST (10 mL TBS 10X, 50 µL Tween-20, 3 g milk, 100 µL

sodium azide 20% in 100 mL). Sponges were slowly hydrated in 1 X Laemmli and placed between electrodes and the gel. The gel and nitrocellulose were sandwiched between two thick pieces of paper, and topped with additional Laemmli-soaked sponges. I rolled a truncated pipet over the sponges and gel to remove all air bubbles. Forceps were always used to handle the nitrocellulose. The proteins were transferred at 45 volts over 2-3 hours. The nitrocellulose was transferred to Ponceau Red solution to visualize protein, rinsing gently with distilled water until bands were clearly seen and photographed, photocopied, or scanned as needed. A pencil was used to mark the location of molecular weight standards. Then the membrane was rocked 1 hour at room temperature in 3% Milk/TBST to block. A fresh primary antibody solution was added 1:3000 in milk/ TBST, using 4 mL for a gel-sized membrane. This was rocked at 4 degrees overnight or at room temperature for one hour. Then it was rinsed twice 10 minutes each with 1X TBST, once 5 minutes with 1X TBST with 0.5 M NaCl, and once 10 minutes with 1X TBST to reduce background signal. I diluted secondary antibody GAM-HRP 1:3000 in TBST and rocked at room temperature one hour, repeating rinses as above. Using tweezers the membrane was moved onto plastic wrap, squeezing out excess liquid. Equal amounts of each ECL solution were added and allowed to sit 10-30 seconds. I squeezed out the excess liquid, wrapped the membrane in plastic tightly, and taped it into place in a developer. This was developed with film in darkroom for 30 seconds, shiny side up. The film was fed into processor. The molecular weight location was marked onto the processed film with Sharpie before removing the membrane from the developer.

Bradford Protein Concentration Assay

A commercial 4X Bradford solution was diluted to 1X with distilled water. A commercial standard vial of bovine serum albumin at 2 mg/mL was used for the standards. Six calibration plastic cuvettes plus sample cuvettes were each aliquotted 1 mL of Bradford 1X solution. The calibration standards were given 0 µL, 2 µL, 3 µL, 4 µL, 5 µL, and 6 µL for a total content of 0 µg, 4 µg, 6 µg, 8 µg, 10 µg, and 12 µg protein. For each sample 10 µL was typically added. The OD 595 was blanked to the 0 standard and recorded for each standard and sample. Then sample content in µg was automatically calculated by linear regression, and divided by the volume of sample added to find the original sample concentration.

Spectrophotometric turbidity assay for clathrin self-assembly

Hub or chimera assembly was assayed as increased OD 320 in pH-induced assembly as described previously (Liu et al., 1995). Extent was recorded after 5 minutes assembly, at which point pH was increased to reversibility of the assembly reaction. In Chapter Three, all preparations were scaled relative to Hub assembly assayed on the same day and were averaged over multiple purificiations (Hub n=8, 4 batches; Proximal-li and Distal-li, n=4, 2 batches). Proximal leg segment at 0.3 mg/mL was used as a control (n=3, 2 batches).

Specifically, I used 100 µL of clathrin heavy chain or clathrin hub or hub mutant at 0.3 mg/mL (or higher) in Tris pH 7.9 10 mM buffer. If assembly in the

presence of light chain is desired, add 11.9 µg per assembly reaction for 0.3 mg/mL of hub or clathrin. This sample was placed on ice to allow the light chain to saturate the heavy chain binding sites for a few minutes. The sample was pipetted into a limited volume microcuvette, wiping the sides clean carefully with a Kimwipe. The spectrophotometer was set up to monitor kinetics at 320 nm, collecting data every 12 seconds for 8 minutes. The spectrophotometer was blanked to the unassembled solution. To prepare Assembly Buffer, I combined 9.76 g MES with 0.19 g EGTA, q.s. to 40 mL, adjusted pH to 6.2, 6.7 or other desired level, then finished by q.s. to 50 mL. In some cases, 0.55 g calcium chloride dehydrate was added to the assembly buffer to heighten extent of assembly, but at a cost of high irreversibility. Data used in this dissertation did not contain calcium in the assembly buffer unless specifically noted. The assembly buffer was stored at 4 degrees unless components were insoluble at the chosen pH. To prepare Disassembly Buffer I combined 7.88 g Tris, diluted to 40 mL, adjust pH to 9.0, q.s. to 50 mL, and stored at 4 degrees.

At time zero, I began monitoring kinetics. At time one minute, I added 4 µL of Assembly Buffer to the cuvette between measurements and slid the pipet tip carefully from side to side in the cuvette to mix. I allowed the assembly reaction to proceed for 5 minutes, or until the time reached 6 minutes. Then I added 4 µL of Disassembly Buffer and monitored disassembly for the remaining two minutes.

Special circumstances- Hypoassembly- in this situation, the extent of assembly did not level off with time but continues to rise. Attempts to re-mix in

some cases corrected the problem. I disassembled after 5 minutes, interpreting extent of assembly with caution. Hyperassembly- in this situation, after an initial rise in signal due to assembly, the OD decreased before disassembly buffer was added. To avoid this, calcium should be removed from the assembly buffer, and care should be taken to ensure pure, fresh protein is used in the assay. Irreversibility- Assembly is distinguishable from aggregation by its reversibility. If irreversibility is observed, attempts to re-mix with the disassembly buffer or addition of more disassembly buffer often corrected the problem. Efforts to minimize the concentration of divalent metal ions or increase the EGTA content in the assembly buffer may also assist.

Note that if the full course of assembly is not observed, i.e. if the sample initial absorbance is recorded, the solution is removed from the cuvette, and the solution is replaced in the cuvette again to record extent of assembly at 5 minutes as per previously published results from the laboratory (Liu et al., 1995), much time is saved but the overall result is less accurate because pipetting disrupts the nucleation phase of assembly and reduces the overall extent and/ or increases aggregation.

DEPC Chemical Modification of Histidines

Spectrophotometry is useful for monitoring conjugation of proteins with DEPC (diethyl pyrocarbonate) (Chapter Two). DEPC reaction is fairly specific for covalent modification of histidine, resulting in an increase in OD237. The

reaction of DEPC with protein is quenched by adding imidazole, and is reversed by adding hydroxylamine. By allowing the reaction to proceed to varying extents of time before quenching, a variable number of histidines may be modified, with the most surface-accessible histidines reacting first (Miles, 1977). In these experiments, the DEPC reagent was first calibrated by addition of 5 µL of 5 mM DEPC to a cuvette of 100 µL of 10 mM imidazole at pH 7.9. Absorbance at 230 nm was monitored for 10 minutes, and the concentration of DEPC was calculated according to c = (change in OD 230)/ (1 cm) (3000/cm /M extinction coefficient) and found to be within 50% of the correct value, indicating relatively fresh and potent DEPC. For the control reaction, first clathrin Hub assembly was monitored at 237 and 320 nm. For the experiments, a cuvette of 100 µL of 0.3 mg/mL clathrin Hub was blanked and monitored at 237 and 320 nm over 10 minutes as the following reagents were added: At 1 minute, 5 µL of 5 mM DEPC was added to modify the histidines. At 3 minutes, 5 µL of 60 mM imidazole was added to quench unreacted DEPC reagent. At 4 minutes, 4 µL of MES pH 6.2 1 M was added to induce clathrin Hub assembly. At 9 minutes, 5 µL of 1 M Tris pH 9 was added to reverse clathrin Hub assembly. This sequence was repeated using 10 mM Tris pH 7.9 buffer rather than clathrin Hub as a second control.

RasMol Visualization of Crystal Structure

Many images (Chapters Two and Four) were generated using RasMol (Sayle and Milner-White, 1995).

GRASP Visualization of Surface Charge and Accessible Surface Area

UCSF CGL (Computer Graphics Laboratory) facilities and equipment were used to generate images and make calculations using GRASP (Nicholls et al., 1991) (Chapter Two). After the structure file of the dimeric model was read, the molecular surface for each molecule was build individually, adding the second surface. A surface-to surface distance array was calculated between these two constructed surfaces, and the accessible surface area of the molecule with a probe radius of 1.4 angstroms was calculated to be 16639 square angstroms per proximal leg molecule. Then the excluded surface area (the area of the interface between the two proximal leg segments) was calculated as 1553 square angstroms. Surface area between CHCR6 (residues 1274- 1416) of the proximal leg segment was calculated to be 746 square angstroms.

To build the surface potential map, the default full.crg radius/charge file was used and a new potential map was calculated. A molecular surface was built, excluding waters. The potential on the surface was then calculated for all atoms and all surfaces, and the scale was adjusted to reflect equal intensity of color for equal regions of net charge, whether the charge is acidic (red) or basic (blue). Hydrophobicity maps and histidine maps were generated by manually typing r=hyd to select all hydrophobic residues or r=his to select histidines, and coloring them cyan (hydrophobic) or magenta (histidines). This color was then mapped to the constructed molecular surface with the command vc = a.

MolScript and Raster 3D Visualization of Crystal Structures

UCSF Macromolecular Structure Group (MSG) facilities and equipment were used to generate the images of the crystal structures of clathrin fragments (Chapter Two, Appendix). Terminal domain crystal coordinates were downloaded from the Protein Data Bank (PDB) structure with identification 1BPO. Molscript (Kraulis, 1991) and Raster3d (Merritt and Murphy, 1994) were the programs used to generate these figures.

Analytical ultracentrifugation

A Beckman XL-A or XL-I was used in absorbance mode at 280 nm for both sedimentation velocity and equilibrium assays (Chapters Two and Three). For equilibrium assays, proximal leg segments at 0.1, 0.2, and 0.4 mg/mL in 10mM Tris/ 40 mM MES pH 6.2 were equilibrated at 4 degrees Celsius in a six-channel centerpiece with data collected at 7,000 rpm, 10,000 rpm, and 14,000 rpm. These nine data files were simultaneously fit using WinNonLin (Johnson et al., 1981) to find σ and compare it to a calculated σ for the monomer under those conditions. A forced fit to a dimer was used to estimate lnK2. Hub trimer tested under similar conditions at pH 7.9 fit to an n-mer of 2.9, validating the methodology used. The partial specific volume of proximal leg and trimeric hubs was estimated at 0.732 g/mL using Ultrascan (Demeler and Saber, 1998) and solvent density was estimated at 1.00 g/mL.

For velocity experiments, a 3 mg/mL sample of proximal leg segment was used at pH 6.2 and a 0.3 mg/mL sample at pH 6.7. Data collection proceeded in a two-channel centerpiece at 4 degrees at 40K at pH 6.2 or 8 degrees at 50K for pH 6.7. Each data set was fitted using continuous c(S) distribution analysis in SedFit (Schuck, 2000). The proximal-Ii and distal-Ii chimeras and Hubs were tested at 40K, 8 degrees at pH 7.9 10mM Tris at 0.3 mg/mL.

Proximal legs were tested in the presence of 40 mM MES as above to simulate assembly assays, and also in the presence of equivalent concentrations of citrate, EPPS, phosphate, and Tris to assess the observed molecular weight under multiple pH conditions. EPPS was also tested with 5 micromolar or 5 mM zinc, and MES was tested with 50 mM salt or with formate. In the presence of mM levels of zinc, the proximal leg aggregated (data not shown). However, all other buffers were averaged to generate a smoothly fitted line for the pH sensitivity of the observed molecular size of proximal leg during equilibrium centrifugation (Chapter Two).

Yeast-two-hybrid assays

For yeast-two-hybrid experiments to test interactions between proximal legs (Chapter Three), clathrin heavy chain fragments were PCR amplified from bovine brain cDNAs (Liu et al., 1995) and cloned into *Bam*HI- *Sal*I sites of pGBT9 or pGAD424 (Clontech). Yeast transformation, qualitative liquid β-galactosidase (using o-nitrophenyl β-D-galactopyranoside (ONPG) as substrate), filter assays

were performed using Clontech yeast protocols handbook, as previously described (Chen et al., 2002).

Surface plasmon resonance

An IAsys Plus biosensor (Thermolabs) was used to determine associative interaction of clathrin proximal leg domains (Chapter Three). Proximal leg segments (at 100 μg/mL in 20 mM sodium phosphate, pH 7.5) and light chain LCb (at 50 μg/mL) were each immobilized to a surface of an IAsys dual-well carboxylate cuvette, using EDC/NHS coupling chemistry. Each prepared surface was tested for binding of 160 μg/mL proximal leg segment in PBS, and monitored for 10 minutes. The surfaces were regenerated by washes with 3M sodium chloride, and then re-equilibrated to PBS.

Creation of Alignment and Calculation of pI Using GCG and Vector NTI

NCBI BLAST searches at http://www.ncbi.nlm.nih.gov/BLAST, GCG searches (GCG, 1994), searches of databases for sequencing of particular genomes, and literature searches were used to identify and download clathrin fragments in various species (Chapter Four). Some sequence data used in this analysis were produced by the Brugia malayi Sequencing and Mapping Project, the Aspergillus fumigatus sequencing project, the Entamoeba histolytica Whole Genome Shotgun, the Plasmodium falciparum Genome Project, the Theileria annulata Sequencing Projects, the Toxoplasma gondii Sequencing Project, the

Trypanosoma brucei Genome Project, the Leishmania major Friedlin Genome Project, and the C. albicans pilot project at the Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens, and some sequences were from the Virginia Commonwealth University/ Tufts University School of Veterinary Medicine Cryptosporidium parvum Genome Sequencing Project at http://www.parvum.mic.vcu.edu/, used with permission from Dr. Gregory Buck. For the plant sequences, the Plant Genome Information Resource (PGDIC) WWW site, from the National Agricultural Library of the US Department of Agriculture – ARS at http://www.nal.usda.gov/pgdic/, and the NCBI Plant Genomes Central at http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/PlantList.html were useful. Sequence data for many parasites and plants were obtained from The Institute for Genomic Research website at http://www.tigr.org. DNA and protein accession codes are noted in Table 4.5 (clathrin heavy chain) and Table 4.6 (clathrin light chain).

Alignments were made using the UCSF Computer Graphics Laboratory (CGL) access to GCG (GCG, 1994) PileUp or LineUp and Vector NTI Suite 7.0 (Infor Max Inc.) AlignX, both of which are based on ClustalX (Thompson et al., 1997). Alignments were then exported in MSF format and imported and converted to NEXUS format for use with PAUP*4.0 (Sinauer Associates, 2000) for construction of parsimony and NJ trees and with MrBayes for construction of Bayesian trees (Huelsenbeck and Ronquist, 2001).

GCG (GCG, 1994) was also used to calculate the isoelectric point of various clathrin domains using PeptideMap. VectorNTI (Infor Max Inc.) performs the same function using the Analysis feature in the main protein database.

Creation of Phylogenetic Trees

Creation of phylogenetic trees for clathrin heavy chain and light chain (Chapter Four) followed the guidelines set forth by Hall (Hall, 2001). From PAUP the alignment was exported in NEXUS format and a script was added for construction of Bayesian trees using the MrBayes program (Huelsenbeck and Ronquist, 2001).

Trees were generated using various protists or fungi as the outgroup, with *N. crassa* working best for the light chain tree and *C. albicans* , *T. brucei,* or *P. brasilensis* for the heavy chain tree. MrBayes (Huelsenbeck and Ronquist, 2001) was then used to infer the ancestral sequence at specified nodes of the trees. Parsimony trees via a heuristic search from PAUP and Bayesian trees from MrBayes gave similar results, while NJ trees frequently failed to categorize known sequences appropriately, presumably because of the large number of incomplete sequences and its reliance on a pairwise comparison. The consensus tree (majority rule) was used for presentation from parsimony or Bayesian output.

Bayesian Algorithms

MrBayes was used to create a Bayesian phylogenetic tree for clathrin heavy chain and clathrin light chain (Chapter Four). Scripts and parameters

were used for the Bayesian algorithm using MrBayes (Huelsenbeck and Ronquist, 2001). Bayesian algorithms are much like maximum likelihood trees except that it uses posterior probabilities and finds the best set of trees rather than a single tree (Mau et al., 1999; Rannula and Yang, 1996). The program uses the metropolis-coupled Markov Chain Monte Carlo method.

nst=6 sets the number of substitution types according to GTR (General Time Reversible), a general model of DNA substitution. **Rates** sets the model for among-site rate variation. **Equal** is no rate variation, while **gamma** is a gamma-distributed rate variation across sites. Other options include **sitespec** for site specific variation or **covarion** for the covarion model. **nchains** sets the number of simultaneous chains running. **ngen** is the number of generations the program runs for, and the current tree is saved to the file every **samplefreq** generations. **Sumt** summarizes the results into a tree. Because the program converges on a solution slowly, the first trees are discarded. The number discarded is set by the **burnin** parameter. **Enforcecon** enforces the constraints set by the user to group species together and **inferanc** allows inferance of the ancestral sequences for each constraint. For determination of trees, the following lines were included at the beginning and end of a NEXUS formatted interleaved alignment (some of these are standard for NEXUS files and others particular to MrBayes):

Before the alignment the file begins with

#NEXUS

BEGIN DATA;

dimensions ntax={insert number of sequences} nchar={length of alignment in characters};

format datatype=PROTEIN missing=X interleave=yes gap=.;

matrix

After the alignment the file ends with

;

end;

begin mrbayes;

outgroup {insert name of outgroup species here};

lset nst=6 rates=gamma;

set autoclose=yes;

mcmcp ngen=150000 printfreq=100 samplefreq=100 nchains=4 savebriens=yes

filename=LCp;

mcmc;

sumt filename=LCp.t burnin=150;

end;

In order to infer ancestral sequences the file begins the same but ends with a

different algorithm as follows:

;

end;

begin mrbayes;

outgroup {insert outgroup species name here];

usertree={paste usertree from LCp.t file above, the format looks like lots of numbers and parenthesis};

constraint con1={list all CHC22 or LCb sequences together from a clade here. The program will infer the node where this branch radiated from the main tree};

lset nst=6 aamodel=GTR rates=equal enforcecon=yes inferanc=yes;

set autoclose=yes;

mcmcp ngen=150000 printfreq=100 samplefreq=100

nchains=4 savebrlens=yes startingtree=user filename=Lcinfer;

mcmc;

sumt filename=Lcinfer.t burnin=150;

end;

when memory constraints were encountered the number of chains was reduced to nchains=2.

Creation of Paralogue Maps

Paralogous regions identified by an automated analysis by Wolfe and colleagues (McLysaght et al., 2002) were used as a starting point for the analysis in Chapter Four. All paralogons noted by this analysis using the ENSEMBL human genome database versions 1.0 and 5.28 were reviewed at the Wolfe website, http://wolfe.gen.tcd.ie/dup. As the human genome sequencing and annotation neared completion, I updated this list using the NCBI human genome webpage http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi and the UCSC Genome page at http://genome.ucsc.edu. A full list of genes surrounding each

clathrin gene was made, and attempts were made to identify structural domains or functional assignments for each. The genes closest to clathrin in the Wolfe analysis were searched by BLAST (Altschul et al., 1997) to determine its function and most closely related family members. Proteins with similar domains or functions between the two chromosomes were then tested as possible paralogons by using NCBI Pairwise BLAST (Tatusova and Madden, 1999) at http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html with a cutoff E-value of 10 to find if the sequence similarity extended over most of the primary sequence of the protein. For some promising paralogue candidates, an alignment and a phylogenetic tree were generated, if a suitable tree did not already exist in literature.

Chapter Six: Conclusions and Perspectives

Twenty-five years after clathrin was first identified as the major component in vesicle coats, many questions remain unanswered about the mechanism of clathrin self-assembly and the functional differences between duplicate clathrin forms.

The molecular mechanism of clathrin self-assembly is still unknown, but in this dissertation the nature of the interactions involved in assembly has been defined. Clathrin heavy chains interact in a coordinated entropy-driven process, which relies on hydrophobic contacts for the affinity between leg segments. The overall electrostatic contribution to clathrin assembly is repulsive, due to the negative charge on both clathrin heavy chain and clathrin light chain. Specificity in clathrin assembly may be contributed by salt bridges, metal ion interactions, or shape and size considerations. Future work will certainly involve new efforts to locate the residues involved in clathrin self-assembly interactions and elucidate the mechanism of this reversible assembly. While a crystal structure of the dimeric interface would be an ideal starting point, dimerization of individual clathrin heavy chain domains has proven impossible (Chapter Three), and efforts to crystallize trimerized heavy chains have not yet resulted in a structure (Ybe et al., 1999-2003). An improved model of the interface from a higher resolution cryo-electronmicroscopy map or improved molecular docking into the existing

map may be sufficient for the selection of residues for future mutagenesis studies.

It would perhaps be prudent to take a step back from our ambitious long-term goals to first identify the interacting face of the proximal leg segment, and only later identify the specific residues involved. Here I consider various methods that might be used for such a study.

Crosslinking combined with mass spectrometry has been used to localize the interface for some interacting proteins (Bennett et al., 2000; Rappsilber et al., 2000). However, because this is a self-assembly interaction, there will be significant difficulty in separating intramolecular crosslinks from intermolecular crosslinks. Moreover, the sheer size of the complex makes the problem less amenable to this approach.

FRET, or fluorescent resonance energy transfer, is another method that may allow identification of the interface. Previously reported experiments indicate that transfer of energy between tryptophan residues and unknown coated vesicle components covalently labeled with fluorescent N-(1-anilinonaphthalene)maleimide increases greatly at assembly pH (Maezawa et al., 1989; Prasad et al., 1984; Yoshimura et al., 1987). Since clathrin proximal leg segment has no surface-accessible sulfhydryl groups, the fluorescent label was likely attached to either a different clathrin domain or to clathrin light chains or other vesicle components. It seems promising that more specifically localized FRET probes conjugated to abundant lysines or other residues on the surface of the proximal leg would be able to accept energy transferred from surface

accessible tryptophans such as W1340, W1358, or W1386. W1340 is located on the opposite side of the molecule from the other two of these. Creation of clathrin heavy chain mutants with a single surface-accessible tryptophan to determine which side of the molecule contributes fluorescence as it approaches its binding partner, or to find one whose fluorescence is quenched in the interface, would unambiguously determine the orientation of proximal legs in the assembled basket. Alternatively, the light chains could be used as an energy donor, since the modeled interface between them is predicted to involve hydrophobic interactions between tryptophans and heavy chain cations (Chen et al., 2002), allowing additional experimental evidence to refine the model of light chains binding to the heavy chain surface. FRET seems like a promising approach for study of this interface.

A novel approach now in development also holds a great deal of promise for characterization of the interface between clathrin heavy chains. Algorithms exist to search the conformational space of two molecules using ambiguous constraints (such as mutagenesis data or NMR peak shifts) such that the interface may be determined (Dominguez et al., 2003). These efforts are being extended to larger molecules in which the NMR peak assignment of residues would be exceedingly difficult (Reese et al., 2003). Since clathrin heavy chain is too large to solve its structure using NMR but resistant to efforts to find smaller dimerizing domains, treating assembled clathrin as a macromolecular complex using this technique explores the contacts in the assembled basket. This approach has the advantage of identifying the face of the interaction and the

specific residues involved in one step. This does not require assumptions about which region of the molecule is involved that are required for targeted mutagenesis. It also removes the intermediate steps between the identification of the geometry of the interface and the assignment of specific residues involved.

Selection of residues for future mutagenesis should take into account the cooperative nature of the interactions between clathrin segments (Chapter Two). Individual salt bridge interactions within a large interface are likely to have a negligible effect on affinity between two proximal leg segments, unless they are critical for specificity or unless they disrupt the fit between the two surfaces. Given the size of the extended interface and the low number of salt bridges typically contributing to a predominantly hydrophobic interface, this seems a slow way to locate the critical residues in the interface. A better approach for mutagenesis might be to mutate the surface-accessible histidine residues. Whether this answers the intriguing questions about the role of histidines in the pH sensitivity of the assembly process and their possible interactions with metal ions or not, it would be a good method to eliminate a large hydrophobic patch from the surface of the molecule, and its replacement with a smaller residue would certainly change the surface shape complementarity, resulting in a greater decrease in affinity between proximal legs than most salt bridge mutations. The elucidation of the molecular mechanism of clathrin assembly will lead naturally to studies to determine how cellular co-factors utilize this mechanism to regulate clathrin self-assembly in-vivo.

In this era of genome sequencing, the function of CHC22 and the functional differences between LCa and LCb in vertebrates are subjects that can be approached using bioinformatics tools. The preliminary work described in this dissertation has set a groundwork for future study by establishing the time frame of gene duplication and shedding light on the likely mechanism by which the duplications occurred. Completion of the annotation of the human genome will establish with finality the presence or absence of a light chain-like regulatory subunit for CHC22. The human genome project may also lead to the discovery of additional pseudogenes in the human genome, ancestral copies of clathrin genes that were mutated and inactivated over evolutionary time. Finally, with complete genome annotation, the genes flanking the clathrin genes in the human genome can be fully characterized to more firmly establish whether clathrin heavy chain and light chain duplicated by a large scale event such as en-bloc genome duplication. Zebrafish, which are known to have both light chain genes and are likely to contain two heavy chain genes, may be an ideal organism for the cell biology study of functional differences between duplicated genes. Bioinformatics coupled with biochemistry and cell biology will lead to a deeper understanding of how clathrins have specialized over evolutionary time for enhanced function.

Efforts to further narrow down the time frame of clathrin duplication or to prove conclusively that clathrin genes duplicated in a large-scale duplication event such as genome duplication are ambitious projects. To narrow down the timeframe from the rather wide 750 to 400 million years ago to something more

specific would re

the cephalochord

lampreys, hagfis

are a subject of

whole genome s

require cloning a

species. A simi

genome duplica

paralogues is n

Wolfe and colle

genome is fully

that the autom

the analysis do

trees for the p

potential para

family membe

The m

just beginnin

a well-known

mechanistic

assembly. I

manipulate

insights tha

specific would require determining the clathrin sequences from vertebrates from the cephalochordata, agnatha, and chondrichthyes branches, such as lancelets, lampreys, hagfish, sharks, skates, rays, or dogfish sharks. While the lancelets are a subject of considerable interest, none of these organisms currently have whole genome sequencing projects underway. Thus, refining the timeline would require cloning and sequencing of clathrin CHC17 and CHC22 from these species. A similar study of other proteins generated significant evidence of genome duplication (Abi-Rached et al., 2002). A thorough analysis of potential paralogues is not possible at this time. It is likely that the automated analysis by Wolfe and colleagues (McLysaght et al., 2002) will be repeated once the human genome is fully annotated. Then further studies would be appropriate to ensure that the automated analysis did not miss obvious proteins of interest, repeating the analysis done in this paper and extending it to generate full phylogenetic trees for the protein family of each paralogue and determining whether the potential paralogues lie on closer branches of the tree than any other protein family member in the same organism.

The molecular era of endocytosis has begun and the bioinformatics age is just beginning, offering powerful new tools to dissect the molecular machinery of a well-known process. I anticipate that in the near future, clathrin will reveal its mechanistic and functional secrets and allow exploration of the regulation of assembly. I am hopeful that in the future researchers will find ways to manipulate the regulation of clathrin assembly in-vivo, and that this might lead to insights that can be applied therapeutically.

References

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. and Inoko, H. (2002) Evidence of *en bloc* duplication in vertebrate genomes. *Nat. Genet.*, **31**, 100-105.

Acton, S. and Brodsky, F.M. (1990) Predominance of clathrin light chain LCb correlates with the presence of a regulated secretory pathway. *J. Cell Biol.*, **111**, 1419-1426.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., et.al. and Venter, J.C. (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185-2195.

(AGI), T.A.G.I. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**, 796-815.

Agulnik, S.I., Papaioannou, V.E. and Silver, L.M. (1998) Cloning, Mapping, and Expression Analysis of TBX15, a New Member of the T-Box Gene Family. *Genomics*, **51**, 68-75.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) *Molecular biology of the cell.* Garland Science, New York.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

Anderson, C.L., Monni, O., Wagner, U., Kononen, J., Barlund, M., Bucher, C., Haas, P., Nocito, A., Bissig, H., Sauter, G. and Kallioniemi, A. (2002)

High-throughput copy number analysis of 17q23 in 3520 tissue specimens by fluorescence *in situ* hybridization to tissue microarrays. *Am. J. Path.*, **161**, 73-79.

Anderson, R.G. (1998) The caveolae membrane system. *Annu Rev Biochem*, **67**, 199-225.

Anderson, R.G.W., Goldstein, J.L. and Brown, M.S. (1976) Localization of low density lipoprotein receptors on plasma membrane of normal human fibroblasts and their absence in cells from a familial hypercholesterolemia homozygote. *Proc. Natl. Acad. Sci. USA*, **73**, 2434-2438.

Angle, B., Yen, F., Hersh, J.H. and Gowans, G. (2003) Patient with terminal duplication 3q and terminal deletion 5q: comparison with the 3q duplication syndrome and distal 5q deletion syndrome. *Am. J. Med. Genet.*, **116A**, 376-380.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., et.al. and Brenner, S. (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science*, **297**, 1301-1310.

Baggett, J.J. and Wendland, B. (2001) Clathrin function in yeast endocytosis. *Traffic*, **2**, 297-302.

Barlund, M., Monni, O., Kononen, J., Cornelison, R., Torhorst, J., Sauter, G., Kallioniemi, O.-P. and Kallioniemi, A. (2000) Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res.*, **60**, 5340-5344.

Bell, C.J., Dixon, R.A., Farmer, A.D., Flores, R., Inman, J., Gonzales, R.A., Harrison, M.J., Paiva, N.L., Scott, A.D., Weller, J.W. and May, G.D. (2001) The Medicago Genome Initiative: a model legume database. *Nucleic Acids Res*, **29**, 114-117.

Bellingham, C.M., Woodhouse, K.A., Robson, P., Rothstein, S.J. and Keeley, F.W. (2001) Self-aggregation characteristics of recombinantly expressed human elastin polypeptides. *Biochim Biophys Acta*, **1550**, 6-19.

Bennett, K.L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P. and Roepstorff, P. (2000) Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein Sci*, **9**, 1503-1518.

Blackbourn, M.E., Anderton, B.H. and Jackson, A.P. (1996) Plant clathrin heavy-chain: sequence analysis and restricted distribution in rapidly growing pollen tubes. *J. Cell Sci.*, **109**, 777-787.

Blank, G.S. and Brodsky, F.M. (1986) Site-specific disruption of clathrin assembly produces novel structures. *EMBO J.*, **5**, 2087-2095.

Blewitt, M., Matz, E.C., Davy, D.F. and Burr, B. (1999) ESTs from developing cotton fiber. *Unpublished.*

Blobel, G. (2000) Protein targeting. *Biosci Rep*, **20**, 303-344.

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., et.al. and Hubbard, S.J. (2002) A Comprehensive Collection of Chicken cDNAs. *Curr Biol*, **12**, 1965-1969.

Boehm, M. and Bonifacino, J.S. (2001) Adaptins: the final recount. *Mol Biol Cell*, **12**, 2907-2920.

Bohnert, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., et.al. and Zepeda, G.R. (2000) Functional Genomics of Plant Stress Tolerance. *Unpublished.*

Borza, D.B., Tatum, F.M. and Morgan, W.T. (1996) Domain structure and conformation of histidine-proline-rich glycoprotein. *Biochemistry*, **35**, 1925-1934.

Botta, A., Lindsay, E.A., Jurecic, V. and Baldini, A. (1997) Comparative mapping of the DiGeorge syndrome region in mouse shows inconsistent gene order and differential degree of gene conservation [published erratum appears in Mamm Genome 1998 Apr;9(4):344]. *Mamm Genome*, **8**, 890-895.

Bridgland, L., Footz, T.K., Kardel, M.D., Riazi, M.A. and McDermid, H.E. (2003) Three duplicons form a novel chimeric transcription unit in the pericentromeric region of chromosome 22q11. *Hum. Genet.*, **112**, 57-61.

Brodsky, F.M. (1988) Living with clathrin: Its role in intracellular membrane traffic. *Science*, **242**, 1396-1402.

Brodsky, F.M. (1997) New fashions in vesicle coats. *Trends Cell Biol.*, **7**, 175-179.

Brodsky, F.M., Chen, C.Y., Knuehl, C., Towler, M.C. and Wakeham, D.E. (2001) Biological basket weaving: formation and function of clathrin-coated vesicles. *Annu Rev Cell Dev Biol*, **17**, 517-568.

Brodsky, F.M., Galloway, C.J., Blank, G.S., Jackson, A.P., Seow, H.F., Drickamer, K. and Parham, P. (1987) Localization of clathrin light-chain

sequences mediating heavy-chain binding and coated vesicle diversity. *Nature*, **326**, 203-205.

Brodsky, F.M., Hill, B.L., Acton, S.L., Näthke, I., Wong, D.H., Ponnambalam, S. and Parham, P. (1991) Clathrin light chains: Arrays of protein motifs that regulate coated vesicle dynamics. *Trends Biol. Sci.*, **16**, 208-213.

Brodsky, F.M. and Parham, P. (1983) Polymorphism in clathrin light chains from different tissues. *J. Mol. Biol.*, **167**, 197-204.

Brown, C.M. and Petersen, N.O. (1999) Free clathrin triskelions are required for the stability of clathrin- associated adaptor protein (AP-2) coated pit nucleation sites. *Biochem Cell Biol*, **77**, 439-448.

Cantor, C.R. and Schimmel, P.R. (1980a) *The behavior of biological macromolecules.* W. H. Freeman & Co.

Cantor, C.R. and Schimmel, P.R. (1980b) *The conformation of biological macromolecules.* W. H. Freeman and Company.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., et.al. and Carucci, D.J. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature*, **419**, 512-519.

Carninci, P. and Hayashizaki, Y. (1999) High-efficiency full-length cDNA cloning. *Methods Enzymol*, **303**, 19-44.

Cazzaniga, G., Daniotti, M., Tosi, S., Giudici, A., Aloisi, E., Pogliani, L., Kearney, A. and Biondi, A. (2001) The paired box domain gene PAX5 is fused to

ETV6/TEL in an acute lymphoblastic leukemia case. *Cancer Res.*, **61**, 4666-4670.

Celera. (2002) Direct submission. Unpublished work. The Anopheles Genome Sequencing Consortium, Celera Genomics.

Chen, C.Y. and Brodsky, F.M. (2001) Unpublished results.

Chen, C.Y., Reese, M.L., Hwang, P.K., Ota, N., Agard, D. and Brodsky, F.M. (2002) Clathrin light and heavy chain interface: alpha-helix binding superhelix loops via critical tryptophans. *Embo J*, **21**, 6072-6082.

Christianson, D.W. (1991) Structural biology of zinc. *Adv Protein Chem*, **42**, 281-355.

Christianson, D.W. and Alexander, R.S. (1989) Carboxylate-Histidine-Zinc Interactions in Protein Structure. *J Am Chem Soc*, **111**, 6412-6419.

Chu, D.S., Pishvaee, B. and Payne, G.S. (1999) A modulatory role for clathrin light chain phosphorylation in Golgi membrane protein localization during vegetative growth and during the mating response of *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, **10**, 713-726.

Clague, M.J. (2002) Membrane transport: a coat for ubiquitin. *Curr Biol*, **12**, R529-531.

Clark, G., Tannich, E. and Hall, N. (2003) Unpublished data. These sequence data were produced by the Entamoeba histolytica Whole Genome Shotgun Sequencing Group at the Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/".

Coe, E., Cone, K., McMullen, M., Chen, S.S., Davis, G., Gardiner, J., et.al. and Wing, R. (2002) Access to the maize genome: an integrated physical and genetic map. *Plant Physiol*, **128**, 9-12.

Coleman, J., Eaton, S., Merkel, G., Skalka, A.M. and Laue, T. (1999) Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation. *J Biol Chem*, **274**, 32842-32846.

Conibear, E. and Stevens, T.H. (1998) Multiple sorting pathways between the late Golgi and the vacuole in yeast. *Biochimica et Biophysica Acta*, **1404**, 211-230.

Cresswell, P. (1996) Invariant chain structure and MHC class II function. *Cell*, **84**, 505-507.

Crowther, R.A. and Pearse, B.M.F. (1981) Assembly and packing of clathrin into coats. *J. Cell Biol.*, **91**, 790-797.

Dabrowski, M.J., Yanchunas, J., Jr., Villafranca, B.C., Dietze, E.C., Schurke, P. and Atkins, W.M. (1994) Supramolecular self-assembly of glutamine synthetase: mutagenesis of a novel intermolecular metal binding site required for dodecamer stacking. *Biochemistry*, **33**, 14957-14964.

de los Reyes, B.G., McGrath, J.M. and Myers, S. (2001) Differential gene expression in sugar beet seedlings (Beta vulgaris) germinated under stress conditions. *Unpublished.*

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., et.al. and Rokhsar, D.S. (2002) The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science*, **298**, 2157-2167.

DeLuca-Flaherty, C., McKay, D.B., Parham, P. and Hill, B.L. (1990) Uncoating protein (hsc70) binds a conformationally labile domain of clathrin light chain LC_a to stimulate ATP hydrolysis. *Cell*, **62**, 875-887.

Demeler, B. and Saber, H. (1998) Determination of molecular parameters by fitting sedimentation data to finite-element solutions of the Lamm equation. *Biophys J*, **74**, 444-454.

Denning, D.W., Anderson, M.J., Turner, G., Latge, J.P. and Bennett, J.W. (2002) Sequencing the Aspergillus fumigatus genome. *Lancet Infect Dis*, **2**, 251-253.

Dodge, G.R., Kovalszky, I., McBride, O.W., Yi, H.J., Chu, M., saitta, B., Stokes, D.G. and Ioozo, R.V. (1991) Human clathrin heavy chain (CLTC): Partial molecular cloning, expression and mapping of the gene to human chromosome 17q11-qter. *Genomics*, **11**, 174-178.

Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J Am Chem Soc*, **125**, 1731-1737.

Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Changanti, R.S.K., Magenis, E., Shprintzen, R.J. and Morrow, B.E. (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Gen.*, **8**, 1157-1167.

Eisenberg, I., Avidan, N., Potikha, T., Hochner, H., Chen, M., Olender, T., Barash, M., Shemesh, M., Sadeh, M., Grabov-Nardini, G., Shmilevich, I., Friedmann, A., Karpati, G., Bradley, W.G., Baumbach, L., Lancet, D.,

Asher, E.B., Beckmann, J.S., Argov, Z. and Mitrani-Rosenbaum, S. (2001) The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat Genet*, **29**, 83-87.

Eisenberg, I., Hochner, H., Levi, T., Yelin, R., Kahan, T. and Mitrani-Rosenbaum, S. (2002a) Cloning and conservation of a novel human gene RNF38 encoding a conserved putative protein with a RING finger domain. *Biochem. Biophys. Res. Comm.*, **294**, 1169-1176.

Eisenberg, I., Hochner, H., Sadeh, M., Argov, Z. and Mitrani-Rosenbaum, S. (2002b) Establishment of the genomic structure and identification of thirteen single-nucleotide polymorphisms in the human RECK gene. *Cytogenet. Genome Res.*, **97**, 58-61.

Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Edwards, Y.J., Bouchireb, N., Cottage, A., Yeo, G.S., Umrania, Y., Williams, G. and Brenner, S. (1999) Generation and analysis of 25 Mb of genomic DNA from the pufferfish Fugu rubripes by sequence scanning. *Genome Res*, **9**, 960-971.

Estivill, X., Cheung, J., Pujana, M.A., Nakabayashi, K., Scherer, S.W. and Tsui, L.C. (2002) Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. *Hum. Mol. Gen.*, **11**, 1987-1995.

Felipe, M.S., Andrade, R.V., Petrofeza, S.S., Maranhao, A.Q., Torres, F.A., et.al. and Brigido, M.M. (2003) Transcriptome characterization of the dimorphic

and pathogenic fungus Paracoccidioides brasiliensis by EST analysis. *Yeast*, **20**, 263-271.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M. and Winter, G. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, **314**, 235-238.

