1

ASSAY TECHNOLOGIES FOR PROTEASES

Anuradha Roy

Del Shankel Structural Biology Center, High Throughput Screening Laboratory, Lawrence, Kansas

Gerald H. Lushington

Molecular Graphics and Modeling Laboratory, University of Kansas, Lawrence, Kansas; LiS Consulting, Lawrence, Kansas

James McGee

Quantitative Biology, Eli Lilly and Company, Indianapolis, Indiana

Rathnam Chaguturu

Del Shankel Structural Biology Center, High Throughput Screening Laboratory, Lawrence, Kansas; SRI International, Harrisonburg, Virginia

I. INTRODUCTION

Proteases are ubiquitously expressed enzymes which catalyze hydrolysis of peptide bonds and work under a wide range of conditions using diverse catalytic mechanisms [1]. Proteases specifically cleave protein substrates either from the N- or C-termini (aminopeptidases and carboxypeptidases, respectively) or in the middle of the molecule (endopeptidases) [2]. Proteolytic enzymes modulate many physiological processes ranging from nonspecific hydrolysis of dietary proteins to highly specific and regulated proteolysis in cell cycle regulation, tissue remodeling,

Enzyme Technologies: Pluripotent Players in Discovering Therapeutic Agents, First Edition.
Edited by Hsiu-Chiung Yang, Wu-Kuang Yeh, and James R. McCarthy.

blood coagulation, blood pressure control, angiogenesis, apoptosis, inflammation, ovulation, fertilization, and embryonic development [3,4]. Over 500 proteases each from humans, rat, mouse, and chimpanzee have been annotated and compiled in the Degradome database (http://degradome.uniovi.es) [5,6]. Information on all known proteases and their substrates/inhibitors is listed in the MEROPS database [7]. Based on the amino acid or metal that catalyzes the nucleophilic attack on substrate peptide bonds, the proteases are classified into five major types: aspartic (Asp), metallo-, cysteine (Cys), serine (Ser), and threonine (Thr) proteases. Aspartic and metalloproteases use an activated water molecule as a nucleophile to attack the peptide bond of the substrate, whereas in Cys, Ser, and Thr proteases, a catalytic amino acid residue (Cys, Ser, or Thr, respectively) serves as a nucleophile (Fig. 1). As a result, acyl–enzyme intermediates are formed only in the reactions catalyzed by Ser/Thr and Cys peptidases. Within each class of protease type are several enzymes that may have overlapping or distinct substrate recognition sites. Rawlings and Barrett proposed a classification of proteases into families based on amino acid sequence similarity, and families with similar three-dimensional folding are assembled into clans, indicating common ancestry [7,8]. The focus of this article is mainly on mammalian proteases and retroviral proteases which are of significant therapeutic relevance.

While pepsin in gastric juices digests a variety of proteins with broad specificity, renin is an example of Asp protease that shows high substrate specificity. Most proteases bind their substrates in fairly similar manner, first elucidated for papain by Schechter and Berger [9–11]. The catalytic site is flanked on one or both sides by sites that confer specificity of substrate binding to the protease and accommodate a side chain of an amino acid residue of the substrate. The enzymatic binding sites toward the N-terminus of the substrate are the non-prime side designated as S1, S2, …, S*n* from the catalytic site, and the residues C-terminal to the cleavage site are the prime side designated as S1′, S2′, …, S*n*′ [8,12,13]. The amino acid residues in the protein substrate which correspond to their respective subsites are numbered P1, P2, …, P*n* and P1′, P2′, …, P*n*′ (Fig. 1). Only few of the substrate binding sites have stringent specificities. For instance, site S1 confers specificity for Ser proteases and caspases, whereas the site S2, a hydrophobic subsite, defines specificity for the papain family of Cys proteases. In addition to the sites close to the catalytic site of the enzyme, distant sites on the enzyme may also contribute to the binding of substrates to the protease [9]. The specificity and biological activity of caspases are also determined by S4, which is distant from the catalytic site [14]. Proteolytic processing is being recognized as a mechanism for regulation of enzymatic activities, localization, and fate of proteins that are activated by limited and specific hydrolysis of peptide bonds. Dysregulation of proteolytic activity, structure, or expression results in major pathologies in the areas of cardiovascular diseases, cancer, neurodegenerative disorders, osteoporosis, diabetes type II, pancreatitis, inflammation, arthritis, and infectious diseases [4]. A large number of marketed drugs target the proteolytic enzymes that are involved in pathogenesis of various diseases [15] (Table 1). Although only a relatively small number of proteases are currently targeted for

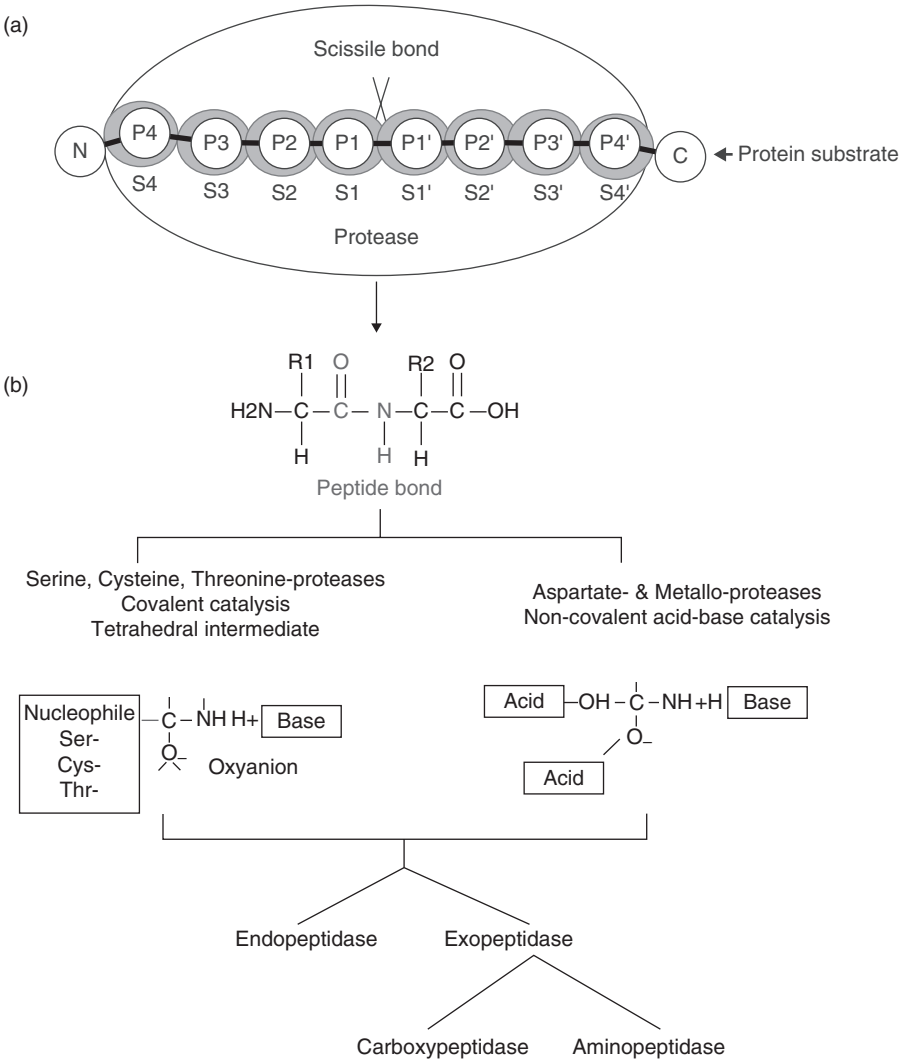

FIGURE 1 Schematic representation of binding of substrate to a protease site. (a) The binding sites of the protease are numbered on either side of the scissile bond, with the non-primed sites (S1, S2, …, S*n*) located toward the amino-terminus of the substrate and S1′ … S*n*′ or the primed subsites toward the carboxy-terminus. (b) Structure of the peptide bond which is hydrolyzed by proteases and the two basic catalytic mechanisms for all types of protease hydrolysis. In Ser, Cys, and Thr proteases, an amino acid at the active site serves as the nucleophile forming a transient covalent intermediate, whereas in metallo- and Asp proteases, an active water molecule functions as nucleophile (adapted from Reference [13]). The base in covalent catalysis is usually a His, and in non-covalent intermediate, Asp/Glu and zinc (metalloproteinases) serve as acids and bases. The proteases are also classified as endo- and/or exoproteinases based on their ability to cleave within or at the amino-/carboxy-terminus of the peptide chain.

TABLE 1 FDA-Approved Drugs for Select Proteases

Protease	Class	Compound	Company	Indication
HIV-1 protease	Asp	Atazanavir	Bristol-Myers Squibb	AIDS
		Darunavir	Prezista	
		Fosamprenavir	GlaxoSmithKline	
		Indinavir	Merck	
		Lopinavir	Abbott	
		Nelfinavir mesylate	Pfizer	
		Ritonavir	Abbott	
		Saquinavir	Hoffmann-La Roche	
		Tipranavir	Boehringer Mannheim	
Renin	Asp	Aliskiren (Tekturna)	Novartis	Hypertension
ACE	Metallo	Captopril	Bristol-Myers Squibb	Hypertension
		Enalapril	Merck	
		Lisinopril	AstraZeneca	
Carboxypeptidase A	Metallo	Penicillamine		Hypertension
MMP-1 and MMP-2 collagenases	Metallo	Periostat	Galderma Labs	Periodontitis
Enkephalinase	Metallo	Racecadotril		
Thrombin	Ser	Ximelagatran	AstraZeneca	Thrombosis
		Argatroban	Mitsubishi Pharma	
		Lepirudin	Aventis	
		Desirudin	Novartis	
Human FXa	Ser	Fondaparinux	Sanofi-Synthélabo	Thrombosis
Human neutrophil elastase	Ser	Sivelestat	Ono	Respiratory disease
Trypsin-like protease	Ser	Camostat mesilate	Ono	Pancreatitis I
Broad-spectrum protease	Ser	Nafamostat mesilate	Japan Tobacco	Pancreatitis inflammation
Plasminogen activator	Ser	Streptokinase		
Proteasome	Thr	Bortezomib (Velcade)	Millennium	Multiple myeloma
DPPIV	Ser	Pioglitazone	Takeda	Diabetes mellitus type II
		Saxagliptin	Bristol-Myers Squibb	
		Linagliptin	Boehringer Ingelheim	
Cathepsin K	Cys	Odanacatib	Merck	Osteoporosis/ bone cancer
		ONO5334	Ono	Osteoporosis

drug development, the commercial success of angiotensin-converting enzyme (ACE) inhibitors and human immunodeficiency virus (HIV) protease inhibitors makes the protease family a valuable target for disease treatment [16]. We have worked on a variety of protease assay formats using proprietary substrates for HTS and will present an overview of common protease assay technologies. We will end the review by discussing computational approaches to designing substrates for protease binding sites.

II. PROTEASE ACTIVITY ASSAYS

Identification of appropriate substrate is the first major step toward characterizing a protease and developing an assay to monitor its activity. A protein containing a short recognition sequence for a protease may work well in an *in vitro* assay, but may not be a physiologically relevant substrate for the proteolytic enzyme. In cases where the peptide sequences are unknown or ill defined, positional scanning experiments or phage display methods are used for screening combinatorial libraries of peptides for specificity determination. Bioinformatics-based prediction of cleavage sites and determination of protease preferences on synthetic substrates are important techniques in predicting natural protease substrates. Figure 2 summarizes the techniques employed in mapping substrates, a rapidly evolving field which is beyond the scope of this article and has been extensively reviewed [17–20]. The design of sensitive and selective synthetic peptide-substrate cleavage assays and the comprehensive mapping of active site specificity determinants are crucial for developing protease inhibitor drugs. The assays for proteolytic enzymes are generally continuous, homogeneous assays that can be performed in medium- to high-throughput formats. The assays are more relevant when the enzyme activity is measured with their specific native substrates, but assays with long native proteins have low hydrolytic rates and are time-consuming and costly. The substrate peptide sequences for most of the common proteases are very well established, and the fact that most proteases catalyze hydrolysis of small peptides has led to the development of technically simple and sensitive assays using fluorometric, colorimetric, and bioluminescent methods in which a single specific peptide bond is cleaved and the cleavage is monitored spectrometrically. A large number of kits are available from various vendors that serve to assay protease families using generic substrate peptides. In general, the minimal defined peptide substrate (average three to six amino acid residues) is synthesized based on the information on the binding fragments of natural substrates or inhibitors of proteases. In the case of many proteases like caspases, matrix metalloproteinases (MMPs), cathepsins, or HIV-1 proteases, the same substrate peptide or a minor variant of a substrate sequence is conjugated with either fluorophores or chromogenic groups or tagged with aminoluciferin to allow development of fluorescence-, absorbance-, or luminescence-based assay formats. Both cell-based and rapid mix-and-read biochemical methods have been developed for assaying the protease activity. The biochemical assays based on purified recombinant enzymes being

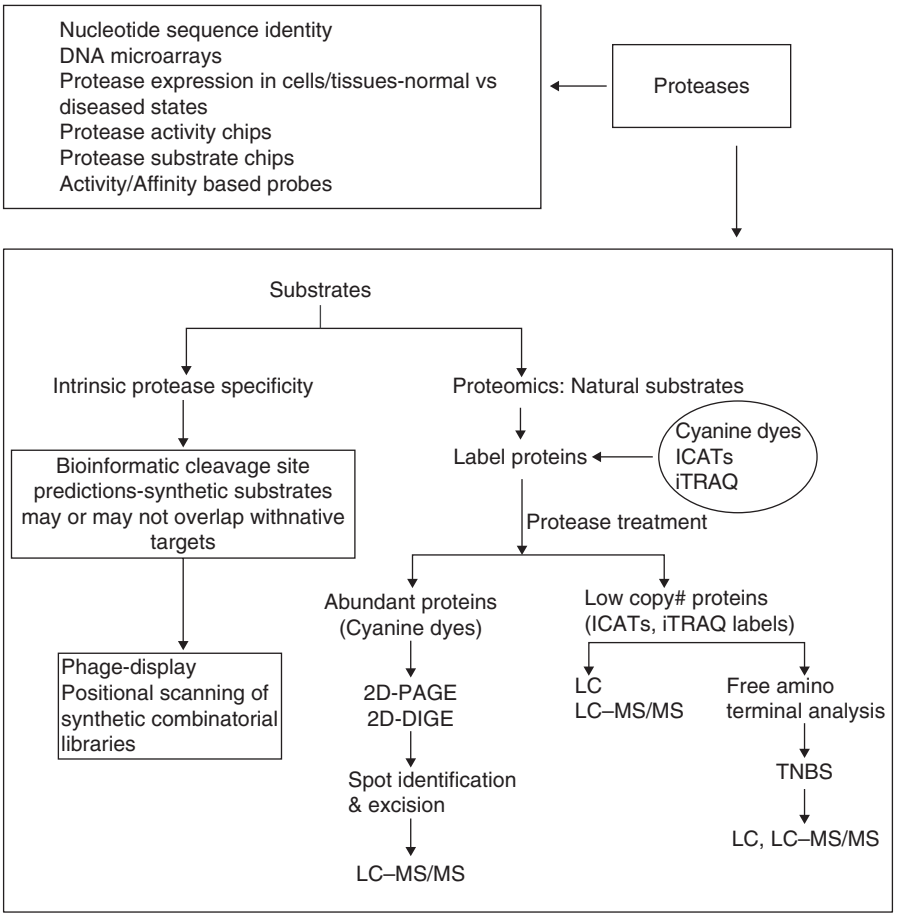

FIGURE 2 Overview of substrate and protease characterization. Protease identification/characterization may initiate from sequence identity/homology, from expression and activity chips, and from loss- or gain-of-function studies in cells or mouse models to establish their role in disease. *In vitro* biochemical hydrolysis and binding experiments based on short synthetic sequences of amino acids are often used to determine protease site binding and hydrolysis. Bioinformatics is useful for building theoretical predictions of substrate sequences, but may differ from real *in vivo* biological targets. Proteomics tools help in the identification of native substrates in cell environment. The protease-treated or protease-untreated cells are labeled with different cyanine dyes (Cy3, Cy5), and the proteins are resolved on two-dimensional (2D) gels (difference gel electrophoresis (DIGE) or 2D DIGE). The substrate and cleavage products are obtained from superimposition of the images from protease treated and controls. The proteins of interest are excised from gels and analyzed by liquid chromatography and mass spectrometry (LC–MS, multidimensional liquid chromatography, MuDPIT). Low-copy-number proteins can be labeled with more sensitive isotope mass tags (isotope-coded affinity tags (ICAT)), or label all primary amines in a trypsinized proteome (iTRAQ), followed by LC and LC–MS/MS. A comprehensive overview of protease substrate identification is available in Overall et al. [1].

more specific are useful for studying kinetics and for the screening of small-molecule libraries or peptidomimetic modulators. Figure 3 shows general aspects of activity assay optimization like defining optimal buffer components, pH, and stability; temperature determination for various substrates of interest; determining the concentration range for the purified protease; and determining the linearity of the reaction. The continuous assays are well suited for miniaturization and adaptation for HTS for identification of protease activity modulators. The modulators identified from the *in vitro* or cell-based assays are generally tested in orthogonal assays to ascertain specificity and selectivity (Fig. 3). The detailed guidelines for development and optimization of any enzyme including proteases for HTS are given in NIH assay guideline manual (http://assay.nih.gov/assay/), which, among others, is a useful reference manual. PubChem (http://pubchem.ncbi.nlm.nih.gov/assay) database lists over 3000 protease activity assays, which were used to identify modulators of protease activity. A few representative assays for proteases compiled from PubChem database are listed in Table 2. The following three major assay formats are widely employed for assaying the activity of various members of protease family:

A. Colorimetric Assays

The colorimetric assays, though less sensitive, have traditionally found more utility in active site titrations for estimating active protease concentrations or for detection of activity in biological fluids. Colorimetric assays based on substrates like amino acid–ethyl, amino acid–methyl, or amino acid–alpha-naphthyl esters were used earlier for quantification of activity of proteases such as trypsin, thrombin, chymotrypsin, pronase, plasmin, and urokinase [21]. These reactions could be used for kinetic analysis and for zymogram studies, but many of these substrates were unstable. The use of tripeptide or other polypeptide substrates afforded higher specificity and stability. In majority of direct colorimetric assays, a substrate peptide (three to five residues) is linked to a chromophore, usually *para*-nitroanilide (*p*-NA). Enzymatic hydrolysis of the substrate releases free *p*-NA, a yellow-colored compound which has high absorbance at 405 nm and is detected spectrophotometrically (Fig. 4a). In the presence of excess substrate concentration, an increase in the rate of absorbance of chromophore released is linearly related to enzyme concentration. A large number of chromogenic substrates are available from Spectrozyme and American Diagnostica for assaying general Ser proteases as well as coagulation factors and kallikreins. For example, ACTICHROME® factor X (fX) is a chromogenic assay for the measurement of fX activity in human plasma. The assay involves activation of fX in the plasma to factor Xa (FXa) by Russell's viper venom. Activated FXa hydrolyzes the Spectrozyme® chromogenic substrate and releases the chromophore, *p*-NA. The color of the reaction solution is read spectrophotometrically at 405 nm. In a large number of colorimetric caspase activity assays, short four-residue peptides linked to *p*-NA serve as substrates. For instance, the DEVD-*p*-NA chromogenic substrate is used to assay caspase-3 and caspase-8 activities in cell lysates. The colorimetric QuantiCleave Protease Assay

