Problem 1. Interleukin 1 β (IL-1 β) has been implicated in the pathogenesis of acute chronic inflammatory diseases including septic shock, rheumatoid arthritis etc. IL-1 β must be processed from a pro-peptide (pro-IL-1 β) form of 31 kDa to the active form of 17.5 kDa by a cysteine protease designated ICE or interleukin converting enzyme. The cleavage of IL-1 β occurs between Asp116 and Ala117. ICE is a unique cysteine protease, but appears to use a similar mechanism to other cysteine proteases requiring a histidine and a cysteine for catalysis. To study the enzyme, Merck had to develop an assay. As you can imagine, assays with a protein such as IL-1 β is cumbersome, and thus they did a number of experiments using peptides to see if they could find an effective peptide analog(s) to report on the substrate specificity and cleavage efficiency. They initially tried to determine the binding specificity length on either side of the bond to be cleaved (Figure 1), as well as the optimized amino acid binding at each subsite (Figure 2). The cleaved bond is between P1 and P1' (see igure 1, right) with aspartate at P1, very important for cleavage. The assay required monitoring the reaction by HPLC where the peptide substrates and products needed to be separated and quantitated as a function of time. This assay is very slow and not very sensitive.

Amino-terminal truncations

Peptide	P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'	P7'	Relative V_{max}/K_m
1	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn	1.00 ± 0.05
2		Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn	0.64 ± 0.01
3			Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn	0.29 ± 0.02
4				Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn	0.12 ± 0.00
5			Ac-Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn		0.81 ± 0.01
6				Ac-Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn		< 0.005
7					His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn		< 0.005
8						Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn		< 0.005

Carboxy-terminal truncations

Peptide	P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'	P7'	Relative V_{max}/K_m
1	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu	Asn	1.00 ± 0.05
9	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser	Leu		1.22 ± 0.16
10	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg	Ser			0.80 ± 0.16
11	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val	Arg				0.88 ± 0.04
12	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro	Val					1.02 ± 0.09
13	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala	Pro						1.25 ± 0.15
14	Asn	Glu	Ala	Tyr	Val	His	Asp	Ala							< 0.05
15	Asn	Glu	Ala	Tyr	Val	His	Asp	-CO-NH$_2$							0.93 ± 0.03
16			Ac-Tyr	Val	His	Asp-NH-CH$_3$									12.00 ± 2.0
17			Ac-Tyr	Val	His	Asp-AMC									1.9 ± 0.1
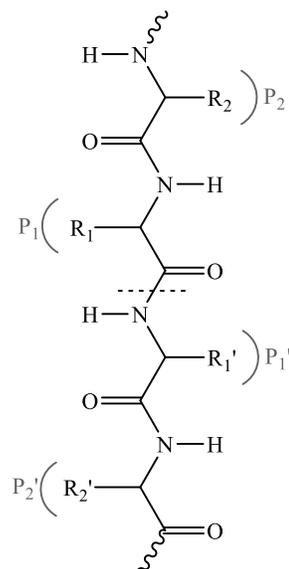
Fig 1. Analogs to better understanding the substrate specificity of ICE

Taking advantage of information gained by the studies in Figure 1, they then turned to develop a more high throughput (HTP), sensitive, assay using coumarin analogs, which upon peptide bond hydrolysis, become highly fluorescent. Thus the reaction could be monitored continuously in a spectrofluorimeter. Alternatively, they used p-nitroanilide in place of the coumarin (see structure below).

   p-nitroanilide
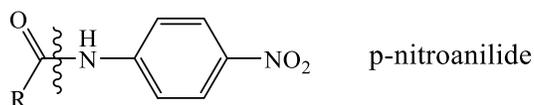
Using this assay they looked further at the substrate specificity and the results are reported in Figure 2.