Feuermann, M., Simeonova, L., Souciet, J.L. and Potier, S. (1997) Analysis of 21.7 kb DNA sequence from the left arm of chromosome VII reveals 11 open reading frames: two correspond to new genes. *Yeast*, **13**, 475-477.

Flood, G., Rowe, A.J., Critchley, D.R. and Gratzer, W.B. (1997) Further analysis of the role of spectrin repeat motifs in alpha-actinin dimer formation. *Eur Biophys J*, **25**, 431-435.

Floyd, S. and De Camilli, P. (1998) Endocytosis proteins and cancer: a potential link? *Trends in Cell Biology*, **8**, 299-301.

Ford, M.G., Mills, I.G., Peter, B.J., Vallis, Y., Praefcke, G.J., Evans, P.R. and McMahon, H.T. (2002) Curvature of clathrin-coated pits driven by epsin. *Nature*, **419**, 361-366.

Ford, M.G., Pearse, B.M., Higgins, M.K., Vallis, Y., Owen, D.J., Gibson, A., Hopkins, C.R., Evans, P.R. and McMahon, H.T. (2001) Simultaneous binding of PtdIns(4,5)P2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science*, **291**, 1051-1055.

Freire, E. and Coelho-Sampaio, T. (2000) Self-assembly of laminin induced by acidic pH. *J Biol Chem*, **275**, 817-822.

Fujita, T., Ralston, G.B. and Morris, M.B. (1998) Purification of erythrocyte spectrin alpha- and beta-subunits at alkaline pH and structural and hydrodynamic properties of the isolated subunits. *Biochemistry*, **37**, 272-280.

Furlong, R.F. and Holland, P.W.H. (2002) Were vertebrates octoploid? *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **357**, 531-544.

Gaidarov, I., Santini, F., Warren, R.A. and Keen, J.H. (1999) Spatial control of coated-pit dynamics in living cells. *Nat. Cell Biol.*, **1**, 1-7.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., et.al. and Barrell, B. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498-511.

GCG. (1994) GCG Program Manual for the Wisconsin Package. Genetics Computer Group, Madison, Wisconsin.

Geli, M.I. and Riezman, H. (1998) Endocytic internalization in yeast and animal cells: similar and different. *J Cell Sci*, **111**, 1031-1037.

Glockner, G., Eichinger, L., Szafranski, K., Pachebat, J.A., Bankier, A.T., Dear, P.H., Lehmann, R., Baumgart, C., Parra, G., Abril, J.F., Guigo, R., Kumpf, K., Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A. and Noegel, A.A. (2002) Sequence and analysis of chromosome 2 of Dictyostelium discoideum. *Nature*, **418**, 79-85.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes,

H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. *Science*, **274**, 546, 563-547.

Goh, C.S. and Cohen, F.E. (2002) Co-evolutionary analysis of interacting proteins reveals insights into protein-protein interactions. *J. Mol. Biol.*, **324**, 177-192.

Goldstein, J.L., Anderson, R.G.W. and Brown, M.S. (1979) Coated pits, coated vesicles, and receptor mediated endocytosis. *Nature*, **279**, 679-685.

Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., et.al. and Hyman, A.A. (2000) Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. *Nature*, **408**, 331-336.

Gong, W., Emanuel, B.S., Collins, J., Kim, D.H., Wang, Z., Chen, F., Zhang, G., Roe, B. and Budarf, M.L. (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum. Mol. Gen.*, **5**, 789-800.

Greene, B., Liu, S.H., Wilde, A. and Brodsky, F.M. (2000) Complete reconstitution of clathrin basket formation with recombinant protein fragments: adaptor control of clathrin self-assembly. *Traffic*, **1**, 69-75.

Grunwald, D.J. and Eisen, J.S. (2002) Headwaters of the zebrafish-- emergence of a new model vertebrate. *Nat. Rev. Genet.*, **3**, 717-724.

Gu, X., Wang, Y. and Gu, J. (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.*, **31**, 205-209.

Hall, B.G. (2001) *Phylogenetic Trees Made Easy*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Hanakam, F., Gerisch, G., Lotz, S., Alt, T. and Seelig, A. (1996) Binding of hisactophilin I and II to lipid membranes is controlled by a pH-dependent myristoyl-histidine switch. *Biochemistry*, **35**, 11036-11044.

Hansen and Garner. (1987) *Prog Biophys Molec Biol*, **50**, 47.

Harris, T.W., Lee, R., Schwarz, E., Bradnam, K., Lawson, D., Chen, W., et.al.and Stein, L.D. (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res*, **31**, 133-137.

Harrison, P.M. and Gerstein, M. (2002) Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. *J Mol Biol*, **318**, 1155-1174.

Harrison, S.C. and Kirchhausen, T. (1983) Clathrin, cages, and coated vesicles. *Cell*, **33**, 650-652.

Helenius, A., Kartenbeck, J., Simons, K. and Fries, E. (1980) On the entry of Semiliki forest virus into BHK-21 cells. *J. Cel Biol.*, **84**, 404-420.

Heuser, J. (1989) Effects of cytoplasmic acidification on clathrin lattice morphology. *J. Cell Biol.*, **108**, 401-411.

Heuser, J.E., Keen, J.H., Amende, L.M., Lippoldt, R.E. and Prasad, K. (1987) Deep-etch visualization of 27S clathrin: a tetrahedral tetramer. *J. Cell Biol.*, **105**, 1999-2009.

Hill, B.L., Drickamer, K., Brodsky, F.M. and Parham, P. (1988) Identification of the phosphorylation sites of clathrin light chain LC_b. *J. Biol. Chem.*, **263**, 5499-5501.

Hinshaw, J.E. (2000) Dynamin and Its Role in Membrane Fusion. *Ann. Rev. Cell. Dev. Biol.*, **16**, 483 -520.

Holland, P.W.H. (2002) Ciona. *Curr. Biol.*, **12**, R609.

Holmes, S.E., Riazi, M.A., Gong, W., McDermid, H.E., Sellinger, B.T., Hua, A., Chen, F., Wang, Z., Zhang, G., Roe, B., Gonzalez, I., McDonald-McGinn, D.M., Zackai, E., Emanuel, B.S. and Budarf, M.L. (1997) Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. *Hum Mol Genet*, **6**, 357-367.

Horvath, J.E., Schwartz, S. and Eichler, E.E. (2000) The Mosaic Structure of Human Pericentromeric DNA: A Strategy for Characterizing Complex Regions of the Human Genome. *Genome*, **10**, 839-852.

Hu, Y., Barzilai, A., Chen, M., Bailey, C.H. and Kandel, E.R. (1993) 5-HT and cAMP induce the formation of coated pits and vesicles and increase the expression of clathrin light chain in sensory neurons of aplysia. *Neuron*, **10**, 921-929.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.

Jackson, A.P. and Parham, P. (1988) Structure of human clathrin light chains. Conservation of light chain polymorphism in three mammalian species. *J Biol Chem*, **263**, 16688-16695.

Jackson, A.P., Seow, H.-F., Holmes, N., Drickamer, K. and Parham, P. (1987) Clathrin light chains contain brain-specific insertion sequences and a region of homology with intermediate filaments. *Nature*, **326**, 154-159.

Jasanoff, A., Wagner, G. and Wiley, D.C. (1998) Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein Ii. *Embo J*, **17**, 6812-6818.

Jenkins, C., Samudrala, R., Anderson, I., Hedlund, B.P., Petroni, G., Michailova, N., Pinel, N., Overbeek, R., Rosati, G. and Staley, J.T. (2002) Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 17049-17054.

Jin, A.J. and Nossal, R. (1993) Topological mechanisms involved in the formation of clathrin-coated vesicles. *Biophys J*, **65**, 1523-1537.

Jin, A.J. and Nossal, R. (2000) Rigidity of triskelion arms and clathrin nets. *Biophys J*, **78**, 1183-1194.

Johnson, M.L., Correia, J.J., Yphantis, D.A. and Halvorson, H.R. (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least- squares techniques. *Biophys J*, **36**, 575-588.

Kaneseki, T. and Kadota, K. (1969) The 'vesicle in a basket.' A morphological study of the coated vesicle isolated from the nerve endings of the guinea

pig brain, with special reference to the mechanism of membrane movements. *J. Cell Biol.*, **42**, 202-220.

Kartmann, B. and Roth, D. (2000) Novel roles for mammalian septins: from vesicle trafficking to oncogenesis. *J. Cell Sci.*, **114**, 839-844.

Kasahara, M., Yawata, M. and Suzuki, T. (1999) The MHC Paralogous Group: listing of members and a brief overview. In Kasahara, M. (ed.), *Major Histocompatibility Complex: Evolution, Structure, and Function*. Springer-Verlag Tokyo, Hong Kong, pp. 27-44.

Kedra, D., Peyrard, M., Fransson, I., Collins, J.E., Dunham, I., Roe, B.A. and Dumanski, J.P. (1996) Characterization of a second human clathrin heavy chain polypeptide gene (*CHL-22*) from chromosome 22q11. *Hum. Mol. Gen.*, **5**, 625-631.

Keen, J.H., Willingham, M.C. and Pastan, I.H. (1979) Clathrin-coated vesicles: Isolation, dissociation, and factor-dependent reassociation of clathrin baskets. *Cell*, **16**, 303-312.

Kentsis, A., Gordon, R.E. and Borden, K.L. (2002) Self-assembly properties of a model RING domain. *Proc Natl Acad Sci U S A*, **99**, 667-672.

Kikuno, R., Nagase, T., Ishikawa, K., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. (1999) Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.*, **6**, 197-205.

Kirchhausen, T. (2000a) Clathrin. *Annu. Rev. Biochem.*, **69**, 699-727.

Kirchhausen, T. (2000b) Three ways to make a vesicle. *Nat. Rev. Mol. Cell Biol.*, **1**, 187-198.

Kirchhausen, T., Harrison, S.C., Chow, E.P., Mattaliano, R.J., Ramachandran, K.L., Smart, J. and Brosius, J. (1987a) Clathrin heavy chain: Molecular cloning and complete primary sequence. *Proc. Natl. Acad. Sci. USA*, **84**, 8805-8809.

Kirchhausen, T., Harrison, S.C. and Heuser, J. (1986) Configuration of clathrin trimers: Evidence from electron microscopy. *J. Ultrastructure and Molec. Structure Res.*, **94**, 199-208.

Kirchhausen, T., Scarmato, P., Harrison, S.C., Monroe, J.J., Chow, E.P., Mattaliano, R.J., Ramachandran, K.L., Smart, J.E., Ahn, A.H. and Brosius, J. (1987b) Clathrin light chains LC_a and LC_b are similar, polymorphic, and share repeated heptad motifs. *Science*, **236**, 320-324.

Kirkpatrick, B.W., Byla, B., Kurar, E. and Warren, W.C. (2002) Development of microsatellite markers and comparative mapping for bovine chromosome 19. *An. Gen.*, **33**, 65-68.

Kissinger, J.C., Gajria, B., Li, L., Paulsen, I.T. and Roos, D.S. (2003) ToxoDB: accessing the Toxoplasma gondii genome. *Nucleic Acids Res*, **31**, 234-236.

Klein, J., Sato, A. and Mayer, W.E. (2000) Jaws and AIS. In Kasahara, M. (ed.), *Major Histocompatibility Complex: Evolution, Structure, and Function.* Springer-Verlag Tokyo, Hong Kong, pp. 3-26.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W. and Richardson, P. (2002) Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. *Dev Dyn*, **225**, 384-391.

Kocsis, E., Trus, B.L., Steer, C.J., Bisher, M.E. and Steven, A.C. (1991) Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. *J Struct Biol*, **107**, 6-14.

Kokoza, V.A., Snigirevskaya, E.S. and Raikhel, A.S. (1997) Mosquito clathrin heavy chain: analysis of protein structure and developmental expression in the ovary during vitellogenesis. *Insect Mol Biol*, **6**, 357-368.

Kraulis, P.J. (1991) Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst.*, **24**, 946-950.

Kuliawat, R., Klumperman, J., Ludwig, T. and Arvan, P. (1997) Differential sorting of lysosomal enzymes out of the regulated secretory pathway in pancreatic beta-cells. *J. Cell Biol.*, **137**, 595-608.

Kusunoki, T. and Amemiya, F. (1995) The brain of the Agnatha. *Kaibogaku Zasshi*, **70**, 436-447.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et. al. and Chen, Y.J. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Laporte, J., Blondeau, F., Buj-Bello, A., Tentler, D., Kretz, C., Dahl, N. and Mandel, J.L. (1998) Characterization of the myotubularin dual specificity phosphatase gene family from yeast to human. *Hum. Mol. Gen.*, **7**, 1703-1712.

Lauffer, M.A. (1975) Entropy-driven processes in biology. *Mol Biol Biochem Biophys*, **20**, 1-264.

Lemmon, S.K. (2001) Clathrin uncoating: Auxilin comes to life. *Curr. Biol.*, **11**, R49-R52.

Lemmon, S.K. and Jones, E.W. (1987) Clathrin requirement for normal growth of yeast. *Science*, **238**, 504-509.

Lemmon, S.K., Pellicena-Palle, A., Conley, K. and Freund, C.L. (1991) Sequence of the clathrin heavy chain from Saccharomyces cerevisiae and requirement of the COOH terminus for clathrin function. *J. Cell Biol.*, **112**, 65-80.

Liu, S.-H., Mallet, W.G. and Brodsky, F.M. (2001a) Clathrin-mediated endocytosis. In Marsh, M. (ed.), *Endocytosis*. Oxford University Press, Vol. in press.

Liu, S.-H., Marks, M.S. and Brodsky, F.M. (1998) A dominant-negative clathrin mutant differentially affects trafficking of molecules with distinct sorting motifs in the class II major histocompatibility complex (MHC) pathway. *J. Cell Biol.*, **140**, 1023-1037.

Liu, S.H., Towler, M.C., Chen, E., Chen, C.Y., Song, W., Apodaca, G. and Brodsky, F.M. (2001b) A novel clathrin homolog that co-distributes with cytoskeletal components functions in the trans-Golgi network. *Embo J*, **20**, 272-284.

Liu, S.H., Wong, M.L., Craik, C.S. and Brodsky, F.M. (1995) Regulation of clathrin assembly and trimerization defined using recombinant triskelion hubs. *Cell*, **83**, 257-267.

Long, K.R., Trofatter, J.A., Ramesh, V., McCormick, M.K. and Buckler, A.J. (1996) Cloning and characterization of a novel human clathrin heavy chain gene (CLTCL). *Genomics*, **35**, 466-472.

Lund, J., Chen, F., Hua, A., Roe, B., Budarf, M., Emanuel, B.S. and Reeves, R.H. (2000) Comparative sequence analysis of 634 kb of the mouse chromosome 16 region of conserved synteny with the human velocardiofacial syndrome region on chromosome 22q11.2. *Genomics*, **63**, 374-383.

Maezawa, S., Yoshimura, T., Hong, K., Düzgünes, N. and Papahadjopoulos, D. (1989) Mechanism of protein-induced membrane fusion: fusion of phospholipid vesicles by clathrin associated with its membrane binding and conformational change. *Biochemistry*, **28**, 1422-1428.

Marti, M. and Hehl, A.B. (2002) Putative clathrin heavy chain from Giardia intestinalis. *Unpublished.*

Matthews, D.E., Carollo, V.L., Lazo, G.R. and Anderson, O.D. (2003) GrainGenes, the genome database for small-grain crops. *Nucleic Acids Res*, **31**, 183-186.

Matthews, D.J. (1995) Interfacial metal-binding site design. *Curr Opin Biotechnol*, **6**, 419-424.

Mau, B., Newton, M.A. and Larget, B. (1999) Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics*, **55**, 1-12.

Maxfield, F.R., Schlessinger, J., Shechter, Y., Pastan, I. and Willingham, M.C. (1978) Collection of insulin, EGF and alpha2-macroglobulin in the same patches on the surface of cultured fibroblasts and common internalization. *Cell*, **14**, 805-810.

McLysaght, A., Hokamp, K. and Wolfe, K.H. (2002) Extensive genomic duplication during early chordate evolution. *Nat. Genet.*, **31**, 200-204.

McMahon, H.T. (1999) Endocytosis: an assembly protein for clathrin cages. *Curr. Biol.*, **9**, R332-335.

Meinertzhagen, I.A. and Okamura, Y. (2001) The larval ascidian nervous system: the chordate brain from its small beginnings. *Tr. Neurosci.*, **24**, 401-410.

Meirovitch, H. and Scheraga, H.A. (1980) Empirical Studies of Hydrophobicity. 2. Distribution of the Hydrophobic, Hydrophilic, Neutral, and Ambivalent Amino Acids in the Interior and Exterior Layers of Native Proteins. *Macromolecules*, **13**, 1406-1414.

Mejean, A., Bodenreider, C., Schuerer, K. and Goldberg, M.E. (2001) Kinetic characterization of the pH-dependent oligomerization of R67 dihydrofolate reductase. *Biochemistry*, **40**, 8169-8179.

Merritt, E.A. and Murphy, M.E.P. (1994) Raster3d Version 2.0 - a Program For Photorealistic Molecular Graphics. *Acta Crystallogr. D-Biol. Cryst.*, **50**, 869-873.

Miles, E.W. (1977) Modification of histidyl residues in proteins by diethylpyrocarbonate. *Methods Enzymol*, **47**, 431-442.

Mishra, S.K., Keyel, P.A., Hawryluk, M.J., Agostinelli, N.R., Watkins, S.C. and Traub, L.M. (2002) Disabled-2 exhibits the properties of a cargo-selective endocytic clathrin adaptor. *Embo J*, **21**, 4915-4926.

Misura, K.M., Scheller, R.H. and Weis, W.I. (2000) Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. *Nature*, **404**, 355-362.

Molinete, M., Dupuis, S., Brodsky, F.M. and Halban, P. (2001) Role of clathrin in the regulated secretory pathway of pancreatic beta-cells. *J Cell Sci*, **114**, 3059-3066.

Momany, M., Zhao, J., Lindsey, R. and Westfall, P.J. (2001) Characterization of the *Aspergillus nidulans* Septin *(asp)* Gene Family. *Genetics*, **157**, 969-977.

Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M.L. and Kallioniemi, A. (2001) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5711-5716.

Morgan, G.W., Allen, C.L., Jeffries, T.R., Hollinshead, M. and Field, M.C. (2001) Developmental and morphological regulation of clathrin-mediated endocytosis in Trypanosoma brucei. *J Cell Sci*, **114**, 2605-2615.

Mullet, J.E., Klein, R.R. and Klein, P.E. (2002) Sorghum bicolor - an important species for comparative grass genomics and a source of beneficial genes for agriculture. *Curr Opin Plant Biol*, **5**, 118-121.

Munn, A.L., Silveira, L., Elgort, M. and Payne, G.S. (1991) Viability of clathrin heavy-chain-deficient Saccharomyces cerevisiae is compromised by mutations at numerous loci: implications for the suppression hypothesis. *Mol. Cell. Biol.*, **11**, 3868-3878.

Musacchio, A., Smith, C.J., Roseman, A.M., Harrison, S.C., Kirchhausen, T. and Pearse, B.M.F. (1999) Functional organization of clathrin in coats: Combining electron cryomicroscopy and x-ray crstallography. *Mol. Cell*, **3**, 761-770.

Myler, P.J., Beverley, S.M., Cruz, A.K., Dobson, D.E., Ivens, A.C., McDonagh, P.D., Madhubala, R., Martinez-Calvillo, S., Ruiz, J.C., Saxena, A., Sisk, E., Sunkin, S.M., Worthey, E., Yan, S. and Stuart, K.D. (2001) The Leishmania genome project: new insights into gene organization and function. *Med Microbiol Immunol (Berl)*, **190**, 9-12.

Nakamura, N., Hirata, A., Ohsumi, Y. and Wada, Y. (1997) Vam2/Vps41p and Vam6/Vps39p are components of a protein complex on the vacuolar membranes and involved in the vacuolar assembly in the yeast Saccharomyces cerevisiae. *Journal of Biological Chemistry*, **272**, 11344-11349.

Nandi, P.K. and Edelhoch, H. (1984) The effects of lyotropic (hofmeister) salts on the stability of clathrin coat structure in coated vesicles and baskets. *J. Biol. Chem.*, **259**, 11290-11296.

Nathke, I., Hill, B.L., Parham, P. and Brodsky, F.M. (1990) The calcium-binding site of clathrin light chains. *J Biol Chem*, **265**, 18621-18627.

Nathke, I.S., Heuser, J., Lupas, A., Stock, J., Turck, C.W. and Brodsky, F.M. (1992) Folding and trimerization of clathrin subunits at the triskelion hub. *Cell*, **68**, 899-910.

Newmyer, S.L. and Schmid, S.L. (2001) Dominant-interfering Hsc70 mutants disrupt multiple stages of the clathrin-coated vesicle cycle in vivo. *J. Cell Biol.*, **152**, 607-620.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. and Tabata, S. (1994) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.*, **1**, 27-35.

Nossal, R. (2001) Energetics of clathrin basket assembly. *Traffic*, **2**, 138-147.

Nossal, R. and Zimmerberg, J. (2002) Endocytosis: Curvature to the ENTH Degree. *Curr. Biol.*, **12**, R770-R772.

O'Halloran, T.J. and Anderson, R.G.W. (1992) Characterization of the clathrin heavy chain from *Dictyostelium discoideum*. *DNA Cell Biol.*, **11**, 321-330.

Odorizzi, G., Cowles, C.R. and Emr, S.D. (1998) The AP-3 complex: a coat of many colours. *Trends Cell Biol.*, **8**, 282-288.

Ohno, S. (1970) *Evolution by gene duplication.* Springer, New York.

Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirata, M., Suto, Y., Hirai, M., Terao, K., Suzuki, Y., Sugano, S., Hashimoto, K. and Kususda, J. (2001) Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes. *Gene*, **275**, 31-37.

Parham, P., Brodsky, F.M. and Drickamer, K. (1989) The occurrence of disulphide bonds in purified clathrin light chains. *Biochem J*, **257**, 775-781.

Parker, M.H., Brouillette, C.G. and Prevelige, P.E., Jr. (2001) Kinetic and calorimetric evidence for two distinct scaffolding protein binding populations within the bacteriophage P22 procapsid. *Biochemistry*, **40**, 8962-8970.

Payne, G.S. (1990) Genetic analysis of clathrin function in yeast. *J. Membr. Biol.*, **116**, 93-105.

Payne, G.S., Hasson, T.B., Hasson, M.S. and Schekman, R. (1987) Genetic and Biochemical characterization of clathrin-deficient *Saccharomyces cerevisiae*. *Molec. Cell. Biol.*, **7**, 3888-3898.

Pearse, B.M.F. (1975) Coated vesicles from pig brain: purification and biochemical characterization. *J. Mol. Biol.*, **97**, 93-98.

Pearse, B.M.F. (1976) Clathrin: A unique protein associated with intracellular transfer of membrane by coated vesicles. *Proc. Natl. Acad. Sci. USA*, **73**, 1255-1259.

Pearse, B.M.F. and Bretscher, M.S. (1981) Membrane recycling by coated vesicles. *Ann. Rev. Biochem.*, **50**, 85-101.

Pereira-Leal, J.B. and Seabra, M.C. (2001) Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol*, **313**, 889-901.

Pishvaee, B., Munn, A. and Payne, G.S. (1997) A novel structural model for regulation of clathrin function. *EMBO J.*, **16**, 2227-2239.

Pishvaee, B. and Payne, G.S. (1998) Clathrin coats--threads laid bare. *Cell*, **95**, 443-446.

Pleasure, I.T., Black, M.M. and Keen, J.H. (1993) Valosin-containing protein, VCP, is a ubiquitous clathrin-binding protein. *Nature*, **365**, 459-462.

Pley, U.M., Hill, B.L., Alibert, C., Brodsky, F.M. and Parham, P. (1995) The interaction of calmodulin with clathrin-coated vesicles, triskelions, and light chains. Localization of a binding site. *J Biol Chem*, **270**, 2395-2402.

Ponnambalam, S., Jackson, A.P., LeBeau, M.M., Pravtcheva, D., Ruddle, F.H., Alibert, C. and Parham, P. (1994) Chromosomal location and some structural features of human clathrin light-chain genes. *Genomics*, **24**, 440-444.

Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., et.al. (1998) Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.*, **18**, 345-349.

Prasad, K., Alfsen, A., Lippoldt, R.E., Nandi, P.K. and Edelhoch, H. (1984) Structural characterization of labeled clathrin and coated vesicles. *Arch Biochem Biophys*, **235**, 403-410.

Prevelige, P.E., Jr. (1998) Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotechnol*, **16**, 61-65.

Quackenbush, J., Liang, F., Holt, I., Pertea, G. and Upton, J. (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res*, **28**, 141-145.

Rannula, B. and Yang, Z. (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.*, **43**, 304-311.

Rappsilber, J., Siniossoglou, S., Hurt, E.C. and Mann, M. (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem*, **72**, 267-275.

Raven, P.H. and Johnson, G.B. (1986) *Biology*. Times Mirror/ Mosby College Pub., St. Louis.

Reese, M.L., Lai, H. and Doetsch, V. (2003) Manuscript in preparation.

Rieder, S.E. and Emr, S.D. (1997) A novel RING finger protein complex essential for a late step in protein transport to the yeast vacuole. *Molecular Biology of the Cell*, **8**, 2307-2327.

Robinson, M.S. (1987) Coated vesicles and protein sorting. *J. Cell Sci.*, **87**, 203-204.

Robinson, M.S. and Bonifacino, J.S. (2001) Adaptor-related proteins. *Curr Opin Cell Biol*, **13**, 444-453.

Roseman, A.M. (2000) Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr*, **56**, 1332-1340.

Roth, T.F. and Porter, K.R. (1964) Yolk protein uptake in the oocyte of the mosquito *Aedes aegypti. L. J. Cell Biol.*, **20**, 313-332.

Rothman, J.E. and Schmid, S.L. (1986) Enzymatic recycling of clathrin from coated vesicles. *Cell*, **49**, 5-9.

Rudolf, E., Rudolf, K., Radocha, J., Peychl, J. and Cervinka, M. (2003) The role of intracellular zinc in modulation of life and death of Hep-2 cells. *Biometals*, **16**, 295-309.

Ryan, A.K., Goodship, J.A., Wilson, D.I., N., P., Levy, A., Seidel, H., et.al. and Scambler, P.J. (1997) Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *J. Med. Genet.*, **34**, 798-804.

Sali, D., Bycroft, M. and Fersht, A.R. (1988) Stabilization of protein structure by interaction of alpha-helix dipole with a charged side chain. *Nature*, **335**, 740-743.

Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S. (2001) Structural analysis of a Lotus japonicus genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 mb regions of the genome. *DNA Res*, **8**, 311-318.

Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci*, **20**, 374.

Scheele, U. and Holstein, S.E. (2002) Functional evidence for the identification of an Arabidopsis clathrin light chain polypeptide. *FEBS Lett.*, **514**, 355-360.

Schledzewski, K., Brinkmann, H. and Mendel, R.R. (1999) Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of coatomer COPI. *J. Mol. Evol.*, **48**, 770-778.

Schnizer, R., Van Heeke, G., Amaturo, D. and Schuster, S.M. (1996) Histidine-49 is necessary for the pH-dependent transition between active and inactive states of the bovine F1-ATPase inhibitor protein. *Biochim Biophys Acta*, **1292**, 241-248.

Schreiber, G. and Fersht, A.R. (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol*, **3**, 427-431.

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J*, **78**, 1606-1619.

Schulte, U., Becker, I., Mewes, H.W. and Mannhaupt, G. (2002) Large scale analysis of sequences from Neurospora crassa. *J Biotechnol*, **94**, 3-13.

Schwarz, F.P., Steer, C.J. and Kirchhoff, W.H. (1989) A differential scanning calorimetric study of brain clathrin. *Arch Biochem Biophys*, **273**, 433-439.

Scott, L.M., Mueller, L. and Collins, S.J. (1996) E3, a hematopoietic-specific transcript directly regulated by the retinoic acid receptor alpha. *Blood*, **88**, 2517-2530.

Shaikh, T.H., Kurahashi, H. and Emanuel, B.S. (2001) Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. *Genet Med*, **3**, 6-13.

Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L. and Emanuel, B.S. (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet*, **9**, 489-501.

Shi, X.W., Fitzsimmons, C.J., Genet, C., Prather, R., Whitworth, K., Green, J.A. and Tuggle, C.K. (2001) Radiation hybrid comparative mapping between human chromosome 17 and porcine chromosome 12 demonstrates conservation of gene order. *Anim Genet*, **32**, 205-209.

Silveira, L.A., Wong, D.H., Masiarz, F.R. and Schekman, R. (1990) Yeast clathrin has a distinctive light chain that is important for cell growth. *J. Cell Biol.*, **111**, 1437-1449.

Simons, T.J. (1991) Intracellular free zinc and zinc buffering in human red blood cells. *J Membr Biol*, **123**, 63-71.

Sirotkin, H., Morrow, B., DasGupta, R., Goldberg, R., Patanjali, S.R., Shi, G., Cannizzaro, L., Shprintzen, R., Weissman, S.M. and Kucherlapati, R. (1996) Isolation of a new clathrin heavy chain gene with muscle-specific expression from the region commonly deleted in velo-cardio-facial syndrome. *Hum. Mol. Genet.*, **5**, 617-624.

Smith, C.J., Grigorieff, N. and Pearse, B.M.F. (1998) Clathrin coats at 21 Å resolution: A cellular assembly designed to recycle multiple membrane receptors. *EMBO J.*, **17**, 4943-4953.

Sorkin, A. and Waters, C.M. (1993) Endocytosis of growth factor receptors. *Bioessays*, **15**, 375-382.

Sprague, J., Doerry, E., Douglas, S. and Westerfield, M. (2001) The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. *Nucleic Acids Res*, **29**, 87-90.

Spring, J. (2002) Genome duplication strikes back. *Nat. Genet.*, **31**, 128-129.

Stellwag, E.J. (1999) Hox gene development in fish. *Sem. Cell. Dev. Biol.*, **10**, 531-540.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B. and Lundeberg, J. (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5, 692 expressed sequence tags. *Proc Natl Acad Sci U S A*, **95**, 13330-13335.

Straubinger, R.M., Hong, K., Friend, D.S., Jr. and Papahadjopoulos, D. (1983) Endocytosis of liposomes and intracellular fate of encapsulated molecules: encounter with a low pH compartment after internalization in coated vesicles. *Cell*, **32**, 1069-1079.

Strong, W.B. and Nelson, R.G. (2000) Preliminary profile of the Cryptosporidium parvum genome: an expressed sequence tag and genome survey sequence analysis. *Mol Biochem Parasitol*, **107**, 1-32.

Suhy, D.A., Simon, K.D., Linzer, D.I. and O'Halloran, T.V. (1999) Metallothionein is part of a zinc-scavenging mechanism for cell survival under conditions of extreme zinc deprivation. *J Biol Chem*, **274**, 9183-9192.

Tait, A. and Hall, N. (2003) Unpublished data. These sequence data were produced by the Thelieria annulata Sequencing Group at the Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/".

Tait, E., Simon, M.C., King, S., Brown, A.J., Gow, N.A. and Shaw, D.J. (1997) A Candida albicans genome project: cosmid contigs, physical mapping, and gene isolation. *Fungal Genet Biol*, **21**, 308-314.

Tatusova, T.A. and Madden, T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*, **174**, 247-250.

Taylor, J.S., Van de Peer, Y., Braasch, I. and Meyer, A. (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **356**, 1661-1679.

ter Haar, E., Harrison, S.C. and Kirchhausen, T. (2000) Peptide-in-groove interactions link target proteins to the beta- propeller of clathrin. *Proc. Natl. Acad. Sci. U S A*, **97**, 1096-1100.

ter Haar, E., Musacchio, A., Harrison, S.C. and Kirchhausen, T. (1998) Atomic structure of clathrin: A β propeller terminal domain joins an α zigzag linker. *Cell*, **95**, 563-573.

Tettelin, H., Agostoni Carbone, M.L., Albermann, K., Albers, M., Arroyo, J., Backes, U., Barreiros, T., Bertani, I., Bjourson, A.J., Bruckner, M., Bruschi, C.V., Carignani, G., Castagnoli, L., Cerdan, E., Clemente, M.L., Coblenz, A., Coglievina, M., Coissac, E., Defoor, E., Del Bino, S., Delius, H., Delneri, D., de Wergifosse, P., Dujon, B., Kleine, K. and et al. (1997) The nucleotide sequence of Saccharomyces cerevisiae chromosome VII. *Nature*, **387**, 81-84.

Theologis, A., Ecker, J.R., Palm, C.J., Federspiel, N.A., Kaul, S., et. al. and Davis, R.W. (2000) Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. *Nature*, **408**, 816-820.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876-4882.

Thomsen, P.D., Wintero, A.K. and Fredholm, M. (1998) Chromosomal assignments of 19 porcine cDNA sequences by FISH. *Mamm Genome*, **9**, 394-396.

Towler, M.C., Rahkila, P., Parton, R. and Brodsky, F.M. (2002) A Potential Role for the Second Isoform of Clathrin Heavy Chain in Skeletal Muscle. *ASCB*

Annual Meeting. Molecular Biology of the Cell, San Francisco, Vol. 13, p. 2059.

Tsao, P. and von Zastrow, M. (2000) Downregulation of G protein-coupled receptors. *Curr Opin Neurobiol*, **10**, 365-369.

Uekita, T., Itoh, Y., Yana, I., Ohno, H. and Seiki, M. (2001) Cytoplasmic tail-dependent internalization of membrane-type 1 matrix metalloproteinase is important for its invasion-promoting activity. *J. Cell Biol.*, **155**, 1345-1356.

Ungewickell, E. and Branton, D. (1981) Assembly units of clathrin coats. *Nature*, **289**, 420-422.

Van Jaarsveld, P.P., Nandi, P.K., Lippoldt, R.E., Saroff, H. and Edelhoch, H. (1981) Polymerization of clathrin promoters into basket structures. *Biochemistry*, **20**, 4129-4135.

van Kerkhof, P., Sachse, M., Klumperman, J. and Strous, G.J. (2000) Growth hormone receptor ubiquitination coincides with recruitment to clathrin-coated membrane domains. *J Biol Chem*, **19**, 19.

Vasyukevich, K. and Bazinet, C. (1999) A Drosophila clathrin light-chain gene: sequence, mapping, and absence of neuronal specialization. *DNA Cell Biol*, **18**, 235-241.

Wakeham, D.E., Ybe, J.A., Brodsky, F.M. and Hwang, P.K. (2000) Molecular structures of proteins involved in vesicle coat formation. *Traffic*, **1**, 393-398.

Walsh, S.T., Cheng, H., Bryson, J.W., Roder, H. and DeGrado, W.F. (1999) Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc Natl Acad Sci U S A*, **96**, 5486-5491.

Warner, T.S., Sinclair, D.A., Fitzpatrick, K.A., Singh, M., Devlin, R.H. and Honda, B.M. (1998) The light gene of Drosophila melanogaster encodes a homologue of VPS41, a yeast gene involved in cellular-protein trafficking. *Genome*, **41**, 236-243.

Waterston, R. and Consortium, T.C.e.S. (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. *Science*, **282**, 2012-2018.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., et.al. and Lander, E.S. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520-562.

Wettey, F.R., Howard, J.C. and Jackson, A.P. (2000) A vertebrate cell line conditionally deficient in clathrin heavy-chain expreission: (II) Elimination of clathrin heavy-chain expression inhibits endocytosis and transferrin recycling, but not lysosomal traffic. *Mol. Biol. Cell*, **11**.

Whetten, R., Sun, Y.H., Zhang, Y. and Sederoff, R. (2001) Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Mol Biol*, **47**, 275-291.

Wilbur, J., Hwang, P.K., Fletterick, R.J. and Brodsky, F.M. (2002) Unpublished observations.

Williams, H., Brenner, S. and Venkatesh, B. (2002) Identification and analysis of additional copies of the platelet-derived growth factor receptor and colony stimulating factor1 receptor genes in fugu. *Gene*, **295**, 255-264.

Williams, S.A., Lizotte-Waniewski, M.R., Foster, J., Guiliano, D., Daub, J., Scott, A.L., Slatko, B. and Blaxter, M.L. (2000) The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of Brugia malayi. *Int J Parasitol*, **30**, 411-419.

Willingham, M.C., Maxfield, F.R. and Pastan, I.H. (1979) α_2Macroglobulin binding to the plasma membrane of cultured fibroblasts. *J. Cell Biol.*, **82**, 614-625.

Winkler, F.K. and Stanley, K.K. (1983) Clathrin heavy chain, light chain interactions. *EMBO J.*, **2**, 1393-1400.

Wintero, A.K., Fredholm, M. and Davies, W. (1996) Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones. *Mamm Genome*, **7**, 509-517.

Wolfe, K.H. (2001) Yesterday's Polyploids and the Mystery of Diploidization. *Nat. Rev. Gen.*, **2**, 333-341.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., et. al. and Cerrutti, L. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature*, **415**, 871-880.

Woolford, C.A., Dixon, C.K., Manolson, M.F., Wright, R. and Jones, E.W. (1990) Isolation and characterization of PEP5, a gene essential for vacuolar biogenesis in Saccharomyces cerevisiae. *Genetics*, **125**, 739-752.

Wu, X., Zhao, X., Baylor, L., Kaushal, S., Eisenberg, E. and Greene, L.E. (2001) Clathrin exchange during clathrin-mediated endocytosis. *J Cell Biol*, **155**, 291-300.

Ybe, J.A., Brodsky, F.M., Hofmann, K., Lin, K., Liu, S.H., Chen, L., Earnest, T.N., Fletterick, R.J. and Hwang, P.K. (1999) Clathrin self-assembly is mediated by a tandemly repeated superhelix. *Nature*, **399**, 371-375.

Ybe, J.A., Greene, B., Liu, S.H., Pley, U., Parham, P. and Brodsky, F.M. (1998) Clathrin self-assembly is regulated by three light-chain residues controlling the formation of critical salt bridges. *Embo J*, **17**, 1297-1303.

Ybe, J.A., Hwang, P.K., Tran, T.Q.H., Fletterick, R.J. and Brodsky, F.M. (1999-2003) Unpublished observations.

Ybe, J.A., Wakeham, D.E., Brodsky, F.M. and Hwang, P.K. (2000) Molecular structures of proteins involved in vesicle fusion. *Traffic*, **1**, 474-479.

Yoshimura, T., Kameyama, K., Maezawa, S. and Takagi, T. (1991) Skeletal structure of clathrin triskelion in solution: experimental and theoretical approaches. *Biochemistry*, **30**, 4528-4534.

Yoshimura, T., Maezawa, S. and Hong, K. (1987) Exposure of hydrophobic domains of clathrin in its membrane fusion-inducible pH region. *J. Biochem.*, **101**, 1265-1272.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., et.al.and Yang, H. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science*, **296**, 79-92.

Appendix

Annu. Rev. Cell Dev. Biol. 2001. 17:517–68

BIOLOGICAL BASKET WEAVING: Formation and Function of Clathrin-Coated Vesicles

Frances M. Brodsky[1,2,3], Chih-Ying Chen[3], Christine Knuehl[3], Mhairi C. Towler[3], and Diane E. Wakeham[1,3]

[1]Department of Biopharmaceutical Sciences, [2]Department of Pharmaceutical Chemistry, and [3]The G.W. Hooper Foundation, Department of Microbiology and Immunology, University of California, San Francisco, California; e-mail: fmarbro@itsa.ucsf.edu, chiying@itsa.ucsf.edu, knuehl@itsa.ucsf.edu, mhairi1@itsa.ucsf.edu, and wakeham@itsa.ucsf.edu

Key Words adaptor, endocytosis, secretion, cytoskeleton, sorting

■ **Abstract** There has recently been considerable progress in understanding the regulation of clathrin-coated vesicle (CCV) formation and function. These advances are due to the determination of the structure of a number of CCV coat components at molecular resolution and the identification of novel regulatory proteins that control CCV formation in the cell. In addition, pathways of (*a*) phosphorylation, (*b*) receptor signaling, and (*c*) lipid modification that influence CCV formation, as well as the interaction between the cytoskeleton and CCV transport pathways are becoming better defined. It is evident that although clathrin coat assembly drives CCV formation, this fundamental reaction is modified by different regulatory proteins, depending on where CCVs are forming in the cell. This regulatory difference likely reflects the distinct biological roles of CCVs at the plasma membrane and *trans*-Golgi network, as well as the distinct properties of these membranes themselves. Tissue-specific functions of CCVs require even more-specialized regulation and defects in these pathways can now be correlated with human diseases.

