A Simulation of the Catalytic Mechanism of Aspartylglucosaminidase Using *ab Initio* Quantum Mechanics and Molecular Dynamics

Mikael Peräkylä[†] and Peter A. Kollman*

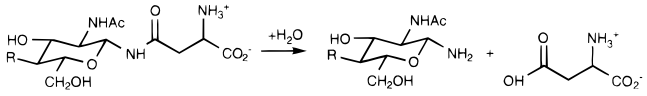
Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received August 19, 1996[⊗]

Abstract: *Ab initio* quantum mechanical (QM) and molecular dynamics (MD) methods have been used to study the catalytic mechanism of aspartylglucosaminidase (AGA)-catalyzed hydrolysis of an amide bond. QM active-site models were used to calculate the reaction sequences of the acylation and deacylation parts of the catalytic reaction at the MP2/6-31G*//HF/6-31G* level. MD simulations, in which the structures of the active sites were constrained to the geometries obtained from the QM model calculations, were carried out for all the species of the reaction sequence. Structural information from the MD simulations was used in the subsequent QM calculations in which the charge distribution of the protein environment and solvent was included as a set of atomic point charges. These QM calculations provided an estimate of the effects of the protein environment on the energetics of the catalytic reaction. The reaction mechanism was also simulated using a continuum model for the aqueous solvent, in order to differentiate general stabilization of polar and ionic groups from specific stabilization by enzyme groups. Using the results of our QM and MD simulations, we present a detailed mechanism of AGA-catalyzed hydrolysis reactions. The catalytic mechanism of AGA was found to be similar to the well-known mechanism of serine proteases. A novel feature of the mechanism of AGA is that AGA uses Oγ of *N*-terminal threonine as a nucleophile and the α-amino group of the terminus as a base in the catalytic reaction. These calculations showed that the protein environment differentially stabilizes the transition states of the reaction. The key transition state-stabilizing residues were Asp47, Lys207, and Arg211. Water molecules made a large contribution to the stabilization of the deacylation part of the reaction. In addition, we discuss how the catalytic nucleophile is activated in the case of AGA and other enzymes with similar catalytic machineries.

Introduction

Aspartylglucosaminidase (AGA) is a lysosomal enzyme that catalyzes the cleavage of *N*-linked oligosaccharide chains from asparagine.[1]



AGA is a medically interesting enzyme because mutations in the protein amino acid sequence cause the genetic disease aspartylglucosaminuria.[1,2] A recently determined X-ray structure[3] and mutational[4−6] and other biochemical studies[1,7,8] have shed light on the functional properties of AGA. However, the detailed catalytic mechanism of AGA is still largely unknown. AGA belongs to the newly-recognized structural superfamily of *N*-terminal nucleophile (Ntn) amidohydrolases.[9] The other Ntn hydrolases identified so far are penicillin acylase (PA),[10] proteasome (PRO),[11] and glutamine-PRPP-amidotransferase (GAT).[12] The enzymes of the superfamily share similar three-dimensional folds and have analogies in their functional mode. They also have similar architectures of their active sites with corresponding catalytic elements arranged in analogous ways.[9] On the basis of the X-ray structures and biochemical studies, the well-known serine protease-like catalytic mechanism[13−18] has been suggested for Ntn amidohydrolases.[10,11] A recent *ab initio* quantum mechanical study on the AGA-catalyzed reaction supported the idea that the serine protease-like catalytic mechanism is feasible for Ntn amidohydrolases.[19]

[†] Permanent address: Department of Chemistry, University of Joensuu, P.O. Box 111, FIN-80101, Joensuu, Finland.

[⊗] Abstract published in *Advance ACS Abstracts,* January 1, 1997.

(1) Mononen, I; Fischer, K. J.; Kaartinen, V.; Aronson, N. N., Jr. *FASEB J.* **1993**, *7*, 1247−1256.

(2) Ikonen, E.; Aula, P.; Grön, K.; Tollersrud, O.; Halila, R.; Manninen, T.; Syvänen, A.-C.; Peltonen, L. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11222−11226.

(3) Oinonen, C.; Tikkanen, R.; Rouvinen, J.; Peltonen, L. *Nature, Struct. Biol.* **1995**, *2*, 1102−1108.

(4) Tikkanen, R.; Riikonen, A.; Oinonen, C.; Rouvinen, J.; Peltonen, L. *EMBO J.* **1996**, *15*, 2954−2960.

(5) Riikonen, A.; Tikkanen, R.; Jalanko, A.; Peltonen, L. *J. Biol. Chem.* **1995**, *270*, 4903−4907.

(6) Fisher, K. J.; Klein, M.; Park, H.; Vettese, M. B.; Aronson, N. A., Jr. *FEBS Lett.* **1993**, *323*, 271−275.

(7) Kaartinen, V.; Mononen, T.; Laatikainen, R.; Mononen, I. *J. Biol. Chem.* **1992**, *267*, 6855−6858.

(8) Kaartinen, V.; Williams, J. C.; Tomich, J.; Yates, J. R. III; Hood, L. E.; Mononen, I. *J. Biol. Chem.* **1991**, *266*, 5860−5869.

(9) Brannigan, J. A.; Dodson, G.; Duggleby, H. J.; Moody, P. C. E.; Smith, J. L.; Tomchick, D. R.; Murzin, A. G. *Nature* **1995**, *378*, 416−419.

(10) Duggleby, H. J.; Tolley, S. P.; Hill, C. P.; Dodson, E. J.; Moody, P. C. E. *Nature* **1995**, *373*, 264−268.

(11) Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. *Science* **1995**, *268*, 533−539.

(12) Smith, J. L.; Zaluzec, E. J.; Wery, J.-P.; Niu, L.; Switzer, R. L.; Zalkin, H.; Satow, Y. *Science* **1994**, *264*, 1427−1433.

(13) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, *28*, 3629−3637.

(14) Daggett, V.; Schröder, S.; Kollman, P. *J. Am. Chem. Soc.* **1991**, *113*, 8926−8935.

(15) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569−6579.

(16) Nakagawa, S.; Yu, H.-A.; Karplus, M.; Umeyama, H. *Proteins* **1993**, *16*, 172−194.

(17) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331−358.

(18) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman & Co.: New York, 1985.

(19) Peräkylä, M; Rouvinen, J. *Chem. Eur. J.* **1996**, *2*, 1548−1551.