Comparison of fluorogenic substrates for ICE	
Substrate	10^{-5} x k_{cat} / $K_{m'}$ (M^{-1} s^{-1})
Ac-WEHD-AMC	33.4 ± 0.3
Ac-WVHD-AMC	15.7 ± 0.1
Ac-YEHD-AMC	9.56 ± 0.20
Ac-WEAD-AMC	7.55 ± 0.07
Ac-YVHD-AMC	2.81 ± 0.14
Ac-WVAD-AMC	2.41 ± 0.60
Ac-YEAD-AMC	1.85 ± 0.07
Ac-YVAD-AMC	0.66 ± 0.14
pro-IL-1 β	1.5
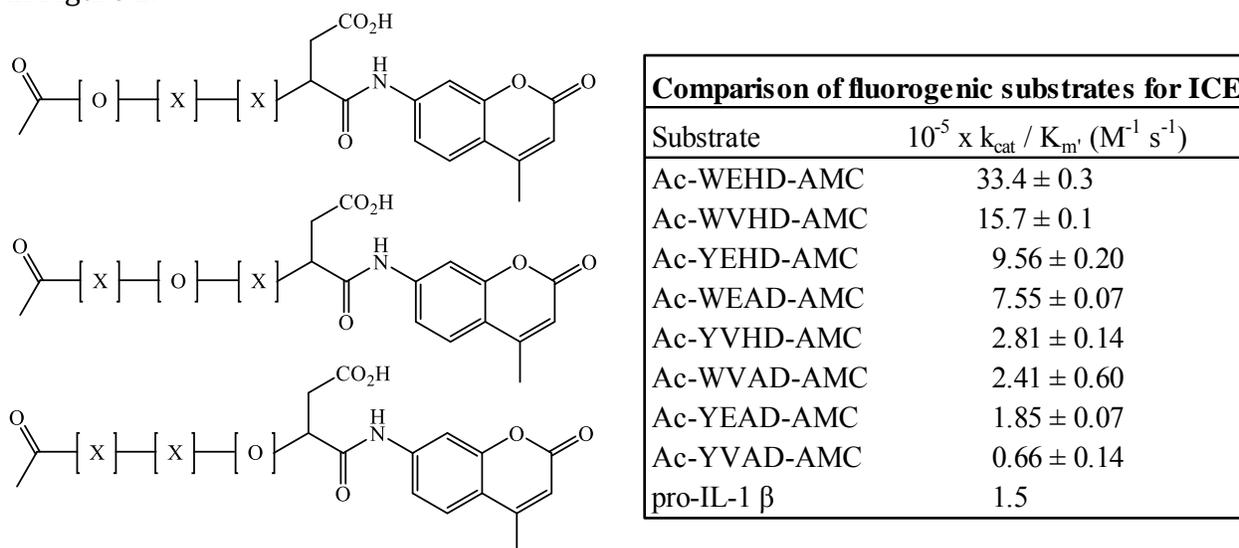
Figure 2. Further substrate specificity studies using a coumarin fluorescence assay.

