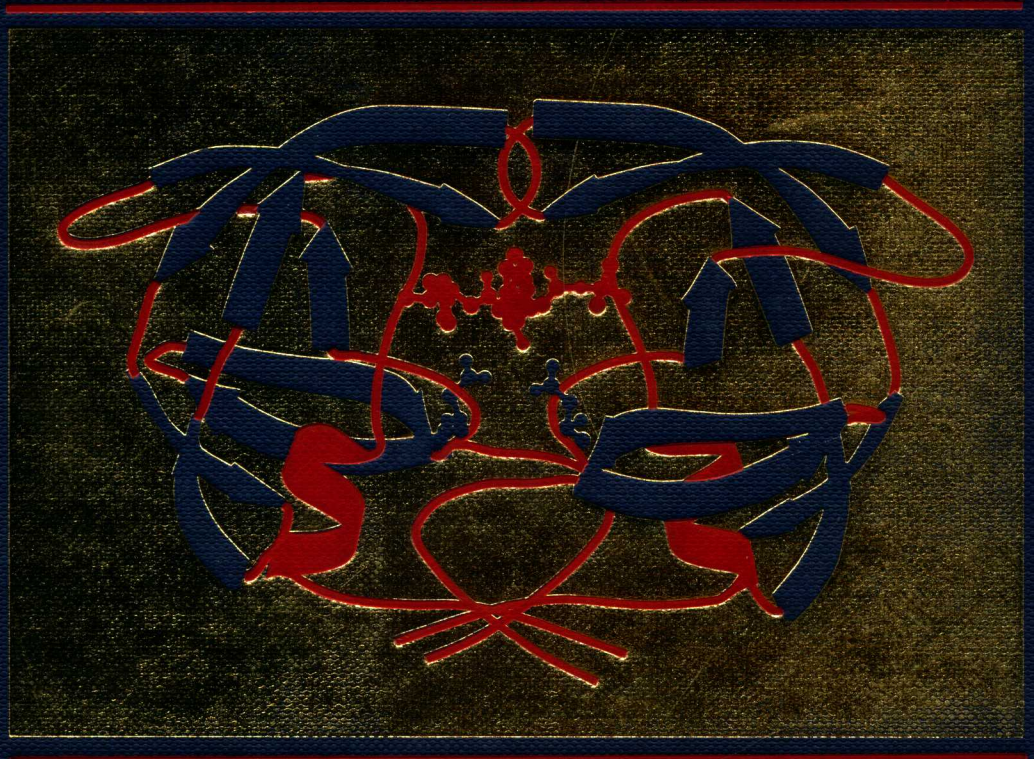
Handbook of Proteolytic Enzymes



EDITED BY
ALAN J. BARRETT NEIL D. RAWLINGS
& J. FRED WOESSNER

ACADEMIC PRESS

Handbook of Proteolytic Enzymes

Edited by

Alan J Barrett
MRC Peptidase Laboratory
The Babraham Institute
Cambridge
UK

Neil D Rawlings
MRC Peptidase Laboratory
The Babraham Institute
Cambridge
UK

J Fred Woessner
Department of Biochemistry and Molecular Biology
University of Miami School of Medicine
Miami
USA

ACADEMIC PRESS

A Harcourt Science and Technology Company

SAN DIEGO SAN FRANCISCO NEW YORK BOSTON
LONDON SYDNEY TOKYO

This book is printed on acid-free paper

Academic Press
A Harcourt Science and Technology Company
Harcourt Place, 32 Jamestown Road, London NW1 7BY, UK
http://www.academicpress.com

Academic Press
A Harcourt Science and Technology Company
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

ISBN 0-12-079370-9

Library of Congress Cataloging-in-Publication Data

Handbook of proteolytic enzymes / edited by Alan J. Barrett, Neil D. Rawlings, F. Fred Woessner, Jr.
p. cm.
Includes index.
ISBN 0-12-079370-9 (alk. paper)
1. Proteolytic enzymes–Handbooks, manuals, etc. I. Barrett, Alan J. II. Rawlings. Neil D. III. Woessner, J.F.
OP609.P78H36 1998
572′.76–dc21 98-23079
CIP

Typeset in 9.5/11pt Times by Laser Words, Madras, India
Printed in Great Britain by The Bath Press, Bath

00 01 02 03 04 05 BP 9 8 7 6 5 4 3 2

Handbook of Proteolytic Enzymes

The cover design is based on the tertiary structure of a complex between the protease of the human immunodeficiency virus and one of the inhibitors that are proving to be valuable drugs in the fight against the virus.

Contributors

Chapter numbers are shown in parenthesis. Full addresses of the contributors are to be found at the end of each chapter.