S0002-7863(96)02896-X CCC: $14.00

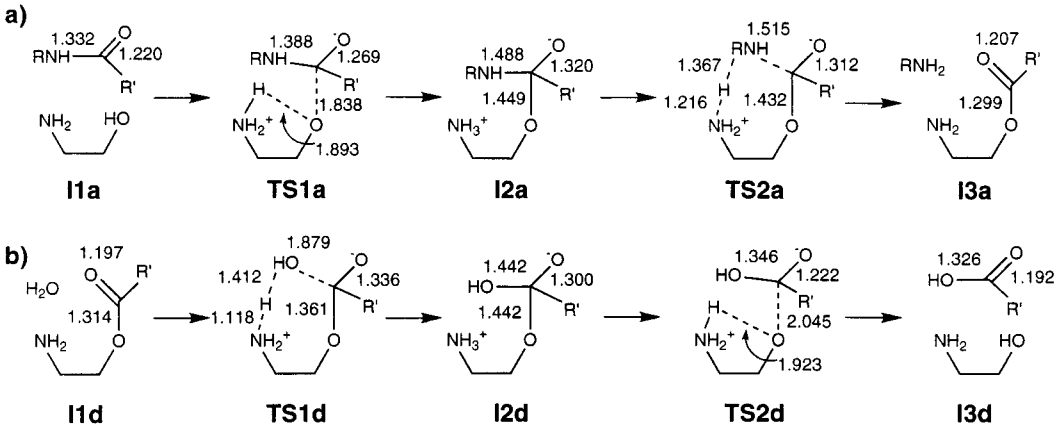


Figure 1. Reaction sequences of the (a) acylation and (b) deacylation steps of the AGA-catalyzed hydrolyses of an amide bond. Selected geometric parameters calculated at the HF/6-31G* level are shown.

A novel feature of the catalytic machineries of the Ntn amidohydrolases is that the nucleophile, which attacks the carbonyl carbon of the scissile amide bond, and the base, which facilitates the attack by accepting the proton from the nucleophile, are located in the same *N*-terminal amino acid. In the case of AGA and PRO the nucleophilic residue is the *N*-terminal threonine, in the case of PA, serine, and in the case of GAT, the *N*-terminal cysteine. The neutral α-amino group of the *N*-terminal amino acid, corresponding to histidine of the serine proteases, is thought to act as a base in the catalytic mechanisms of Ntn amidohydrolases. This is supported by the fact that the pH optima of the rate of the catalytic reactions are between 7 and 9 for these enzymes.[8,10,20] The oxyanion binding site (i.e., oxyanion hole) has also been identified from the structures of Ntn amidohydrolases. Since PRO catalyzes the hydrolyses of amide bonds of structurally diverse peptides,[21−23] this type of catalytic machinery represents a new fifth type of protease—the earlier discovered serine protease, cysteine protease, metallo-protease, and aspartic protease being the first four.[24]

In this work we have investigated the catalytic mechanism of the AGA-catalyzed hydrolysis of an amide bond by applying *ab initio* quantum mechanical (QM) and molecular dynamics (MD) methods. We have first calculated the reaction sequences of the acylation and deacylation parts of the enzyme-catalyzed reaction (Figure 1) utilizing a QM model of the protein's active site (Figure 2). The QM model consisted of the model substrate, *N*-terminal amino acid, and oxyanion hole. This model is large enough to include all the direct interactions that are present at the active site of AGA between the scissile amide bond and the enzyme. Second, in order to study the effect of the protein environment on the enzyme reaction, we carried out MD simulations for all the intermediates and transition states of the reaction. Structural information from the MD simulations was used in the subsequent QM calculations in which the charge distribution of the protein was included as a set of atomic point charges. These calculations provided an estimate of the effects of the protein environment and water on the energetics of the enzyme-catalyzed reaction.

On the basis of the computer simulations, which included a reasonably accurate QM description of the active-site region and detailed description of the protein environment, we present a detailed model of the mechanism of AGA-catalyzed cleavage of an amide bond. Since the Ntn amidohydrolases share similar functional and structural properties, this study, in addition to revealing the mechanism of the AGA-catalyzed reaction, can be taken as a representative model for the enzymes belonging to the Ntn amidohydrolase superfamily.

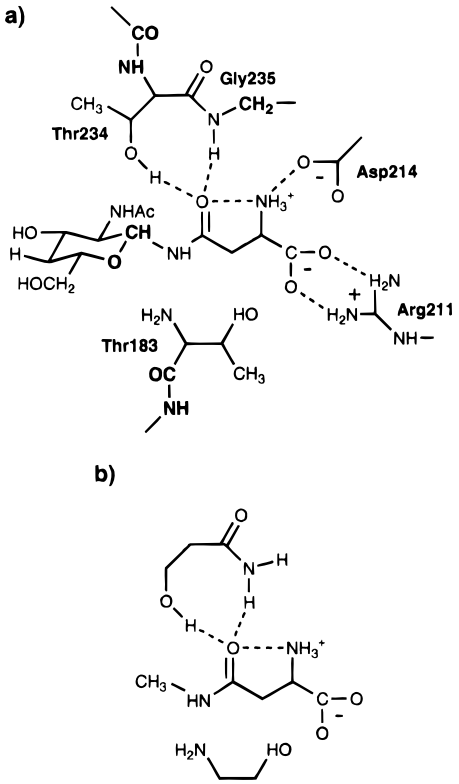


Figure 2. (a) A schematic picture of the active site of the AGA−*N*-acetylglucosamine-L-asparagine complex. Atoms shown in bold were omitted in the calculations of the electrostatic interaction energies between the active-site atoms and the protein environment. (b) Active-site model used in the QM calculations of the enzyme reaction.

Computational Details

The initial coordinates of AGA were obtained from the X-ray crystal structure of human lysosomal AGA complexed with the natural reaction product aspartate determined at 2.0 Å resolution.[3] The quaternary structure of AGA is a heterotetramer with two identical heterodimer halves. In order to reduce computer time, MD simulations were done using a heterodimer half of the protein. QM calculations of the enzyme reaction were done using an active-site model which was comprised

(20) Seemüller, E.; Lupas, A.; Zühl, F.; Zwickl, P.; Baumeister, W. *FEBS Lett*. **1995**, *359*, 173−178.
(21) Seemüller, E.; Lupas, A.; Stock, D.; Löwe, J.; Huber, R.; Baumeister, W. *Science* **1995**, *268*, 579−582.
(22) Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W. *FEBS Lett*. **1994**, *349*, 205−209.
(23) Wenzel, T.; Baumeister, W. *FEBS Lett*. **1993**, *326*, 215−218.
(24) Creighton, T. E. *Protein Structures and Molecular Principles*; W. H. Freeman & Co.: New York, 1983.

