Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 3097–3106
Bioorganic & Medicinal Chemistry

Intramolecular orbital alignments in serine protease/protein inhibitor complexes ☆

Yun-Hua Fan,[a,†] Claire-Anne Grégoire[b] and John Haseltine[b,c,*]

[a]*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA*
[b]*Département de chimie, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1*
[c]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02115-4307, USA*

Received 6 February 2003; accepted 9 February 2004
Available online 24 April 2004

This paper is dedicated to the memory of Professor Jerry Donohue of the University of Pennsylvania

Abstract—By an analysis of PDB crystal structures, the mean conformations of protein strands bound in serine protease active sites are shown to contain extensively aligned atomic orbitals. The active-serine-bearing segment of each enzyme (subtilisin BPN′ and β-trypsin) also contains such alignments. The participating orbitals are almost identical in each system. All of the alignments converge on the targeted linkage. They suggest that a kind of through-strand polarizability is being optimized by evolution, presumably due to corresponding benefits in proteolysis rate. Such polarizability would help to explain the high values of k_{cat} seen for long oligopeptide substrates. The idea predicts long substrates to be relatively reactive even under non-enzymatic conditions, which in fact they are.

1. Introduction

Members of the chymotrypsin protease family are composed of two antiparallel β-barrel domains, while members of the subtilisin family are composed of a parallel β-sheet surrounded by four α-helices. Despite this difference in global structure, the active sites of α-chymotrypsin and subtilisin BPN′ are highly geometrically similar (see Fig. 1).[1] The similarity is extensive throughout the relatively large active sites. The systems' common elements are almost superimposable in three dimensions. Kraut proposed that such enzymes, having large active sites and showing high values of k_{cat} for substrates that can fill those sites, may have evolved 'to bind the extended transition-state geometry but not the ground-state geometry' of their substrates.[2] By this idea,

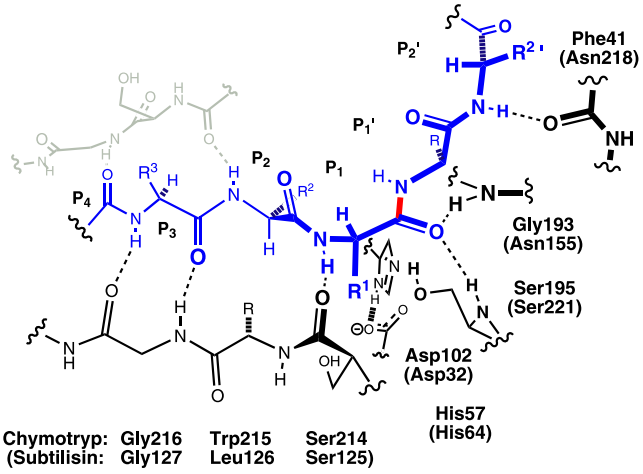


Figure 1. Some shared enzyme/ligand elements of the chymotrypsin and subtilisin systems. The enzyme is shown in black, the bound strand is in blue, and the targeted bond is in red. Subtilisin numbering appears in parentheses after chymotrypsin's numbering. Subtilisin also uses a second H-bonding strand, shown in gray.

the similarity of the active sites suggests that both enzymes are anticipating very similar extended TS geometries and that those geometries are especially advantageous. Kraut's proposal is appealing as a logical

Keywords: Serine proteases; Orbital alignments; Polarizability; Hyperconjugation.
☆ Taken in part from the doctoral dissertation of Y.-H. Fan, Georgia Institute of Technology, 1998.
* Corresponding author at present address: Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, GA 30144, USA. Tel.: +1-678-797-2048; fax: +1-770-423-6744; e-mail: jhaselti@kennesaw.edu
† Present address: Department of Cell Biology, Harvard Medical School, Cambridge, MA 02139, USA.

0968-0896/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.bmc.2004.02.044

deduction, but it is not clear why an optimal transition structure for an acyl transfer mechanism should involve a particular extended geometry or indeed any extended geometry.

If the idea of an extended TS geometry is correct, further comparison of the two serine protease families should turn up specific connections between extended geometry and TS character. In this paper we analyze the conformations of enzyme-bound protein strands and the active-serine-bearing strand in trypsin and subtilisin complexes from the Protein Data Bank (PDB).[3] The analysis suggests that these two enzymes have evolved to optimize a kind of through-strand polarizability in their reacting complexes. This idea leads to a new explanation of the known dependence of k_{cat} on substrate length. It also correctly implies that long-range structure is important to substrate reactivity even under non-enzymatic conditions.

2. Results and discussion

2.1. The mean conformation of enzyme-bound strands

Complexes between serine proteases and their natural protein inhibitors must resemble in certain ways the low-energy structures of normal enzyme/substrate complexes, as has often been noted.[4] Most important for us, each inhibitor places a loop of enzyme-preferred residues in the active site. Each complex shows good structural complementarity, with the bound residues satisfying many H-bonding and van der Waals constraints. For our analysis, we chose crystal structures of such complexes from the PDB by the following criteria. Each enzyme would be a well studied form of trypsin, chymotrypsin, elastase, thrombin, or subtilisin. Each inhibitor would consist of a polypeptide chain and place an intact peptide linkage in the active site. No covalent bonds would exist between enzyme and inhibitor. Resolutions would be 2.5 Å or better with *R*-values of 0.20 or better. Most of the qualifying structures involved β-trypsin or subtilisin BPN′ (the thirteen structures described in Tables 1 and 2), so we focussed on those systems.[5]

We will use the 'P_n' notation of Schechter and Berger to specify the subsites in the enzyme-bound loop.[17] That is, P_1 and P_1' are the residues forming the targeted linkage, P_2 is the next residue upstream, P_2' is the next residue downstream, and so forth (see Fig. 1). We determined the mean conformation, torsion by torsion, for the set of eight trypsin-bound loops and for the set of five subtilisin-bound loops. The results are illustrated in Figure 2. This is a comparison of two oligo-alanine strands that have been given the mean torsional values of those trypsin versus subtilisin data sets. When the targeted linkages are superimposed, one finds that the two conformations are almost isosteric from P_3 to P_2' (subtilisin's binding site seems to be longer than trypsin's, allowing it to interact with the backbone of at least one additional residue, P_4). The details of these mean conformations may be skewed by the narrow samples of residue identities, especially at the P_3–P_1 positions (see Table 2). However, because the binding sites of the two enzymes are very geometrically similar, and because the protein inhibitors satisfy the binding-site constraints very similarly, subtilisin- versus trypsin-bound substrates, as sets, must have similar mean conformations.[18] Because the inhibitors satisfy the binding-site constraints very well, we will assume for now that the conformations of Figure 2 are similar to means for bound substrates.