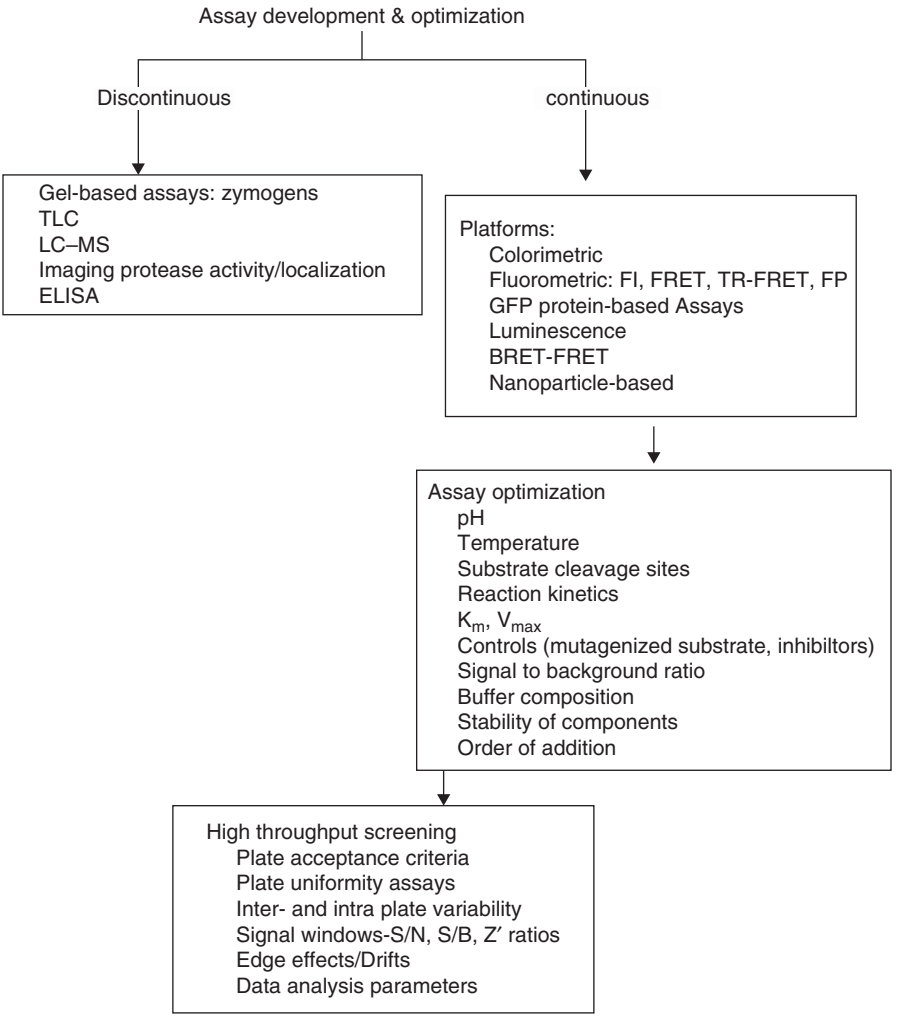

FIGURE 3 Protease activity assay development and optimization. Protease assays can be discontinuous or homogeneous. Examples of discontinuous, low throughput assays include protease activity determination in zymogens, analysis of peptide fragments by TLC or via LC–MS, and imaging of proteases in cell environment by using tagged peptides and ABPs. Biochemical or cell-based assays are continuous assays which are based on colorimetric, fluorometric, or luminescence platforms. The three basic formats can be used in conjunction with nanoparticles or used as such with substrates labeled with appropriate tags. The continuous assays can be further miniaturized and adapted for HTS by further optimizations to conform to stringent statistical acceptance criterion of HTS assays.

TABLE 2 Representative Protease Screens Compiled from PubChem Database

Protease	Assay format	Disease/significance
Cathepsin L	Fluorescence increase by hydrolysis of Z-Phe-Arg-AMC	Proteolysis by cathepsin L required for entry and replication of SARS and Ebola virus
High-temperature-responsive antigen (HTRA-1) Ser protease	FP with FP-TAM probe	Osteoarthritis, AD, and age-related macular degeneration
Insulin-degrading enzyme (IDE), an Abeta-degrading zinc metalloprotease	Fluorescence polarization: fluorescein-Abeta-(1–40)-Lys-biotin (FAbetaB) peptide incubated with IDE dissociates biotin moiety from fluorescein. Addition of avidin to the reaction increases the mass of intact substrate, slowing their rotation rate and decreasing depolarization of plane-polarized light. The low-molecular-weight cleaved substrate rotates rapidly and causes strong depolarization. Ratios determine the relative amounts of cleaved and intact forms of the FAbetaB substrate	AD is characterized by accumulation of amyloid beta-protein
Proteasome, ATP-dependent protease	Flow cytometry: FLAG-tagged, fluorescent proteasomes were captured on anti-FLAG beads, and disassembly was monitored by loss of bead fluorescence in the presence of ATP. The beads were sorted via high-throughput flow cytometer	Inhibitors of proteasome assembly and activity
Sentrin-specific protease-8 (SENP-8)	Luminescence: SENP8-dependent RLRGG-aminoluciferin peptide-substrate hydrolysis – releases aminoluciferin, measurable in a coupled luminescence detection assay	Involved in maturation of SUMO precursors (endopeptidase cleavage) and deconjugation of the targets (isopeptidase cleavage)
SARS coronavirus 3C-like protease	FRET assay: HiLyte Fluor TM 488 is attached at the N-terminus of a 3CLpro peptide substrate (HiLyte Fluor 488-ESATLQSGLRKAK (QXL520)-NH2 (AnaSpec) and is quenched by a QXLTM520 moiety at the C-terminus. Following cleavage, separation of the fluorescent compound and quencher leads to an increase in fluorescence	Cold, lower respiratory tract infections, and diarrhea

(*Continued*)

TABLE 2 (*Continued*)

Protease	Assay format	Disease/significance
Ub-specific peptidase 2 isoform a	Ub-IsoPro1 kit (Progenera, LifeSensors, Inc.): a Ub or ubiquitin-like (UBL) protein conjugated to a reporter enzyme, phospholipase A2 (PLA2), which has an absolute requirement for a free amino-terminus. Fusion of a UBL to the N-terminus of PLA2 inactivates PLA2. When the UBL–PLA2 reporter enzyme is cleaved by USP2, the activated reporter can subsequently act on its substrate, available NBD C6-HPC (Invitrogen), with a fluorescent readout	Ub homeostasis

Source: http://pubchem.ncbi.nlm.nih.gov/assay.

Kit (Thermo Scientific) is based on the usage of modified casein as substrate, in which all surface primary amines have been succinylated. When succinylated casein is cleaved at peptide bonds by proteases, free alpha-amino-terminal groups of peptide fragments are released. The primary amines react with trinitrobenzene sulfonic acid (TNBSA) to form a yellow-colored TNB–peptide adduct [22] which is measured spectrophotometrically at 450 nm (Fig. 4b). The non-succinylated casein serves as a control substrate in these assays. Replacement of the peptide cleavage bond with a thioester bond results in creation of free sulfhydryl in an assay reported for MMPs [23,24]. The free sulfhydryl group reacts with 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB, Ellman's reagent) to form 5-nitro-5-thiobenzoate, a colored compound with absorbance at 412 nm. A thiopeptide (Ac-PLG-[2-mercapto-4-methyl-pentanoyl]-LG-OC2H5)-based assay kit for MMP-9 is available from Enzo Life Sciences. Most colorimetric assays have been replaced by more sensitive fluorometric- or luminescence-based formats.

B. Fluorescence Assays

In recent years, homogeneous fluorescence- and luminescence-based assay formats have gained wider acceptance due to higher sensitivity, larger signal-to-noise windows, high adaptability to enzymes from various sources, and requirement for very low volumes of reagents. Fluorescence assays are most commonly employed for developing assay platforms for protease activity [25], and most of the assays are based on conjugating extrinsic fluorophores like fluorescein, rhodamine, and BODIPY dyes to a substrate of interest [26]. The fluorophores with distinct excitation (Ex) or emission (Em) properties can be multiplexed within the same reaction, enabling simultaneous measurements of multiple endpoints. Sensitivity of fluorescence-based assays is dependent on accumulation of cleaved product and may be limited by

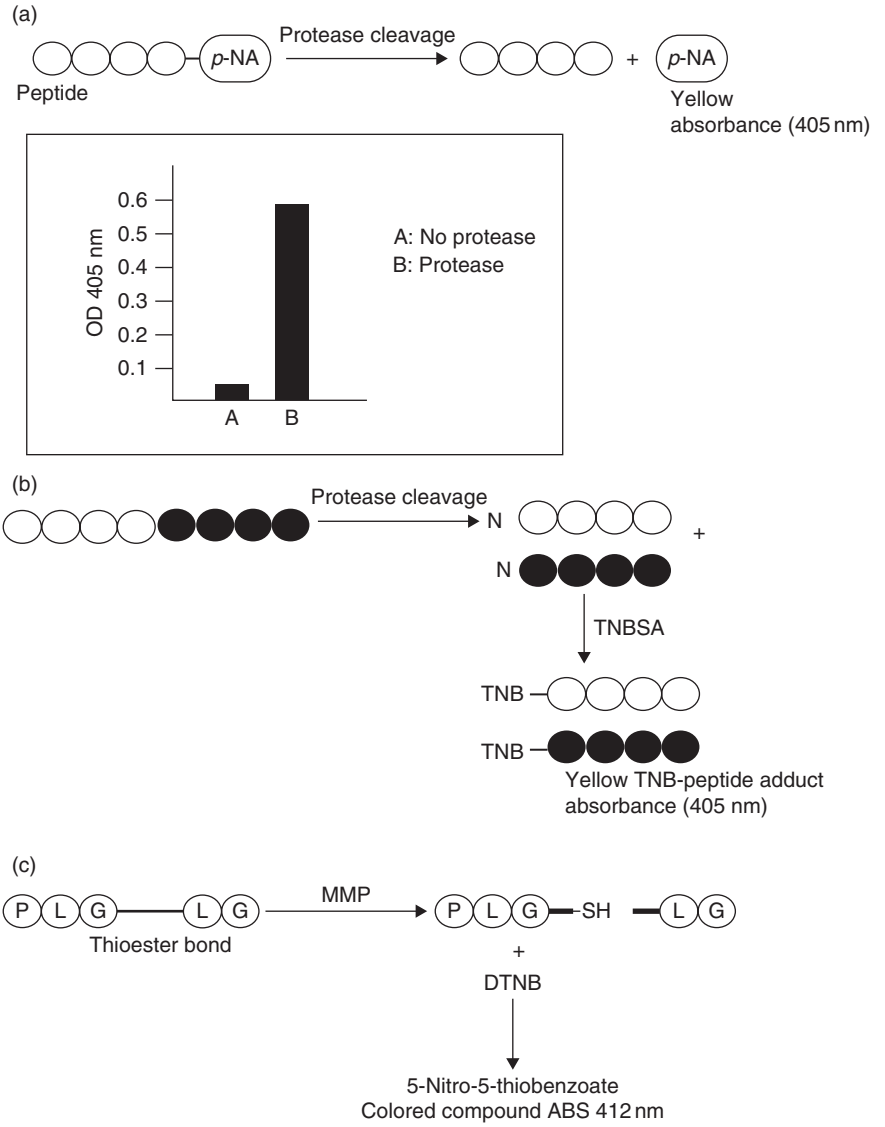

FIGURE 4 Colorimetric assays for proteases. Examples of some colorimetric methods in which protease cleavage is followed by increased absorbance in direct or in coupled assays. (a) Short peptides are conjugated with *p*-NA, which is released following hydrolysis of amino acid–*p*-NA bond, resulting in increased absorbance at 405 nm. (b) TNBSA reacts with the amino-termini of the protease hydrolyzed peptides to increase absorbance at 405 nm. (c) Thioester bond at the peptide cleavage site releases a peptide fragment with reduced sulfhydryl group which reacts with DTNB to form a colored compound with absorbance at 412 nm.

residual fluorescence of peptide-conjugated fluorophores or spectral overlap of cleaved fluorescent products. Fluorescence assays may involve either direct fluorescence intensity measurements [27], Förster resonance energy transfer (FRET), fluorescence polarization (FP) [28,29], or time-resolved measurements (homogeneous time-resolved fluorescence (HTRF), time-resolved Förster resonance energy transfer (TR-FRET)). The measurement method is dependent upon the basic design of the synthesized peptide substrate. In addition, FRET measurements are also performed using fluorescent protein-based substrates. In the simplest substrate type, a short peptide substrate is linked to a fluorophore, which is quenched when bound to the peptide, but fluorescence is detected upon peptide cleavage by a protease (Fig. 5). Many Ser and Cys protease assays are based on simple peptide-conjugated fluorophores. In these proteolytic reactions, the substrate specificity is defined by S subsites, and an aminoacyl intermediate is formed during proteolysis. Fluorogenic substrates for these proteases contain aromatic amines like 7-amino-4-methylcoumarin (AMC) or 4-methyl-7-coumarylaminde (NH-Mec), which are conjugated to short peptides that confers specificity. The cleavage of amide bond by the proteases increases fluorescence. The fluorescent group occupies the S1′ pocket of the proteases and undergoes a change in its fluorescence characteristics on acylation of the amino acid. Many of the caspase activity assays are based on such substrates in which a tetrapeptide is linked to the fluorophore like 7-amino-4-trimethylcoumarin. Cleavage of these substrates by purified caspases or apoptosis-induced cell lysates results in an increase in fluorescence. One of the disadvantages of such substrates is an associated high background signal and proper storage and intactness of the peptide substrate. This format may greatly reduce the sensitivity range for an assay. In other cases, the peptide substrates are synthesized such that the fluorophore Em is quenched by proximity of the second group, the quencher. The peptidase activity physically separates the quencher from the fluorophore and helps restore detectable Em of fluorescence. In FRET, FP, and HTRF/TR-FRET, the peptide substrate is designed to possess two fluorophores, one of which is the donor, which transfers energy to the second fluorophore, the acceptor. The Em spectrum of the donor overlaps with the absorption spectrum of the acceptor (Fig. 5). The acceptor either can emit energy (fluorescence increase) or may serve as a quencher. The same peptide substrate is utilized for FRET and FP assays, and many such peptides are commercially available for generic protease activities. In FRET, direct fluorescence intensity is measured at defined Ex and Em wavelengths and is sensitive to absorptive interference from colored compounds. This spectral interference can be bypassed using FP assay, in which the ratio of orthogonal fluorescence intensities is measured. An interesting FP assay was developed for assaying human cytomegalovirus protease activity using a peptide substrate which was biotinylated at amino-terminus and was conjugated to a fluorophore at the C-terminus [28,29]. After incubation of the substrate with recombinant cytomegalovirus protease, avidin was added. Binding of avidin to the biotin end of the probe produced a polarization signal which was a function of relative amounts of cleaved and uncleaved substrate. The higher-molecular-weight uncleaved substrate bound avidin to generate a high polarization value, whereas the cleaved, low-molecular-weight peptide produced a low polarization value.

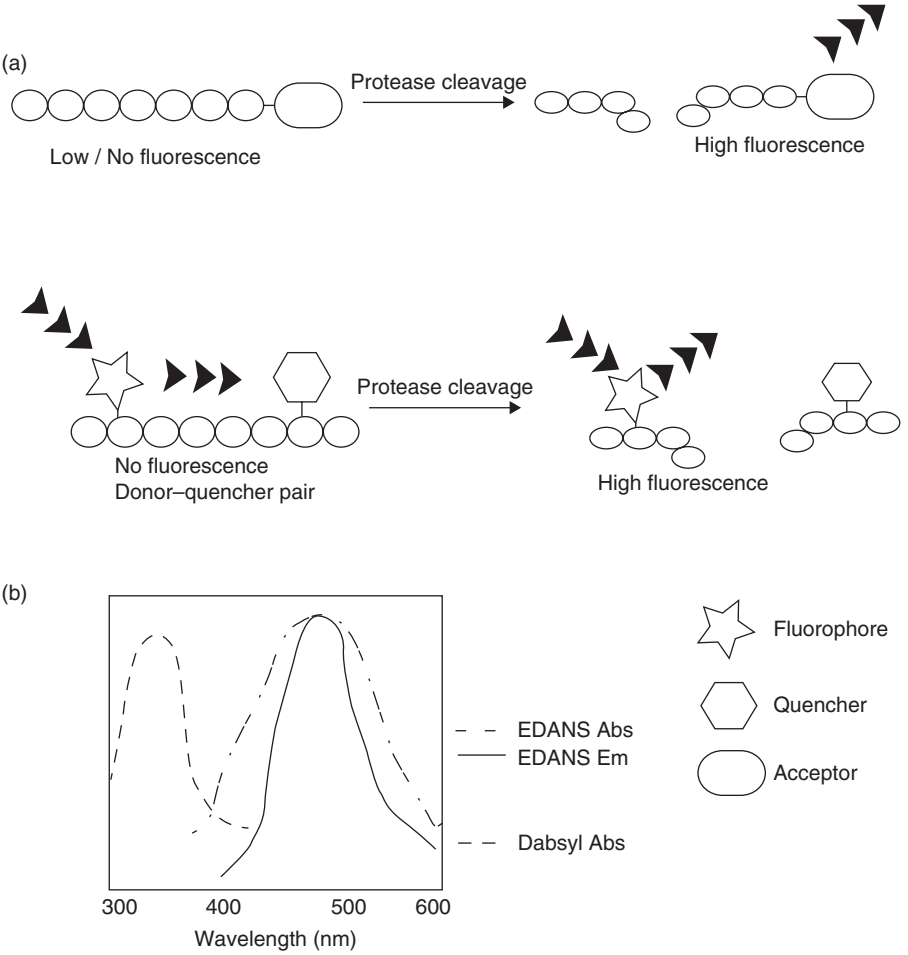

FIGURE 5 Fluorometric assays for proteases. Fluorescent groups are added to peptide substrates to generate substrates for fluorescence-based assays. The simplest substrate is synthesized by adding a coumarin class of fluorophores to linear peptide sequence. An increase in signal is registered with the release of fluorophore post-protease-mediated hydrolysis. (a) Donor–quencher combination in peptide substrate ensures higher signal above background. (b) The overlap of Em spectra of EDANS and Ex spectra of DABSYL characterizes a suitable FRET pair useful for many fluorescence-based platforms.