Hideki Adachi (501)
Osao Adachi (523)
Ibrahim M. Adham (27)
Patrizia Aducci (111)
David A. Agard (82)
Alexey A. Agranovsky (237)
Kyunghye Ahn (363, 364)
Yoshinori Akiyama (516)
Sean M. Amberg (88)
Norma W. Andrews (127)
Stuart M. Arfin (476, 477)
Gérard J. Arlaud (43, 44)
Nathan N. Aronson, Jr (168)
Yasuhisa Asano (146, 147)
Paolo Ascenzi (111)
Marina T. Assakura (440, 446)
David S. Auld (451, 452)
Francesc X. Avilés (455)
William M. Awad, Jr (490)
Lilia M. Babé (24)
S. Paul Bajaj (52)
Susan C. Baker (229)
Marcus D. Ballinger (94)
Alan J. Barrett (6, 86, 255, 267, 343, 371, 372, 531)
Isolda P. Baskova (552)
Karl Bauer (340, 527, 529)
Ulrich Baumann (386)
Wolfgang Baumeister (160, 169, 170)
Ann Beaumont (351, 352)
J. David Becherer (449)
Robert Belas (388)
Teruhiko Beppu (298)
Ernst M. Bergmann (240)
Neil A. Bersani (471)
Ch. Betzel (95)
Robert J. Beynon (351, 352)
Gabriele Bierbaum (109, 259)
Joseph G. Bieth (11, 12, 13, 15)
Jón B. Bjarnason (416, 418, 420, 423, 430, 441, 443)
Roy A. Black (449)
August Böck (561)
Albert Bolhuis (154)
Judith S. Bond (406, 407, 413)
Frank Bordusa (257)
Javier Bordallo (484, 496)
Jacques Bouvier (383)
Barry J. Bowman (289)
Ralph A. Bradshaw (476, 477)
Hans-Peter Braun (469)
Klaus Breddam (79, 80, 132)
Paul J. Brindley (254)
Anne Broadhurst (165)
Keith Brocklehurst (197, 221)
Peter E. Brodelius (279)
Johanna Broer (353)
Dieter Brömme (212)
Sierd Bron (154)
Murray Broom (550)
Leena Bruckner-Tuderman (414, 415)
Barbara A. Burleigh (127)
Michael J. Butler (114, 140, 346, 382)
David J. Buttle (188, 189, 190, 196)
Sandrine Cadel (348)
Javier Caldentey (557)
Daniel R. Caprioglio (344)
Juan Carbonell (199)
Ruth E. Carter (492)
Francis J. Castellino (59)
Joseph J. Catanese (421)
George H. Caughey (18, 20)
Niamh X. Cawley (304)
Tim E. Cawston (389)
Juan José Cazzulo (203)
Manuel Cercós (199)
Trinad Chakraborty (353)
Julie Chao (29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 61)
Marie-Pierre Chapot-Chartier (216)
Jinq-May Chen (371)
Valérie Chesneau (468)
Bernard Chevrier (491)
Jean-François Chich (136)
Kwan Yong Choi (244)
Michel Chrétien (118, 119, 120, 122, 123)
Chin Ha Chung (171, 178, 184)
Christine Clamagirand (540, 541)
Steven Clarke (502)
Paul P. Cleary (100)
Philippe Clauziat (268)
D. Coates (360)
Paul Cohen (348, 468, 540, 541)
D. Collen (58)
Ivan E. Collier (402)
Robert W. Colman (50)
Gregory E. Conner (277)
Andreas Conzelmann (256)
Tim Coolbear (99)
Graham H. Coombs (204)
Antony Cooper (569)
Jonathan B. Cooper (296)
Max. D. Cooper (339)
Maria C. Cordeiro (279)
Pierre Corvol (359)
Joseph T. Coyle (492)
Charles S. Craik (3, 7)
John W.M. Creemers (117)

Makoto Kaneda (110)
Sung Gyun Kang (4)
Mary Kania (169, 170)
Shun-ichiro Kawabata (66, 67, 68, 69, 116, 425, 427, 433)
John Kay (164, 166, 275)
Kenneth C. Keiler (157)
Robert M. Kelly (558)
Susanna R. Keller (342)
Michael A. Kerr (46, 47)
Jukka Kervinen (278)
Efrat Kessler (357, 411, 506, 507)
Mogens Kilian (87)
Do-Hyung Kim (244)
In Seop Kim (4)
Yukio Kimura (535)
Heidrun Kirschke (210, 211, 213)
Ana Kitazono (139)
Naomi Kitamura (63)
Jürgen R. Klein (345, 376, 486, 487)
Hideyuki Kobayashi (299)
Jan Kok (365)
Pappachan E. Kolattukudy (105, 514)
Harald Kolmar (83)
Jan Konvalinka (320)
Alexander A. Kortt (101)
Hans-Georg Kräusslich (320)
Lawrence F. Kress (421)
Takashi Kumazaki (361)
Kotoku Kurachi (60)
Imelda J. Lambkin (179)
Gayle K. Lamppa (470)
Bernard F. Le Bonniec (55)
Kye Joon Lee (4)
Jonathan Leis (313)
Jo Ann C. Leong (564)
Thierry Lepage (400)
Stephen H. Leppla (511)
Ivan A. Lessard (503)
Guy Lhomond (400)
H.R. Lijnen (58)
Shie-Jea Lin (103)
Xinli Lin (332)
William N. Lipscomb (473)
John W. Little (152)
Mark O. Lively (156)
Y. Peng Loh (304, 306)
George P. Lomonossoff (242)
Yvan Looze (191)
Deshun Lu (14)
Andrei N. Lupas (169, 170, 172)
Vivian L. MacKay (303)
Richard E. Mains (121)
Kauko K. Mäkinen (354)
Fajga Mandelbaum (440, 446)
Walter F. Mangel (176)
James M. Manning (130)
Francis S. Markland, Jr (70, 71, 72, 417, 442)
Takeharu Masaki (85)
Sususmu Maruyama (532)
Carla Mason (564)
José Matos (481)
Lynn M. Matrisian (394)
Hiroshi Matsuzawa (103)
Yoshihiro Matsuda (384)
Michael R. Maurizi (177)
Majambu Mbikay (120)
Dewey G. McCafferty (503)
J. Ken McDonald (138, 182, 183, 266, 530, 546, 547, 548)
Gerard McGeehan (398)
Ultan G. McKeon (269)
David McMillan (166)
Robert P. Mecham (515)
Darshini P. Mehta (207)
Alan Mellors (505)
Roger G. Melton (483)
Armelle Ménard (315)
Robert Ménard (187)
Luis Menéndez-Arias (312, 317, 318)
Gregor Meyers (245)
Susan Michaelis (366)
Wojciech P. Mielicki (262)
Igor Mierau (365)
Charles G. Miller (375, 485, 544, 562)
John Mills (164, 165, 166)
Michel-Yves Mistou (216)
Yoshio Misumi (128, 129)
Shigehiko Mizutani (341)
William L. Mock (478)
Véronique Monnet (378, 481)
Cesare Montecucco (509, 510)
Kazuyuki Morihara (387)
James H. Morrissey (53)
John S. Mort (209)
Uffe H. Mortensen (132)
Hidemasa Motoshima (498, 499, 504)
Jerry C. Mottram (204, 205)
Kazuo Murakami (284, 285)
Kimiko Murakami-Murofushi (293)
Sawao Murao (290, 328, 329)
Gillian Murphy (401)
Tatsushi Muta (66, 67, 68, 69)
Hideaki Nagase (393)
Rajesh R. Naik (104)
Hiroshi Nakazato (341)
Yukio Nakamura (284, 285)
Toshiaki Nikai (426, 428, 431, 435)
Makoto Nishiyama (298)
Gordon W. Niven (495)
Shigemi Norioka (5)
Michael J. North (205)
Donald L. Nuss (228)
Richard J. Obiso, Jr (403)
Brendan O'Connor (269)
Kohei Oda (326, 330)
Douglas H. Ohlendorf (81)
Dennis E. Ohman (357, 507)
Junji Ohnishi (62)
Kenji Okuda (367)
Kjeld Olesen (132)
Stephen Oroszlan (312, 317, 318)
Michael Ovadia (438)
Jeremy C.L. Packer (155)