So why would trypsin and subtilisin each evolve to bind this particular conformation of their substrates? The enzymes have broad specificities, so it must be attainable by a broad range of substrates. Could it be especially stable? Or, for advocates of ground-state strain as a catalytic tactic, is it somehow especially *unstable*? Most of the individual bound residues occupy well populated Ramachandran space.[19] The Ψ torsion in the P_1 residue has rather unusual values (means of 37° and 43° in the

Table 1. β-Trypsin/ and subtilisin BPN′/protein inhibitor complexes selected from the PDB for the determination of mean strand conformations

#[a]	Code	Enzyme	Inhibitor/family		Resolution, Å	*R*-value	Authors	Ref.
I	1MCT	Porc. β-trypsin	MCTI-A	MCTI	1.6	0.167	Huang et al.	6a
II	1F2S	Bov. β-trypsin	MCTI-A mutant	MCTI	1.8	0.177	Huang et al.	6b
III	1PPE	Bov. β-trypsin	CMTI-I	MCTI	2.0	0.151	Bode et al.	7
IV	2PTC	Bov. β-trypsin	Pancreatic trypsin inhibitor	BPTI	1.9	0.187	Huber et al.	8
V	1TAW	Bov. β-trypsin	APPI	BPTI	1.8	0.184	Hynes et al.	9
VI	1TAB	Bov. β-trypsin	AB-1 (Bowman-Birk)	AB	2.3	0.200	Tsunogae et al.	10
VII	1SFI	Bov. β-trypsin	Bowman-Birk-type peptide	AB	1.7	0.175	Luckett et al.	11
VIII	1SMF	Bov. β-trypsin	Bowman-Birk mimic	—	2.1	0.190	Huang et al.	12
IX	2SIC	Subtilisin BPN′	*Strept.* subtilisin inhibitor	SSI	1.8	0.177	Mitsui et al.	13
X	3SIC	Subtilisin BPN′	SSI (M73K)	SSI	1.8	0.178	Mitsui et al.	14
XI	5SIC	Subtilisin BPN′	SSI (M70G, M73K)	SSI	2.2	0.176	Mitsui et al.	14
XII	1SBN	Subtilisin BPN′	Eglin C (L45R)	EGC	2.1	0.186	Grütter et al.	15
XIII	2SNI	Subtilisin BPN′	Chymo. inhibitor 2 (CI-2)	CI-2	2.1	0.154	McPhalen et al.	16

[a] The inhibitors are grouped by family.

Table 2. Residues in the enzyme-bound loop from P_5 to $P_{4'}$ for complexes I–XIII

#[a]	P_5	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
I	Arg	Ile	Cys	Pro	Arg	Ile	Trp	Met	Glu
II	Arg	Ile	Cys	Pro	Arg	Ile	Trp	Met	Glu
III	Arg	Val	Cys	Pro	Arg	Ile	Leu	Met	Glu
IV	Thr	Gly	Pro	Cys	Lys	Ala	Arg	Ile	Ile
V	Thr	Gly	Pro	Cys	Arg	Ala	Met	Ile	Ser
VI	Cys	Ser	Cys	Thr	Lys	Ser	Met	Pro	Pro
VII	Gly	Arg	Cys	Thr	Lys	Ser	Ile	Pro	Pro
VIII	Cys	Arg	Cys	Thr	Lys	Ser	Ile	Pro	Pro
IX	Val	Met	Cys	Pro	Met	Val	Tyr	Asp	Pro
X	Val	Met	Cys	Pro	Lys	Val	Tyr	Asp	Pro
XI	Val	Gly	Cys	Pro	Lys	Val	Tyr	Asp	Pro
XII	Ser	Pro	Val	Thr	Arg	Asp	Leu	Arg	Tyr
XIII	Thr	Ile	Val	Thr	Met	Glu	Tyr	Arg	Ile

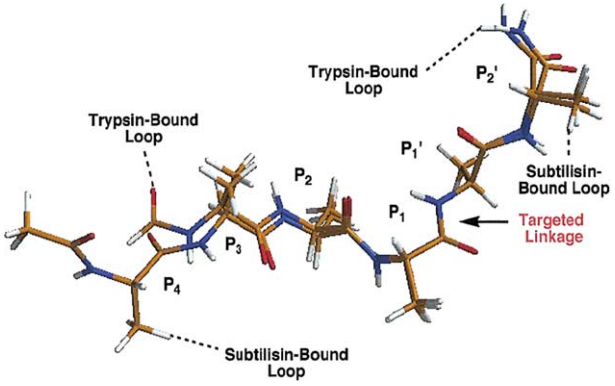


Figure 2. A comparison of the mean conformations for the trypsin- versus subtilisin-bound loop of the protein inhibitors in complexes I–VIII and IX–XIII, respectively. The targeted linkages are superimposed.

trypsin and subtilisin complexes, respectively; 'Ψ' refers to the backbone torsion N–C_α–C–N). However, ab initio minimizations of an isolated dipeptide, Ac-Ala-Ala-NH_2, show no strong preference for or aversion to these values.[20] To date, we cannot explain the bound conformation in terms of ground-state binding other than to say that it complements very well the extended active site of each enzyme.

By Kraut's proposal, the conformation must anticipate some feature of an extended TS geometry. Sometimes in chemistry, long-range structure can be important by defining an orbital alignment for long-range charge dynamics (e.g., ring currents, conducting polymers, Grob-type fragmentations). Pursuing our analysis from this point of view was more fruitful.

2.2. Orbital alignments in the bound loop

The mean conformation of the subtilisin-bound loop effectively contains two extensive alignments of sp^3, sp^2, and p atomic orbitals (Fig. 3). The longest alignment runs from C_α of the P_1' subsite through at least P_4 (Fig. 3a). It is defined by six mobile torsions and four ω torsions as listed in the figure (ω is the backbone torsion C_α–C–N–C_α; Φ is the backbone torsion C–N–C_α–C). The imperfections in the ten local orbital alignments defined by these 10 torsions are each less than 20°. A second, shorter alignment extends from the side chain of P_1 to C_α of P_3' (Fig. 3b). It is defined by four mobile torsions and two ω torsions. The imperfections in the local alignments defined by these six torsions are each less than 20° except for that of 30° in the torsion $C_{\beta(P1')}$–C_α–N–C_{P1}.

The mean trypsin-bound loop (Fig. 4) contains two extensive orbital alignments that are almost identical to those just described. The main difference is that the upstream alignment in the trypsin-bound loop does not extend beyond P_3 (Fig. 4a). All six of the torsions that define this alignment, as listed in the figure, sponsor local imperfections of less than 20°. Five of the six torsions that define the second alignment also sponsor local imperfections of less than 20°, but the remaining torsion is off by 31° (again it is the torsion $C_{\beta(P1')}$–C_α–N–C_{P1}; Fig. 4b).