CONTENTS

INTRODUCTION 518
GENERAL PRINCIPLES OF COATED VESICLE FORMATION AND THOSE SPECIFIC TO A CCV 519
STRUCTURE AND BIOCHEMISTRY OF CCV COMPONENTS AND REGULATORS 521
Clathrins 526
Adaptors and Related Proteins 528
Coat Proteins Influencing Assembly and Disassembly 530
Dynamin and Binding Partners 532
MECHANICS OF CCV FORMATION AND MEMBRANE INTERACTIONS 533

1081-0706/01/1115-0517$14.00

Clathrin Assembly by Adaptors and Lattice Rearrangement 533
Lipid Interactions During CCV Formation 536
Role for Dynamin in Vesicle Scission 538
Coat Disassembly 539
INTRACELLULAR LOCALIZATION OF CCV FORMATION AND CARGO INTERACTION 540
Plasma Membrane CCV Nucleation and Receptor Sequestration 540
TGN CCV Nucleation and Receptor Sorting 543
CCV-Cargo Interactions at Other Cellular Membranes 544
RELATIONSHIP OF CCVS TO THE CYTOSKELETON 545
TISSUE-SPECIFIC CCV FUNCTION AND SPECIALIZED CLATHRIN FUNCTION 546
Cell Polarity 546
Neuronal Function 547
Regulated Secretion 548
Immune System Function 548
Muscle Function 549
CCV MALFUNCTION: INHIBITORS, MUTANTS, AND DISEASES 550
Molecular Inhibitors 550
Deletion Mutants in Model Organisms 551
Human Diseases 551

INTRODUCTION

Clathrin-coated vesicles (CCVs) mediate sorting and selective transport of membrane-bound proteins for several pathways of intracellular membrane traffic. They are responsible for receptor-mediated endocytosis (RME) at the plasma membrane (PM) and sorting of proteins at the *trans*-Golgi network (TGN) during the biogenesis of lysosomes and secretory granules. CCVs are the first such transport vesicles for membrane proteins to be identified, and consequently the understanding of their biochemistry and function is quite sophisticated. The dense protein coat of the CCV and its bristle-like morphology was first described by Roth & Porter (1964), who noted that these vesicles appeared to be involved in RME of yolk proteins in mosquito oocytes. Subsequently, the remarkable morphology of the clathrin coat was noted for vesicles purified from pig brain (Kaneseki & Kadota 1969). Clathrin was identified as one of the major coat proteins of CCVs by Pearse (1975), and the clathrate or basket-like appearance of assembled clathrin was recognized in the naming of the protein (Pearse 1975). The other major protein component of CCVs, the adaptors [also called assembly proteins (APs)], were identified by Keen et al. (1979) and characterized biochemically in the early 1980s (reviewed in Pearse & Robinson 1990). Recent major advances in the characterization of CCVs include the determination of the structure of a number of CCV coat components at crystallographic resolution, the identification of novel regulatory proteins that control CCV formation in the cell, and the discovery that phosphorylation and receptor signaling influence CCV formation. In addition, the relationship of CCV

to cellular physiology is becoming better defined. An important role for membrane lipids in CCV formation is emerging, the interplay between CCVs and the actin cytoskeleton is being elucidated, and the intracellular pathways in which CCVs participate are being established through the study of genetic and protein mutations affecting molecules involved in CCV formation. Lastly, as other types of coated membrane vesicles that mediate intracellular transport have become better characterized, common and unique features of membrane vesicle transport systems are emerging, and both the general and specialized roles of CCVs in intracellular transport are better understood. This chapter focuses on the recent advances in understanding CCV formation, function, and regulation, as well as their relationship to cellular physiology and human disease.

GENERAL PRINCIPLES OF COATED VESICLE FORMATION AND THOSE SPECIFIC TO A CCV

It has become clear that CCVs are not unique in their ability to select protein cargo for transport from one membrane to another, nor are CCVs unique in their ability to mediate endocytosis from the PM. Therefore it is important to define CCV-mediated transport pathways relative to other vesicle-mediated membrane transport pathways (Figure 1). Cargo selection for anterograde transport from the endoplasmic reticulum through the Golgi apparatus during protein export is mediated by COPII-coated vesicles. Cargo selection for retrograde transport, returning resident proteins to the endoplasmic reticulum when they escape into the Golgi apparatus, is mediated by COPI-coated vesicles (Barlowe 2000). At least two endocytic pathways are independent of CCVs. Potocytosis of small molecules can be mediated by caveolae associated with cholesterol-rich plasma membrane rafts (Anderson 1998). Endocytosis of a subset of signaling receptors and toxins has been shown to be independent of both caveolae and CCVs (Lamaze et al. 2001, Skretting et al. 1999).

The biochemical principles that control CCV formation and function can be clearly defined. Some apply to the formation of other coated-vesicle transport pathways and some are unique to CCVs. Such a comparison has been made recently in two review articles (Kirchhausen 2000b, Springer et al. 1999). The self-assembling property of the clathrin coat is the key to the ability of CCVs to selectively sequester protein cargo into a membrane vesicle. Clathrin has a triskelion (three-legged pinwheel) shape, and intrinsic to the molecule is its ability to form a polyhedral lattice (reviewed in Brodsky 1988) (Figure 1). During CCV formation, clathrin lattice formation is nucleated on cellular membranes by adaptors (AP1 and AP2), which are drawn into the lattice and trigger CCV formation at the TGN and PM, respectively. In turn adaptors incorporate transmembrane molecules into the lattice by association with the cytoplasmic domains of these molecules (Figure 2). Thus in the simplest conception of CCV formation, the polymerization of clathrin provides the organizing function for protein sorting by concentrating

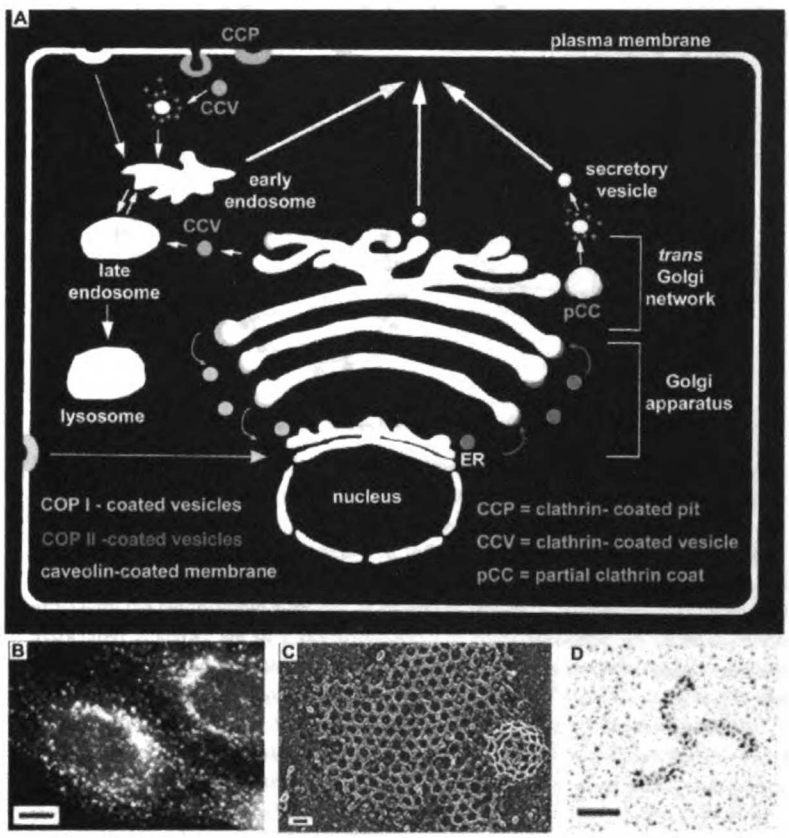


Figure 1 Intracellular location and morphology of clathrin-coated vesicles (CCV). (*A*) Intracellular transport steps mediated by CCV during endocytosis and secretion are delineated in blue, in comparison with the intracellular locations of other transport vesicles (other colors, as defined). (*B*) Distribution of CCV in a HeLa cell labeled with the X22 anti-clathrin heavy chain monoclonal antibody and fluorescent anti-immunoglobulin. (*C*) A membrane-associated clathrin lattice and emerging clathrin-coated pit. [Reproduced with permission from Heuser et al. (1987) and with copyright permission from Rockefeller University Press.] (*D*) A clathrin triskelion purified from bovine brain CCVs and visualized by platinum shadowing. (*B* and *D*) Images from our laboratory [reproduced from (Liu et al. 2001a) with the copyright permission of Oxford University Press]. The bars indicate the following dimensions: (*B*) 5 μm, (*C*) 33 nm, (*D*) 20 nm.

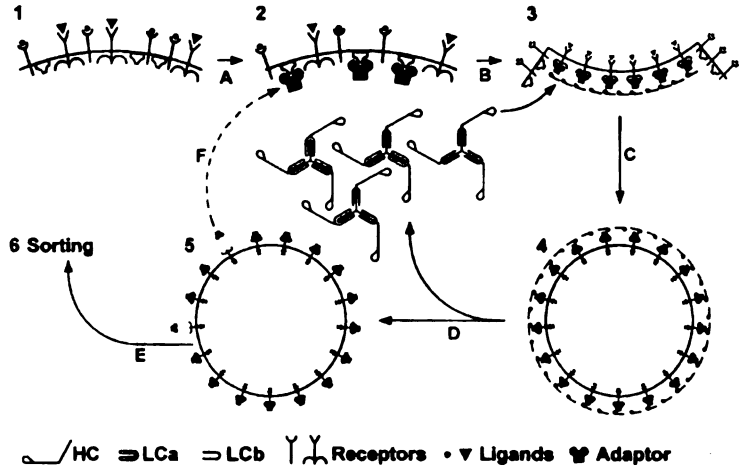


Figure 2 Basic steps in the nucleation, budding, scission, and uncoating of a clathrin-coated vesicle (CCV) from either the plasma membrane (PM) or the *trans*-Golgi network (TGN). (*1*) Receptors with adaptor recognition signals are present in the donor membrane along with determinants that influence adaptor localization. (*2*) Adaptors bind to the localization determinants with high affinity. (*3*) Adaptors interact (low affinity) with the cytoplasmic domains of receptors present in the membrane. Membrane-associated adaptors are in a dephosphorylated form favorable for clathrin binding. Clathrin is recruited from the cystosol, and assembly of the clathrin lattice at the membrane is triggered by adaptors. (*4*) The fully coated CCV detaches from the donor membrane. (*5*) Clathrin is uncoated, followed by uncoating of adaptor molecules. The adaptor molecules in the cytosol are phosphorylated, preventing nonproductive interaction with clathrin. The timing of adaptor phosphorylation within the uncoating process is not defined. (*6*) The uncoated vesicle fuses with an acceptor compartment. In the case of budding from the PM and TGN, the acceptor compartment is in the endocytic pathway. The molecules involved in the regulatory steps indicated by letters are best defined for CCV formation at the PM. In step *A*, receptor uptake can be constitutive or it can be stimulated by signaling. In the latter, phosphorylation, arrestin binding, and ubiquitination can all play a role. Binding of assembly protein AP2 to the PM can be mediated by lipid recognition, synaptotagmin, and stonin 2. Orientation of nucleation relative to the actin cytoskeleton could involve Hip1R. Note that mammalian clathrin is shown with random combinations of light chain (LC)a and LCb. Step *B* involves AP180 or CALM, eps15, and epsin and may incorporate auxilin. Step *C* involves dynamin, amphiphysin, and endophilin, and actin tails may form at this stage via pacsin/syndapin. Intersectin may also be involved in this step. Step *D* involves synaptojanin, hsc70, and auxilin. Step *E* probably involves the Rab5-guanine-nucleotide dissociation inhibitor (GDI) complex, which is also implicated in step *C*. For CCV formation at the TGN, step *A* involves activation of ARF1 and, in some cases, may involve SCAMP recognition. It is not yet established whether for CCV formation at the TGN, steps *B–D* involve regulatory proteins that are the same or equivalent to those described for PM CCV formation.

associated proteins into a regular protein array, and adaptors provide a cargo selection function.

For the other types of coated vesicles that have been defined, the cargo selection function has been attributed to various coat subunits: γ-COP or ARF-GAP (ADP ribosylation factor-GTPase activating protein) for COPI, and Sec23/24p for COPII (Kirchhausen 2000b, Springer et al. 1999). There are also two versions of protein complexes related to CCV adaptors (AP3 and AP4, both present on perinuclear membrane vesicles but not at the PM) that apparently can function independently of clathrin as cargo selectors (Kirchhausen 1999). For these other types of vesicle coats, the organizing function is defined in some cases but not others. Cargo recognition by COPII coats triggers recruitment of the other COPII subunits, which then form a concentrated array of protein, constituting a coat (Springer et al. 1999) and presumably consolidating the cargo and the recognition molecules. For COPI and the clathrin-independent adaptors, what drives coat formation is not completely clear. However, it appears that for COPI, cargo selection and self-assembly are likely to be mediated by the same large protein complex. The clathrin-independent adaptors may self-assemble and select cargo, or they may be organized by proteins not yet identified. For COP coats, cargo recognition is associated with a priming event in which a small GTP-binding protein associated with the target membrane is activated. Assembly of the coat is triggered only when the activated GTP-binding protein interacts with the cargo recognition unit (Sar1p in the case of COPII and ARF1 in the case of COPI) (Kirchhausen 2000b, Springer et al. 1999). This type of priming is also a step in CCV formation at the TGN, where ARF1 plays a role, but such a priming step has not yet been defined for CCV formation at the PM, although an increasing number of regulatory proteins have been implicated in this process. Whether the components of caveolae coats are recruited by priming and cargo recognition is not yet clear, although caveolin appears to self-assemble into a regular array in cholesterol-rich regions of the PM (Anderson 1998). However, cells that lack caveolin can still internalize proteins and lipids from rafts, indicating that caveolin-independent or analogous pathways exist (Garred et al. 2001).

Finally, as more details emerge regarding CCV formation, it has become evident that although clathrin coat assembly drives this process, this fundamental reaction is modified by different regulatory proteins, depending on where CCVs are forming in the cell (Figure 2). In particular, there are a large number of regulatory differences between CCV formation at the PM and at the TGN. This regulatory difference likely reflects the distinct biological roles of CCVs at the PM and TGN, as well as the distinct properties of these membranes themselves.

STRUCTURE AND BIOCHEMISTRY OF CCV COMPONENTS AND REGULATORS

The identification of new proteins involved in CCV formation (Table 1, Figure 2) and the determination of the structures of some CCV components at molecular resolution (Figure 3) have recently moved our understanding of

TABLE 1 Clathrin-coated vesicle components and interacting partners

Proteins[a]	Mammalian (kDa)	Yeast (kDa) and others[b]	Binding partners (domain)[c]	References
Clathrin				
CHC	CHC17 (192)[d]	Chc1p (190): *A.t., C.e., D.d., D.m., G.g.*	AP1β, AP2β, AP3β, AP180, epsin, β-arrestin, synaptojanin (TD); CLC, auxilin, Hsc70, Hip1R, ankyrin, amphiphysin	(Kirchhausen 2000b, Slepnev & De Camilli 2000)
	CHC22 (180)		AP1, AP3	(Liu et al. 2001b)
CLC	LCa/LCb (25–29)[d, e]	Clc1p (36): *A.c., D.d., D.m.*	CHC17	(Brodsky et al. 1991)
Adaptors	AP1 γ(90)	Apl4p (94); *A.t., D.m.*	γ-synergin	(Kirchhausen 1999)
	β1 (100)	Apl2p (82); *A.t., D.m.*	CHC17 TD (clathrin box), CHC22 TD; cargo LL motif	(Rapoport et al. 1998)
	μ1, μ1B (47)	Apm1p (54); *C.e., D.d., D.m.*	Cargo YXXΦ/LL motifs	(F lsch et al. 1999, Hofmann et al. 1999, Ohno et al. 1999, Rapoport et al. 1998)
	σ1 (19)	Aps1p (18); *A.t.*		
	AP2 αA, αC (110)[e]	Apl3p (115); *D.m.*	Synaptojanin, amphiphysin 1, synaptotagmin, Eps15, epsin, AP180, auxilin (ear); phosphoinositides (core domain)	(Gaidarov et al. 1996, Kirchhausen 1999, Owen et al. 1999, Traub et al. 1999)
	β2 (100)	Apl1p (80); *D.m.*	CHC17 TD (clathrin box); Eps15, epsin, AP180, β-arrestin	(Owen et al. 2000)
	μ2 (50)	Apm4p (55); *C.e., D.d., D.m.*	Cargo YXXΦ/LL motifs	(Hofmann et al. 1999, Rapoport et al. 1998)
	σ2 (17)	Aps2p (17); *D.m.*		
	AP3 δ (125)[d]	Apl5p (107); *D.m.*		(Dell'Angelica et al. 1997, Kirchhausen 1999, Simpson et al. 1997)
	β3A, β3B/ βNAP (100),	Apl6p (91)	CHC17 TD (β3A/B clathrin box), CHC22 TD	
	μ3/p47 (47)	Apm3p (55), *D.d., D.m.*		
	σ3A, σ3B (22)	Aps3p (22); *Z.m.*		
	AP4 ε (127)[d], β4 (83), μ4 (50), σ4 (17)			(Dell'Angelica et al. 1999a, Hirst et al. 1999, Kirchhausen 1999)

β-Arrestin	β-Arrestin1, 2 (45)	*C.e.*, *D.m.*, *O.m.*	Clathrin TD, AP2β, GPCRs	(Ferguson 2001, Laporte et al. 2000)
AP180	AP180-1/F1-20/NP185/ pp155 (91)[d,e]		Clathrin TD (DLL); AP2 α,β-ear (DΦF/W); phosphoinositides (ENTH); phospholipase D	(McMahon 1999, Morgan et al. 2000)
	CALM/AP180-2 (72)	YAP1801p, YAP1802p (72, 64); *C.e.* (UNC-11); *D.m.* (LAP); *L.p.*, *X.l.*	Clathrin TD (clathrin box); Eps15 EH (NPF); phosphoinositides (ENTH)	(Ford et al. 2001, Mao et al. 2001, Tebar et al. 1999, Wendland & Emr 1998)
Dynamin	Dynamin 1 (95)[c]	Vps1p (79); *A.t.*; *C.e.* (Dyn-1); *D.m.* (Shibire)	Phosphoinositides (PH); G-proteins β/γ (PH); amphiphysin SH3, endophilin SH3, intersectin SH3, syndapin SH3, Grb2 SH3, cortactin SH3, Src SH3, PI3Kinase SH3 (PRD); phospholipase Cγ (PRD); profilin	(Hinshaw 2000)
	Dynamin 2 (97)			(Altschuler et al. 1998)
	Dynamin 3 (96)			(Nakata et al. 1993)
Auxilin	Auxilin (100)	Swa2p/Aux1p (73); *C.e.*	Hsc70 (J domain); clathrin, AP2 α-ear	(Lemmon 2001)
	Auxilin 2/GAK (160)		Hsc70 (J domain); clathrin, AP1 γ-ear, AP2 α-ear	(Greener et al. 2000, Umeda et al. 2000)
Hsc70	Hsc70/Hsp70 (72)	Ssa1p (70), Ssa2p (69)	Auxilin J domain, clathrin	(Newmyer & Schmid 2001)
GGA	GGA1 (70), GGA2 (67), GGA3 (75)	Gga1p (62), Gga2p (64)	ARF1/3 (GAT); γ-synergin (AGEH in GGA1, GGA3)	(Boman et al. 2000, Dell'Angelica et al. 2000, Hirst et al. 2000, Takatsu et al. 2000)
γ-synergin	γ-synergin (145)		APIγ, GGA1, GGA3 (γ-adaptin binding domain); SCAMP1 (EH)	(Page et al. 1999, Takatsu et al. 2000)
Amphiphysin	Amphiphysin I (128)	Rvs167p (53), Rvs161p (30); *C.e.*	Dynamin, synaptojanin (PRD); AP2 (DΦF/W); clathrin TD (clathrin box); endophilin, SOS, myc, PLD1, PLD2, p35	(David et al. 1996, Lichte et al. 1992, Slepnev et al. 2000)
	Amphiphysin II (60,85)/ BIN1 (70)/BRAMP-2/ ALP-1/SH3P9			(Ramjaun et al. 1997, Sakamuro et al. 1996, Sparks et al. 1996)

(*Continued*)

TABLE 1 *(Continued)*

Proteins[a]	Mammalian (kDa)	Yeast (kDa) and others[b]	Binding partners (domain)[c]	References
Synaptotagmin	Synaptotagmin I (46)[d]	*A.p., C.e., D.m., D.o., G.g., L.p.*	Ca^{2+}, phospholipids, syntaxin (C2A); P/Q-Ca^{2+} channel, N-Ca^{2+} channel, β-SNAP, phosphoinisitides, AP2 α, AP2 μ (C2B); calmodulin, neurexins, casein KII, Cam KII, PKC	(Haucke & De Camilli 1999, Marqueze et al. 2000, Zhang et al. 1994)
	Synaptotagmin II-XIII (46—76)			(Marqueze et al. 2000)
SCAMP1	SCAMP1 (37)		γ-synergin EH, intersectin EH (NPF)	(Fernandez-Chacon et al. 2000, Page et al. 1999)
Intersectin (Ese) Intersectin 2 (145)	Intersectin 1 (200),	*D.m.* (DAP160), *C.e.*, *X.l.*	Dynamin PRD, synaptojanin PRD (SH3); SCAMP1 NPF, epsin NPF (EH); Eps15; mSos	(Pucharcos et al. 2000, Roos & Kelly 1998, Yamabhai et al. 1998)
Epsin	Epsin 1 (78)[d]/Ibp1	Ent1p (52); *X.l.* (M90); *D.m.* (Lqf)	AP2 α,β (DΦF/W); Eps15 EH, intersectin EH, PLZF EH, POB1 EH (NPF); clathrin TD (clathrin box)	(Chen et al. 1998, Wendland et al. 1999, Itoh et al. 2001, Drake et al. 2000)
	Epsin 2 (74)/Ibp2[e]	Ent2p (72)		(Rosenthal et al. 1999)
	Epsin 3 (75)			(Spradling et al. 2000)
Eps15	Eps15 (108, 76)	Pan1p (127); *C.e.* (EHS-1); *D.m.*	AP2 α,β (DΦF/W); synaptojanin NPF, intersectins NPF, epsin NPF, Numb NPF, Rab/hRip (EH); Crk SH3 (PRD); Hrs, EAST	(de Beer et al. 1998, Enmon et al. 2000, Salcini et al. 1999, Wendland & Emr 1998, Whitehead et al. 1999)
	Eps15R (125)[e]		AP2 α (DΦF/W); Rab (EH)	(Coda et al. 1998)
Ankyrin	Ank 1/ankyrin R (20—206), Ank 2/ankyrin B (440, 220), Ank 3/ankyrin G (119—480)		Clathrin, spectrin, vimentin, tubulin, microtubules, protein 3, Na-channels, Na/K-ATPase, ryanodine receptor, IP_3 receptor, CD44, PKC, cell adhesion molecules/L1-family	(Rubtsov & Lopina 2000)

Syndapin/Pacsin	Syndapin 1/Pacsin 1 (52/50) Syndapin 2/Pacsin 2 (65)	*C.e., G.g.*	Dynamin 1 PRD, synaptojanin PRD (SH3); synapsin 1, N-WASP	(Merilainen et al. 1997, Qualmann & Kelly 2000, Ritter et al. 1999)
	Pacsin 3 (48)	*G.g.* (FAP52)		
Hip1R	Hip1R (120)	Sla2p/End4p/Mop2p (110); *C.e., D.m.*	phosphoinositides (ENTH); F-actin, clathrin	(Engqvist-Goldstein et al. 1999, 2000)
Synaptojanin	Synaptojanin 1 (145[c], 170)	Sj11p, Sj12p, Sj13p/ Inp51p, Inp52p, Inp53p (108,136,124); *C.e.* (UNC-26)	Grb2 SH3, amphiphysin SH3, endophilin SH3, intersectin SH3, syndapin SH3 (PRD); Eps15 EH (NPF); AP2 α-ear, clathrin TD	(Haffner et al. 2000, Harris et al. 2000, Hughes et al. 2000)
	Synaptojanin 2 (140)		Mitochondrial OMP25 PDZ	(Nemoto & De Camilli 1999)
Endophilin	Endophilin A1(40), Endophilin A2, Endophilin A3	*C.e., D.m.*	Synaptojanin 1 PRD, dynamin PRD (SH3); amphiphysin	(de Heuvel et al. 1997, Huttner & Schmidt 2000, Micheva et al. 1997, Ringstad et al. 1997)
	Endophilin B1	*C.e., D.m.*		(Huttner & Schmidt 2000)

[a]Abbreviations used: AGEH, adaptor gamma ear homology; Cam KII, Ca^{2+}/calmodulin kinase II; CHC, clathrin heavy chain; CLC, clathrin light chain; EAST, epidermal growth factor receptor-associated protein with SH3 and TAM domains; EH, Eps15 homology; ENTH, epsin N-terminal homology; IP, inositol polyphosphate; LC, light chain; LL, di-leucine motif; PDZ, PSD-95, DLG, ZO-1; PH, pleckstrin homology; PKC, protein kinase C; PLZF, promyelocytic leukemia Zn^{2+} finger protein; POB1, partner of RalBP1; PRD, proline-rich domain; SH3, Src-homology 3; TD, terminal domain; VHS, Vps27, Hrs, STAM; YXXΦ, tyrosine-based motif; WASP, Wiskott-Aldrich syndrome protein.

[b]Other nonmammalian organisms in which a homolog has been identified are abbreviated as follows: *A.c., Aplysia californica*; *A.t., Arabidopsis thaliana*; *C.e., Caenorhabditis elegans*; *D.d. Dictyostelium discoideum*; *D.o., Discopyge ommata*; *D.m., Drosophila melanogaster*; *G.g., Gallus gallus*; *L.p., Loligo pealei*; *O.m., Oncorhynchus mykiss*; *X.l., Xenopus laevis*; *Z.m. Zea mays*. Name of the homolog protein is provided in parentheses following the species abbreviation.

[c]Binding partner(s) for the protein in the first column are listed and, where known, the interaction domain of the binding partner is noted. Parentheses after each partner or set of partners enclose the interaction domain of the protein in the first column that binds the partner(s) and indicated domains.

[d]Apparent molecular weight on sodium dodecyl sulfate-polyacrylamide gel electrophoresis in kDa: AP180, 91; AP3 δ, 160; AP4 ε, 140; CHC17, 180; Epsin1, 94; GGA1, 85; GGA3, 90; LCa/LCb, 30–36; Synaptotagmin I, 65.

[e]Neuronal splice variant(s).

CCV biochemistry to a new level (Kirchhausen 2000b, Slepnev & De Camilli 2000, Wakeham et al. 2000). In this section we review the structure and biochemistry of bona fide members of the vesicle coat, as well as those proteins involved in vesicle assembly, disassembly, and scission from the membrane, to set the stage for understanding the cellular physiology of CCV formation.

Clathrins

The clathrin triskelion is formed from three identical clathrin heavy chains (CHCs) of 1675 residues in mammalian clathrin (192 kDa) (Kirchhausen et al. 1987), with variation of sequence length at the extreme carboxyl (C) terminus in CHCs of other species (Brodsky 1999). CHCs have variable migration by sodium dodecyl sulfate–polyacrylamide gel eletrophoresis (SDS-PAGE), such that mammalian CHCs appear to be 180 kDa, whereas yeast CHC, which terminates at the equivalent of mammalian residue 1653, runs closer to 190 kDa (Tan 1993). CHCs can be divided functionally and structurally into thirds (Figure 4). The central portion of the clathrin molecule, known as the hub (Liu et al. 1995), is formed by the C-terminal third (mammalian CHC residues 1074–1675) and includes sequences mediating CHC trimerization (between mammalian residues 1560–1615) and binding sites for the clathrin light chain (CLC) subunits, which regulate clathrin assembly (Brodsky et al. 1991). The proximal portion of the triskelion leg (mammalian residues 1074–1522) extending from the trimerization domain is a superhelix of α-helices (Figure 3). These helices are grouped in a repeating structural motif comprising 10 short α-helices linked by a pattern of conserved salt bridges (Ybe et al. 1999). The general sequence motif characteristic of the structural motif is repeated seven times in all CHC sequences, which suggests that this CHC repeat (CHCR) accounts for the entire linear portion of the triskelion leg (proximal and distal) and contributes to CHC trimerization. The CHCR also appears in other proteins (singly or in pairs) involved in membrane traffic, including Pep3p, Pep5p, Vps39p, and Vps41p. A single CHCR would be about 55 Å in length and may represent a general protein interaction domain or possibly a clathrin-binding domain (Ybe et al. 1999). The central third of the CHC forms the distal segment of the triskelion leg and is predicted to have a similar structure to the proximal leg. A flexible bend separates the proximal from the distal leg segments (Musacchio et al. 1999), and even when this connection is severed, the resulting separated fragments still self-assemble to form a lattice (Greene et al. 2000). The N-terminal third of the CHC comprises a globular terminal domain (TD), which forms a beta-propeller structure, with binding sites for various clathrin-interacting proteins on the "blades" of the propeller (ter Haar et al. 1998) (Table 1, Figure 3). The TD is connected to the distal leg segment by a linker region of short alpha helices, which are not organized into a CHCR but appear to be more flexible. The TD has been cocrystallized with peptides containing clathrin-binding sequences both from the β-chain of the AP3 adaptor and from β-arrestin, which bind to the same binding site (ter Haar et al. 2000). It is likely that this limited binding site

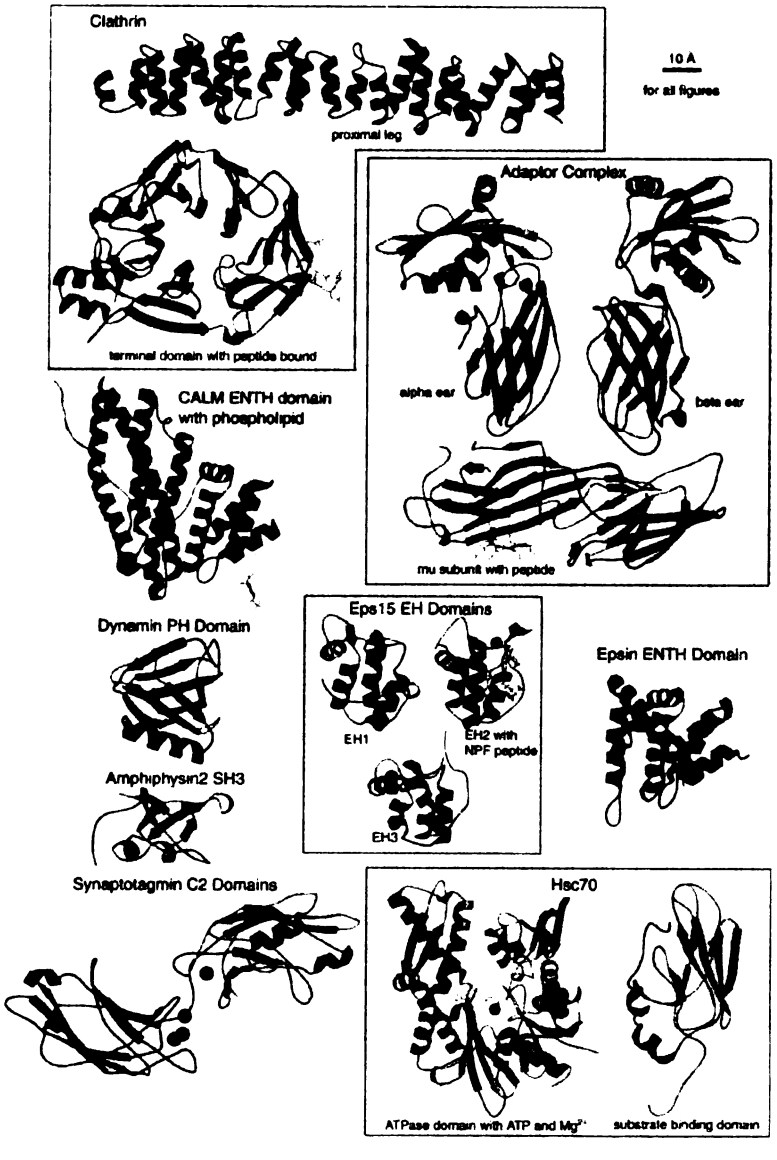


Figure 3 (See legend on next page)

Figure 3 (See figure on previous page) (opposite page) Molecular structures of clathrin-coated vesicle components and regulatory proteins, as of April 2001. (For information on where each domain lies in the primary structure of the protein, see Figure 4; for information on the domain's interactions, see Table 1.) Images were generated from coordinates deposited in the Protein Data Bank (Berman et al. 2000) using Molscript (Kraulis 1991) and Raster3D (Merritt & Murphy 1994). The colors are used to highlight different structural features. Ligands are shown in red, and red balls depict bound calcium atoms, or magnesium in the case of hsc70 and synaptotagmin. The sources for the coordinates are as follows: clathrin proximal leg PDB ID, 1B89 (Ybe et al. 1999); clathrin terminal domain with bound β-adaptin 3 peptide PDB ID, 1C9I (ter Haar et al. 2000); adaptor complex α-ear PDB ID, 1B9K (Owen et al. 1999); adaptor complex β2-ear PDB ID, 1E42 (Owen et al. 2000); adaptor complex μ2-subunit with bound TGN38 peptide PDB ID, 1BXX (Owen & Evans 1998); CALM ENTH domain with bound PI(4,5)P2 PDB ID, 1HFA (Ford et al. 2001); dynamin1 PH domain PDB ID, 2DYN (Timm et al. 1994); amphiphysin2 SH3 domain PDB ID, 1BB9 (Owen et al. 1998); synaptotagmin III C2A and C2B domains PDB ID, 1DQV (Sutton et al. 1999); Eps15 EH1 domain PDB ID, 1QJT (Whitehead et al. 1999); Eps15 EH2 domain with bound NPF-containing peptide PDB ID, 1FF1 (de Beer et al. 2000); Eps15 EH3 domain PDB ID, 1C07 (Enmon et al. 2000); epsin1 ENTH domain PDB ID, 1EDU (Hyman et al. 2000); hsc70 ATPase domain with bound ATP and Mg^{2+} ions PDB ID, 3HSC (Flaherty et al. 1990); and hsc70 substrate binding domain PDB ID, 7HSC (Morshauser et al. 1999).

represents only a fraction of an extended site for either protein to interact with clathrin.

CHC22, a second form of CHC (1640 residues), was identified during human genome analysis of human chromosome 22 (reviewed in Brodsky 1997). The conventional CHC described above is encoded on human chromosome 17 and for comparative purposes is referred to as CHC17. CHC22 is 84% identical to CHC17, and the differences are scattered throughout the protein sequence such that no particular domain is dramatically different from CHC17. CHC22 forms a trimer but does not associate with CLCs as avidly as does CHC17 (Liu et al. 2001b). CLC binding to CHC22 is not detected biochemically and has been observed only through yeast two-hybrid interactions, whereas CLCs form a stable subunit of CHC17 triskelia and are dissociated only under extreme denaturing conditions. It is possible that CHC22 has an additional subunit replacing CLCs, but no such subunit has yet been identified. CHC22 is highly expressed in skeletal muscle and can be detected at a low level in other cell types. In nonmuscle cells, CHC22 is associated with the TGN but not the PM and correspondingly associates with adaptors AP1 and AP3 but not AP2 (Liu et al. 2001b). The function of CHC22 appears to be distinct from CHC17 and is discussed further in the sections on tissue-specific function and cytoskeletal interactions below. It is of interest that a gene encoding CHC22 is not present in mice, whose corresponding chromosomal region seems to have been deleted during inversion (Lund et al. 2000). Southern blot analysis has suggested its presence in primates, rabbits, and dogs (Holmes et al. 1997), but in nonmammalian species the sequences of CHCs are too divergent to determine whether the CHC resembles human CHC22 or CHC17.

The CLC subunits bind to the hub region of the triskelion, one associated with each leg. There is a single CLC in yeast, invertebrates, and insects, and there are two forms (LCa and LCb) in vertebrate species. These are encoded on different chromosomes in humans and have about 60% protein sequence identity. There are also neuronal splicing variants of both LCa and LCb (Brodsky et al. 1991). Thus LCa and LCb vary from 25–29 kDa, but by SDS-PAGE they appear to be 30–36 kDa, as they are highly negatively charged (Brodsky 1988). CLCs bind clathrin through a central domain, flanked on the C-terminal side by the splice sites for neuronal sequences. which introduce a hydrophobic patch into the proteins. On the N-terminal side, both LCa and LCb have a calcium-binding site and a sequence unique to each CLC, which in LCa can stimulate the uncoating protein hsc70 to disassemble polymerized clathrin (Brodsky et al. 1991). Approximately 20 residues from the N terminus in both light chains is a completely conserved sequence of 22 residues shared by LCa and LCb in all mammalian species and which represents the region of highest homology with yeast CLC (Pley & Parham 1993). The first three residues of the conserved mammalian sequence are negatively charged, with two charges conserved in CLCs of all species analyzed to date. These charged residues regulate the pH sensitivity of clathrin assembly. Without CLC bound, triskelions of CHCs will polymerize into a lattice at physiological pH. With CLC bound, polymerization will only occur below pH 6.5, and this effect depends

on the negatively charged residues at the beginning of the conserved CLC sequence (Ybe et al. 1998). Adaptors AP1 and AP2 can reverse this inhibition and stimulate assembly of triskelia at physiological pH. Thus one role of CLCs is to moderate the tendency of CHCs to self-assemble, rendering clathrin susceptible to regulation by adaptors in the cell. In vitro, clathrin self-assembly can be stimulated by calcium, and this effect is reduced when CLCs are dissociated from triskelia. It is not clear whether calcium plays a role in cellular clathrin assembly because concentrations required for CLC binding of calcium (K_d of 25 μM) are considerably higher than steady state intracellular calcium levels (Pley & Parham 1993).

In vertebrate cells, LCa and LCb compete with each other for CHC binding (Brodsky 1988). Although LCb has a slightly higher affinity in vitro, cellular triskelia comprise a random distribution of LCa and LCb according to their relative expression level in the cell, such that four types of triskelia (aaa, aab, bba, and bbb) are always present. LCa and LCb are expressed at different relative levels in cells of different tissue origins. Notably LCb is in excess of LCa in cells with a regulated secretory pathway (Acton & Brodsky 1990). However, major differences in function of LCa and LCb have not been evident from analysis of PC12 cells lacking LCa, which suggests that LCb may represent the more-specialized form and/or that the differential function of LCa and LCb is only important for cells participating in organized tissue. LCb and yeast LC are targets for phosphorylation (Brodsky et al. 1991, Chu et al. 1999), which may reflect regulation of their function.