of the model substrate *N*δ-methylasparagine, *N*-terminal threonine (Thr183), as modeled by 2-aminoethanol, and the oxyanion hole, as modeled by 3-hydroxypropionamide (Figure 2). The starting structures of the QM calculations were based on the active-site geometry of the X-ray structure and the earlier *ab initio* QM calculations of the AGA-catalyzed reaction.[19] *Ab initio* quantum mechanical calculations were done with the Gaussian 94 program.[25] The geometries of the species studied (Figure 1) were optimized at the HF/6-31G* level, and the energies were further calculated at the MP2/6-31G*//HF/6-31G* level. Earlier calculations on the AGA-catalyzed reaction have shown that this level produces results that are similar to those found at the MP2/6-31+G*//HF/6-31+G* level.[19] Although the earlier calculations were done with a substrate model lacking the NH_3^+ and CO_2^- tails of the substrate, the geometries and the locations of the transition states are similar in the two studies. The effect of aqueous solvation on the energetics of the model enzyme reaction was estimated by using the continuum solvation model of Tomasi[26−28] as implemented in Gaussian 94 (IPCM option of Gaussian). In these calculations the solute cavity is determined from the electron density of the solute. A dielectric constant (ϵ) of 78.3 (water) and the gas-phase geometries were used in the solvation calculations.

The MD calculations were done for all the species of the acylation and deacylation pathways with the AMBER version 4.1[29] using the Cornell *et al*. force field.[30] Parameters for *N*-acetylglucose, which was used as a glycan moiety of the substrate, were taken from the work of Woods *et al.*[31] The van der Waals and torsion parameters of the active-site atoms, which were comprised of the same atoms as in the gas-phase calculations (Figure 1b), were taken from the corresponding amino acids of the Cornell *et al*. force field. The point charges of the active-site atoms were calculated with the RESP method using the gas-phase optimized geometries.[32] In the MD calculations the bond distances and angles of the active-site atoms were constrained to the corresponding quantum mechanically calculated values with large force constants (1000 kcal/(mol Å^2) for bond distances and 200 kcal/(mol rad^2) for angles). In addition, in order to keep the substrate in a proper position at the active site during the MD simulations, several distances were constrained in **I1a**, **I3a**, **I1d**, and **I3d** by harmonic forces of 10 kcal/(mol Å^2). In the Michaelis complex of the enzyme−substrate system (**I1a**) the substrate *β*-*N*-acetylglucosamine-L-asparagine was kept in a productive binding geometry by constraining the distances between the nucleophilic oxygen (Oγ of Thr183) and the carbonyl carbon of the substrate to 2.5 Å and the distance between the hydroxyl hydrogen and α-amino nitrogen of Thr183 to 2.1 Å. In this position, the nucleophilic Oγ is ready to attack the carbonyl carbon and the hydrogen is suitably positioned to be transferred to the base, the α-amino group of Thr183. In **I3a** the distance corresponding to the cleaved amide bond was constrained to 3 Å. The purpose of this constraint was to prevent the glycan moiety from leaving the active site. In **I1d** the water molecule which hydrolyzes the covalent intermediate was kept in a position where it is ready to deliver its proton to the α-amino group of Thr183 and attack the carbonyl carbon of the acyl-enzyme intermediate. This was done by constraining the distance between the water hydrogen and Nα of Thr183 to 2.0 Å and the distance between the water oxygen and carbonyl carbon of the covalent intermediate to 2.0 Å. Earlier three-dimensional Laue crystallography studies of trypsin[33] and MD simulations of the acyl-chymotrypsin intermediate[16] have indicated the attacking water would be located at this position prior to attacking the carbonyl carbon of the acyl-enzyme intermediate. Finally, in the enzyme−reaction product complex (**I3d**) the product was kept in the active site by a distance constraint of 3.0 Å between the Oγ of Thr183 and Cγ of the product asparagine.

In the MD simulations only the residues which contain atoms within 12 Å of the nucleophilic oxygen of the *N*-terminal Thr183 were allowed to move. A spherical 20 Å cap of TIP3P[34] water molecules centered on the Oγ of Thr183 was added, and harmonic forces of 1.5 kcal/(mol Å^2) were applied to any water leaving the 20 Å boundary. The SHAKE[35,36] algorithm was used to constrain the positions of hydrogens in the MD simulations. The simulations were performed at 300 K using a nonbonded cutoff of 10 Å and time step of 1 fs. The protein system, which had a total charge of −7.0, was neutralized by adding Na^+ ions in the positions of the largest negative electrostatic potential using program CION of the AMBER package. All the Na^+ ions were located at least 20 Å from the active site.

The optimized gas-phase structures of **I1a** and **I1d**, the first structures of the acylation and deacylation parts of the reaction, respectively, were first built at the active site of AGA. After Na^+ ions and the waters were added, the energy of the system was minimized for 2000 cycles using the conjugate gradient method. The initial MD simulations of 30 ps at 300 K were done for the models. These structures were used as starting points in the 130 ps MD simulations of **I1a** and **I1d**. The structure of **I1a** from the 30 ps simulation was used in building the initial models for the other species of the acylation part of the catalytic reaction. In analogy, the 30 ps structure of **I1d** was used in building models for the species of the deacylation part. These initial models were first minimized, and 30 ps equilibrium simulations were run. After that 130 ps MD simulations were performed. Structures at 70, 90, 110, and 130 ps were taken from the 130 ps simulations and used in the subsequent QM calculations in which the electrostatic interaction energies between the active site and the protein environment were calculated.

The active-site models, similar to those used in QM gas-phase calculations but lacking the CH_3 group bonded to the amide nitrogen of the model substrate (Figure 2b), were cut from the enzyme−ligand systems and terminated with hydrogens. All the residues which had atoms within 12 Å and water molecules which had atoms within 14 Å from the Oγ of Thr183 were included as point charges in the QM calculations. In order to avoid artifacts which might arise from the short distances between the QM atoms and point charges, several protein atoms which were directly bonded to the atoms of the QM active site were not included as point charges in the quantum mechanical calculations. These were the main chain amides (CONH) of Thr183 and Thr234, CαH_2 group of Gly235, and C1, H1, and O6 atoms of the carbohydrate ring. These atoms are shown in bold in Figure 2a. Charges from the Cornell *et al*. force field[30] were used for the protein atoms. The point-charge model had a total charge of −3.0 which was neutralized by increasing the charges on the carboxylate oxygens of Glu42, Asp81, and Asp238 by 0.5e. The carboxylate groups of these amino acids were the negatively charged groups furthest away (>14 Å) from the active site in the point-charge model. Point-charge calculations were done with Gaussian 92.[37] The electrostatic interaction energies of the protein environment with the quantum mechanical fragments (ΔE_{ES}) are presented in Table 1. It should be noted that

(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.;. Robb, M. A; Cheeseman, J. R.; Keith, T. A.; Petterson, G. A.; Montgomery, J. A.; Rachavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J. A. *Gaussian 94*, Revision B.3., Gaussian Inc., 1995.