2.3. Orbital alignments in the serine-bearing segment

The mean conformation of trypsin's active-site-serine-bearing segment contains two extensive orbital alignments (Fig. 5). The relevant residues are Ser195 to Pro198. The longest alignment (Fig. 5a) extends from at least C_β of serine to proline's nitrogen. While serine's H_δ was not observed in the complexes, a simple modeling of its position for optimal H-bonding with His57 suggests that serine's O–H bond continues the orbital alignment of Figure 5a (mean H_δ–O_γ–C_β–C_α torsion of −164°).[21] The net alignment is then defined by four mobile torsions and two ω torsions as listed in the figure. The imperfections in the six local orbital alignments defined by these six torsions are each less than 20°. A second, shorter alignment extends from serine's O_γ into the carbonyl of Gly196 (Fig. 5b). This alignment is defined by two mobile torsions and one ω torsion. Two of these

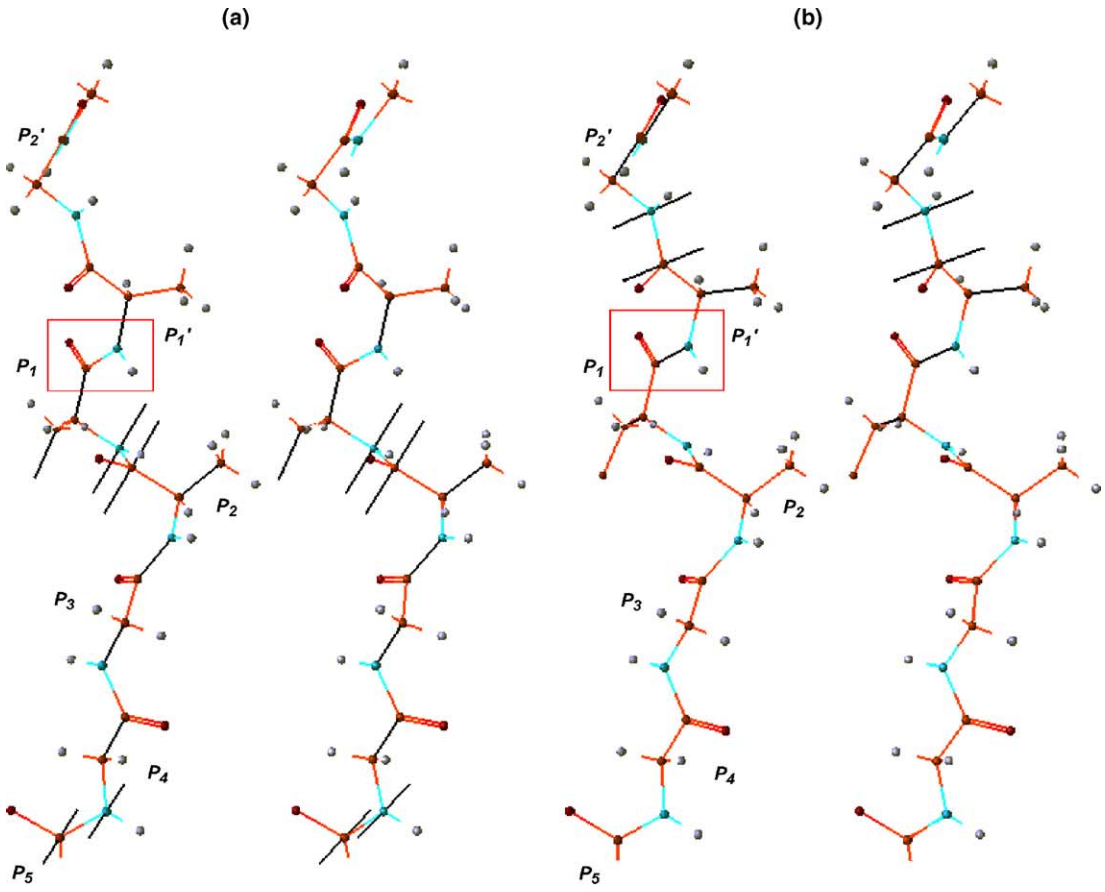


Figure 3. Stereo drawings of two extensive orbital alignments, (a) and (b), in the mean conformation of the *subtilisin-bound loop*. The participating bonds between heavy atoms and the participating p_z orbitals are drawn in black. The torsions that define each local orbital alignment are listed in order with their mean values. The values in parentheses are the approximate optimal values for alignment. The targeted linkage is boxed.

torsions sponsor local imperfections of less than 20°, but the third carries an imperfection of 32° (O_γ–C_β–C_α–C of serine itself). This imperfection might be fully corrected during an acyl transfer event, as we will discuss below.

Inter-system similarity again supports an importance for these alignments. That is, in the subtilisin system (Fig. 6), there are two extensive alignments that match closely the alignments just described for trypsin. The relevant subtilisin residues are Ser221, Met222, and Ala223. The long alignment (Fig. 6a) is slightly less extensive than its counterpart in trypsin (Fig. 5a). The p_z orbitals of the Ser221/Met222 amide are rotated by 180° relative to their counterparts in trypsin at Ser195/Gly196, and so are distinguishing, but these p_z orbitals are aligned with the same adjacent sp^3 orbitals in each enzyme. The high overall quality of the alignments in subtilisin's serine-bearing segment is about the same as that of the trypsin alignments. The most unobliging torsion is again O_γ–C_β–C_α–C of serine itself (Fig. 6b), with a mean value of 135° versus its hypothetical optimum of 180°.

Since the trypsin and subtilisin systems share these extensive alignments in the serine-bearing strand, the inter-system similarity of active sites is greater than previously realized. In fact, save only for the differences

