Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site

Stefano Piana,[‡,§] Daniel Sebastiani,[†,∥] Paolo Carloni,*,[‡] and Michele Parrinello[†,⊥]

Contribution from the Scuola Internazionale Superiore di Studi Avanzati and Istituto Nazionale di Fisica della Materia, Via Beirut 2-4, 34014 Trieste, Italy, and the Max-Planck Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

Received August 23, 2000. Revised Manuscript Received May 11, 2001

Abstract: A recent ^{13}C NMR experiment (Smith et al. *Nature Struct. Biol.* **1996**, *3*, 946−950) on the Asp 25-Asp25′ dyad in pepstatin A/HIV-1 protease measured two separate resonance lines, which were interpreted as being a singly protonated dyad. We address this issue by performing ab initio molecular dynamics calculations on models for this site accompanied by calculations of ^{13}C NMR chemical shifts and isotopic shifts. We find that already on the picosecond time-scale the model proposed by Smith et al. is not stable and evolves toward a different monoprotonated form whose NMR pattern differs from the experimental one. We suggest, instead, a different protonation state in which both aspartic groups are protonated. Despite the symmetric protonation state, the calculated ^{13}C NMR properties are in good agreement with the experiment. We rationalize this result using a simple valence bond model, which explains the chemical inequality of the two C sites. The model calculations, together with our calculations on the complex, allow also the rationalization of ^{13}C NMR properties on other HIV-1 PR/inhibitor complexes. Both putative binding of the substrate to the free enzyme, which has the dyad singly protonated (Piana, S.; Carloni, P. *Proteins*: *Struct*., *Funct*., *Genet*. **2000**, *39*, 26−36), and pepstatin A binding to the diprotonated form are consistent with the inverse solvent isotope effect on the onset of inhibition of pepsin by pepstatin and the kinetic iso-mechanism proposed for aspartic proteases (Cho, T.-K.; Rebholz, K.; Northrop, D.B. *Biochemistry* **1994**, 33, 9637−9642).

Introduction

The human immunodeficiency virus protease (HIV-1 PR) is a homodimeric enzyme essential for the metabolism of the virus.[1,2] HIV-1 PR cleaves polypeptide segments at specific locations. These segments subsequently fold to form structural proteins and enzymes, including the protease itself.[2−5]

Since the discovery that binding to pepstatins (Chart 1) leads to the production of immature, noninfectious viral particles,[3,6−8]

HIV-1 protease has become one of the major targets for therapy against the AIDS epidemic.[9−15] Enzyme/inhibitor interactions depend dramatically on the ionization state of the cleavage site, which contains an aspartyl dyad (Asp25/Asp25′).[9,16−20] Thus, the determination of the protonation state of the aspartyl groups of inhibitor/enzyme complexes may help computer-aided design of related inhibitors. Furthermore, it may provide fundamental insights into the enzymatic mechanism.[21,22]

A frequently used method for determining the exact charge state of carboxylic groups in proteins is ^{13}C NMR spectroscopy. Chemical shifts on ^{13}C-enriched HIV-1 PR have been measured

* Corresponding author. Phone: +39-040-3787407 Fax: +39-040-3787-528. E-mail: carloni@sissa.it.

‡ Scuola Internazionale Superiore di Studi Avanzati e Istituto Nazionale di Fisica della Materia.

† Max-Planck Institut für Festkörperforschung.

§ Present address: Inorganic Chemistry Department, ETH Zurich, Universitatstrasse 6, 8092 Zurich, Switzerland.

∥ Max-Plank Institute for Polymer Research, Hackermann 10, D-55128 Mainz, Germany.

⊥ Present address: CSCS-Centro Svizzero di Calcolo Scientifico-ETH Zurich, Galleria 2, Via Cantonale, CH-6928 Manno, Switzerland.

(1) Debouck, C.; Gorniack, J. G.; Strickler, J. E.; Meek, T. D.; Metcalf, B. W.; Rosenberg, M. *Proc. Natl. Acad. Sci.* **1987**, *84*, 8903−8906.

(2) Louis, J. M.; Weber, I. T.; Totzser, J.; Clore, G. N. Gronenborn, A. N. *Adv. Pharm.* **2000**, *49*, 111−146.

(3) Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6612−6616.

(4) Fitzgerald, P. M. D.; Springer, J. P. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 299−320.

(5) Debouck, C. *AIDS Res. Hum. Retroviruses* **1992**, *8*, 153−164.

(6) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686−4690.

(7) McQuade, T. J.; Tomasselli, A. G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T. K.; Heinrikson, R. L.; Tarpley, W. G. *Science* **1990**, *247*, 454−456.

