

# Folding-Reaction Coupling in a Self-Cleaving Protein

Göran Wallin,<sup>†</sup> Torleif Härd,<sup>‡</sup> and Johan Åqvist<sup>\*,†</sup><sup>†</sup>Department of Cell and Molecular Biology, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden<sup>‡</sup>Department of Molecular Biology, Swedish University of Agricultural Sciences (SLU), Box 590, SE-751 24 Uppsala, Sweden

## S Supporting Information

**ABSTRACT:** Backbone torsional strain has been implicated as a cause of rate enhancement in a class of autoprocessing proteins performing proteolysis and protein splicing. In the autoproteolytic protein domain SEA, folding and proteolytic activity have experimentally been shown to be coupled with about 7 kcal/mol of folding free energy available for catalysis. Here, we have examined the catalytic strategy of SEA with molecular dynamics simulations, potential of mean force free energy profiles, and B3LYP/6-311G(d,p) density functional calculations. A quantitative estimate of the free energy stored as protein strain (about 8 kcal/mol), that is available for catalyzing the cleavage reaction, is obtained and found to be in excellent agreement with thermodynamic and kinetic data. It is further shown that there is strong coupling between folding and reaction coordinates leading to reactant state destabilization in the direction of folding and transition state stabilization along the reaction coordinate. This situation is different from the preorganized active site model in that the fully folded transition state stabilizing structure is not realized until the reaction barrier is surmounted.

## ■ INTRODUCTION

The model that has arguably been most successful in explaining the catalytic power of enzymes is that of the electrostatically preorganized active site.<sup>1,2</sup> The preorganization model posits that the enzyme spends part of its favorable folding free energy in arranging dipolar and/or charged groups in such a way that they become predisposed for favorable electrostatic interactions with the transition state.<sup>1</sup> This has been shown by computer simulations to allow both the stabilization of charged intermediates (for multistep reactions) and transition states, where the reduction of reorganization energy is a key feature.<sup>2–4</sup> It is, however, important to realize that electrostatic preorganization of functional enzyme groups toward particular substrate conformations and charges (i.e., transition states) is likely to result in a locally unfavorable structural arrangement in the absence of substrate(s). In fact, experimental data indeed show that enzymes sacrifice some of their thermal stability in creating the catalytic environment, and an inverse relationship between activity and thermal stability has been demonstrated for active site mutants of the T4 lysozyme.<sup>5</sup> This could thus be viewed as a type of local electrostatic strain, rather than steric, that pertains to the protein rather than to the substrate.

A similar relation between folding free energy and catalytic activity has recently been reported for the autoproteolytic protein domain SEA.<sup>6–9</sup> Proteolysis of the SEA domain is, however, a unimolecular reaction that does not require the binding of a substrate, and it appears to be assisted by backbone torsional strain. Furthermore, folding and proteolysis are coupled in SEA, which raises the interesting possibility that the rate enhancement in SEA may occur by a catalytic mode that is different from classical preorganization (Figure 1). The principles of catalysis in this case also presumably extend to a broader class of autoprocessing proteins that perform proteolysis and protein splicing by the same mechanisms.<sup>10,11</sup> NMR experiments indicate that structural rearrangements take place as the reaction progresses from the reactant to the

product state.<sup>6</sup> The situation thus appears to be different from active site preorganization in that the folding event and the proteolysis reaction are so tightly coupled that the transition state stabilizing fold is not realized until the proteolysis reaction actually passes through the rate-limiting transition state.

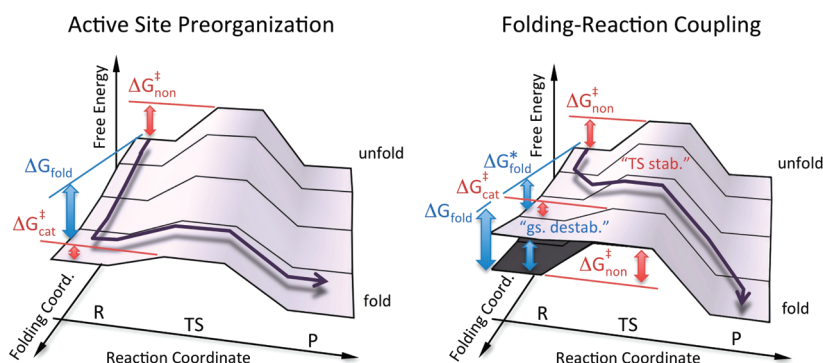
There are, of course, examples of other proteases that undergo autoproteolysis through “standard” mechanisms (e.g., HIV-1 protease<sup>12</sup>), involving sequence-specific peptide chain binding into subsites and peptide bond cleavage by aspartic acid or serine protease mechanisms.<sup>13</sup> The autoprocessing proteins studied here are different in that they share specific sequence motifs with a serine, threonine, or cysteine residue that is able to cleave its own preceding peptide bond through a distinct reaction mechanism: the N → O (or N → S) acyl shift.<sup>14,15</sup> In SEA, the N → O acyl shift is followed by hydrolysis of the ester bond (Figure 2A). In protein splicing, the initial N → O or N → S acyl shift is followed by additional reaction steps (that have been studied computationally by QM/MM methods<sup>16,17</sup>). The cleavage site in SEA is furthermore characterized by a conserved sequence G↓SVVV (cleavage at the arrow) that is located in a tight  $\beta$ -hairpin loop at the protein surface (inset of Figure 2B). The proteolysis reaction in SEA starts with the serine hydroxyl in the scissile peptide bond performing a nucleophilic attack on the preceding glycine carbonyl carbon, after which the intermediate hydroxyoxazolidine ring breaks up into an ester that is subsequently hydrolyzed to complete the peptide cleavage (Figure 2A).<sup>11,18,19</sup>

The rate enhancement is thought to be caused by torsional strain because (i) it has been shown experimentally that the serine hydroxyl and the tight  $\beta$ -hairpin loop are the only requirements for autoproteolysis, i.e., no other polar side chains

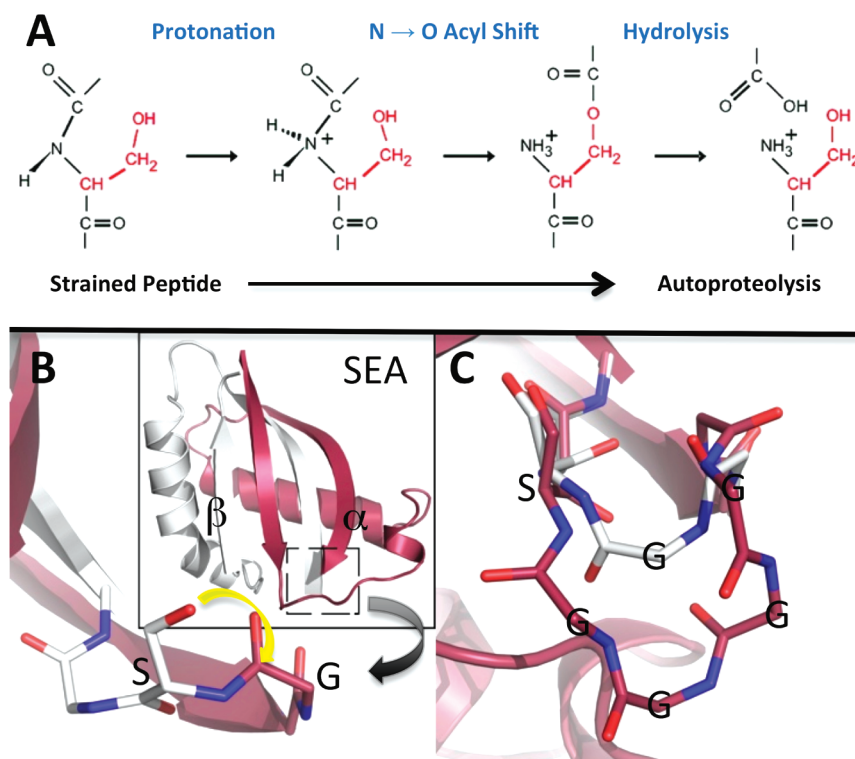
**Special Issue:** Wilfred F. van Gunsteren Festschrift