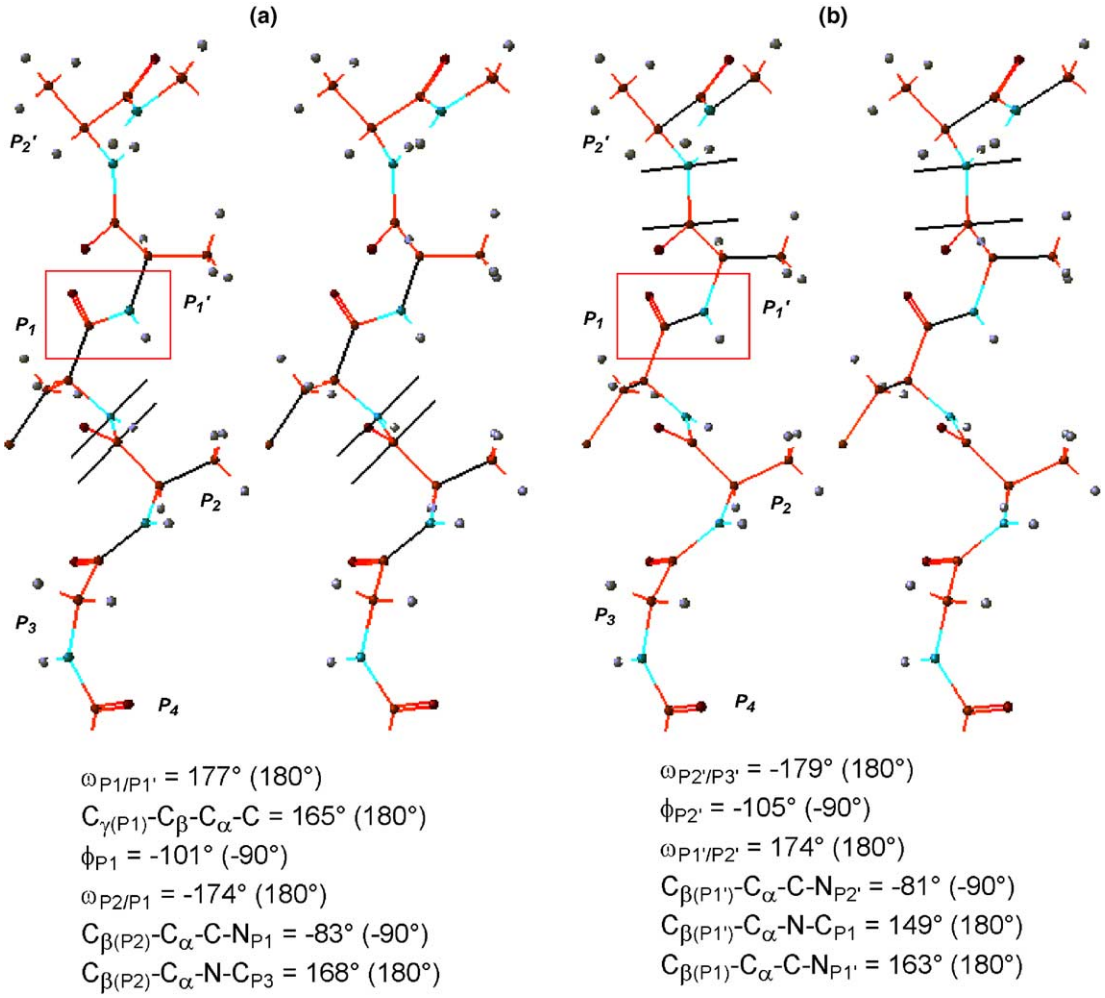


Figure 4. Stereo drawings of two extensive orbital alignments in the mean conformation of the *trypsin-bound loop*. The targeted linkage is boxed. Compare with Figure 3.

in the lengths of alignment, the atomic orbitals that make up all four extensive alignments (Figs. 3–6) are identical in the trypsin versus subtilisin systems. Note too that all four alignments converge at the targeted linkage and that their two main flaws are directly adjacent to the reacting atoms.

2.4. Correcting the alignments

If the orbital alignments have an importance to reaction kinetics, some of their imperfections might indicate part of the inhibitors' mechanism(s) of action. At present, we cannot argue for this. Another possibility is that certain mean imperfections are proper in a normal ground state and then substantially corrected in one or more transition structures. Consider the imperfection defined by serine's O_γ–C_β–C_α–C torsion. An improved orbital alignment at this torsion would be consistent with serine's attack on P_1. The mean $O_\gamma\cdots C(=O)_{P1}$ distance in the eight trypsin complexes is 2.7 Å (Fig. 7). When Swiss-PdbViewer[5] is used to widen the O_γ–C_β–C_α–C torsion in each trypsin complex to 180°, O_γ approaches the p_z axis of the targeted carbon and the mean $O_\gamma\cdots C$ distance shortens to 2.1 Å. When we widen the same torsion in the subtilisin complexes, O_γ approaches the p_z axis of the targeted carbon and the mean $O_\gamma\cdots C$ distance shortens from 2.8 to 2.0 Å.

As for the enzyme-bound loop, $C_{\beta(P1')}$–C_α–N–C_{P1} is the only torsion defining a mean imperfection greater than 20° in its local orbital alignment. This torsion involves the targeted C–N bond. The nature of the torsion will change during proteolysis, as will the nature of Ψ_{P1} and $\omega_{P1/P1'}$, since the hybridization of the targeted C and N atoms must change. Can we therefore imagine the imperfections at these torsions to be corrected during serine's attack?

Figure 8 shows a hypothetical transition conformation with 'perfected' values of $C_{\beta(P1')}$–C_α–N–C_{P1} and Ψ_{P1}. It was created simply from the trypsin-bound conformation of Figure 4 by: (1) giving an sp^3 geometry to each atom of the targeted bond, (2) maintaining the antiperiplanarity of the main chain along that bond, and (3) adjusting $C_{\beta(P1')}$–C_α–N–C_{P1} and Ψ_{P1} to their nearest fully staggered values. This process requires the least motion to remove the flaws at $C_{\beta(P1')}$–C_α–N–C_{P1} and