(26) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(27) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.

(28) Wiberg, K. B.; Rablen, P. R.; Rush, D. J.; Keith, T. A. *J. Am. Chem. Soc.* **1995**, *117*, 4261−4270.

(29) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K., Kollman, P. A AMBER 4.1, University of California, San Francisco, 1995.

(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc*. **1995**, *117*, 5179−5197.

(31) Woods, R. J.; Dvek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.

(32) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(33) Singer, P. T.; Smalås, A.; Carty, R. P.; Mangel, W. F.; Sweet, R. M. *Science* **1993**, *259*, 669−673.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(35) Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, 1311−1327.

(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

Table 1. Electrostatic Interaction Energies (ΔE_{ES}) between the Active-Site Atoms and the Protein Environment at Points 70, 90, 110, and 130 ps of the MD Simulations

	ΔE_{ES} (kcal/mol)[a]				
	70 ps	90 ps	110 ps	130 ps	av
I1a	−109.1	−119.2	−128.4	−123.6	−120.1
TS1a	−143.2	−145.0	−134.5	−136.1	−139.7
I2a	−139.1	−130.8	−123.8	−145.7	−134.8
TS2a	−135.4	−131.6	−127.8	−130.5	−131.3
I3a	−136.2	−132.0	−118.3	−143.8	−132.6
I1d	−136.2	−136.0	−131.3	−149.4	−138.2
TS1d	−177.7	−161.2	−179.1	−166.7	−171.2
I2d	−167.7	−160.2	−173.2	−179.8	−170.2
TS2d	−148.0	−170.3	−150.0	−154.4	−155.7
I3d	−122.0	−147.0	−135.4	−142.9	−136.8

[a] $\Delta E_{ES} = E_{QM}$ (with point charges) − E_{QM} (without point charges).

Table 2. Relative Energies of the AGA-Catalyzed Reaction in the Gas Phase (ΔE(gas)), in Aqueous Solution (ΔE(aq)), and within the Enzyme (ΔE(enz) = ΔE(gas) + ΔE_{ES})

	ΔE(gas)	ΔE(aq)	ΔE(enz,av)[a]
I1a	0.0	0.0	0.0 (0.0)
TS1a	31.3	15.7	11.7 (21.5)
I2a	13.4	8.9	−1.3 (6.0)
TS2a	12.7	13.7	1.5 (7.1)
I3a	10.0	19.6	−2.5 (3.8)
I1d	0.0	0.0	0.0 (0.0)
TS1d	22.2	28.1	−10.7 (5.7)
I2d	13.5	3.5	−18.5 (−2.5)
TS2d	27.4	23.4	9.9 (18.6)
I3d	11.0	18.3	12.3 (11.7)

[a] ΔE(enz,av) is calculated using the average of ΔE_{ES} values from four snapshots of the MD simulation. Values in parentheses are calculated using half of the ΔE_{ES} values.

ΔE_{ES} includes both the electrostatic interaction of the rest of the protein with the active site atoms and the electronic polarization of the active-site atoms due to the electrostatic interaction with the rest of the protein atoms. The fluctuations of ΔE_{ES} values of Table 1 are 10−20 kcal/mol. In order to check the magnitude of the fluctuations of the electrostatic interaction energies during the MD simulations, we took 10 snapshots separated by 10 ps from the MD simulations of each of the reaction intermediates (total of 100 points) and calculated Coulombic interaction energies for the structures using the CARNAL module of AMBER 4.1. The fluctuations of the electrostatic interaction energies in these calculations were in the same range as in Table 1. Thus, although the total energy of the whole protein system fluctuates considerably during the MD simulation, the variations in the electrostatic interaction energies are as expected much smaller. In addition, it must be noted that the variations in the intramolecular energies of the solutes are not exactly zero with the constraints used. However, the effect of this on the relative electrostatic interaction energies is assumed to be small. Table 2 presents the relative quantum mechanical energies for the active site atoms in vacuo (first column), with the continuum model of solvation (second column), and including the average protein−active-site interaction energy from Table 1.

The contribution of each of the amino acid residues to the active-site−enzyme interactions was estimated by calculating the electrostatic interaction energies between the individual amino acid residues and the active-site atoms. Such residue-by-residue analyses were carried out for **I1a**, **TS1a**, **I1d**, and **TS1d**. The electrostatic interaction energies of the residue analyses are classical Coulombic energies between the active-site region and the amino acid residues and are averages of the four structures taken from the MD trajectory. Atomic charges of the active-site region were calculated with the RESP method for each of the structures. These purely Coulombic electrostatic interaction energies were found to be 10−15% smaller than the corresponding energies calculated using the QM active site, the latter of which included electronic polarization.

Results

The reaction sequences for the acylation and deacylation steps of the AGA-catalyzed reaction, with selected geometric parameters optimized at the HF/6-31G* level, are shown in Figure 1. The relative energies of the intermediates and the transition states in the gas-phase (ΔE(gas)) and with solvation energies included (ΔE(aq)) are presented in Table 2. Energies (ΔE(enz,av)) which include the electrostatic interaction energies between the active site and the protein environment are also presented in Table 2. In Table 2 ΔE(enz,av) values are reported both by using full ΔE_{ES} and half of ΔE_{ES} values. Our use of the electrostatic point charges directly in the quantum mechanical Hamiltonian enables the inclusion of both electrostatic and polarization effects of the reacting species due to the environment. However, this approach does not take into account the energy cost on the environment of reorienting its partial charges to stabilize the reactant/transition state/product. For example, when an ion is solvated in water, its solvation energy (and free energy) is about 1/2 that of the ion−water interaction energy because the water−water energy rises.[38] The factor 1/2 comes from classical electrostatics and linear responce theory.[39,40] Of course, enzyme active sites are "preorganized" so the factor 1/2 is likely to be the upper bound for the magnitude of the energetic cost of environmental reorientational polarization. That is why we present both the full inclusion of the enzyme environment (ΔE_{ES}) and multiplication of this value by 1/2 (in parentheses) in Table 2. Figure 3 shows the relative energies of the reaction sequences in the gas phase, in aqueous solution, and within the enzyme in which we use the electrostatic interaction energies multiplied by 1/2.