**Received:** February 17, 2012

**Published:** April 23, 2012



**Figure 1.** (Left) Simplified picture of the active site preorganization model of catalysis (not showing the substrate binding and release steps). In this model, the enzyme sacrifices some of its folding energy to establish an intrinsically (electrostatically) strained active site that is catalytically favorable upon substrate binding and enhances the reaction rate by transition state stabilization. In the ideal case of this model, the folding reaction runs perpendicular to the catalyzed reaction. In the autoproteolytic proteins studied here, the coupling between the folding and reaction coordinate is, however, much stronger. (Right) This results in an apparent ground state destabilization in the direction of folding, with an observable difference<sup>8</sup> in folding free energy of  $\Delta G_{\text{fold}}^* \sim 9 \text{ kcal mol}^{-1}$  on a strained noncleaving S1098A mutant of SEA and  $\Delta G_{\text{fold}} \sim 16 \text{ kcal mol}^{-1}$  measured on the fully folded cleaved wildtype and uncleaved 4G mutant. This results in a measured contribution<sup>8</sup> to the rate enhancement corresponding to  $\Delta G_{\text{strain}} \sim 7 \text{ kcal mol}^{-1}$ . This strain energy is reproduced in the umbrella sampling calculations presented here.



**Figure 2.** (A) Schematic overview of the amide protonation and N → O acyl shift in this class of autoproteolytic proteins. (B) The G↓S cleavage site (dashed box) in the fully folded and cleaved heterodimer of the SEA protein domain (inset) forms a strained  $\beta$ -hairpin loop in the uncleaved precursor. In this conformation, the serine hydroxyl is poised for attack on the carbonyl carbon of the preceding glycine residue (yellow arrow). The N-terminal fragment is shown in red and the C-terminal chain in white. (C) The GG↓SVVV and [G]<sub>4</sub>G↓SVVV precursor models are shown in white and red, respectively.

are conserved (or required) to catalyze the reaction, and the cleavage site is entirely solvent exposed;<sup>6,7</sup> (ii) structural modeling suggests a distorted conformation of the amide plane of the breaking peptide bond that is supported by extensive  $\beta$ -sheet hydrogen bonding and steric packing of conserved valine side chains;<sup>7</sup> (iii) removal of strain by insertion of a flexible glycine loop inhibits autoproteolysis;<sup>6</sup> (iv) the uncatalyzed N → O acyl shift becomes favorable at low pH<sup>15,18</sup> and amide  $pK_a$  increases with distortion (nonplanarity).<sup>20–24</sup> In a previous study, we combined experimental data on the pH dependence

of SEA autoproteolysis with quantum chemical calculations to show that the pH dependence can be explained if the amide nitrogen is upshifted from a  $pK_a$  of about  $-7$  (planar amide)<sup>25</sup> to  $pK_a = 4.2$  in a slow-cleaving SEA mutant, or  $pK_a = 6.3$  in wildtype SEA.<sup>9</sup> In fact, the  $pK_a$  was predicted to become  $>10$  for distorted secondary  $-\text{NH}_2-$  amines with  $\text{sp}^3$ -hybridized nitrogen atoms.<sup>9</sup>

Just as in the active site preorganization model, SEA has to sacrifice folding energy that is primarily of electrostatic origin (suboptimal H-bonding) to distort the amide plane (Figure 1).

Thermodynamic and kinetic measurements show that SEA has 7 kcal mol<sup>-1</sup> of folding free energy available for catalysis.<sup>8</sup> With the unfolded nascent polypeptide as the thermodynamic reference, the kinetic models further imply that this is due to a destabilization of the reactant state rather than a stabilization of the transition state. SEA is in fact arrested in a partially unfolded state in the absence of proteolysis (when, e.g., the nucleophilic serine is mutated).<sup>7,8</sup> This situation should be distinguished from the popular hypothesis that enzymes could, through nonbonded forces, sterically distort their substrates in such a way that the substrate conformation comes closer to that of the reaction transition state. In such models, rate enhancement would be attained by increasing the ground state energy of the *substrate* (by straining it), rather than by decreasing the energy of the transition state as postulated by Pauling.<sup>26,27</sup> An early example of the substrate strain hypothesis was put forward to explain the catalytic power of lysozyme based on crystallographic data that indicated a strained transition state analog bound to the enzyme.<sup>28</sup> More recent variants of the proposal are Menger's "split-site" model and related works,<sup>29,30</sup> as well as Bruice's so-called NAC (near attack conformer) hypothesis.<sup>31</sup> Warshel has, however, convincingly argued that enzymes are generally too flexible to exert strong steric strain on their substrates through nonbonded interactions and that the hypothetically stored steric strain therefore would rapidly be dissipated into the protein.<sup>2,32</sup> In the case of triosephosphate isomerase, where the split-site model has been invoked,<sup>29</sup> computer simulations also showed that the contribution of substrate strain to the catalytic effect was negligible.<sup>33</sup> It has further been shown that the NAC phenomenon, whose main signature is that the available substrate conformational space is reduced and pushed toward the transition state (the exact definition of a NAC remains somewhat unclear), is rather an effect of electrostatic transition state stabilization than the cause of catalysis and, thus, not a steric effect.<sup>34</sup> Autoproteolytic proteins such as the SEA domain are, however, different in that they do not bind a substrate and that their folding is arrested in the absence of proteolysis.<sup>6,7,9</sup>

The amount of catalytic strain available in SEA has been measured in two ways.<sup>8</sup> First, the free energy of unfolding was compared between a folding arrested Ser→Ala mutant (where there is no N → O acyl shift) and a four-glycine insertion mutant [G]<sub>4</sub>G↓SVVV (where strain is removed). The 4G loop-insertion mutant was found to be 7.4 ± 1.5 kcal mol<sup>-1</sup> more stable than the strained precursor mutant. Second, it was found that protein destabilization by hydrophobic core mutations or denaturant slowed autoproteolysis of wildtype SEA. This linkage between protein stability and cleavage kinetics could be accounted for if the conformational energy (strain) that limits precursor folding is 7.1 ± 0.2 kcal mol<sup>-1</sup>.<sup>8</sup>

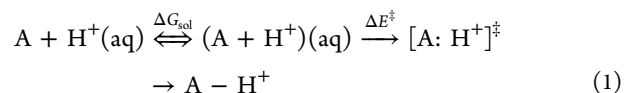
NMR spectra and kinetic and thermodynamic measurements, however, do not reveal how the strain is stored and, in particular, if the structure of the cleavage site in the folding arrested precursor can hold all of the observed 7 kcal mol<sup>-1</sup> of straining energy. Here, we use molecular dynamics simulations, free energy perturbation, umbrella sampling, and density functional theory (DFT) quantum chemical methods to estimate the rate enhancement in the SEA domain. We find that the calculated catalytic effect and the amount of strain present at the cleavage site are in excellent agreement with the thermodynamic and kinetic measurements. Furthermore, the coupling between the folding and reaction coordinate is shown to be too strong for SEA to be described by the preorganized

model of catalysis. Instead, SEA is described by a folding-reaction coupled mode of catalysis in which there is reactant state destabilization acting along the folding coordinate and transition state stabilization in the direction of the reaction. These two free energy components are also quantified.