Mark J.I. Paine (429, 432)
Salomé M. Pais (279)
Himadri B. Pakrasi (158)
Kirk L. Parkin (217)
Pyong Woo Park (515)
Sudhir Paul (180)
Duanqing Pei (397)
Alan D. Pemberton (21)
Iva Pichová (316)
Michael Pieper (390)
Lothaire Pinck (243)
Euclides M.V. Pires (280, 281)
Hubert Pirkle (70, 71)
Henry C. Pitot (261)
Andrew G. Plaut (508)
Jan Potempa (224, 258, 538)
Knud Poulsen (87)
Annik Prat (468)
David G. Pritchard (379)
Alexey V. Pshezhetsky (133)
Mohammad Abul Qasim (77, 78)
Stephen D. Rader (82)
Robert R. Rando (260)
Oscar D. Ratnoff (49)
Neil D. Rawlings (563)
Karen E. Reed (567)
Antonia P. Reichl (440, 446)
Julian R. Reid (99)
Kenneth B.M. Reid (45)
S. James Remington (134)
Wolfgang J. Rettig (131)
Mohamed Rholam (540)
Charles M. Rice (88, 89, 567)
Todd W. Ridky (313)
Howard Riezman (256)
Lori A. Rinckel (321)
Alison Ritchie (165)
Janine Robert-Baudouy (268)
R. Michael Roberts (307, 308)
Michael Robinson (165)
Adela Rodriguez-Romero (198)
John C. Rogers (200)
Markus Rohrwild (171)
Philip J. Rosenthal (201)
J. Rosing (73)
Cesare Rossi (245)
Richard A. Roth (466, 467)
Andrew D. Rowan (192, 193, 194, 195)
Mikhail N. Rozanov (232)
Tillmann Rümenapf (565)
Martin D. Ryan (566)
Thomas J. Ryan (264)
J. Evan Sadler (14)
W. Saenger (106)
Hans-Georg Sahl (109, 259)
Fumio Sakiyama (5, 85)
Guy S. Salvesen (16,22,23)
Eladio F. Sanchez (424)
QingXiang Amy Sang (392)
Krishnan Sankaran (333)
Yoshikiyo Sasagawa (518)
Kimiyuki Satoh (159)
Robert T. Sauer (157)
Jennifer J. Schiller (229)
Manfred Schlösser (27)
Patrick M. Schlievert (81)
Brian F. Schmidt (24)
Udo K. Schmitz (469)
Henning Scholze (202)
Erika Seemüller (172)
Nabil G. Seidah (118, 119, 120, 122, 123)
Motoharu Seiki (399)
Giorgio Semenza (338)
Robert M. Senior (395)
Solange M.T. Serrano (74, 440, 446)
Peter Setlow (559)
M. Shanks (242)
Stephen D. Shapiro (395)
Lei Shen (56)
Roger F. Sherwood (483)
Hitoshi Shimoi (84)
Yun-Bo Shi (392)
William H. Simmons (135, 479, 480)
Robert B. Sim (48)
Alison Singleton (196)
Tatiana D. Sirakova (105, 514)
Tim Skern (239, 241)
Randal Skidgel (459, 460)
Jeffrey Slack (206)
Mark J. Smyth (25)
Eric J. Snijder (92, 234, 235, 236)
Hiroyuki Sorimachi (218, 219, 220)
Eric B. Springman (454)
Robert Stark (565)
James L. Steele (217, 545)
Henning R. Stennicke (79, 80)
Johan Stenflo (54)
Valentin M. Stepanov (456)
Erwin E. Sterchi (408)
Frank Steven (75)
Kenneth J. Stevenson (108)
David J. Stewart (101)
Raymond J. St. Leger (107)
Walter Stöcker (405)
Stuart R. Stone (55)
Andrew C. Storer (187)
Ellen G. Strauss (230)
James H. Strauss (230)
Norbert Sträter (473)
Paz Suárez-Rendueles (484, 496)
Kyoko Suefuji (103)
Fumiaki Suzuki (284, 285)
Koichi Suzuki (218, 219, 220)
Eugene D. Sverdlov (552)
Stephen Swenson (72, 417, 442)
Pal B. Szecsi (273, 274)
László Szilágyi (8, 9)
Muhamed-Kheir Taha (560)
Kenji Takahashi (291, 327)
Takayuki Takahashi (62)
Koji Takio (537)
Amy Tam (366)