The mechanism of cysteine proteases is proposed to be similar to serine proteases that you have or will be discussing in recitation (Figure 3).
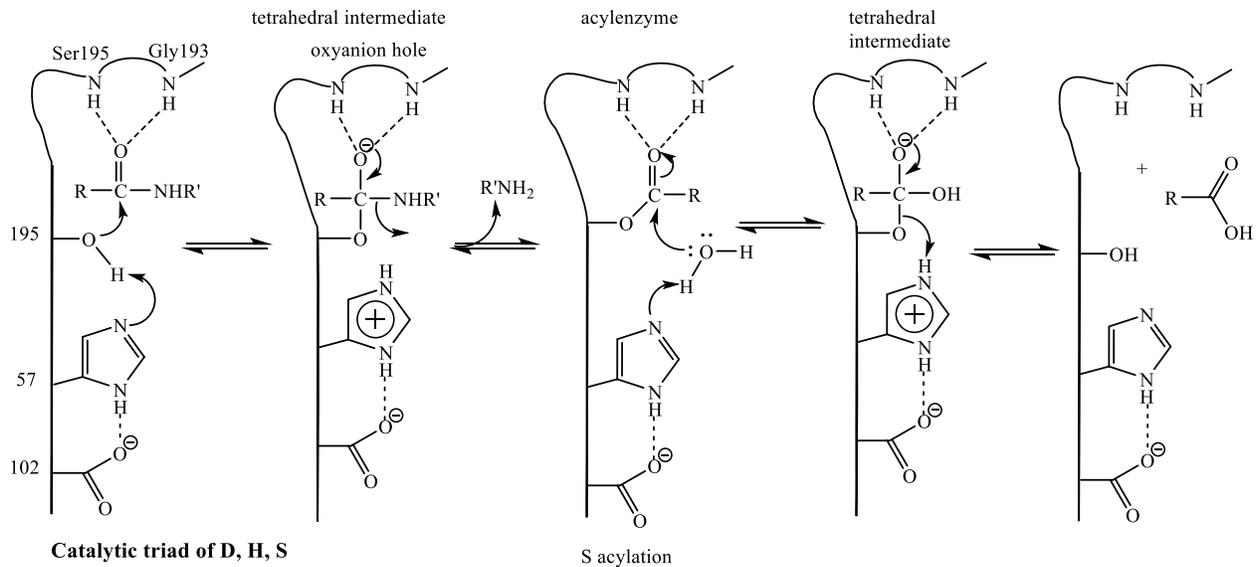
Figure 3 is a generic mechanism for serine proteases.

However with cysteine proteases we know now from crystallographic studies and studies with many covalent inhibitors of cysteine proteases, that there are only two amino acids that play an essential role in catalysis: a cysteine and a histidine.

Questions:
1. What does the data in Figure 1 tell you about the substrate binding site to ICE? How does the k_{cat}/K_m or (V_{max}/K_m) for the real substrate pro-IL-1β meet your expectations relative to the other data? Explain why?

The specificity question with proteins as substrates is challenging as parts of the protein far removed from the active site and the few subsites on either side of that site, may play a crucial role in binding and catalysis. In the approach taken above, only the sites adjacent to the site where cleavage occurs can be studied. A second issue is that changing a residue say at P2' of P3' could affect the binding of a residue at P3. Thus changing one amino acid at a time may not be the best way to understand specificity of protein substrates. [Aside: due to the importance of post-translational modifications, many proteins are substrates for enzymes and thus this is a general issue that needs a creative solution.] In Figure 1, k_{cat}/K_m values tell you about the effectiveness of the different peptide constructs for binding to the enzyme (E + S), that is, the magnitude of the second order rate constant is limited by diffusion control and this number tells you how effectively your enzyme interacts with the S. In Figure 1 the amino acids in the prime (') site of cleavage are held constant. Thus what you learn from this exercise is that you need at least 4 amino acids (P1 through P4) and that the N-terminal amino acid when acylated interacts more efficiently. In this one experiment the amino acids in sites P1 through P4 are not explored. [Selection of these amino acids would be aided by knowing "all" cleavage sites of ICE in vivo. Only recently have technologies been successfully developed to provide us with this information. This is a challenging problem.] In the bottom half of the Table in Figure 1, the prime (') site is

investigated. This site is important as it allows us to determine if we can modify the leaving group (fluorescent probe or colored probe upon peptide bond hydrolysis) so that our assays can be carried out with continuous monitoring and with greater sensitivity. In fact the data suggest that the entire prime (') peptide can efficiently be replaced by $NHCH_3$. This leaving group, however, is NOT easy to detect and your assay would still be challenging. Fortuitously the AMC (see Figure 2) is fluorescent on hydrolysis of the peptide bond and its k_{cat}/K_m is as efficient as the 14-mer peptide. Thus this piece of information is very important for a high throughput assay where one is looking for inhibitors.

The gold standard for peptide bond hydrolysis would be how it compares to the actual pro-IL-1β substrate. That information is provided in Figure 2. Its k_{cat}/K_m is 1.5×10^5 $M^{-1}s^{-1}$ compared with the acylated YVHD-AMC that is 2.8×10^5 $M^{-1}s^{-1}$.