Model Enzyme Reaction. The first step in the AGA-catalyzed enzyme reaction is the attack of the nucleophilic Oγ of Thr183 on the carbonyl carbon of the substrate's amide bond. At the transition state of this attack (**TS1a**, Figures 1 and 4) the hydrogen of Oγ is almost fully transferred to the proton-accepting base, the α-amino group of Thr183, and the distance between Oγ and the carbonyl carbon is 1.84 Å. At **TS1a** the nitrogen of the amide bond has its lone-pair electrons antiperiplanar to the developing C−Oγ bond.[41] This transition state is similar to the one found by Daggett *et al.*[14] in the QM model calculations of the trypsin-catalyzed reaction. In the cases of PA and PRO, it has been suggested that proton transfer from the nucleophile to the base could be bridged by a water molecule.[10,21] However, the calculations presented here indicate that there is no room for such bridging water at the active site of AGA. Next in the reaction sequence, the tetrahedral intermediate **I2a** collapses to the acyl-enzyme intermediate through transition state **TS2a**. At **TS2a** the nitrogen of the amide bond has its lone pair pointing toward the α-amino group of Thr183, and the proton, which protonates the leaving amine group, is between the two nitrogens.

The deacylation step of the catalytic reaction is the reverse to the acylation. In this part of the reaction water first attacks the carbonyl carbon of the acyl-enzyme intermediate (**I1d** → **TS1d**). At the transition state **TS1d** (Figures 1 and 5), the proton from the attacking water is almost completely transferred to the α-amino group and the bond which is formed between

(37) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92*, Revision G.4, Gaussian, Inc., 1992.

(38) Sun, Y.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 3599−3604.

(39) Hansson, T.; Åqvist, J. *Protein Eng.* **1995**, *8*, 1137−1144.

(40) Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385−391.

(41) Dutler, H; Bizzozero, S. A. *Acc. Chem. Res.* **1989**, *22*, 322−327.

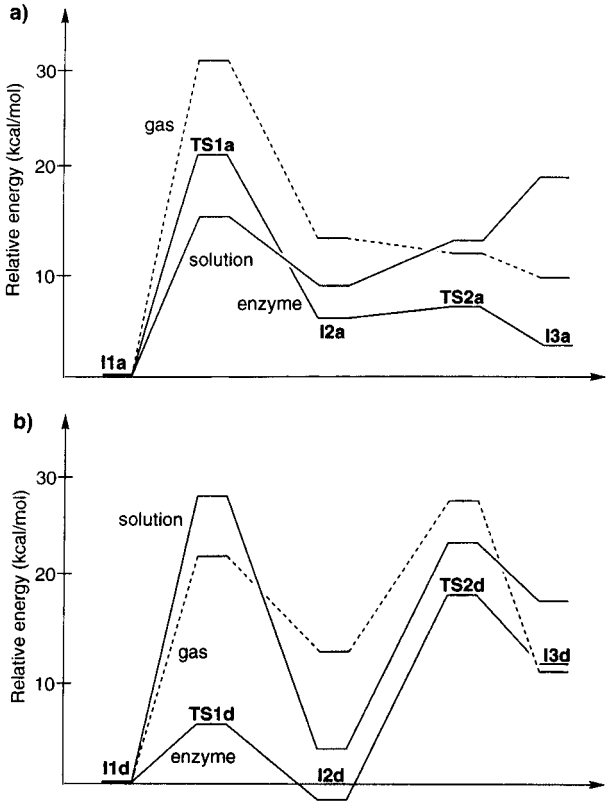


Figure 3. Calculated relative energies of the (a) acylation and (b) deacylation steps of the AGA-catalyzed reaction in the gas phase (ΔE(gas)), in aqueous solution (ΔE(aq)), and within the enzyme (ΔE(enz,av).

the water oxygen and carbonyl carbon is 1.88 Å. Tetrahedral intermediate **I2d** can go on to reaction product via the transition state **TS2d**. This step corresponds structurally to the step from **I1a** to **TS1a** of the acylation reaction.

During the gas-phase optimizations the relative orientations of the *N*-terminus, substrate model, and oxyanion hole part of the model stayed close to the structure that the catalytic amino acids have at the enzyme's active site. The only major geometrical change occurred in the position of the oxyanion hole in the structures in the deacylation part of the reaction. In these structures the carbonyl oxygen binding hydrogen bond of the oxyanion hole made by NH of Gly235 was significantly longer than found either in the X-ray structure or in the structures found in the acylation reaction.

Energetically the first transition state of the acylation step, **TS1a**, is the species that has the highest gas-phase reaction barrier, 31.3 kcal/mol, of the model enzyme reaction. In the gas phase, **I2a**, **TS2a**, and **I3a**, which follow **TS1a** in the reaction sequence, are 18−21 kcal/mol more stable than **TS1a**. The inclusion of aqueous solvation energies with the continuum solvation model changes the order of stability of these species. Solvation stabilizes **TS1a** and the tetrahedral intermediate **I2a** while destabilizing **TS2a** by 1.0 kcal/mol and **I3a** by 9.6 kcal/mol. As a consequence, the energies of the transition states **TS1a** and **TS2a** are within 2 kcal/mol and **I3a** becomes the least stable structure. In the deacylation part of the reaction the two transition states, **TS1d** and **TS2d**, have high gas-phase energy barriers of 22.2 and 27.4 kcal/mol, respectively. Since solvation destabilizes transition state **TS1d** and stabilizes tetrahedral intermediate **I2d** the two barriers are somewhat larger in aqueous solution than in the gas phase. The model calculations indicate that the deacylation part would be the rate-limiting reaction step in the aqueous phase.