## METHODS

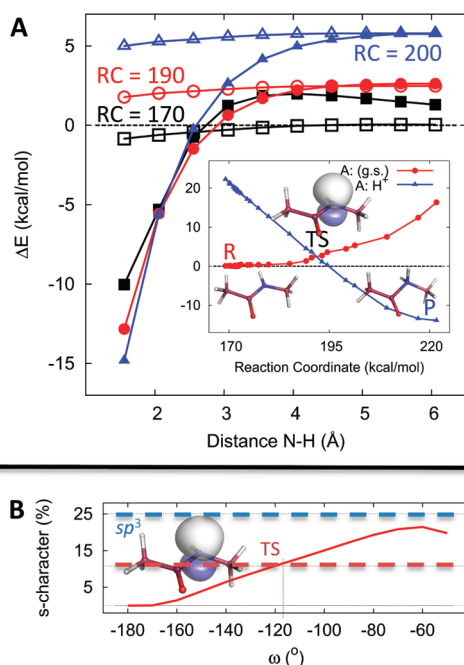
Molecular dynamics and free energy perturbation (FEP) simulations were used to investigate the extent to which there is mechanical strain acting across the breaking G↓S peptide bond of the uncleaved SEA domain precursor. To this end, quantum chemical DFT methods on a minimal model were used to determine where on the  $\omega$  torsion angle  $\angle C_\alpha^i C N C_\alpha^{i+1}$  of the scissile peptide bond the transition state of the rate limiting amide nitrogen protonation is found. Then, the relative free energies along this angle were estimated with umbrella sampling. The umbrella sampling calculations were carried out for an existing model of the uncleaved precursor,<sup>7</sup> a glycine-serine dipeptide in water and models of two additional glycine insertion mutants [G]<sub>n</sub>G↓SVVV,  $n = 1$  and 4, where this mechanical strain is partly and completely removed.

**Transition State Search.** The SEA domain is suggested to enhance the rate of proteolysis by straining the conformation of the amide plane of the scissile peptide bond. Breaking the amide plane by applying constraints such that it approaches the protonated conformation was shown previously to shift the pK<sub>a</sub> to the physiological range.<sup>9</sup> What was not shown, however, was the link between protonation of the amide nitrogen and the  $\omega$  angle of the peptide bond. Amide protonation is a downhill process after initial formation of the (elongated) covalent N–H bond. There is, however, an intrinsic barrier occurring after the formation of the solvent cage that is associated with the initial ionic electron configuration between the amide nitrogen and the partially desolvated proton.



Here, the peptide bond is represented by an N-methylacetamide, and the electronic [A:H<sup>+</sup>]<sup>‡</sup> transition state was located as a function of the torsion angle  $\angle CCNC$  in the following way: First, the charged singlet state of the protonated N-methylacetamide was optimized in the *trans* conformation (B3LYP/6-311G\*\*). Then, one of the amide hydrogen atoms was removed from this structure while retaining the basis set and integration grid points in its place. This structure was then relaxed to the sp<sup>2</sup> hybridized deprotonated and uncharged singlet state of methylacetamide in small steps (<0.05 Bohr or rad). The set of intermediate structures from this optimization was used as an approximate reaction path connecting the adiabatic ground states of the deprotonated reactant state (left structure in the inset of Figure 3A) and the protonated product state in the *trans* conformation (right structure in the inset of Figure 3A). The reaction coordinate on this path is taken to be the difference in electronic energy between the adiabatic deprotonated A: state and the protonated covalent A – H<sup>+</sup> state given these nuclear coordinates after partial optimization of the N–H bond distance. The diabatic electronic energy of the A:H<sup>+</sup> state was estimated indirectly by computing the wave function of the adiabatic A: state and then freezing the electronic configuration in the presence of the hydrogen atom. The diabatic energy in this approximation was seen to give the correct separation energy at long N–H distances (>100 Å).





**Figure 3.** (A) Protonation of the  $sp^2$  hybridized amide nitrogen in the *trans* conformation of N-methylacetamide passing over a 2.0 kcal mol $^{-1}$  barrier (filled black squares) with respect to the adiabatic ground state (open black squares, energies relative to the A: reactant state). This barrier is associated with the approach of the proton (abscissa) and subsequent formation of an ionic electron configuration A:H $^+$ . The transition state (center molecule in the inset) is found where the diabatic A:H $^+$  state crosses the adiabatic ground state (inset). At the reaction coordinate of the transition state (RC = 190 kcal mol $^{-1}$ , c.f. inset abscissa), the barrier vanishes (red filled circles) against the destabilization of the ground state (red open circles). This transition state cannot be further stabilized beyond this point by moving along the reaction coordinate (c.f. filled and open blue triangles, RC = 200 kcal mol $^{-1}$ ). The electronic state of the transition state is characterized by  $sp$  hybridization of the nitrogen lone pair orbital (inset). The gradual increase of basis set superposition error as the proton approaches is apparent in the adiabatic ground state energies (open markers). (B) Degree of  $s$  character (ordinate) of the amide nitrogen lone-pair orbital (inset) versus the  $\angle$ CCNC torsion angle corresponding to the  $\omega$  angle in a peptide bond (abscissa). The electronic state corresponding to the transition state of protonation on the amide nitrogen (in A) is found by identifying the degree of  $sp$  hybridization, here seen to be at  $-120^\circ$  (inset). The maximum value of  $sp$  hybridization that can be accomplished by twisting along this angle is  $sp^{3.7}$  at  $-60^\circ$ .

The energy of the protonation barrier is found to be 2.0 kcal mol $^{-1}$  and is seen to gradually vanish the further along the reaction coordinate that protonation is attempted (Figure 3A). The electronic transition state is identified as the crossing between the adiabatic A: state and the indirectly estimated diabatic A:H $^+$  state (inset of Figure 3A). There is no further gain in energy if protonation occurs after this point (Figure 3A). Natural bond orbitals<sup>35,36</sup> show that the lone-pair orbital of the amide nitrogen has 11%  $s$  character in this state, so it is, in fact, about half-way to becoming  $sp^3$  hybridized. Finally, a relaxed scan of the  $\angle$ CCNC torsion angle of the deprotonated N-methylacetamide was performed from *trans* to *cis*. The electronic transition state as a function of this torsion angle was located by monitoring the extent of  $sp$  hybridization of the N lone-pair orbital: in the *trans* conformation, it participates fully in the aromatic  $\pi$  bond with a 100%  $p$  character. When twisting

along the torsion angle, the lone-pair orbital picks up  $s$  character in a nearly linear fashion (Figure 3B) and is seen to match the  $sp$  hybridization of the transition state at a torsion angle of  $\pm 120^\circ$ . The planarity angle of the amide is  $134^\circ$  in this state, which is near the  $138^\circ$  of the transition state. The planarity angle in protonated and fully  $sp^3$  hybridized N-methylacetamide is  $124^\circ$ . These calculations were done with density functional theory using the B3LYP functional<sup>37–41</sup> and Pople's split valence 6-311G(d,p) basis set<sup>42–44</sup> in Gaussian 09.<sup>45</sup>

As a test, the above methodology was further examined adding diffuse functions to hydrogens and first-row atoms, which has a very slight effect. The transition state is moved 0.5 Å closer to the amide nitrogen where the energy barrier is 0.3 kcal/mol higher. As expected, the BSSE at this point all but vanishes for this basis set (0.02 kcal/mol). The functional dependence was also tested. This was done with the MPWB1K functional that has been shown to give good results for hydrogen transfer barriers.<sup>58</sup> With this functional, the barrier is lowered by 0.10 kcal/mol.