For the FRET/FP assays, the labeling moiety does not need to be placed at the catalytic site and permits design of substrates with optimal residues at the active site. Since FRET is based on electromagnetic energy transfer between the donor and acceptor pairs of fluorophores separated by very short distance of 10–100 Å, the substrates are designed such that fluorescence increases on cleavage of peptide as a result of loss of intramolecular fluorescence quenching. The selection of donor–acceptor fluorophores requires a high degree of overlap between the Em spectrum of

the donor and absorption spectrum of the acceptor, preferably in the long wavelength region; the high quantum yield of the donor; and an effective quenching of donor fluorescence [27]. Several donor–acceptor pairs have been developed like EDANS/DABSYL (4-(4′-dimethylaminophenylazo) benzoic acid, Dansyl/Trp, and naphthalene/anthracene (5 (2′aminoethylamino) naphthalenesulfonic acid). In EDANS/DABSYL pair, quenching of EDNAS fluorescence in substrates with amino-terminal DABSL group is very efficient because of almost complete overlap between the excited state of EDANS and the DABSYL absorbance and the high molar extinction coefficient of DABSYL (Fig. 5). Many aspartyl proteases and metalloproteases require amino acids P′ to the cleavage site for enzyme recognition, and their peptide substrates are designed such that the fluorophore and quencher are separated by less than three or four residues, allowing short-range intramolecular interactions. Many proteasome-mediated events like conformational changes in protein or protein degradation, which change the distance between the acceptor and donor pairs, are efficiently assayed using FRET-based techniques.

The small stokes shifts and near-UV absorption/Em spectra of donor–acceptor pairs in FRET assays have significant drawbacks in HTS of compound libraries since FRET reactions are influenced by high backgrounds from compounds, microplates, reagents, low signal-to-background ratios, and short fluorescence lifetimes. Using a combination of fluorophores with widely separated wavelengths, for example, a protease assay using dual substrates labeled with fluorescent probes with nonoverlapping Ex/Em wavelengths like rhodamine 110- and coumarin-based fluorophores, in the same well allowed for identification of true inhibitors which were active at both fluorescence settings [30]. These problems of FRET assays are circumvented by TR-FRET assays or HTRF assays. The TR-FRET/HTRF assays, introduced by *Cisbio* International, are based on using rare earth lanthanides like dysprosium (Dy), samarium (Sm), terbium (Tb), and europium (Eu). The lanthanides have poor intrinsic fluorescence but can be complexed to rare earth chelates and cryptates to enhance their fluorescence and prevent decay (cisbio.com). The complexed lanthanides have large stokes shifts and long Em half-lives of 100–1000 μs and help avoid signal contamination by shorter-lived signals contributed by other components of the assay. The HTRF assays generally use complexed cryptates like EuK as an energy donor and XL665 protein as an acceptor (Em 665 nm) [31]. The Em spectra of donor and the Ex spectra of the acceptor overlap and the donor–acceptor pairs are placed at 10 nm or less from one another. A highly sensitive time-resolved fluorescence quench assay (TR-FQA) was developed for caspase-3 [32] using a hexapeptide substrate labeled with fluorescent Eu chelate at one end and a Dabacyl quencher at the other end. Cleavage with caspase-3 allowed measurement of time-resolved Eu signal to the acceptor allophycocyanin. Time resolution allowed the separation of the fluorescence signal by spectral and temporal filtering and resulted in very high signal-to-background ratios. In addition to the extrinsic fluorophores, many protease assays are based on the naturally fluorescent proteins, like the green fluorescent protein (GFP), which can be used for both *in vitro* and *in vivo* assay formats.

Protein-Based Fluorescence Reporter Substrates The reporter constructs in which the protease substrate sites are introduced between the coding sequences of the fluorescent proteins are used in many cell-based protease assays. The cell-based

protease assays require transient or stable expression of the GFP which absorbs blue light and emits green light, in the absence of any cofactor requirement. The GFP chromophore arises from posttranslational modifications and comprises of *p*-hydroxybenzylideneimidazolinone formed by cyclization of Ser65, Tyr66, and Gly67 and 1,2-dehydrogenation of the tyrosine (Tyr) [33].The wild-type GFP has been extensively mutagenized (Fig. 6a, [33,34]) to improve spectral characteristics and quantum yields and generate more useful variants like eGFP, blue fluorescent protein (BFP), cyan fluorescent protein (CFP), and yellow fluorescent protein (YFP) [33,34]. Many protease assays use FRET-based reporters for monitoring protease activity in cellular environment in which a protease cleavage site is placed between two fluorescent variants of GFP or between a GFP and a fluorophore, with overlapping donor Em and acceptor Ex spectra [35]. Induction of protease activity in cells expressing the fusion protein results in peptide cleavage followed with the loss of the FRET signal between the donor and acceptor, when excited at the wavelength of donor absorbance. In general, both live cells expressing FRET constructs as well as cell lysates can be used for studying protease activity. Onuki et al. [36] reported a caspase-8 assay, in which its native substrate, BID, a BH3-interacting domain death agonist peptide, was flanked by CFP, the cyan fluorescent variant of GFP, and YFP, the yellow fluorescent variant. The absorbance of light at 433 nm by CFP results in energy transfer to YFP, which emits fluorescence at 527 nm (Fig. 6b). The hydrolysis of the protease site by caspase-8 results in loss of FRET Em at 527 nm, with a gain of Em at 475 nm from CFP [36]. The ratio of light Em at the two wavelengths is used for FRET quantification and is proportional to protease activity. Most assays are based on energy transfers between CFP/BFP and GFP or YFP separated by specific protease cleavage sites. Since most cells express a wide variety of proteases which have redundant activities, a direct correlation between the protease type and measured activity is obtained only when cells are transfected with highly specific peptidase sites in the FRET expression constructs. The FRET-based reporters can be used for imaging cell systems with detectable GFP expression. In addition to the reporter assays expressing fluorescent proteins, short fluorescence-based cell-permeable probes have been used for studying specific protease activities and their localization in live or fixed cells and in the small animal models.

Imaging of Protease Activities Fluorescence-based probes have been used in many *in vitro* and *in vivo* imaging techniques for monitoring and detecting active proteases within live cells and small animal models. For instance, the degradation of extracellular matrix (ECM) by collagenases, which are required for migration of cancer cells, was studied using BT549 breast cancer cells grown in the presence of fluorescent ECM protein substrate, FITC-labeled collagen IV. The areas containing cell clusters degraded the matrix and, after fixing, were imaged as nonfluorescent, cleared areas, showing that the ECM substrates have relevance in the cell migration through the tumor matrix. However, since high background fluorescence and cell fixation preclude the ability for real-time data imaging, the technical problems were circumvented by using the dye-quenched fluorescent protein substrates. The dye-quenched substrates contain high density of FITC molecules which are quenched due to high molecular proximity [37]. Protease-mediated cleavage of the substrate releases fluorescent degradation products. BioMol has developed cell-based assays which

(a)

Fluorescent Protein (FP)	Mutations	Excitation Max (nm)	Emission Max (nm)
Wt GFP	None	395, 495	511
Green (eGFP)	S65T, F64L	489	509
Blue (BFP)	Y66H, Y145F, F64L	385	445
Cyan (CFP)	Y66W, S65A, S72A, M153T, V163A	434	477
Yellow (YFP)	T203Y, S65G, S72A, V68L	514	527
T-Sapphire	T203I, Q69M, C70P, V163A, S17G	399	511

(b)

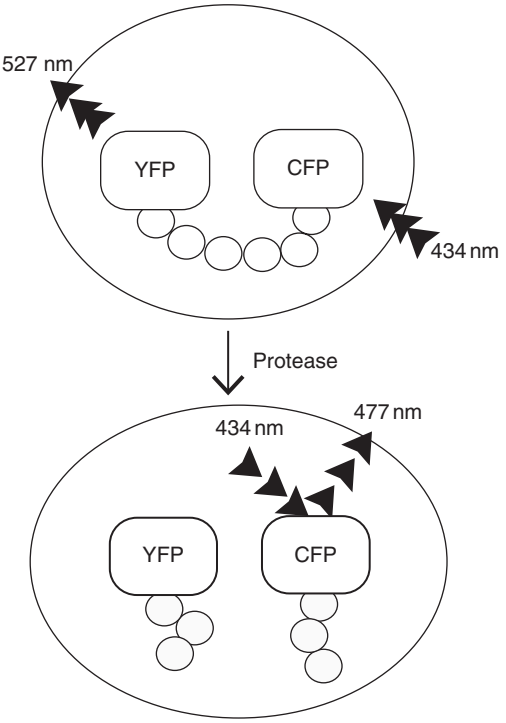

FIGURE 6 Fluorescent protein-based FRET assays. (a) Green fluorescent protein, an intrinsically fluorescent protein, and its variants derived from extensive mutagenesis of the wild-type GFP. (b) A schematic showing a FRET assay using YFP and CFP as acceptor/donor in a cell-based reporter assay. The peptide sequence between the two fluorescent proteins could represent any sequence specific for caspases or cathepsins or metalloproteinases.

utilize cell-permeable fluorescent substrate probes for *in situ* localization of active proteases. The cells with active Ser proteases stain green (FFCK-FAM) and cells with active caspases stain red (SR-VAD-FMK). Cells displaying either or both activities are easily detected using fluorescence microscopy and FACS. A number of

polymer-based near-infrared fluorescence (NIRF) probes have been introduced for imaging protease activity [38] in whole organism models. The NIRF probes consist of a high-molecular-weight carrier like polyethylene glycol (PEG), a cleavable peptide spacer, and a near-infrared fluorophore, indocyanine green, which is an FDA-approved tricarbocyanine dye. Many such probes have been used to monitor activity of caspases [39], metalloproteases, cathepsin, and coagulation enzymes [40]. In one such example, an NIRF peptide substrate was developed to monitor general cathepsin activity *in vivo* [38]. Since most lysosomal cathepsins cleave the polylysine peptides, fluorophores were coupled at high density to an inert copolymer composed of polylysine backbone stabilized with PEG side chains, and the NIRF peptide was injected along with a dye in mice. The high density of fluorophores quenched fluorescence Em, but *in vivo* cathepsin-mediated proteolysis of the peptide decreased the fluorophore density and increased the Em signal in the near-infrared range, the long wavelength which was not subject to scattering or absorption by tissues [38].

FRET-based probes are not specific and can generally be hydrolyzed by many cellular proteases. The intracellular responses to a variety of physiological and pathological conditions can be monitored with the help of more specific and stable, fluorescently labeled activity-based probes (ABPs) [41]. Activity-based probes form covalent complex with the protease active site since they are often based on specific protease inhibitors that occupy protease active sites. The highly hydrophobic ABPs are membrane permeable and can detect specific protease activity in a cellular microenvironment. All ABPs consist of three parts: reactive group, linker, and tag. The reactive functional group, the warhead, is an electrophilic group, which covalently binds to the catalytic nucleophiles (Ser, Cys, or Thr in Ser, Cys, or Thr proteases, respectively) located at the active site of the enzymes [41,42]. The steric hindrance between the warheads from the tag is reduced by a variable linker, which may be a nonspecific spacer of alkyl or PEG groups or may be designed after a substrate. All ABPs contain a tag to enable detection and/or purification of labeled proteins. The tags may be radiolabels (I-125, H-3), fluorescent groups (cyanins, BODIPY, rhodamines, TAMRA), or biotin, which have been used for activity profiling of Cys proteases like cathepsins [43], caspases, and proteasomes. The design of specific ABPs requires that the target proteases have well-defined catalytic mechanisms and known inhibitors for which structural and kinetic data are available [41]. The ABPs have been used extensively for Cys and Ser proteases. DCG-04 was designed based on E-64, a natural product inhibitor of papain family of Cys proteases, and was used to determine the role of papain-like proteases in tumor progression, parasitic invasion, and cell cycle regulation [44]. The caspase-1 of Cys protease family was identified using ABPs containing a biotin-tagged specific peptide sequence coupled to acyloxymethyl ketone (AOMK) and aldehyde reactive groups [45]. The ABP based on specific Ser protease inhibitor, diisopropyl fluorophosphonate (DFP), was used to profile Ser hydrolases for biomarkers for cancer, and activities of at least seven Ser hydrolases were found to be differentially expressed in MDA-MB-231 breast cancer cells in cell culture and in xenograft model [46]. The metalloproteases, with their zinc-containing active sites, hydrolyze the substrate peptide bond without forming acyl–enzyme intermediates. An ABP against metalloproteinase, based on its inhibitor peptide hydoxamate [47], a zinc chelator, and a photocross-linking group were used to profile MMP activity in melanoma cell lines. The study

uncovered the role of neprilysin, a membrane protein, in progression of melanoma. The ABPs have proven useful in unraveling both previously unknown drug targets and novel players in various pathophysiological processes. In summary, fluorescence-based formats have extensive applications in development of many *in vitro* and *in vivo* imaging assays. At the same time there are several drawbacks in fluorescence-based approaches for biochemical and cell-based assays. Although fluorescence assays are very sensitive, they are susceptible to interference from inherent absorption characteristics of chemical compounds in the *in vitro* assays. In cell-based systems, the quantum yields of many GFP derivative proteins are weak, and many such fluorescent protein-based constructs have limited sensitivity, and the large size of fluorescent proteins limit the spatial resolution of the constructs. Many FRET-independent photochemical changes can result from external light source to initiate fluorescence transfer causing background noise or photobleaching [48,49]. In addition, the use of ultraviolet light for Ex of BFP in cell-based reporter assays has cytotoxic effects. The problems with fluorescent assays can often be circumvented using luminescence assay formats, which are also not affected by compound interference, photobleaching, and phototoxic effects.

C. Bioluminescence Assays

Bioluminescent assays are based on naturally occurring bioluminescence reactions catalyzed by the enzyme luciferase derived from *Photinus pyralis*. Bioluminescent assays are rapid, are sensitive, and are based on the reaction catalyzed by luciferase enzyme, which acts on its substrate luciferin or aminoluciferin in the presence of ATP and generates energy-rich peroxide intermediates, which spontaneously decompose and reach the ground state accompanied with Em of photons at 550–605 nm. Any modification of the amino or carboxy group of aminoluciferin makes the resultant aminoluciferin derivative insensitive to luciferase reactions [50]. Protease substrates are covalently conjugated via a peptide bond to the amino group or to the modified carboxy-terminal derivative of aminoluciferin. The protease-mediated cleavage of the peptide bond linking the substrate and aminoluciferin generates free aminoluciferin, which is available as a viable substrate for luciferase enzyme [51]. These assays are nonhomogeneous reactions that require compatibility between the protease assay conditions and the luciferase assay buffer especially with respect to pH requirements (Fig. 7a). The amount of aminoluciferin released is a direct measure of protease activity. A detailed comparison of the bioluminescent assays with corresponding fluorescent assays for caspase-3/7, caspase-8/9, DPPIV, and calpain showed that the luminescence assays were more sensitive (100–1000-fold) [50]. Ser and Cys proteases which do not require P′ sites for cleavage work well with aminoluciferin–peptides. A single-step homogeneous assay was reported for caspase-3, in which the aminoluciferin-tagged peptide substrate was premixed with luciferase enzyme before adding to wells containing caspase-3 enzyme [52]. Many FRET-based approaches have been replaced with BRET (bioluminescence resonance energy transfer) in which the fluorescent donor is replaced by a bioluminescent luciferase. The luciferase produces an initial photon Em, which is transferred to an acceptor (a fluorophore) that absorbs the donor energy and emits light at a longer

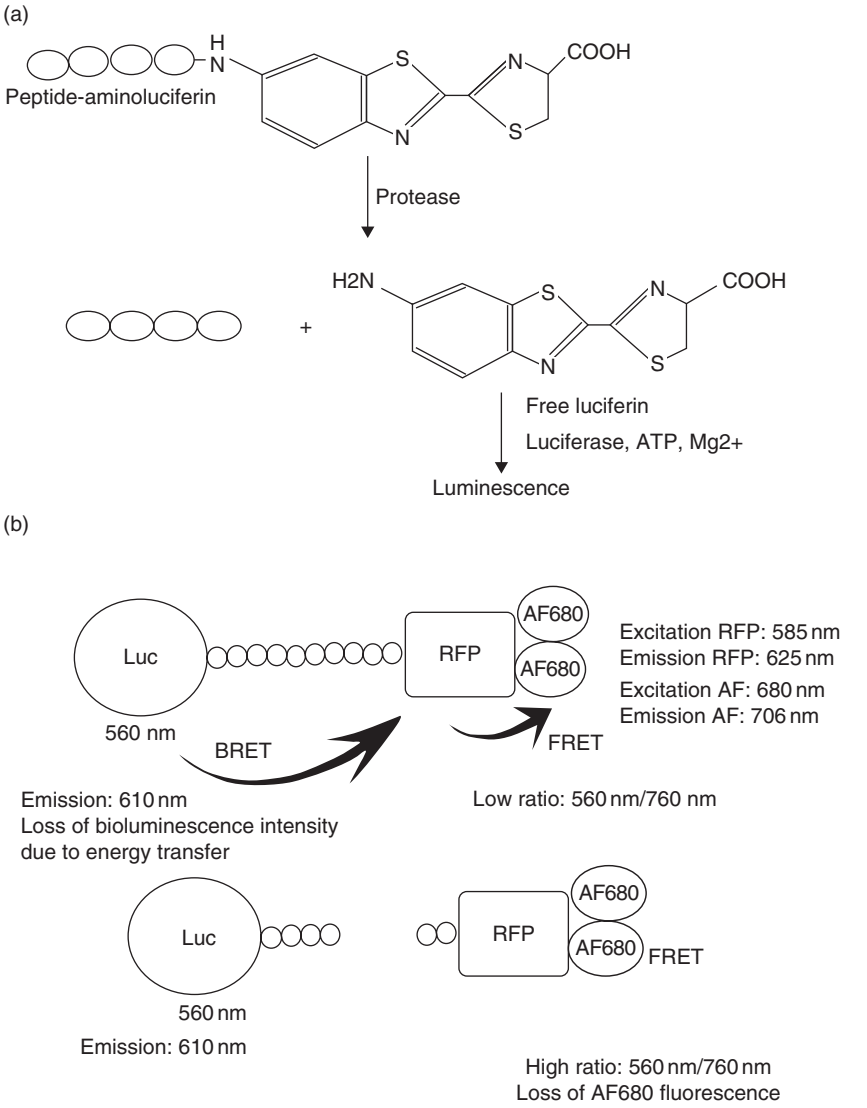

FIGURE 7 Luminescence-based protease assays. (a) Basic construction of the luciferase substrate conjugated to a protease-specific peptide sequence. The peptide–aminoluciferin has no detectable activity. Hydrolysis of the aminoluciferin–peptide bond by protease releases free aminoluciferin which is a usable substrate for luciferase in the presence of ATP and magnesium. (b) A luciferase-based bioluminescence transfer assay (BRET) in which luminescence is detected only after proteolysis.

wavelength. In one example of BRET format, the protease substrate decapeptides containing FXa, thrombin, or caspase-3 recognition sequences were flanked by the bioluminescent luciferase protein (Em 610 nm) and a red fluorescent protein (Ex 585 nm/Em 625 nm) covalently labeled with an NIRF dye, AF680 (Ex 680 nm/Em 705 nm). The energy transfers were possible due to overlap between the red bioluminescence Em at 617 nm of the Luc variant and the AF680 absorbance at 680 nm (Fig. 7b). The ratios of 560 to 760 nm, determined following protease cleavage of the protein substrates, were monitored by recording Em spectra and plotting the change in peak ratios over time [53]. The bioluminescent reporters have high sensitivity and have very high signal windows due to minimal autoluminescence of cells and tissues. The three basic assay formats described earlier are also being developed in conjunction with new technologies like nanoparticle-based assay systems.