Enzyme Reaction. The electrostatic interaction energies (Table 1) between the QM active site and the enzyme environment reveal how the protein environment can electrostatically

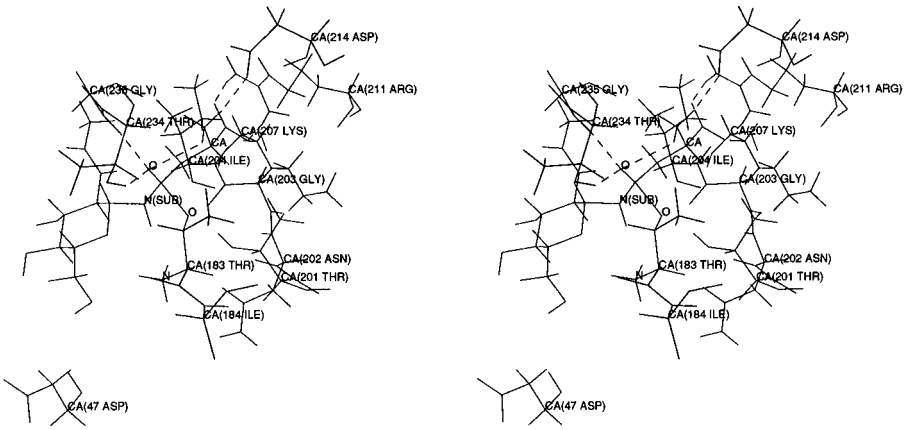


Figure 4. Stereo presentation of **TS1a** at the end point of the 130 ps MD simulation.

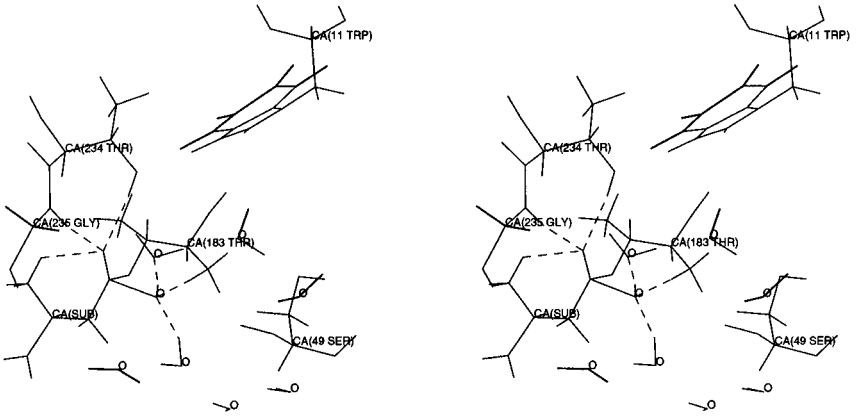


Figure 5. Stereo presentation of **TS1d** at the end point of the 130 ps MD simulation.

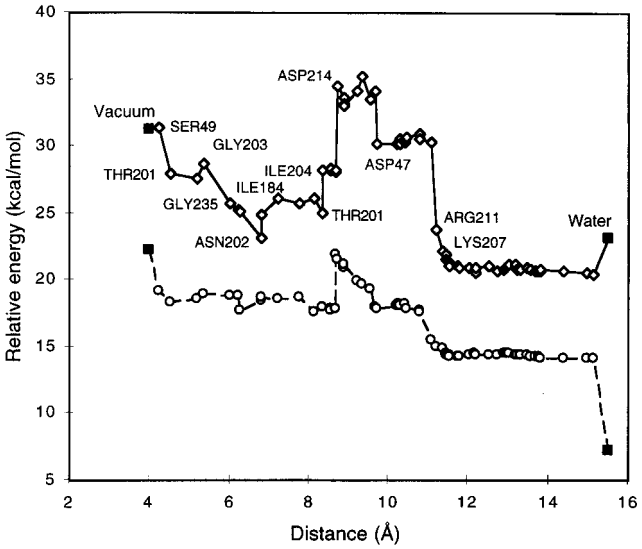


Figure 6. Contributions of the amino acid residues to the energies of the step from **I1a** to **TS1a** (upper curve) and from **I1d** to **TS1d** (lower curve). Amino acid residues are arranged in the order of the distance between their center of mass to Oγ of Thr183. Contributions of all the water molecules to the energies of the two steps are added together (points "water" in the figure).

contribute to enzyme catalyses. The electrostatic interactions stabilize the transition state **TS1a** considerably. In contrast to the effect of the aqueous solvation, which destabilizes species following **TS1a** considerably, the enzyme environment stabilizes **I2a**, **TS2a**, and **I3a**. The tetrahedral intermediate of the acylation reaction is calculated to be 13−16 kcal/mol lower in energy than **TS1a** and separated from the acyl-enzyme intermediate by a small barrier of 1−3 kcal/mol. In the deacylation reaction the tetrahedral intermediate (**I2d**) is calculated to be energetically as stable as the acyl-enzyme intermediate (**I1d**). The transition states of the deacylation step (**TS1d** and **TS2d**) are stabilized by the enzyme environment. The transition state **TS2d**, which separates **I2d** and the reaction product **I3d**, has the largest barrier (21−28 kcal/mol) in the deacylation step. The calculations predict that the transition-state barriers of the acylation and deacylation parts of the enzyme-catalyzed reaction are comparable, with the acylation step having a slightly larger barrier.

The interactions between the active site and protein environment showed that the protein environment differentially stabilizes the transition states in a way that enhances the enzyme reaction. We thus analyzed the contribution of the individual amino acid residues to the transition-state stabilization. This was done by calculating the electrostatic interaction energies between each of the amino acid residues and the active site. This kind of analysis has been earlier used in studying the mechanisms of chorismate mutase[42] and triosephosphate isomerase.[43] Results from residue analyses are presented in Figure 6. The figure shows how the electrostatic interactions between individual amino acid residues and the active-site atoms affect the energy barriers of steps from **I1a** to **TS1a** and from **I1d** to **TS1d**. The former step has the highest energy barrier of the acylation part of the reaction, and in the latter step, water attacks the acyl-enzyme intermediate. The structures of **TS1a** and **TS1d** with the most important transition state-stabilizing amino acid residues are shown in Figures 4 and 5.