**Loop Modeling.** Glycine insertion mutants  $[G]_nG\downarrow SVVV$  of the uncleaved precursor with  $n = 1$  and 4 glycines were modeled starting from the lowest energy NMR structure<sup>6</sup> of cleaved wildtype SEA (PDB ID 2ACM). These mutants will be referred to as the 1G and 4G mutants. The initial models were subjected to simulated annealing simulations in a dielectric continuum with  $\epsilon = 80$ . In these simulations, the  $[G]_nG\downarrow S$  residues were fully flexible, and residues within a 10 Å radius were restrained to the initial configuration by 80 kcal mol $^{-1}$  Å $^{-2}$  positional restraints. Everything outside this region was kept under strong positional restraints, and nonbonded interactions were excluded. The system was heated to 3200 K and kept there for 200 ps using a time step of 0.5 fs. A total of 3000 snapshots were extracted from this run that were subsequently quenched to 50 K and optimized. To get an idea of the diversity of solutions found with this method, 100 low energy 4G loop models from these runs are shown in the Supporting Information.

We selected the loop models from these 3000 conformers based on the lowest optimized solute and restraint energies. In the case of the 4G mutants, all of the lowest energy conformers were in the *trans* conformation with respect to the scissile  $G\downarrow S$  peptide bond. For the 1G mutants, there were some low energy *cis* peptide bond conformers that were excluded from this scoring, because we ultimately want to compare with the *trans* conformer of the uncleaved wildtype precursor.

The eight lowest energy *trans* conformers for each mutant were re-equilibrated and rescored with respect to the average solute energy in short MD runs at 300 K with explicit TIP3P waters in a 20 Å sphere. These MD runs were equilibrated for 68 ps followed by 30 ps production runs. The average solute energies for these models are shown in Supporting Information Figures 2 and 3. The final 1G and 4G loop models are shown in Figure 2C.

**Umbrella Sampling.** In umbrella sampling,<sup>46</sup> a biasing potential is used to drive the system from one thermodynamic state to another. Either these data can then be reweighted with the microstate probabilities of the biasing potential to produce an unbiased ensemble average<sup>47</sup> or free energy perturbation can be used on the potential energy differences to estimate the free energy profile on the desired energy surface.<sup>3</sup> In the latter approach, the generalized reaction coordinate  $q = U^A - U^B$  is defined to be the difference in microstate energy between the

end states, A and B. Here, we associate these energies with the force field potential as

$$\begin{aligned} U^A &= U^{\text{FF}} - U_{\text{tor}}^{\text{FF}}(\omega) + U_{\text{bias}}^A(\omega) \\ U^B &= U^{\text{FF}} - U_{\text{tor}}^{\text{FF}}(\omega) + U_{\text{bias}}^B(\omega) \end{aligned} \quad (2)$$

The force field terms associated with the  $\omega$  torsion,  $U_{\text{tor}}^{\text{FF}}(\omega)$ , are replaced with the biasing potential. We let the biasing potential be a harmonic torsion potential

$$U_{\text{bias}} = \frac{k}{2}(\omega - \omega_0)^2 \quad (3)$$

with a force constant of 75 kcal mol<sup>-1</sup> rad<sup>-2</sup> that yields good sampling of the reaction coordinate. The starting state is the *trans* conformation of the breaking peptide bond  $\omega_0 = \pm 180^\circ$ , and the final state is close to the *cis* conformation  $\omega_0 = -10^\circ$ . By this choice, the transformation is limited to be clockwise along the  $\omega$  angle. We then map between the end states A and B in  $n$  discrete steps by running molecular dynamics simulations on the effective potential  $U^{\text{eff}}$ .

$$U^{\text{eff}} = (1 - \lambda_n)U^A + \lambda_n U^B, \lambda_n \in [0, 1] \quad (4)$$

During the simulations, the energies  $U^A$ ,  $U^B$ , and  $U^{\text{FF}}$  are recorded every 50th time step. In the postprocessing of these runs, we collect the recorded energies in bins with respect to the generalized reaction coordinate spanning the maximum and minimum recorded gap energy  $U^A - U^B$ . The umbrella sampling free energy is calculated for every bin  $q$  of width  $\Delta q$  as a weighted average over the collection of configurations  $\{i\}$

$$e^{-\beta \Delta G_{\text{US}}(q)} = \langle e^{-\beta \Delta G^{\text{eff}}(\lambda(i))} \langle e^{-\beta (U_i^{\text{FF}} - U^{\text{eff}}(\lambda(i)))} \rangle_{\lambda(i)} \rangle_{q(i) \in [q, q + \Delta q]} \quad (5)$$

where  $\beta = 1/RT$ . Note that configurations from different steps  $\lambda_n$  on the mapping potential may appear in the same bin. Each of these configurations is weighted with its associated effective potential and free energy from step  $\lambda_n = \lambda(i)$ . The free energy of each mapping potential  $U^{\text{eff}}(\lambda_n)$  in eq 2 is estimated from an average of the forward and reverse FEP free energy.

$$\Delta G^{\text{eff}}(\lambda_n) = \frac{1}{2}(\Delta G_{\text{fwd}}^{\text{eff}}(\lambda_n) + \Delta G_{\text{rev}}^{\text{eff}}(\lambda_n)) \quad (6)$$

The forward energy is given as

$$e^{-\beta \Delta G_{\text{fwd}}^{\text{eff}}(\lambda_n)} = \sum_{i=0}^{n-1} \langle e^{-\beta (U^{\text{eff}}(\lambda_{i+1}) - U^{\text{eff}}(\lambda_i))} \rangle_{\lambda_i}, n = 1, 2, \dots \quad (7)$$

and the reverse free energy is

$$e^{+\beta \Delta G_{\text{rev}}^{\text{eff}}(\lambda_n)} = \sum_{i=0}^{n-1} \langle e^{-\beta (U^{\text{eff}}(\lambda_i) - U^{\text{eff}}(\lambda_{i+1}))} \rangle_{\lambda_{i+1}}, n = 1, 2, \dots \quad (8)$$

With the choice of end state potentials in eq 2, the umbrella sampling exponent reads

$$\begin{aligned} U_i^{\text{FF}} - U^{\text{eff}}(\lambda(i) = \lambda_n) \\ = U_{\text{tor}}^{\text{FF}}(\omega_i) - ((1 - \lambda_n)U_{\text{A}}^{\text{bias}} + \lambda_n U_{\text{B}}^{\text{bias}}) \end{aligned} \quad (9)$$

For the  $\{i\}$  configurations belonging to the first step of the umbrella sampling, where  $\lambda_0 = 0$  and  $\Delta G^{\text{eff}}(\lambda_0) \equiv 0$ , the umbrella sampling expression measures the free energy

difference between the starting state biasing potential and the torsion potential.

$$U_i^{\text{FF}} - U_{\text{eff}}(\lambda(i) = \lambda_0) = U_{\text{tor}}^{\text{FF}}(\omega_i) - U_{\text{A}}^{\text{bias}} \quad (10)$$

For the remainder of steps, the right-hand biasing term in eq 5 cancels with the relative free energy of the mapping potential  $\Delta G^{\text{eff}}(\lambda_n)$  in eq 6, and the umbrella sampling expression reduces to the free energy measured on the  $\omega$  angle torsion potential.

$$e^{-\beta \Delta G_{\text{tor}}(q)} = e^{+\beta \Delta G^{\text{eff}}(\lambda_0)} e^{-\beta \Delta G_{\text{US}}(q)} \quad (11)$$

The point of this setup is to find out whether there is a stabilizing catalytic influence of the SEA domain fold along the  $\omega$  torsion angle.