Adaptors and Related Proteins

The heterotetrameric adaptors AP1 (γ-, β1-, μ1A-, and σ1-subunits) and AP2 (α-, β2-, μ2-, and σ2-subunits) are established components of the coat of CCVs (Kirchhausen 1999). These adaptors localize CCV formation to the TGN (AP1) or the PM (AP2) (Pearse & Robinson 1990) by induction of clathrin assembly through the β1- or β2-subunits (Gallusser & Kirchhausen 1993). They bind to cargo with the μ1- or μ2-subunits. In particular, the μ-subunits recognize sequence motifs YXXΦ (where Φ is a bulky hydrophobic residue) in the cytoplasmic domains of receptors that are sequestered in CCVs (Ohno et al. 1995). The structure of the μ2-subunit has been determined, cocrystallized with a few peptides (Owen & Evans 1998). It consists of two β-sandwich subdomains joined so that the whole structure is one convex surface with a peptide-binding site along the edge of one subdomain. The peptide (in an extended conformation) mimics an additional β-strand, and there are defined pockets that can accommodate the critical tyrosine and bulky hydrophobic residues in the internalization motif. An extended binding site was revealed when an extended peptide from p-selectin (Owen et al. 2001) was cocrystallized. The recognition of two other CCV cargo motifs by adaptors is less well defined. The dileucine-based motif (LL, LI, or ML) can be demonstrated to interact with the β1-subunit by cross-linking assays but can be demonstrated to interact with μ1- and μ2-subunits by surface plasmon resonance (Hofmann et al. 1999, Rapoport et al. 1998). Recognition of the NPXY motif directly by the TD of

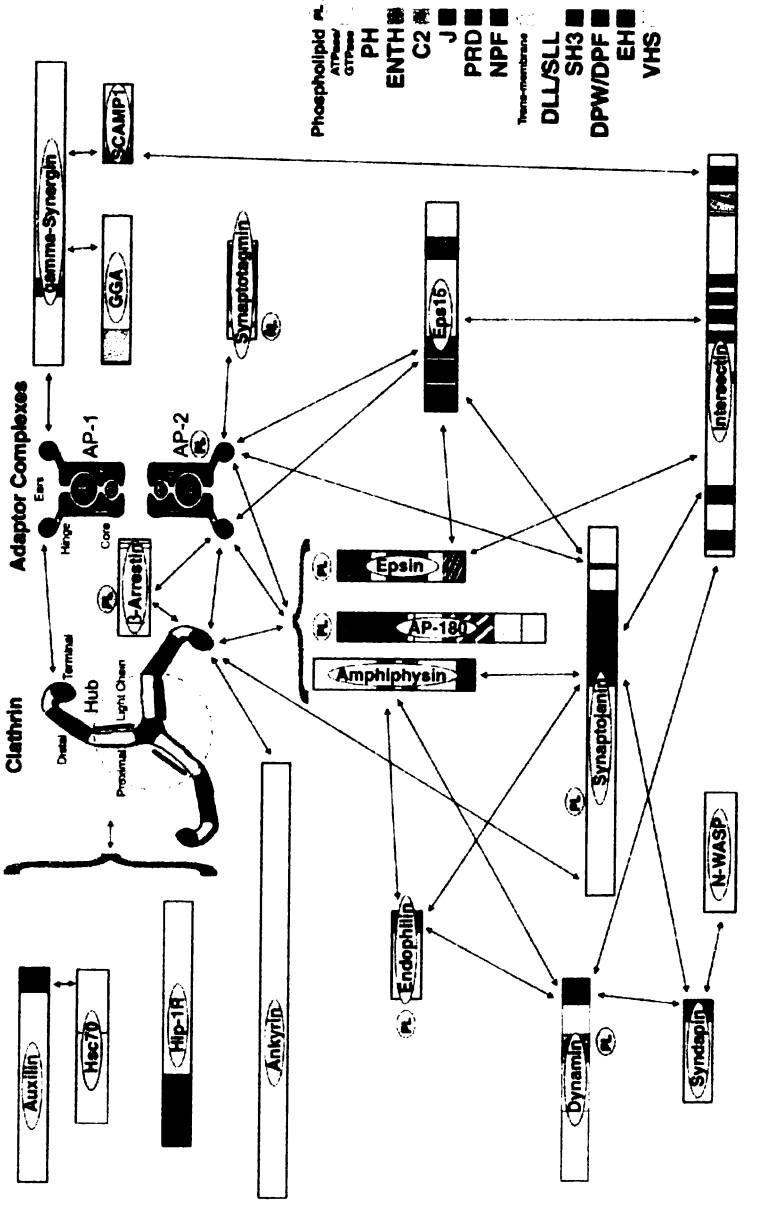


Figure 4 (See legend on next page)

Figure 4 (See figure on previous page) Molecular interactions between clathrin-coated vesicle components and regulatory proteins. Clathrin and adaptors interact with myriad factors that influence assembly, vesicle budding, cytoskeleton interaction, signaling, and phospholipid interaction. The central cartoons represent clathrin and adaptors, and a bar proportional to protein length represents other proteins. Arrows represent binding interactions between proteins. Colors represent the domain location and length within the primary structure of the protein. Where two domains overlap, the two colors are shown as diagonal stripes. In cases where proteins are alternatively spliced or there are family variants, the longest member of the protein family is depicted. (Note that arrows do not necessarily indicate which domains are interacting; for that information, see Table 1). Large gray brackets indicate binding of multiple proteins, e.g., auxilin, hsc70, and Hip1R all bind to clathrin. Pink ovals enclosing PL represent phospholipid binding sites.

clathrin has been detected by nuclear magnetic resonance spectroscopy (Kibbey et al. 1998), but there is not yet any independent experimental evidence that this occurs in vivo.

Electron microscopic analysis has revealed that the AP2 adaptors have a characteristic morphology resembling a head with ears (Heuser & Keen 1988). The head, or core domain, comprises the N-terminal two thirds of the large subunits and the μ2- and σ2-subunits and contains the determinants that localize AP1 and AP2 to the TGN and PM, respectively. As localization properties associate with the N-terminal sequences of the α- and γ-subunits, which also control which μ- and σ-subunits are bound, localization could be a function of any or all of these three subunits. The N-terminal segment of the α-subunit of AP2 in the core domain has a binding site for phosphoinositides (residues 5–80) (Gaidarov et al. 1996).

The ears, or appendages, are formed by the C-terminal domains of the α- and β2-subunits for AP2 (residues 701–938 for α and 705–937 for β). The appendage domains are attached to the core via proline-rich hinge regions. In the β-subunit, this region has a sequence motif, L (L,I) (D,E,N)(L,F)(D,E), that has been called the clathrin box (Dell'Angelica et al. 1998). Segments of the protein containing this sequence have been shown to bind to clathrin, and indeed, the hinge-appendage fragment of the β2-subunit stimulates assembly of purified clathrin and recombinant clathrin fragments (Greene et al. 2000). Clathrin box sequences have been identified in other clathrin-binding proteins, including arrestins, amphiphysins, the AP180 family, and the β-subunits of AP1 and adaptor-related complex AP3. Peptides containing this motif interact with clathrin TDs, as described above (ter Haar et al. 2000). Recent sequence analysis of the AP180 family of proteins (see below) suggests that the actual clathrin-binding motif may contain features of the clathrin box but is more extended and somewhat more degenerate (Morgan et al. 2000).

The structures of the appendage domains of both the α-subunit (Owen et al. 1999, Traub et al. 1999) and β2-subunit (Owen et al. 2000) of AP2 have been determined at crystallographic resolution. Despite their low sequence identity, these two structures are similar (Figure 3). Each appendage can be divided into two subdomains, with a distinctly different orientation relative to each other in the α- or β-subunit. The N-terminal domain is a β-sandwich (similar to an immunoglobulin domain with nine strands in the α-appendage and eight strands in the β-appendage). The C-terminal domain is a single β-sheet, flanked by one α-helix on one side and two α-helices on the other face. The C-terminal subdomain of both structures has a hydrophobic patch, centered around a critical tryptophan residue, that is the binding site for proteins with the DΦF/W motif. Each appendage binds a different subset of these proteins (Table 1, Figure 4), and their binding has different affinities, which suggests a sequence of association of these proteins with the appendage domains during CCV formation.

The γ-subunit of AP1 has an abbreviated sequence that suggests it lacks the protein interaction domain of the other appendages. This appendage binds to a protein called γ-synergin, which may supply interaction sites for other proteins (Page et al.

1999). γ-Synergin contains an EH (Eps15 homology) domain, which suggests the potential to interact with the NPF motif found in proteins implicated in vesicle formation at the PM. In addition, γ-synergin binds to the GGA proteins (Takatsu et al. 2000), which have homology with the appendage domain of the AP1 γ-subunit and interact with ARF3 in two-hybrid assays (Boman et al. 2000, Dell'Angelica et al. 2000, Hirst et al. 2000). It is likely that binding of γ-synergin to GGAs or to AP1 is mutually exclusive. The GGAs are localized to the TGN and are involved in transport from the TGN to lysosomes in yeast (Black & Pelham 2000). They bind to the TGN via an interaction with ARF1 and have been visualized in a dense protein coat at the TGN. The GGAs interact with clathrin in vivo, and GGAs have been shown to promote recruitment of clathrin to liposomes in vitro and to TGN membranes in vivo (Puertollano et al. 2001). GGAs are monomeric but contain multiple domains that could potentially perform all the functions of the heterotetrameric adaptors including ARF binding, cargo recognition, and clathrin recruitment (Puertollano et al. 2001). Thus GGAs appear to be able to nucleate clathrin coat formation at the TGN and thereby represent a novel form of adaptor molecule. A specialized form of the AP1 adaptor having the μ1B-subunit instead of μ1A has been identified, with expression limited to polarized cells (Fölsch et al. 1999, Ohno et al. 1999). Furthermore, two additional complexes with subunit compositions similar to AP1 and AP2 have been identified and characterized (Kirchhausen 1999). These are known as AP3 (δ-, β3A or β3B-, μ3-, and σ3A or σ3B-subunits) and AP4 (ε-, β4-, μ4-, and σ4-subunits). AP3 has been implicated in specialized sorting pathways in the TGN, particularly in melanosome formation. By immunoprecipitation it has been shown to interact with CHC17 and CHC22 (Liu et al. 2001b). The former interaction is through the predicted clathrin box in the β3-hinge region (Dell'Angelica et al. 1998). However, AP3 does not copurify with CCVs (Simpson et al. 1996), and in yeast, AP3-mediated transport pathways function independently from clathrin-mediated transport pathways (Vowels & Payne 1998). Thus whether AP3 functions completely independently of clathrin in mammalian cells or perhaps less dependently than AP1 and AP2 is not clear. The μ3-subunit does interact with cargo motifs in yeast two-hybrid analysis (Dell'Angelica et al. 1997). Expression of AP4 is limited to a low level and is localized to the TGN region of cells (Dell'Angelica et al. 1999a, Hirst et al. 1999). There are no predicted interactions with clathrin for AP4. Adaptor complexes equivalent to AP1, AP2, and AP3 have been found in nonmammalian species, including yeast, which have 13 potential adaptor subunits (four μ-subunits and three of each of the others corresponding to homologues of subunits from mammalian AP1, AP2 and AP3) (Cowles et al. 1997).

Coat Proteins Influencing Assembly and Disassembly

AP180 and auxilin are two clathrin-binding proteins involved in CCV coat assembly and disassembly, which were identified as coat components. AP180 (see Table 1 for aliases) is a neuronal protein, related in structure and function to a

nonneuronal family member CALM (McMahon 1999). Versions of AP180 have been identified in *Drosophila melanogaster*, *Caenorhabditis elegans*, and squid with variable degrees of homology and inserted sequences. All the proteins of this family have at their N termini an ENTH domain, which binds phosphatidyl inositol 4,5-bisphosphate (PIP2). Although the C termini are highly divergent, all retain the ability to bind clathrin. Rat AP180, bovine AP180, *Xenopus* AP180, and human AP180 are similar, whereas CALM from rats and humans are similar to each other. Both AP180 and CALM bind clathrin and are involved in endocytosis, but they coexist in some cell types (Kusner & Carlin 2000). The other members of this family identified in yeast (Yap1801, Yap1802), *C. elegans* (Unc-11), *D. melanogaster* (LAP), and squid (AP180) appear to be more closely related to CALM than to AP180, although homology in the C terminus is low (Morgan et al. 1999). For mammalian AP180, deletion and mutagenesis studies on the clathrin-binding domain implicate a DLL or SLL motif in clathrin binding (Morgan et al. 2000). This motif is present in multiple (12) copies in both AP180 and in other clathrin-binding molecules, including all the β-subunits (one to four copies) of the AP1, AP2, AP3, and AP4. AP180 also makes a complex with AP2 via DΦF/W repeats that can bind both the α- and β-subunit appendages (Hao et al. 1999, Owen et al. 2000). This complex would have multiple DLL motifs that could potentially cross-link clathrin and promote assembly. The predicted "clathrin box" motif in CALM and LAP is actually buried and inaccessible in the structure of the molecule (Ford et al. 2001, Mao et al. 2001). However, the clathrin box motif, defined in other analyses of clathrin-binding proteins, such as that of the β-subunits of AP1, AP2 and AP3, overlaps with DLL or SLL motifs so that it is not yet possible to definitively identify what actually constitutes a clathrin-binding motif.

Auxilin, originally identified as a neuron-specific component of CCVs (Ahle & Ungewickell 1990), has now been found in two forms in mammalian cells, the neuronal form and the ubiquitous form, auxilin 2. The latter form is also a cyclin G–associated kinase (GAK) that has serine/threonine kinase activity and can phosphorylate the μ-subunit of both AP1 and AP2 (Greener et al. 2000, Umeda et al. 2000). The N terminus of both auxilins has a phosphatase and tensin homology (PTEN) domain, followed by a clathrin-binding domain and a J domain at the C terminus (Greener et al. 2000, Umeda et al. 2000, Ungewickell et al. 1995). The PTEN domain could have actin-binding activity through its tensin homology, as tensin binds actin in focal adhesion plaques (Kanaoka et al. 1997). The phosphatase associated with PTEN tumor suppressor domains can dephosphorylate phosphatidyl inositol 3,4,5-triphosphate (PIP3) (Maehama & Dixon 1998). Although the phosphatase activity of auxilins remains to be determined, through a similar activity they could potentially influence the phosphoinositide binding of other CCV components. The J domain is a defining feature of the hsp40 cochaperone family and is essential for stimulating the ATPase activity of the hsc70/hsp70 chaperones, as occurs in their *Escherichia coli* homologues DnaJ and DnaK, respectively (Kelley 1998). Thus auxilins can recruit hsc70 to the clathrin coat through their ability to interact with clathrin and APs and stimulate uncoating activity. Auxilin

in *C. elegans* and yeast (Swa2p/Aux1p) has highest homology to mammalian auxilins in the J domains and little homology elsewhere in the protein (reviewed in Lemmon 2001). Mutations in auxilin-encoding genes in yeast and worms cause accumulation of uncoated CCVs and block endocytosis. Overexpression of auxilin and GAK in mammalian cells disrupts clathrin localization to membranes. These in vivo experiments support a role for auxilin as a cofactor in CCV uncoating (Umeda et al. 2000, Zhao et al. 2001). It is of interest that auxilin was initially discovered as a stimulator of clathrin assembly in vitro (Ahle & Ungewickell 1990). It may be that auxilin/GAK has assembly-promoting properties that allow its incorporation into the CCV coat, thereby priming vesicles for uncoating.

Dynamin and Binding Partners

Dynamin is a GTPase that oligomerizes into tetramers that can stack into open rings and form tubules (Hinshaw 2000). Its function in scission of an assembled CCV from the PM was identified through the phenotype of the *shibire D. melanogaster* mutant, which accumulates CCVs attached to the PM in synaptic regions (Kosaka & Ikeda 1983). Dynamin's role in endocytosis has been confirmed by the construction of many mutants that affect endocytosis, most notably the K44A mutant, with a defect in GTP binding and hydrolysis. Overexpression of K44A-dynamin inhibits clathrin-mediated endocytosis, as well as caveolae function (McNiven et al. 2000). There is some debate as to the exact role of dynamin during CCV fission (discussed below in the section on mechanics of CCV formation). There are numerous splice variants and three genes encoding human dynamins. Dynamin 1 functions at the PM and is the most well-characterized. Dynamin 2 has been proposed to play a role in budding of CCVs at the TGN, although conflicting results discussing dynamin 2 function are reported in the literature (Altschuler et al. 1998, Kreitzer et al. 2000, McNiven et al. 2000). Dynamin 3 has restricted tissue distribution and is most strongly expressed in testis (McNiven et al. 2000). Additional roles for mammalian dynamins in membrane traffic have been suggested by expression and mutagenesis studies and include an influence on membrane-cytoskeleton interactions (Witke et al. 1998) and intracellular signaling (Ahn et al. 1999, Fish et al. 2000, Whistler & von Zastrow 1999). Dynamin homologues have been identified in *D. melanogaster*, *C. elegans*, and yeast. In the latter, the closest homologue, Vps1p, is involved in TGN-to-vacuole membrane traffic; a more distant homologue, Dnm1p, has been implicated in endosome-to-vacuole transport (Gammie et al. 1995) and mitochondrial morphology (Bleazard et al. 1999); and a third, Mgm1p, is also involved in mitochondrial function (Baggett & Wendland 2001).

Dynamin 1 interacts with numerous protein partners, all of which play a role in CCV-mediated endocytosis and some of which, such as profilin and cortactin, can interact directly with the actin cytoskeleton. The majority of dynamin's protein interactions occur through a proline-rich domain at the C terminus, which binds SH3 domains in other proteins (Table 1, Figure 4) (Simpson et al. 1999). Adjacent to the proline-rich domain, dynamin has a GTPase effector domain (Muhlberg

et al. 1997) that increases dynamin's own GTPase activity (a function of the N-terminal domain) by 50- to 100-fold. This GTPase effector domain activity can influence adjacent dynamin tetramers during dynamin self-assembly into stacked rings of tetramers. Dynamin also has a pleckstrin homology (PH) domain that interacts with PIP2 (Salim et al. 1996) to mediate membrane binding and that could also mediate binding to the β- and γ-subunits of heterotrimeric G proteins (Liu et al. 1997). The structure of the dynamin PH domain (Timm et al. 1994) resembles that of PH domains in other phosphoinositide-binding proteins (Rebecchi & Scarlata 1998), such as Bruton's tyrosine kinase (Btk).

Amphiphysins bind dynamin through a C-terminal SH3 domain, and some forms bind clathrin and AP2, thereby linking dynamin to these two coat proteins (Slepnev et al. 2000). Amphiphysins also interact with lipid-modifying proteins endophilin, synaptojanin, and phospholipase D1 and 2 and with proteins involved in signaling (Table 1) (Wigge & McMahon 1998). In mammals, amphiphysin I is expressed in neurons, testis, and neuroendocrine cells, whereas amphiphysin II has notable expression in skeletal muscle and brain (Butler et al. 1997, Wigge & McMahon 1998). Amphiphysin homologs in yeast Rvs167p and Rvs161p are functionally required for endocytosis, and genetic evidence suggests an interaction with components of the actin cytoskeleton (Amberg et al. 1995). In mammals, amphiphysin II plays a role in macrophage phagocytosis, which further suggests a cytoskeletal connection (Gold et al. 2000).

MECHANICS OF CCV FORMATION AND MEMBRANE INTERACTIONS

Clathrin Assembly by Adaptors and Lattice Rearrangement

The structure of the lattice produced by clathrin assembly in the presence of adaptor molecules has been determined to a resolution of 21 Å by cyroelectron microscopy and image averaging (Figure 5) (Smith et al. 1998). The features of this lattice are that every edge is formed by the interaction of two proximal leg domains and two distal leg domains, each contributed by four different triskelia. The geometry is such that underneath the trimerization domain at the vertex of each triskelion is the conjunction of three "knee bends" from adjacent triskelia and below that is the conjunction of three TDs from triskelia centered two vertices away. The structure of the TDs has been modeled within the 21 Å map, and it is clear that they extend inside the lattice to interact with adaptor molecules, which in turn would be interacting with membrane components (Musacchio et al. 1999, Smith et al. 1998).

The CLC subunits inhibit spontaneous assembly of the heavy chains at physiological pH (Liu et al. 1995). This inhibition can be overcome in vitro, at low pH. In vivo, this inhibition is overcome by adaptor interaction with clathrin. The influence of adaptors on clathrin lattice assembly has been studied in an in vitro

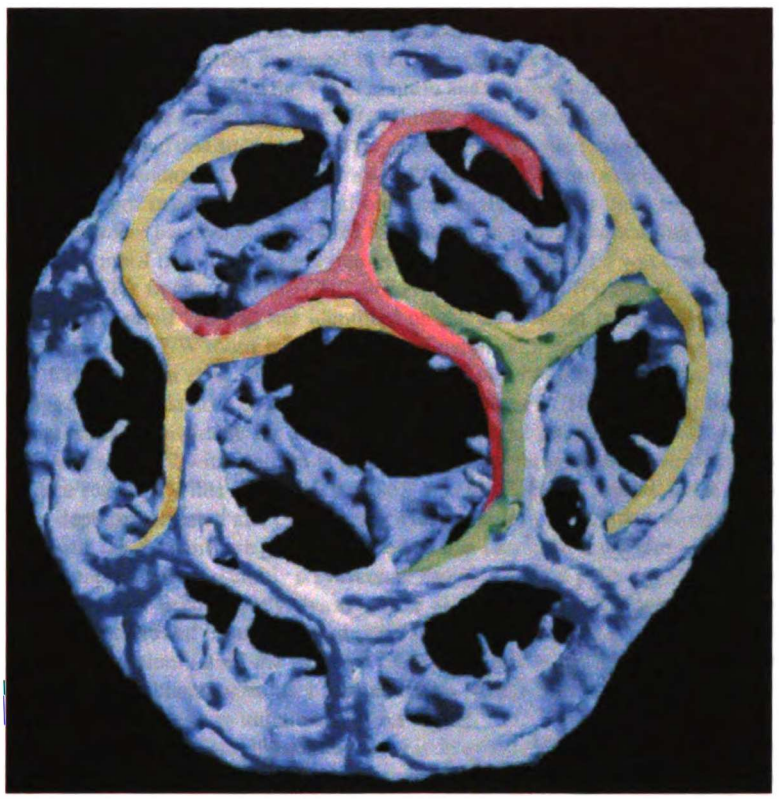


Figure 5 A cryo-electron microscopy image of a whole clathrin basket [reproduced from Smith et al. (1998) with permission of the authors and with copyright permission from Oxford University Press]. The image has been colored to show the location of individual triskelia, similar to the version of this image reproduced by Marsh & McMahon (1999). One side of a polygon of the clathrin lattice is composed of segments of clathrin legs from four different triskelia. Two proximal legs from the red and green triskelia make up the top layer of the lattice. The yellow distal legs of two triskelia centered at adjacent vertices in the lattice can be seen curving underneath the proximal legs to form a deeper layer of the lattice. Note that the terminal domains of the triskelia are not included in this image. The terminal domains curve into the center of the polyhedron, under the vertices (Musacchio et al. 1999).

reconstitution system in which the entire process can be reproduced with fragments of clathrin and adaptors produced in bacteria (Greene et al. 2000). Earlier work showed that adaptors bind the TD of clathrin and that the β-chain of both AP1 and AP2 can induce clathrin assembly (Gallusser & Kirchhausen 1993). Subsequent studies revealed that an N-terminal fragment of the β2-chain comprising the hinge and ear (appendage) domain (Figure 4) will stimulate clathrin assembly (Shih et al. 1995). From structural studies and sequence analysis, it has been suggested that this fragment has two binding sites for clathrin. A peptide containing a clathrin-binding motif from the hinge region of the β3-subunit of AP3 was cocrystallized with the TD fragment of the CHC (ter Haar et al. 2000) (Figure 3). In addition, a DΦF/W-motif binding site identified on the ear domain is predicted to interact with a DΦF/W motif in the CHCR1 of the distal triskelion leg, near the TDs of CHCs (Owen et al. 2000). The presence of two binding sites in one β-chain suggests that adaptors could potentially cross-link two CHCs to orient them in a conformation favorable for assembly. In the cell, the interaction of the β-chains of AP1 and AP2 with clathrin is regulated by phosphorylation of serines in their hinge regions (Wilde & Brodsky 1996). When membrane bound, these regions are dephosphorylated and can interact with clathrin to stimulate assembly. In the cytosol, the β-chains are phosphorylated so that they cannot interact with clathrin and nonproductive assembly is prevented. Properties of the kinase and phosphatase involved have been recently characterized, although the specific proteins have not yet been identified (Lauritsen et al. 2000). The timing of phosphorylation and dephosphorylation within the cycle of CCV formation is not known, although the kinase is associated with fully formed CCVs (Wilde & Brodsky 1996). Thus its activation may help to initiate coat disassembly.

Studies on clathrin self-assembly have indicated that the interaction of proximal triskelion legs is important for self-assembly. Monoclonal antibodies to this region inhibit assembly, and the regulatory CLCs bind in this region (Blank & Brodsky 1987). Studies with recombinant fragments have indicated that the hub fragment, comprising only the proximal leg and the trimerization domain of the triskelion, can self-assemble, but only pseudo-lattices (poorly formed and not closed) are formed. This latter reaction is not influenced by adaptor molecules or fragments. However, when adaptor fragments are combined with a fragment of the triskelion leg containing the distal leg and TD and these are added to hubs with bound CLCs, an intact spherical lattice can be formed that reproduces clathrin coat morphology (Greene et al. 2000). This assembly reaction is influenced by adaptors and requires the presence of the distal leg, as well as the TD of CHC. These observations suggest that the adaptors influence triskelion assembly through their ability to bind both the TD and the distal domain. Given that the distal part of the triskelion leg interacts so intimately with the proximal legs in a clathrin lattice (Figure 5), these segments are in a prime position to also interact with the CLCs and thereby reverse their negative effect on assembly. CLCs do not dissociate from CHCs during clathrin assembly and disassembly, but it is not known whether they undergo any conformational change during these reactions that might alter their influence on CHC assembly

(Pishvaee et al. 1997). Studies on the disposition of the CLCs along the proximal leg of the CHCs have suggested two different potential conformations (Kirchhausen & Toyoda 1993, Näthke et al. 1992), and detailed mapping of the CLC-binding site within the proximal leg is in progress. The answer to this problem will lie in the resolution of the structures of CLCs bound to CHCs in assembled and disassembled clathrin.

At the PM, large patches of clathrin lattices composed only of hexagons are frequently observed. In order for a coat to form around membranes, 12 pentagons must be introduced (based on the mathematical requirements for sphere formation). It has been argued that because of the extensive "weaving" interactions of triskelia, rearrangement of triskelia to form pentagons from hexagons can only occur at the edge of a hexagonal lattice by using the outermost triskelia as a donor pool for a novel reassembly reaction that can introduce pentagons (Kirchhausen 2000a). However, recent analysis of the energetics of clathrin basket formation in the presence of adaptors indicates that although this is an energetically favorable reaction, the threshold energy of assembly and disassembly is such that the lattice can "breathe" and potentially lead to local rearrangement (Nossal 2001). What actually happens at the cell membrane remains unresolved; however, it seems reasonable to hypothesize that rearrangement of a clathrin lattice would require the participation of additional molecules.

The potential need for lattice rearrangement appears to be a feature of CCV formation at the PM, where apparently stable hexagonal arrays of assembled clathrin are observed (Heuser & Kirchhausen 1985). In the TGN, vesicles appear to form from tubules, and the CCVs in the TGN are less stable. Correspondingly, nucleation of CCVs in the TGN is regulated differently (see sections below). There are distinct differences between CCV formation at the PM or at the TGN with respect to both the source of clathrin (membrane bound versus cytosolic) and the nature of the donor membrane (flat and cholesterol rich versus tubulated). These differences could explain why candidate molecules for clathrin rearrangement and lipid deformation have been identified mainly in endocytic-coated vesicles, without the implication of strictly analogous participants in CCV formation at the TGN. It is likely not a coincidence that many such accessory molecules have been discovered by studying CCV formation at nerve termini, where extensive endocytosis from the PM occurs after a burst of synaptic activity (Slepnev & De Camilli 2000).

Candidates for involvement in a lattice rearrangement reaction at the PM are molecules such as Eps15 (Salcini et al. 1999) or amphiphysin I (Wigge & McMahon 1998). Both molecules interact with clathrin and adaptors. Amphiphysin binds directly to both clathrin and adaptors, as well as to dynamin. Eps15 binds the AP2 adaptor directly and clathrin via its binding partner epsin, a protein that also binds AP2. Eps15 was originally identified as a substrate for the epidermal growth factor receptor (EGFR) tyrosine kinase (Salcini et al. 1999). Through expression of dominant negative mutant fragments of Eps15, it is clear that it plays a critical role in constitutive as well as receptor-stimulated endocytosis (Benmerah et al.

1998). Eps15 forms homodimers of parallel stalks and tetramers of antiparallel dimers (Cupers et al. 1997), or it can heterodimerize with intersectins, scaffold proteins that provide links to signaling molecules (Sengar et al. 1999). Eps15 is localized to the necks of CCVs forming at the PM and interacts avidly with the AP2 adaptor, as well as with synaptojanin, epsin, and intersectin (Salcini et al. 1999). Epsins (Slepnev & De Camilli 2000) also bind AP2, as well as clathrin, expanding the potential for multimeric interaction between CCV coat proteins during assembly.

The AP180 and CALM proteins also apparently play a PM-specific role in stimulating clathrin assembly (McMahon 1999, Tebar et al. 1999). It is not established whether either of these proteins localizes to CCVs in the TGN, although overexpression of CALM by transfection showed a TGN, as well as a PM, localization (Tebar et al. 1999). AP180 stimulates clathrin assembly and influences the size of a clathrin coat, a factor that may be particularly important at the neuronal PM, corresponding with the high neuronal expression of AP180. CCV formation at the synapse plays a major role in recapture of synaptic vesicle (SV) contents. SV can reform either from CCVs or by resorting of SV proteins from endosomes (Slepnev & De Camilli 2000). In the former situation, regulation of the size of the recapture vesicle by AP180 is critical for generation of SV of the appropriate size (Zhang et al. 1999). Studies on the function of CALM indicate that it also plays a role in mammalian cell endocytosis (Tebar et al. 1999). CALM may also regulate vesicle size, but its function may have more to do with localization of CCV formation through interaction with PM-specific lipids (see next section). Thus the two related proteins may have diverged to play slightly different functions in stimulation of CCV formation at the PM in different cell types (Kusner & Carlin 2000). AP180 is not sufficient on its own to stimulate CCV formation. It can induce clathrin lattice formation on lipid monolayers, but these lattices do not invaginate unless AP2 is also present (Ford et al. 2001).

Finally, the late stages of CCV formation at the PM seem to depend on and incorporate the Rab5-GDI complex into the CCVs (McLauchlan et al. 1997). This complex is involved in fusion of uncoated CCVs with each other and with early endosomes. It is not known whether the formation of CCVs at the TGN requires incorporation of an equivalent Rab5-GDI complex to facilitate transport to target membrane compartments.

Lipid Interactions During CCV Formation

At the PM, regions of CCV formation are not as rich in cholesterol as are caveolae and rafts, but cholesterol content is important for CCV formation at the PM and cholesterol depletion interferes with vesicle budding (Subtil et al. 1999). The energetics of clathrin assembly with adaptor molecules is favorable at physiological pH and could potentially provide enough energy to cause spontaneous vesicle budding (Nossal 2001). However, it appears that lipid deformation both at the PM and TGN during vesicle formation has molecular assistance and that lipids play a

role in nucleation of vesicle formation at both cellular sites. A number of proteins associated with PM CCVs, including the α-subunit of AP2, dynamin, amphiphysin I, and, most recently, AP180 and epsin, have been shown to bind phosphatidyl inositol polyphosphates (PIPs) (Cremona & De Camilli 2001, Ford et al. 2001, Itoh et al. 2001). Using in vitro assays, it has been shown that interaction of dynamin and amphiphysin (separately and in combination) can deform lipid membranes, and AP180 can cause initiation of clathrin assembly at artificial membranes (Ford et al. 2001). In the case of AP180 and AP2, lipid recognition may help to initiate clathrin assembly at the PM, and for AP2, lipid binding has been implicated in clathrin assembly on lysosomal membranes (Arneson et al. 1999). In the case of dynamin and amphiphysin, lipid interactions may contribute to membrane deformation during budding and vesicle scission. Clathrin binds a class II PI3-kinase through its TD and thereby stimulates PIP formation, amplifying binding sites for coat proteins at both the PM and TGN (Gaidarov et al. 2001).

CCV-interacting proteins synaptojanin, amphiphysin, and dynamin bind endophilins. Endophilin I has lysophosphatidic acid acyl transferase activity (Huttner & Schmidt 2000, Schmidt et al. 1999). This reaction generates phosphatidic acid, potentially causing a change in membrane curvature and contributing to budding and scission. Furthermore, Eps15, AP2, the TD of clathrin, intersectins, syndapin, and endophilins (Haffner et al. 2000, Micheva et al. 1997) all have binding sites for interaction with the phosphatase synaptojanin. In vitro, synaptojanin dephosphorylates PIPs at the 3, 4, and 5 positions of the inositol ring through dual phosphatase modules (Hughes et al. 2000, Woscholski & Parker 1997). Loss of synaptojanin in mice is lethal, and their neurons accumulate uncoated CCVs (Cremona et al. 1999). This suggests that PIP binding is a critical interaction stabilizing an assembled coat and that inositide-specific phospholipase may play a role in vesicle uncoating. In yeast, mutants of synaptojanin homologs are defective in endocytosis and regulation of the actin cytoskeleton (Hughes et al. 2000). A worm with defective synaptojanin (Unc-26) is impaired in vesicle recruitment, fission, and uncoating (Harris et al. 2000). These phenotypes are consistent with a pleiotropic role for synaptojanins in regulating interactions between proteins and the PM. All these interactions between lipids and CCV-associated proteins depend on PIPs, which are present primarily on the cytosolic leaflet of the PM, in the endocytic pathway, and in the TGN.

There are TGN-specific lipid interactions of coat proteins that are implicated in CCV formation at the TGN (Roth 1999). Binding of AP1 to the TGN requires the formation of ARF1-GTP and its interaction with an additional unknown factor. This activation step for coat formation is shared by other coats in the Golgi region but is not a feature of CCV formation at the PM. ARF1-GTP formation in turn depends on guanine nucleotide-exchange factors (GEFs) of two classes. The high-molecular-weight class includes yeast Sec7p, Gea1p, and Gea2p and mammalian p200 BIG1, BIG2, and GBF1. All but GBF1 are inhibited by the fungal metabolite brefeldin A (BFA). The low-molecular-weight GEFs including ARNO, cytohesin-1, cytohesin-4, and GRP1, have PH domains and are insensitive to BFA.

ARFs and both classes of GEFs bind PIPs, and their recruitment to TGN membrane is responsive to PI3-kinase activity. In addition, ARF1-GTP stimulates phospholipase D activation, potentially resulting in increased PIP synthesis and increased sites of ARF and GEF recruitment, which could serve as a positive feedback loop for coat component recruitment. Amphiphysins inhibit phospholipase D activity, although it is not established whether this contributes to their function at the PM or whether amphiphysins also function at the TGN (Lee et al. 2000).

Role for Dynamin in Vesicle Scission

In the *shibire* mutant of *D. melanogaster*, neuronal synapses are studded with CCVs attached to the PM via collared necks, which suggests a role for dynamin in vesicle scission (Sever et al. 2000). This role is supported by the observation that dynamin can also self-assemble and form coated tubules following interaction with either synthetic or coated-vesicle-derived liposomes. Furthermore, the GTPase activity of dynamin is associated with a conformational change in dynamin tubules (Marks et al. 2001). However, there has been considerable debate about the specific function of dynamin in the cell, primarily focused on the role of the GTPase activity of dynamin. Several models have been proposed based on in vitro data and the phenotype of in vivo mutants in the GTPase and GTPase effector domains (Marks et al. 2001, Sever et al. 2000). At one end of the spectrum is the hypothesis that dynamin functions as a regulatory GTPase, like most characterized GTPases, and that it attracts other proteins that actually mediate vesicle scission. In this model, the function of self-assembly of dynamin is primarily to stimulate the GTPase activity of dynamin and perhaps act as a sensor of vesicle closure. Alternative hypotheses suggest that the self-assembly of dynamin is the mechanical force behind vesicle scission, either through formation of a garrote causing membrane constriction or through intrinsic spring-like action due to a conformational change causing membrane rupture. It is likely that the role of dynamin in CCV scission involves both of these mechanisms. As more dynamin mutants are studied for their effects on endocytosis, as well as their in vitro assembly phenotype, it appears there is supportive data for more than one mechanism of action (Marks et al. 2001). Presumably dynamin self-assembly does contribute to the mechanics of scission, along with a conformational change induced by GTP hydrolysis. However, self-assembly also activates GTPase activity, which could very well play a regulatory role in the recruitment of other proteins involved in scission. In addition, dynamin has recently been implicated in late stages of vesicle invagination during CCV formation at the PM, possibly a function of its interaction with partner proteins that may occur prior to dynamin self-assembly and vesicle scission (Hill et al. 2001). Although there is conflicting data about the role of dynamin in CCV function at the TGN (Altschuler et al. 1998, Kreitzer et al. 2000), it has been suggested that the dynamin that functions at the TGN might be a different splice variant than the ones that function at the PM, again highlighting the different requirements for vesicle formation at these two membranes (McNiven et al. 2000).

Coat Disassembly

Dissociation of coat proteins from lipids and from each other are both needed for CCVs to uncoat. Dissociation of AP2 and AP180/CALM from the PM is likely to be mediated by synaptojanin (Cremona et al. 1999), whereas GTP hydrolysis by ARF may contribute to destabilization of AP1 binding to lipids at the TGN (Roth 1999). Furthermore, the phosphatase activity associated with auxilins (Lemmon 2001), involved in uncoating, could conceivably destabilize coat protein-PIP interaction. It is not clear how lipid dissociation processes are coupled to disassembly of the clathrin lattice. Clathrin and adaptors dissociate from CCVs in separate steps (Hannan et al. 1998), and their dissociation may be promoted by adaptor phosphorylation in the clathrin-binding domain of the β-subunits (Wilde & Brodsky 1996). Following clathrin dissociation, AP2 does not appear to be present in significant levels on endosomes, although it may help mediate the aggregation of CCVs that have lost clathrin (Beck & Keen 1991a). AP1 has recently been shown to bind to a kinesin superfamily protein (KIF13A), and this binding has been implicated in cargo transport in the TGN, where vesicles with AP1 but not clathrin are observed (Nakagawa et al. 2000). Thus AP1 appears to function temporally after clathrin disassembly. However, when cells are treated with BFA, causing ARF1 to dissociate from TGN membranes, AP1 and AP3 also dissociate from membranes and, in the case of AP1, associated clathrin molecules disassemble (Liu et al. 2001b, Robinson & Kreis 1992, Wong & Brodsky 1992).

Hsc70 was shown to be able to trigger clathrin basket disassembly in vitro in the 1980s (Schlossman et al. 1984). Its role in this process in vivo was not substantiated until more recently, via antibody injection studies (Honing et al. 1994) and the use of dominant-negative mutants (Newmyer & Schmid 2001) of hsc70, which disrupt cellular functions of CCVs. The recently defined role in cellular CCV disassembly for auxilin, with its DnaJ homology, also strengthens the implication of hsc70 as a regulator of clathrin disassembly (Lemmon 2001). Hsc70 is an ATP-dependent chaperone, which binds relatively hydrophobic peptides or exposed protein sequences (Bukau & Horwich 1998). The structures of the peptide-binding site and the ATPase domain have been determined independently (Flaherty et al. 1990, Zhu et al. 1996), and the latter has an ATPase domain that resembles that of actin (Flaherty et al. 1991) (Figure 3). The recruitment of hsc70 to CCVs by auxilin followed by ATP binding and hydrolysis by hsc70 may cause a conformational change in assembled clathrin, triggering disassembly. In vitro studies revealed that a sequence unique to clathrin LCa, exposed in the presence of calcium, can stimulate hsc70 ATPase activity (DeLuca-Flaherty et al. 1990). Subsequent studies indicated that hsc70, in conjunction with auxilin, can cause clathrin disassembly in the absence of CLCs (Lemmon 2001). However, the LCa sequence could still contribute a regulatory role in in vivo uncoating. Indeed, cells lacking LCa have a reduced rate of CCV uncoating (Acton et al. 1993).