D. Nanoparticle-Based Protease Activity Assays

Nanoparticles are nanometer-sized particles that have found wide utility in biology and other fields of study. Since nanoparticles are similar to subcellular components like proteins, they provide an interface to study events at nanoscale [54]. The nanoparticles are tagged with antibodies, fluorescent or chromogenic tags that enable detection and quantification of an enzyme activity, and in many cases the material composition of nanoparticles itself contributes to optical detection properties of the assays. Nanoparticle-based assays are very sensitive and can be used to monitor protease activity in real-time both *in vitro* and *in vivo*. The protease substrates immobilized onto a nanoparticle surface are acted upon by active enzymes resulting in a change of nanoparticle environment, which is measured via fluorescence, absorbance, and imaging or by biophysical methods. Gold nanoparticles (AuNP) exhibit characteristic size- and shape-dependent electronic and spectral properties, which have been exploited to develop protease activity assays. Colloidal AuNP tethered to peptides containing two Cys residues flanking protease cleavage site acquires blue color due to aggregation [55,56]. Protease-mediated cleavage of the activity on the immobilized substrate disperses the aggregates, changing the color to red [57]. In addition to colorimetric assays, AuNP and other metallic nanoparticles are known to quench fluorescence of transferred excited electrons [58]. The electrons transferred from fluorophore-containing peptides are quenched when attached to AuNP. After the cleavage of peptide by protease, the physical separation from the AuNP restores the previously quenched fluorescence [59]. An MMP-2 assay was established using Cy5.5-labeled substrate in which association with AuNP quenched the signals. With MMP-2 activity, the fluorescence signal recovered both *in vitro* and in mice experiments [60]. A new method to assay MMPs expressed in tumors was reported recently using composite gold–iron oxide (Au–Fe_3O_4) nanoparticles. An optical probe containing Cy5.5-GPLGVRG-TDOPA was immobilized on the iron oxide surface and SH-PEG (5000) for *in vivo* imaging on the gold surface. The methodology exploited the quenching properties of AuNP and stable surface chemistry of iron oxide nanoparticles. The MMP activity was assayed by fluorescence imaging of both *in vitro* and *in vivo* mouse tumor models [61]. In addition to AuNPs, magnetic [62,63] and

polymeric- and silica-based nanoparticles [64,65] are also used in protease sensing assays. In addition to the fluorescence quenching properties, the iron oxide-based magnetic nanoparticles [64] are useful for tracing distribution of injected samples by magnetic resonance technology. The nanoparticle-based protease assay systems are quantitative, specific, and sensitive and will soon find wide applications in drug screening campaigns and in routine assay development.

III. ASSAYS FOR SOME CLINICALLY SIGNIFICANT PROTEASES

Because of their therapeutic relevance, members from each of the protease families have been targeted as significant targets for drug discovery. A successful identification of protease activity in disease and its development as a therapeutic target requires an understanding of the *in vitro* chemical mechanism and kinetics of proteolytic activity as well as the complexity of *in vivo* function and biological processes regulating proteolysis. Some of these proteases and the assays used for activity determination and for identification of small-molecule/peptidomimetic modulators are discussed briefly in the following sections.

A. Aspartic Proteases

Aspartyl proteases play an important role in several aspects of our overall health and physiology, including blood pressure (renin), digestion (pepsin and chymosin), and in the maturation of the HIV-1 protease [66]. Human Asp proteases are a small class of proteases with only 15 members, most of which are significant drug targets [67]. The aspartate proteases are monomeric enzymes, except for the HIV protease, which is a homodimer [68]. Most aspartyl proteases can accommodate up to 9-amino acid-long peptides into the active site binding cleft and specifically cleave dipeptide bonds that have hydrophobic residues [67]. The aspartate proteases consist of two domains arising from gene duplication, and the two halves of the enzyme are independent but similarly folded units that move relative to each other [67]. In contrast, the HIV protease consists of two identical subunits which are interconnected with six-stranded antiparallel β-sheets. Most of the amino acid sequences of aspartate proteases are divergent except for the highly conserved catalytic site motif comprising of Asp-Thr-Gly. The amino- and carboxy-terminal domains each contribute one catalytic Asp acid residue to the active site, each of which has different pK value. Peptide bond cleavage occurs by a general acid–base catalytic mechanism. One of the two catalytic Asp residues acts as a general acid and is protonated in the enzyme–substrate complex. The other Asp residue acts as a general base activating a water molecule which then attacks the carbonyl carbon of the amide bond, resulting in the formation of a tetrahedral diol intermediate. Subsequent deprotonation of the hydroxyl group by one of the catalytic aspartates and simultaneous activation of the leaving amine by the other protonated Asp residue ultimately lead to peptide bond cleavage. At least three types of Asp proteases are discussed in the succeeding text.

HIV Protease The HIV genome encodes a single Asp protease responsible for the cleavage of the viral polyprotein, which is an essential step for maturation of infectious viral particles [69]. Interference with the proteolytic activity of HIV protease renders viral particles immature and noninfectious. Consequently, inhibitors of HIV protease are important drug candidates as these compounds reduce the viral load in AIDS patients, and at least 15 anti-HIV drugs are known to act against the proteases. The rapid development of inhibitors of HIV proteases was guided by extensive structural information on the target peptide binding to its active site as well as from fluorescence-based assays [68,70]. The extensive modeling of the cleavage sites in the glycosaminoglycan (gag) and gag–polymerase (pol) precursor proteins in the active site of HIV protease led to the design of HIV protease inhibitor screens. All of the protease inhibitors like saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, and lopinavir in the market for the treatment of HIV infection share a hydroxyethylene core, which mimics the peptide linkage in polyprotein gag–pol, and are uncleavable peptidomimetic competitive substrate analogs for the HIV protease [16]. However, drug–drug interactions, overlapping resistance patterns, and long-term side effects resulted in development of new, nonpeptidic inhibitors like tipranavir and atazanavir based on scaffolds such as 4-hydroxycoumarin, L-mannaric acid, or 4-hydroxy-5,6-dihydro-2-pyrone. In addition to hydrolyzing large precursor polyproteins (gag and gag–pol) *in vivo*, the HIV protease can also hydrolyze much smaller peptides, which allows convenient assay development for HTS campaigns. The naturally processed substrate site in the gag–pol polyprotein for HIV-1 protease is very well characterized and forms the basis for the design of small peptide substrates for homogeneous FRET assays. This substrate SQNYPIVWL or SQNYPIVQ is labeled at the N-terminus with DABSYL and at the C-terminus with EDANS. The cleavage of the peptide with HIV-1 protease resulted in a 40-fold increase in DABSYL fluorescence [71], allowing initial reaction velocities to be determined. The same or related peptides have been tagged with chromophores to develop various other assay formats. Many retroviral proteases cleave bonds adjacent to the prolyl residues (phenylalanine–proline (Phe-Pro) bonds). A colorimetric assay was developed, in which the amino-terminal Pro residue in the product was shown to react with isatin, in the presence of 2-(4-chlorobenzoyl) benzoic acid, an aromatic carboxylic acid catalyst to form a blue product that was quantitated spectrophotometrically [72]. The reaction is discontinuous but simple, sensitive, and specific for routine evaluation of potential protease inhibitors. The N- and C-termini of substrate peptides were protected to make them specific for HIV protease. The substrates Ac-ARALAEA-NH2 and Ac-ATIMMQR-NH2 are cleaved between Leu-Ala and Met-Met, exposing primary amino groups that can be detected in a reaction with 2,4,6-trinitrobenzene sulfonic acid, yielding yellow products.

Renin Renin, an Asp protease, catalyzes the first and the rate-limiting step in an enzymatic cascade controlling hypertension and, unlike other Asp proteases like pepsin and cathepsin D, has high specificity for its only substrate angiotensinogen. Renin is produced by kidney juxtaglomerular cells and released into the circulation following regulated and constitutive pathways [67]. Renin cleaves angiotensinogen

to release the N-terminal decapeptide angiotensin I (Ang I), which is cleaved by ACE, a carboxypeptidase, to release a vasoconstrictor hormone, angiotensin II. A number of compounds based on the natural substrate of renin were found to be potent inhibitors (peptide-based, CGP29287; peptide-like, CPG 38560) *in vitro* but could not be developed as drugs due to poor pharmacokinetic properties. The use of homology models and docking studies using various conformers of CPG38560 allowed substitution of the peptide backbone with a tetrahydroquinoline and, eventually, a phenyl-based scaffold extended by a methoxy group. Derivatives of this phenoxy scaffold led to the development of aliskiren, the third generation of nonpeptidic inhibitors of renin [73]. Plasma Renin Activity (PRA) is a well-established biomarker for assessing the circulating renin levels and to test the activity of antihypertensive agents. Most of the PRA measurements utilize assays to detect Ang I, an unstable decapeptide released when plasma renin cleaves the circulating angiotensinogen. The Ang I peptide is captured and stabilized by specific antibodies. The amount of Ang I is quantified by either radioimmunoassay or enzyme immunoassay [74]. These nonhomogeneous methods, though very specific, are tedious and require large amounts of plasma. Many homogeneous formats have been reported in literature for detection of renin activity by quantifying its product, Ang I. A high-throughput immunoassay based on AlphaScreen (Amplified Luminescent Proximity Homogeneous Assay) technology (Perkin-Elmer) was reported by Cauchon et al. [75]. The AlphaScreen is based on the use of latex-based donor and acceptor beads coated with a hydrogel which provides reactive aldehyde groups for conjugating biomolecules to the bead surface. Donor beads contain a phthalocyanine, a photosensitizer, which converts ambient oxygen to singlet oxygen upon illumination at 680 nm. If acceptor bead is in close proximity, energy is transferred from singlet oxygen to thioxene derivatives within the acceptor beads. This generates chemiluminescence at 370 nm, which activates a fluorophore in the acceptor bead that subsequently emits light. Biotin–Ang I competes with unlabeled Ang I for binding to its antibody, which is immobilized on acceptor beads. The antibody-bound Biotin–Ang I is captured by streptavidin-coated donor beads. The assay was reported to be very sensitive (1–50 nM Ang I), robust, and fast and did not require wash steps. Many FRET-based peptide assays have been developed to monitor renin activity and for the identification of renin inhibitors. The FRET peptide-substrate sequence was derived from the native renin cleavage site in the N-terminal region of angiotensinogen, Ile-His-Pro-Phe-His-Leu-Val-Ile-His-Thr. The octapeptide, labeled with either Mca-Ala/Dnp or EDANS/DABSYL, was used as FRET peptide. Another continuous assay of renin activity was developed using a FRET peptide with a fluorophore 5-FAM and quencher, QXL™ 520 (R-E(EDANS)-IHPFHLVIHT-K(DABCYL)-R, SensoLyte® 520 Renin Assay Kit, AnaSpec). Upon cleavage into two separate fragments by renin, the fluorescence of 5-FAM is recovered and can be monitored at Ex/Em of 490/520 nm. This assay is about 50-fold more sensitive than an EDANS/DABSYL-based assay.

BACE-1 β-Amyloid precursor protein cleaving enzyme-1 (BACE-1, or β-secretase) is a transmembrane protein containing pepsin-like Asp protease activity [76]. BACE-1 is considered the initial and rate-limiting enzyme responsible for the proteolytic

cleavage of amyloid precursor protein (APP) to a membrane-bound APP C-terminal fragment (APPCTFβ or C99), which is further processed by presenilins (γ-secretase) to generate amyloidogenic peptides Aβ40 and Aβ42. The latter peptides constitute the major components of amyloid plaques accumulating in the brain of Alzheimer's disease (AD) patients. BACE-1 is expressed as a preproenzyme, and its maturation involves removal of signal peptide, followed by the cleavage of the prodomain and sorting of BACE-1 by the secretory system to plasma membrane and clusters within lipid rafts. Various inhibitors to BACE-1 have been designed based on the X-ray crystallographic structure, which shows a high degree of homology to that of other human Asp proteases like pepsin and renin. The cleavage site specificity sequences of BACE-1 resides mainly at the S1′ subsite, which requires small side chains such as alanine (Ala), Ser, and Asp. Most of the assays for BACE-1 use short peptides based on an extensively characterized Swedish mutation of the 695 amino acid of the APP isoform (KM to NL), a variant that is a 100-fold better substrate for BACE-1 cleavage compared with the wild-type sequence. The FRET substrate specific for BACE-1, Rh-EV**NL**DAEFKQuencher [77] or EVNLDAEF peptide labeled with EDANS/DABSYL pair, has been used to determine the potency of synthetic BACE inhibitors. After cleavage by BACE, the product (peptide–EDANS) is brightly fluorescent and can be easily analyzed using a fluorescence plate reader or a fluorometer with Ex wavelengths of 335–345 nm and Em wavelengths of 485–510 nm. Highly hydrophobic BACE-1 FRET peptide has low solubility, resulting in their use at much below Km value and low assay sensitivities and use of high enzyme concentrations. Alternate protease substrates like casein–FITC and BPNA–FITC have also been reported for use as BACE-1 substrates for *in vitro* activity assays [78]. Pietrak et al. [79] reported an optimized peptide cleavage sequence, NFEV, which was cleaved by very low picomolar levels of BACE-1 about 10 times more efficiently than was the substrate containing the Swedish (NLDA) sequence [80]. An HTRF-based assay was reported for assaying BACE-1 activity using an 18-amino acid peptide substrate, EuK-APPsw (EuK-KTEEISEVNLDAEFRHDKC-biotin), that was labeled with Eu cryptate at the N-terminus and was biotin labeled at the C-terminus [81]. The streptavidin-coupled cross-linked allophycocyanin (SA-XL665) protein was used as an energy acceptor. In these assays, the EuK-peptide was incubated with recombinant BACE-1 and SA-XL665 protein. The EuK fluorophore is excited at 337 nm, and after a 50 μs delay, the 620 and 665 nm Em were measured. When the peptide is intact, resonance energy transfer occurs between EuK and XL665, resulting in quenched 620 nm fluorescence and increased 665 nm and a high 665/620 nm ratios. FRET is disrupted by cleavage at the scissile bond by BACE-1. Several cell culture models for assaying secretase activity were described in a recent report by Volbracht et al. [76]. In one such model, the BACE-1 activity in HEK293 cells was determined in cells transfected with constructs expressing secreted alkaline phosphatase (SEAP)–APP in which the extracellular domain of APP was substituted with SEAP. Secreted alkaline phosphatase was quantitated in the cell culture supernatants after cleavage of SEAP–APP by BACE-1. The SEAP hydrolyzes *para*-nitrophenyl-phosphate (*p*NPP) to yield phosphate and *p*-nitrophenolate, a yellow chromophore that absorbs at 405 nm. An ELISA-based assay is available for assaying BACE-1 activity in brain cell extracts or cell lysates.

This solid-phase sandwich ELISA uses two kinds of high specific antibodies against BACE-1. The plates coated with one of the anti-BACE-1 antibody (N42) is incubated with cell extracts and, after several washes, is incubated with HRP-labeled BACE-1 antibody (C) (Immuno-Biological Laboratories Co., Ltd). After washes, the bound HRP activity is measured using tetra methyl benzidine (TMB), and the strength of color development is in proportion to the amount of BACE1.

B. Metalloproteases

Matrix metalloproteinases are a family of zinc-dependent endoproteinases secreted by cells and are responsible for much of the turnover, remodeling of components of the ECM and basement membrane. MMPs cleave a peptide bond before a hydrophobic amino acid residue (Leu, Ile, Met, Phe, Tyr) [82]. MMPs are modular proteins which are secreted as inactive proenzymes and contain a prodomain, a catalytic domain, and hemopexin domains. The prodomain harbors a Cys switch motif PRCGXPD, in which the Cys sulfhydryl group interacts with the zinc ion and blocks active site. Removal of the 10 kDa propeptide by proteolysis leads to activation of the enzyme [83]. The catalytic domain contains catalytic zinc which is coordinated by three histidine (His) residues in the active site motif HEXXHXXGXXH. The carbonyl group of the scissile peptide bond coordinates with the zinc atom, displacing water. A glutamate adjacent to the first His in the active site motif serves as a general base to draw a proton from the displaced water molecule on the carbonyl carbon. The S1′ pocket which accommodates the side chain of the substrate residue varies in size among the MMPs and determine the substrate specificity. All MMPs, except MMP-7 and MMP-26, have regulatory α-propeller hemopexin domains, which mediate protein–protein interactions and contribute to substrate specificity. MMP activities are measured using full-length collagens/gelatins, short synthetic linear peptides, or synthetic mini-collagens as substrates [84]. Zymography, based on collagens/gelatin substrates copolymerized in polyacrylamide gels, is the most commonly employed technique for in-gel detection of enzyme, proenzyme forms of MMP activity [85]. Gelatins are used for detection of MMP-2, MMP-9, and to a lower extent MMP-8. Collagen-embedded gels are used for MMP-1 and MMP-13, while casein–SDS-PAGE are the preferred substrate gels for MMP-11 and MMP-1, MMP-3, MMP-7, and MMP-12. Addition of heparin to the sample either at time of gel loading or post-electrophoresis improves detection of very low levels (low ngs to pgs) of MMP-7 (matrilysin) and, for collagenases, MMP-1 and MMP-13, respectively. A number of short peptides derived from MMP sites on collagen were initially designed for fluorescence-based assays: Dnp-Pro-Leu-Gly-Ile-Ala-Gly for MMP-8 and Dnp-Pro-Gln-Gly-Ile-Ala for MMP-9, containing the MMP cleavage site, the Gly-Ile bond. A large number of assays based on the use of FRET substrates [86] have been developed with various donor and quencher combinations (Fig. 8). Most of the linear peptides are labeled at the N-terminus or in the P3 position with Dnp as the quenching group. The donor groups are varied: Mca ((7-methoxycoumarin-4-yl)acetyl) and Nma (*N*-methylanthranilic acid). The peptide substrate, now marketed by Enzo, is extensively used for assaying almost all MMPs of commercial interest. The solubility