Tomohiro Tamura (160)
Fulong Tan (137)
Nget Hong Tan (434)
Jordan Tang (272, 276, 332)
Martha M. Tanizaki (439)
Nayuki Taniguchi (174)
Egbert Tannich (202)
Anthony L. Tarentino (412)
Peter J. Tatnell (275)
Hiroki Tatsumi (513)
Norbert Tautz (91)
Markus F. Templin (143, 555, 556)
Gonzales Thierry (268)
Heinz-Jürgen Thiel (91, 245, 565)
Nicole M. Thielens (43, 44)
Nancy A. Thornberry (248, 249)
Peter E. Thorsness (517)
Natalie J. Tigue (164)
Harold Tjalsma (154)
Diana L. Toledo (176)
Birgitta Tomkinson (113)
Akito Tomomura (10)
Fiorella Tonello (510)
H.S. Toogood (96, 97, 331)
Paolo Tortora (112, 494)
József Tözsér (312, 317, 318)
James Travis (224, 258, 538)
Harald Tschesche (390)
Christopher A. Tsu (7)
Masafumi Tsujimoto (341)
Daisuke Tsuru (126)
Anthony T. Tu (419, 422)
Boris Turk (214)
Vito Turk (214)
Anthony J. Turner (337, 362)
Veli-Jukka Uitto (102)
Dirk Ullmann (257)
Wim J.M. Van de Ven (117)
Marco Vanoni (112, 494)
Harold E. Van Wart (368)
Gregory M. Vath (81)
Josep Vendrell (455)
Gerard Venema (154)
István Venekei (8)
Jeanmarie Verchot (90, 227)
John E. Volanakis (42)
Fred W. Wagner (436)
Kenneth W. Walker (476, 477)
Stephen J. Walker (156)
Christopher T. Walsh (503)
Peter N. Walsh (51)
Jiyang Wang (339)
Alfred L.M. Wassenaar (92, 235, 236)
Aaron B. Watts (197, 221)
Joseph M. Weber (251)
Daniel R. Webster (549)
Jürgen Wehland (353)
Stephen J. Weiss (397)
James A. Wells (94)
Judith M. White (447)
Gene L. Wijffels (208)
Sherwin Wilk (524, 533, 551)
Jean-Marc Wilkin (142, 148)
Keith D. Wilkinson (222, 223)
Tracy D. Wilkins (403)
Tracy A. Williams (359)
Jakob R. Winther (283)
J. Fred Woessner (149, 282, 358, 381, 396, 521, 525, 526)
Dieter H. Wolf (488)
Tyra G. Wolfsberg (447)
David O. Wood (474)
Gerry D. Wright (554)
Sancai Xie (307, 308)
Kenjiro Yamagami (409, 410)
Yoshio Yamakawa (437, 445)
Koichiro Yamamoto (356)
Noriko Yasuda (535)
Shigeki Yasumasu (409, 410)
Toshimasa Yasuhara (489)
Soon Ji Yoo (171)
Tadashi Yoshimoto (125, 139)
Ludmila L. Zavalova (552)
Nada Zein (568)

Preface

Ever since the discovery of pepsin in the late eighteenth century there has been continuing investigation of the chemistry and activities of the proteolytic enzymes, or peptidases. But recent years have seen a remarkable acceleration of the pace of this research, fuelled by numerous practical applications in biotechnology, and the realization that the peptidases are major therapeutic targets. A striking example of the link between peptidases and drug design is the basis of the cover design of the *Hankbook*. This depicts the structure of retropepsin, the essential processing peptidase of the human immunodeficiency virus, in complex with a potent inhibitor of the class that are proving to be effective drugs against the virus (drawn from Brookhaven Protein Databank entry 1SBG).

The many ways in which proteolytic enzymes impinge on the health and welfare of mankind have made it essential for biological scientists in many fields to have ready access to data on peptidases, but the sheer numbers of these enzymes pose a problem. Analysis of complete sequences of several genomes has shown about 2% of all gene products to be peptidases, indicating that this is one of the larger functional groups of proteins. The great number of known peptidases creates many practical problems for those needing to work with them. For example, it is difficult to know how one peptidase can be distinguished from another and referred to unambiguously, and how a scientist can tell when he/she has discovered a novel peptidase. It is precisely this kind of question that the present *Handbook* is designed to answer. The rapid expansion of the field of proteolytic enzymes is bringing into the field new investigators who will find a comprehensive reference book a particularly valuable resource.

The present volume has grown out of a long-standing interest of the editors in preparing readily accessible compilations of data on peptidases. Its genesis can be traced back to the 2nd International Symposium on Intracellular Protein Catabolism held in Ljubljana, Slovenia, in 1975. Two of the editors (AJB and JFW) recognized that there was considerable confusion in the field concerning the various intracellular proteolytic enzymes and how they might be distinguished. A chapter was prepared with the collaboration of I. Kregar and V. Turk entitled 'Present knowledge of proteolytic enzymes and their inhibitors', and published in *Intracellular Protein Catabolism II* (V. Turk & N. Marks eds, Plenum Press, New York, 1977). A grand total of 23 intracellular enzymes were tabulated! This modest beginning was followed by *Mammalian Proteases: a Glossary and Bibliography* by AJB with J.K. McDonald in two volumes (Academic Press, London, 1980, 1986). Here, 173 peptidases were described in concise summaries supplemented by extensive bibliographies.

These two books have long been out of print and the number of known peptidases has risen steadily, so that the total across all kinds of organisms now exceeds 500. This growth has been reflected in a short doubling-time of the literature, so that publications on proteolytic enzymes now approach 8000 a year. Those responsible for subsection 3.4 of the EC List (the enzyme nomenclature of the International Union of Biochemistry and Molecular Biology) that deals with peptidases have been striving to keep up with this flood and have succeeded in including nearly 300 peptidases to date. Although the EC List has considerable value, notably as a source of unambiguous approved names, it has significant limitations. Only enzymes that have been subject to rigorous enzymological characterization can be included, and the endopeptidases are allocated to just four mechanistic classes, in which they are listed in random order.