Coordinate snapshots are recorded in every 100th step. Average structures for each umbrella sampling bin are calculated by matching the coordinate snapshots to the points collected in the bin. The bin average structures typically have contributions from several  $\lambda$  trajectories and span all of the simulated replicas.

**Data Collection.** The FEP calculations were staged in 100  $\lambda$  windows of 70 ps each for the protein runs and the dipeptide in water. With eight replicas per FEP run, three different loop models, and one dipeptide, this totals 224 ns of MD simulation time. Furthermore, the simulations of each  $\lambda$  window were divided into an equilibration and a production run. The changing part of the topology is referred to as the Q-atom system,<sup>48</sup> and the system is equilibrated until the total Q-atom energy becomes stationary after switching the effective potential. Trial calculations showed that 30 ps per  $\lambda$  step were necessary in these systems. The equilibration of the next  $\lambda$  step starts from the last step of the equilibration from the previous window. The idea is to first run a short equilibration simulation from the first  $\lambda$  window to the last to ensure phase space overlap and then collect data in separate runs that start from the last frame of the equilibration of that  $\lambda$  window. The lengths of the data collection runs were optimized separately from the equilibration runs, here 40 ps per  $\lambda$  step.

The idea with this approach is to run the data collection for each  $\lambda$  window independently and in parallel, and still maintain as much phase space overlap as possible. An alternative to this is Hamiltonian replica exchange<sup>49</sup> (HREMD) where the exchange probability represents a measure of the degree of total free energy overlap between the simulations. The drawback of HREMD is that the simulations are dependent on each other and need to be run simultaneously. The approach taken here is simpler: the data collection for the different  $\lambda$  windows can be run independently of each other. Drift is found by error estimation, e.g., forward and reverse hysteresis on the relative free energies. It should also be noted that phase space overlap issues are rather a question of sampling than the underlying method.

**MD Simulations.** MD simulations were run with the Q software package<sup>48</sup> in a 20 Å sphere centered on the C $_{\alpha}$  of the G $\downarrow$ S serine. The OPLS-AA force field<sup>50</sup> was used throughout, and the systems were solvated with TIP3P waters.<sup>51</sup> The solvated systems were heated to 300 K in seven consecutive steps, while at the same time releasing positional restraints applied to the heavy atoms of the enzyme and the solvent. The solvent heavy atom restraints were kept until 280 K to avoid the low density amorphous glass transition at  $\sim 125$  K ( $T_g$  measured for SPC/E water) and the extremely slow relaxation

times below this temperature.<sup>52</sup> An equilibration phase of 10 ps was performed without positional restraints before entering the umbrella sampling phase described above.

The SCAAS model<sup>48,53</sup> was applied to solvent molecules close to the border to reproduce the density and dipole angular distribution of bulk water. The nonbonded cutoff was set to 10 Å for all atoms inside the sphere, except for the G↓S atoms for which all nonbonded interactions were explicitly calculated. Long-range electrostatic interactions were treated with the local reaction field multipole expansion approximation,<sup>54</sup> whereas atoms outside the simulation sphere, which only interacted through bonded terms, were subjected to strong positional restraints.

**Error Estimation.** The sensitivity of these results was estimated by running eight umbrella sampling replicas with varying initial solvent configuration. If this is a small enough variation given the length of the simulations, these data could be combined directly into a single Boltzmann average. If, on the other hand, the initial perturbation is large enough to produce a number of distinct observations of the underlying ensemble distribution, the absolute free energy  $G_i^\circ$  of each simulation,  $i$ , would have to be estimated in order to evaluate the ensemble average directly.

$$e^{-\beta \overline{G^\circ}} e^{-\beta \Delta \overline{G}} = \langle e^{-\beta G_i^\circ} e^{-\beta \Delta G_i} \rangle_{\{i\}} \quad (12)$$

Here, we assume that each replica is distinct enough to be a unique observation of the underlying ensemble distribution  $P(X_i) = e^{+\beta \overline{G^\circ}} e^{-\beta G_i^\circ}$ , yet is still a good representative of the ensemble.

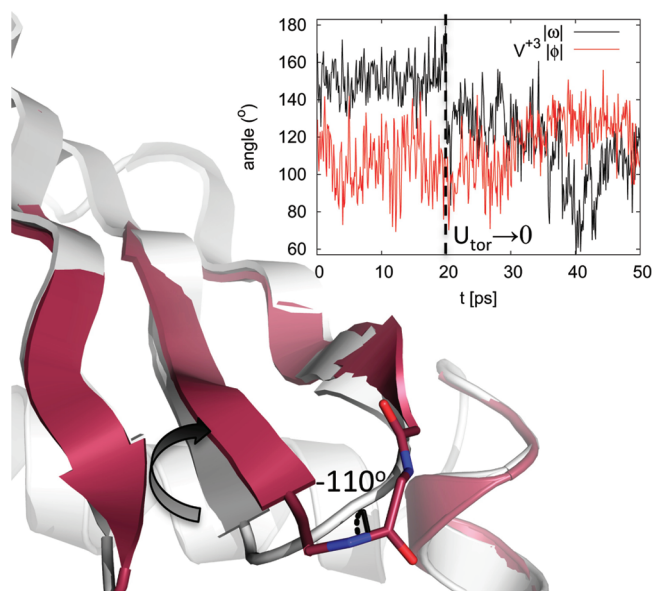
$$e^{-\beta \Delta \overline{G}} = \sum_i P(X_i) e^{-\beta \Delta G_i} \quad (13)$$

Under this assumption, the underlying ensemble distribution is estimated by resampling. Each individual umbrella sampling run is treated as a distinct observation of the ensemble, and the data  $\{U^A, U^B, U^{FF}\}_i$  are combined in a delete-4 jackknife estimate<sup>55</sup> on eq 5.

## RESULTS AND DISCUSSION

The free energy that is stored locally in the backbone  $\omega$  torsion angle of the scissile peptide bond of the autoproteolytic SEA domain precursor has been estimated with umbrella sampling. This was done with a previously published model of the precursor and glycine insertion mutants that can fold independently of proteolysis. The question is whether the free energy of this strain can be used to catalyze proteolysis. Already, an unbiased MD simulation of the solvated wildtype precursor model shows that the equilibrium value of  $\omega$  is significantly shifted from the common *trans* conformation ( $\pm 180^\circ$ ) to about  $-151^\circ$  (Figure 4), indicating that the torsion is unusually strained and approaching the  $sp^3$  hybridized transition state geometry found at  $\omega = -120^\circ$  (Figure 3). To determine the energy needed to obtain this conformation, and further twist along the  $\omega$  angle, free energy calculations were performed with umbrella sampling on the wildtype and  $[G]_n G \downarrow$  SVVV glycine insertion mutant models of SEA. If there is strain energy stored in the  $\omega$  torsion, it should be energetically favorable to twist along this angle in the uncleaved partially folded precursor compared to the glycine insertion mutants and a GS (glycine-serine) dipeptide in solution.