Following ATP hydrolysis and coat dissociation, ADP-bound hsc70 and clathrin form a stable complex, presumably maintaining the pool of cytosolic clathrin

in a disassembled state. Included in this complex is valosin-containing protein (100 kDa), an ATP-binding protein with a suggested chaperone function (Pleasure et al. 1993). A second protein, which may also contribute to sequestration of cytosolic clathrin, is the giant protein (p619) with numerous regulatory protein domains (Rosa & Barbacid 1997). These include a domain homologous to cell cycle regulator RCC1, a guanine nucleotide exchange factor for Ran, seven β-repeats characteristic of the β-subunit of heterotrimeric G proteins, three SH3 domains, a leucine zipper, and a domain homologous to the E3 ubiquitin-protein ligases. The giant protein forms a cytosolic ternary complex with clathrin and hsp70, but it also stimulates guanine nucleotide exchange on ARF1 and rab proteins, perhaps contributing to clathrin recruitment at the TGN.

INTRACELLULAR LOCALIZATION OF CCV FORMATION AND CARGO INTERACTION

In mammalian cells, intracellular localization of CCV formation is specified by the localization of the adaptor molecules AP1 and AP2. Binding of both adaptors to cellular membranes is independent of clathrin function. Studies of adaptor interactions with membranes suggest that both protein determinants (Mahaffey et al. 1990, Mallet & Brodsky 1996, Seaman et al. 1996) and lipid determinants are recognized (see previous section on lipid interactions). In yeast, CCVs can form and function in the complete absence of adaptors or yeast AP180 (Huang et al. 1999, Yeung et al. 1999). Thus nucleation of clathrin assembly on yeast membranes can occur through other mediators of interaction between clathrin and lipids. Presumably the yeast adaptors share some of the properties of mammalian adaptors, and their incorporation into yeast CCVs might at least have a role in cargo recognition, even if adaptors are not absolutely required for CCV formation. It is notable that key cargo recognition signals in yeast differ from those that play a key role in mammalian receptor sequestration by AP1 and AP2 into CCVs, which suggests that other types of "adaptor" molecules may also be involved in receptor recognition in yeast. As with the mechanics of clathrin assembly and disassembly discussed above, the process of CCV nucleation and cargo recognition is different for CCVs at the PM or TGN in both yeast and mammalian cells. For this reason, the events at each membrane are discussed separately.

Plasma Membrane CCV Nucleation and Receptor Sequestration

AP2 binding to PM fragments in vitro can be ablated by proteolysis, which suggests that AP2 recognizes protein determinants in addition to PIPs (Mahaffey et al. 1990). Candidate determinants for nucleation of AP2-PM binding include the synaptotagmins (von Poser et al. 2000). Neuronal and nonneuronal synaptotagmins bind AP2 through their C2B domains. Synaptotagmins associate with the PM via calcium-dependent phospholipid binding and by recognition of PIPs. Expression of a

nonneuronal synaptotagmin lacking the AP2-binding domain can inhibit endocytosis in nonneuronal cells as a dominant-negative mutant. In addition, cells expressing this mutant were observed to have fewer coated pits. However, synaptotagmin is expressed at high levels in neuronal cells and the nonneuronal synaptotagmins are expressed at considerably lower levels. Thus it is not clear whether these proteins play a general role in nucleating AP2 binding.

Two additional protein families have been identified that may play a role in nucleation or localization of CCVs to the PM. Both proteins interact with Eps15, and overexpression of functional fragments affects endocytosis. These are the Numb proteins and stonin 2, both ubiquitously expressed mammalian homologues of *D. melanogaster* genes involved in neuronal function. The Numb family of proteins localize to CCVs at the PM and in the TGN and bind the α-appendage of AP2 (Santolini et al. 2000). Stonin 2 has some homology to the μ-subunits of adaptors and interacts both with Eps15 and with synaptotagmin, which suggests it could serve as a link between PM-associated proteins and CCV components (Martina et al. 2001).

CCVs mediate endocytosis of nonsignaling receptors, such as receptors for nutrients, whether or not ligand is bound by these receptors. The role of such cargo in coated pit nucleation is not completely established. Peptides containing the YXXΦ endocytosis motif present in a number of these receptors can enhance the interaction of AP2 with synaptotagmin in vitro and can enhance the binding of AP2 to membrane fragments of nonneuronal cells (Haucke & De Camilli 1999). However, at very high concentrations, soluble cytoplasmic domains of receptors can inhibit AP2 binding to membranes (Chang et al. 1993). Conversely, interaction of AP2 with PIP3 can increase the affinity of AP2 for cytoplasmic tail peptides (Gaidarov & Keen 1999). Overexpression of transferrin receptor in chicken cells has been reported to increase coated pit numbers (Miller et al. 1991). However, in other cell types, no effect of receptor overexpression on coated pit formation has been observed (Brown et al. 1999, Warren et al. 1997). It has recently been noted that phosphorylation of the μ-subunit of AP2 increases its affinity for peptides with tyrosine-based motifs and can increase AP2 binding to membrane fragments (Fingerhut et al. 2001). The physiological pathways that influence μ-subunit phosphorylation have not been characterized. A reasonable conclusion based on the sum of these observations about cargo-AP2 interactions is that AP2 membrane interactions partially involve recognition of receptor cytoplasmic domains, but that these latter interactions are secondary to independent interactions mediating AP2-membrane binding, although the two binding events may be mutually stimulatory.

Receptors that trigger signaling pathways on ligand binding are generally endocytosed in a ligand-dependent fashion. In mammalian cells, these fall into two well-characterized categories—receptor tyrosine kinases (RTKs) and G-protein coupled receptors (GPCR). In yeast, several cell surface proteins are internalized upon ubiquitination. Ste6p (a peptide transporter) and pheromone receptors Ste2p and Ste3p are monoubiquitinated in response to phosphorylation induced by ligand binding. This monoubiquitination is the cargo recognition signal for receptor

down-regulation by CCVs (Wendland et al. 1998), but whether this ubiquitin-mediated uptake depends on clathrin is not clear. In mammalian systems, ubiquitination has also been associated with the uptake of growth hormone receptor (van Kerkhof et al. 2001). Thus ubiquitination may represent one of several general mechanisms by which receptors are sequestered in transport carriers, possibly CCV, in response to ligand binding. The adaptor molecule that recognizes the ubiquitin signal in yeast has not been identified, although mutation studies in mammalian cells suggest that ubiquitin may have sequences related to the dileucine motif recognized by AP2 and AP1 adaptors (Nakatsu et al. 2000). Uptake of yeast a-factor receptor Ste3p could also rely on the presence of the sequence NPFXD in the cytoplasmic domain, which is known to be recognized by CCV components. How recognition of this sequence is ligand gated has yet to be established (Tan et al. 1996).

For mammalian signaling receptors, it does not appear that ligand binding stimulates CCV formation (Santini et al. 1998). However, ligand binding does stimulate recognition of the receptor by AP2 molecules. In the case of RTKs, phosphorylation of the receptor cytoplasmic domain upon ligand binding promotes a conformational change that exposes a binding site for AP2 (Chen et al. 1989). For epidermal growth factor receptor (EGFR), this binding site has particularly high affinity, unlike the AP2-binding site in constitutively internalized receptors, and receptor-AP2 interaction is detectable by immunoprecitation following receptor activation (Sorkin & Carpenter 1993). Coprecipitation of constitutively internalized receptors with AP2 is not easily detected. The high-affinity AP2-binding site can be eliminated from EGFR and the receptor can still be endocytosed by CCVs, which suggests that additional interactions with the endocytic machinery function in EGFR uptake. Consistent with this result is the observation that mutations in the μ-subunit of AP2 that block recognition of tyrosine-containing internalization motifs can abrogate transferrin receptor internalization but not affect EGFR uptake (Nesterov et al. 1999). Taken together, these data on the RTK EGFR suggest that the endocytosis of signaling receptors in response to ligand binding is controlled by several independent interactions, in addition to μ-subunit recognition of the receptor's cytoplasmic domain. Ligand binding to both EGFR and nerve growth factor receptor results in stimulation of src family kinases, which phosphorylate CHC on residue 1477, in the hub domain (Beattie et al. 2000, Wilde et al. 1999). After ligand binding to either RTK, clathrin recruitment to the PM is observed, and for EGFR, this recruitment can be abrogated by inhibition of clathrin phosphorylation (Wilde et al. 1999). The implication of these observations is that ligand binding leads to clathrin phosphorylation, which enhances clathrin localization to the PM for participation in ligand-gated endocytosis.

Enhanced interaction with CCV components is also observed during internalization of ligand-activated GPCR. In this case, ligand binding leads to phosphorylation of the GPCR, which is then recognized by cytoplasmic proteins of the β-arrestin family (Ferguson 2001). The β-arrestin members of the arrestin family interact directly with the TD of clathrin (Goodman et al. 1997) and are involved

in promoting endocytosis of a subset of GPCRs. The β-arrestin sequence contains a predicted clathrin box (Krupnick et al. 1997), and a peptide representing this sequence has been cocrystallized with CHC TDs (ter Haar et al. 2000). The fact that its binding site overlaps with that of the clathrin box peptide from the β3-subunit of AP3 initially suggested that arrestins might be substitute adaptors for GPCR cargo in CCVs. However, it appears that arrestins also bind AP2, independently of clathrin, and that AP2 binding is critical for their role in GPCR uptake. Thus arrestins are linker proteins for GPCR sequestration in clathrin-coated pits with AP2 at the PM. In conjunction with a role at the PM, β-arrestin has been shown to have a PIP-binding site, which must be intact for arrestin function in GPCR uptake (Gaidarov et al. 1999a).

Following ligand binding, many signaling receptors are initially localized in lipid rafts at the PM, where kinases that mediate downstream signaling are also localized (Dykstra et al. 2001). Some of these receptors are internalized by clathrin-independent mechanisms following signaling, whereas others are eventually down-regulated in CCVs (Mineo et al. 1999). The mechanisms operating for internalization pathways that are not mediated by clathrin have yet to be established, as does the relationship between receptor signaling from rafts and eventual uptake by CCVs.

TGN CCV Nucleation and Receptor Sorting

AP1 membrane-binding dynamics are different from those of AP2 in a number of features. ARF1 activation by GTP binding is required for AP1 to bind membranes at the TGN (see previous section on lipid interactions). How this signal is physiologically linked to the requirement to sort proteins is not determined. It has also been proposed, based on indirect evidence, that there is an AP1 docking protein whose interaction with AP1 depends on ARF activation (Traub et al. 1993). Several membrane-associated, AP1-binding proteins have been identified, but none has been demonstrated to have a docking function (Mallet & Brodsky 1996, Seaman et al. 1996).

γ-Synergin binds to the ear domain of the γ-subunit of AP1 and has the potential to mediate interaction of AP1 with NPF-containing proteins through its EH domain (Page et al. 1999). One such family of proteins that can bind γ-synergin is the SCAMP family (Fernandez-Chacon et al. 2000). These proteins were originally discovered as components of exocrine secretory vesicles, and some members of the family have a ubiquitous tissue expression pattern. There is evidence that the SCAMPs cycle between the PM, the TGN, and endosomes. The distribution of SCAMPs correlates with a potential role in AP1-dependent CCV nucleation. Perhaps, similar to synaptotagmin, their role is critical in such specialized cell function as CCV nucleation during secretory granule formation, but their involvement in general CCV function is not yet defined. However, expression of a dominant-negative mutant fragment of SCAMP1 can inhibit transferrin endocytosis, indicating that SCAMP interacts with proteins critical for the transferrin uptake pathway

and can disrupt its cycling (Fernandez-Chacon et al. 2000). γ-Synergin itself does not act as a nucleating protein for CCV formation at the TGN because its highest affinity interaction in the cell appears to be its binding to the γ-appendage of AP1 (Page et al. 1999). However, it is still a candidate for involvement in nucleation by interaction with other proteins. The recent demonstration that GGA proteins, which bind γ-synergin, can recruit clathrin to TGN membranes in an ARF-dependent fashion suggests these proteins play a role in nucleation of CCV formation at the TGN (Puertollano et al. 2001). Binding of GGA to ARF can displace AP1. It is possible that GGAs and AP1 participate in formation of different CCVs or act sequentially during CCV nucleation.

Cargo recognition by CCVs in the TGN can be direct, by μ1-subunit interactions with cytoplasmic domains bearing the YXXΦ motif (Marks et al. 1997). In addition, the AP1-binding protein PACS1 acts as a connector, recognizing phosphorylated cytoplasmic domain sequences containing an acidic cluster of amino acids with two serines that are a target for casein kinase II phosphorylation (...EECPSDSEEDE..) (Wan et al. 1998). PACS1 is involved in CCV-mediated TGN sequestration of furin and sorting of both forms of mannose-6-phosphate receptor (M6PR) through recognition of their phosphorylated cytoplasmic domains, and its function has been compared with that of β-arrestin during uptake of GPCR at the PM. It is notable that both cation-independent and cation-dependent M6PRs interact with other tail-binding proteins, including TIP47, which may be involved in their recycling to the TGN (a clathrin-independent step) and with AP2, which recognizes a site that overlaps with TIP47 binding (Orsel et al. 2000). Overexpression of M6PR has been observed to increase AP1 localization to the TGN (Le Borgne et al. 1993), which may be a function of PACS1 binding. In yeast, CCVs play a major intracellular role in maintaining sequestration of proteins in the TGN, generally measured by the TGN targeting of Kex2p, a protease similar to furin (Payne & Schekman 1989). CLC is a target for extensive phosphorylation in yeast, and expression of mutant CLCs that cannot be phosphorylated partially disrupts Kex2p localization (Chu et al. 1999). Thus in yeast, CLC phosphorylation may play a role in regulation of TGN CCV function, whereas in mammalian cells, CHC phosphorylation regulates CCV function at the PM.

CCV-Cargo Interactions at Other Cellular Membranes

Clathrin coats have been observed on both endosomes and lysosomes (Stoorvogel et al. 1996, Traub et al. 1996). Endosomal clathrin coats appear to contain either AP1 or neither AP1 nor AP2. The lysosomal coats are asssociated with AP2. It had been proposed that the endosomal CCVs might be sorting proteins for recycling. However, there is no kinetic argument that receptor recycling is facilitated by CCVs, as the rates of transferrin receptor recycling correspond to rates of bulk flow lipid recycling (Mayor et al. 1993). In addition, transferrin receptors lacking their cytoplasmic domains recycle at the same rates as wild-type receptors (McGraw & Maxfield 1990). Consistent with these observations is the finding that expression

of the clathrin hub in cells, which has a dominant-negative mutant effect on CCV-mediated pathways, has a minimal effect on transferrin receptor recycling rates (Bennett et al. 2001). However, in HeLa cells expressing the hub molecule, the intracellular distribution of endosomes is altered. This suggests that clathrin coats on endosomes serve a transport function that influences organelle localization, but that they do not function in direct sorting of receptors from endosomes to the PM. Recycling to the basolateral membrane in polarized cells appears to be affected by BFA, which suggests that CCVs might play a role in influencing the directionality of transport from endosomes (Futter et al. 1998). The function of clathrin coats observed on lysosomes remains to be determined. The fact that these coats can form under physiological conditions in semi-intact cells suggests that such CCVs might mediate some kind of retrograde transport from lysosomes (Traub et al. 1996).

RELATIONSHIP OF CCVS TO THE CYTOSKELETON

Mammalian PM has cortical actin, but yeast has a more elaborate actin cytoskeleton at the PM. Correspondingly, endocytosis in yeast is highly dependent on the actin cytoskeleton, and CCVs appear to play a less essential role (Baggett & Wendland 2001, Wendland et al. 1998). Yeast lacking CHCs or CLCs exhibit reduced, but not completely impaired, endocytosis (Baggett & Wendland 2001). In contrast, mutations affecting the actin cytoskeleton generally abrogate endocytosis in yeast, so it has not been possible to clearly establish the relationship between CCV-mediated uptake and actin-dependent uptake. Analysis of *Dictyostelium discoideum* expressing CHCs labeled with green fluorescent protein at the C terminus reveals that a subset of labeled CCVs moves along linear intracellular tracks, which again suggests a cytoskeletal interaction for CCV proteins (Damer & O'Halloran 2000).

In mammalian cells, CCV components interact with actin-binding proteins, but the interplay between CCV formation and the actin cytoskeleton also has yet to be clearly defined (Qualmann et al. 2000). Inhibitors of actin function have variable effects on endocytosis, depending on the cell type treated (Fujimoto et al. 2000). However, endocytic CCVs in HeLa and other cell types are observed to align with actin filaments in the cell periphery, and actin depolymerization disrupts this distribution pattern (E.M. Bennett & F.M. Brodsky, unpublished data). Furthermore, a study of the dynamics of clathrin labeled with green fluorescent protein shows that recruitment of clathrin to fixed sites at the PM is dependent on an intact actin cytoskeleton (Gaidarov et al. 1999b). These interactions of CCVs with the actin cytoskeleton could be explained by a number of molecular links between CCV components and cytoskeletal components. Mammalian clathrin binds ankyrin (Michaely et al. 1999), which has a well-established function in red blood cells, linking the actin-spectrin network to the plasma membrane. Clathrin also binds to the Hip1R (huntingtin-interacting protein 1–related) protein, whose yeast

homologue, Sla2p, functions in actin-dependent endocytosis in yeast (Engqvist-Goldstein et al. 1999, 2000). Hip1R also has an ENTH domain, which could mediate its binding to PIPs, and it has been localized to CCVs at the PM and TGN. These molecular interactions predict a role for actin in defining sites of CCV formation by delivery of coat components to nucleation sites and/or by establishing nucleation sites relative to the cortical actin network so that vesicle budding can occur in between actin filaments (Fujimoto et al. 2000). Two more roles for actin in CCV function are also suggested by molecular data. First, actin could potentially play a role in vesicle scission, via the interaction of amphiphysin and dynamin (Wigge & McMahon 1998). The yeast homologs of amphiphysin are both actin-interacting proteins and their deletion affects endocytosis (Munn et al. 1995). A third potential role for actin in CCV function could be to propel CCVs away from the membrane by formation of actin tails. CCVs are short-lived, particularly at the PM; however, an actin tail could play a transient role in detachment or translocation of vesicles. In support of this hypothesis, the syndapin/pacsin proteins that interact with dynamin and synaptojanin also interact with Wiskott-Aldrich syndrome protein family members (Qualmann & Kelly 2000). The Wiskott-Aldrich syndrome proteins can nucleate actin tail formation (Modregger et al. 2000, Qualmann et al. 1999).

The structure of the TGN is dependent on the microtubule cytoskeleton. The fact that AP1 and a kinesin interact suggests that coated or uncoated vesicles could move along microtubules at the TGN. However, CCVs with actin tails have been observed in the vicinity of the TGN (Frischknecht et al. 1999), indicating that actin may also play a role in CCV excision or transport at the TGN. Finally, the novel form of clathrin highly expressed in muscle, CHC22, appears to be a form of clathrin that has evolved stronger interactions with the actin cytoskeleton (Liu et al. 2001b) than those exhibited by conventional clathrin CHC17. In nonmuscle cells, the intracellular distribution of CHC22 at the TGN is dramatically altered by disruption of the actin cytoskeleton, whereas CHC17 distribution is relatively stable. When the function of CHC22 in muscle tissue is established, it is likely to reveal more about the relationship between the clathrins and actin.

TISSUE-SPECIFIC CCV FUNCTION AND SPECIALIZED CLATHRIN FUNCTION

Cell Polarity

In addition to fundamental roles in receptor-mediated endocytosis and in sorting proteins destined for the endocytic pathway (such as lysosomal proteases) from the TGN, CCVs play a number of specialized roles in different tissues. Most tissue culture cell lines are not polarized, as the same cells might be in situ. When polarization is induced in epithelial cell lines, it is clear that CCVs play important roles in maintaining polarity. CCVs have been implicated in basolateral targeting of

some receptors. Basolateral targeting signals for these receptors are similar to and frequently overlap with recognition signals for binding of the μ-subunits of AP1 and AP2 (Mostov et al. 2000). As mentioned in the section on adaptor biochemistry, there is a μ1B alternate subunit of AP1, creating an AP1 that functions specifically in the basolateral sorting pathway for the low-density lipoprotein and transferrin receptors (Fölsch et al. 1999). Defects in μ1B do not affect apical targeting of proteins. The apical membrane of polarized cells has more raft-like properties than does the basolateral membrane, but experiments with the clathrin hub mutant indicate that at least some of the endocytosis that occurs from the apical membrane of cells is clathrin mediated (Altschuler et al. 1999).

Neuronal Function

In neurons, CCVs recapture synaptic vesicle (SV) proteins, and CCV formation at the neuronal synapse has been well-characterized. In fact, many of the proteins involved in regulation of CCV formation were identified through analysis of molecular components involved in neuronal CCV formation, and their functions have been reviewed recently (Slepnev & De Camilli 2000). In keeping with specialized features of CCV formation in neurons, many CCV components and regulators have neuron-specific forms. Both forms of mammalian CLCs have neuronal splice variants, as do the α- and β2-subunits of AP2 (Pley & Parham 1993). The additional sequences in the neuronal forms of these proteins may mediate their transport to the synapse or may even influence their regulation. It is notable that only CCVs with neuronal CLCs interact with calmodulin (Pley et al. 1995). For many proteins involved in regulation of CCV formation, their neuronal forms are encoded by different genes, with some modified by variable splicing. These include AP180, auxilin, amphiphysin, endophilin, intersectin, synaptojanin, syndapin, pacsin, SCAMPs, and synaptotagmin (see Table 1 for references). As discussed above, in the section on the mechanics of CCV formation, one of the requirements for CCV at the synapse is control of vesicle size (Slepnev & De Camilli 2000). An additional requirement that distinguishes neurons from other cells is a need for synchronicity in vesicle formation and transport. The existence of specialized neuronal forms of CCV proteins likely reflects these specialized needs, although comparative studies between neuronal and nonneuronal forms of these proteins have yet to be done. Although it is not clear whether AP3 adaptors participate in CCV formation, they have been implicated in SV formation. The β3-subunit has two forms, one of which is primarily expressed in neurons (Newman et al. 1995). In keeping with the importance of clathrin at the neuronal synapse, many CCV components have been identified in *C. elegans* and *D. melanogaster* by isolation of mutants in neuronal function. Furthermore, CLC is one of the proteins that is up-regulated in the sensory neurons of aplysia during the development of long-term facilitation (Hu et al. 1993), and CHC is up-regulated following stimulation of recognition memory in rats (Solomonia et al. 1997).

Regulated Secretion

CCVs play an important role in cells that undergo regulated secretion. Partial clathrin coats are observed on immature secretory granules (ISGs), which also have binding sites for ARF1 and AP1 (Austin et al. 2000). ISGs form from the TGN and contain aggregated protein or proprotein destined for the mature secretory granule (SG). It has been proposed that the clathrin coats on ISGs represent CCVs that are budding to remove TGN membrane proteins that should be sorted to destinations other than the SG. A number of morphological studies and studies of the relative biosynthetic transport pathways of lysosomal enzymes and SG proteins support this hypothesis (Dittié et al. 1996, Kuliawat et al. 1997). Introduction of the clathrin hub mutant into insulin-producing cells results in enhanced proteolysis of insulin fragments (Molinete et al. 2001). This observation suggests that the role of a CCV is to remove lysosomal enzymes from an ISG as it matures into an SG and that this removal is inhibited in hub-expressing cells. The formation of melanosomes appears to involve transport pathways that are distinct from those contributing to SG formation in that they depend on AP3 function (see human diseases and mouse mutants section on CCV malfunction).

Immune System Function

In the immune system, CCVs are involved in regulation of the cell surface expression of antigen receptors and signaling molecules and in sorting molecules to the pathway of antigen processing and presentation by class II histocompatibility molecules. CCVs have been implicated in the uptake of both the T cell receptor (TCR) (Boyer et al. 1991) and the B cell receptor (BCR) (surface immunoglobulin) (Salisbury et al. 1980) following cross-linking by receptor-specific antibodies. For T cells, endocytosis and recycling of the TCR during contact with antigen is believed to contribute to continuous signaling by a T cell and sustaining T cell activation (Liu et al. 2000). For B cells, uptake of the BCR can deliver antigen to the antigen processing pathway, leading to antigen display on the cell surface and stimulation of helper T cells needed for B cell differentiation. Recent studies have suggested that the membrane region where TCRs are concentrated during T cell stimulation by a target cell (the so-called immunological synapse) has raft-like properties (Langlet et al. 2000). Signaling B cell receptors have also been localized to rafts (Dykstra et al. 2001). Therefore, the mechanism of internalization of stimulated TCR and BCR should be further investigated as mechanisms of uptake from rafts are better defined. CCVs also control the surface expression of CTLA-4, a co-receptor on T cells that negatively modulates T cell activation and contributes to the development of T cell tolerance. During T cell activation, CTLA-4 is phosphorylated on a tyrosine residue that forms part of an AP2 recognition site (Nakaseko et al. 1999). The phosphorylation therefore blocks uptake of the CTLA-4 molecule, allowing it to perform its signaling function and modulate T cell activation. On loss of TCR signaling, CTLA-4 is no longer phosphorylated

and can be endocytosed, leading to its down-regulation and resting state localization in the TGN. CCVs play a major role in the intracellular transport of class II major histocompatibility complex (MHC) molecules, whose recognition by helper T cells leads to stimulation of an antibody-mediated immune response. Class II molecules must intersect the endocytic pathway during their biosynthetic assembly, so that they can acquire an antigenic peptide. The intracellular compartment in which class II molecules are loaded with peptides has lysosome-like properties, and similar to lysosomes, CCVs are involved in targeting molecules to this organelle (Geuze 1998). The invariant chain molecule that chaperones class II molecules through their unique biosynthetic pathway is sorted in CCVs (Hofmann et al. 1999, Liu et al. 1998). In addition, the HLA-DM molecule that catalyzes the interaction between class II molecules and peptides is localized to the peptide loading compartment by transport in CCVs (Liu et al. 1998). The specialized class I MHC molecule analog CD1 is also targeted to the peptide loading compartment through a signal for uptake in CCVs (Sugita et al. 2000). This intracellular localization facilitates the binding of glycolipid antigen by CD1, a pathway involved in the immune response to mycobacteria.

Muscle Function

The human CHC encoded on chromosome 22, CHC22 is most highly expressed in skeletal muscle (Liu et al. 2001b). CHC22 is also expressed at low levels in nonmuscle cells, where it is associated with the TGN via interaction with AP1 and AP3, as well as with elements of the actin cytoskeleton. Expression of the hub fragment of CHC22 in nonmuscle cells disrupts protein sorting in the TGN, but whether this is due to a direct effect on protein sorting or whether CHC22 is involved in some structural aspect of TGN organization is not known. CHC17 is also expressed in differentiated skeletal muscle, which suggests the hypothesis that CHC22 performs a muscle-specific function. There is also a form of amphiphysin II that is highly expressed in muscle cells (Butler et al. 1997, Wigge & McMahon 1998) that is a potential interaction partner for CHC22. However, this form lacks known clathrin binding sequences. Specialized pathways of membrane transport in muscle cells include the biogenesis of the T-tubule system. A muscle-specific form of caveolin has already been implicated in this process (Parton et al. 1997), and amphiphysin II is localized to T-tubules (Butler et al. 1997), so it would not be surprising if specialized clathrin might also play a role. In addition, the flux of glucose transporters is highly controlled in muscle cells, and a specialized form of clathrin could be involved in their sequestration in the endosomal pathway or their regulated expression on the cell surface (Simpson et al. 2001). It is also possible that muscle-specific forms of CCV proteins play a role in establishing protein configurations at post-synaptic sites in muscle. Alternatively, CHC22 might not function primarily in intracellular membrane transport pathways in muscle but could potentially play a role in the structural organization of muscle proteins.

CCV MALFUNCTION: INHIBITORS, MUTANTS, AND DISEASES

Molecular Inhibitors

For many years molecular inhibition of CCV function was limited to the use of low-pH shock treatment, potassium depletion, and treatment of cells with BFA (Robinson & Kreis 1992, Wong & Brodsky 1992). None of the treatments is ideal for implicating CCV in transport, as all have pleiotropic effects on cell function. Low pH and potassium depletion do appear to cause nonproductive assembly of clathrin cages in the cytoplasm adjacent to coated pits (Heuser 1989, Heuser & Anderson 1989). However, BFA affects the binding of AP3, COPI, and COPII to membranes in the Golgi and TGN, as well as disrupting the binding of AP1 (Roth 1999). A more recent approach in a substantial number of the studies in which the function of CCV-associated proteins in mammalian cells has been observed utilizes the overexpression of fragments or mutants of these proteins to test for dominant-negative effects. A potential problem with such approaches is the continued existence of an endogenous pool of normal protein and the possibility that a protein fragment binding a regulatory protein may be inhibitory by an indirect rather than a direct effect. Because so many of the regulatory proteins interact with a network of other proteins (Figure 4), the hierarchy of interaction and the direct effects of a mutant fragment are difficult to define. For example, expression of the SH3 domains of proteins that regulate CCV formation generally affects endocytosis, but because these domains can interact with so many partners, the direct and indirect effects are difficult to distinguish. Several molecular inhibitors have been widely used or are particularly well-characterized, so their inhibitory behavior can be more easily interpreted. Dynamin I, with the same mutation as the *shibire D. melanogaster* mutant dynamin, has been widely used. Initially, it was characterized as an inhibitor of CCV-mediated endocytosis (Damke et al. 1994), and later it was shown to also inhibit internalization from caveolae (McNiven et al. 2000). It is reasonable to assume that if a transport step is not affected by this dominant-negative mutant, then it involves neither of these pathways. One caution regarding this mutant is that cells expressing it over time tend to compensate by up-regulating non-CCV-mediated uptake pathways (Damke et al. 1994). The hub fragment of clathrin has been used in a number of studies to block CCV-mediated uptake and CCV sorting at the TGN (Altschuler et al. 1999, Liu et al. 1998, Lu et al. 1998, Trejo et al. 2000). This inhibition appears to be a direct effect of CCV function, as the mechanism of action involves hub fragments binding up endogenous CLCs so that CHC assembly and disassembly can no longer be regulated. Expression of the μ2-subunit of AP2 with a mutation in the site recognizing tyrosine-containing internalization motifs results in very specific inhibition of AP2-mediated endocytosis of constitutively internalized receptors (Nesterov et al. 1999). As noted above, this mutation does not affect the uptake of EGFR, which appears to have a specialized interaction with endocytic CCV components,

independent of the tyrosine-containing motifs present in the cytoplasmic domain of the receptor.

Deletion Mutants in Model Organisms

Genes encoding proteins involved in CCV formation have been identified and deleted in *Saccharomyces cerevisiae, D. melanogaster, D. discoideum, C. elegans*, and *Mus musculus*. Of interest is the fact that deletions of many of the regulatory proteins are not lethal. However, given the extensive redundant interactions of CCV proteins with proteins linked to the cytoskeleton (Figure 4), lack of lethality is perhaps not so surprising. It was initially a shock when the first yeast strain deleted for CHC was found to survive (Payne & Schekman 1985). However, now that we know that membrane traffic in yeast is highly dependent on the actin cytoskeleton, it appears that clathrin is used primarily to fine-tune transport, and yeast membrane traffic can occur to a limited extent, sufficient to sustain life, without clathrin (Baggett & Wendland 2001). In other strains of yeast, clathrin deletion has been found to be lethal, and the identification of suppressor mutations has been useful for identifying additional players in membrane traffic. As mentioned above, deletion mutants lacking elements of CCV formation in *D. melanogaster* and *C. elegans* often present with a neuronal phenotype. Some of the nonneuronal phenotypes are noted here. *D. discoideum* without clathrin has defects in many processes that are also associated with the actin and microtubule cytoskeleton. These include cytokinesis, establishment of cell polarity, pseudopod formation, and uropod stability and motility (O'Halloran 2000, Wessels et al. 2000). The cytokinesis defect could be due to a need for clathrin to transport membrane proteins to sites where cytokinesis is initiated (O'Halloran 2000). Clathrin has been observed in mammalian mitotic spindles (Okamoto et al. 2000), and there is a strong connection between membrane traffic and cytokinesis (O'Halloran 2000). *D. melanogaster* without clathrin has defective spermatogenesis (Fabrizio et al. 1998), likely owing to the importance of membrane compartmentalization during this process. Organisms with defects in AP3 proteins (*D. melanogaster* and *M. musculus*) have pigmentation defects (Odorizzi et al. 1998), whereas defects in AP1 subunits in *M. musculus* are lethal (Meyer et al. 2000, Zizioli et al. 1999).

Human Diseases

As might be expected, there are many human diseases associated with defects in CCV formation. In fact studies of familial hypercholesterolemia by Anderson et al. (1977) provided a key observation that contributed to defining the role of CCVs in receptor-mediated endocytosis. In one form of this disease, the recognition signal for AP2 in the cytoplasmic domain of the low-density-lipoprotein receptor is mutated, leading to a lack of endocytosis and accumulation of low-density lipoprotein in the blood. Another metabolic disease that apparently results from altered receptor trafficking by CCVs is hereditary hemochromatosis (Enns 2001). The

defective gene in hereditary hemochromatosis is a member of the extended family of molecules with homology to class I MHC molecules. This protein, called HFE, binds transferrin receptor and regulates its ability to internalize iron-loaded transferrin. The defective HFE is unable to interact with transferrin receptor, and regulation of iron metabolism is disrupted.

CCVs also play a critical role in human viral infections. CCVs are the route of entry for infection by some viruses, including influenza and Semliki Forest virus, but not others, such as polio virus (DeTulleo & Kirchhausen 1998). In the case of influenza, entry into the acidic endosome induces a conformational change in the viral hemagglutinin, causing it to become a fusion protein, which mediates viral envelope fusion and delivery of viral RNA into the cytosol. CCVs also play a role in viral pathogensis. The HIV nef protein induces internalization of the CD4 molecules on helper T cells and causes reduction of expression of class I MHC molecules on infected T cells. Both effects appear to be part of the HIV immune evasion strategy. The reduction of class I molecules abrogates recognition of infected cells by cytotoxic T cells (Collins et al. 1998), and the reduction of CD4 expression interferes with helper T cell function. For alteration of CD4 traffic, it appears that nef functions by causing CD4 internalization in CCVs (Piguet et al. 1998). For reduction of class I MHC molecule expression, nef seems to alter class I traffic in the TGN in a PACS1-dependent pathway (Piguet et al. 2000). The exact mechanisms by which nef affects membrane traffic are still under investigation and have been the subject of much debate (Oldridge & Marsh 1998). Nef binds to a subunit of the proton pump, responsible for acidifying endosomes, which itself can bind AP2 (Lu et al. 1998). Nef has also been shown in vitro to interact directly with AP1 and, albeit weakly, with AP2 (Bresnahan et al. 1999, Greenberg et al. 1998). Another viral protein that interacts with AP1 is the E6 protein of bovine papilloma virus (Tong et al. 1998). The E6 protein is involved in transformation, and a homolog is present in human papilloma virus. Its AP1-binding activity could either be important for the virus life cycle or somehow participate in induction of cellular transformation.

A number of leukemias have associated defects in genes of proteins involved in CCV formation (Floyd & De Camilli 1998). CALM becomes fused to the AF10 gene in a translocation that occurs in both acute myeoblastic leukemia and acute lymphoblastic leukemia (Dreyling et al. 1996). Two translocations in acute myeoblastic leukemia result in fusion of Eps15 with HRX (Salcini et al. 1999). Endophilin is also fused to HRX in another case of acute leukemia (So et al. 2000). The donor gene in these two cases could give rise to a fusion protein causing transformation, as AF10 and HRX are both putative transcription factors. However, it is entirely conceivable that the endocytosis-related portion of the fusion protein could play a role in transformation because of the importance of endocytosis in growth factor receptor regulation. A number of oncogenes are constitutively activated receptor tyrosine kinases (RTKs) (Riese & Stern 1998), and expression of these receptors is regulated by CCV-mediated endocytosis. Thus CCV gene products are likely to play a tumor suppressor function. It is notable that tumor

suppression of meningioma has been associated with a gene encoding the β1-subunit of AP1 (Peyrard et al. 1994). The AP1-mediated pathway could potentially affect the presence of signaling molecules in endosomes from which RTKs can stimulate signaling pathways (Beattie et al. 2000).

As in model genetic systems, disruptions in the CCV pathways in humans are associated with neurological disorders. Histological distribution of CLCs is altered in brain tissue from patients with Alzheimer's disease and Pick's disease (Nakamura et al. 1994a,b). A significant reduction of expression of AP180 and AP2 in particular regions of brain from patients with Alzheimer's disease has also been reported (Yao et al. 1999, 2000). In patients with Huntington's disease, the huntington protein is altered by an expanded polyglutamine tract in the N terminus. Huntington has been reported to be enriched in CCVs, and Hip1R has been implicated in CCV-actin interactions (Engqvist-Goldstein et al. 1999, Velier et al. 1998). The dynamin/synaptojanin-binding protein intersectin is encoded in the Down's syndrome region of chromosome 21 (Pucharcos et al. 1999), and its overexpression could potentially be responsible for brain defects. It is interesting that genetic defects in the human β3A-subunit of the AP3 adaptor, which occurs in some cases of Hermansky-Pudlak syndrome, do not have a neurological phenotype (Dell'Angelica et al. 1999b). Similarly to AP3 defects in *M. musculus* and *D. melanogaster*, Hermansky-Pudlak syndrome patients present with albinism because of melanosome abnormalities and have platelet storage deficiency because of defects in formation of platelet dense granules.

Muscle defects are also associated with CCV malfunction. CCVs are present in the autophagic vacuoles of patients with distal myopathy with rimmed vacuoles (Kumamoto et al. 2000). Patients with DiGeorge and/or velocardiofacial syndromes are heterozygous for a deletion in chromosome 22, which includes the gene encoding CHC22. These patients also exhibit muscle weakness and hypocalcemic tetany, characterized by muscle seizures, which occur because of lower-than-normal levels of calcium (Liu et al. 2001b). With the mapping of the human genome completed, it is certain that many more disease genes will turn out to be components of the CCV formation pathway.