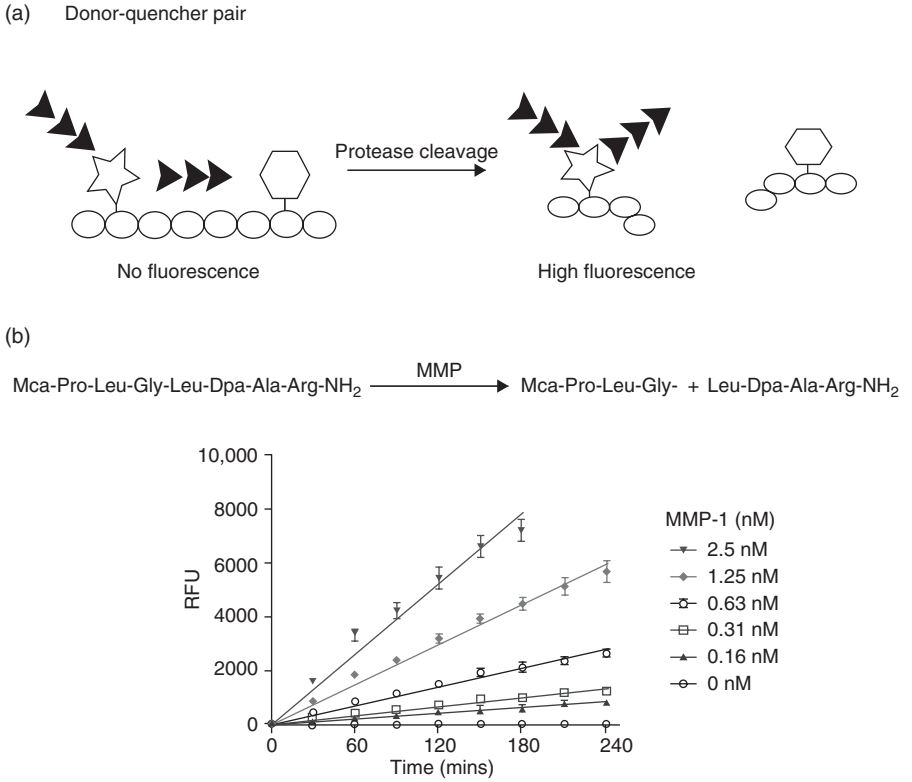

FIGURE 8 Matrix metalloproteinase assay. (a) Donor–quencher pair. Most of the MMP assays use FRET peptides in which the fluorescence from the donor is quenched by the quencher. (b) Detection of MMP-1 activity using an MCA/ Dpa FRET peptide as substrate. The graph shows kinetics of increase in fluorescence of hydrolyzed peptide which is monitored at Ex/Em = 340/490 nm. The increase in luminescence after hydrolysis is shown for different concentrations of MMP-1. (*See insert for color representation of part b.*)

of this peptide was improved by the addition of an N-terminal lysine (Lys). The cleavage by MMP-1, MMP-2, MMP-7, MMP-8, MMP-9, MMP-10, MMP-12, MMP-13, or MMP-14 at the Gly-Leu bond separates the Dpa, resulting in an increase in fluorescence (Fig. 8b). FRET peptides labeled with EDANS/DABSYL as donor–quencher pair are also used as MMP-assay substrates (Dabcyl-Gaba-Pro-Gln-Gly-Leu-Cys-Ala-Lys-NH2) with EDANS conjugated to the side chain of Glu or at the C-terminus of the peptides [87]. A derivative of the nonspecific metalloproteinase inhibitor, actinonin, was used as a universal probe to develop a generic selectivity profiling platform for HTS of MMP inhibitors. Actinonin, which binds to most of the MMP active sites, was used to design a probe for use in FP-based competition assay to obtain a selectivity profiling of inhibitors against a panel of MMP enzymes [88]. A colorimetric assay was developed by Weingarten et al. [23,24] by introducing a

thioester bond at the MMP cleavage site. MMP-mediated hydrolysis of the thiopeptide produced a sulfhydryl group, which reacts with DTNB, Ellman's reagent, to form a colored compound, 2-nitro-5-thiobenzoic acid, detectable at 412 nm. An Enzo kit based on the concept of the thiopeptide chromogenic substrate (Ac-PLG-[2-mercapto-4-methyl-pentanoyl]-LG-OC2H5) allows activity determination on majority of MMP-3, MMP-12, MMP-1, MMP-7, MMP-2, MMP-9, MMP-13, MMP-8, MMP-14, MMP-10, and MMP-20 enzymes. AlphaScreen™-based MMP assay has been developed for monitoring degradation of the substrate aggrecan, an important structural component of joint cartilage [89]. At least four antibodies specific to the carbohydrate side chains of aggrecan were used to create a macromolecular complex whereby aggrecan could form a cross-link between donor and acceptor AlphaScreen detector beads. Digested aggrecan, which failed to form a cross-link, generated no signal, so that inhibitors of the digestion could be detected as a restoration of signal. The assay was rapid and well suited for HTS and gave signal-to-background values around 20:1 and an overall Z′ factor of 0.8.

C. Serine Proteases

Around one-third of the known proteases are Ser proteases, which are predominantly endoproteases. All Ser proteases share common structural features of their active sites: a catalytic triad, a selectivity S1 pocket, a proximal hydrophobic S2 pocket, and a distal hydrophobic S3 pocket. All Ser proteases can be grouped into three types: trypsin-like, which cleave proteins after basic amino acids; chymotrypsin-like, which cleave after hydrophobic amino acids; and elastase-like, which cleave after small hydrophobic amino acids. The hydroxyl group of Ser acts as a nucleophile that attacks the carbonyl moiety of the substrate peptide bond to form an acyl–enzyme intermediate. In addition to Ser, the catalytic machinery requires a proton donor (His) and a third residue, like Asp or another His. The catalytic triad is a hallmark of Ser proteases though many Ser proteases employ a simpler dyad mechanism where Lys (proton donor) or His is paired with the catalytic Ser. Antithrombotic agents like anticoagulants and antiplatelet agents targeting Ser proteases, the coagulation factors thrombin and FXa, are very prominent drugs for cardiovascular disease. Factor Xa catalyzes the hydrolysis of the Arg-Thr and then Arg-Ile bonds in prothrombin to yield active thrombin, with specificity for the recognition sequence Ile-Glu (or Asp)-Gly-Arg-X [90,91]. Several short peptide substrates are available for FXa assays: *N*-benzoyl-Ile-Glu-Gly-Arg-*p*-Na, Boc-Ile-Glu-Gly-Arg-7-amido-4-methylcoumarin (Sigma), methoxycarbonyl-cyclohexylglycyl-Gly-Arg-NHPhNO2 (Spectrozyme Xa), methanesulfonyl-D-Leu-Gly- Arg-NHPhNO2 (CBS 31.39), and Bz-Ile-Glu-(piperidine amide)-Gly-Arg-NHPhNO2 (S2337). Thrombin is the last protease in the coagulation cascade and is activated by a cleavage of its prodomain by FXa. Thrombin triggers clot formation by cleaving N-termini of A and B chains of fibrinogen and releases two corresponding oligopeptides, fibrinopeptides A and B. It also activates factor XIII which cross-links fibrinogen monomers and stabilizes the clot. Activity of thrombin is assayed colorimetrically using chromogenic peptide

substrates like D-Phe-Pip-Arg-NHPhNO2, S-2238 (Chromogenix), Tos-Gly-Pro-Arg-NHPhNO2, Chromozym-TH (Boehringer Mannheim), or AFC thrombin substrate (AnaSpec). Argatroban is an arginine (Arg)-based peptidomimetic small-molecule inhibitor of thrombin.

D. Threonine Proteases: Proteasome Degradation Enzymes

Ubiquitination Threonine proteases are very similar to Ser proteases since the catalytic mechanism is identical, but a Thr is used in place of Ser. Threonine proteases are components of the proteasome, a multi-subunit complex which plays a central role in ubiquitin (Ub)-mediated degradation of Ub-tagged cytoplasmic proteins. Proteasome activity regulates cell cycle progression, apoptosis, and inflammation and has a basic housekeeping function of removing damaged or misfolded proteins. Proteasome subunits are upregulated in many cancers, and proteasome inhibitors are known to block cell cycle proliferation. The proteasome has 7α and 7β subunits (proteolytic chamber), and the core is made up of six proteolytic sites with three distinct substrate specificities [92,93]. The catalytic activity in β5-subunits of the proteasome has chymotrypsin-like activity and can hydrolyze bonds on the carboxyl side of Tyr or Phe at the peptide carbonyl or P1 position. The β2-subunits harbor trypsin-like activity with preference for basic amino acids Arg or Lys at the P1 site. The β1-subunits have caspase-like activity with preference for glutamate or acidic (peptidyl-glutamyl peptide bond hydrolyzing) amino acid residue [94,95]. The proteasome uses the N-terminal Thr of its catalytic β-subunit as a nucleophile for peptide bond hydrolysis. The cleavage of the propeptide upon activation exposed Thr as the N-terminal residue. Misfolded or other target proteins targeted for degradation are first acted upon by ubiquitination enzymes which add polyubiquitin chains to internal Lys. The polyubiquitin chains interact with the degradation signals such as misfolded domains, small motifs, or posttranslational signals. The N-end rule and ubiquitin fusion degradation (UFD) signals [96,97] are modular domains that were inserted in GFP [98] for developing fluorescence-based *in vitro* assays for monitoring proteasome activity (Fig. 9). The presence of one or several degradation (UFD) signals fused to GFP reporter, along with an N-terminal Ub moiety as the acceptor for polyubiquitin chains, shortens the half-life since the proteasome degrades the GFP, resulting in loss of fluorescence. The treatment of UFD–GFP-expressing cells with proteasome inhibitors led to stabilization of the GFP reporter with a resultant up to a 1000-fold increase in fluorescent intensity [35]. This increase correlated with cell cycle arrest and apoptosis, which is an established result of proteasome activity arrest. The fact that these fluorescent proteins follow the same pathway as endogenous substrates allows functional analysis of the Ub–proteasome system in cells. The UFD–GFP system was also used for *in vivo* analysis of the Ub–proteasome system in mouse strains transgenic for (GFP) reporter fused to constitutively active degradation signal [99]. *In vitro* assays for proteasome activity have been based on utilization of fluorescently AMC-tagged tri-/tetrapeptides (Table 2). Substrates like LLVY-AMC are also acted upon by other nonspecific protease activities in whole cell

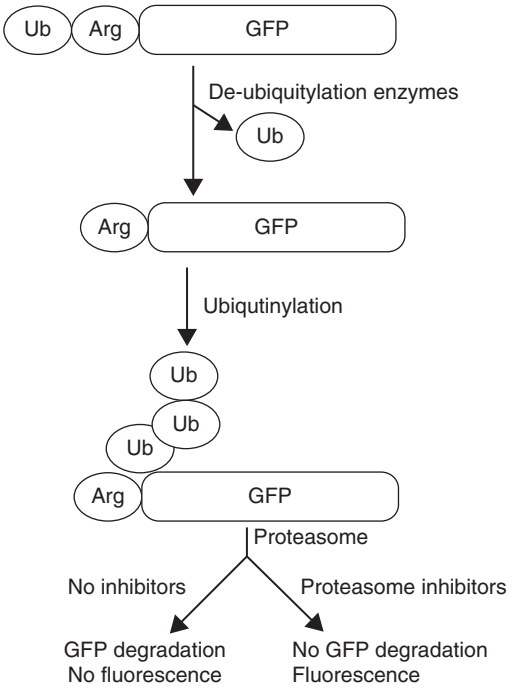

FIGURE 9 Proteasome assay using GFP. The Ub at the amino-terminus of the fusion protein is removed by deubiquitinating enzymes, thus exposing the N-terminal Arg residue. The Arg residue promotes ubiquitination and proteasome-mediated degradation of GFP.

lysates [100] but are specific for use in assaying circulating proteasome activity in the blood of chronic lymphocytic leukemia patients [101]. FRET peptides have been used to assay proteasome activity *in vitro*. The FRET peptides contain donor fluorescein and acceptor tetramethylrhodamine (TMR), which have spectral overlaps and are separated by six amino acids. In the undegraded peptide, Ex of fluorescein results in energy transfer to TMR and measurable acceptor fluorescence is obtained. After proteasome-mediated degradation, the donor and acceptor pair separates accompanied with an increased Em from the donor and a concomitant loss of acceptor fluorescence. The fluorescent peptides that contain a donor and a quencher have also been used for assaying proteasome activity. Cleavage of such peptides by proteasome separates the donor and quencher with a resultant increase in fluorescence of the donor over very low backgrounds. The small peptides can also be injected into cells and proteolysis followed by imaging techniques. Since the fluorescent assays are less sensitive than bioluminescent assays, many luminogenic substrates have been developed by Promega Corp. for testing proteasome activity in homogeneous bioluminescent assays. The Suc-LLVY-aminoluciferin, Z-LRR-aminoluciferin, and Z-norleucineLPnLD-aminoluciferin have been used to monitor chymotrypsin-like, trypsin-like, and caspase-like activities of proteasome. The rate of proteasome

cleavage of the substrate was found to be equal to the rate of luciferase utilization of the released aminoluciferin [92].

E. Cystine Proteases

Cysteine proteases use a catalytic dyad comprising of a thiolate ion in active site Cys as a nucleophile which is activated by imidazole group of a His (proton donor). The Cys proteases stabilize an oxyanion of tetrahedral transition state via hydrogen bond interactions with active sites Cys and glycine (Gly). Cys proteases are grouped into six clans based on their structure of the active site and include calpain families as well as Ub-processing peptidases, the cytosolic cysteinyl aspartate-specific proteases, or caspases. Caspases are Cys proteases responsible for proteolysis associated with programmed cell death or apoptosis, whereas lysosomal cathepsins are involved in protein degradation.

Caspases Caspases are a family of cytosolic Cys proteases involved in the initiation and execution of apoptosis by targeting a discrete set of proteins for proteolysis which occurs after an Asp residue [102]. The caspase family in mammals has at least 15 members, which are either inflammatory (caspase-1, caspase-4, caspase-5, caspase-12, caspase-13, and caspase-14) or apoptotic caspases. The apoptotic caspases are further characterized as either initiator caspases (caspase-2, caspase-8, caspase-9, and caspase-10) or effector caspases (caspase-3, caspase-6, and caspase-7). Caspases are present as inactive proenzymes, most of which are activated by proteolytic cleavage or via autoproteolysis at the Asp-X bonds. X-ray crystallography studies and sequence analysis have shown that the caspases are modular with a 17–21 kDa catalytic domain (p20), a 10–13 kDa small catalytic domain (p10), and a 3–24 kDa amino-terminal prodomain (death domain). Both the large (p20) and the small (p10) catalytic subunits contribute residues to form substrate binding sites (S4, S3, S2, S1, S1′). All caspases show an absolute requirement for the aspartate residue in position P1 for catalysis [103]. However, the substrate specificity of human caspases is determined by their preferences for residues at the P4 site. The group I pro-inflammatory caspases (caspase-1, caspase-4, caspase-5, and caspase-13) prefer large aromatic/hydrophobic amino acids in P4, group II initiator caspases (caspase-2, caspase-8, caspase-9, and caspase-10) require Asp at P4, and the group III apoptotic caspases (caspase-3, caspase-6, and caspase-7) prefer branched-chain aliphatic amino acids at P4 [103]. The general cleavage site for all caspases is X-Glu-X-Asp: caspase-1, caspase-4, and caspase-5 prefer WEHD; caspase-2, caspase-3, and caspase-7 cleave DEXD; and caspase-6, caspase-8, and caspase-9 cleave specifically at (L/V)EXD [103]. Caspase activity measurement is widely used in apoptosis, and numerous *in vitro* and cell-based assay methods are available. In cell line apoptotic models, the relative abundance and activity of caspases in a cell actually dictates which activity is being measured primarily in the cell lysates [104]. It has been reported that caspase-3 is most active for many nonspecific substrates and is also the most abundant caspase in cell lysates [103]. The activity of initiator caspases is minimal in dilute cell extracts due to dimer dissociation. As a result, most of the activity measured in

cell lysates is due to one or more caspases and does not allow specifically associating a cleavage event with any one caspase. The short tetrapeptide substrates are more relevant in the context of purified recombinant caspases and for determination of enzyme kinetic parameters. Recombinant caspase-2, caspase-3, caspase-6, and caspase-7 exist as active dimers in solution, whereas the recombinant caspase-8 and caspase-9 exist as mixtures of inactive monomers and active dimers in solution [105]. Various kosmotropic salts like sodium or ammonium citrate or dithiothreitol are added to the buffers to stabilize these protein mixtures as dimers or change protein conformation to increase substrate binding and decrease the Km values. Caspase activity is measured with homogeneous platforms using labeled peptides and in cell FRET assays [104,106]. Although most caspases have overlapping substrate specificities, a number of assay kits are available from various vendors that use different peptides for different members of caspase family. The various tetrapeptide sequences have been defined as optimal for different caspases: DEVD is used for assays with caspase-3 and caspase-7, YVAD for caspase-1, VDVAD for caspase-2, LEVD for caspase-4, WEHD for caspase-5, VEID for caspase-6, IETD for caspase-8, LEHD for caspase-9, AEVD for caspase-10, and ATAD for caspase-10. The peptides are labeled with chromogenic group like *p*-NA for colorimetric assays; fluorofluors like AFC, AMC, EDANS, and MCA for fluorescence assays; or aminoluciferin for bioluminescent assays (Fig. 10). Luminescence-based assays developed by Promega Corp. utilize the same tetrapeptide sequences for conjugation with aminoluciferin. The peptide is solubilized in a buffer containing luciferase, $MgSO_4$, and ATP. The purified recombinant caspases or caspase mixtures from lysates of cells undergoing apoptosis are added to the mix and luminescence is measured.

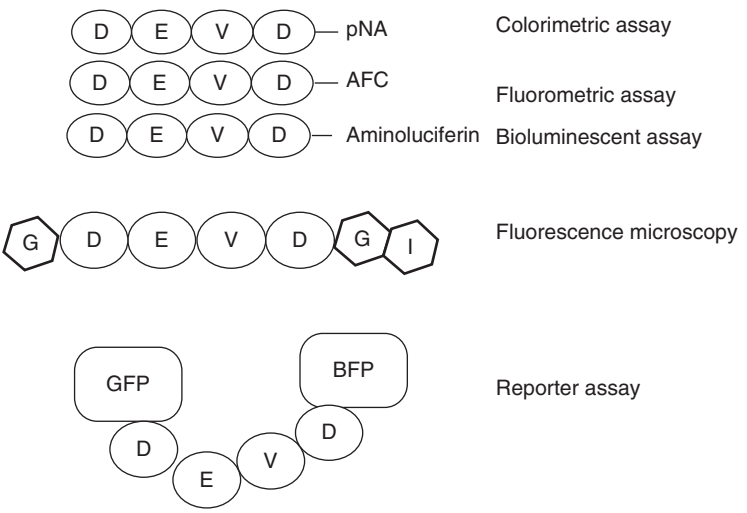

FIGURE 10 Substrates for caspase assay. Various conjugations with chromogenic, fluorescent, and luminescence tags to the same peptide generate substrates for all major assay formats.