The flexible 4G mutant has been shown experimentally to be proteolytically inactive, whereas the 1G mutant cleaves at a rate that is about a 100-fold slower than the wildtype.<sup>7</sup> The glycine



**Figure 4.** Time series of the absolute values of two angles from a 50 ps MD simulation of the wildtype precursor (inset). After 20 ps, the  $\omega$  angle torsion potential is switched off, which results in a relaxation from the strained reactant equilibrium  $\omega$  angle of  $-151^\circ$  to a product angle of  $-110^\circ$ . This has a direct impact on the secondary structure of the  $\beta$ -hairpin loop, which relaxes simultaneously (black arrow), a change that is quantified by a representative Ramachandran angle: the  $\varphi$  angle of the third residue from the site of cleavage.

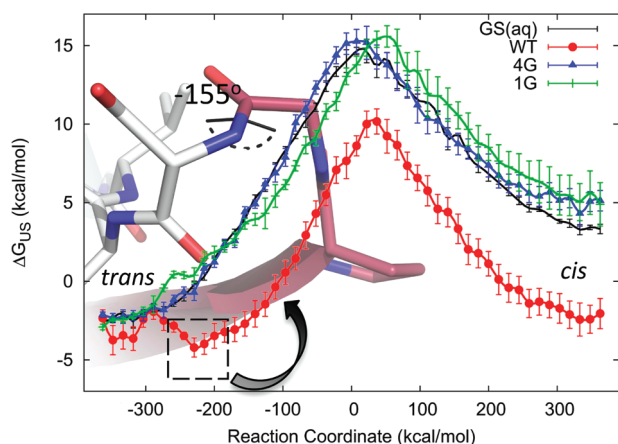
loops were in fact designed to prevent local mechanical strain from building up during protein folding. Of course, in the GS dipeptide, there is no strain present at all. The question is also whether a rate enhancement due to strain can be understood in terms of existing catalytic models: those of transition state stabilization by active site preorganization or ground state destabilization. This discussion leads us to propose a distinct catalytic strategy where the reaction coordinate of folding and the reaction coordinate of catalysis are strongly coupled during the reaction.

Since there are no 3D structures of the glycine insertion mutants 1G and 4G, they were produced with simulated annealing in an implicit solvent model. A total number of 6000 models were calculated, 3000 for each mutant, and they were scored with respect to their optimized solute energies. Then, the eight lowest energy models from this scoring were rescored with respect to the average solute energies from short MD simulations in explicit solvent with the same conditions as the umbrella sampling runs. The final models are shown together with the previously published model of the uncleaved wildtype precursor<sup>6</sup> in Figure 2.

The simulated torsion free energy profiles for the wildtype, the glycine insertion mutants, and the dipeptide in water are shown in Figure 5. The torsion free energy profile for the 4G mutant is statistically indistinguishable from that of the GS dipeptide. There is clearly no strain in the 4G mutant. In fact, there is no stabilizing effect at all, so the surrounding protein environment does not make any particular contribution to the free energy profile along this coordinate in the 4G mutant.

In contrast, the torsion free energy for the uncleaved wildtype precursor is consistently lower than that of the 4G and GS dipeptide in water. The uncleaved wildtype precursor stabilizes a state that is twisted  $25^\circ$  away from the *trans* conformation of the peptide bond. The DFT calculations show

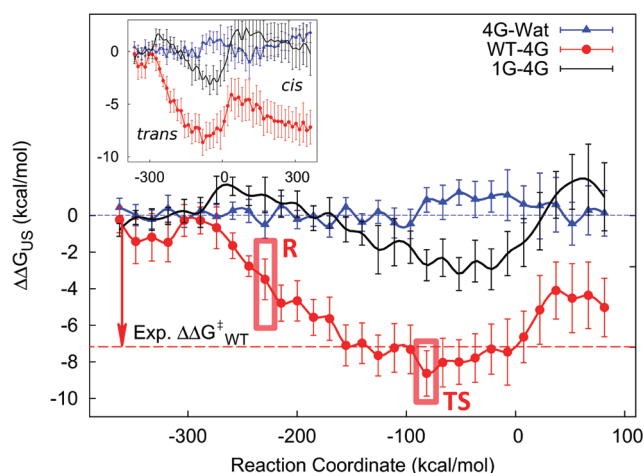




**Figure 5.** Free energy profiles (at 300 K) of the  $\omega$  torsion from *trans* to *cis* for the wildtype uncleaved precursor (red circles), the two glycine insertion mutants 1G (green) and 4G (blue triangles), and a GS dipeptide in water (black). The inset shows an average structure of the generalized reaction coordinate corresponding to the minimum free energy state of the uncleaved wildtype precursor (dashed box). In this state, the  $\omega$  torsion is shifted by  $25^\circ$  toward the *cis* conformation, and the lone-pair orbital of the nitrogen atom is half way of becoming  $sp^3$  hybridized.

that the N lone pair has picked up 4% s character at this  $\omega$  angle (Figure 3B), so the amide nitrogen has not yet fully acquired the electronic structure that favors protonation in this state. A bin average structure of this point on the reaction coordinate is shown in the inset of Figure 5 ( $\langle\omega\rangle_{US} = -155^\circ$ ). The unbiased MD simulation of the wildtype precursor (Figure 4) gives an average  $\langle\omega\rangle_{MD} = -151^\circ \pm 10^\circ$ , which is consistent with the umbrella sampling results. The relative free energy of this state is 5 kcal/mol lower than the corresponding state in the GS dipeptide and the proteolytically inactive 4G mutant, as shown in Figure 5. The free energy at the transition state,  $\langle\omega\rangle_{US} = -120^\circ$ , is downshifted by  $8.6 \pm 1.2$  kcal  $mol^{-1}$  (Figure 6), which is in excellent agreement with the experimentally measured value of  $7.1 \pm 0.2$  kcal  $mol^{-1}$ . The free energy profile for the 1G mutant, also shown in Figure 5, shows a slight shift of the torsional free energy barrier toward the *cis* conformation similar to that of the wildtype. The *trans* conformation is the energetically favored state in this mutant, as in the uncatalyzed 4G mutant and GS dipeptide. It can be noted here that the free energy difference between the *cis* and *trans* conformation in the wildtype precursor is much smaller than that generally found in peptides. This is in line with the occurrence of *cis* conformers in, e.g., the crystal structure of GyrA that undergoes post-translational protein splicing.<sup>56</sup>

We have previously shown that the pH dependence of autoproteolysis is consistent with an upshift of the  $pK_a$  for amide nitrogen protonation in the scissile peptide bond.<sup>9</sup> Quantum chemical methods were used in that study to show that breaking the amide plane forces the nitrogen to become  $sp^3$  hybridized and that this accounts for the shift in  $pK_a$ .<sup>9</sup> We show here that twisting the  $\omega$  angle by  $60^\circ$  is enough to realize the electronic structure favoring protonation (Figure 3), and the results from the umbrella sampling calculations indicate that this is also where the catalytic optimum of the SEA domain is found (Figure 6). There is clearly enough free energy to accomplish this  $pK_a$  shift in the uncleaved precursor. The previously published DFT calculations also showed that the ring strain in the hydroxyoxazolidine-like transition state of the



**Figure 6.** Relative free energy profiles (at 300 K) as a function of the generalized reaction coordinate  $U^A - U^B$  when twisting the  $\omega$  torsion angle from *trans* to *cis* (full profile in the inset). There is no catalytic energy found in this torsion angle for the uncleaved 4G insertion mutant compared with a solvated GS dipeptide (blue curve). There is, however, enough energy stored in the strained wildtype precursor,  $8.6 \pm 1.2$  kcal  $mol^{-1}$ , to account for the experimentally observed rate enhancement of  $7.1 \pm 0.2$  kcal  $mol^{-1}$  (red arrow) measured on the wildtype precursor (red curve). The strained reactant state,  $\langle\omega\rangle_{US} = -155^\circ$ , and the  $sp$  hybridized transition state of the protonation reaction,  $\langle\omega\rangle_{US} = -120^\circ$ , are shown for the wildtype precursor (R and TS, red boxes). Both reactant state destabilization and transition state stabilization are apparent in the wildtype. The 1G mutant however has  $3.2 \pm 0.9$  kcal  $mol^{-1}$  of strain available in this torsion that is seen to stabilize the transition state rather than destabilize the reactant state (black curve).