Two conclusions: the study with the endogenous protein substrate is important as it suggests that sites "far removed" from the site of cleavage are not essential in specificity of cleavage. While we have not discussed this in class, most proteases are made in a pro-protease form that is inactive. They are only activated after cleavage. This is one mechanism that prevents the proteases from randomly degrading proteins inside the cell. The activity of each protease is carefully regulated. Chymotrypsin and trypsin are synthesized by the ribosome in pro-inactive forms that need to be activated.

Finally many proteases have been targeted for therapeutic intervention and in general peptides with several sites (P1---Pn) provide excellent substrates based on k_{cat}/K_m for studying the enzyme and for high throughput screens.

2. Explain why the investigators decided to use the substrate analogs shown in Figure 2 to continue their substrate specificity studies. Can you think of how a more informative assay could be carried out? How would the corresponding p-nitroanilide analogs function in the assay? What criteria would you use to determine if the coumarin or the nitroanilide would be best for high throughput assays?

In Figure 2 the sites P2 through P4 are being explored. The P1 site is held constant as studies of biological protein substrates all implicate the importance of D at this site. [This is similar to trypsin where R or K are very important in the P1 site.] The studies are changing one or two amino acids at a time and measuring the consequences using the continuous coumarin fluorescence assay. They have found peptides 10 to 20 times more efficient. The importance of proteases in regulation and as therapeutic targets have resulted in many technologies to better define their specificity. For those interested in combinatorial approaches you can read about them online. Lim, Mark D, and Charles S Craik. 2009. "Using Specificity to Strategically Target Proteases." *Bioorganic & Medicinal Chemistry* 17 (3): 1094–1100. doi:10.1016/j.bmc.2008.03.068.
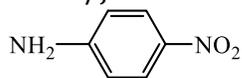
The p-nitroanilide which is yellow colored subsequent to cleavage of the amide bond and has a high exctinction coefficient of 15000 $M^{-1}cm^{-1}$. Think about the resonance structures that can be drawn subsequent to cleavage. One would need to

measure k_{cat}/K_m with this leaving group and compare the results with the AMC analog with the same peptide in sites P1 through P4. You would also want to think about which one is more sensitive, allowing you to use very small amounts of samples and 360 well microplate assay trays.

3. Explain why k_{cat}/K_m is the kinetic parameter used in the studies described in Figure 1 and not k_{cat}.

In carrying out an assay for inhibition, one needs to know what concentrations of substrates to employ. If for example your enzyme is saturated (k_{cat} is reached) with 0.2 mM substrate and you used 10 mM substrate in your assays, all of your enzyme is in the ES form and you still have a huge amount 9.8 mM S not bound. Thus when looking for an inhibitor that binds at the active site, the S would almost always outcompete the I and one would never see inhibition. Thus understanding k_{cat}/K_m helps you to understand the specificity of substrate binding and how to set up assays for inhibitors.

4. Using the mechanism shown in Figure 3 for serine proteases, draw a corresponding mechanism for cysteine proteases. Based on what you have learned about serine proteases and the chemical difficulties associated with catalysis, provide an explanation for why an aspartate residue is not found in the active sites of cysteine proteases.