ACKNOWLEDGMENTS

Research from the Brodsky laboratory described here is supported by National Institutes of Health grants GM38093, GM61954, GM57657, and AI45865. Chih-Ying Chen is supported by the Arthritis Foundation of America and Mhairi C. Towler is supported by the Wellcome Trust. We thank the following colleagues for providing preprints and reprints of their work and for helpful discussions: Elizabeth Bennett, Caroline Enns, Pietro DeCamilli, Antony Jackson, James Keen, Harvey McMahon, Sandra Schmid, Corinne Smith, Alexander Sorkin, and Bing Zhang. We also thank Corinne Smith and John Heuser for permission to reproduce their published work. We apologize that we were unable to include all the primary references describing the research discussed here and hope that we at least managed

to provide links through review articles to all the outstanding work that has laid the foundation for the recent developments we have reviewed.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

Acton S, Brodsky FM. 1990. Predominance of clathrin light chain LCb correlates with the presence of a regulated secretory pathway. *J. Cell Biol.* 111:1419–26

Acton SL, Wong DH, Parham P, Brodsky FM, Jackson AP. 1993. Alteration of clathrin light chain expression by transfection and gene disruption. *Mol. Biol. Cell* 4:647–60

Ahle S, Ungewickell E. 1990. Auxilin, a newly identified clathrin-associated protein in coated vesicles from bovine brain. *J. Cell Biol.* 111:19–29

Ahn S, Maudsley S, Luttrell LM, Lefkowitz RJ, Daaka Y. 1999. Src-mediated tyrosine phosphorylation of dynamin is required for β2-adrenergic receptor internalization and mitogen-activated protein kinase signaling. *J. Biol. Chem.* 274:1185–88

Altschuler Y, Barbas SM, Terlecky LJ, Tang K, Hardy S, et al. 1998. Redundant and distinct functions for dynamin-1 and dynamin-2 isoforms. *J. Cell Biol.* 143:1871–81

Altschuler Y, Liu S-H, Katz L, Tang K, Hardy S, et al. 1999. ADP-ribosylation factor 6 and endocytosis at the apical surface of Madin-Darby canine kidney cells. *J. Cell Biol.* 147: 7–12

Amberg DC, Basart E, Botstein D. 1995. Defining protein interactions with yeast actin in vivo. *Nat. Struct. Biol.* 2:28–35

Anderson RGW. 1998. The caveolae membrane system. *Annu. Rev. Biochem.* 67:199–225

Anderson RGW, Brown MS, Goldstein JL. 1977. Role of the coated endocytic vesicle in the uptake of receptor-bound low density lipoprotein in human fibroblasts. *Cell* 10:351–64

Arneson LS, Kunz J, Anderson RA, Traub LM. 1999. Coupled inositide phosphorylation and phospholipase D activation initiates clathrin-coat assembly on lysosomes. *J. Biol. Chem.* 274:17794–805

Austin C, Hinners I, Tooze SA. 2000. Direct and GTP-dependent interaction of ADP-ribosylation factor 1 with clathrin adaptor protein AP-1 on immature secretory granules. *J. Biol. Chem.* 275:21862–69

Baggett JJ, Wendland B. 2001. Clathrin function in yeast endocytosis. *Traffic* 2:297–303

Barlowe C. 2000. Traffic COPs of the early secretory pathway. *Traffic* 1:371–77

Beattie EC, Howe CL, Wilde A, Brodsky FM, Mobley WC. 2000. NGF signals through TrkA to increase clathrin at the plasma membrane and enhance clathrin-mediated membrane trafficking. *J. Neurosci.* 20:7325–33

Beck KA, Keen JH. 1991a. Self-association of the plasma membrane-associated clathrin assembly protein AP-2. *J. Biol. Chem.* 266: 4437–41

Benmerah A, Lamaze C, Bègue B, Schmid SL, Dautry-Varsat A, Cerf-Bensussan N. 1998. AP-2/Eps15 interaction is required for receptor-mediated endocytosis. *J. Cell Biol.* 140:1055–62

Bennett EM, Lin SX, Towler MC, Maxfield FR, Brodsky FM. 2001. Clathrin hub expression affects early endosome distribution with minimal impact on receptor sorting and recycling. *Mol. Biol. Cell.* In press

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–42

Black MW, Pelham HR. 2000. A selective transport route from Golgi to late endosomes that requires the yeast GGA proteins. *J. Cell Biol.* 151:587–600

Blank GS, Brodsky FM. 1987. Clathrin assembly involves a light chain-binding region. *J. Cell Biol.* 105:2011–19

Bleazard W, McCaffery JM, King EJ, Bale S, Mozdy A, et al. 1999. The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. *Nat. Cell Biol.* 1:298–304

Boman AL, Zhang C, Zhu X, Kahn RA. 2000. A family of ADP-ribosylation factor effectors that can alter membrane transport through the *trans*-Golgi. *Mol. Biol. Cell* 11:1241–55

Boyer C, Auphan N, Luton F, Malburet JM, Barad M, et al. 1991. T cell receptor/CD3 complex internalization following activation of a cytolytic T cell clone: evidence for a protein kinase C-independent staurosporine-sensitive step. *Eur. J. Immunol.* 21:1623–34

Bresnahan PA, Yonemoto W, Greene WC. 1999. Cutting edge: SIV Nef protein utilizes both leucine- and tyrosine-based protein sorting pathways for down-regulation of CD4. *J. Immunol.* 163:2977–81

Brodsky FM. 1988. Living with clathrin: its role in intracellular membrane traffic. *Science* 242:1396–402

Brodsky FM. 1997. New fashions in vesicle coats. *Trends Cell Biol.* 7:175–79

Brodsky FM. 1999. Clathrin. In *Guidebook to Cytoskeletal and Motor Proteins*, ed. T Kreis, R Vale, pp. 512–16. Oxford, UK: Oxford Univ. Press

Brodsky FM, Hill BL, Acton SL, Näthke I, Wong DH, et al. 1991. Clathrin light chains: arrays of protein motifs that regulate coated vesicle dynamics. *Trends Biol. Sci.* 16:208–13

Brown CM, Roth MG, Henis YI, Petersen NO. 1999. An internalization-competent influenza hemagglutinin mutant causes the redistribution of AP-2 to existing coated pits and is colocalized with AP-2 in clathrin free clusters. *Biochemistry* 38:15166–73

Bukau B, Horwich AL. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* 92:351–66

Butler MH, David C, Ochoa GC, Freyberg Z, Daniell L, et al. 1997. Amphiphysin II (SH3P9; BIN1), a member of the amphiphysin/Rvs family, is concentrated in the cortical cytomatrix of axon initial segments and nodes of ranvier in brand and around T tubules in skeletal muscle. *J. Cell. Biol.* 137:1355–97

Chang MP, Mallet WG, Mostov KE, Brodsky FM. 1993. Adaptor self-aggregation, adaptor-receptor recognition and binding of α-adaptin subunits to the plasma membrane contribute to recruitment of adaptor (AP2) components of clathrin-coated pits. *EMBO J.* 12:2169–80

Chen H, Fre S, Slepnev VI, Capua MR, Takei K, et al. 1998. Epsin is an EH domain-binding protein implicated in clathrin-mediated endocytosis. *Nature* 394:793–97

Chen WS, Lazar CS, Lund KA, Welsh JB, Chang C-P, et al. 1989. Functional independence of the epidermal growth factor receptor from a domain required for ligand-induced internalization and calcium regulation. *Cell* 59:33–43

Chu DS, Pishvaee B, Payne GS. 1999. A modulatory role for clathrin light chain phosphorylation in Golgi membrane protein localization during vegetative growth and during the mating response of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 10:713–26

Coda L, Salcini AE, Confalonieri S, Pelicci G, Sorkina T, et al. 1998. Eps15R is a tyrosine kinase substrate with characteristics of a docking protein possibly involved in coated pits-mediated internalization. *J. Biol. Chem.* 273:3003–12

Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D. 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* 391:397–401

Cowles CR, Odorizzi G, Payne GS, Emr SD. 1997. The AP-3 adaptor complex is essential for cargo-selective transport to the yeast vacuole. *Cell* 91:109–18

Cremona O, De Camilli P. 2001. Phosphoinositides in membrane traffic at the synapse. *J. Cell Sci.* 114:1041–52

Cremona O, Di Paolo G, Wenk MR, Lüthi A, Kim WT, et al. 1999. Essential role of phosphoinositide metabolism in synaptic vesicle recycling. *Cell* 99:179–88

Cupers P, ter Haar E, Boll W, Kirchhausen

T. 1997. Parallel dimers and anti-parallel tetramers formed by epidermal growth factor receptor pathway substrate clone 15. *J. Biol. Chem.* 272:33430–34

Damer CK, O'Halloran TJ. 2000. Spatially regulated recruitment of clathrin to the plasma membrane during capping and cell translocation. *Mol. Biol. Cell* 11:2151–59

Damke H, Baba T, Warnock DE, Schmid SL. 1994. Induction of mutant dynamin specifically blocks endocytic coated vesicle formation. *J. Cell Biol.* 127:915–34

David C, McPherson PS, Mundigl O, De Camilli P. 1996. A role of amphiphysin in synaptic vesicle endocytosis suggested by its binding to dynamin in nerve terminals. *Proc. Natl. Acad. Sci. USA* 93:331–35

de Beer T, Carter RE, Lobel Rice KE, Sorkin A, Overduin M. 1998. Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science* 281:1357–60

de Beer T, Hoofnagle AN, Enmon JL, Bowers RC, Yamabhai M, et al. 2000. Molecular mechanism of NPF recognition by EH domains. *Nat. Struct. Biol.* 7:1018–22

de Heuvel E, Bell AW, Ramjaun AR, Wong K, Sossin WS, McPherson PS. 1997. Identification of the major synaptojanin-binding proteins in brain. *J. Biol. Chem.* 272:8710–16

Dell'Angelica EC, Klumperman J, Stoorvogel W, Bonifacino JS. 1998. Association of the AP-3 adaptor complex with clathrin. *Science* 280:431–34

Dell'Angelica EC, Mullins C, Bonifacino JS. 1999a. AP-4, a novel protein complex related to clathrin adaptors. *J. Biol. Chem.* 274: 7278–85

Dell'Angelica EC, Ohno H, Ooi CE, Rabinovich E, Roche KW, Bonifacino JS. 1997. AP-3: an adaptor-like protein complex with ubiquitous expression. *EMBO J.* 16:917–28

Dell'Angelica EC, Puertollano R, Mullins C, Aguilar RC, Vargas JD, et al. 2000. GGAs: a family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. *J. Cell Biol.* 149:81–94

Dell'Angelica EC, Shotelersuk V, Aguilar RC, Gahl WA, Bonifacino JS. 1999b. Altered trafficking of lysosomal proteins in Hermansky-Pudlak syndrome due to mutations in the β3A subunit of the AP-3 adaptor. *Mol. Cell* 3:11–21

DeLuca-Flaherty C, McKay DB, Parham P, Hill BL. 1990. Uncoating protein (hsc70) binds a conformationally labile domain of clathrin light chain LC_a to stimulate ATP hydrolysis. *Cell* 62:875–87

DeTulleo L, Kirchhausen T. 1998. The clathrin endocytic pathway in viral infection. *EMBO J.* 17:4585–93

Dittié AS, Hajibagheri N, Tooze SA. 1996. The AP-1 adaptor complex binds to immature secretory granules from PC12 cells, and is regulated by ADP-ribosylation factor. *J. Cell Biol.* 132:523–36

Drake MT, Downs MA, Traub LM. 2000. Epsin binds to clathrin by associating directly with the clathrin-terminal domain. Evidence for cooperative binding through two discrete sites. *J. Biol. Chem.* 275:6479–89

Dreyling MH, Martinez-Climent JA, Zheng M, Mao J, Rowley JD, Bohlander SK. 1996. The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc. Natl. Acad. Sci. USA* 93:4804–9

Dykstra ML, Cherukuri A, Pierce SK. 2001. Floating the raft hypothesis for immune receptors: access to rafts controls receptor signaling and trafficking. *Traffic* 2:160–66

Engqvist-Goldstein AEY, Kessels MM, Chopra VS, Hayden MR, Drubin DG. 1999. An actin-binding protein of the Sla2/Huntingtin interacting protein 1 family is a novel component of clathrin-coated pits and vesicles. *J. Cell Biol.* 147:1503–18

Engqvist-Goldstein AEY, Warren RA, Kessels MM, Heuser JE, Keen JE, Drubin DG. 2000. The actin-binding protein Hip1R binds clathrin and functions together with clathrin in the early stage of endocytosis. *Mol. Biol. Cell* 11(Suppl.):A147

Enmon JL, de Beer T, Overduin M. 2000. Solution structure of Eps15's third EH domain

reveals coincident Phe-Trp and Asn-Pro-Phe binding sites. *Biochemistry* 39:4309–19

Enns CA. 2001. Pumping iron: the strange partnership of the hemochromatosis protein, a class I MHC homolog, with the transferrin receptor. *Traffic* 2:167–74

Fabrizio JJ, Hime G, Lemmon SK, Bazinet C. 1998. Genetic dissection of sperm individualization in *Drosophila melanogaster*. *Development* 125:1833–43

Ferguson SS. 2001. Evolving concepts in G protein-coupled receptor endocytosis: the role in receptor desensitization and signaling. *Pharmacol. Rev.* 53:1–24

Fernandez-Chacon R, Achiriloaie M, Janz R, Albanesi JP, Südhof TC. 2000. SCAMP1 function in endocytosis. *J. Biol. Chem.* 275:12752–56

Fingerhut A, von Figura K, Honing S. 2001. Binding of AP2 to sorting signals is modulated by AP2 phosphorylation. *J. Biol. Chem.* 276:5476–82

Fish KN, Schmid SL, Damke H. 2000. Evidence that dynamin-2 functions as a signal-transducing GTPase. *J. Cell Biol.* 150:145–54

Flaherty KM, DeLuca-Flaherty C, McKay DB. 1990. Three-dimensional structure of the ATPase fragment of a 70 K heat-shock cognate protein. *Nature* 346:623–28

Flaherty KM, McKay DB, Kabsch W, Holmes KC. 1991. Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc. Natl. Acad. Sci. USA* 88:5041–45

Floyd S, De Camilli P. 1998. Endocytosis proteins and cancer: a potential link? *Trends Cell Biol.* 8:299–301

Fölsch H, Ohno H, Bonifacino JS, Mellman I. 1999. A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells. *Cell* 99:189–98

Ford MG, Pearse BM, Higgins MK, Vallis Y, Owen DJ, et al. 2001. Simultaneous binding of PtdIns(4,5)P2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* 291:1051–55

Frischknecht F, Cudmore S, Moreau V, Reckmann I, Rottger S, Way M. 1999. Tyrosine phosphorylation is required for actin-based motility of vaccinia but not Listeria or Shigella. *Curr. Biol.* 9:89–92

Fujimoto LM, Roth R, Heuser JE, Schmid SL. 2000. Actin assembly plays a variable, but not obligatory role in receptor-mediated endocytosis in mammalian cells. *Traffic* 1:161–71

Futter CE, Gibson A, Allchin EH, Maxwell S, Ruddock LJ, et al. 1998. In polarized MDCK cells basolateral vesicles arise from clathrin-γ-adaptin-coated domains on endosomal tubules. *J. Cell Biol.* 141:611–23

Gaidarov I, Chen Q, Falck JR, Reddy KK, Keen JH. 1996. A functional phosphatidylinositol 3,4,5-trisphosphate/phosphoinositide binding domain in the clathrin adaptor AP-2 α subunit. Implications for the endocytic pathway. *J. Biol. Chem.* 271:20922–29

Gaidarov I, Keen JH. 1999. Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits. *J. Cell Biol.* 146:755–64

Gaidarov I, Krupnick JG, Falck JR, Benovic JL, Keen JH. 1999a. Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *EMBO J.* 18:871–81

Gaidarov I, Santini F, Warren RA, Keen JH. 1999b. Spatial control of coated-pit dynamics in living cells. *Nat. Cell Biol.* 1:1–7

Gaidarov I, Smith ME, Domin J, Keen JH. 2001. The class II phosphoinositide 3-kinase C2α is activated by clathrin and regulates clathrin-mediated membrane trafficking. *Mol. Cell* 7:443–49

Gallusser A, Kirchhausen T. 1993. The β1 and β2 subunits of the AP complexes are the clathrin coat assembly components. *EMBO J.* 12:5237–44

Gammie AE, Kurihara LJ, Vallee RB, Rose MD. 1995. DNM1, a dynamin-related gene, participates in endosomal trafficking in yeast. *J. Cell Biol.* 130:553–66

Garred O, Rodal SK, van Deurs B, Sandvig K. 2001. Reconstitution of clathrin-independent

endocytosis at the apical domain of permeabilized MDCK II cells: requirement for a Rho-family GTPase. *Traffic* 2:26–36

Geuze HJ. 1998. The role of endosomes and lysosomes in MHC class II functioning. *Immunol. Today* 19:282–87

Gold ES, Morrissette NS, Underhill DM, Guo J, Bassetti M, Aderem A. 2000. Amphiphysin IIm, a novel amphiphysin II isoform, is required for macrophage phagocytosis. *Immunity* 12:285–92

Goodman OB Jr, Krupnick JG, Gurevich VV, Benovic JL, Keen JH. 1997. Arrestin/clathrin interaction. Localization of the arrestin binding locus to the clathrin terminal domain. *J. Biol. Chem.* 272:15017–22

Greenberg M, DeTulleo L, Rapoport I, Skowronski J, Kirchhausen T. 1998. A dileucine motif in HIV-1 Nef is essential for sorting into clathrin-coated pits and for downregulation of CD4. *Curr. Biol.* 8:1239–42

Greene B, Liu S-H, Wilde A, Brodsky FM. 2000. Complete reconstitution of clathrin basket formation with recombinant protein fragments: adaptor control of clathrin self-assembly. *Traffic* 1:69–75

Greener T, Zhao X, Nojima H, Eisenberg E, Greene LE. 2000. Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *J. Biol. Chem.* 275:1365–70

Haffner C, Paolo GD, Rosenthal JA, de Camilli P. 2000. Direct interaction of the 170 kDa isoform of synaptojanin 1 with clathrin and with the clathrin adaptor AP-2. *Curr. Biol.* 10:471–74

Hannan LA, Newmyer SL, Schmid SL. 1998. ATP- and cytosol-dependent release of adaptor proteins from clathrin-coated vesicles: a dual role for Hsc70. *Mol. Biol. Cell* 9:2217–29

Hao W, Luo Z, Zheng L, Prasad K, Lafer EM. 1999. AP180 and AP-2 interact directly in a complex that cooperatively assembles clathrin. *J. Biol. Chem.* 274:22785–94

Harris TW, Hartwieg E, Horvitz HR, Jorgensen EM. 2000. Mutations in synaptojanin disrupt synaptic vesicle recycling. *J. Cell Biol.* 150:589–600

Haucke V, De Camilli P. 1999. AP-2 recruitment to synaptotagmin stimulated by tyrosine-based endocytic motifs. *Science* 285:1268–71

Heuser J. 1989. Effects of cytoplasmic acidification on clathrin lattice morphology. *J. Cell Biol.* 108:401–11

Heuser JE, Anderson RGW. 1989. Hypertonic media inhibit receptor-mediated endocytosis by blocking clathrin-coated pit formation. *J. Cell Biol.* 108:389–400

Heuser JE, Keen J. 1988. Deep-etch visualization of proteins involved in clathrin assembly. *J. Cell Biol.* 107:877–86

Heuser JE, Keen JH, Amende LM, Lippoldt RE, Prasad K. 1987. Deep-etch visualization of 27S clathrin: a tetrahedral tetramer. *J. Cell Biol.* 105:1999–2009

Heuser JE, Kirchhausen T. 1985. Deep-etch view of clathrin assembly. *J. Ultrastruct. Res.* 92:1–27

Hill E, van Der Kaay J, Downes CP, Smythe E. 2001. The role of dynamin and its binding partners in coated pit invagination and scission. *J. Cell Biol.* 152:309–24

Hinshaw JE. 2000. Dynamin and its role in membrane fusion. *Annu. Rev. Cell. Dev. Biol.* 16:483–520

Hirst J, Bright NA, Rous B, Robinson MS. 1999. Characterization of a fourth adaptor-related protein complex. *Mol. Biol. Cell* 10:2787–802

Hirst J, Lui WW, Bright NA, Totty N, Seaman MN, Robinson MS. 2000. A family of proteins with γ-adaptin and VHS domains that facilitate trafficking between the *trans*-Golgi network and the vacuole/lysosome. *J. Cell Biol.* 149:67–80

Hofmann MW, Honing S, Rodionov D, Dobberstein B, von Figura K, Bakke O. 1999. The leucine-based sorting motifs in the cytoplasmic domain of the invariant chain are recognized by the clathrin adaptors AP1 and AP2 and their medium chains. *J. Biol. Chem.* 274:36153–58

Holmes SE, Riazi MA, Gong W, McDermid

HE, Sellinger BT, et al. 1997. Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. *Hum. Mol. Genet.* 6:357–67

Honing S, Kreimer G, Robenek H, Jockusch BM. 1994. Receptor-mediated endocytosis is sensitive to antibodies against the uncoating ATPase (hsc70). *J. Cell Sci.* 107:1185–96

Hu Y, Barzilai A, Chen M, Bailey CH, Kandel ER. 1993. 5-HT and cAMP induce the formation of coated pits and vesicles and increase the expression of clathrin light chain in sensory neurons of aplysia. *Neuron* 10:921–29

Huang KM, D'Hondt K, Riezman H, Lemmon SK. 1999. Clathrin functions in the absence of heterotetrameric adaptors and AP180-related proteins in yeast. *EMBO J.* 18:3897–908

Hughes WE, Cooke FT, Parker PJ. 2000. Sac phosphatase domain proteins. *Biochem. J.* 350(2):337–52

Huttner WB, Schmidt A. 2000. Lipids, lipid modification and lipid-protein interaction in membrane budding and fission—insights from the roles of endophilin A1 and synaptophysin in synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* 10:543–51

Hyman J, Chen H, Di Fiore PP, De Camilli P, Brunger AT. 2000. Epsin 1 undergoes nucleocytosolic shuttling and its eps15 interactor NH(2)-terminal homology (ENTH) domain, structurally similar to Armadillo and HEAT repeats, interacts with the transcription factor promyelocytic leukemia Zn(2)+ finger protein (PLZF). *J. Cell Biol.* 149:537–46

Itoh T, Koshiba S, Kigawa T, Kikuchi A, Yokoyama S, Takenawa T. 2001. Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* 291:1047–51

Kanaoka Y, Kimura SH, Okazaki I, Ikeda M, Nojima H. 1997. GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. *FEBS Lett.* 402:73–80

Kaneseki T, Kadota K. 1969. The "vesicle in a basket." A morphological study of the coated vesicle isolated from the nerve endings of the guinea pig brain, with special reference to the mechanism of membrane movements. *J. Cell Biol.* 42:202–20

Keen JH, Willingham MC, Pastan IH. 1979. Clathrin-coated vesicles: isolation, dissociation, and factor-dependent reassociation of clathrin baskets. *Cell* 16:303–12

Kelley WL. 1998. The J-domain family and the recruitment of chaperone power. *Trends Biochem. Sci.* 23:222–27

Kibbey RG, Rizo J, Gierasch LM, Anderson RG. 1998. The LDL receptor clustering motif interacts with the clathrin terminal domain in a reverse turn conformation. *J. Cell Biol.* 142:59–67

Kirchhausen T. 1999. Adaptors for clathrin-mediated traffic. *Annu. Rev. Cell Dev. Biol.* 15:705–32

Kirchhausen T. 2000a. Clathrin. *Annu. Rev. Biochem.* 69:699–727

Kirchhausen T. 2000b. Three ways to make a vesicle. *Nat. Rev. Mol. Cell Biol.* 1:187–98

Kirchhausen T, Harrison SC, Chow EP, Mattaliano RJ, Ramachandran KL, et al. 1987. Clathrin heavy chain: molecular cloning and complete primary sequence. *Proc. Natl. Acad. Sci. USA* 84:8805–9

Kirchhausen T, Toyoda T. 1993. Immunoelectron microscopic evidence for the extended conformation of light chains in clathrin trimers. *J. Biol. Chem.* 268:10268–73

Kosaka T, Ikeda K. 1983. Reversible blockage of membrane retrieval and endocytosis in the garland cell of the temperature-sensitive mutant of *Drosophila melanogaster, shibire*. *J. Cell Biol.* 97:499–507

Kraulis PJ. 1991. Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24:946–50

Kreitzer G, Marmorstein A, Okamoto P, Vallee R, Rodriguez-Boulan E. 2000. Kinesin and dynamin are required for post-Golgi transport of a plasma-membrane protein. *Nat. Cell Biol.* 2:125–27

Krupnick JG, Goodman OB Jr, Keen JH, Benovic JL. 1997. Arrestin-clathrin interaction.

Localization of the clathrin binding domain of nonvisual arrestins to the carboxyl terminus. *J. Biol. Chem.* 272:15011–16

Kuliawat R, Klumperman J, Ludwig T, Arvan P. 1997. Differential sorting of lysosomal enzymes out of the regulated secretory pathway in pancreatic β-cells. *J. Cell Biol.* 137:595–608

Kumamoto T, Ito T, Horinouchi H, Ueyama H, Toyoshima I, Tsuda T. 2000. Increased lysosome-related proteins in the skeletal muscles of distal myopathy with rimmed vacuoles. *Muscle Nerve* 23:1686–93

Kusner LL, Carlin C. 2000. Expression of clathrin assembly protein AP180 in non-neuronal cells. *Mol. Biol. Cell* 11(Suppl.): A323

Lamaze C, Dujeancourt A, Baba T, Lo CG, Benmerah A, Dautry-Varsat A. 2001. Interleukin 2 receptors and detergent-resistant membrane domains define a clathrin-independent endocytic pathway. *Mol. Cell* 7:661–71

Langlet C, Bernard AM, Drevot P, He HT. 2000. Membrane rafts and signaling by the multichain immune recognition receptors. *Curr. Opin. Immunol.* 12:250–55

Laporte SA, Oakley RH, Holt JA, Barak LS, Caron MG. 2000. The interaction of β-arrestin with the AP-2 adaptor is required for the clustering of β2-adrenergic receptor into clathrin-coated pits. *J. Biol. Chem.* 275:23120–26

Lauritsen JP, Menne C, Kastrup J, Dietrich J, Odum N, Geisler C. 2000 β2-Adaptin is constitutively de-phosphorylated by serine/threonine protein phosphatase PP2A and phosphorylated by a staurosporine-sensitive kinase. *Biochim. Biophys. Acta* 1497:297–307

Le Borgne R, Schmidt A, Mauxion F, Griffiths G, Hoflack B. 1993. Binding of AP-1 Golgi adaptors to membranes requires phosphorylated cytoplasmic domains of the mannose 6-phosphate/insulin-like growth factor II receptor. *J. Biol. Chem.* 268:22552–56

Lee C, Kim SR, Chung JK, Frohman MA, Kilimann MW, Rhee SG. 2000. Inhibition of phospholipase D by amphiphysins. *J. Biol. Chem.* 275:18751–58

Lemmon SK. 2001. Clathrin uncoating: auxilin comes to life. *Curr. Biol.* 11:R49–52

Lichte B, Veh RW, Meyer HE, Kilimann MW. 1992. Amphiphysin, a novel protein associated with synaptic vesicles. *EMBO J.* 11:2521–30

Liu H, Rhodes M, Wiest DL, Vignali DA. 2000. On the dynamics of TCR:CD3 complex cell surface expression and downmodulation. *Immunity* 13:665–75

Liu JP, Yajima Y, Li H, Ackland S, Akita Y, et al. 1997. Molecular interactions between dynamin and G-protein $\beta\gamma$-subunits in neuroendocrine cells. *Mol. Cell. Endocrinol.* 132:61–71

Liu S-H, Mallet WG, Brodsky FM. 2001a. Clathrin-mediated endocytosis. In *Endocytosis*, ed. M Marsh, pp. 1–25. Oxford, UK: Oxford Univ. Press

Liu S-H, Marks MS, Brodsky FM. 1998. A dominant negative clathrin mutant differentially affects trafficking of molecules with distinct sorting motifs in the class II MHC pathway. *J. Cell Biol.* 140:1023–37

Liu S-H, Towler MC, Chen E, Chen C-Y, Song W, et al. 2001b. A novel clathrin homolog that co-distributes with cytoskeletal components functions in the *trans*-Golgi network. *EMBO J.* 20:272–84

Liu S-H, Wong ML, Craik CS, Brodsky FM. 1995. Regulation of clathrin assembly and trimerization defined using recombinant triskelion hubs. *Cell* 83:257–67

Lu X, Yu H, Liu SH, Brodsky FM, Peterlin BM. 1998. Interactions between HIV1 Nef and vacuolar ATPase facilitate the internalization of CD4. *Immunity* 8:647–56

Lund J, Chen F, Hua A, Roe B, Budarf M, et al. 2000. Comparative sequence analysis of 634 kb of the mouse chromosome 16 region of conserved synteny with the human velocardiofacial syndrome region on chromosome 22q11.2. *Genomics* 63:374–83

Maehama T, Dixon JE. 1998. The tumor suppressor, PTEN/MMAC1, dephosphorylates

the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* 273:13375–78

Mahaffey DT, Peeler JS, Brodsky FM, Anderson RGW. 1990. Clathrin-coated pits contain an integral membrane protein that binds the AP-2 subunit with high affinity. *J. Biol. Chem.* 265:16514–20

Mallet WG, Brodsky FM. 1996. A membrane-associated protein complex with selective binding to the clathrin coat adaptor AP1. *J. Cell Sci.* 109:3059–68

Mao Y, Chen J, Maynard JA, Zhang B, Quiocho FA. 2001. A novel all helix fold of the AP180 amino-terminal domain for phosphoinositide binding and clathrin assembly in synaptic vesicle endocytosis. *Cell* 104:433–40

Marks B, Stowell MH, Vallis Y, Mills IG, Gibson A, et al. 2001. GTPase activity of dynamin and resulting conformation change are essential for endocytosis. *Nature* 410:231–35

Marks MS, Ohno H, Kirchhausen T, Bonifacino JS. 1997. Protein sorting by tyrosine based signals: adapting to the Ys and wherefores. *Trends Cell Biol.* 7:124–28

Marqueze B, Berton F, Seagar M. 2000. Synaptotagmins in membrane traffic: which vesicles do the tagmins tag? *Biochimie* 82:409–20

Marsh M, McMahon HT. 1999. The structural era of endocytosis. *Science* 285:215–20

Martina JA, Bonangelino CJ, Aguilar RC, Bonifacino JS. 2001. Stonins 2: an adaptor-like protein that interacts with components of the endocytic machinery. *J. Cell Biol.* 153:1111–20

Mayor S, Presley JF, Maxfield FR. 1993. Sorting of membrane components from endosomes and subsequent recycling to the cell surface occurs by a bulk flow process. *J. Cell Biol.* 121:1257–69

McGraw TE, Maxfield FR. 1990. Human transferrin receptor internalization is partially dependent upon an aromatic amino acid in the cytoplasmic domain. *Cell Reg.* 1:369–77

McLauchlan H, Newell J, Morrice N, Osborne A, West M, Smythe E. 1997. A novel role for Rab5-GDI in ligand sequestration into clathrin-coated pits. *Curr. Biol.* 8:34–45

McMahon HT. 1999. Endocytosis: an assembly protein for clathrin cages. *Curr. Biol.* 9:R332–35

McNiven MA, Cao H, Pitts KR, Yoon Y. 2000. The dynamin family of mechanoenzymes: pinching in new places. *Trends Biochem. Sci.* 25:115–20

Merilainen J, Lehto VP, Wasenius VM. 1997. FAP52, a novel, SH3 domain-containing focal adhesion protein. *J. Biol. Chem.* 272:23278–84

Merritt EA, Murphy MEP. 1994. Raster3d Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D* 50:869–73

Meyer C, Zizioli D, Lausmann S, Eskelinen EL, Hamann J, et al. 2000. μ1A-adaptin-deficient mice: lethality, loss of AP-1 binding and rerouting of mannose 6-phosphate receptors. *EMBO J.* 19:2193–203

Michaely P, Kamal A, Anderson RG, Bennett V. 1999 A requirement for ankyrin binding to clathrin during coated pit budding. *J. Biol. Chem.* 274:35908–13

Micheva KD, Kay BK, McPherson PS. 1997. Synaptojanin forms two separate complexes in the nerve terminal. Interactions with endophilin and amphiphysin. *J. Biol. Chem.* 272:27239–45

Miller K, Shipman M, Trowbridge IS, Hopkins CR. 1991. Transferrin receptors promote the formation of clathrin lattices. *Cell* 65:621–32

Mineo C, Gill GN, Anderson RG. 1999. Regulated migration of epidermal growth factor receptor from caveolae. *J. Biol. Chem.* 274:30636–43

Modregger J, Ritter B, Witter B, Paulsson M, Plomann M. 2000. All three PACSIN isoforms bind to endocytic proteins and inhibit endocytosis. *J. Cell Sci.* 113:4511–21

Molinete M, Dupuis S, Brodsky FM, Halban P. 2001. Role of clathrin in the regulated secretory pathway of pancreatic β-cells. *J. Cell Sci.* In press

Morgan JR, Prasad K, Hao W, Augustine GJ, Lafer EM. 2000. A conserved clathrin assembly motif essential for synaptic vesicle endocytosis. *J. Neurosci.* 20:8667–76

Morgan JR, Zhao X, Womack M, Prasad K, Augustine GJ, Lafer EM. 1999 A role for the clathrin assembly domain of AP180 in synaptic vesicle endocytosis. *J. Neurosci.* 19: 10201–12

Morshauser RC, Hu W, Wang H, Pang Y, Flynn GC, Zuiderweg ER. 1999. High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. *J. Mol. Biol.* 289: 1387–403

Mostov KE, Verges M, Altschuler Y. 2000. Membrane traffic in polarized epithelial cells. *Curr. Opin. Cell Biol.* 12:483–90

Muhlberg AB, Warnock DE, Schmid SL. 1997. Domain structure and intramolecular regulation of dynamin GTPase. *EMBO J.* 16:6676–83

Munn AL, Stevenson BJ, Geli MI, Riezman H. 1995. *end5*, *end6*, and *end7*: mutations that cause actin delocalization and block the internalization step of endocytosis in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 6:1721–42

Musacchio A, Smith CJ, Roseman AM, Harrison SC, Kirchhausen T, Pearse BMF. 1999. Functional organization of clathrin in coats: combining electron cryomicroscopy and X-ray crystallography. *Mol. Cell* 3:761–70

Nakagawa T, Setou M, Seog D, Ogasawara K, Dohmae N, et al. 2000. A novel motor, KIF13A, transports mannose-6-phosphate receptor to plasma membrane through direct interaction with AP-1 complex. *Cell* 103: 569–81

Nakamura Y, Takeda M, Yoshimi K, Hattori H, Hariguchi S, et al. 1994a. Involvement of clathrin light chains in the pathology of Alzheimer's disease. *Acta Neuropathol.* 87:23–31

Nakamura Y, Takeda M, Yoshimi K, Hattori H, Hariguchi S, et al. 1994b. Involvement of clathrin light chains in the pathology of Pick's disease: implication for impairment of axonal transport. *Neurosci. Lett.* 180:25–28

Nakaseko C, Miyatake S, Iida T, Hara S, Abe R, et al. 1999. Cytotoxic T lymphocyte antigen 4 (CTLA-4) engagement delivers an inhibitory signal through the membrane-proximal region in the absence of the tyrosine motif in the cytoplasmic tail. *J. Exp. Med.* 190:765–74

Nakata T, Takemura R, Hirokawa N. 1993. A novel member of the dynamin family of GTP-binding proteins is expressed specifically in the testis. *J. Cell Sci.* 105:1–5

Nakatsu F, Sakuma M, Matsuo Y, Arase H, Yamasaki S, et al. 2000. A di-leucine signal in the ubiquitin moiety: possible involvement in ubiquitination-mediated endocytosis. *Mol. Biol. Cell* 11(Suppl.):A147

Näthke IS, Heuser J, Lupas A, Stock J, Turck CW, Brodsky FM. 1992. Folding and trimerization of clathrin subunits at the triskelion hub. *Cell* 68:899–910

Nemoto Y, De Camilli P. 1999. Recruitment of an alternatively spliced form of synaptojanin 2 to mitochondria by the interaction with the PDZ domain of a mitochondrial outer membrane protein. *EMBO J.* 18:2991–3006

Nesterov A, Carter RE, Sorkina T, Gill GN, Sorkin A. 1999. Inhibition of the receptor-binding function of clathrin adaptor protein AP-2 by dominant-negative mutant μ2 subunit and its effects on endocytosis. *EMBO J.* 18:2489–99

Newman LS, McKeever MO, Okano HJ, Darnell RB. 1995. β-NAP, a cerebellar degeneration antigen, is a neuron-specific vesicle coat protein. *Cell* 82:773–83

Newmyer SL, Schmid SL. 2001. Dominant-interfering Hsc70 mutants disrupt multiple stages of the clathrin-coated vesicle cycle in vivo. *J. Cell Biol.* 152:607–20

Nossal R. 2001. Energetics of clathrin basket assembly. *Traffic* 2:138–47

Odorizzi G, Cowles CR, Emr SD. 1998. The AP-3 complex: a coat of many colours. *Trends Cell Biol.* 8:282–88

O'Halloran TJ. 2000. Membrane traffic and cytokinesis. *Traffic* 1:921–26

Ohno H, Stewart J, Fournier M-C, Bosshart H, Rhee I, et al. 1995. Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science* 269:1872–75

Ohno H, Tomemori T, Nakatsu F, Okazaki Y, Aguilar RC, et al. 1999. μ1B, a novel adaptor medium chain expressed in polarized epithelial cells. *FEBS Lett.* 449:215–20

Okamoto CT, McKinney J, Jeng YY. 2000. Clathrin in mitotic spindles. *Am. J. Physiol. Cell Physiol.* 279:C369–C74

Oldridge J, Marsh M. 1998. Nef—an adaptor adaptor? *Trends Cell Biol.* 8:302–5

Orsel JG, Sincock PM, Krise JP, Pfeffer SR. 2000. Recognition of the 300-kDa mannose 6-phosphate receptor cytoplasmic domain by 47-kDa tail-interacting protein. *Proc. Natl. Acad. Sci. USA* 97:9047–51

Owen DJ, Evans PR. 1998. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* 282:1327–32

Owen DJ, Setiadi H, Evans PR, McEver RP, Green SA. 2001. A third specificity-determining site in μ2 adaptin for sequences upstream of YxxΦ sorting motifs. *Traffic* 2:105–10

Owen DJ, Vallis Y, Noble ME, Hunter JB, Dafforn TR, et al. 1999. A structural explanation for the binding of multiple ligands by the α-adaptin appendage domain. *Cell* 97:805–15

Owen DJ, Vallis Y, Pearse BM, McMahon HT, Evans PR. 2000. The structure and function of the β2-adaptin appendage domain. *EMBO J.* 19:4216–27

Owen DJ, Wigge P, Vallis Y, Moore JD, Evans PR, McMahon HT. 1998. Crystal structure of the amphiphysin-2 SH3 domain and its role in the prevention of dynamin ring formation. *EMBO J.* 17:5273–85

Page LJ, Sowerby PJ, Lui WW, Robinson MS. 1999. γ-synergin: an EH domain-containing protein that interacts with γ-adaptin. *J. Cell Biol.* 146:993–1004

Parton RG, Way M, Zorzi N, Stang E. 1997. Caveolin-3 associates with developing T-tubules during muscle differentiation. *J. Cell Biol.* 136:137–54

Payne GS, Schekman R. 1985. A test of clathrin function in protein secretion and cell growth. *Science* 230:1009–14

Payne GS, Schekman R. 1989. Clathrin: a role in the intracellular retention of a Golgi membrane protein. *Science* 245:1358–65

Pearse BMF. 1975. Coated vesicles from pig brain: purification and biochemical characterization. *J. Mol. Biol.* 97:93–98

Pearse BMF, Robinson MS. 1990. Clathrin, adaptors and sorting. *Annu. Rev. Cell Biol.* 6:151–71

Peyrard M, Fransson I, Xie Y-G, Han F-Y, Ruttledge MH, et al. 1994. Characterization of a new member of the human β-adaptin gene family from chromosome 22q12, a candidate meningioma gene. *Hum. Mol. Genet.* 3:1393–99

Piguet V, Chen YL, Mangasarian A, Foti M, Carpentier JL, Trono D. 1998. Mechanism of Nef-induced CD4 endocytosis: Nef connects CD4 with the μ chain of adaptor complexes. *EMBO J.* 17:2472–81

Piguet V, Wan L, Borel C, Mangasarian A, Demaurex N, et al. 2000. HIV-1 Nef protein binds to the cellular protein PACS-1 to downregulate class I major histocompatibility complexes. *Nat. Cell Biol.* 2:163–67

Pishvaee B, Munn A, Payne GS. 1997. A novel structural model for regulation of clathrin function. *EMBO J.* 16:2227–39

Pleasure IT, Black MM, Keen JH. 1993. Valosin-containing protein, VCP, is a ubiquitous clathrin-binding protein. *Nature* 365: 459–62

Pley U, Parham P. 1993. Clathrin: its role in receptor-mediated vesicular transport and specialized functions in neurons. *Crit. Rev. Biochem. Mol. Biol.* 28:431–64

Pley UM, Hill BL, Alibert C, Brodsky FM, Parham P. 1995. The interaction of calmodulin with clathrin-coated vesicles, triskelions, and light chains. *J. Biol. Chem.* 270: 2395–402

Pucharcos C, Estivill X, de la Luna S. 2000. Intersectin 2, a new multimodular protein involved in clathrin-mediated endocytosis. *FEBS Lett.* 478:43–51