It can be seen from Figure 6 that most of the residues stabilize transition states **TS1a** and **TS1d** relative to ground states **I1a** and **I1d** or have only a very small effect on the energetics of the two reactions. This is especially notable for the step from **I1d** to **TS1d**. The charged groups have, as expected, the largest interactions with the active-site region. Asp214, which binds the α-amino group of the substrate and is close to the nucleophilic oxygen and the oxyanion of the substrate, has an unfavorable interaction of 7 kcal/mol with the active-site atoms in **TS1** as compared to **I1a** and 3 kcal/mol in **TS1d** as compared to **I1d**. In contrast to Asp 214, Asp47, which is located at the mouth of the active-site funnel and is close to the α-amino group of the *N*-terminal threonine, makes a favorable contribution to transition-state stabilization. Arg211, which binds the carboxylate end of the substrate, has the single most favorable differential interaction with the transition state. When the residue contributions of the steps from **I1a** to **TS1a** and from **I1d** to **TS1d** are compared, two main differences can be seen. First, the magnitudes of the individual residue contributions are clearly smaller in the latter step. However, when the total contribution of the protein residues are compared, they are of comparable magnitude in the two cases. In the step from **I1a** to **TS1a**, the protein atoms stabilize the transition state by 9 kcal/mol, and in step from **I1d** to **TS1d**, the stabilization is 7 kcal/mol. Second, while water molecules make a slightly unfavorable contribution to the step from **I1a** to **TS1a**, they facilitate the step from **I1d** to **TS1d** by 7 kcal/mol. The larger water contribution in the step belonging to the deacylation part of the reaction is due to the fact that in the acylation step the glycan moiety prevents water from entering the active site, whereas in the deacylation part of the reaction the active-site funnel is filled with water (Figure 5). We can see from Figure 5 that several water molecules make hydrogen bonds with the α-amino group of Thr183 and the attacking water, which delivers proton to the α-amino group. Analyses of the contribution of individual water molecules on the transition-state energies of the two steps revealed that 4−6 water molecules closest to the active-site atoms are responsible for the major part of the more favorable water contribution in the step from **I1d** to **TS1d** as compared to the step from **I1a** to **TS1a**. A similar observation was made by Jensen *et al.*[44] in computer simulation work on ferredoxin, an iron−sulfur protein. In that case specific water molecules were found to make an important contribution to the redox potential of ferredoxin. As discussed by Jensen *et al.*, the explicit modeling of water molecules is clearly important in the systems where waters play a specific role in the structures and energetics; the deacylation reaction studied here also appears to be such a system.

Discussion

The results from computer simulations presented here support the earlier suggestion[10,11,21] that a serine protease-like catalytic mechanism is feasible for AGA-catalyzed hydrolyses of amide bonds. The chemical steps of the mechanism presented here are almost identical to those presented in earlier theoretical studies of serine protease-catalyzed reactions.[13−15,45] The calculated energetics of the catalytic mechanism show that the transition-state barriers are similar for the acylation and deacylation parts of the overall catalytic mechanism. Energetically, the highest point in the acylation step is the attack of nucleophilic Oγ to the carbonyl carbon of the substrate. In the

(42) Lyne, P. D.; Mulholland,A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345−11350.
(43) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826−5832.
(44) Jensen, G. M.; Warshel, A.; Stephens, P. J.; *Biochemistry* **1994**, *33*, 10911−10924.
(45) Weiner, S. J.; Seibel, G. L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 649−653.

deacylation step the corresponding transformation, the formation of the reaction product from the tetrahedral intermediate, has the highest barrier. The calculated transition-state energies were for acylation 12−22 kcal/mol and for deacylation 21−28 kcal/mol. The acylation value is in the range of the experimental transition-state free energy ($\Delta G^{\ddagger}$) of 16.4 kcal/mol for the AGA-catalyzed hydrolyses of *β*-*N*-acetylglucosamine-L-asparagine (it has been estimated from the measured k_{cat} of 5.5 s^{-1} [7] using the transition-state theory approximation with a transmission factor of unity), and the calculated deacylation barrier is reasonable but a bit too large. The calculations also predicted that, in the acylation part of the reaction, the tetrahedral intermediate would collapse to an acyl-enzyme intermediate through a small barrier of 1−3 kcal/mol. In the deacylation part, the tetrahedral intermediate is almost as stable as the acyl-enzyme intermediate and it is separated from the acyl-enzyme intermediate by a barrier of 7 kcal/mol.

Although our calculations only include the environmental effects of the enzyme in a semiquantitative way, we feel the compromise is reasonable because it is still too computationally difficult to simulate a complex, multistep reaction in an enzyme with fully coupled *ab initio* QM/MD or MC/FEP, as has been done by Jorgensen[46,47] (*ab initio* MC/FEP) and by Gao *et al.*[48,49] (semiempirical MC/FEP) for organic reactions and by Warshel *et al.*[13,15,50,51] for more limited reaction profiles. It is clear that our calculations show enzyme stabilization by about the appropriate magnitude so that our activation energy barriers are in the right range of experiment. What is particularly intriguing about our results is that a simple continuum model gives an excellent representation of "enzyme-like" stabilization of the acylation transition state, but actually leads to a higher barrier than in the gas phase for deacylation, whereas the explicit enzyme model indicates significant barrier lowering for both acylation and deacylation. This can be understood through the essential role of explicit water molecules in the latter case. Although the calculation with enzyme environment suggest that **I2d**, the tetrahedral intermediate for deacylation, is more stable than the acyl-enzyme with H_2O, **I1d**, the accuracy of our calculations are not definitive enough to suggest it can be observed. However, this issue is worth exploring at a higher level of theory.

The present calculations allow us to conclude that the single *N*-terminal amino acid of AGA, and the Ntn amidohydrolases in general, can function as a nucleophile and a base in a way similar to that of serine and histidine residues in the active sites of serine proteases. The fact that the neutral α-amino group of an *N*-terminal amino acid functions as a base in a way similar to that of histidine cannot be regarded as a surprise due to their similar pK_a values: the α-amino group of an *N*-terminal amino acid has a pK_a value of 6.8−7.9, and histidine's first pK_a is 6.2.[52] Also, as the simulations of the reaction within the enzyme showed, the *N*-terminal threonine of AGA can function effectively as a single catalytic amino acid from the structural point of view as well, with both acid and base functionalities.