The cysteine proteases in general have evolved to have the pKas of their cysteines lowered from 8.5 to about 6 or in some cases even lower (distinct with every protease). Thus the resting form of the enzyme unusually is the thiolate and imidazolium shown as shown in 1 below. This is distinct from the serine proteases that need to use a different strategy to allow deprotonation of the serine. The pKa of the serine in solution is 15 to 16 (on the enzyme it may be lowered to 14). How can a histidine with a pKa of 6.5 to 7 deprotonate the serine required for the chemistry? The pKa of the cysteine is 8.5 in solution and can be lowered to on the enzyme from 3 to 6. The aspartate is necessary for the serine proteases to force the His to become more basic (using binding energy, see figure below). Since in the cysteine proteases the pKa of the cysteine is so much lower, the pKas are not as mismatched.
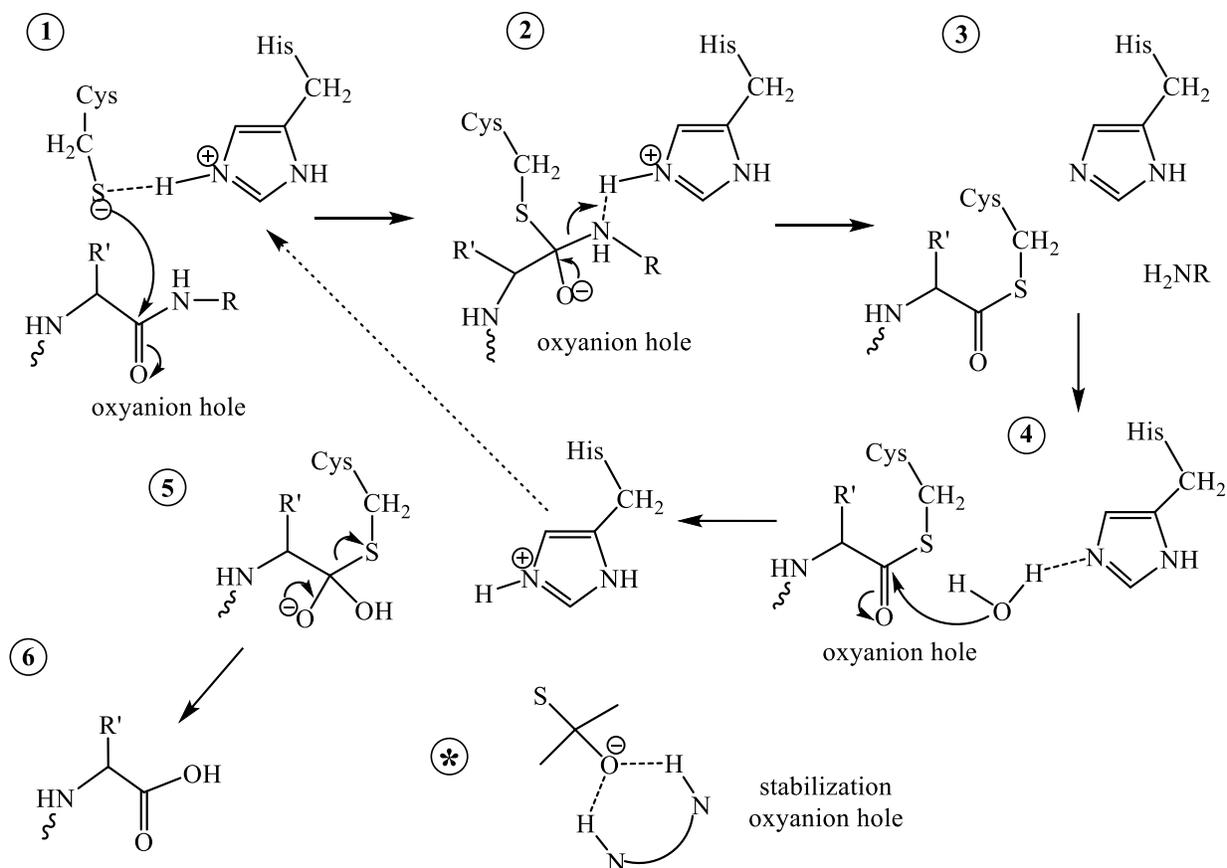
5. Peptide bond hydrolysis catalyzed by the enzyme is still 10^{10} x that of the non-enzymatic reaction. Discuss the different mechanisms of catalysis and their contributions to this rate acceleration.