Cathepsins Many cathepsins are lysosomal proteases that play an important role in antigen presentation, tissue remodeling, and enzyme activation/inactivation [107]. Upregulation of one or more cathepsins has been implicated in cancer, AD, and rheumatoid arthritis. Structurally, cathepsins A and G are members of the Ser peptidase family, cathepsins D and E are Asp proteases, and cathepsins B, C (DPPI), F, L, S, H, O, V, X, W, and K belong to the Cys protease family [108–110]. All cathepsins belong to papain family and are synthesized as inactive zymogens. Lysosomal Cys proteinases are active in weakly acidic environment, and the substrate binds in an extended conformation with specificity defined by S2 and S1′ in the active sites. Cathepsin B, which has been studied most extensively, has both endo- and exopeptidase activities and has less activity for substrates with Pro or Arg at P1′ [110]. Figure 11 lists minimal peptide substrates for assaying some of the cathepsins. The commonly used formats include fluorescence intensity measurements using short peptides labeled with AFC or AMC. In addition, many Mca/Dnp FRET peptides are available commercially (AnaSpec) for measurement of cathepsin activity. Cathepsin B is assayed using substrates like Z-Arg-Arg-AFC or AMC, as cathepsin B specifically cleaves the Arg-fluorophore bond in the fluorometric assay. Cathepsins H, L, and S show very little activity against these peptides. Cathepsin B also preferentially cleaves Z-Phe-Arg-AFC at pH 7.5–8. The substrate specificity for cathepsin H has not been clearly defined. Cathepsin C, an aminodipeptidase which requires halide ions for optimal activity, cleaves dipeptides from N-terminal of peptides and is inactive for substrates with Arg, Lys, or Pro at N-terminal positions. Cathepsin L has broad substrate tolerance and prefers substrates with hydrophobic amino acids in P2 and P3 residues [110]. Cathepsin L is active at pH 3–6.5 in the presence of thiol compounds. A generic benzyloxycarbonyl-protected dipeptide Leu-Arg-AMC (Peptide Institute) is used as a fluorogenic substrate for monitoring activity of cathepsins B, S, K, and V using Ex/Em wavelengths of 355/465 nm. Longer peptides like Mca-Gly-Lys-Pro-Ile-Leu-Phe-Phe-Arg-Leu-Lys(Dnp)-D-Arg-NH2 are substrates for cathepsin D/E. Some of the cathepsins have been targeted for drug development. Cathepsin C or dipeptidyl protease I (DPPI) activity is critical for differentiation of precursor promyelocytes into mature neutrophils as DPPI mediates production of neutrophil elastase, proteinase 3, and cathepsin G, which are potential drug targets for the treatment of chronic obstructive pulmonary disease (COPD). In most cases, cathepsin

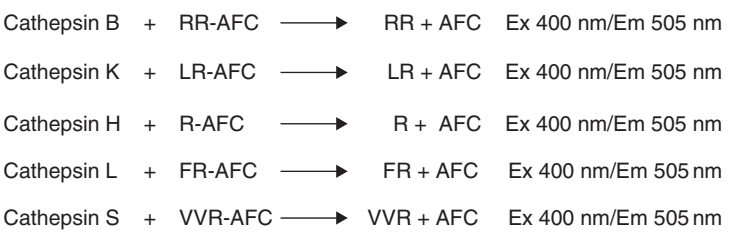

FIGURE 11 Substrates for cathepsin assay. The preferred dipeptide sequence for cathepsin assays is linked to the fluorescent group to generate substrates for fluorescence intensity assays.

C activity has been measured using fluorescent substrates like Gly-Phe-AFC. Cathepsin K, a key enzyme in bone resorption, is a new therapeutic target for the treatment of osteoporosis. Odanacatib (Merck) is a powerful, reversible nonpeptidic biaryl inhibitor of cathepsin K which was highly selective over cathepsins B, L, and S in enzyme assays. Based on structure–activity data, a series of epoxysuccinate derivatives have been reported as Cys cathepsin-specific inhibitors [111].

IV. COMPUTATIONAL APPROACHES FOR PROTEASE IDENTIFICATION AND CHARACTERIZATION

Just as the analytical technologies required in the identification and characterization of proteases and their function are diverse, a substantial variety of different computational tools provide complementary insight. In silico methods can be applied to *de novo* prediction of biochemical structure and function, but their greatest power is found when they are used in concert with laboratory analyses. Key contributions include helping to interpret or deconvolute complex data, to extrapolate observations from a limited set of systems toward a more general species manifold, and to assimilate observations made from multiple different analytical platforms. Such analysis helps to address numerous specific questions about proteases, including:

- Which of the myriad of characterized and postulated proteins exhibit proteolytic activity?
- What type of catalytic function is any given protease likely to support?
- What substrate (or range of substrates) can a given protease process?
- What is the proteolytic mechanism of action?
- What types of chemicals are likely to serve as competitive inhibitors?
- Are there foreseeable mechanisms for noncompetitive inhibition or agonist effects?
- If so, which chemicals are candidates for accomplishing such modulation?

The importance of having computational tools to address toward such questions is well established. For example, fewer than 4% of putative enzymes currently have experimentally characterized catalytic sites or binding data [112], but for those that have experimentally characterized crystallographic or NMR structure, it is possible to computationally perceive specific cavities that have a size and shape potentially suitable for ligand binding [113]. Furthermore, a much larger number of proteins have been sequenced and annotated within large systematic sequence databases, and a reasonable fraction have been at least qualitatively characterized in terms of general protein family and general biochemical function. Cross-references between sequence and function have enabled bioinformatics discovery of plausible sequence motifs (including either phylogenetically conserved sequence patterns or rigorously conserved sequence substrings) that can be used to identify with reasonable accuracy which proteins are likely to support several of the known enzymatic functions [114].

Consequently today there are many examples of specific sequence substrings that are indicative (with varying degrees of certainty) of a given enzyme family. The most comprehensive collection of such motifs is probably housed within the PROSITE database (PROSITE: Database of protein domains, families and functions; http://www.expasy.ch/prosite/). Querying a sequence in PROSITE gives one a reasonable first-pass prediction of whether a given protein has a significant chance of belonging to a known enzyme family and, if so, which class would it belong to. Such analysis also identifies the specific amino acid sequence strings on which the protein family class assignment was made. A specific motif may either be labeled as an *active site*, in which case the constituent amino acids are expected to play a direct role in enzyme catalytic function (information that can be useful to elucidating the mechanism of action), or as a *signature*, in which case the amino acids may still play an important role in function but they may be somewhat spatially peripheral to the active site or have indirect implications toward the reactive mechanism (e.g., perhaps facilitating substrate admission to the active site or fixing it in place for reaction or expedite product evacuation). In general one can assume that enzymes within a certain class will accept substrate species that all share some similar chemical functionality. In other words, proteases are expected to admit amino acid chains for processing, whereas phosphatases should be selective toward phosphate-containing substrates, kinases act on species that are capable of receiving a phosphate group, hydrolases process ligands that can be made susceptible to aqueous hydrolysis, etc. Identification of a key sequence motif can thus be used to project not only that a given protein is likely to be an enzyme but also that it is likely to process some members of given chemotype, for example, esters, phosphates, lipids, and specific amino acids. In general, experimental characterization or computational prediction of the precise range of substrates that a given enzyme will process (e.g., alkyl vs. aryl esters, ATP vs. GTP) and the exact nature of the enzyme functionality/mechanism can be significantly more challenging. Some members of the protease family have unusually consistent function (e.g., trypsin, which is capable of digesting an immense variety of different peptides as long as they have at least one nonterminal Lys or Arg residue), while others are much more limited in their scope of action. The fact that there can be a tremendous amount of variation in the extent of observed substrate specificity from one enzyme family to another, and even within some specific families, makes it difficult to confidently state whether or not a given substrate-like molecule will or will not be processed by a given enzyme unless that molecule has been subjected to direct assay in the presence of that enzyme. This challenge has not prevented the computational community from attempting to address the issue however. Since a key component of enzyme specificity is determined by the three-dimensional structure of the active site cavity, it is very helpful to have a plausible spatial receptor model that can guide a computational assessment of whether a given chemical has a size and shape that is amenable to occupying the active site. Chromotrypsin, for example, has a sizeable hydrophobic pocket located strategically close to the catalytically active Ser residue. This location strongly favors the binding of lipophilic amino acids such as tryptophan, Phe, and Tyr, thus leading to reasonably specific cleavage points adjacent to such residues. This specificity contrasts with elastase, which preferentially

cleaves on the carboxy side of much smaller neutral residues such as Gly, Ala, and valine. Such size- and shape-dependent specificity is not always sufficient evidence to proclaim that a given molecule will definitely be an enzyme–substrate, but the converse can be stated more confidently: a molecule that is too large, or has the wrong shape, to occupy an active site will very likely not be a substrate. Applying these criteria toward an understanding of substrate limits of specific proteases has straightforward prerequisites: a reasonable representation of the three-dimensional receptor structure (i.e., Cartesian spatial coordinates and elemental types for all enzyme atoms) and a computational algorithm for sampling and energetically assessing conformations of prospective substrate species interacting with the receptor. The best enzyme receptor spatial models will obviously be those that have been experimentally resolved via crystallographic or NMR studies. Such species constitute a fairly minor fraction of the total population of known or putative enzymes and are often those about which substrate activity and specificity are already well characterized. One computational tool, comparative protein modeling, is particularly helpful in extending knowledge on this limited subset base toward a broader range of analogous species. The guidelines and protocols for use of comparative modeling in producing spatial models of putative enzymes are surveyed elsewhere [115], but it should be mentioned that in order to generate a model capable of supporting high-confidence structure–activity assessments, it is generally assumed that you need an experimentally resolved template structure whose amino acid sequence is at least 70% identical to the putative enzyme of interest, with minimal sequence gaps between target and template. Computational studies on species that do not have a template with that high degree of homology may still be possible (and are frequently pursued) but must proceed with the caveat of lower confidence in their resulting analytical and predictive capacity. Comparative models based on templates with less than 30% sequence identity to the target may theoretically yield plausible structures, but the chances that they will offer meaningful structure–activity guidance are considered to be thin.

There are numerous molecular docking software programs capable of computationally evaluating the capacity of small-molecule or peptidic ligands (including prospective substrate species) to access the enzyme receptor model. A comprehensive review of these programs is beyond the scope of this chapter, but it is reasonable to assert that most of them are theoretically capable of addressing the relatively simple-minded negative test we specified earlier (i.e., whether one can show that the putative substrate does not have a suitable shape and size for accessing the receptor model site). In order to approach such a test rigorously, however, one may wish to modify the docking approach from the default settings which are typically geared toward effecting the converse (i.e., positive) test for evaluating the probability that a molecule *will* occupy a receptor. If calculations based on the default settings do lead to conformers docked within the enzyme active site cavity, then the test is over in that you have failed to rule out the possibility that your molecule is a substrate and have successfully suggested that it might be. If the default settings do not report a suitable docked conformer, however, then a rigorous negative test will require more stringent conditions. Such a test can be achieved in a number of ways, with the simplest being to give the molecule more

opportunities to find a cavity-centered pose, that is, by simply requesting a larger number (≥100) pose trials. Many docking programs also provide some mechanism for biasing the pose toward a specific receptor location, for example, by specifying a specific H-bonding or hydrophobic interaction within the receptor that must be satisfied in order for a pose to be considered or by permitting one to pre-position the ligand directly within the cavity and test to see if a subsequent pose evolution retains the molecule in this cavity instead of preferentially diffusing outward.

If molecular docking can be used for a negative test to rule out species as possible substrates, it is natural to ask whether one can use similar methods to accurately quantify the likelihood that the species actually is a substrate. The simple answer is no. Molecular docking can provide some insight into whether a given ligand might be a viable enzyme–substrate, but the task of predicting reactive viability or quantitative catalytic kinetics requires treatment of multiple underlying factors, for which docking alone cannot fully account. In the broadest enzymological sense, key factors include:

1. The relative amount of attraction between potentially relevant enzyme peripheral sites and specific substrate candidates (best quantified as a free energy of interaction that accounts for the fact that many enzymes electrostatically or lipophilically attract substrate molecules to a region close to their active site, from which then can recruit specific molecules into their active site for catalytic processing).
2. The suitability of substrate candidates for progressing from a peripheral location into the active site (kinetic transport rate constant; note that some enzymes rely on passive substrate diffusion, while others engage in proactive dynamics that help to guide substrate into position for catalysis).
3. The capacity of a substrate molecule within the active site to orient itself in a position suitable for catalysis (measured by free energy of interaction).
4. The propensity for covalent reaction between enzyme and substrate, leading to the desired catalytic products (determined by the activation barrier computed as an enthalpy difference between the reactant species and transition states progressing toward and relevant intermediate states and the final product state).
5. The capacity of the products to evacuate the enzyme, leaving it free to process additional substrate species (again, quantified as a transport rate constant). In truth, the dominant factor underlying the mechanism by which some proteolytic substrates access their intended catalytic site may be passive diffusion, in which case some of these considerations would be superfluous. However, for the sake of more sophisticated proteolytic systems with tangible intra-enzyme coupling effects, it is worth examining more generalizable analytical techniques.

Molecular docking techniques are very well designed for addressing factor 1. In terms of a compromise between reasonable prediction accuracy and reasonable computational efficiency, they are very likely the best currently available tools for this step, offering a major savings in computational expense relative to the potentially more quantitatively accurate option of molecular dynamics simulations. Most

molecular docking methods are not suited for the probing transport kinetics underlying factors 2 and 5, because they rely on algorithmic shortcuts (e.g., random positional sampling and fragment-based searches to identify a binding site). Programs such as AutoDock [116] that treat ligands as whole molecules whose translational, rotational, and torsional coordinates are subjected to pseudodirectional variation as dictated by energy-weighted stochastic sampling techniques (e.g., genetic algorithms, Monte Carlo) may be able to shed some insight into the relative propensity of different ligands to effect such transport in that the programs can be trained to yield a trajectory that semi-realistically traces the progress of a ligand through a macromolecular environment [117], but we are unaware of any systematic validation of whether this approach can quantitatively reproduce transport-related factors in protease kinetics. Molecular dynamics simulations represent a more obvious rigorous avenue for transport kinetics modeling, although one encounters the practical problem of how to bridge the time scale gap between practical simulations (fully atomistic MD simulations of proteins are currently limited by computational expense to time scales on the order <1.0 μS) and real applications (depending on the specific enzyme, transport processes may take orders of magnitude more time). One reasonable option for semiquantitative estimates of transport rate constants is statistical extrapolation of MD trajectories. As an illustration of this extrapolation, we performed a molecular dynamics simulation [118] of the acetylcholine substrate interacting with the active site gorge of *Torpedo californica* acetylcholinesterase, a Ser hydrolase enzyme closely functionally related to Ser proteases, modeled from crystal structure [119] using the AMBER suite (AMBER charges and force field [120], fully explicit TIP3P solvent model [121] periodicity modeled via Ewald sums, 20 ps volumetric equilibration followed by 20 ps thermal equilibration and 1.0 nS analysis) to probe the time scale within which the reactive ligand ester carbonyl C would approach within a covalent-bonding distance ($R_{CO} < 1.5$ Å) of the catalytic Ser200 side chain oxygen atom. Figure 12 reports a statistical sampling of the R_{CO} distance recorded every 100 fs over the course of the 1.0 nS analysis run. Visual inspection of the plot suggests a biphasic distribution, for which each phase corresponds fairly nicely to a quadratic curve fit. Taking into account the logarithmic scaling of the abscissa, this curve suggests that the simulation yielded two main conformational families, both of which displayed approximately normal R_{CO} distributions. The more populous conformational family further to the right in this figure is a reasonable representation of the positional distribution experienced by the ligand as it strives to transit from a peripherally bound location near the active site before ultimately penetrating into the catalytic receptor region (smaller leftmost conformational family). The quadratic curve fit to the rightmost distribution yields a logarithmic frequency expression that can be expressed as $\ln[F(R_{CO})] = -4.06R_{CO}^2 + 41.62R_{CO} - 20.78$. Assuming that the barrier between this conformational family and the one represented by shorter R_{CO} distances occurs approximately at $R_{CO} = 4.0$ Å, this expression yields a value of $\ln[F(R_{CO})] = 0.35$, which translates to a frequency of 1.41/nS, that is, the molecule would be expected to pass this barrier en route to accessing the catalytic site approximately 1.4 times per nS, which is qualitatively reasonable since we observed such a transit to occur only once during the underlying 1.0 nS simulation. The primary

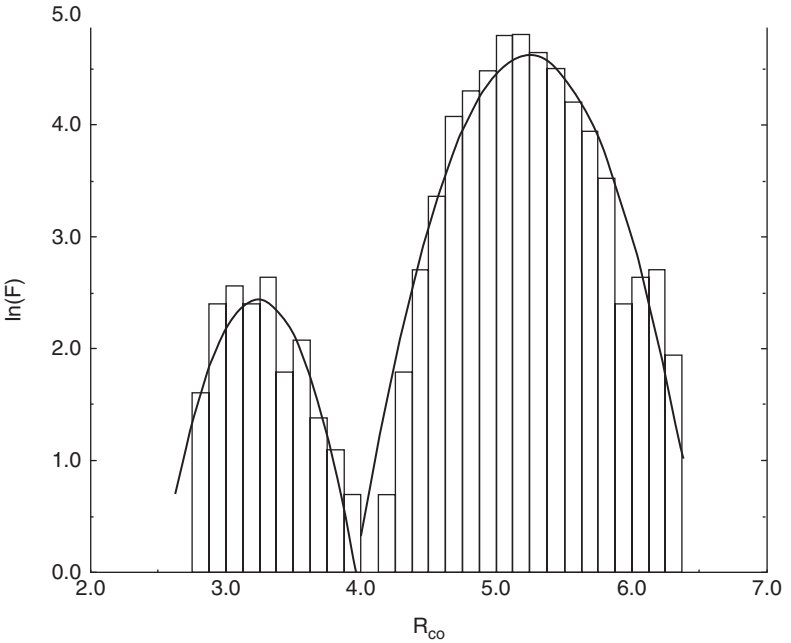

FIGURE 12 Logarithmic frequency of the distance between acetylcholine reactive C to acetylcholinesterase catalytic Ser 200 O during a 1 nS MD simulation.

drawback for such calculations is that even a relatively modest time scale such as 1 nS requires a significant investment in computer time (several days of CPU time on a reasonably efficient processor circa 2009) and thus is not amenable to high-throughput processing of a diverse range of prospective substrate candidates. Given a reasonable level of quantitative reliability in carefully constructed molecular dynamics calculations, however, it may be possible to perform simulations on a moderate number of prospective substrate systems and use the results as the basis for statistically training a quantitative structure–property relationship (QSPR) model to reproduce MD-derived transport rate constants. Some suitable descriptors for such QSPR models might entail volumetric and surfacial properties such as VolSurf metrics [122] (broadly used for transport and permeability properties), as well as terms describing intra-chain interactions that can impact peptide molecular flexibility.

The MD simulation described earlier also lends itself very well to redress of factor 3 (assessment of the capacity of a substrate molecule to orient itself within the active site in a conformation suitable for covalent reaction) in that the leftmost curve in Figure 12 tracks the conformationally dictated distribution of R_{CO} distances fairly close to the covalent limit (~1.5 Å), essentially representing substrate binding modes that are possible precursors to the Michaelis complex. This latter curve yields a logarithmic frequency quantified as $\ln(R_{CO}) = -3.45R_{CO^2} - 0.33R_{CO} + 3.96$. By this relationship, the incidence rate for $R_{CO} < 1.5$ Å separation is predicted to be ~0.1/nS (i.e., once every 10 nS), which can be assumed to provide a rough estimate for the rate

at which substrate will approach the reactive Ser200 oxygen closely enough to form a covalent bond. This rate is obviously a coarse approximation since the classical MD force field represents the atoms as fairly firm spheres governed by repulsive van der Waals potentials, while more realistic close contacts of reactive species should exhibit polarized electron densities that are often more permissive toward further approach. However, within a family of prospective substrate molecules with similar core electronic structure, the errors should be systematic; thus, it is reasonable to expect that the trends dictated strictly by conformational access to the Michaelis complex should correlate fairly well with observed substrate reactivity. It should also be noted that MD simulations such as these can yield quite accurate interaction free-energy prediction via rigorous analysis of variations in the computed classical electrostatic and van der Waals enthalpy terms between the ligand and receptor (e.g., [123]).