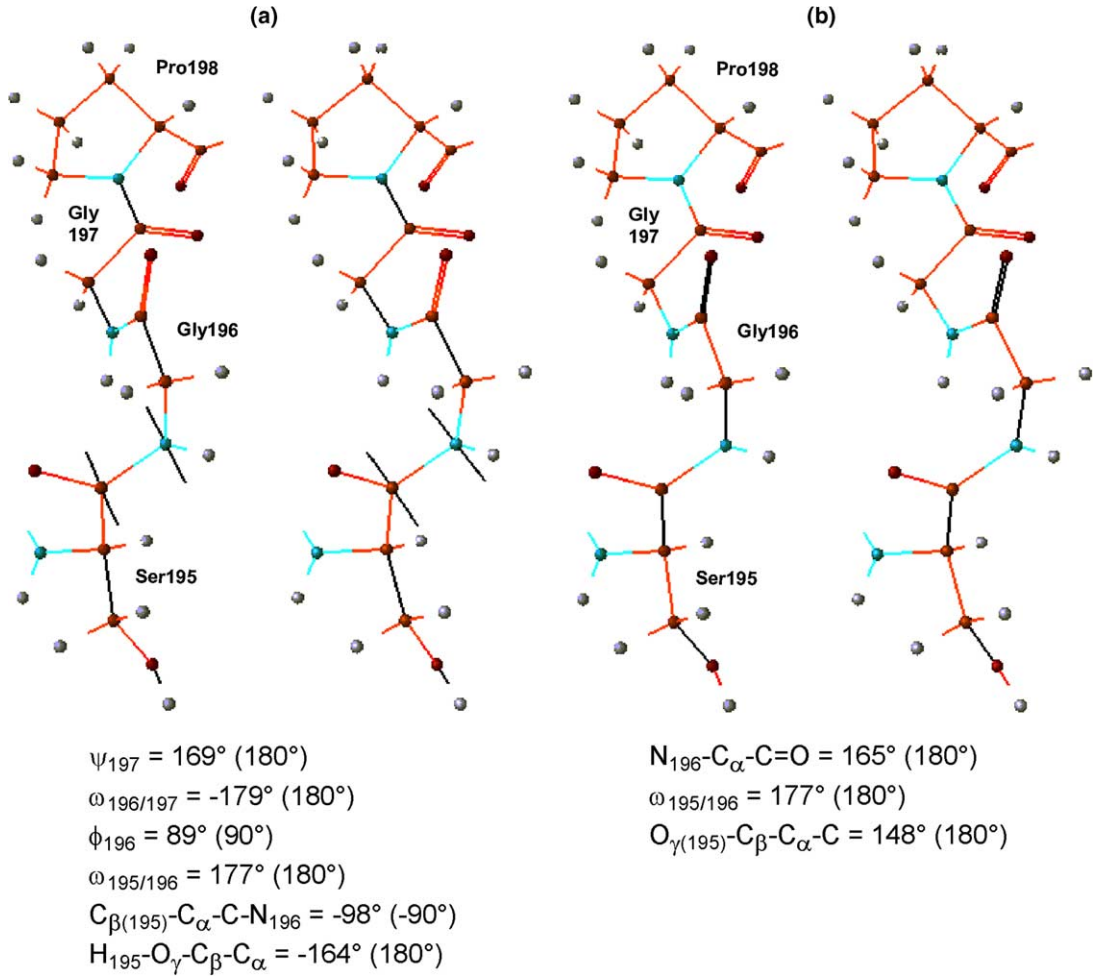


Figure 5. Stereo drawings of two extensive orbital alignments in residues 195–198 for *trypsin* complexes I–VIII. The side-chain oxygen of *serine 195* is trypsin's nucleophile in the proteolytic mechanism. Serine's H–O–C–C torsion was modeled (see text).

Ψ_{P1}. It also has two other notable traits. First, it disposes the nonbonded sp^3 orbital on each atom of the targeted bond toward the active serine hydroxyl group. Second, it creates an *additional* extensive orbital alignment reaching from P_3 to C_α of P_2'. The new alignment is defined by eight torsions. None of its local imperfections are greater than 20°. Bear in mind that its imperfections at $\phi_{P1'}$ and N_{P1}–C_α–C–O are each 0° by our definition of the conformation.

2.5. What do the alignments mean?

Are the alignments in Figures 3–6 perhaps just coincidental? Looking at models of a peptide strand, there are many conformations that contain extensive orbital alignments. We should not be surprised to find them in real systems. Yet, Figures 3–6 concern two genetically unrelated systems with nearly identical alignments in the same two key strands, with the same focus on the reacting atoms, with the same two main flaws, and with those flaws being directly adjacent to the reacting atoms. The alignments share twenty defining torsions between the two systems. For these twenty, each mean local alignment differs by an average of just 5° between the two systems. The largest difference is only 14°. This high similarity argues against coincidence. As an alternative reading of the similarity, perhaps the strand conformations are somehow important but the orbital alignments themselves are not. This possibility is less compelling since the pertinent segment of the serine-bearing strand does not have the same conformation in the two enzymes, as noted above. In any case, the more shared conformation of the bound strand has so far not inspired an interpretation other than the orbital alignments.

So what benefit could result from multiple extensive orbital alignments focussed on the reacting atoms? To us, the alignments imply an optimized 'through-strand' polarizability. We propose that the alignments' evolution has helped to optimize TS charge distributions as follows. The mean orbital alignments of the bound strand in each system (Figs. 3 and 4) are aligned with sp^2 lone-pair densities of the targeted carbonyl. During serine's attack, these alignments may help to delocalize the carbonyl lone-pair densities, thus helping the carbonyl oxygen to accept the increasing density from its own polarizing π bond (e.g., Fig. 9). The building charge on the carbonyl oxygen is thus an 'incentive' for lone-pair delocalization while the bound strand itself repre-

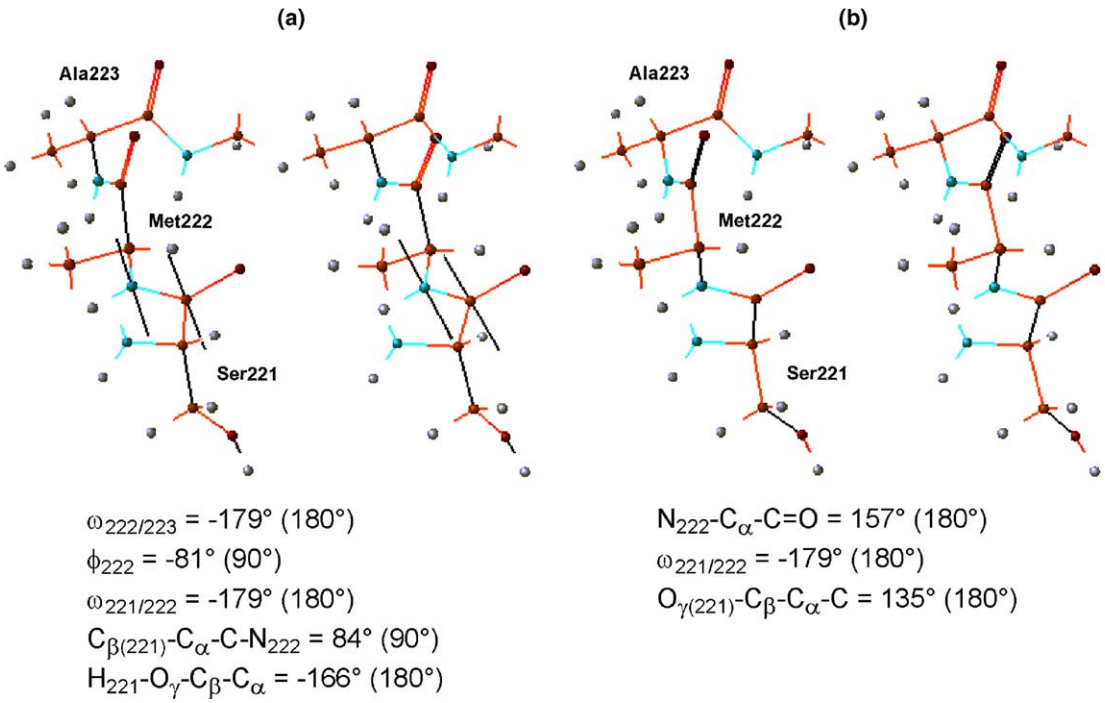


Figure 6. Stereo drawings of two extensive orbital alignments in residues 221–223 for *subtilisin* complexes IX–XIII. The side-chain oxygen of *serine 221* is subtilisin's nucleophile in the proteolytic mechanism. Serine's H–O–C–C torsion was modeled (see text). The side chain of Met222 has been truncated to provide a better view. Compare with Figure 5.