The mechanisms of rate acceleration can be provided by binding energy (BE) released on interaction between the enzyme and the substrate. This BE can be used to stabilize the transition state of the reaction (have stronger H bonds to the tetrahedral intermediate or ts, than to the ground state carbonyl of the amide) and to desolvate and orient the substrate in the active site so all the groups involved in catalysis are precisely aligned (<0.1 Å). The other two mechanisms are general base and general acid catalysis and covalent catalysis. The cysteine and serine proteases use all three mechanisms to obtain the observed huge rate accelerations relative to the non-enzyme catalyzed hydrolysis. In the case of the serine proteases the hypothesis has been that when the peptide substrate binds, the binding energy in part is used to juxtapose the negatively charged aspartate with the imidazole. Imidazole/imidazolate has a pKa of 16, thus there is a huge pKa mismatch (4.5 vs 16). Thus the His rapidly picks up a proton from the serine to overcome this unfavorable interaction imposed by substrate binding.
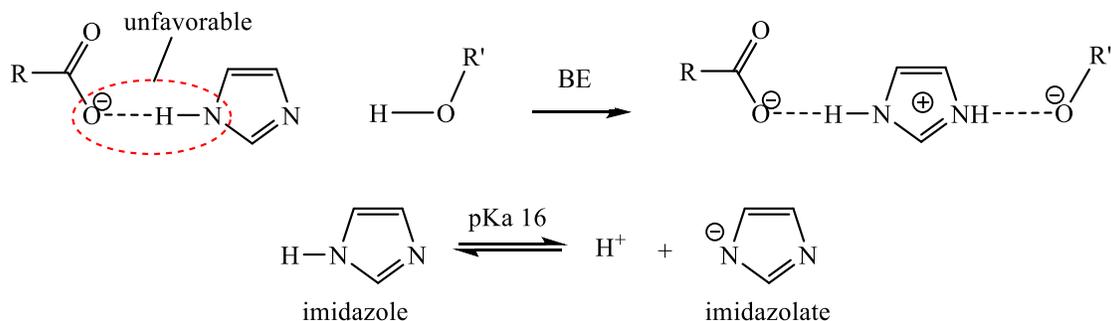
In the case of both proteases, the His can act as a general base and a general acid. As a general base one needs to make the Ser or Cys into a good nucleophile. Also in breakdown of the tetrahedral transition state or intermediate, one needs to protonate the leaving group. In this case the imidazolium state of the His is used. Thus as in most enzymes, a single group is often used as both a general acid and a general base during a single conversion of substrate to product. In the case of both serine and cysteine proteases covalent catalysis is also used. It should be noted however that there are two other general mechanisms of proteolysis: one involves zinc (Zn^{2+}) for catalysis and a second involves two aspartates that function one as a general acid and one as a general base in catalysis. Neither involves covalent catalysis.

Problem 2. As with serine proteases, a number of different successful approaches have been achieved to target cysteine proteases once their specificity is understood. Shown below are two distinct approaches to inhibit cysteine proteases in general. The data below is shown for inhibition of ICE. The first is peptide aldehyde analogs where R in Figure 4 is replaced by an aldehyde (CHO).

R	K_i (nM)
	0.76
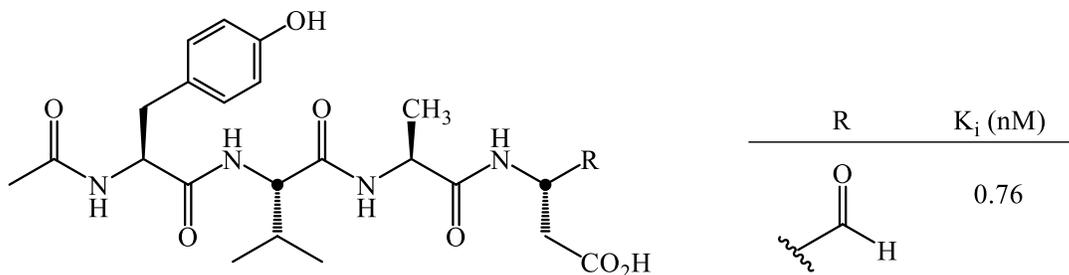
Figure 4. An example of the peptide aldehyde inhibitor

The second type of inhibitor is acyloxymethyl ketone analogs shown in Table 1.