Although MD simulations are clearly superior to molecular docking for the purpose of addressing transport-related factors 2 and 5, it is not clear whether this advantageous performance is also observed for prediction of molecular propensity for forming the Michaelis complex. Many conventional docking programs provide a suitable medium for probing the capacity of a putative substrate molecule to orient itself within the active site cavity in a position susceptible to covalent interaction with the receptor and for gauging the relative stability of covalently bound enzyme–substrate intermediates. Essentially all small-molecule docking programs support the generation of precovalent enzyme–ligand Michaelis interaction complexes, and a selection of docking programs (e.g., AutoDock [116], FlexX [124], GOLD [125], and MacDOCK [126]) enables the projection of covalently bound complexes based on specification of one ligand atom that must bind to one particular receptor atom. While such calculations do not take into account the covalent bond-formation energy, they otherwise enable qualitative assessment of whether a given putative substrate has the right shape and electrostatic profile for effecting a covalently bound complex and enable comparative evaluation of enthalpy trends across a series of putative substrates. Such docking experiments cannot yield quantitative kinetic rate constants in the manner of MD simulations; however, their score models have been trained (via a QSPR-like process across diverse sets of ligands and receptors) to reproduce binding free-energy trends. One possible drawback with these models is that they are primarily geared toward treating protein inhibitor or protein agonist interactions, and at present we are unaware of rigorous validation of such score models for reproducing trends among enzyme–substrate complexes.

Computationally quantifying the covalent activated enzyme–substrate complex kinetics underlying factor 4 is most rigorously achieved via quantum chemical transition state calculations [127]. While these methods can be quite accurate and are a tempting tool for the pursuit of insight into specific protease–substrate pairings, an important caveat for such approaches is that they are very computationally demanding and low throughput. As can be surmised from the earlier discussion, computationally derived insight into the specific factors underlying the process of protease function is challenging to achieve. Furthermore, since many of the specific factors are also difficult to experimentally quantify, one frequently lacks quality analytical points of validation with which to evaluate specific models. By sacrificing fundamental detail,

however, it is nonetheless possible to obtain fairly reliably predictive computational models of enzyme kinetics by focusing directly on readily measurable rate constants such as the Michaelis constant, K_M, that reflect the overall substrate processing capacity of the enzyme. Such observables can serve as the basis for the successful development of general, computationally efficient quantitative structure–activity relationship (QSAR) models (e.g., [128–130]). Constructing a reasonable model that accounts for substrate transport, Michaelis complex formation, and covalent reactivity requires a flexible basis of QSAR descriptors. Highly generalized descriptors such as the VolSurf parameters are very useful for such a requirement, although many instances with significant variation in the electronic structures of the reactive core may also require quantification of substrate atomic charge, bond strength, and polarizability terms via quantum chemical calculations. In cases where protease–substrate specificity appears to hinge on subtle aspects of the ligand–receptor interaction, the comparative binding energy (COMBINE) method [131,132] could provide important additional structure–activity insight. QSAR is a potentially very powerful tool for efficient and reliable prediction and characterization of enzyme–substrate processes, but nonetheless has fundamental limitations in practice. The predictive value of QSAR models can vary greatly as a function of many technical aspects relating to the manner in which they were trained. The general level of predictivity of QSAR models can be gauged through assessment of the statistical significance of the model and by performing cross-validation and blind-test studies using the model; however, an important additional caveat arises in assessing the scope of the model, that is, whether a model that has been trained based on one set of molecules is suitable for predicting the behavior of a distinct set of molecules. The general rule is that one is not advised to attempt a QSAR-based prediction on a molecule for which there are no reasonably similar analogs within the model training set, but at present we are unaware of studies that attempt to rigorously quantify how many analogs should be present in a training set and what is a reasonable similarity threshold by which to consider a given training set molecule to be an analog.

Most of the methods discussed this far with potential relevance to understanding protease catalytic function, kinetics, and substrate selectivity are equally applicable to computational assessment or design of ligand-based protease modulation and inhibition. Modeling studies on the latter are actually far more common than those applied toward assessment and prediction of substrate kinetics, perhaps in part due to the substantially greater availability of quantitative inhibition and modulation kinetics data. We can thus reprise the general list of key factors underlying enzyme kinetics discussed earlier, except this time in the context of non-substrate interactions:

1. Peripheral site interactions quantified as ligand–receptor interaction free energies are likely to be as important for quantifying non-substrate interactions.
2. Ligand transport from peripheral sites into the active site (for inhibitors) or adjunct subsites (modulators or noncompetitive inhibitors) is likely to be as qualitatively important.
3. The interaction free energy for a substrate molecule binding to the active site or adjunct subsite remains a key component.

4. The propensity for ligand–receptor covalent binding is only relevant in those cases (generally in the minority) when the inhibitor or modulator acts according to a covalent binding mode.
5. Ligand evacuation from the receptor can sometimes (although need not always) be important in inhibitor and modulator design or characterization.

The aforementioned considerations are largely sufficient for understanding the fundamental basis for competitive inhibition of proteases. Sometimes competitive inhibitors act by complexing with a peripheral or intermediate site in such a way as to bar access by substrate species to points further along (in particular the catalytic active site). In such cases the inhibitor mechanism of action is dictated largely by factor 1 and perhaps also by factor 2 (in scenarios where the inhibitor transits part of the way toward the active site). More often, however, the inhibitor possesses physical attributes adequately similar to those of substrate molecules so as to permit it to occupy the active site, but not so similar as to allow it to be catalytically processed and expelled in the same manner as a substrate. In this case the key actions by the inhibitor are determined by factors 1–3 and possibly also (in the case of a covalent inhibitor) by factor 4. The only other significant modeling considerations that may need to be entertained arise from the question of how the inhibitor is delivered into the general vicinity of the enzyme. In the case of *in vivo* delivery, it may thus be important to take into account the full range of absorption, distribution, and metabolic properties of the prospective ligand. Such properties are generally modeled quite well for standard drug-like and natural product-like compounds via established QSAR/QSPR models that have been assembled for drug design. The aforementioned VolSurf suite [122], for example, provides well-validated models for ligand solubility, albumin binding, blood–brain barrier permeation, volume distribution, cell permeation, etc.

In silico characterization and prediction of modulator and noncompetitive inhibitor kinetics pose a somewhat greater challenge in that the mechanism by which they affect enzyme function can be more complex than just physically blocking access by the substrate to its corresponding active site. Some cases of noncompetitive inhibition entail cases where inhibitors occupy enzyme sites intended for enzymatic cofactors, such as is being considered for targeting the NS2B–NS3 dengue virus protease [133]. In this case, kinetic characterization is really no more complicated than the competitive inhibition scenario since it can be adequately quantified by assessing the capacity of the inhibitor to approach, enter, and persist in the cofactor subsite. In other cases, however, the mechanism typically entails induction by the inhibitor/modulator of an enzymatic conformation change that modifies the capacity (reduced in the case of inhibitors, amplified in the case of activators) of substrate molecules to access the catalytic site and be covalently processed. QSAR-based studies of enzyme modulation are commonly carried out, with reasonable levels of predictive performance [134,135]; however, detailed understanding of the underlying biochemical basis for such effects is really only possible either from analytical characterization (i.e., crystallography or NMR of the enzyme in a modulated state) or via molecular dynamics simulations that may extrapolate a plausible modulated state from a ligand–enzyme complex assembled

and initiated from an apo state structure, provided the molecular conformational transformation occurs within a computationally feasible time frame.

Finally, any comprehensive treatment of computational enzymology should include at least some mention of mutagenesis. Along the amino acid chain from which an enzyme is composed, many residue positions have little bearing on the ultimate biomolecular structure and function since amino acid substitution (and sometimes also deletion or insertion) at these points has little effect on quantitative observations of substrate catalytic, kinetics, or inhibition/modulation profiles; however, some sites yield significant (sometimes critical) sensitivity. Determination of the structural effects underlying such mutational sensitivity can shed very important light into the basic mechanism of protease function or modulation, and computational tools can play important roles in unraveling this insight. The results of molecular docking simulations can sometimes demonstrate sensitivity commensurate with experimental observations [136] especially in cases where the mutation is in close proximity to a ligand binding site and directly impinges on the capacity of a substrate, inhibitor, or modulator to complex with the site. Molecular dynamics simulations are even better suited to such mutations in that they can translate such structural modifications into revised projections for ligand transport kinetics and binding site complexation and can capture small- to medium-sized protein-wide conformational shifts that might arise from relaxation around the mutation site [137], provided the shift occurs in a computationally feasible time frame (<1.0 μS). One of the most promising computational resources for guiding the conception and deconvoluting the results of mutagenetic studies is the aforementioned COMBINE technique whose analytical strength, in addition to providing potentially valuable information about the underlying ligand structure–activity relationship, is the capacity for framing this SAR in the context of specific ligand–residue interactions. Thus, a COMBINE study on activity trends observed for a wild-type protease can identify specific residues that appear to play a critical role in determining the SAR and thus provide the basis for subsequent mutagenetic profiling. Furthermore, the COMBINE model can be retrained to account for trends observed in subsequent screening results on the various resulting mutational variants, thus leading to a comprehensive model that should yield reasonably predictive insight into the target system both as a function of variations in receptor and ligand structure. Some examples of COMBINE models applied to specific proteases include work done on HIV-1 protease [138,139] and trypsin [140] and our own work on proteasome 20S where we extended the basic COMBINE technique to account for covalent inhibitors [141].

V. CONCLUSIONS

Proteases perform a wide range of functions essential for cell survival and homeostasis. Since the peptide bond hydrolysis is an irreversible process, all protease activities are tightly regulated. Dysregulation of protease activity has been reported in cancer, inflammatory, and neurodegenerative diseases and in cardiovascular anomalies. A number of drugs targeting proteases are available for indications where the protease

activity is unequivocally associated with the disease. The development of safe and efficacious drugs targeting therapeutically relevant proteases requires detailed characterization of physiological protease substrates as well as protease expression and activation in normal and diseased states. A variety of biochemical, cell-based, and genomic approaches as well as *in vivo* animal models to study gain or loss of function have been useful in complete characterization of proteases. The mechanistic studies on proteolytic catalysis on different substrates require assay development and optimization. Majority of commonly used activity assays for proteases are based on short peptide substrates which may be generic or specific for a protease or for the protease family. Short peptide substrates designed for colorimetric-, fluorometric-, and luminescence-based assays give significant information on protease activity and its modulation by peptidomimetics or small molecules. For proteases with well-defined active site chemistry, ABPs have been utilized to detect active proteases in the cell after treatment with inhibitors or during disease progression or for determination of *in vivo* selectivity. Another modification of the assay platform uses quenched fluorescent peptide substrates immobilized on AuNP, in which fluorescence is observed post proteolysis. While the *in vitro* assays are useful for understanding mechanisms of proteolysis and their modulators, structural studies on protease–substrate interactions have helped refine specificity requirements in potential inhibitor scaffolds. In addition, computational tools clearly have a significant role to play in the characterization of normal and modulated protease function. While we must still take care to consider their current practical limitations, it should also be noted that active algorithmic development and the constant increase in available computational power are rendering in silico techniques ever more accurate and effective and are increasing the scope of applications to which they may be addressed. The day when computational analysis will no longer require complementary experimental analysis as a source of parameterization and validation is not yet conceivable; however, in silico methods are distinguishing themselves as ever more equal partners with experiment toward the complex goals of unraveling protease structure–function relationships and designing novel homologs with specially tailored functional attributes. An all-encompassing approach ranging from biochemical aspects of substrate specificity and assay development to optimization in conjunction with information from other genomics, proteomics, and degradomics will prove valuable in understanding proteases, designing assays, and ultimately identifying specific or pan-inhibitors which can be developed into safe and efficacious drugs in the future.

Disclaimer: Any reference to vendor products is merely for information purposes and is not a statement of endorsement.

REFERENCES

1. AJ Barrett, ND Rawlings, JF Woessner. *Handbook of Proteolytic Enzymes*. Amsterdam: Academic Press, 1998.
2. AJ Barrett, ND Rawlings. Types and families of endopeptidases. Biochem Soc Trans 19:707–715, 1991.

3. NM Hooper. Proteases: a primer. Essays Biochem 38:1–8, 2002.
4. C Lopez-Otin, JS Bond. Proteases: multifunctional enzymes in life and disease. J Biol Chem 283:30433–30437, 2008.
5. XS Puente, LM Sanchez, CM Overall, C Lopez-Otin. Human and mouse proteases: a comparative genomic approach. Nat Rev Genet 4:544–558, 2003.
6. V Quesada, GR Ordonez, LM Sanchez, XS Puente, C Lopez-Otin. The degradome database: mammalian proteases and diseases of proteolysis. Nucleic Acids Res 37:D239–D243, 2009.
7. ND Rawlings, AJ Barrett, A Bateman. MEROPS: the peptidase database. Nucleic Acids Res 38:D227–D233, 2010.
8. AJ Barrett. Classification of peptidases. Meth Enzymol 244:1–15, 1994.
9. I Schechter. Mapping of the active site of proteases in the 1960s and rational design of inhibitors/drugs in the 1990s. Curr Protein Pept Sci 6:501–512, 2005.
10. I Schechter, A Berger. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27:157–162, 1967.
11. I Schechter, A Berger. On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. Biochem Biophys Res Commun 32:898–902, 1968.
12. A Berger, I Schechter. Mapping the active site of papain with the aid of peptide substrates and inhibitors. Philos Trans R Soc Lond B Biol Sci 257:249–264, 1970.
13. B Turk. Targeting proteases: Successes, failures and future prospects. Nat Rev Drug Discov 5:785–799, 2006.
14. I Chowdhury, B Tharakan, GK Bhat. Caspases – an update. Comp Biochem Physiol B Biochem Mol Biol 151:10–27, 2008.
15. CS Craik, MJ Page, EL Madison. Proteases as therapeutics. Biochem J 435:1–16, 2011.
16. M Drag, GS Salvesen. Emerging principles in protease-based drug discovery. Nat Rev Drug Discov 9:690–701, 2010.
17. SS Cotrin, L Puzer, WA de Souza Judice, L Juliano, AK Carmona, MA Juliano. Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: assays with human cathepsin b. Anal Biochem 335:244–252, 2004.
18. WZ Mari Enoksson, GS Salvesen, ed. *Identification of Protease Substrates by Mass Spectrometry Approaches-1*. New York: Springer Science, 2008, pp. 67–81.
19. CM Overall, CP Blobel. In search of partners: linking extracellular proteases to substrates. Nat Rev Mol Cell Biol 8:245–257, 2007.
20. O Schilling, CM Overall. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 26:685–694, 2008.
21. M Niinobe, Y Hitomi, S Fujii. A sensitive colorimetric assay for various proteases using naphthyl ester derivatives as substrates. J Biochem 87:779–783, 1980.
22. AK Hazra, SP Chock, RW Albers. Protein determination with trinitrobenzene sulfonate: a method relatively independent of amino acid composition. Anal Biochem 137:437–443, 1984.
23. H Weingarten, J Feder. Spectrophotometric assay for vertebrate collagenase. Anal Biochem 147:437–440, 1985.
24. H Weingarten, R Martin, J Feder. Synthetic substrates of vertebrate collagenase. Biochemistry 24:6730–6734, 1985.
25. DM Jameson, JC Croney, PD Moens. Fluorescence: basic concepts, practical aspects, and some anecdotes. Meth Enzymol 360:1–43, 2003.

26. RP Haugland. *Hand Book of Fluorescent Probes and Chemical Research*. Eugene: Molecular Probes, Inc., 1996.
27. AJ Pope, UM Haupts, KJ Moore. Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. Drug Discov Today 4:350–362, 1999.
28. DM Jameson, JC Croney. Fluorescence polarization: past, present and future. Comb Chem High Throughput Screen 6:167–173, 2003.
29. LM Levine, ML Michener, MV Toth, BC Holwerda. Measurement of specific protease activity utilizing fluorescence polarization. Anal Biochem 247:83–88, 1997.
30. SK Grant, JG Sklar, RT Cummings. Development of novel assays for proteolytic enzymes using rhodamine-based fluorogenic substrates. J Biomol Screen 7:531–540, 2002.
31. J Karvinen, P Hurskainen, S Gopalakrishnan, D Burns, U Warrior, I Hemmila. Homogeneous time-resolved fluorescence quenching assay (LANCE) for caspase-3. J Biomol Screen 7:223–231, 2002.
32. S Mizukami, K Tonai, M Kaneko, K Kikuchi. Lanthanide-based protease activity sensors for time-resolved fluorescence measurements. J Am Chem Soc 130:14376–14377, 2008.
33. AB Cubitt, R Heim, SR Adams, AE Boyd, LA Gross, RY Tsien. Understanding, improving and using green fluorescent proteins. Trends Biochem Sci 20:448–455, 1995.
34. BA Pollok, R Heim. Using GFP in FRET-based applications. Trends Cell Biol 9:57–60, 1999.
35. J Neefjes, NP Dantuma. Fluorescent probes for proteolysis: tools for drug discovery. Nat Rev Drug Discov 3:58–69, 2004.
36. R Onuki, A Nagasaki, H Kawasaki, T Baba, TQ Uyeda, K Taira. Confirmation by FRET in individual living cells of the absence of significant amyloid beta -mediated caspase 8 activation. Proc Natl Acad Sci USA 99:14716–14721, 2002.
37. K Moin, M Samen, C Jedeszko, Q Li, MB Olive, RR Mattingly, BF Sloane. Images of cleavage: tumor proteases in action.In: D Edwards, G Høyer-Hansen, F Blasi, BF Sloane, eds. *The Cancer Degradome: Proteases and Cancer Biology*. New York: Springer Science, 2008, pp. 137–154.
38. DD Nolting, JC Gore, W Pham. Near-infrared dyes: probe development and applications in optical molecular imaging. Curr Org Synth 8:521–534, 2011.
39. D Maxwell, Q Chang, X Zhang, EM Barnett, D Piwnica-Worms. An improved cell-penetrating, caspase-activatable, near-infrared fluorescent peptide for apoptosis imaging. Bioconjug Chem 20:702–709, 2009.
40. M Funovics, R Weissleder, CH Tung. Protease sensors for bioimaging. Anal Bioanal Chem 377:956–963, 2003.
41. M Fonovic, M Bogyo. Activity based probes for proteases: applications to biomarker discovery, molecular imaging and drug screening. Curr Pharm Des 13:253–261, 2007.
42. WP Heal, TH Dang, EW Tate. Activity-based probes: discovering new biology and new drug targets. Chem Soc Rev 40:246–257, 2011.
43. D Kato, KM Boatright, AB Berger, T Nazif, G Blum, C Ryan, KA Chehade, GS Salvesen, M Bogyo. Activity-based probes that target diverse cysteine protease families. Nat Chem Biol 1:33–38, 2005.
44. D Greenbaum, KF Medzihradszky, A Burlingame, M Bogyo. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. Chem Biol 7:569–581, 2000.
45. NA Thornberry, HG Bull, JR Calaycay, KT Chapman, AD Howard, MJ Kostura, DK Miller, SM Molineaux, JR Weidner, J Aunins, et al. A novel heterodimeric cysteine

protease is required for interleukin-1 beta processing in monocytes. Nature 356:768–774, 1992.