N  $\rightarrow$  O acyl shift (Figure 2A) becomes significantly reduced once the amide nitrogen is protonated.<sup>9</sup> This in turn lowers the enthalpy barrier of the reaction.

The umbrella sampling calculations presented here show that there is a locally unfavorable environment surrounding the G $\downarrow$ S cleavage site in the uncleaved precursor. Judging from the free energy profiles in Figure 5, the protein fold is seen to stabilize the torsion free energy in the *cis* direction along the  $\omega$  angle, favoring an  $sp^3$  hybridized amide nitrogen (Figure 3B). The question is then whether folding destabilizes the  $sp^2$  ground state of the amide plane or if the fold stabilizes the  $sp^3$  hybridized transition state of protonation. The thermodynamic measurements show, e.g., that the folding arrested Ser $\rightarrow$ Ala precursor mutant is destabilized compared to the 4G glycine insertion mutant with respect to the folding free energy. Kinetic measurements on SEA mutants where the fold was destabilized by mutations in the hydrophobic core are also consistent with a catalytic straining free energy of  $7.1 \pm 0.2$  kcal  $mol^{-1}$ .

To analyze this situation further, it is helpful to define the different thermodynamic reference states: the unfolded state UF, the fully folded and uncleaved state F, the partially unfolded state  $F^*$ , and the fully folded and cleaved state FC. If the folded state F is taken as the reference state, the observable difference in folding free energy between F in the fully folded 4G mutant,  $F_{4G}$ , and  $F^*$  in the wildtype precursor  $F_{WT}^*$  is interpreted as a destabilization of the ground state potential. When  $F^*$  is realized in the wildtype precursor, the  $\omega$  angle is shifted toward the rate limiting transition state in this reaction (Figures 3 and 5), which leads to the conclusion that the ground state is destabilized toward the transition state of protonation and that this enhances the rate of proteolysis. If,

however, the reaction is treated independently of the folding event and the *trans* proteolysis reactant states in  $F_{WT}^*$  and  $F_{4G}$  are defined to have the same free energy, the transition state appears to be entirely stabilized by the  $F^*$  fold (Figure 5). Note that the GS dipeptide also has virtually the same free energy profile as 4G in the absence of folding free energy. The choice of reference state can thus lead to different interpretations.

To resolve this problem it is useful to define an additional thermodynamic reference state: the  $F_{4G}^*$  state, where the 4G mutant has the secondary structure of the uncleaved wildtype precursor but does not harbor strain. Since the locally relaxed but partially unfolded  $F_{4G}^*$  state represents the orthogonal projection of the  $F_{WT}^*$  state onto the folding coordinate, this allows for a decomposition of the free energy contributions from reactant state destabilization in the direction of folding and transition state stabilization acting in the direction of the reaction, in line with the folding-reaction coupled model in Figure 1,  $\Delta\Delta G_{gs-destab}(F^*) = \Delta G_{fold}(F_{WT}^*) - \Delta G_{fold}(F_{4G}^*)$  and  $\Delta\Delta G^\ddagger(F^*) = \Delta G^\ddagger(F_{WT}^*) - \Delta G^\ddagger(F_{4G}^*)$ . The  $F_{4G}^*$  state is not experimentally observable, but the orthogonal projection  $\Delta\Delta G_{gs-destab}(F^*)$  is straightforward to calculate: it is the free energy difference when switching off the  $\omega$  torsion potential in the reactant state of the uncleaved wildtype precursor (i.e.,  $\langle\omega\rangle_{US} = -155^\circ$  in Figure 5) while maintaining its secondary structure. This does not require any additional simulations. The projected free energy difference can be obtained directly from the umbrella sampling calculations,  $\Delta\Delta G_{gs-destab} = \Delta G_{US}^{WT}(-155^\circ) - \Delta G_{US}^{WT}(-155^\circ, U_{tor}^{FF}(\omega) = 0) = 2.96 \pm 1.22$  kcal mol<sup>-1</sup>. Since the 4G mutant experiences the same uncatalyzed reaction barrier regardless of where it is on the folding coordinate, this leaves 4.5 kcal mol<sup>-1</sup> of the straining free energy for transition state stabilization  $\Delta\Delta G^\ddagger(F^*)$ . This transition state stabilizing energy is also seen in the relative free energy profile of the wildtype precursor in Figure 6. If the rate enhancement were entirely due to ground state destabilization, there would be no further stabilizing  $\Delta\Delta G^\ddagger$  free energy contribution in the wildtype precursor beyond the strained equilibrium angle  $\langle\omega\rangle_{US} = -155^\circ$ , which there clearly is (Figure 6). In fact, instantaneously switching off the  $\omega$  torsion potential term in the unbiased MD simulation of the strained wildtype precursor shifts the  $\omega$  angle by another  $50^\circ$  (Figure 4). In the absence of the resisting reactant state torsion potential, this  $\beta$ -hairpin loop further stabilizes an amide geometry that is closer to the transition state of protonation at  $\omega = -120^\circ$  (Figure 3B). The conclusion from this analysis is thus that there is both transition state stabilization and ground state destabilization acting in the system.

The 1G mutant has been shown experimentally to cleave at a rate that is 100-fold slower than the wildtype. This corresponds to a difference in activation free energy of about 3 kcal mol<sup>-1</sup>. In Figure 5, we see that the modeled 1G loop does not show any degree of reactant destabilization. It does however harbor  $3.2 \pm 0.9$  kcal mol<sup>-1</sup> worth of free energy for transition state stabilization (Figure 6). According to these calculations, the rate enhancement found in the 1G mutant is hence entirely due to transition state stabilization.

The above decomposition between transition state stabilization and reactant state destabilization is, however, only meaningful if the folding event is strongly coupled to the catalyzed reaction. If the secondary structure is realized independently of the catalyzed reaction (e.g., before the reaction takes place as in the recent folding-catalysis study by Roca et al.<sup>57</sup>), it is either transition state stabilization by

secondary structure preorganization or reactant state destabilization by arrested folding, depending on the choice of reference states. It is only when the secondary structure varies along the reaction coordinate that significant free energy components in folding-reaction space arise. Whether this is the case is straightforward to test, e.g., by monitoring the response in the  $\beta$ -sheet Ramachandran angles in the strained wildtype precursor after switching off the  $\omega$  torsion potential. This was done for the MD simulation discussed above, and the time series of a representative Ramachandran angle, the  $\varphi$  angle of the second valine residue in the cleavage site motif ( $V^{+3}$  in  $G^{-1}S^{+1}VV^{+3}V$ ), is shown in Figure 4. This particular angle changes from approximately  $-100^\circ$  to  $-130^\circ$ , i.e., toward a value that is more favorable in an antiparallel  $\beta$  sheet. In fact, the changes in Ramachandran angles are largely anticorrelated with the  $\omega$  angle during the short duration of this simulation, so the folding-reaction coupling is rather strong in SEA. In the 4G glycine insertion mutant, the local strain has been entirely removed. In terms of rate enhancement by folding-reaction coupling, the glycine insertions decouple the autoproteolysis reaction from the folding coordinate, since the 4G precursor can fold independently of the reaction coordinate. Wherever on the folding coordinate it tries to cleave, on the other hand, it is always facing the same uncatalyzed barrier (Figure 1).