(8) Kaplan, A. H.; Zack, J. A.; Knigge, M.; Paul, D. A.; Kempf, D. J.; Norbeck, D. W.; Swanstrom, M. *J. Virol.* **1993**, *67*, 4050−4055.

(9) Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.

(10) Gulnik, S.; Erickson, J. W.; Xie, D. *Vitam. Horm.* **2000**, *58*, 213−256.

(11) Wlodawer, A.; Erickson, J. W. *Annu. Rev. Biochem.* **1993**, *62*, 543−585.

(12) Vacca, J. P. In *Methods in Enzymology*; Kuo, L. C., Shafer, J. A., Eds.; Academic Press: San Diego, CA, 1994; pp 311−334.

(13) Vacca, J. P.; Condra, J. H. *Drug Des. Discovery* **1997**, *2*, 261−272.

(14) Flexner, C. *N. Engl. J. Med.* **1998**, *338*, 1281−1292.

(15) Erickson, J. W.; Eissenstat, M. E. In *Proteases of Infectious Agents*; Dunn, B., Ed.; Academic Press: San Diego, CA, 1999; pp 1−60.

(16) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Erickson, J. W. *J. Am. Chem. Soc.* **1998**, *120*, 8856−8863.

(17) Harte, W. E.; Beveridge, D. L. *J. Am. Chem. Soc.* **1993**, *115*, 3883−3886.

(18) Chen, X.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 42−48.
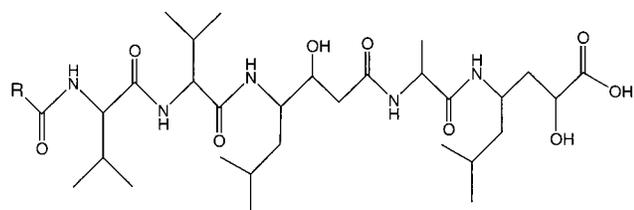
(19) Trylska, J.; Antosiewicz, J.; Geller, M.; Hodge, C. N.; Klabe, R. M.; Head, M. S.; Gilson, M. K. *Protein Sci.* **1999**, *8*, 180−195.

(20) Tropsha, A.; Hermans, J. *Protein Eng.* **1992**, *5*, 29−33.

(21) Rebholz, K.; Northrop, D. B. *Biochem. Biophys. Res. Comm.* **1991**, *176*, 65−69.

(22) Cho, T.-K.; Rebholz, K.; Northrop, D. B. *Biochemistry* **1994**, *33*, 9637−9642.

Chart 1

Pepstatin A, R = Methyl
Acetyl Pepstatin, R = Iso-propyl

in the free enzyme[23] and in inhibitor/enzyme complexes;[23−31] however, the interpretation of the signals has proved to be challenging.[27,29]

An example of this difficulty is given by the ^{13}C NMR spectrum of the adduct with the transition-state analogue pepstatin A (Chart 1). The complex exhibits two distinct signals at 172.4 and 178.8 ppm that remain unchanged in the pH range between 2.5 and 6.5,[23] thus indicating that the protonation state of the Asp dyad does not change in this pH range. Of the two peaks, the low-field one undergoes an isotopic shift (defined as the changes in chemical shifts upon replacement of hydrogen with deuteration), but the other remains unchanged upon deuteration.