During recent years, two of the editors (AJB, NDR) have been developing an alternative approach to the classification of peptidases that takes advantage of the new wealth of structural information. In this scheme the enzymes are allocated to clans and families within each major mechanistic class. The system is based on genetic relationships among the enzymes and has been presented in two recent volumes of *Methods in Enzymology* (244 & 248, Academic Press, San Diego, 1994, 1995). But this treatment was not designed to be comprehensive, and is not in convenient form for rapid reference. In contrast, it is the hope of the editors of the present *Handbook* to provide a convenient classification of all the known peptidases that meet minimal criteria for inclusion (see the Introduction). This *Handbook* provides a ready reference to the 500 or so peptidases known to date, but it also provides a framework for the addition of the many more proteolytic enzymes that may be expected to emerge during the coming years of intensive genome research. It is the hope of the editors that this work will be readily accessible to the multitude of workers in the field, so that it is truly a 'hand' book. To this end, we have arranged with the Publishers that a CD-ROM be made available in addition to the printed bound volume.

The chapters on individual peptidases have been contributed by over 500 expert authors who have worked hard and well to describe their favorite peptidases in the concise and strict format required for the *Handbook*. For practical reasons, the editors have had to assume full responsibility for the final editing and proofreading of each chapter, and sometimes have made substantive changes without providing the authors with much opportunity for rebuttal. It cannot be expected that every author will be in complete agreement with the way in which his or her work is presented, but we trust that they will understand that any apparent slights were unintended and are merely the consequences of our effort to produce a systematic and consistent overview of the peptidases. We thank them all for their good work.

The editors thank their secretaries, Desi O'Rourke and Michelle Gonzalez, for their skilled assistance in the task of soliciting, collecting and collating the many manuscripts for this project. We are particularly grateful for the assistance of the staff at Academic Press, London, including Susan Lord, Sarah Stafford, Roopa Baliga and Emma Parkinson, and to Harriet Stewart-Jones for meticulous copy-editing and Ian Ross for expert proof-reading. Finally, we express our heartfelt, personal appreciation to Jinq-May Chen, Cheow-Yong Rawlings and Nina Woessner, who generously tolerated our preoccupation with the *Handbook* over much of the past two years.

AJB, NDR, JFW
February 1998

Introduction

Terminology

The scientists who work on the large and important group of enzymes that hydrolyze peptide bonds currently allow themselves a great deal of freedom in the terms they use for their objects of study. The effect of this is that there are commonly several names in use for much the same thing. This is seen when one looks for a collective word for all of these enzymes. They are commonly termed ***proteases***, ***proteinases*** and ***peptidases***, as well as ***proteolytic enzymes***. Historically, these terms had slightly different meanings (Barrett & McDonald, 1986), but these are now forgotten by many. The editors felt that ***proteolytic enzymes*** was perhaps the most generally understood term in the current usage and therefore adopted this for the title of the *Handbook*. The reader should note, however, that even this is not unambiguous, since many of the enzymes that hydrolyze peptide bonds (and are included in the *Handbook*) do not act on proteins directly.

In our editing of the present volume, we wished to encourage movement towards more rational and systematic terminology in the study of proteolytic enzymes. Fortunately, sound and authoritative recommendations are available in the form of the EC List. This is still named after the Enzyme Commission that compiled the first editions (Webb, 1993), but has now been curated by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) for many years. The EC List was last printed in full as *Enzyme Nomenclature 1992* (NC-IUBMB, 1992), but the part dealing with peptidases has subsequently been amended by regular supplements and can be found in its revised form on the World Wide Web (WWW) at http://www.chem.qmw.ac.uk/iubmb/enzyme/index.html. The EC List recommends the term ***peptidase*** as the general term for all the enzymes that hydrolyze peptide bonds, and the present editors support this recommendation. Such terms as protease and proteinase will be found to occur frequently in the *Handbook*, but peptidase is the term that is preferred by the editors and is used in the editorial chapters that introduce the major groups of these enzymes. Most peptidases are either ***exopeptidases*** cleaving one or a few amino acids from the N- or C-terminus, or ***endopeptidases*** that act internally in polypeptide chains. The EC List also provides terms for subtypes of exopeptidases and endopeptidases. The exopeptidases that act at a free N-terminus liberate a single amino acid residue (***aminopeptidases***) or a dipeptide or a tripeptide (***dipeptidyl-peptidases*** and ***tripeptidyl-peptidases***). Those acting at a free C-terminus liberate a single residue (***carboxypeptidases***) or a dipeptide (***peptidyl-dipeptidases***). Other exopeptidases are specific for dipeptides (***dipeptidases***) or remove terminal residues that are substituted, cyclized or linked by isopeptide bonds (peptide linkages other than those of α-carboxyl to α-amino groups) (***omega peptidases***). The endopeptidases are divided on the basis of catalytic mechanism into ***serine endopeptidases***, ***cysteine endopeptidases***, ***aspartic endopeptidases*** and ***metalloendopeptidases***. The term ***oligopeptidase*** is used to refer to endopeptidases that act optimally on substrates smaller than proteins.

The muddled situation that exists in the language currently used in reference to peptidases is seen at its worst in the naming of the individual enzymes. It is all too common for scientists working on a single enzyme in different laboratories to use quite different names for it, even in their published work. Such a liberal practice must come as a surprise to scientists from other disciplines – it is as if biologists were to use their own local names for organisms rather then the scientific binomials!

Two reasons for the present relaxed attitude to the naming of peptidases are not hard to see. One is that the number of known peptidases is commonly perceived as small, so that confusion seems unlikely. And secondly, the naming of peptidases is notoriously difficult, so that it has never been possible to derive simple, objective rules for forming names that can be applied by anyone, yielding the same result. A consequence of this is that all of the names may tend to look arbitrary, none being better than another.