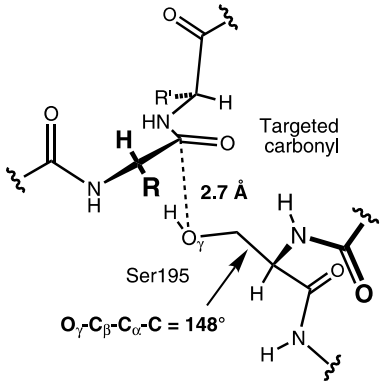


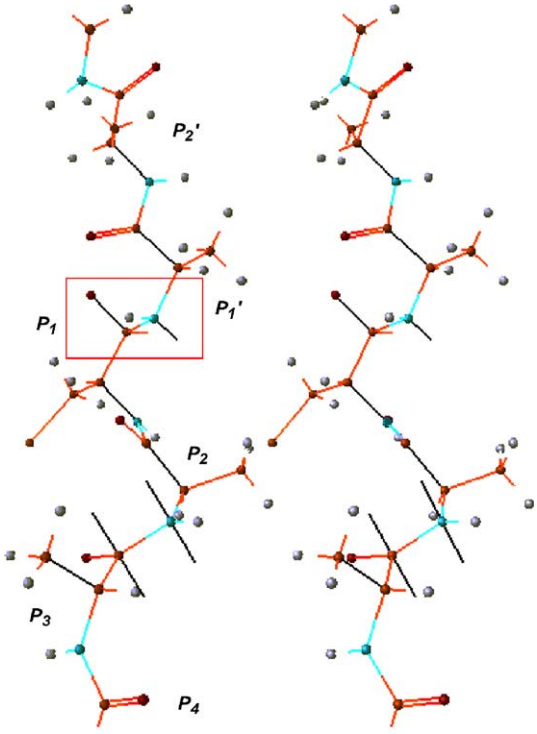
Figure 7. The mean values of the targeted carbonyl/Ser O_γ distance and the O_γ–C_β–C_α–C torsion in trypsin complexes I–VIII.

sents a charge sink. Meanwhile, the alignment shown in Figures 5a and 6a may help to delocalize the density of serine's breaking H–O_γ bond, that is, the serine strand acts as a charge sink to make the H–O_γ bond more labile. The other alignment in the serine strand (Figs. 5b and 6b) might help to export other density from O_γ to minimize electron/electron repulsion, it might help to import density to the emerging $O_\gamma\cdots$substrate bond, or it might even do both simultaneously at different electronic energy levels.

Delocalization that makes use of σ or mixed σ/π bond frameworks is referred to as 'hyperconjugation'.[23] The concept is very useful as a way of explaining relative stabilities and reactivities, and the impact of orbital alignments on these. It can be applied to many well-known phenomena such as the anomeric effect,[24] the Grob fragmentation,[25] and the β-effect of silicon.[26] It is in principle very general.[23,27] In cases involving oxyanion stabilization, the importance of hyperconjugation must logically depend on, among other things, the capacity of the available charge sink(s) and the alignment of relevant orbitals.

In the present context, it was the orbital alignments themselves that suggested the protein backbone as a possible charge sink. At least three more of protein's basic features encourage the idea: (1) 50% of its backbone heavy atoms are relatively electronegative; (2) 75% of its backbone heavy atoms make up amide groups (which offer low-energy π^* orbitals as components for larger antibonding MOs); (3) since those amide groups are closely linked to one another along the backbone and hydrogen bonded to other strands in their respective secondary structure, they must Coulombically sense one another and thus collaborate and cooperate in the global vibronics of their secondary structure; they might therefore act as a unified network in charge inductions or accommodations. The idea of individual strands operating as charge sinks is consistent with reactivity data mentioned below.

As with any theory, our interpretation of the alignments in Figures 3–6 cannot be proven correct. It is a working hypothesis that lets us make testable predictions. Consider the following. The rate constant k_{cat} can increase strongly with substrate length in the proteolysis of oligopeptides.[28] The dependence is seen with many proteases and substrates. Bizzozero and co-workers proposed that adding residues to a small substrate might have some electronic influence ('pure substitution effect') on



$\omega_{P1'/P2'}$ = 174° (180°)
$\phi_{P1'}$ = -60° (by defn.)
$\omega_{P1/P1'}$ = 177° (180°)
N_{P1}-C_α-C-O = 180° (by defn.)
$\omega_{P2/P1}$ = -174° (180°)
ϕ_{P2} = -70° (-90°)
$\omega_{P3/P2}$ = 175° (180°)
$C_{\beta(P3)}$-C_α-C-N_{P2} = -76° (-90°)

Figure 8. Stereo drawing of a hypothetical transition structure for the trypsin-bound loop. The atoms of the targeted linkage are pyramidalized and highlighted with a box. A new extensive orbital alignment is created with this conformation. Its component bonds and orbitals are drawn in black.

the reactivity of the scissile bond.[29] Kraut took the dependence as evidence that the enzyme is selectively binding an extended TS geometry.[2] We can now be more specific. By our interpretation, long oligopeptides of proper bound conformation will enjoy especially extensive accommodations of charge during the acyl transfer steps of proteolysis. And that interpretation forces us to an important prediction: *the enzyme should not be needed to observe a rate/length dependence*. Reaction rate should vary with substrate length and conformation even in the non-enzymatic solvolysis of oligopeptides.

After the present structural analysis was done, we performed some non-enzymatic solvolyses of simple oligopeptide esters. The results showed that an enzyme is not needed to observe a significant dependence of reaction rate on substrate length (Table 3). For example, in methanolic Hünig's base at ambient temperature, the

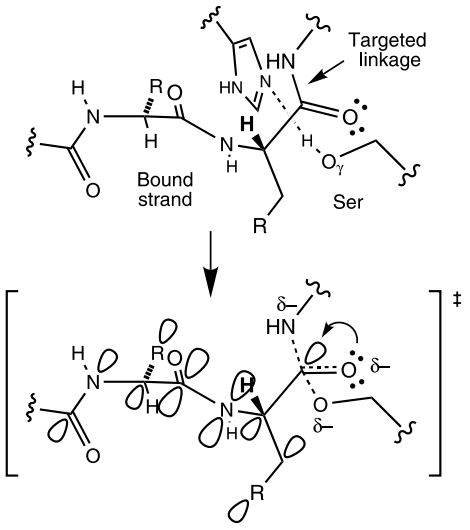


Figure 9. A hypothetical transition-state delocalization of carbonyl lone-pair density that is suggested by the atomic orbital alignments of Figures 3a and 4a. Such a delocalization should help the carbonyl oxygen to accept electron density from its polarizing π bond.

substrate Piv-Pro-Pro-OEt underwent transesterification (to its methyl ester) eight times faster than did Piv-Pro-OEt. Piv-Sar-Sar-OEt converted 15 times faster than Piv-Sar-OEt (Piv = pivaloyl, Pro = prolyl, Sar = sarcosyl). The results were also consistent with a strong dependence of rate on substrate conformation. For example, Ac-Pro-OEt converted to its methyl ester 720 times faster than did Piv-Pro-OEt, Piv-Sar-Pro-OEt converted 150 times faster than Piv-Pro-Pro-OEt, and Ac-Pro-Pro-OEt converted seven times faster than Piv-Pro-Pro-OEt.[31] The polarity of the medium argues against field effects as a major factor behind these results. Through-strand polarizability is not thereby proven, of course, but the data are hard to explain otherwise.