## CONCLUSIONS

In summary, we studied here the autoproteolytic SEA protein domain, pinpointed the coordinate in which catalytic strain is stored, and quantified the amount of free energy available to catalyze the cleavage reaction. The results are in excellent agreement with thermodynamic and kinetic measurements on wildtype and glycine insertion mutants of SEA. A further analysis reveals a strategy for rate enhancement that none of the existing models of enzyme catalysis are fully capable of describing. The strong coupling between the folding and autoproteolysis reaction gives rise to significant free energy components in the direction of folding as reactant state destabilization and in the direction of the reaction as transition state stabilization. This emphasizes the role of folding-reaction coupling as a strategy to enhance the rate of proteolysis in this domain and other autoprocessing proteins where catalytic strain is found.

## ASSOCIATED CONTENT

### Supporting Information

Figures showing low energy loop mutant conformers and their two stage scoring results. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Phone: +46 18 471 4109. Fax: +46 18 53 69 71. E-mail: [aqvist@xray.bmc.uu.se](mailto:aqvist@xray.bmc.uu.se).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Support from the Swedish Research Council (VR) is gratefully acknowledged.



## ■ DEDICATION

It is a pleasure to dedicate this paper to Prof. Wilfred van Gunsteren on the occasion of his 65th birthday. Wilfred was the molecular dynamics mentor of J.Å. and has been a continuous source of inspiration for us over the years. He was also instrumental in getting the first Swedish protein dynamics simulation running back in 1983, when we were able to collect 5 ps a day for a small (58 residue) protein in vacuo on a VAX 750.

## ■ REFERENCES

- (1) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250–5254.
- (2) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.
- (3) Warshel, A.; Hwang, J. K.; Åqvist, J. *Faraday Discuss.* **1992**, *93*, 225–238.
- (4) Feierberg, I.; Åqvist, J. *Theor. Chem. Acc.* **2002**, *108*, 71–84.
- (5) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 452–456.
- (6) Macao, B.; Johansson, D. G.; Hansson, G. C.; Härd, T. *Nat. Struct. Mol. Biol.* **2006**, *13*, 71–6.
- (7) Johansson, D. G. A.; Macao, B.; Sandberg, A.; Härd, T. *J. Mol. Biol.* **2008**, *377*, 1130–1143.
- (8) Sandberg, A.; Johansson, D. G. A.; Macao, B.; Härd, T. *J. Mol. Biol.* **2008**, *377*, 1117–1129.
- (9) Johansson, D. G. A.; Wallin, G.; Sandberg, A.; Macao, B.; Åqvist, J.; Härd, T. *J. Am. Chem. Soc.* **2009**, *131*, 9475–9477.
- (10) Shao, Y.; Paulus, H. J. *Peptide Res.* **1997**, *50*, 193–198.
- (11) Paulus, H. *Chem. Soc. Rev.* **1998**, *27*, 375–386.
- (12) Babe, L. M.; Craik, C. S. *Cell* **1997**, *91*, 427–430.
- (13) Fersht, A. *Structure and Mechanism in Protein Science*; W.H. Freeman: New York, 1999.
- (14) Bergmann, B.; Brand, E.; Weinmann, F. *Hoppe-Seyler's Z. Physiol. Chem.* **1923**, *131*, 1–17.
- (15) Iwai, K.; Ando, T. *Methods Enzymol.* **1967**, *11*, 263–282.
- (16) Mujika, J. I.; Lopez, X.; Mulholland, A. J. *J. Phys. Chem. B* **2009**, *113*, S607–S616.
- (17) Mujika, J. I.; Lopez, X.; Mulholland, A. J. *Org. Biomol. Chem.* **2012**, *10*, 1207–1218.
- (18) Sakakibara, S.; Shin, K.; Hess, G. *J. Am. Chem. Soc.* **1962**, *84*, 4921–4928.
- (19) Levitin, F.; Stern, O.; Weiss, M.; Gil-Henn, C.; Ziv, R.; Prokocimer, Z.; Smorodinsky, N. I.; Rubinstein, D. B.; Wreschner, D. H. *J. Biol. Chem.* **2005**, *39*, 33374–33386.
- (20) Wang, Q. P.; Bennet, A. J.; Brown, R. S.; Santasiero, B. D. *J. Am. Chem. Soc.* **1991**, *113*, 5757–5765.
- (21) Greenberg, A.; Moore, D. T.; Dubois, T. D. *J. Am. Chem. Soc.* **1996**, *118*, 8658–8668.
- (22) Choo, S. J.; Cui, C.; Lee, J. Y.; Park, J. K.; Suh, S. B.; Park, J.; Kim, B. H.; Kim, K. S. *J. Org. Chem.* **1997**, *62*, 4068–4071.
- (23) Lopez, X.; Mujika, J. I.; Blackburn, G. M.; Karplus, M. *J. Phys. Chem. A* **2003**, *107*, 2304–2315.
- (24) Mujika, J. I.; Mercero, J. M.; Lopez, X. *J. Phys. Chem. A* **2003**, *107*, 6099–6107.
- (25) Fersht, A. R. *J. Am. Chem. Soc.* **1971**, *93*, 3504–3515.
- (26) Pauling, L. *Chem. Eng. News* **1946**, *24*, 1375–1377.
- (27) Pauling, L. *Nature* **1948**, *161*, 707–709.
- (28) Ford, L. O.; Johnson, L. N.; Machin, P. A.; Phillips, D. C.; Tjian, R. *J. Mol. Biol.* **1974**, *88*, 349–371.
- (29) Menger, F. M. *Biochemistry* **1992**, *31*, 5368–5373.
- (30) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. *J. Am. Chem. Soc.* **1999**, *121*, 11831–11846.
- (31) Bruce, T. C. *Acc. Chem. Res.* **2002**, *35*, 139–148.
- (32) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.
- (33) Åqvist, J.; Fothergill, M. *J. Biol. Chem.* **1996**, *271*, 10010–10016.
- (34) Shurki, A.; Strajbl, M.; Villa, J.; Warshel, A. *J. Am. Chem. Soc.* **2002**, *124*, 4097–4107.
- (35) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- (36) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (37) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (38) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (40) Stephens, P. J.; Devlin, F. J.; Ashvar, C. S.; Chabalowski, C. F.; Frisch, M. J. *Faraday Discuss.* **1994**, *99*, 103–119.
- (41) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (42) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (43) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (44) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–4.
- (45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Norm, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
- (46) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (47) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (48) Marelus, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics Modell.* **1998**, *16*, 213–225.
- (49) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (50) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (51) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Rw, I.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (52) Starr, F. W.; Sciortino, F.; Stanley, H. E. *Phys. Rev. E* **1999**, *60*, 6757–6768.
- (53) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647–3661.
- (54) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.
- (55) Efron, B. *Ann. Stat.* **1986**, *14*, 1301–1304.
- (56) Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R.; Sacchettini, J. C. *Nat. Struct. Biol.* **1998**, *5*, 31–36.
- (57) Roca, M.; Messer, B.; Hilvert, D.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13877–13882.
- (58) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43–52.