Pucharcos C, Fuentes JJ, Casas C, de la Luna S, Alcantara S, et al. 1999. Alu-splice cloning of human intersectin (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down's syndrome. *Eur. J. Hum. Genet.* 7:704–12

Puertollano R, Randazzo PA, Presley JF, Hartnell LM, Bonifacino JS. 2001. The GGAs promote ARF-dependent recruitment of clathrin to the TGN. *Cell* 105:93–102

Qualmann B, Kelly RB. 2000. Syndapin isoforms participate in receptor-mediated endocytosis and actin organization. *J. Cell Biol.* 148:1047–62

Qualmann B, Kessels MM, Kelly RB. 2000. Molecular links between endocytosis and the actin cytoskeleton. *J. Cell Biol.* 150:F111–16

Qualmann B, Roos J, DiGregorio PJ, Kelly RB. 1999. Syndapin I, a synaptic dynamin-binding protein that associates with the neural Wiskott-Aldrich syndrome protein. *Mol. Biol. Cell* 10:501–13

Ramjaun AR, Micheva KD, Bouchelet I, McPherson PS. 1997. Identification and characterization of a nerve terminal-enriched amphiphysin isoform. *J. Biol. Chem.* 272:16700–6

Rapoport I, Chen YC, Cupers P, Shoelson SE, Kirchhausen T. 1998. Dileucine-based sorting signals bind to the β chain of AP-1 at a site distinct and regulated differently from the tyrosine-based motif-binding site. *EMBO J.* 17:2148–55

Rebecchi MJ, Scarlata S. 1998. Pleckstrin homology domains: a common fold with diverse functions. *Annu. Rev. Biophys. Biomol. Struct.* 27:503–28

Riese DJ II, Stern DF. 1998. Specificity within the EGF family/ErbB receptor family signaling network. *BioEssays* 20:41–48

Ringstad N, Nemoto Y, De Camilli P. 1997. The SH3p4/Sh3p8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain. *Proc. Natl. Acad. Sci. USA* 94:8569–74

Ritter B, Modregger J, Paulsson M, Plomann M. 1999. PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter proteins. *FEBS Lett.* 454:356–62

Robinson MS, Kreis TE. 1992. Recruitment of coat proteins onto Golgi membranes in intact and permeabilized cells: effects of Brefeldin A and G-protein activators. *Cell* 69:129–38

Roos J, Kelly RB. 1998. Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing protein that interacts with *Drosophila* dynamin. *J. Biol. Chem.* 273:19108–19

Rosa JL, Barbacid M. 1997. A giant protein that stimulates guanine nucleotide exchange on ARF1 and Rab proteins forms a cytosolic ternary complex with clathrin and Hsp70. *Oncogene* 15:1–6

Rosenthal JA, Chen H, Slepnev VI, Pellegrini L, Salcini AE, et al. 1999. The epsins define a family of proteins that interact with components of the clathrin coat and contain a new protein module. *J. Biol. Chem.* 274:33959–65

Roth MG. 1999. Lipid regulators of membrane traffic through the Golgi complex. *Trends Cell Biol.* 9:174–79

Roth TF, Porter KR. 1964. Yolk protein uptake in the oocyte of the mosquito *Aedes aegypti L. J. Cell Biol.* 20:313–32

Rubtsov AM, Lopina OD. 2000. Ankyrins. *FEBS Lett.* 482:1–5

Sakamuro D, Elliott KJ, Wechsler-Reya R, Prendergast GC. 1996. BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat. Genet.* 14:69–77

Salcini AE, Chen H, Iannolo G, De Camilli P, Di Fiore PP. 1999. Epidermal growth factor pathway substrate 15, Eps15. *Int. J. Biochem. Cell Biol.* 31:805–9

Salim K, Bottomley MJ, Querfurth E, Zvelebil MJ, Gout I, et al. 1996. Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase. *EMBO J.* 15:6241–50

Salisbury JL, Condeelis JS, Satir P. 1980. Role of coated vesicles, microfilaments, and

calmodulin in receptor-mediated endocytosis by cultured B lymphoblastoid cells. *J. Cell Biol.* 87:132–41

Santini F, Marks MS, Keen JH. 1998. Endocytic clathrin-coated pit formation is independent of receptor internalization signal levels. *Mol. Biol. Cell* 9:1177–94

Santolini E, Puri C, Salcini AE, Gagliani MC, Pelicci PG, et al. 2000. Numb is an endocytic protein. *J. Cell Biol.* 151:1345–52

Schlossman DM, Schmid SL, Braell WA, Rothman JE. 1984. An enzyme that removes clathrin coats: purification of an uncoating ATPase. *J. Cell Biol.* 99:723–33

Schmidt A, Wolde M, Thiele C, Fest W, Kratzin H, et al. 1999. Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* 401:133–41

Seaman MNJ, Sowerby PJ, Robinson MS. 1996. Cytosolic and membrane-associated proteins involved in the recruitment of AP-1 adaptors onto the *trans*-Golgi network. *J. Biol. Chem.* 271:25446–51

Sengar AS, Wang W, Bishay J, Cohen S, Egan SE. 1999. The EH and SH3 domain Ese proteins regulate endocytosis by linking to dynamin and Eps15. *EMBO J.* 18:1159–71

Sever S, Damke H, Schmid SL. 2000. Garrotes, springs, ratchets, and whips: putting dynamin models to the test. *Traffic* 1:385–92

Shih W, Gallusser A, Kirchhausen T. 1995. A clathrin-binding site in the hinge of the $\beta 2$ chain of mammalian AP-2 complexes. *J. Biol. Chem.* 270:31083–90

Simpson F, Bright NA, West MA, Newman LS, Darnell RB, Robinson MS. 1996. A novel adaptor-related protein complex. *J. Cell Biol.* 133:749–60

Simpson F, Hussain NK, Qualmann B, Kelly RB, Kay BK, et al. 1999. SH3-domain-containing proteins function at distinct steps in clathrin-coated vesicle formation. *Nat. Cell Biol.* 1:119–24

Simpson F, Peden AA, Christopoulou L, Robinson MS. 1997. Characterization of the adaptor-related protein complex, AP3. *J. Cell Biol.* 137:835–45

Simpson F, Whitehead JP, James DE. 2001. GLUT4—at the cross roads between membrane trafficking and signal transduction. *Traffic* 2:2–11

Skretting G, Torgersen ML, van Deurs B, Sandvig K. 1999. Endocytic mechanisms responsible for uptake of GPI-linked diphtheria toxin receptor. *J. Cell Sci.* 112:3899–909

Slepnev VI, De Camilli P. 2000. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nat. Rev. Neurosci.* 1:161–72

Slepnev VI, Ochoa GC, Butler MH, De Camilli P. 2000. Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. *J. Biol. Chem.* 275:17583–89

Smith CJ, Grigorieff N, Pearse BMF. 1998. Clathrin coats at 21 Å resolution: a cellular assembly designed to recycle multiple membrane receptors. *EMBO J.* 17:4943–53

So CW, So CK, Cheung N, Chew SL, Sham MH, Chan LC. 2000. The interaction between EEN and Abi-1, two MLL fusion partners, and synaptojanin and dynamin: implications for leukaemogenesis. *Leukemia* 14:594–601

Solomonia RO, McCabe BJ, Jackson AP, Horn G. 1997. Clathrin proteins and recognition memory. *Neuroscience* 80:59–67

Sorkin A, Carpenter G. 1993. Interaction of activated EGF receptors with coated pit adaptins. *Science* 261:612–15

Sparks AB, Hoffman NG, McConnell SJ, Fowlkes DM, Kay BK. 1996. Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins. *Nat. Biotechnol.* 14:741–44

Spradling KD, Burke TJ, Lohi J, Pilcher BK. 2000. Cloning and initial characterization of human epsin 3, a novel matrix-induced keratinocyte specific transcript. *J. Invest. Dermatol.* 115:332

Springer S, Spang A, Schekman R. 1999. A primer on vesicle budding. *Cell* 97:145–48

Stoorvogel W, Oorschot V, Geuze HJ. 1996. A

novel class of clathrin-coated vesicles budding from endosomes. *J. Cell Biol.* 132:21–33

Subtil A, Daidarov I, Kobylarz K, Lampson MA, Keen JH, McGraw TE. 1999. Acute cholesterol depletion inhibits clathrin-coated pit budding. *Proc. Natl. Acad. Sci. USA* 8:6775–80

Sugita M, Peters PJ, Brenner MB. 2000. Pathways for lipid antigen presentation by CD1 molecules: nowhere for intracellular pathogens to hide. *Traffic* 1:295–300

Sutton RB, Ernst JA, Brunger AT. 1999. Crystal structure of the cytosolic C2A-C2B domains of synaptotagmin III. Implications for Ca(+2)-independent SNARE complex interaction. *J. Cell Biol.* 147:589–98

Takatsu H, Yoshino K, Nakayama K. 2000. Adaptor γ ear homology domain conserved in γ-adaptin and GGA proteins that interact with γ-synergin. *Biochem. Biophys. Res. Commun.* 271:719–25

Tan PK. 1993. Clathrin facilitates the internalization of seven transmembrane segment receptors for mating pheromones in yeast. *J. Cell Biol.* 123:1707–16

Tan PK, Howard JP, Payne GS. 1996. The sequence NPFXD defines a new class of endocytosis signal in *Saccharomyces cerevisiae*. *J. Cell Biol.* 135:1789–800

Tebar F, Bohlander SK, Sorkin A. 1999. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol. Biol. Cell* 10:2687–702

ter Haar E, Harrison SC, Kirchhausen T. 2000. Peptide-in-groove interactions link target proteins to the β-propeller of clathrin. *Proc. Natl. Acad. Sci. USA* 97:1096–100

ter Haar E, Musacchio A, Harrison SC, Kirchhausen T. 1998. Atomic structure of clathrin: a beta propeller terminal domain joins an alpha linker. *Cell* 95:563–73

Timm D, Salim K, Gout I, Guruprasad L, Waterfield M, Blundell T. 1994. Crystal structure of the pleckstrin homology domain from dynamin. *Nat. Struct. Biol.* 1:782–88

Tong X, Boll W, Kirchhausen T, Howley PM. 1998. Interaction of the bovine papillomavirus E6 protein with the clathrin adaptor complex AP-1. *J. Virol.* 72:476–82

Traub LM, Bannykh SI, Rodel JE, Aridor M, Balch WE, Kornfeld S. 1996. AP-2-containing clathrin coats assemble on mature lysosomes. *J. Cell Biol.* 135:1801–14

Traub LM, Downs MA, Westrich JL, Fremont DH. 1999. Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. *Proc. Natl. Acad. Sci. USA* 96:8907–12

Traub LM, Ostrom JA, Kornfeld S. 1993. Biochemical dissection of AP-1 recruitment onto Golgi membranes. *J. Cell Biol.* 123:561–73

Trejo J, Altschuler Y, Fu H-W, Mostov KE, Coughlin SR. 2000. A mutant HeLa cell line suggests novel requirements for protease-activated receptor-1 phosphorylation and recruitment to clathrin-coated pits. *J. Biol. Chem.* 275:31255–65

Umeda A, Meyerholz A, Ungewickell E. 2000. Identification of the universal cofactor (auxilin 2) in clathrin coat dissociation. *Eur. J. Cell Biol.* 79:336–42

Ungewickell E, Ungewickell H, Holstein SEH, Lindner R, Prasad K, et al. 1995. Role of auxilin in uncoating clathrin-coated vesicles. *Nature* 378:632–35

van Kerkhof P, Sachse M, Klumperman J, Strous GJ. 2001. Growth hormone receptor ubiquitination coincides with recruitment to clathrin-coated membrane domains. *J. Biol. Chem.* 276:3778–84

Velier J, Kim M, Schwarz C, Kim TW, Sapp E, et al. 1998. Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways. *Exp. Neurol.* 152:34–40

von Poser C, Zhang JZ, Mineo C, Ding W, Ying Y, et al. 2000. Synaptotagmin regulation of coated pit assembly. *J. Biol. Chem.* 275:30916–24

Vowels JJ, Payne GS. 1998. A dileucine-like

sorting signal directs transport into an AP-3-dependent, clathrin-independent pathway to the yeast vacuole. *EMBO J.* 17:2482–93

Wakeham DE, Ybe JA, Brodsky FM, Hwang PK. 2000. Molecular structures of proteins involved in vesicle coat formation. *Traffic* 1:393–98

Wan L, Molloy SS, Thomas L, Liu GP, Xiang Y, et al. 1998. PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization. *Cell* 94:205–16

Warren RA, Green FA, Enns CA. 1997. Saturation of the endocytic pathway for the transferrin receptor does not affect the endocytosis of the epidermal growth factor receptor. *J. Biol. Chem.* 272:2116–21

Wendland B, Emr SD. 1998. Pan1p, yeast eps15, functions as a multivalent adaptor that coordinates protein-protein interactions essential for endocytosis. *J. Cell Biol.* 141:71–84

Wendland B, Emr SD, Riezman H. 1998. Protein traffic in the yeast endocytic and vacuolar protein sorting pathways. *Curr. Opin. Cell Biol.* 10:513–22

Wendland B, Steece KE, Emr SD. 1999. Yeast epsins contain an essential N-terminal ENTH domain, bind clathrin and are required for endocytosis. *EMBO J.* 18:4383–93

Wessels D, Reynolds J, Johnson O, Voss E, Burns R, et al. 2000. Clathrin plays a novel role in the regulation of cell polarity, pseudopod formation, uropod stability and motility in *Dictyostelium*. *J. Cell Sci.* 113:21–36

Whistler JL, von Zastrow M. 1999. Dissociation of functional roles of dynamin in receptor-mediated endocytosis and mitogenic signal transduction. *J. Biol. Chem.* 274:24575–78

Whitehead B, Tessari M, Carotenuto A, Henegouwen PMPVE, Vuister GW. 1999. The EH1 domain of Eps15 is structurally classified as a member of the S100 subclass of EF-hand-containing proteins. *Biochemistry* 38:11271–77

Wigge P, McMahon HT. 1998. The amphiphysin family of proteins and their role in endocytosis at the synapse. *Trends Neurosci.* 21:339–44

Wilde A, Beattie EC, Lem L, Riethof DA, Liu S-H, et al. 1999. EGF receptor signaling stimulates SRC kinase phosphorylation of clathrin, influencing clathrin redistribution and EGF uptake. *Cell* 96:677–87

Wilde A, Brodsky FM. 1996. In vivo phosphorylation of adaptors regulates their interaction with clathrin. *J. Cell Biol.* 135:635–45

Witke W, Podtelejnikov AV, Di Nardo A, Sutherland JD, Gurniak CB, et al. 1998. In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J.* 17:967–76

Wong DH, Brodsky FM. 1992. 100-kD proteins of Golgi and *trans*-Golgi network-associated coated vesicles have related but distinct membrane binding properties. *J. Cell Biol.* 117:1171–79

Woscholski R, Parker PJ. 1997. Inositol lipid 5-phosphatases—traffic signals and signal traffic. *Trends Biochem. Sci.* 22:427–31

Yamabhai M, Hoffman NG, Hardison NL, McPherson PS, Castagnoli L, et al. 1998. Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains. *J. Biol. Chem.* 273:31401–7

Yao PJ, Morsch R, Callahan LM, Coleman PD. 1999. Changes in synaptic expression of clathrin assembly protein AP180 in Alzheimer's disease analysed by immunohistochemistry. *Neuroscience* 94:389–94

Yao PJ, Weimer JM, O'Herron TM, Coleman PD. 2000. Clathrin assembly protein AP-2 is detected in both neurons and glia, and its reduction is prominent in layer II of frontal cortex in Alzheimer's disease. *Neurobiol. Aging* 21:921–29

Ybe JA, Brodsky FM, Hofmann K, Lin K, Liu S-H, et al. 1999. Clathrin self-assembly is mediated by a tandemly repeated superhelix. *Nature* 399:371–75

Ybe JA, Greene B, Liu S-H, Pley U, Parham P, Brodsky FM. 1998. Clathrin self-assembly is regulated by three light chain residues controlling the formation of critical salt bridges. *EMBO J.* 17:1297–303

Yeung BG, Phan HL, Payne GS. 1999. Adaptor complex-independent clathrin function in yeast. *Mol. Biol. Cell* 10:3643–59

Zhang B, Ganetzky B, Bellen HJ, Murthy VN. 1999. Tailoring uniform coats for synaptic vesicles during endocytosis. *Neuron* 23:419–22

Zhang JZ, Davletov BA, Südhof TC, Anderson RG. 1994. Synaptotagmin I is a high affinity receptor for clathrin AP-2: implications for membrane recycling. *Cell* 78:751–60

Zhao X, Greener T, Al-Hasani H, Cushman SW, Eisenberg E, Greene LE. 2001. Expression of auxilin or AP180 inhibits endocytosis by mislocalizing clathrin: evidence for formation of nascent pits containing AP1 or AP2 but not clathrin. *J. Cell Sci.* 114:353–65

Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, et al. 1996. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 272:1606–14

Zizioli D, Meyer C, Guhde G, Saftig P, von Figura K, Schu P. 1999. Early embryonic death of mice deficient in γ-adaptin. *J. Biol. Chem.* 274:5385–90

Traffic 2000 1: 393-398
Munksgaard International Publishers

ISSN 1398-9219

Toolbox

Molecular Structures of Proteins Involved in Vesicle Coat Formation [1]

Diane E. Wakeham [a], Joel A. Ybe[a], Frances M. Brodsky [a*] and Peter K. Hwang [b]

[a] *G.W. Hooper Foundation, Box 0552, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA*
[b] *Department of Biochemistry and Biophysics, Box 0448, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA*
* *Corresponding author: F.M. Brodsky, fmarbro@itsa.ucsf.edu*

This review includes 16 structures nents and accessory proteins and roles in vesicle budding or coat disassembly.

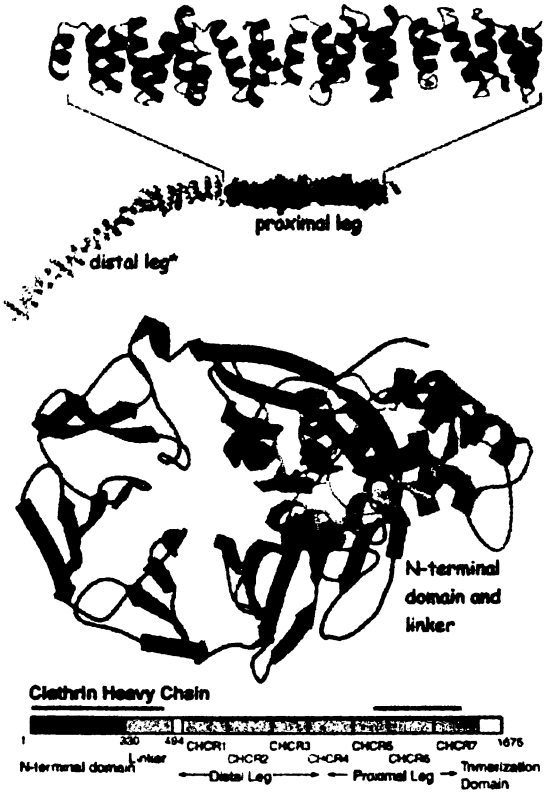
Key words: Adaptin, amphiphysin, GEF, ARNO, arrestin, CHCR, Clathrin, main, Eps15, Hsc70, Nef, PAPb, PH structure, vesicle coat

Received 11 February 2000, revised publication 14 February 2000

Figures in this toolbox were generated from co-ordinates deposited in the Protein Data Bank (1). This article is a composite of work from many laboratories referred within.

Clathrin heavy chain proximal leg (1210-1516) and N-terminal domain and linker (1-494). Clathrin is a trimer of 192 kDa heavy chains, each with an associated regulatory light chain. Adaptor complexes recruit clathrin to the membrane at the cell surface or *trans*-Golgi network where it self-assembles into a lattice coat (2). The proximal leg mediates lattice assembly, which is regulated by binding of acidic light chain and phosphorylation of Y1477 (red) (3). The proximal leg (4) (*upper**,115 x 28x 24Å,) is an elongated rod made up of an extended α/α superhelix, comprised of tandemly repeated 146-residue motifs (CHCRs). Alignments indicate that the structure of clathrin legs is generated by seven CHCRs, each of which contains a stack of 5 helix hairpin pairs. A conserved basic groove (blue) may be the binding site for clathrin light chains. The globular N-terminal domain projects vesicles toward the vesicle membrane (5) to interact with the β-hinge domain of the adaptor complex (6) and other accessory proteins. This domain (7) (*lower*, 47x 40x 75 Å) is a seven-bladed β-propeller structure with an α-helical flexible linker domain (gray). Each blade of the propeller is a slightly twisted antiparallel β-sheet. β-arrestin and β3 adaptin bind in a groove between two blades of the propeller (red) (8,51).

[1] *For all Figures: Black bar indicates portion of protein included in molecular structure solution. Abbreviations: CC, coiled coil domain, CCP, clathrin coated pit, CHCR, clathrin heavy chain repeat, EGF, epidermal growth factor, GAP, GTPase activating protein, GEF, guanine nucleotide exchange factor, PPII helix, polyproline II helix*

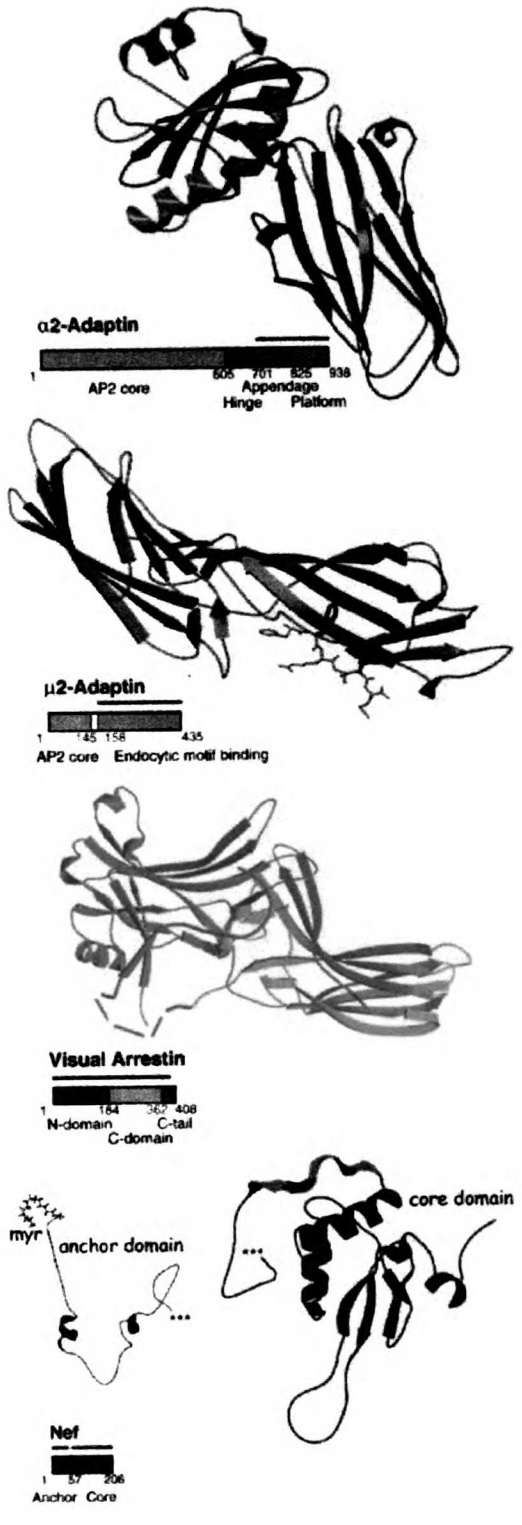


AP2 adaptor α subunit C-terminal appendage domain (701-938). The AP2 adaptor complex (~ 300 kDa) selects molecules for sorting into clathrin-coated vesicles (CCVs)and recruits clathrin to the plasma membrane (16). The ~ 100 kDa α subunit's appendage (47x 53x 59 Å) is a tightly-packed two-lobed structure (17,18). One domain (blue) is a two-sheet β-sandwich of antiparallel strands, resembling the immunoglobulin (IG) superfamily. The second `platform' domain, similar to the yeast TATA-box, is an antiparallel β-sheet `platform' with two buttressing helices below it (yellow, red), and a helix crossing over top. A conserved hydrophobic patch on the platform face centered at W840 (violet) is required for binding to accessory proteins Eps15, epsin, amphiphysin, auxilin, or AP180.

AP2 adaptor μ2 subunit internalization signal binding domain (122-435). The μ2 subunit of AP2 (~ 50 kDa) adaptor binds the endocytic sequence motif of cargo proteins, coupling them to the clathrin coat (16). The elongated, banana-shaped endocytic binding motif (80x 25x 20 Å) has two β-sandwich subdomains (19) (left and right). Signal peptides of *trans*-Golgi network protein TGN38 (DYQRLN) and EGF receptor (FYRALM, shown) both bind in an identical manner as extended β-strands. Hydrophobic cavities binding the Tyr and ϕ residues (blue) are positioned on either side of an edge β strand (pink). μ2 dimerization may contribute to selective recognition of adjacent signal peptides in dimeric receptors.

Visual arrestin (8-402). Arrestins (~ 45 kDa) bind to G-protein coupled receptors (GPCRs)and block G protein binding to terminate signaling (20). Non-visual arrestins, presumably similar in structure, bind clathrin N-terminal domain and can function as adaptors for the internalization of β2-adrenergic receptor (21). Arrestin (95x 45x 60 Å) is composed of seven-stranded β-sandwich N (blue) and C (yellow) domains, and a C tail (orange) that packs up against their interface (22,23). Arrestin's proximity to its phosphorylated receptor may disrupt electrostatic interactions to induce conformation changes that favor GPCR binding. The non-visual arrestins have an LIEFE insertion in an exposed loop (green, dashed) for clathrin binding (21). β-Arrestin localization is regulated by phosphorylation (24) and a phosphoinositide binding site (violet) (25).

HIV-1 protein negative factor (Nef) anchor domain (1-57) and core (57-203). Nef (27 kDa) is crucial for disease progression (26,27). Nef binds to μ-adaptin and vacuolar ATPase NBP1, and accelerates CD4 internalization by localizing CD4 to CCPs. Nef may then bind to endosomal β-COP, leading to CD4 degradation (28). (*Left*, 60x 45x 60 Å) The myristoylated N-terminus anchors Nef to the membrane (29). Kinase interaction is mediated through an α-helix (*left*, green) and a PPII helix RPQVPLR(*right*, orange) in the loosely-packed α/β core (50x 50x 30 Å) (30- 32). A protruding loop (*right*, red) containing a dileucine motif binds μ-adaptin. Residues 57-58 (indicated ***, *left* and *right*) in turn bind to the CD4 dileucine motif.

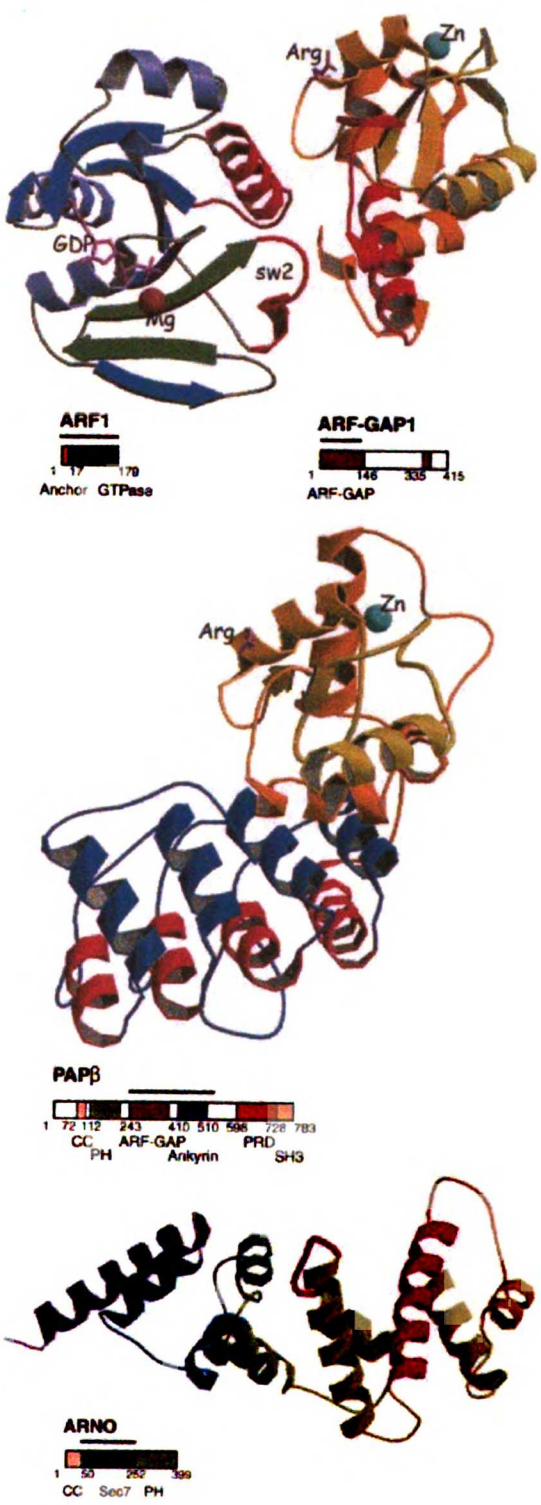


ADP-ribosylation factor 1 (ARF1) (1-179). ARF1 is a membrane-associating 21-kDa GTPase that controls coating and uncoating of COP1 formed in the secretory pathway or on endosomes and clathrin-coated vesicles of cells (33). Coat assembly is initiated when ARF1 binds GTP and recruits coat proteins. After vesicles pinch off, the coat disassembles as ARF hydrolyzes GTP and releases from the membrane. ARF1 is dissociated from the membranes by treatment of cells with Brefeldin A. The myristoylated N-terminus (1-17, not shown) anchors ARF to the membrane (34). (*Upper left*, 30**x** 35**x** 25 Å) ARF1 GTPase core (34,35) shares the structural fold of Ras, an eight stranded β-sheet surrounded by five helices. Strands β2 and β3 (green) and adjacent sw2 (switch2) loop (red) move about 7Å between ARF structures with bound GDP (shown) or bound GTP-analog, suggesting how nucleotide exchange regulates exposure of the myristoylated N-terminus.

ARF GTPase activating protein 1 (ARFGAP1) (1-136). ARFGAP1 (45 kDa) binding to ARF is required to accelerate GTP hydrolysis (33). (*Upper right*, 35**x** 35**x** 25 Å) ARFGAP1 (35) features a GATA-like Cys4 zinc finger ($CX_2CX_{16}CX_2$, yellow in vicinity of zinc) nested against six helices and a β strand. Binding between ARF1 and ARFGAP1 involves the structural components highlighted in red. The role of a conserved Arg (violet) remains unsettled, but it appears essential for GTPase activity. However, in the crystallized complex (as shown), this residue is too distant from the GTP site to serve as catalytic Arg finger.

PYK2 tyrosine kinase activating protein β-subunit (PAPβ): ARF-GAP domain and ankyrin repeats (112-522). PAPβ (88 kDa) activates ARF1 GTP hydrolysis and contains C-terminal ankyrin repeats common in other proteins with ARF-GAP activity (36). ARF-GAP domains of PAPβ (37) (*center*, 42**x** 28**x** 26 Å) and ARFGAP1 (*upper right*) are similar, although divergent in C-terminal portions buttressing the back of the zinc finger module. Comparison of structures suggest that either the ankyrin repeats (alternating blue and red helices, 40**x** 15**x** 20 Å) are dislodged from the ARF-GAP domain before binding ARF1 or that the PAPβ ARF-GAP binds ARF1 differently from the previous structure.

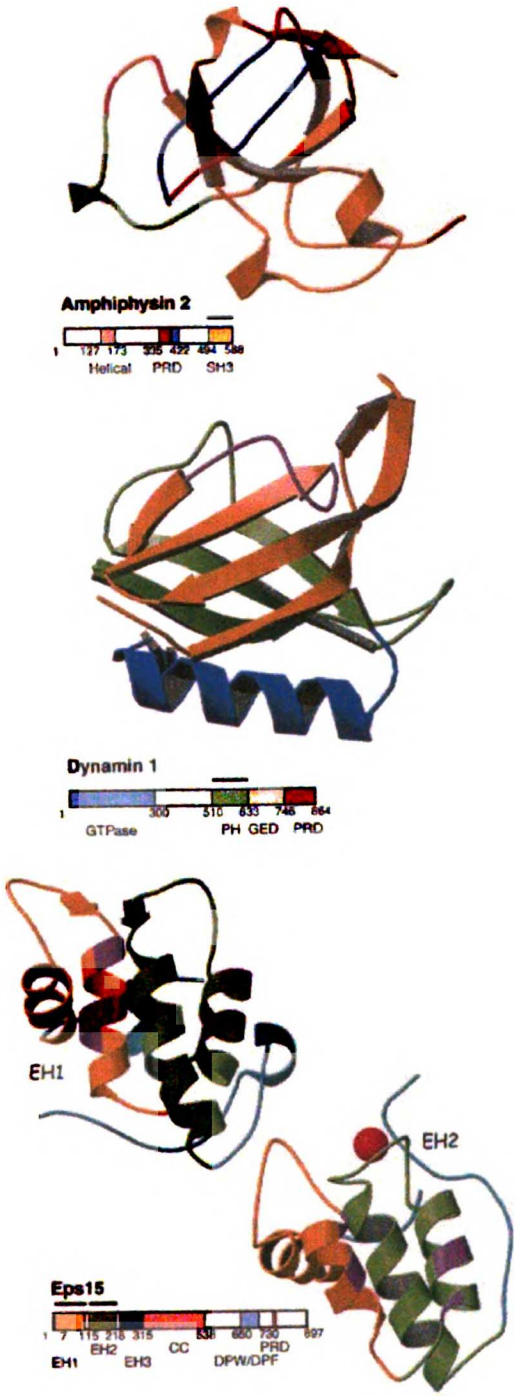
ARF nucleotide-binding-site opener (ARNO): Sec7 ARF-GEF domain (50-252). The Sec7 domain of the 47-kDa protein, ARNO, catalyzes nucleotide exchange in the G-protein ARF1 (38). After GDP to GTP exchange, ARF1 initiates the coating of vesicles formed in the endoplasmic reticulum, Golgi apparatus, TGN and endosomes. The crystal structure (2 Å) of ARNO-Sec7 (39), resembling a flared cylinder, is a series of α-helices folded in a distorted right-handed superhelix (70**x** 40**x** 40 Å). Two highly conserved regions, Motif1-loop and Motif2-helix (red), define opposite walls of a groove into which both Brefeldin A and the switch 2 region of ARF1 fit (34,39). Substitutions in either motif block ARNO-Sec7 nucleotide exchange activity.



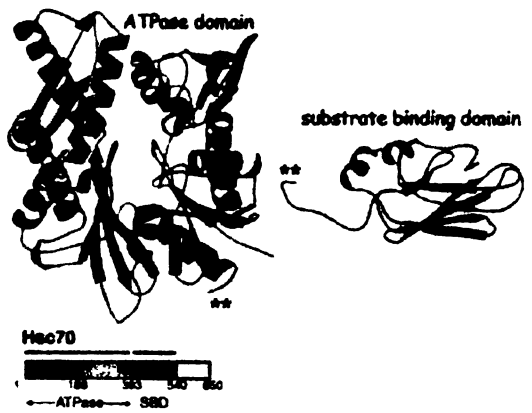
Amphiphysin 2: *Src*-homology 3 (SH3) domain (494-588). Amphiphysin isoforms 1 and 2 (49% sequence identity, ~ 95 kDa) form a heterodimer that binds to the C-terminal appendage domain of the AP2 adaptor α-chain in CCVs (40). Both isoforms have SH3 domains which recognize the PSRPNR sequence within dynamin's proline rich domain (PRD). Amph1 and Amph2 also bind clathrin, synaptojanin, and endophilin. Isolated Amph SH3 domains disrupt endocytosis by preventing multimerization of dynamin. The Amph2 SH3 domain (41) (35**x** 25**x** 30 Å) comprises a five-stranded β-barrel with a hydrophobic binding face formed by the RT (blue) and n-Src (green) loops for interacting with Pro-rich ligand sequences. The unique n-Src loop in Amph 2 introduces acidic residues, specific for the two R residues in its dynamin binding site (red) and its extended size explains the steric interference with dynamin multimerization. Thus, amphiphysin recruits dynamin to CCVs, but negatively regulates dynamin assembly until the two proteins dissociate.

Dynamin 1: pleckstrin homology (PH) domain (518-630). Dynamin, a GTPase of ~ 100 kDa that is essential for endocytosis, is recruited to a CCV during scission from the plasma membrane. Dynamin self-assembles into a collar at the vesicle-membrane attachment site, a process that regulates its GTPase activity and controls membrane scission (42,43). The dynamin 1 PH domain (44,45) is a β-sandwich (40**x** 40**x** 35 Å) of two orthogonally-oriented β-sheets (yellow and green), flanked on one side by an α-helix (blue). On the other side of the sandwich, protruding loops form a positively charged surface for binding to proteins or to phosphoinositide (violet) (46). PH-mediated phosphoinositide binding is not essential for dynamin membrane localization, but may be important for function (47).

EGF receptor substrate 15 (Eps15): EH1 (7-115) and EH2 (115-218) domains. Eps15 (~ 100 kDa) binds the C-terminal appendage domain of the AP2 adaptor α-subunit in CCVs (48). Eps15 function is essential to endocytosis, mediating the interaction of AP2 with proteins containing NPF or W/FW sequences through its three EH domains. Binding proteins include epsin, CALM: AP180 and synaptojanin, all implicated in regulation of receptor-mediated endocytosis. Eps15 forms oligomers through a coiled-coil domain in the center of the molecule, suggesting a structural or cytoskeletal role. The EH1 (49) (*upper*, 25**x** 35**x** 30 Å) and EH2 (50) (*lower*, 25**x** 35**x** 30 Å) domains of eps15 each comprise two helix-loop-helix EF hand motifs (green and yellow), which are connected by a short anti-parallel β-sheet in EH1. Ca^{2+} binding (red sphere) in EH2 likely has a structural role. The NPF-binding site of each is formed by hydrophobic residues along helical faces contributed by both EF hands (violet).



Heat-shock cognate 70 kDa protein (Hsc70) ATPase (1-384) and substrate binding domain (SBD) (383-540). The molecular chaperone Hsc70 (9) is the uncoating ATPase for the disassembly of clathrin lattices and also contributes to adaptor uncoating (10). The portion of the protein represented by these two structures together is sufficient for uncoating activity (11). The DnaJ homologue auxilin binds to the proximal leg of assembled clathrin and to ATP-bound Hsc70 (12) to mediate clathrin uncoating. (*Left*, 50x 50x 20 Å) The Hsc70 ATPase domain (13) is a member of the hexokinase:actin superfamily of structures (14). ATP binds in a deep cleft (red) between two subdomains (yellow, cyan). The connection point between the ATPase and SBD domains is indicated (**). (*Right*, 55x 30x 20 Å) SBD (15) contains a β-sandwich domain (purple) with a helix latched on top (orange), where L539 (green) blocks the substrate binding groove. Nucleotide-dependent movement of this helical latch may make the groove accessible (16).