On the basis of these results and the X-ray structure of the complex with the analogue acetyl pepstatin[32] (Figure 1 and Chart 1), the authors proposed the H-bond pattern **A**1 shown in Figure 2. In this structure, both aspartyl groups interact with the inhibitor's hydroxyl hydrogen, and Asp25′ forms an additional H-bond to the hydroxyl oxygen. Thus, the two carboxylates appear to be chemically nonequivalent, which might explain the NMR signals. Furthermore, because one of the aspartyl groups is ionized, the pattern can also explain the different isotopic behavior.[33−35] However, also in the eyes of the authors, this assignment is surprising. In fact, the chemical shifts are reversed with respect to those of the carboxylic acids in aqueous solution[36] and in the adduct with the chemically similar inhibitor KNI-272.[27]
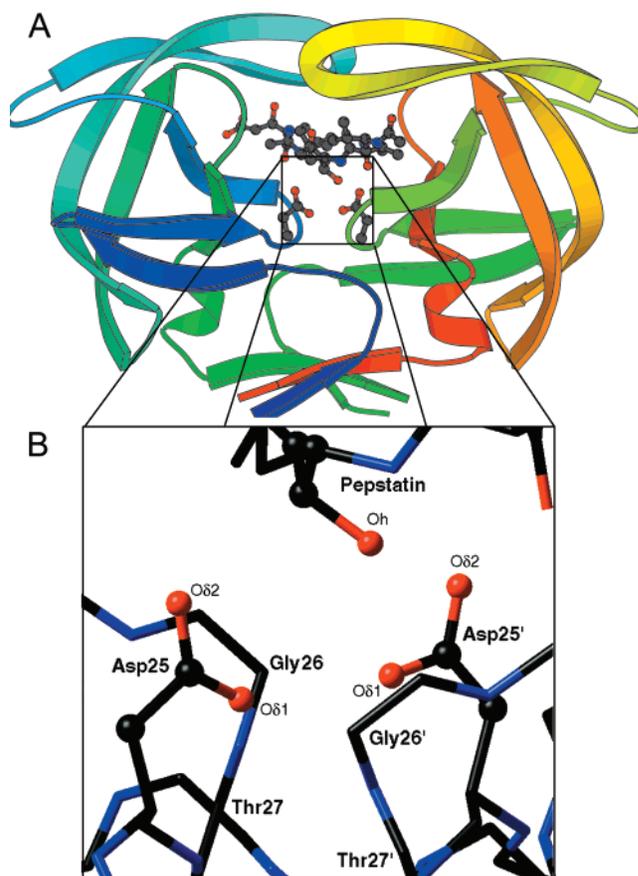
Figure 1. Acetyl pepstatin/HIV-1 PR complex:[32] (a) the entire adduct and (b) a close-up of the cleavage site region (b). The central hydroxyl group (Oh) of this asymmetric inhibitor interacts with the catalytic Asp dyad.

Clearly, an ab initio calculation of the NMR chemical shift[37−56] could be of great value in explaining these results. Here, we shall use some recent developments that allow NMR

(23) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. *Nature Struct. Biol.* **1996**, *3*, 946−950.

(24) Ishima, R.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Torchia, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 10534−10542.

(25) Freedberg, D. I.; Wang, Y. X.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Kiso, Y.; Torchia, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 7916−7923.

(26) Katoh, E.; Yamazaki, T.; Kiso, Y.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Torchia, D. A. *J. Am. Chem. Soc.* **1999**, *121*, 2607−2608.

(27) Wang, Y. X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. *Biochemistry* **1996**, *35*, 9945−9950.

(28) Wang, Y. X.; Freedberg, D. I.; Grzesiek, S.; Torchia, D. A.; Wingfield, P. T.; Kaufman, J. D.; Stahl, S. J.; Chang, C. H.; Hodge, C. N. *Biochemistry* **1996**, *35*, 12694−12704.

(29) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C. H.; Weber, P. C. *J. Am. Chem. Soc.* **1994**, *116, 6*, 10791−10792.

(30) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Domaille, P. J.; Campbell-Burk, S. *Eur. J. Biochem.* **1994**, *219*, 707−712.

(31) Grzesiek, S.; Bax, A.; Nicholson, L. K.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hodge, C. N.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C. H. *J. Am. Chem. Soc.* **1994**, *116*, 1581−1582.

(32) Fitzgerald, P. M. D.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C.-T.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. *J. Biol. Chem.* **1990**, *265*, 14209−14219.

(33) Northrop, D. B. *Methods* **2001**, 117−124.

(34) Led, J. J.; Petersen, S. B. *J. Magn. Res.* **1979**, *33*, 603−617.

(35) Ladner, H. K.; Led, J. J.; Grant, D. M. *J. Magn. Res.* **1975**, *20*, 530−534.

(36) Hagen, R.; Roberts, J. D. *J. Am. Chem. Soc.* **1969**, *91*, 4504−4506.

(37) Abildgaard, J.; Bolvig, S.; Hansen, P. E. *J. Am. Chem. Soc.* **1998**, *120*, 9063−9069.

(38) Rauhut, G.; Puyear, S.; Wolinski, K.; Pulay, P. *J. Phys. Chem.* **1996**, *100*, 6310−6316.

(39) Alder, R. W.; Blake, M. E.; Oliva, J. M. *J. Phys. Chem. A* **1999**, *103*, 11200−11211.

(40) Karger, N.; da Costa, A. M. A.; Ribeiro-Claro, P. J. A. *J. Phys. Chem. A* **1999**, *103*, 8672−8677.

(41) Asakawa, N.; Sato, D.; Sakurai, M.; Inoue, Y. *J. Phys. Chem. A* **2000**, *104*, 2716−2723.