Neither of these reasons is adequate to justify the present situation. There are indeed fewer peptidases so far recorded than living organisms, say, but the reader of the present volume will soon realize that they are quite numerous. Even if each of the peptidases known today were referred to by only one name, there would be quite enough to tax any but the complete specialist. And although it is indeed difficult to name peptidases, there is no doubt that there are good names and bad ones. Some of the names in common use are very poorly suited to the functions that a name should serve, and there is every reason to encourage general use of one of the better names. The situation in which most peptidases have several names in common use can only be an obstacle to the progress of this field of study. Vast amounts of data about these enzymes are now being published in printed form and in computer databases, but the accessing of that information is made more difficult and less efficient when multiple names are used, especially when some of the names are ambiguous, being applied to several different enzymes. And as the number of known peptidases rises into the thousands, multiple names must be a recipe for chaos.

In assembling the present volume, we have attempted to encourage the rationalization of terminology that is needed in the field, and again we support the recommendations of the EC List. Although we have not insisted that only the names recommended in the EC List be used, we have encouraged their use, and have inserted cross-references to them where necessary. We have also invited each author of a chapter in the *Handbook* to start with a section on the history and naming of the enzyme.

Scope of the Handbook

The *Handbook* contains information on almost every known peptidase. Amongst these there are formal entries on all the peptidases that are included in the current EC List, but it should be noted that a few of these now seem unlikely to represent distinct enzymes and may disappear in later revisions of the List (as is explained in the text). There are also chapters on many peptidases that have not yet been included in the EC List, as well as accounts of some gene products that are shown by their amino acid sequences to be close relatives of known peptidases but have not yet been demonstrated directly to be peptidases in their own right. Also included for their interest are a few entities that show peptidase-like activity, but would not normally be thought of as peptidases proper. Amongst these are a number of self-processing proteins, an antibody molecule, and a histone-splitting chromoprotein. By use of these broad criteria, we assembled a list of over 500 peptidases for inclusion in the *Handbook*, and invited a similar number of expert authors to write about them.

Classification

A fundamental aspect of the *Handbook* is the way in which it is organized. The peptidases are grouped on the basis of primary and tertiary structures into families and clans, and these are further grouped by catalytic mechanism. The classification used in the *Handbook* is therefore very different from that adopted in the EC List (see above), where the peptidases, which form subclass 3.4 of all enzymes, are divided into 13 sub-subclasses. The sub-subclasses are not further divided, and the peptidases are listed in arbitrary order within each of them. The molecular structures and evolutionary relationships that are of key importance in the present *Handbook* are not taken into account in the EC classification.

The system of classification that we have employed in the *Handbook* is one that was introduced by Rawlings & Barrett (1993), and has subsequently been further developed by these authors. In this scheme, a ***family*** of peptidases is a group in which every member shows a statistically significant relationship in amino acid sequence to at least one other member of the family in the part of the molecule that is responsible for peptidase activity. Strict statistical criteria are applied so that we can be confident that any two peptidases that are placed in the same family have evolved from a common ancestor and thus are homologous by the definition of Reeck *et al*. (1987). The restriction of the comparison to the catalytically active part of the molecule is an important one, since many peptidases are chimeric proteins containing additional, nonpeptidase domains that are shared with other groups of proteins. We do not consider the relationships of these domains to be directly relevant to the classification of the peptidase, and for this reason, the peptidase families do not correspond closely to families of proteins recognized in other systems, such as the PIR protein sequence database (Barker *et al*., 1990). A few of the peptidase families contain two or more rather distinct groups of peptidases (shown by a deep divergence in the dendrogram: see below), and for these, subfamilies are recognized.

Each peptidase family is named with a letter denoting the catalytic type (S, T, C, A, M or U, for serine, threonine, cysteine, aspartic, metallo- or unknown), followed by an arbitrarily assigned number. When a family disappears, usually because it is merged with another, the family name is not reused, and for this reason there are interruptions in the numerical sequences of families that are of no current significance.

Clan is the term used to describe a group of families the members of which have evolved from a single ancestral protein, but have diverged so far that we can no longer prove their relationship by comparison of the primary structures (Rawlings & Barrett, 1993; Barrett & Rawlings, 1995). The clearest kind of evidence for clan-level relationship between families is similarity in three-dimensional structures, but the arrangement of catalytic residues in the polypeptide chains and limited similarities in amino acid sequence around the catalytic amino acids can also be revealing. The name of a clan is formed from the letter for the catalytic type (as for families) followed by an arbitrary second capital letter. If a clan disappears, the name is not reused. Not all families can yet be assigned to clans. When a formal clan assignment is needed for these it is given as SX (for a serine peptidase family), or CX, AX, etc.

Catalytic type depends upon the chemical nature of the groups responsible for catalysis in a way that can be traced back to Hartley (1960) and was adapted for the EC List in 1972. The major catalytic types are Serine, Cysteine, Aspartic, Metallo and as yet Unclassified, and these initial letters can be seen on the right-hand margin of the pages of the present volume, to provide easy access to the relevant sections of the book. Generally, the use of these catalytic types as the top level of the hierarchical classification of peptidases works well, but there are a few anomalies. Most notably, the serine peptidases of clan SA undoubtedly share a common origin with the cysteine peptidases of clan CB (see Chapter 238). It should also be mentioned here that for practical convenience a few enzymes such as the proteasome in which threonine rather than serine forms the nucleophile of catalysis have been placed in the 'serine' section, and the 'aspartic' section contains several acid-acting endopeptidases that may possibly contain catalytic glutamic residues.