References

1. Protein Data Bank (PDB) identification for structures shown: Clathrin heavy chain proximal leg: 1B89; Clathrin heavy chain N-terminal domain: 1BPO; Hsc70 ATPase domain: 3HSC; Hsc70 SBD: 7HSC; α-Adaptin appendage: 1B9K, 1QTP, 1QTS; μ2-Adaptin binding domain: 1BW8; Visual arrestin: 1CF1; HIV-1 Nef Anchor: 1QA5; HIV-1 Nef Core: 2NEF; PAPβ ARF-GAP and ankyrin domains: 1DCQ; ARNO ARF-GEF: 1PBV; Amphiphysin 2 SH3 Domain: 1BB9; Dynamin PH Domain: 2DYN; Eps15 EH1 Domain: 1QJT; Eps15 EH2 Domain: 1EH2. ARF1-ARF-GAP complex: The authors thank J. Goldberg for the coordinates of this structure.
2. Schmid SL. Clathrin-coated vesicle formation and protein sorting: an integrated process. Annu Rev Biochem 1997;66: 511-548.
3. Wilde A, Beattie EC, Lem L, et al. EGF receptor signaling stimulates SRC kinase phosphorylation of clathrin, influencing clathrin redistribution and EGF uptake. Cell 1999;96: 677-687.
4. Ybe JA, Brodsky FM, Hofmann K, et al. Clathrin self-assembly is mediated by a tandemly repeated superhelix. Nature 1999;399: 371-375.
5. Smith CJ, Grigorieff N, Pearse BMF. Clathrin coats at 21 Å resolution: a cellular assembly designed to recycle multiple membrane receptors. EMBO J 1998;17: 4943-4953.
6. Shih W, Gallusser A, Kirchhausen T. A clathrin-binding site in the hinge of the β2 chain of mammalian AP-2 complexes. J Biol Chem 1995;270: 31083-31090.
7. ter Haar E, Musacchio A, Harrison SC, Kirchhausen T. Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. Cell 1998;95: 563-573.
8. Goodman OB Jr, Krupnick JG, Gurevich VV, Benovic JL, Keen JH. Arrestin:clathrin interaction. Localization of the arrestin binding locus to the clathrin terminal domain. J Biol Chem 1997;272: 15017-15022.
9. Hartl FU, Hlodan R, Langer T. Molecular chaperones in protein folding: the art of avoiding sticky situations. Trends Biochem Sci 1994;19: 20-25.
10. Hannan LA, Newmyer SL, Schmid SL. ATP- and cytosol-dependent release of adaptor proteins from clathrin-coated vesicles: a dual role for Hsc70. Mol Biol Cell 1998;9: 2217-2229.
11. Ungewickell E, Ungewickell H, Holstein SE. Functional interaction of the auxilin J domain with the nucleotide- and substrate-binding modules of Hsc70. J Biol Chem 1997;272: 19594-19600.
12. Holstein SE, Ungewickell H, Ungewickell E. Mechanism of clathrin basket dissociation: separate functions of protein domains of the DnaJ homologue auxilin. J Cell Biol 1996;135: 925-937.
13. Flaherty KM, DeLuca-Flaherty C, McKay DB. Three-dimensional structure of the ATPase fragment of a 70 K heat-shock cognate protein. Nature 1990;346: 623-628.
14. Hurley JH. The sugar kinase:heat shock protein 70:actin superfamily: implications of conserved structure for mechanism. Annu Rev Biophys Biomol Struct 1996;25: 137-162.
15. Morshauser RC, Hu W, Wang H, Pang Y, Flynn GC, Zuiderweg ER. High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. J Mol Biol 1999;289: 1387-1403.
16. Kirchhausen T. Adaptors for clathrin-mediated traffic. Annu Rev Cell Devel Biol 1999;15: 705-732.
17. Owen DJ, Vallis Y, Noble ME, et al. A structural explanation for the binding of multiple ligands by the alpha-adaptin appendage domain. Cell 1999;97: 805-815.
18. Traub LM, Downs MA, Westrich JL, Fremont DH. Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. Proc Natl Acad Sci USA 1999;96: 8907 8912.
19. Owen DJ, Evans PR. A structural explanation for the recognition of tyrosine-based endocytotic signals. Science 1998;282: 1327-1332.
20. Lefkowitz RJ. G protein-coupled receptors. III. New roles for receptor kinases and beta-arrestins in receptor signaling and desensitization. J Biol Chem 1998;273: 18677-18680.
22. Granzin J, Wilden U, Choe HW, Labahn J, Krafft B, Buldt G. X-ray crystal structure of arrestin from bovine rod outer segments. Nature 1998;391: 918-921.
23. Hirsch JA, Schubert C, Gurevich VV, Sigler PB. The 2.8 Å crystal structure of visual arrestin: a model for arrestin's regulation. Cell 1999;97: 257-269.
21. Krupnick JG, Goodman Jr. OB, Keen JH, Benovic JL. Arrestin-clathrin interaction. Localization of the clathrin binding domain of nonvisual arrestins to the carboxyl terminus. J Biol Chem 1997;272: 15011-15016.
24. Lin F-T, Krueger KM, Kendall HE, et al. Clathrin-mediated endocytosis of the β-adrenergic receptor is regulated by phosphorylation dephosphorylation of β-arrestin 1. J Biol Chem 1997;49: 31051-31057.
25. Gaidarov I, Krupnick JG, Falck JR, Benovic JL, Keen JH. Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. EMBO J 1999;18: 871-881.
26. Harris M. HIV: a new role for Nef in the spread of HIV. Curr Biol 1999;9: R459-R461.
27. Oldridge J, Marsh M. Nef - an adaptor adaptor? Trends Cell Biol 1998;8: 302-305.

28. Piguet V, Gu F, Foti M, et al. Nef-induced CD4 degradation: a diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of beta-COP in endosomes. Cell 1999;97: 63-73.
29. Geyer M, Munte CE, Schorr J, Kellner R, Kalbitzer HR. Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. J Mol Biol 1999;289: 123-138.
30. Grzesiek S, Bax A, Hu JS, et al. Refined solution structure and backbone dynamics of HIV-1 Nef. Protein Sci 1997;6: 1248-1263.
31. Lee CH, Saksela K, Mirza UA, Chait BT, Kuriyan J. Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell 1996;85: 931-942.
32. Arold S, Franken P, Strub MP, et al. The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. Structure 1997;5: 1361-1372.
33. Roth MG. Snapshots of ARF1: implications for mechanisms of activation and inactivation. Cell 1999;97: 149-152.
34. Amor JC, Harrison DH, Kahn RA, Ringe D. Structure of the human ADP-ribosylation factor 1 complexed with GDP. Nature 1994;372: 704-708.
35. Goldberg J. Structural and functional analysis of the ARF1-ARFGAP complex reveals a role for coatomer in GTP hydrolysis. Cell 1999;96: 893-902.
36. Andreev J, Simon JP, Sabatini DD, et al. Identification of a new Pyk2 target protein with Arf-GAP activity. Mol Cell Biol 1999;19: 2338-2350.
37. Mandiyan V, Andreev J, Schlessinger J, Hubbard SR. Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-associated protein beta. EMBO J 1999;18: 6890-6898.
38. Chardin P, Paris S, Antonny B, et al. A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. Nature 1996;384: 481-484.
39. Cherfils J, Menetrey J, Mathieu M, et al. Structure of the Sec7 domain of the Arf exchange factor ARNO. Nature 1998;392: 101-105.
40. Wigge P, McMahon HT. The amphiphysin family of proteins and their role in endocytosis at the synapse. Trends Neurosci 1998;21: 339-344.
41. Owen DJ, Wigge P, Vallis Y, Moore JD, Evans PR, McMahon HT. Crystal structure of the amphiphysin-2 SH3 domain and its role in the prevention of dynamin ring formation. EMBO J 1998;17: 5273-5285.
42. Sever S, Muhlberg AB, Schmid SL. Impairment of dynamin's GAP domain stimulates receptor-mediated endocytosis. Nature 1999;398: 481-486.
43. Marsh M, McMahon HT. The structural era of endocytosis. Science 1999;285: 215-220.
44. Ferguson KM, Lemmon MA, Schlessinger J, Sigler PB. Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. Cell 1994;79: 199-209.
45. Timm D, Salim K, Gout I, Guruprasad L, WaterÆeldM, Blundell T. Crystal structure of the pleckstrin homology domain from dynamin. Nat Struct Biol 1994;1: 782-788.
46. Zheng J, Cahill SM, Lemmon MA, Fushman D, Schlessinger J, Cowburn D. Identification of the binding site for acidic phospholipids on the PH domain of dynamin: implications for stimulation of GTPase activity. J Mol Biol 1996;255: 14-21.
47. Schmid SL, McNiven MA, De Camilli P. Dynamin and its partners: a progress report. Curr Opin Cell Biol 1998;10: 504-512.
48. Mayer BJ. Endocytosis: EH domains lend a hand. Curr Biol 1999;9: R70-R73.
49. Whitehead B, Tessari M, Carotenuto A, van Bergen en Henegouwen PM, Vuister GW. The EH1 domain of Eps15 is structurally classified as a member of the S100 subclass of EF-hand-containing proteins. Biochemistry 1999;38: 11271-11277.
50. de Beer T, Carter RE, Lobel-Rice KE, Sorkin A, Overduin M. Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. Science 1998;281: 1357-1360.
51. ter Haar E, Harrison S, Kirchausen T. Peptide-in-groove interactions link target proteins to the β-propeller of clathrin. Proc Natl Acad Sci 2000;97: 1096-1100.

Traffic 2000 **1**: 474-479
Munksgaard International Publishers

ISSN 1398-9219

Toolbox

Molecular Structures of Proteins Involved in Vesicle Fusion

Joel A. Ybe[a], Diane E. Wakeham[a], Frances M. Brodsky[a,*] and Peter K. Hwang[b]

[a] *G.W. Hooper Foundation, Box 0552, Department of Microbiology and Immunology, Department of Biopharmaceutical Sciences, University of California, 513 Parnassus Avenue, San Francisco, CA 94143, USA*
[b] *Department of Biochemistry and Biophysics, Box 0448, University of California, 513 Parnassus Avenue, San Francisco, CA 94143, USA*
** Corresponding author: F.M. Brodsky, fmarbro@itsa.ucsf.edu*

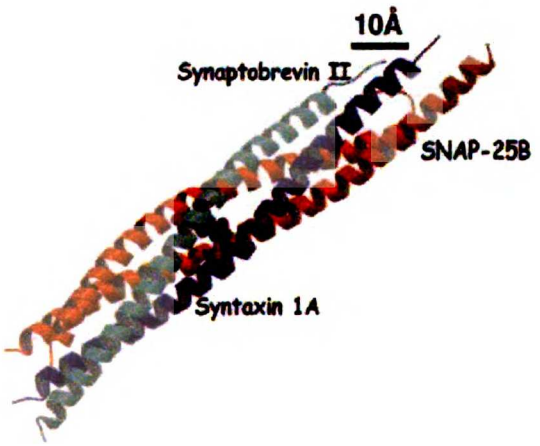
We present a summary of the structures of 13 proteins involved in the docking and fusion of intracellular transport vesicles to their target membranes.

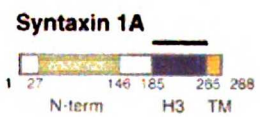
Key words: C2 domain, D2 domain, fusion, NSF, Rab, Rab-GDI, Rabphilin, Sec 17, α-SNAP, SNAP-25, synaptic fusion complex, structure, synaptobrevin, synaptotagmin, syntaxin, vesicle docking

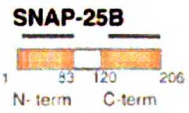
Received and accepted for publication 15 March 2000

Figures in this toolbox were generated from coordinates deposited in the Protein Data Bank (1) using Molscript (2) and Raster 3D (3). The structures described in this article are a composite of work from many laboratories referred to within.[1]

Synaptic fusion complex of syntaxin IA (188-259), synaptobrevin II (25-93) and SNAP-25B (7-83, 131-204). Syntaxin IA, synaptobrevin II and SNAP-25B spontaneously assemble into a stable ternary (SNARE) complex to drive vesicle docking. The crystal structure (2.4 Å) of the fusion complex, consisting of the carboxyl terminal H3 domain of syntaxin 1A (9 kDa) (blue), the cytoplasmic domain of synaptobrevin II (11 kDa) (green) and the N- and C-terminal portions of SNAP-25B (9 and 10 kDa) (orange), is a cable of four intertwined α-helices with their N-termini at one end and their C-termini at the other (4). Each helix in the fusion complex contributes residues (shown in center of complex) that interact in a similar manner to other leucine-zipper proteins, such as transcriptional activator GCN4. The ternary complex is further stabilized by ionic interactions, shielded from solvent by the surrounding leucine-zipper layer (4). This arrangement may have a functional impact on disassembling the SNARE complex, since disrupting the hydrophobic layer by the binding of α-SNAPs:Sec17 will expose buried salt bridges to facilitate disassembly.



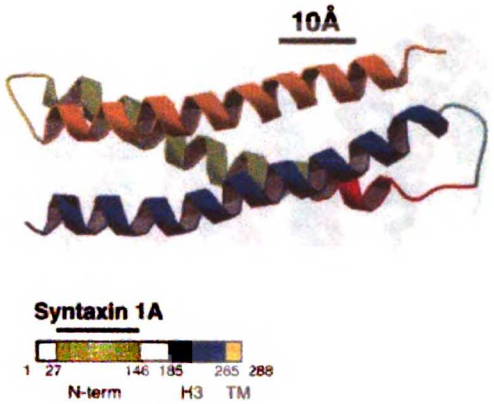




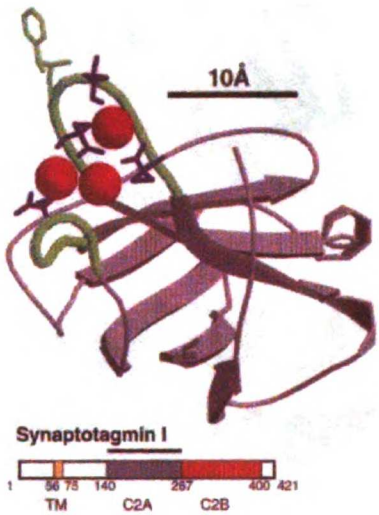
Abbreviations: IP_4, inositol tetraphosphate; IP_5, inositol pentaphosphate; SNAP, soluble NSF attachment protein; SNAP-25, synaptosomal-associated protein of 25 kDa; SNARE, SNAP receptors; Syt, synaptotagmin; TM, transmembrane domain

[1] For all figures: Black bar indicates portion of protein included in molecular structure solution.

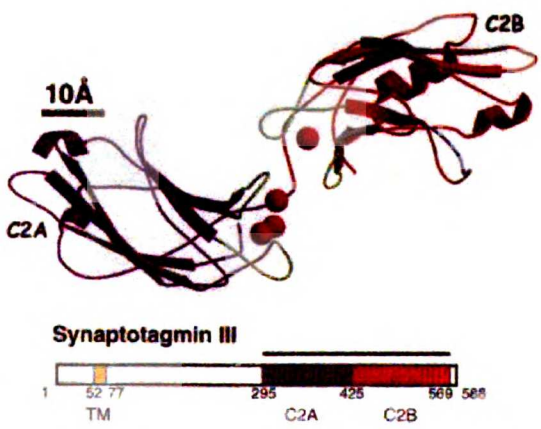
Syntaxin 1A N-terminal domain (27-146). The N-terminal domain of the syntaxin family is highly variable; however, some residues are conserved in syntaxin isoforms involved in neurotransmitter release at the plasma membrane (5). The syntaxin 1A N-terminal domain (13.2 kDa) is essential for the binding of syntaxin-interacting proteins munc13s (6) and munc18-1 (7,8). The NMR structure reveals three long α-helices in a twisted left-handed up-down bundle (9). A long groove lined with conserved residues may be important for specific protein- protein interactions. The N-terminal domain is also marked by a highly charged region (red) that binds to the C2A domain of synaptotagmin I in a calcium-dependent manner and may act as an electrostatic switch to regulate neurotransmitter release (9).



Synaptotagmin I C2A domain (140-267). Synaptotagmins are transmembrane Ca^{2+} sensors regulating synaptic vesicle fusion (10). The 11 known isoforms, seven of which are neuronal, vary in their affinity for calcium or inositol polyphosphates, inositol tetraphosphate (IP_4) and inositiol pentaphosphate (IP_5) (11). Neuronal Syt1 (65 kDa) binds to calcium and IP_4 with moderate affinity. Synaptotagmins form homo- or heteromultimers in the cell to regulate synaptic plasticity (12). The synaptotagmin I (Syt1) C2A domain (13,14) is an eight-stranded β-sandwich which uses aspartates and serine (violet) in two loops (green) to bind three Ca^{2+} (red) following synaptic Ca^{2+} influx. Electrostatic changes then favor insertion of Phe 204 (green) into the phospholipid bilayer (15) and Syt binding to components of the SNARE complex to stimulate exocytosis. Ca^{2+} -independent interaction between Syt1 C2B and the AP-2 clathrin adaptor complex is enhanced by cargo binding to AP-2 (16) and may recruit AP-2 to the membrane for endocytosis.

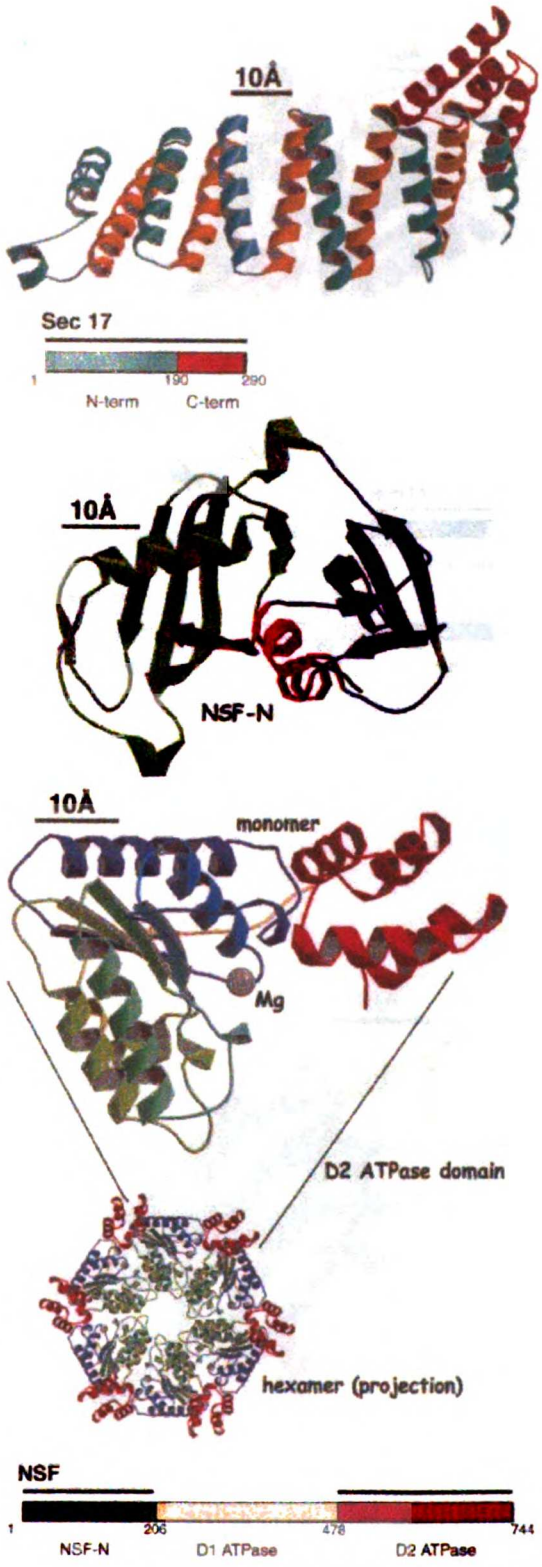


Synaptotagmin III C2A and C2B domains (295-566). This 65 kDa neuronal isoform of synaptotagmin is very calcium sensitive and has no detectable IP_4 affinity (11). The two C2 domains of Syt3 (17) are connected by a flexible linker to bring the calcium binding sites (green) face-to-face. The C2A domain (purple) binds three ions (red), while the lower affinity C2B domain (orange) binds just one (red). Ca^{2+} -dependent binding of Syt to the SNARE fusion complex requires both C2A and C2B. C2B has a basic region (blue) which in some Syt isoforms is important for binding to AP-2, Ca^{2+} channels, inositol polyphosphates, and for dimerization (18). SytC2B binding to IP_5, which accumulates after depolarization, inhibits the fusion step of exocytosis. Hence, Syts initially trigger and subsequently arrest vesicle fusion in response to secondary messengers.



Sec17 vesicular transport protein (1-290) (α-SNAP yeast homologue). Three molecules of yeast SNAP Sec17, each 32 kDa, form a sheath around the SNARE complex to permit the binding of the cytosolic chaperone NSF. The Sec17 crystal structure (2.9 A) is an N-terminal twisted sheet of α-helical hairpins (green and yellow) and a C-terminal α-helical bundle (red) (19). The N-terminus resembles HEAT repeats (20- 23), armadillo repeats (24,25) and clathrin heavy chain repeats (26). Several models based on surface charge profiles have been proposed to understand the SNAP:SNARE interaction. The acidic concave face of the N-terminal sheet of helical hairpins (orange helices) would unlikely interact with highly acidic SNARE complex. In contrast, the opposite helical face has a slightly basic charge distribution, suggesting this side may provide a binding site. The disposition of Sec17 on the SNARE complex suggests they may act as torque arms that transmit forces generated by NSF to pull apart the fusion complex (19).

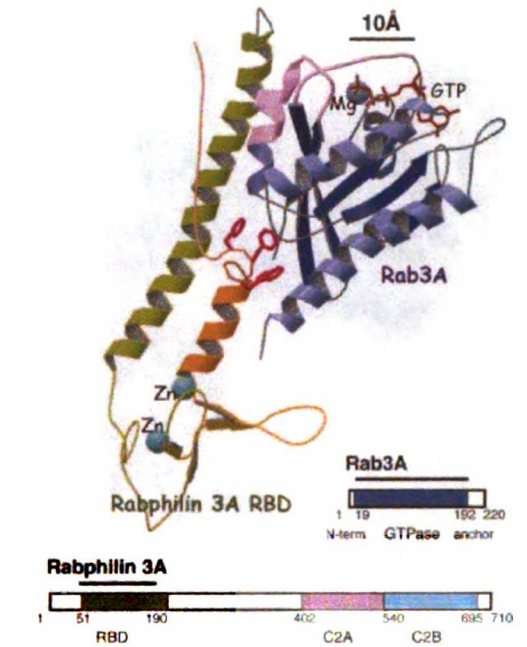
NSF (N-ethylmaleimide-sensitive fusion protein) N-terminal (1-201) and D2 (478-744) domains. NSF, a 20S hexameric protomer, serves as a cytosolic ATPase chaperone to disassemble the SNARE fusion complex. The N-terminal domain (NSF-N) (22 kDa) (*upper*) associates with the C-terminal region of α-SNAP:Sec17 and its ATPase activity is stimulated, in part, by this interaction. The kidney-shaped structure of NSF-N solved independently by two groups (27,28) has two subdomains (blue and green). The fragment crystallized as a trimer (blue to green arrangement), probably its functional form. A highly basic groove (red) may be the α-SNAP:Sec17 binding site (27). The structure of the N-terminal domain of the yeast orthologue (Sec18) closely resembles its mammalian counterpart (29). The low-affinity ATP-binding hexameric D1 domain (structure undetermined) stacks on D2 (high-affinity ATP-binding, low hydrolysis rate), forming a two-layer washer stack underneath NSF-N. The D1 domain drives SNARE complex disassembly, while D2 is important in hexamer formation. The 29 kDa D2 protomer structure (*lower*) (30,31) has a nucleotide-binding domain (blue and green) and a helical region (red) important for hexamer formation. When NSF is bound to the (α-SNAP:Sec17):SNARE complex, the D1±D2 ring stack shows diffuse features, suggested to be NSF-N, splayed away from the D1±D2 pore (ATP hydrolysis dilates the D1 ring and pulls these features in). This is consistent with the '3-in, 3-out' model, which suggests there are two sets of NSF-N trimers that exchange position on the D1±D2 stack by flipping each of their monomers in or out of position, resulting in a 60° rotation of the SNAPs with respect to the SNARE complex (28).

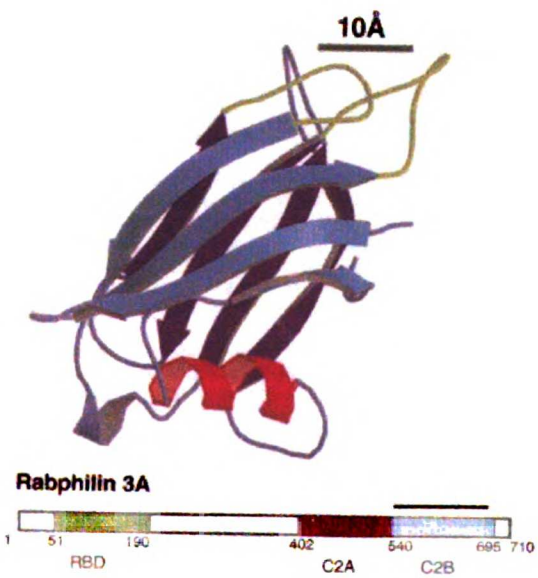


Rab3A (18-192). Rab proteins constitute a large family (over 40 distinct members) of GTPases that participate with SNAREs in vesicle targeting and membrane fusion (32). Rabs form diverse complexes with effectors that regulate and impart specificity to membrane trafficking at the levels of vesicle transport, docking and fusion. These include guanine nucleotide exchange factors (GEFs), GTPase-activating proteins (GAPs), and GDP dissociation inhibitors (GDIs). Rab3A isoforms associate with presynaptic vesicles to regulate Ca^{2+} -dependent exocytosis and neurotransmitter release. The Rab3A GTPase core (19 kDa, blue, *upper right*) shares the structural fold of Ras, with a central six-stranded beta sheet surrounded by five helices (33,34). The isoprenylated C-terminal extension (missing from model) likely serves as the membrane anchor (35). The conserved switch 1 and switch 2 regions (violet) are sensitive to nucleotide-binding state, but specificity determinants for effector interactions may reside elsewhere in the protein.

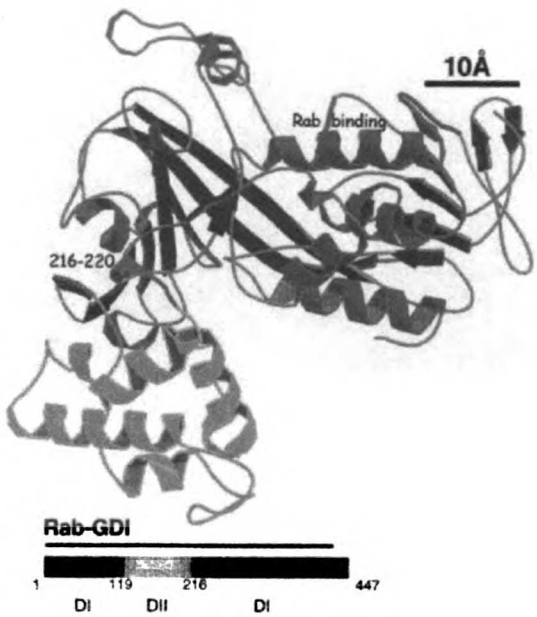
Rabphilin 3A Rab-binding domain (RBD) (44-167). Rabphilin 3A binds to the active GTP-bound state of Rab3A to form a protein complex that mediates docking of synaptic vesicles at the synaptic membrane. The 14 kDa RBD resembles a pistol (*left of Rab3A*), with a 'pistol barrel' composed of an extended helix (green), a 'handle' region consisting of two Zn^{2+} -binding sites and interspersed loops, and a 'trigger' made of a turn containing a conserved SGAWFF motif (WFF side-chains are shown in red). In the crystallized complex, most side-chain interactions between Rab3A and Rabphilin 3A are hydrophobic (34). Of the two distinct interface areas, the first involves the highly conserved switch 1 and switch 2 regions of Rab3A (violet). The second interface involves a deep pocket in Rab3A that interacts with the SGAWFF motif. Sequence variability suggests that this pocket may determine the specificity of interaction between Rabs and their effectors.

Rabphilin 3A C2B domain (541-680). Tandemly repeated C2 domains are common in Ca^{2+} -sensing mediators of interactions between proteins involved in membrane trafficking (36). Both rabphilin and synaptotagmin contain consecutive C2 domains (C2A and C2B) that share a high degree of homology but also have specific differences. The Rabphilin C2B domain (18 kDa) binds two Ca^{2+} ions with micromolar affinity and has the distinctive property of participating in Ca^{2+} -dependent and Ca^{2+} -independent interactions (37). NMR spectroscopy indicates a structure consisting of a β-sandwich formed by two four-stranded β-sheets (37), as observed in other C2 domains (see synaptotagmins above). Ca^{2+} binding sites are formed by residues in loops 1 and 2 (yellow). The C2B structure also contains an α-helix (helix 2, red) at the bottom of the sandwich that is absent in C2A domains. Helix 2 packs against the interface between the two β-sheets and may mediate the Ca^{2+} -independent interactions of C2 domains.





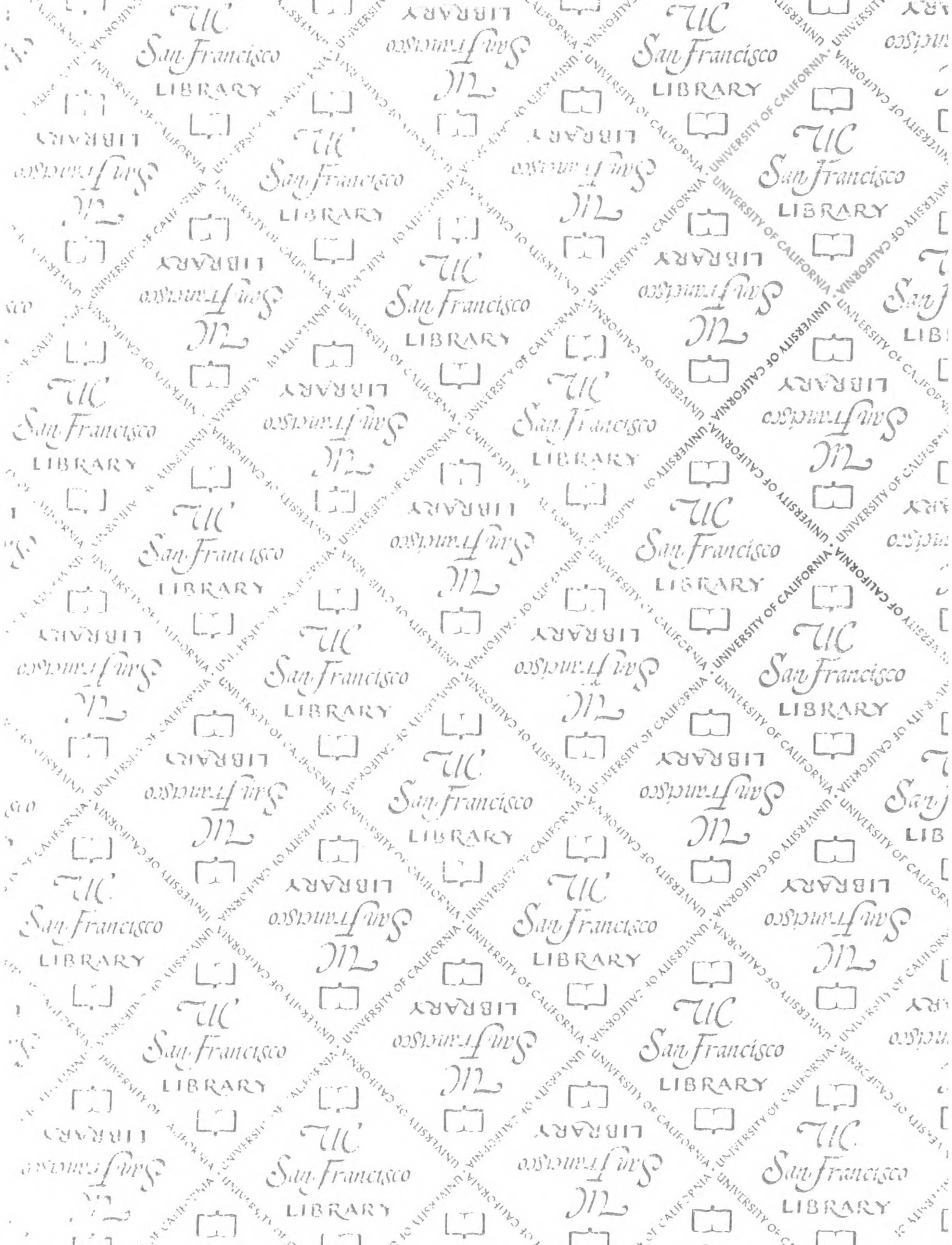
(Rab-GDI) Rab GDP-dissociation inhibitor. Rab-GDI (55 kDa) maintains Rab in a soluble, inactive, GDP-bound state in the cytosol. Following GTP hydrolysis, the binding of GDI to Rab also facilitates the release of GDP-bound Rab from the membrane. GDI is eventually displaced by the GDI displacement factor (GDF), allowing Rab to reassociate with membrane. GDI consists of two major structural domains, a large cylindrical multisheet module (domain I, 340 residues, blue), related in structural organization to FAD-dependent Flavoproteins, and a smaller α-helical domain (domain II, 119-219, green) (38). Site-directed mutagenesis indicates that a region of conserved residues (19, 232-259, red) plays a critical role in the binding of Rab3A. Furthermore, residues in a mobile domain II loop (216-220) may be required for binding GDI to membranes and extracting membrane-bound Rab (39).



References

1. PDB (Protein Data Bank) Identification for structures shown: SNAP-25B:syntaxin 1A:synaptobrevin II fusion complex: 1SFC; Synaptotagmin I C2A: 1BYN; Synaptotagmin III C2A and C2B: 1DQV; Syntaxin 1A N-terminal domain: IBR0; Sec 17: 1QQE; NSF N-terminal domain: 1QCS, 1QDN; NSF D2 domain: 1NSF; Rab3A:Rabphilin 3A complex: 1ZBD; Rabphilin 3A C2B: 3RPB; Rab-GDI: 1GND.
2. Kraulis PJ. Molscript - a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 1991;24: 946-950.
3. Merritt EA, Murphy MEP. Raster3d Version 2.0 - a program for photorealistic molecular graphics. Acta Crystallogr D : Biol Cryst 1994;50: 869-873.
4. Sutton RB, Fasshauer D, Jahn R, Brunger AT. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature 1998;395: 347-353.
5. Bennett MK, Garcia-Arraras JE, Elferink LA, et al. The syntaxin family of vesicular transport receptors. Cell 1993;74: 863-873.
6. Betz A, Okamoto M, Benseler F, Brose N. Direct interaction of the rat unc-13 homologue Munc13-1 with the N terminus of syntaxin. J Biol Chem 1997;272: 2520-2526.
7. Hata Y, Slaughter CA, Sudhof TC. Synaptic vesicle fusion complex contains unc-18 homologue bound to syntaxin. Nature 1993;366: 347-351.
8. Kee Y, Lin RC, Hsu SC, Scheller RH. Distinct domains of syntaxin are required for synaptic vesicle fusion complex formation and dissociation. Neuron 1995;14: 991-998.
9. Fernandez I, Ubach J, Dulubova I, Zhang X, Sudhof TC, Rizo J. Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. Cell 1998;94: 841-849.
10. Sudhof TC, Rizo J. Synaptotagmins: C2-domain proteins that regulate membrane traffic. Neuron 1996;17: 379-388.
11. Mikoshiba K, Fukuda M, Ibata K, Kabayama H, Mizutani A. Role of synaptotagmin, a Ca^{2+} and inositol polyphosphate binding protein, in neurotransmitter release and neurite outgrowth. Chem Phys Lipids 1999;98: 59-67.
12. Littleton JT, Serano TL, Rubin GM, Ganetzky B, Chapman ER. Synaptic function modulated by changes in the ratio of synaptotagmin I and IV. Nature 1999;400: 757-760.
13. Sutton RB, Davletov BA, Berghuis AM, Sudhof TC, Sprang SR. Structure of the first C2 domain of synaptotagmin I: a novel Ca^{2+} : phospholipid-binding fold. Cell 1995;80: 929-938.
14. Shao X, Fernandez I, Sudhof TC, Rizo J. Solution structures of the Ca^{2+} -free and Ca^{2+} -bound C2A domain of synaptotagmin I: does Ca^{2+} induce a conformational change? Biochemistry 1998;37: 16106-16115.
15. Zhang X, Rizo J, Sudhof TC. Mechanism of phospholipid binding by the C2A domain of synaptotagmin I. Biochemistry 1998;37: 12395-12403.
16. Haucke V, De Camilli P. AP-2 recruitment to synaptotagmin stimulated by tyrosine-based endocytic motifs. Science 1999;285: 1268-1271.
17. Sutton RB, Ernst JA, Brunger AT. Crystal structure of the cytosolic C2A-C2B domains of synaptotagmin III. Implications for Ca(+ 2)-independent snare complex interaction. J Cell Biol 1999;147: 589-598.
18. Chapman ER, Desai RC, Davis AF, Tornehl CK. Delineation of the oligomerization, AP-2 binding, and synprint binding region of the C2B domain of synaptotagmin. J Biol Chem 1998;273: 32966-32972.
19. Rice LM, Brunger AT. Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. Mol Cell 1999;4: 85-95.
20. Chook YM, Blobel G. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nature 1999;399: 230-237.
21. Cingolani G, Petosa C, Weis K, Muller CW. Structure of importin-beta bound to the IBB domain of importin-alpha. Nature 1999;399: 221-229.
22. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell 1999;96: 99-110.
23. Vetter IR, Nowak C, Nishimoto T, Kuhlmann J, Wittinghofer A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. Nature 1999;398: 39-46.
24. Huber AH, Nelson WJ, Weis WI. Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell 1997;90: 871-882.

:5. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell 1998;94: 193-204.
:6. Ybe JA, Brodsky FM, Hofmann K, et al. Clathrin self-assembly is mediated by a tandemly repeated superhelix. Nature 1999;399: 371-375.
:7. Yu RC, Jahn R, Brunger AT. NSF N-terminal domain crystal structure: models of NSF function. Mol Cell 1999;4: 97-107.
:8. May AP, Misura KM, Whiteheart SW, Weis WI. Crystal structure of the aminoterminal domain of N-ethylmaleimide-sensitive fusion protein. Nat Cell Biol 1999;1: 175-182.
:9. Babor SM, Fass D. Crystal structure of the Sec18p N-terminal domain. Proc Natl Acad Sci USA 1999;96: 14759-14764.
:0. Yu RC, Hanson PI, Jahn R, Brunger AT. Structure of the ATP-dependent oligomerization domain of N-ethylmaleimide sensitive factor complexed with ATP. Nat Struct Biol 1998;5: 803-811.
:1. Lenzen CU, Steinmann D, Whiteheart SW, Weis WI. Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. Cell 1998;94: 525-536.
32. Novick P, Zerial M. The diversity of Rab proteins in vesicle transport. Curr Opin Cell Biol 1997;9: 496-504.
33. Dumas JJ, Zhu Z, Connolly JL, Lambright DG. Structural basis of activation and GTP hydrolysis in Rab proteins. Structure 1999;7: 413-423.
34. Ostermeier C, Brunger AT. Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3A. Cell 1999;96: 363-374.
35. Peter M, Chavrier P, Nigg EA, Zerial M. Isoprenylation of rab proteins on structurally distinct cysteine motifs. J Cell Sci 1992;102: 857-865.
36. Nalefski EA, Falke JJ. The C2 domain calcium-binding motif: structural and functional diversity. Protein Sci 1996;5: 2375-2390.
37. Ubach J, Garc¬aJ, Nittler MP, Sudhof TC, Rizo J. Structure of the Janus-faced C2B domain of rabphilin. Nat Cell Biol 1999;1: 106-112.
38. Schalk I, Zeng K, Wu SK, et al. Structure and mutational analysis of Rab GDP dissociation inhibitor. Nature 1996;381: 42-48.
39. Luan P, Heine A, Zeng K, et al. A new functional domain of guanine nucleotide dissociation inhibitor (α-GDI) involved in rab recycling. Traffic 2000;1: 270.


UC
San Francisco
LIBRARY
UNIVERSITY OF CALIFORNIA · UNIVERSITY OF CALIFORNIA · UNIVERSITY OF CALIFORNIA