The comparison of the energetics of the model enzyme reaction in the gas phase and with solvation energies included with the energetics of the reaction within the enzyme shows that the enzyme considerably enhances the hydrolyses reaction. This allowed us to address the question of how AGA achieves its catalytic properties. The reaction rate-accelerating properties of serine proteases have been largely attributed to the oxyanion-stabilizing oxyanion hole and the catalytic triad which is formed by amino acids Ser-His-Asp(Glu). Experimental mutagenesis studies and earlier computational works have shown that one hydrogen bond from the oxyanion hole stabilizes the transition state by 3−5 kcal/mol.[53−56] Since we included the oxyanion hole in QM part of the enzyme system, it is difficult to separate its contribution to the enzyme catalysis. In light of structural data on AGA and the vast amount of data on serine proteases, there is no doubt that the oxyanion hole has an important role in enhancing the catalytic activity of AGA as well. In addition, in all the MD simulations the oxyanion hole, which was not constrained intermolecularly to bind the carbonyl oxygen, maintained two hydrogen bonds with the oxygen. In addition to two hydrogen bonds from the OH of Thr234 and the main chain NH of Gly235, the carbonyl oxygen of the scissile amide bond makes an additional intramolecular hydrogen bond with the protonated α-amino group of the substrate. These three hydrogen bonds are arranged tetrahedrally around the oxygen (Figures 4 and 5). A similar kind of oxyanion hole with three oxyanion-stabilizing neutral hydrogen bonds was recently found in the structure of bacterial esterase.[57] It was suggested that since the esterase has a Ser-His dyad in its active site and lacks the third member (Asp of Glu) of the catalytic triad of a typical serine protease, the well-developed oxyanion hole has evolved to compensate for the lack of carboxylic acid in the triad. It is possible that in the case of AGA the oxyanion hole with three components serves a similar function. It has also been suggested that the α-amino group of AGA[19] and the NH_3^+ groups of lysines located in the vicinity of the nucleophilic atom can enhance the catalysis by lowering the pK_a of the nucleophile.[11,21] This question was not addressed in our calculations, but it seems a reasonable assumption. Furthermore, the arrangement of the protonated α-amino group of the substrate and Asp214 is similar to the Lys-Asp(Glu) pairs found in the active site of proteasome[9] and several other enzymes with corresponding enzymatic activities.[9,58−60] So, it seems that together the *N*-terminal base and the Lys-Asp(Glu) pair form a catalytic motif which activates the catalytic center in a manner similar to that of the catalytic triad of serine proteases. These comparisons indicate that in the case of AGA the substrate α-amino group plays an important role in catalysis and can be related to several other enzymes that have protonated amine groups of lysines as a structural counterpart to the α-amino group at their active sites.

The catalytic properties of AGA are certainly largely determined by the exact arrangement of the catalytic *N*-terminal Thr183, the oxyanion hole, and the amino acid residues which directly participate in binding of the substrate. The favorable electrostatic interaction energies (Table 1) between the active-site model and the protein environment and water located further

(46) Joneshertzog, D. K.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1995**, *117*, 9077−9078.
(47) Lim, D. C.; Hrovat, D. A.; Borden, W. T.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1994**, *116*, 3494−3499
(48) Sehgal, A.; Shao, L.; Gao, J. L. *J. Am. Chem. Soc.* **1995**, *117*, 11337−11340.
(49) Gao, J. L. *J. Am. Chem. Soc.* **1995**, *117*, 8600−8607.
(50) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.
(51) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(52) Tanford, C. *Adv. Protein Chem.* **1962**, *17*, 69−165.
(53) Bryan, P.; Pantoliano, M. W.; Quill, S. G.; Hsiao, H.-Y.; Poulos, T. *Proc. Natl. Acad. Sci. U.S.A.* **1986**,*83*,3743−3745.
(54) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature* **1987**, *328*, 551−554.
(55) Hwang, J.-K.; Warshel, A. *Biochemistry* **1986**, *26*, 2669−2673.
(56) Wells, J. A.; Cunningham, B. C.; Craycar, T. P.; Estell. D. A. *Philos. Trans. R. Soc. London* **1986**, *A317*, 415−423.
(57) Wei, Y.; Schottel, J. L.; Derewenda, U.; Swenson, L.; Patkar, S.; Derewenda, Z. S. *Struct. Biol.* **1995**, *2*, 218−223.
(58) Swain, A. L.; Jaskólski, M.; Housset, D.; Rao, J. K. M.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1474−1478.
(59) Miller, M.; Rao, J. K. M.; Wlodawer, A.; Gribskov, M. R. *FEBS Lett.* **1993**, *328*, 275−279.
(60) Ramão, M. J.; Turk, D.; Gomis-Rüth, F.-X.; Huber, R.; Schumacher, G.; Möllering, H.; Rüssmann, L. *J. Mol. Biol.* **1992**, *266*, 1111−1130.

away from the active site also make a large reaction rate-enhancing contribution. The analyses of the residue contributions on the energy barriers of the steps from **I1a** to **TS1a** and from **I1d** to **TS1d** (Figure 6) show that stabilization of the two transition states is a result of a large number of small favorable interactions. However, a large part of the electrostatic interactions can be attributed to charged residues Asp47, Lys207, and Arg211. These residues are located in the protein in such a way that they favorably interact with the active site. Arg211, which binds the α-carboxylate group of the substrate, and Lys207, which is located near Arg211, are closer to the nucleophilic oxygen and the oxyanion than the α-amino group of the *N*-terminus. This evidently favors the ionized transition states and intermediates relative to nonionized states. In a similar way Asp47, which is located in the mouth of the active-site funnel, interacts favorably with the α-amino group. Charged Asp214 makes, however, an unfavorable contribution to the energies of the transition states **TS1a** and **TS1d**. Asp214 was even found to be the only residue that has clearly unfavorable contribution to the stabilization of **TS1d**. As discussed above, Asp214 may serve an important functional role by binding the substrate and especially its α-amino group at a proper location at the active site. In addition, residue analyses revealed that, once the glycan moiety of the substrate has been cleaved in the acylation step and left the active-site funnel, water can replace it and significantly contribute to the stabilization of **TS1d**. Thus, in the deacylation step AGA achieves considerable catalytic power by solvating part of the active site.

Conclusions

In this paper we have presented results from *ab initio* quantum mechanical (QM) and molecular dynamics (MD) calculations of the aspartylglucosaminidase (AGA)-catalyzed hydrolyses of an amide bond. The catalytic mechanism of AGA was found to be similar to that of the serine proteases. Thus, both from structural and energetic points of view the *N*-terminal threonine of AGA can function as a nucleophile and a base in a way similar to that of the serine and histidine residues of the serine proteases. Residue analyses, in which the electrostatic interaction energies between the amino acid residues and the active-site atoms were calculated, were done in order to reveal the contribution of each of the residues to the transition-state stabilization. These analyses showed that charged Asp47, Lys207, and Arg211 are the most important transition state-stabilizing amino acid residues. These amino acids are located in the protein in such a way that they electrostatically stabilize the ionized states of the catalytic reaction. Furthermore, in the deacylation step water molecules can enter the active site and fill the space which in the acylation part of the reaction was occupied by the glycan moiety of the substrate. The water molecules were found to make several hydrogen bonds with the catalytic amino acid and in this way contribute significantly to the stabilization of the ionized species. Further, it was found that 4−6 active-site water molecules made the major part of the favorable water contribution in the deacylation step.

Acknowledgment. M.P. gratefully acknowledges the support from the Academy of Finland, and P.A.K. is grateful to the NIH for support (Grant GM-20972). We thank the UCSF graphics laboratory (T. E. Ferrin, P.I., RR-1081) for computer support.

JA9628967