Table 3. Methanolysis of peptide-esters

	Series/substrate	Half-life[a,b,c]	Rel. rate
I	Piv-Sar-OEt	30 h	1
	Piv-Pro-Sar-OEt	12 h	2.5
	Piv-Sar-Pro-Sar-OEt	5.0 h	6
	Piv-Sar-Sar-OEt	2.0 h	15
II	Piv-Pro-OEt	1700 d	1
	Piv-Pro-Pro-OEt	220 d	8
	Piv-Sar-Pro-Pro-OEt	15 d	110
	Piv-Sar-Pro-OEt	1.4 d	1200
	Ac-Pro-OEt	2.4 d	720
	Ac-Pro-Pro-OEt	30 d	56

[a] Substrates in series I were allowed to stand at ambient temperature in 1.03 M Hünig's base in methanol-d_4. NMR spectra were recorded at measured intervals.
[b] Substrates in series II were allowed to stand at ambient temperature in methanolic 2.06 M Hünig's base. Aliquots were withdrawn at measured intervals and quenched into excess AcOH/THF (10% v/v). Degree of conversion for substrates in this series was determined by GC/MS.
[c] Half-lives were calculated as described in Ref. 30.

2.6. Other predictions

We can make some further predictions based on our hypothesis of through-strand polarizability. First, if the orbital alignments in the serine-bearing segment are affecting reactivity, then simple empirical models should reflect a similar effect. We predict that the rates of non-enzymatic acylation and deacylation of serine-bearing oligopeptides will generally increase with those oligopeptides' lengths and depend on their conformational profile.

Second, if the alignments in Figures 3–6 are assisting in the optimal distribution of TS charges, the quality of alignment should improve as a reacting complex approaches a transition geometry. As theoretical calculations of trypsin or subtilisin transition structures come to treat more and more atoms in a high-level quantum mechanical mode, the results should generally converge toward such improvements.

Third, we predict the alignments of Figures 4 and 5 to be well conserved (or paralleled) by the trypsin(s) of any species as long as the enzyme's role of broad-specificity cleavage of Arg- or Lys-bearing sites is conserved. The other members of the chymotrypsin family should favor extensive alignments in their own Michaelis-like complexes. We also make corresponding predictions for the subtilisin(s) of any species and any other members of the subtilisin family.

Finally, we mentioned that a dependence of k_{cat} on substrate length has been found for many proteases. The present analysis and our simple kinetics study imply to us that all of those proteases should come to favor extensive orbital alignments in their substrates' mean bound conformation. Fairlie et al. have asserted that all four main classes of protease bind their competitive inhibitors in relatively extended β-strand conformations.[32] We expect that all four classes have in fact evolved to favor extensive orbital alignments. We predict that such alignments will be found in the mean conformation of bound peptides or peptide isosteres, and focussed on the targeted linkage, for any protease system showing k_{cat}/substrate-length dependence, and that they also always occur in the key active-site strands.

3. Conclusion

Extensive orbital alignments occur in the mean loop conformations bound by β-trypsin and subtilisin BPN′ and in the active-serine-bearing segment of each enzyme. The alignments in the bound loop are dictated by the three-dimensional fit of the loop in the active site. An importance of the alignments is implied by their near identity in these two systems, by their focus on the reacting atoms, and by the nature of their two main flaws. From the same facts and the known dependence of k_{cat} on substrate length in each system, we propose that evolution is shaping each active site to optimize through-strand polarizability in each system. By this theory, long substrates show high values of k_{cat} because they allow extensive accommodations of TS charge. A dependence of reaction rate on long-range structure should then also exist in non-enzymatic peptide solvolyses, and we have shown such a dependence.

The idea of optimized through-strand polarizability is not inconsistent with the idea of TS stabilization related to other features of trypsin and subtilisin such as the active site's 'oxyanion hole' and His/Asp diad. It seems a natural complement to Koshland's idea of intermolecular orbital steering.[33] We are pursuing the idea that both inter- and intramolecular orbital alignments contribute to optimal transition structures.

Acknowledgements

Acknowledgment is made to the Donors of the Petroleum Research Fund for support of our work (grant #29365-AC4). J.H. also thanks MIT and the University of Sherbrooke for providing outstanding environments in which to pursue the work.

References and notes

1. (a) Wright, C. S.; Alden, R. A.; Kraut, J. *Nature* **1969**, *221*, 235; (b) Kraut, J. *Ann. Rev. Biochem.* **1977**, *46*, 331, and references cited therein.
2. Kraut, J. *Science* **1988**, *242*, 533.
3. Protein Data Bank (http://www.rcsb.org/pdb/).
4. (a) Wright, H. T. *Eur. J. Biochem.* **1977**, *73*, 567; (b) Laskowski, M., Jr; Kato, I. *Ann. Rev. Biochem.* **1980**, *49*, 593; (c) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed.; W.H. Freeman: New York, 1985, pp 28–30; (d) Perona, J. J.; Craik, C. S. *Protein Sci.* **1995**, *4*, 337.
5. The geometries of the complexes were analyzed using Swiss-PdbViewer, version 3.5 (http://www.expasy.ch/spdbv/).
6. (a) Huang, Q.; Liu, S.; Tang, Y. *J. Mol. Biol.* **1993**, *229*, 1022; (b) Zhu, Y.; Huang, Q.; Qian, M.; Jia, Y.; Tang, Y. *J. Protein Chem.* **1999**, *18*, 505.
7. Bode, W.; Greyling, H. J.; Huber, R.; Otlewski, J.; Wilusz, T. *FEBS Lett.* **1989**, *242*, 285.
8. Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr. Sect. B* **1983**, *39*, 480.
9. Scheidig, A. J.; Hynes, T. R.; Pelletier, L. A.; Wells, J. A.; Kossiakoff, A. A. *Protein Sci.* **1997**, *6*, 1806.
10. Tsunogae, Y.; Tanaka, I.; Yamane, T.; Kikkawa, J.-i.; Ashida, T.; Ishikawa, C.; Watanabe, K.; Nakamura, S.; Takahashi, K. *J. Biochem. (Tokyo)* **1986**, *100*, 1637.
11. Luckett, S.; Garcia, R. S.; Barker, J. J.; Konarev, A. V.; Shewry, P.; Clarke, A. R.; Brady, R. L. *J. Mol. Biol.* **1999**, *290*, 525.
12. Li, Y.; Huang, Q.; Zhang, S.; Liu, S.; Chi, C.; Tang, Y. *J. Biochem. (Tokyo)* **1994**, *116*, 18.
13. Takeuchi, Y.; Satow, Y.; Nakamura, K. T.; Mitsui, Y. *J. Mol. Biol.* **1991**, *221*, 309.
14. Takeuchi, Y.; Noguchi, S.; Satow, Y.; Kojima, S.; Kumagai, I.; Miura, K.-i.; Nakamura, K. T.; Mitsui, Y. *Protein Eng.* **1991**, *4*, 501.
15. Heinz, D. W.; Priestle, J. P.; Rahuel, J.; Wilson, K. S.; Grütter, M. G. *J. Mol. Biol.* **1991**, *217*, 353.