The *Handbook* contains many links to the **MEROPS** database. This is a WWW database of information on peptidases (curated by NDR) set up as a means to respond to the constant state of development of the system of families and clans of peptidases that is driven by the flow of new data for primary and tertiary structures. New releases of the database appear as needed, and we hope that readers will obtain updating information from it after publication of the book. The CD-ROM version of the *Handbook* contains a hypertext link to the MEROPS database on the WWW from each chapter. Anyone wishing to browse the MEROPS file for a chapter without the CD can do so by use of a URL in the form http://www.bi.bbsrc.ac.uk/Merops/HPE[chapter].htm, in which [chapter] is the chapter number in the *Handbook* (without the brackets). Each peptidase has a unique identifier (ID) in MEROPS. The ID is constructed of two parts, the family name

(e.g. 'S01') and an arbitrary three-digit number for the individual peptidase within the family, the two parts being separated by a decimal point. The minority of peptidases for which sequences are not yet available cannot be assigned to families, and for these, provisional MEROPS IDs are formed as follows. As usual, the first character is a letter for the catalytic type, but this is always followed by a figure 9, and the third character is a letter indicating the type of peptidase activity: A, aminopeptidase; B, dipeptidase; C, dipeptidyl- or tripeptidyl-peptidase; D, peptidyl-dipeptidase; E, carboxypeptidase; F, omega peptidase, and G, endopeptidase. The second part of the ID is formed as usual. The ID also can be used for direct access to the relevant data file in MEROPS. The URL is exactly as described for the chapter links, but HPE[chapter] is replaced by the seven-character MEROPS ID modified by changing the decimal point to 'p'. Thus, the file for acrosin (ID S01.223) is http://www.bi.bbsrc.ac.uk/Merops/S01p223.htm.

Editors' Introductory Chapters

Since the *Handbook* is organized according to the hierarchical classification: catalytic type, clan and family, families do not necessarily appear in numerical order. Text on any given family can be located by use of the subject index, but additionally, each sequence of chapters on peptidases from a single clan or large family is preceded by an introduction from the editors. The introductory chapter typically starts with a databanks table (see below) that gives an overview of the families within the clan and their constituent peptidases. Data are not repeated from the individual chapters, but cross-referenced. Sequence database accession numbers are, however, given for peptidases or putative peptidases that are not the subjects of separate chapters, including peptidase homologs that are known as sequences but have not yet been characterized biochemically.

Alignments of amino acid sequences have been prepared for most families. A preliminary alignment was constructed by use of the PILEUP program of the GCG package (Genetics Computer Group, 1994). This was improved manually, and when a structural alignment based on three-dimensional structure was available in the literature, the preliminary alignment was edited to match this. For brevity, the alignments are not presented in full, but are shown in one of two ways. For some families, the chimeric nature of the proteins is of interest, and 'domain diagrams' have been produced in which the arrangement of domains of various types, location of active-site residues, and disulfide loops are depicted schematically. On the CD-ROM only, there are also partial alignment diagrams in which key regions of the sequences are picked out, typically to show the conserved sequence motifs around the functionally important amino acid residues.

Dendrograms are shown for many families. These are based on the alignment of the amino acid sequences of the peptidase units for the family. That is to say, the alignment as described above was trimmed to exclude all portions of the sequences that are not part of the peptidase unit, including in some cases inserts within the peptidase unit. For some very large families, only a selection of peptidases was included. A difference matrix of percentage sequence identities was computed from the alignment and this was used to compute the tree. The Fitch–Margoliash algorithm (Fitch & Margoliash, 1967) with contemporary tips was used as implemented in the KITSCH program of the PHYLIP package (Felsenstein, 1989). The trees are used here to give a graphic depiction of the distances between the structures of peptidases within a family and are not intended to represent an accurate reconstruction of the evolutionary history of the family; statistical sampling has therefore not been performed. The trees are drawn so that the longest branches are uppermost. The x-axis is calibrated in PAM (percentage accepted mutations) calculated from the percentage identity at the amino acid level as described by Dayhoff *et al*. (1978).

At least one Richardson diagram depicting the protein tertiary structure has been constructed for every peptidase family for which a structure has been published. Each of these (in the style of Richardson, 1985) has been constructed from a PDB entry, and where possible structures have been selected that include a bound small molecule inhibitor or substrate. The diagrams were constructed by use of a series of programs. First, the RASMOL program (Sayle & Milner-White, 1995) was used to orient the molecule so that all the active-site residues are visible (and to place molecules with similar structures in the same orientation). For the book, the MOLSCRIPT program (Kraulis, 1991) was used to generate the images in grayscale. For the CD-ROM, the MOLSCRIPT program was then used to generate an input file for the RENDER program (Bacon & Anderson, 1988) of the RASTER3D package (Merritt & Murphy, 1994). We have shown helices as (red) coils, sheets as (green) arrows, and coils and turns as (cyan) wires. Active-site residues and bound inhibitors and substrates are shown in ball-and-stick representation, and metal ions as Corey–Pauling–Koltun spheres. In some cases, the PDB entries have been edited so that only one subunit of a multimeric structure is shown.

Chapters on Individual Peptidases

A series of chapters on related peptidases is typically ordered to bring together enzymes that are shown as being closely related in the dendrogram for the family. Each chapter on an individual peptidase starts with a databanks table compiled by NDR. The first line of this gives the peptidase classification, i.e. the name of the clan and family and the MEROPS identifier (ID) of the peptidase. The peptidase classification line is followed by the EC number, the American Tissue Culture Collection (ATCC) clone numbers, and the Chemical Abstracts Service (CAS) Registry number.

The body of the databanks table contains the primary sequence database accession numbers, arranged alphabetically by scientific binomial of the organisms from which sequences are known. The organism name is printed for each distinct gene product. Accession numbers have been collected from the SwissProt protein sequence database (SW) (Bairoch & Apweiler, 1996) (http://expasy.hcuge.ch/sprot/sprot-top.html), the PIR database (Barker *et al*., 1990) (http://www-nbrf.georgetown.edu/pir/), the EMBL nucleic acid sequence database (Kahn & Cameron, 1990)

(http://www.ebi.ac.uk/ebi_docs/embl_db/embl_db.html), and occasionally the GenPept database. Of these databases, only SwissProt is nonredundant. Accession numbers in the EMBL database are the same as those in the GenBank (Benson *et al*., 1998) (http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html) and DDBJ (http://www.ddbj.nig.ac.jp/) databases. The entries for the EMBL database have been arranged in two columns for clarity. One column is headed 'cDNA' and refers to EMBL entries containing only the sequence of the peptidase mRNA (or the gene in the case of a peptidase from a bacterium). The second column, headed 'genomic' includes complete genes, exons, introns, gene promoters, and portions of genomes. The accession number may be followed by a brief comment describing the database entry.