Table 1. Tetrapeptide (Acyloxy)methyl Ketones (below)
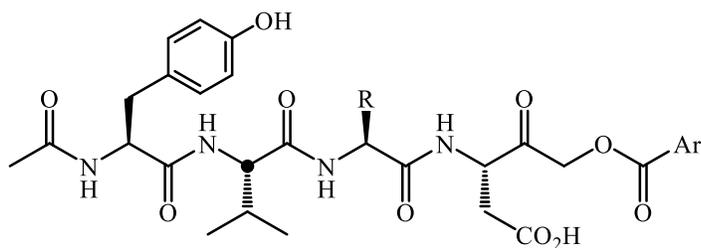
#	Ar	R
1	(aromatic ring with CF₃ and F₃C substituents)	CH_3
2	(aromatic ring with CH₃ and H₃C substituents)	CH_3
3	(aromatic ring with CH₃ and H₃C substituents)	$(CH_2)_4NH$-biotin
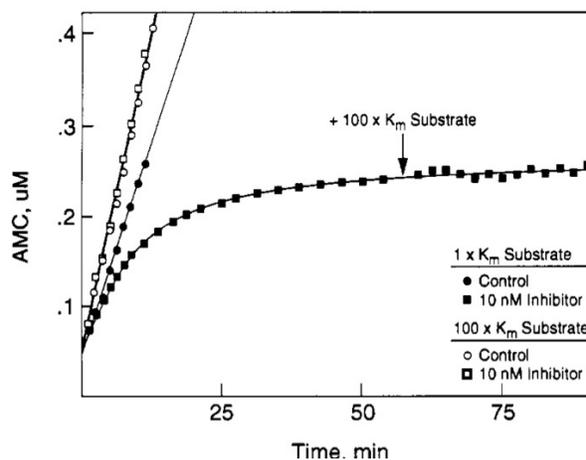
Figure 5. Using AMC analogs in Figure 2, the rate of the reaction was monitored in the presence of inhibitors in Table 1. Conditions are given in the figure along with the control (no inhibitor).

The data from the inhibition studies is shown in Figure 5.

Questions: 1. Propose a mechanism of inhibition of ICE by the peptide aldehydes such as the one shown in Figure 4. This is a mechanism-based inhibitor. What type of inhibition would you expect to see using substrates in Figure 2 to analyze for inhibition? Show the Eq and the kinetic reciprocal plots ($1/v$ vs $1/[S]$) with varying concentrations of substrate and inhibitor.

Peptide Aldehydes
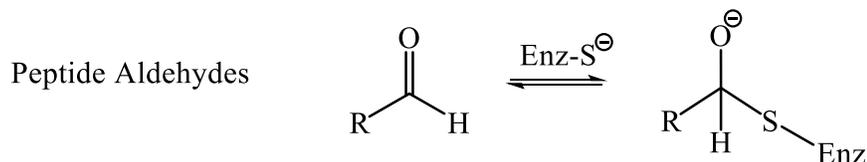
This is a reversible inhibitor in which the normal substrate can compete with the aldehyde with the same R group for binding to the active site nucleophile. As long as the equilibrium shown above is rapid and reversible, one can outcompete the inhibitor by trapping the free

enzyme with substrate as the concentration of the substrate increases. If the equilibrium is not fast and the breakdown of the tetrahedral intermediate is slow, then one might see slow inactivation of the enzyme.

$$v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_{is}} + [S]\right)}$$

with the reciprocal giving

$$\frac{1}{v} = \frac{K_m}{V_{max}}\left(1 + \frac{[I]}{K_{is}}\right)\frac{1}{[S]} + \frac{1}{V_{max}}$$