(42) Wilkens, S. J.; Xia, B.; Volkman, B. F.; Weinhold, F.; Markley, J. L.; Westler, W. M. *J. Phys. Chem. B* **1998**, *102*, 8300−8305.

(43) Hricovini, M.; Malkina, O. L.; Bizik, F.; Nagy, L. T.; Malkin, V. G. *J. Phys. Chem. A* **1997**, *101*, 9756−9762.

(44) Cloran, F.; Carmichael, I.; Serianni, A. S. *J. Phys. Chem. A* **1999**, *103*, 3783−3795.

(45) Farcasiu, D.; Hancu, D.; Haw, J. F. *J. Phys. Chem. A* **1998**, *102*, 2493−2499.

(46) Lampert, H.; Mikenda, W.; Karpfen, A.; Kählig, H. *J. Phys. Chem. A* **1997**, *101*, 9610−9617.

(47) Prakash, G. K. S.; Rasul, G.; Liang, G.; Olah, G. A. *J. Phys. Chem.* **1996**, *100*, 15805−15809.

(48) Farcasiu, D.; Hancu, D. *J. Phys. Chem. A* **1999**, *103*, 754−761.

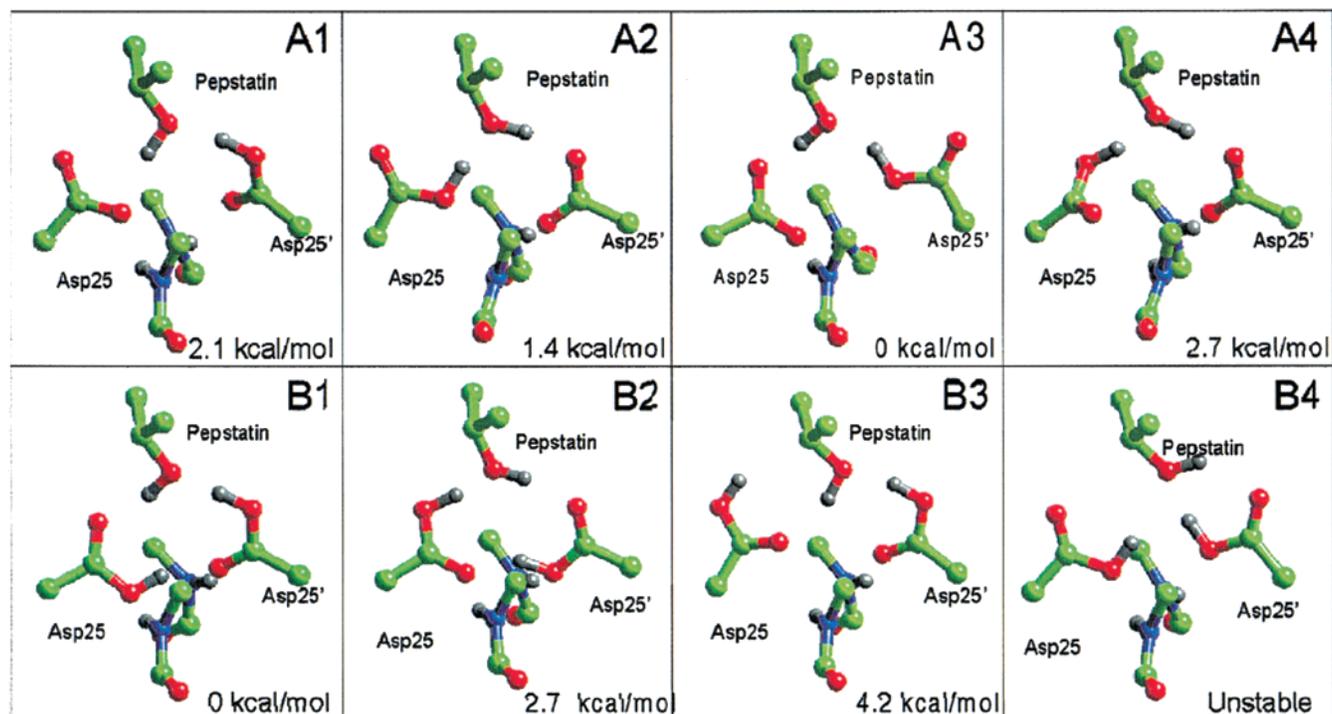(49) Kubicki, J. D.; Sykes, D.; Apitz, S. E. *J. Phys. Chem. A* **1999**, *103*, 903−915.

(50) Sykes, D.; Kubicki, J. D.; Farrar, T. C. *J. Phys. Chem. A* **1997**, *101*, 2715−2722.