A Brookhaven Protein Data Bank table is provided for those peptidases for which the coordinates of a three-dimensional structure have been deposited in the Protein Data Bank (PDB) (Bernstein *et al*., 1977) (http://www.pdb.bnl.gov/). The entries in this table are arranged by scientific binomial and then alphabetical order of the PDB accession number. All structures relating to the peptidase in question are included in the table together with the computed resolution of each structure (in Angstrom units) and a brief description.

The remainder of each chapter has been prepared by the author(s) credited at the end of the chapter, and the chapters have a uniform structure. Following the databanks table there is the author's text describing the Name and History, Activity and Specificity, Structural Chemistry, Preparation, Biological Aspects, Distinguishing Features, Related Peptidases, Further Reading and References.

In the descriptions of the specificity of the peptidases, the symbol '┼' has frequently been used to mark the bond that is hydrolyzed, the scissile bond, in the formulae of substrate molecules. Underlying the visible text on the CD-ROM there is encoded an invisible version of the same information that can be searched (see below).

The terminology used in describing the specificity of peptidases depends on a model in which the catalytic site is considered to be flanked on one or both sides by specificity subsites, each able to accommodate the side chain of a single amino acid residue. These sites are numbered from the catalytic site, S1...Sn towards the N-terminus of the substrate, and S1′...Sn′ towards the C-terminus. The residues they accommodate are numbered P1...Pn, and P1′...Pn′, respectively, as follows (the catalytic site of the enzyme being marked '*'):

Substrate:	–	P3	–	P2	–	P1	┼	P1′	–	P2′	–	P3′	–
Enzyme:	–	S3	–	S2	–	S1	*	S1′	–	S2′	–	S3′	–

This scheme is essentially as described by Berger & Schechter (1970), but slightly simplified in that we print the numbers of subsites and the side chains they accommodate on the line rather than subscript, and place the prime signs after the numbers.

Appendices

The sequence of 569 chapters is followed by two appendices: Appendix 1 provides the common English name (if any) and the type of organism for each of the species from which a peptidase sequence is listed in the book, and Appendix 2 provides address details for many of the suppliers of materials mentioned.

CD-ROM

The CD-ROM version of the *Handbook* contains the full text with a number of enhancements. There are many hypertext links, including links through the WWW to the MEROPS database. The figures are in color, and the text is fully searchable, so that any term not in the printed Index should be easily located. These may include authors' names, words in titles or journal names for the more than 15 000 references cited throughout the text. Also, there is a specially designed search function for the scissile bonds in substrate structures that allows the reader to locate a peptidase with a given specificity. There are facilities for the insertion of bookmarks, and for copying and pasting the text. All of this is described more fully in the leaflet that accompanies the CD.

Feedback to the Editors

The editors would be pleased to hear of any necessary corrections, additions, or information on new enzymes. Whether or not there will be a further edition of the *Handbook*, such information may be used to correct and update the MEROPS database and the EC List. Please address your comments to any of the editors at his email address.

References

Bacon, D. & Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220.

Bairoch, A. & Apweiler, R. (1996) The SWISS-PROT protein-sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* **24**, 21–25.

Barker, W.C., George, D.G. & Hunt, L.T. (1990) Protein sequence database. *Methods Enzymol.* **183**, 31–49.

Barrett, A.J. & McDonald, J.K. (1986) Nomenclature: protease, proteinase and peptidase. *Biochem. J.* **237**, 935.

Barrett, A.J. & Rawlings, N.D. (1995) Families and clans of serine peptidases. *Arch. Biochem. Biophys.* **318**, 247–250.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. & Ouelette, B.F. (1998) GenBank. *Nucleic Acids Res.* **26**, 1–7.

Berger, A. & Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos. Trans. R. Soc. Lond. [Biol.]* **257**, 249–264.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. (1978) A model of evolutionary change in proteins. In: *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.). Washington DC: National Biomedical Research Foundation, pp. 345–362.

Felsenstein, J. (1989) PHYLIP – phylogeny inference package. *Cladistics* **5**, 164–166.

Fitch, W.M. & Margoliash, E. (1967) A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability. *Science* **155**, 281–284.

Genetics Computer Group (1994) *Program Manual for the Wisconsin Package, Version 8, September 1994*, University of Madison, Wisconsin.

Hartley, B.S. (1960) Proteolytic enzymes. *Annu. Rev. Biochem.* **29**, 45–72.

Kahn, P. & Cameron, G. (1990) EMBL data library. *Methods Enzymol.* **183**, 23–31.

Kraulis, P.J. (1991) *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.

Merritt, E.A. & Murphy, M.P. (1994) RASTER3D version 2.0 – a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873.

NC-IUBMB (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology) (1992) *Enzyme Nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. Orlando, FL: Academic Press.

Rawlings, N.D. & Barrett, A.J. (1993) Evolutionary families of peptidases. *Biochem. J.* **290**, 205–218.

Reeck, G.R., de Haën, C., Teller, D.C., Doolittle, R.F., Fitch, W.M., Dickerson, R.E., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H. & Zuckerkandl, E. (1987) 'Homology' in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* **50**, 667.

Richardson, J.S. (1985) Schematic drawings of protein structures. *Methods Enzymol.* **115**, 359–380.

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

Webb, E.C. (1993) Enzyme nomenclature: a personal perspective. *FASEB J.* **7**, 1192–1194.

Alan J Barrett
alan.barrett@bbsrc.ac.uk

Neil D Rawlings
neil.rawlings@bbsrc.ac.uk

J. Fred Woessner
fwoessne@mednet.med.miami.edu