The plot below readily reveals that increasing the concentration of the substrate prevents inhibitor binding to the same site. This requires as noted above, rapid reversible binding of the inhibitor.
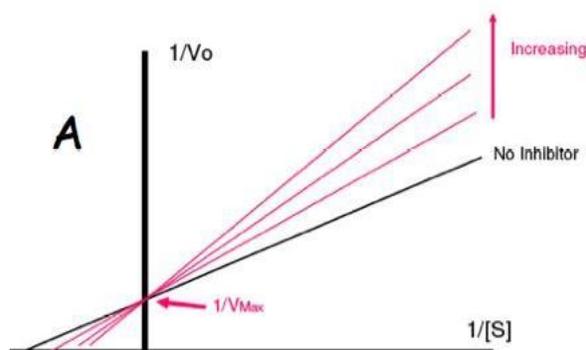
2. From a chemical perspective, provide two reasons why you might not want to use an aldehyde as a therapeutic.

a. Aldehydes often exist in hydrated form in solution in equilibrium with the free aldehyde. Thus the concentration of the substrate available for binding depends on the equilibrium constant.
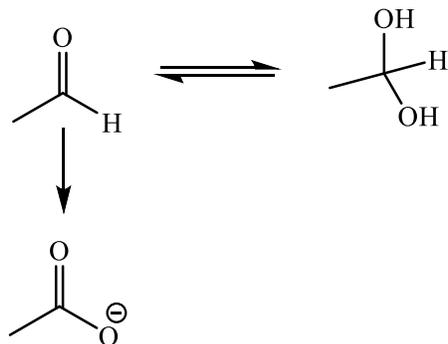
b. Aldehydes are readily oxidized to acids in vivo and given the mechanism of inhibition shown above, would no longer inhibit the enzyme.

3. The acyloxymethylketone analogs shown in Table 1 with data in Figure 5, have a mechanism of inhibition distinct from the peptide aldehydes. Propose a chemical mechanism for their inhibition of ICE. Provide an explanation for the data in Figure 5.

$$E + S \rightleftarrows ES \rightarrow E + P \qquad K_{is} = [E][I]/[EI]$$

$$[I] \downarrow\uparrow$$

$$EI \rightarrow E\text{-}I \ \text{(or } E\bullet I \text{ where I binds with a pM } K_1)$$

α-substituted ketones
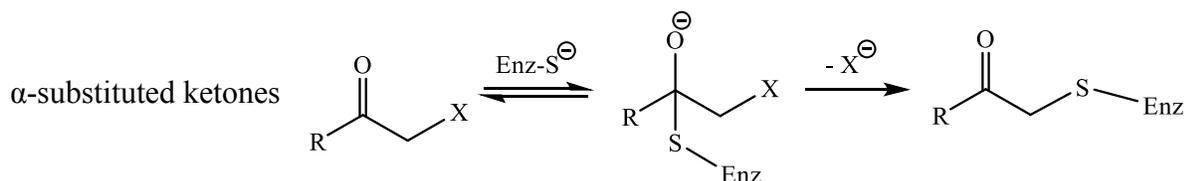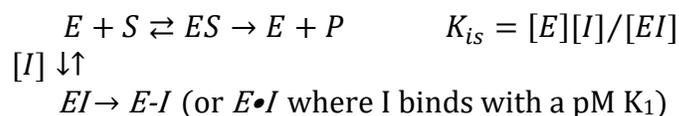
This type of inhibitor is similar to the tosylphenylchloroketone that you discussed in recitation that can inhibit serine proteases by alkylation of the active site histidine. You may or may not have had time to discuss the mechanism, but what is formally required is an SN2 reaction by the imidazole form of the enzyme's active site His with loss of X-. In the case of serine proteases the mechanism involves the active site nucleophile serine that displaces X- to form an epoxide, that then is protonated by the imidazolium form of the His making it more activated for nucleophilic attack by the resulting imidazole. The result, regardless of the mechanistic details, is that the enzyme active site is modified with a covalent bond to the inhibitor. You lose active catalyst and so with time the rate of the normal reaction decreases. This gives rise to the time-dependent phenomenon shown in Figure 5. The inset in the plot shows that if you have enough substrate around (100 x K_m) and low amounts of inhibitor, you can prevent this inhibition. It depends on the rate constants for the different steps in the equations shown above.

5.07SC Biological Chemistry I
Fall 2013