46. N Jessani, M Humphrey, WH McDonald, S Niessen, K Masuda, B Gangadharan, JR Yates, 3rd, BM Mueller, BF Cravatt. Carcinoma and stromal enzyme activity profiles associated with breast tumor growth in vivo. Proc Natl Acad Sci USA 101:13756–13761, 2004.
47. A Saghatelian, N Jessani, A Joseph, M Humphrey, BF Cravatt. Activity-based probes for the proteomic profiling of metalloproteases. Proc Natl Acad Sci USA 101:10000–10005, 2004.
48. V Gurtu, SR Kain, G Zhang. Fluorometric and colorimetric detection of caspase activity associated with apoptosis. Anal Biochem 251:98–102, 1997.
49. A Prinz, G Reither, M Diskar, C Schultz. Fluorescence and bioluminescence procedures for functional proteomics. Proteomics 8:1179–1196, 2008.
50. MA O'Brien, RA Moravec, TL Riss, RF Bulleit. Homogeneous, bioluminescent proteasome assays. Methods Mol Biol 414:163–181, 2008.
51. MA O'Brien, WJ Daily, PE Hesselberth, RA Moravec, MA Scurria, DH Klaubert, RF Bulleit, KV Wood. Homogeneous, bioluminescent protease assays: caspase-3 as a model. J Biomol Screen 10:137–148, 2005.
52. WJD Martha, A O'Brien, P Eric Hesselberth, RA Moravec, MA Scurria, DH Klaubert, RF Bulleit, KV Wood. Homogeneous, bioluminescent protease assays: caspase-3 as a model. J Biomol Screen 10:137–148, 1995.
53. BR Branchini, JC Rosenberg, DM Ablamsky, KP Taylor, TL Southworth, SJ Linder. Sequential bioluminescence resonance energy transfer–fluorescence resonance energy transfer-based ratiometric protease assays with fusion proteins of firefly luciferase and red fluorescent protein. Anal Biochem 414:239–245, 2011.
54. O Salata. Applications of nanoparticles in biology and medicine. J Nanobiotechnology 2:3, 2004.
55. C Guarise, L Pasquato, V De Filippis, P Scrimin. Gold nanoparticles-based protease assay. Proc Natl Acad Sci USA 103:3978–3982, 2006.
56. C Guarise, L Pasquato, P Scrimin, V De Filippis. Reversible aggregation/deaggregation of gold nanoparticles induced by a cleavable dithiol linker gold nanoparticles-based protease assay. Langmuir 21:5537–5541, 2005.
57. A Laromaine, L Koh, M Murugesan, RV Ulijn, MM Stevens. Protease-triggered dispersion of nanoparticle assemblies. J Am Chem Soc 129:4156–4157, 2007.
58. E Dulkeith, AC Morteani, T Niedereichholz, TA Klar, J Feldmann, SA Levi, FC van Veggel, DN Reinhoudt, M Moller, DI Gittins. Fluorescence quenching of dye molecules near gold nanoparticles: radiative and nonradiative effects. Phys Rev Lett 89:203002, 2002.
59. CH Tung, Y Lin, WK Moon, R Weissleder. A receptor-targeted near-infrared fluorescence probe for in vivo tumor imaging. Chembiochem 3:784–786, 2002.
60. BW Lee, MR Olin, GL Johnson, RJ Griffin. In vitro and in vivo apoptosis detection using membrane permeant fluorescent-labeled inhibitors of caspases. Meth Mol Biol 414:109–135, 2008.
61. J Xie, F Zhang, M Aronova, L Zhu, X Lin, Q Quan, G Liu, G Zhang, KY Choi, K Kim, et al. Manipulating the power of an additional phase: a flower-like AU-FE3O4 optical nanosensor for imaging protease expressions in vivo. ACS Nano 5:3043–3051, 2011.
62. L Josephson, MF Kircher, U Mahmood, Y Tang, R Weissleder. Near-infrared fluorescent nanoparticles as combined MR/optical imaging probes. Bioconjug Chem 13:554–560, 2002.

63. M Zhao, L Josephson, Y Tang, R Weissleder. Magnetic sensors for protease assays. Angew Chem Int Ed Engl 42:1375–1378, 2003.

64. T Asefa, CT Duncan, KK Sharma. Recent advances in nanostructured chemosensors and biosensors. Analyst 134:1980–1990, 2009.

65. K Welser, J Grilj, E Vauthey, JW Aylott, WC Chan. Protease responsive nanoprobes with tethered fluorogenic peptidyl 3-arylcoumarin substrates. Chem Commun (Camb): 6: 671–673, 2009.

66. J Anderson, C Schiffer, SK Lee, R Swanstrom. Viral protease inhibitors. Handb Exp Pharmacol 189:85–110, 2009.

67. UH Jörg Eder, F Cumin, B Martoglio, B Gerhartz. Aspartic proteases in drug discovery. Curr Pharma Des 13:271–285, 2007.

68. R Lapatto, T Blundell, A Hemmings, J Overington, A Wilderspin, S Wood, JR Merson, PJ Whittle, DE Danley, KF Geoghegan, et al. X-ray analysis of HIV-1 proteinase at 2.7 a resolution confirms structural homology among retroviral enzymes. Nature 342:299–302, 1989.

69. NE Kohl, EA Emini, WA Schleif, LJ Davis, JC Heimbach, RA Dixon, EM Scolnick, IS Sigal. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA 85:4686–4690, 1988.

70. AP Lea, D Faulds. Ritonavir. Drugs 52:541–546; discussion 547–548, 1996.

71. ED Matayoshi, GT Wang, GA Krafft, J Erickson. Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. Science 247:954–958, 1990.

72. AV Broadhurst, NA Roberts, AJ Ritchie, BK Handa, C Kay. Assay of HIV-1 proteinase: a colorimetric method using small peptide substrates. Anal Biochem 193:280–286, 1991.

73. CA Sanoski. Aliskiren: an oral direct renin inhibitor for the treatment of hypertension. Pharmacotherapy 29:193–212, 2009.

74. JR Stockigt, RD Collins, EG Biglieri. Determination of plasma renin concentration by angiotensin I immunoassay. Diagnostic import of precise measurement of subnormal renin in hyperaldosteronism. Circ Res 28(Suppl. 2):175–191, 1971.

75. E Cauchon, S Liu, MD Percival, SE Rowland, D Xu, C Binkert, P Strickner, J-P Falgueyret. Development of a homogeneous immunoassay for the detection of angiotensin I in plasma using alphalisa acceptor beads technology. Anal Biochem 388:134–139, 2009.

76. C Volbracht, S Penzkofer, D Mansson, KV Christensen, K Fog, S Schildknecht, M Leist, J Nielsen. Measurement of cellular [beta]-site of APP Cleaving enzyme 1 activity and its modulation in neuronal assay systems. Anal Biochem 387:208–220, 2009.

77. C Garino, T Tomita, N Pietrancosta, Y Laras, R Rosas, G Herbette, B Maigret, G Quéléver, T Iwatsubo, J-L Kraus. Naphthyl and coumarinyl biarylpiperazine derivatives as highly potent human β-secretase inhibitors. Design, synthesis, and enzymatic BACE-1 and cell assays. J Med Chem 49:4275–4285, 2006.

78. F Mancini, M Naldi, V Cavrini, V Andrisano. Multiwell fluorometric and colorimetric microassays for the evaluation of beta-secretase (BACE-1) inhibitors. Anal Bioanal Chem 388:1175–1183, 2007.

79. BL Pietrak, M-C Crouthamel, K Tugusheva, JE Lineberger, M Xu, JM DiMuzio, T Steele, AS Espeseth, SJ Stachel, CA Coburn, et al. Biochemical and cell-based assays for characterization of BACE-1 inhibitors. Ana Biochem 342:144–151, 2005.

80. F Mancini, A De Simone, V Andrisano. Beta-secretase as a target for Alzheimer's disease drug discovery: an overview of in vitro methods for characterization of inhibitors. Anal Bioanal Chem 400:1979–1996, 2011.

81. ME Kennedy, W Wang, L Song, J Lee, L Zhang, G Wong, L Wang, E Parker. Measuring human [beta]-secretase (BACE1) activity using homogeneous time-resolved fluorescence. Anal Biochem 319:49–55, 2003.
82. H Nagase, JF Woessner, Jr. Matrix metalloproteinases. J Biol Chem 274:21491–21494, 1999.
83. TL Blundell. Metalloproteinase superfamilies and drug design. Nat Struct Biol 1:73–75, 1994.
84. C Lombard, J Saulnier, J Wallach. Assays of matrix metalloproteinases (mmps) activities: a review. Biochimie 87:265–272, 2005.
85. K Kupai, G Szucs, S Cseh, I Hajdu, C Csonka, T Csont, P Ferdinandy. Matrix metalloproteinase activity assays: importance of zymography. J Pharmacol Toxicol Methods 61:205–209, 2010.
86. S Netzel-Arnett, SK Mallya, H Nagase, H Birkedal-Hansen, HE Van Wart. Continuously recording fluorescent assays optimized for five human matrix metalloproteinases. Anal Biochem 195:86–92, 1991.
87. B Beekman, JW Drijfhout, HK Ronday, JM TeKoppele. Fluorogenic mmp activity assay for plasma including mmps complexed to α2-macroglobulin. Ann NY Acad Sci 878:150–158, 1999.
88. C Antczak, C Radu, H Djaballah. A profiling platform for the identification of selective metalloprotease inhibitors. J Biomol Screen 13:285–294, 2008.
89. J Peppard, F Glickman, Y He, SI Hu, J Doughty, R Goldberg. Development of a high-throughput screening assay for inhibitors of aggrecan cleavage using luminescent oxygen channeling (alphascreen). J Biomol Screen 8:149–156, 2003.
90. RJ Butkowski, J Elion, MR Downing, KG Mann. Primary structure of human prethrombin 2 and alpha-thrombin. J Biol Chem 252:4942–4957, 1977.
91. DA Walz, D Hewett-Emmett, WH Seegers. Amino acid sequence of human prothrombin fragments 1 and 2. Proc Natl Acad Sci USA 74:1969–1972, 1977.
92. RA Moravec, MA O'Brien, WJ Daily, MA Scurria, L Bernad, TL Riss. Cell-based bioluminescent assays for all three proteasome activities in a homogeneous format. Anal Biochem 387:294–302, 2009.
93. S Murata, H Yashiroda, K Tanaka. Molecular mechanisms of proteasome assembly. Nat Rev Mol Cell Biol 10:104–115, 2009.
94. A Liggett, LJ Crawford, B Walker, TCM Morris, AE Irvine. Methods for measuring proteasome activity: current limitations and future developments. Leuk Res 34:1403–1409, 2010.
95. M Unno, T Mizushima, Y Morimoto, Y Tomisugi, K Tanaka, N Yasuoka, T Tsukihara. The structure of the mammalian 20S proteasome at 2.75 Å resolution. Structure 10:609–618, 2002.
96. ES Johnson, PC Ma, IM Ota, A Varshavsky. A proteolytic pathway that recognizes ubiquitin as a degradation signal. J Biol Chem 270:17442–17456, 1995.
97. A Varshavsky. The n-end rule pathway of protein degradation. Genes Cells 2:13–28, 1997.
98. NP Dantuma, K Lindsten, R Glas, M Jellne, MG Masucci. Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. Nat Biotechnol 18:538–543, 2000.
99. K Lindsten, V Menendez-Benito, MG Masucci, NP Dantuma. A transgenic mouse model of the ubiquitin/proteasome system. Nat Biotechnol 21:897–902, 2003.

100. J Reidlinger, AM Pike, PJ Savory, RZ Murray, AJ Rivett. Catalytic properties of 26 s and 20 s proteasomes and radiolabeling of mb1, lmp7, and c7 subunits associated with trypsin-like and chymotrypsin-like activities. J Biol Chem 272:24899–24905, 1997.
101. W Ma, H Kantarjian, S O'Brien, I Jilani, X Zhang, Z Estrov, A Ferrajoli, M Keating, F Giles, M Albitar. Enzymatic activity of circulating proteasomes correlates with clinical behavior in patients with chronic lymphocytic leukemia. Cancer 112:1306–1312, 2008.
102. B Howley, HO Fearnhead. Caspases as therapeutic targets. J Cell Mol Med 12:1502–1516, 2008.
103. NA Thornberry, TA Rano, EP Peterson, DM Rasper, T Timkey, M Garcia-Calvo, VM Houtzager, PA Nordstrom, S Roy, JP Vaillancourt, et al. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. J Biol Chem 272:17907–17911, 1997.
104. C Pop, GS Salvesen, FL Scott. Caspase assays: identifying caspase activity and substrates in vitro and in vivo. Methods Enzymol 446:351–367, 2008.
105. JB Denault, GS Salvesen. Apoptotic caspase activation and activity. Methods Mol Biol 414:191–220, 2008.
106. AL Niles, RA Moravec, TL Riss. Caspase activity assays. Methods Mol Biol 414:137–150, 2008.
107. V Turk, B Turk, D Turk. Lysosomal cysteine proteases: facts and opportunities. Embo J 20:4629–4633, 2001.
108. S Conus, HU Simon. Cathepsins and their involvement in immune responses. Swiss Med Wkly 140:w13042, 2010.
109. MM Mohamed, BF Sloane. Cysteine cathepsins: multifunctional enzymes in cancer. Nat Rev Cancer 6:764–775, 2006.
110. HH Otto, T Schirmeister. Cysteine proteases and their inhibitors. Chem Rev 97:133–172, 1997.
111. N Katunuma. Structure-based development of specific inhibitors for individual cathepsins and their medical applications. Proc Jpn Acad Ser B Phys Biol Sci 87:29–39, 2011.
112. P Liewlom, T Rakthanmanon, K Waiyamai. Prediction of enzyme class by using reactive motifs generated from binding and catalytic sites. Lect Notes Artif Int 4632:442–453, 2007.
113. GP Brady, Jr., PF Stouten. Fast prediction and visualization of protein binding pockets with pass. J Comput Aided Mol Des 14:383–401, 2000.
114. S Jones, JM Thornton. Searching for functional sites in protein structures. Curr Opin Chem Biol 8:3–7, 2004.
115. GH Lushington. Comparative modeling of proteins. Methods Mol Biol 443:199–212, 2008.
116. GM Morris, DS Goodsell, RS Halliday, R Huey, WE Hart, RK Belew, AJ Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. J Comp Chem 19:1639–1662, 1998.
117. P Cunningham, I Afzal-Ahmed, RJ Naftalin. Docking studies show that D-glucose and quercetin slide through the transporter GLUT1. J Biol Chem 281:5797–5803, 2006.
118. JX Guo, Lushington, G.H. Unpublished results. 2004.
119. JL Sussman, M Harel, F Frolow, C Oefner, A Goldman, L Toker, I Silman. Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science 253:872–879, 1991.

120. DA Case, JW Caldwell, TE Cheatham, III, WS Ross, CL Simmerling, TA Darden, KM Merz, RV Stanton, AL Cheng, JJ Vincent, et al. (1999) Amber 6: University of California, San Francisco.

121. WL Jorgensen, C Jenson. Temperature dependence of TIP3P, SPC, AND TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. J Comp Chem 19:1179–1186, 1998.

122. G Cruciani, M Pastor, W Guba. Volsurf: a new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci 11(Suppl. 2):S29–S39, 2000.

123. GD Szklarz, MD Paulsen. Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities. J Biomol Struct Dyn 20:155–162, 2002.

124. M Rarey, B Kramer, T Lengauer, G Klebe. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489, 1996.

125. GW Jones, RC Glen, AR Leach, R Taylor. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748, 1997.

126. X Fradera, J Kaur, J Mestres. Unsupervised guided docking of covalently bound ligands. J Comput Aided Mol Des 18:635–650, 2004.

127. L Ridder, BA Palfey, J Vervoort, IM Rietjens. Modelling flavin and substrate substituent effects on the activation barrier and rate of oxygen transfer by p-hydroxybenzoate hydroxylase. FEBS Lett 478:197–201, 2000.

128. Z Anusevicius, AEMF Soffers, N Cenas, J Sarlauskas, M Martinez-Julvez, IMCM Rietjens. Quantitative structure activity relationships for the electron transfer reactions of Anabaena PCC 7119 ferredoxin-NADP+ oxidoreductase with nitrobenzene and nitrobenzimidazolone derivatives: mechanistic implications. FEBS Lett 450:44–48, 1999.

129. A Costescu, CD Moldovan, G Katona, M Diudea. QSAR modeling of human catechol O-methyltransferase enzyme kinetics. J Math Chem 45:287–294, 2009.

130. J Sipila, AM Hood, MW Coughtrie, J Taskinen. CoMFA modeling of enzyme kinetics: Km values for sulfation of diverse phenolic substrates by human catecholamine sulfotransferase SULT1A3. J Chem Inf Comput Sci 43:1563–1569, 2003.

131. GH Lushington, JX Guo, JL Wang. Whither COMBINE? New opportunities for receptor-based QSAR. Curr Med Chem 14:1863–1877, 2007.

132. AR Ortiz, MT Pisabarro, F Gago, RC Wade. Prediction of drug binding affinities by comparative binding energy analysis. J Med Chem 38:2681–2691, 1995.

133. G Katzenmeier. Inhibition of the ns2b-ns3 protease – towards a causative therapy for dengue virus diseases. Dengue Bull 28:58–67, 2004.

134. M Araaujo de Brito, CR Rodrigues, JJV Cirino, R Bicca de Alencastro, HC Castro, MG Albuquerque. 3D-QSAR COMFA of a series of DABO derivatives as HIV-1 reverse transcriptase non-nucleoside inhibitors. J Chem Inf Model 48:1706–1715, 2008.

135. J Singh, B Shaik, S Singh, S Sikhima, VK Agrawal, PV Khadikar, CT Supuran. QSAR studies on the activation of the human carbonic anhydrase cytosolic isoforms I and II and secretory isozyme VI with amino acids and amines. Bioorg Med Chem 15:6501–6509, 2007.

136. S Repo, J Jyrkkärinne, JT Pulkkinen, R Laatikainen, P Honkakoski, MS Johnson. Ligand specificity of constitutive androstane receptor as probed by induced-fit docking and mutagenesis. J Med Chem 51:7119–7131, 2008.

137. P Kollman. Computer simulation methods as applied to site-specific mutations. Protein Des Dev New Ther Vaccines: 229–236, 1990.

138. S Nakamura, I Nakanishi, K Kitaura. Binding affinity prediction of non-peptide inhibitors of HIV-1 protease using combine model introduced from peptide inhibitors. Bioorg Med Chem Lett 16:6334–6337, 2006.

139. C Pérez, M Pastor, AR Ortiz, F Gago. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. J Med Chem 41:836–852, 1998.

140. MBM Peters, M Kenneth. Semiempirical comparative binding energy analysis (SE-COMBINE) of a series of trypsin inhibitors. J Chem Theor Comput 2:383–399, 2006.

141. JL Wang, A Datta, GH Lushington. Structure-based quantitative structure activity relationship analysis of omuralide analogs in the 20S proteasome: a covalent inhibitor COMBINE study. Lett Drug Des Discov 4:417–421, 2007.