16. McPhalen, C. A.; James, M. N. G. *Biochemistry* **1988**, *27*, 6582.
17. Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157.
18. A conformational similarity of trypsin- versus subtilisin-bound substrates is not a new idea. It followed readily from the discovery by Kraut et al. of the similar binding sites and it has been supported frequently by individual solved structures of many enzyme/inhibitor complexes.
19. Ramachandran, G. N.; Sasisekharan, V. *Adv. Prot. Sci.* **1968**, *23*, 283.
20. Minimizations were performed at the 6-31G* level using the PC Spartan Pro software, version 1.0.3, Wavefunction, Inc., Irvine, California, USA.
21. We modeled the position of serine's H_δ using the WebLab ViewerLite program.[22] The only suitable H-bond acceptor in each structure is the histidine nitrogen, N_ε. A range of values for the C_α–C_β–O_γ–H_δ torsion in each structure was established by giving the O–H bond a length of 1 Å and the C–O–H angle a tetrahedral value of 109.5°, then adjusting the C–C–O–H torsion to sweep the $H_\delta \cdots N_\varepsilon$ distance over a typical H-bonding range of 1.6 Å, if possible, to 2.0 Å. This procedure gave a small range of C–C–O–H torsional values for each structure, from 11° to 44° wide. All values for the thirteen complexes covered only a 50° range (−141° to +169°). For each complex, the position of H_δ was assigned as corresponding to the midpoint of each so-derived torsional range.
22. WebLab ViewerLite version 3.2 (http://www.msi.com/life/products/weblab/viewer/).
23. For a recent discussion of hyperconjugation with many recent references, see: Alabugin, I. V.; Zeidan, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 3175.
24. (a) Deslongchamps, P. *Stereoelectronic Effects in Organic Chemistry*; Pergamon: New York, 1983, Chapter 2; (b) Kirby, A. J. *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*; Springer: Berlin, 1983; (c) *The Anomeric Effect and Associated Stereoelectronic Effects*; Thatcher, G. R. J., Ed.; *ACS Symposium Series 539*; American Chemical Society: Washington, DC, 1993; (d) Juaristi, E.; Guevas, G. *The Anomeric Effect*; CRC: Boca Raton, 1994.
25. (a) Grob, C. A. *Angew. Chem., Int. Ed. Engl.* **1969**, *8*, 535; (b) Impact of bond alignment: see Ref. 24a, pp 257–266.
26. Lambert, J. B.; Zhao, Y.; Emblidge, R. W.; Salvador, L. A.; Liu, X.; So, J.-H.; Chelius, E. C. *Acc. Chem. Res.* **1999**, *32*, 183.
27. For some interesting recent examples concerning diastereoselective nucleophile additions to ketones and rates of (apparently) radical carbocyclizations, respectively, see: (a) Jones, C. D.; Kaselj, M.; Salvatore, R. N.; le Noble, W. J. *J. Org. Chem.* **1998**, *63*, 2758; (b) Deng, K.; Bensari, A.; Cohen, T. *J. Am. Chem. Soc.* **2002**, *124*, 12106.
28. (a) For examples, see: subtilisins BPN′ and Carlsberg: Morihara, K.; Oka, T.; Tsuzuki, H. *Arch. Biochem. Biophys.* **1970**, *138*, 515; (b) β-Trypsin: Izumiya, N.; Uchio, H. *J. Biochem.* **1959**, *46*, 645; Yamamoto, T.; Izumiya, N. *Arch. Biochem. Biophys.* **1967**, *120*, 497; (c) Porcine pancreatic elastase: Thompson, R. C.; Blout, E. R. *Biochemistry* **1973**, *12*, 57; (d) α-Lytic protease: Bauer, C.-A.; Brayer, G. D.; Sielecki, A. R.; James, M. N. G. *Eur. J. Biochem.* **1981**, *120*, 289; (e) *Streptomyces griseus* proteases A and B: Bauer, C.-A.; Thompson, R. C.; Blout, E. R. *Biochemistry* **1976**, *15*, 1291; Bauer, C.-A. *Biochim. Biophys. Acta* **1976**, *438*, 495–502; Bauer, C.-A. *Biochemistry* **1978**, *17*, 375; (f) Carboxypeptidase A: Izumiya, N.; Uchio, H. *J. Biochem.* **1959**, *46*, 235; Abramowitz, N.; Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *29*, 862; (g) papain: Berger, A.; Schechter, I. *Philos. Trans. R. Soc.* **1970**, *B257*, 249; Mattis, J. A.; Fruton, J. S. *Biochemistry* **1976**, *15*, 2191; (h) Aspartyl proteases: Sachdev, G. P.; Fruton, J. S. *Biochemistry* **1970**, *9*, 4465; *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 3424; Sampath-Kumar, P. S.; Fruton, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 1070; Hofmann, T.; Allen, B.; Bendiner, M.; Blum, M.; Cunningham, A. *Biochemistry* **1988**, *27*, 1140; Balbaa, M.; Cunningham, A.; Hofmann, T. *Arch. Biochem. Biophys.* **1993**, *306*, 297.
29. A factor involving the steric control of substrate orientation was also proposed and was seen as probably being more important. The idea of an electronic effect seems to have been eventually abandoned. See: Baumann, W. K.; Bizzozero, S. A.; Dutler, H. *FEBS Lett.* **1970**, *8*, 257–260; *Eur. J. Biochem.* **1982**, *122*, 251.
30. Kinetics data were collected as a series of signal integrations with a Bruker AC-300 NMR spectrometer and/or a VG Micromass ZAB-1F mass spectrometer. The half-lives were taken to be $t_{1/2} = 0.693/k$, where k was calculated by the least-squares method for the simple linear regression equation $\ln(\%RCO_2Et) = -kt + C$. Each half-life in Table 3 is the average of at least two determinations.
31. Ab initio calculations suggest that substrates may react most easily when they contain a cis amide linkage, and that Ac-Pro-OEt is better able to maintain a cis linkage during methanolysis than is Piv-Pro-OEt.
32. Fairlie, D. P.; Tyndall, J. D. A.; Reid, R. C.; Wong, A. K.; Abbenante, G.; Scanlon, M. J.; March, D. R.; Bergman, D. A.; Chai, C. L. L.; Burkett, B. A. *J. Med. Chem.* **2000**, *43*, 1271.
33. (a) Storm, D. R.; Koshland, D. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *66*, 445; For alternative views on orbital steering, see: (b) Bruice, T. C.; Brown, A.; Harris, D. O. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 658; (c) Scheiner, S.; Lipscomb, W. N.; Kleier, D. A. *J. Am. Chem. Soc.* **1976**, *98*, 4770; (d) Menger, F. M. *Acc. Chem. Res.* **1985**, *18*, 128.