(51) Rodriguez-Fortea, A.; Alemany, P.; Ziegler, T. *J. Phys. Chem.* **1999**, *103*, 8288−8294.

(52) Valerio, G.; Goursot, A. *J. Phys. Chem. B* **1999**, *103*, 51−58.

(53) Gregor, T.; Mauri, F.; Car, R. *J. Chem. Phys.* **1999**, *111*, 1815−1822.

(54) Mauri, F.; Pfrommer, B. G.; Louie, S. G. *Phys. Rev. Lett.* **1996**, *77*, 5300−5303.

Figure 2. Proton locations in the negatively charged and neutral forms of the Asp dyad in HIV-1 PR: chemical structures and structural models of the monoprotonated (a) and diprotonated (b) forms. The calculated energies relative to the most stable protomers (**A**3 and **B**1 for the monoprotonated and diprotonated forms, respectively) are also reported.

chemical shifts to be calculated within the framework of Car–Parrinello molecular dynamics (CPMD).[57] This combination is particularly powerful, because it allows us to investigate the effects of temperature on the NMR signals, which are expected to be of crucial relevance in biological systems.[58]

Here we consider relatively large structural models of the enzyme's cleavage site (Figure 2) in which effects of the environment are partially taken into account. Starting from the complex proposed in **A**1, we find that the H-bond pattern is not stable and evolves to protomer **A**2. For this pattern, the calculated ^{13}C chemical shifts are almost equivalent, in contrast to the experimental results. The other possible patterns of the monoprotonated form exhibit NMR properties that disagree with experiment or are highly unstable.

We suggest, therefore, that the Asp dyad is doubly protonated.[17,19] We find that the most stable, diprotonated protomer is **B**1 (Figure 2). The calculated chemical shifts of this protomer are different, in qualitative agreement with experimental data. Furthermore, the calculated isotopic shifts are in good agreement with the experimental data.[23] Such a doubly protonated pattern has already been suggested for the DMP323/HIV-1 PR complex[29] on the basis of NMR data. At first sight, this latter result seems surprising, because the two Asp groups appear to be chemically equivalent. However, we argue that the two aspartyl groups are differently H-bonded to the environment. This leads to a different destabilization of the π-electrons of the carboxylic group resonance structures and, therefore, to different chemical shifts and different isotopic effects. To further substantiate this

point of view, we study the phenomena of H-bond-induced π-resonance destabilization in a series of formic acid/water complexes.

Methods

Structural Models. (a) The Enzyme. Although the structure of the pepstatin A/HIV-1 PR complex has not been solved yet, the structure of the complex with acetyl pepstatin is known at 2.0 Å resolution[32] (5HVP entry in the PDB database[59]). Acetyl pepstatin is chemically and structurally extremely similar to pepstatin A (Chart 1), and the portion binding to the active site is identical to that of pepstatin A. Therefore, the initial configuration of pepstatin A/HIV-1 PR complex was built upon the basis of the known X-ray structure.

All of the possible protomers of both diprotonated and monoprotonated forms were considered (Figure 2). In **A**1–**A**4, one of the two Asp side chains was protonated (overall net charge, −1); in **B**1–**B**4, both Asp groups were protonated (overall charge, 0). It should be noted that **A**1 represents the protonation pattern of HIV-1 PR/pepstatin A proposed on the basis of ^{13}C NMR measurements.[23]

(b) Formic Acid/Water Complexes. Formic acid (**0**) and its conjugated base (**0**-) were constructed assuming standard bond lengths and bond angles (Figure 3). Water/formic acid complexes **I**–**IV** (Figure 3) were built adding an increasing number of water molecules to **0**. (Complexes are named according to the number of water models included in the calculations.) **IIa** (**IIc**) was the same as **III**, except that WATA (WATB) was removed; **IIb** was the same as **IV**, except that WATB and WATC were removed.

Quantum-Mechanical Calculations. The quantum problem was solved within the framework of density functional theory. Exchange and correlation functionals were those of Becke[60] and

(55) Mauri, F.; Pfrommer, B. G.; Louie, S. G. *Phys. Rev. Lett.* **1997**, *79*, 2340–2343.

(56) Pfrommer, B. G.; Mauri, F.; Louie, S. G. *J. Am. Chem. Soc.* **2000**, *122*, 123–129.

(57) Putrino, A.; Sebastiani, D.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **2000**, *113*, 7102–7109.

(58) Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins: a Perspective of Dynamics, Structure, Thermodynamics*; J. Wiley and Sons: New York, 1988.

(59) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G. L.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
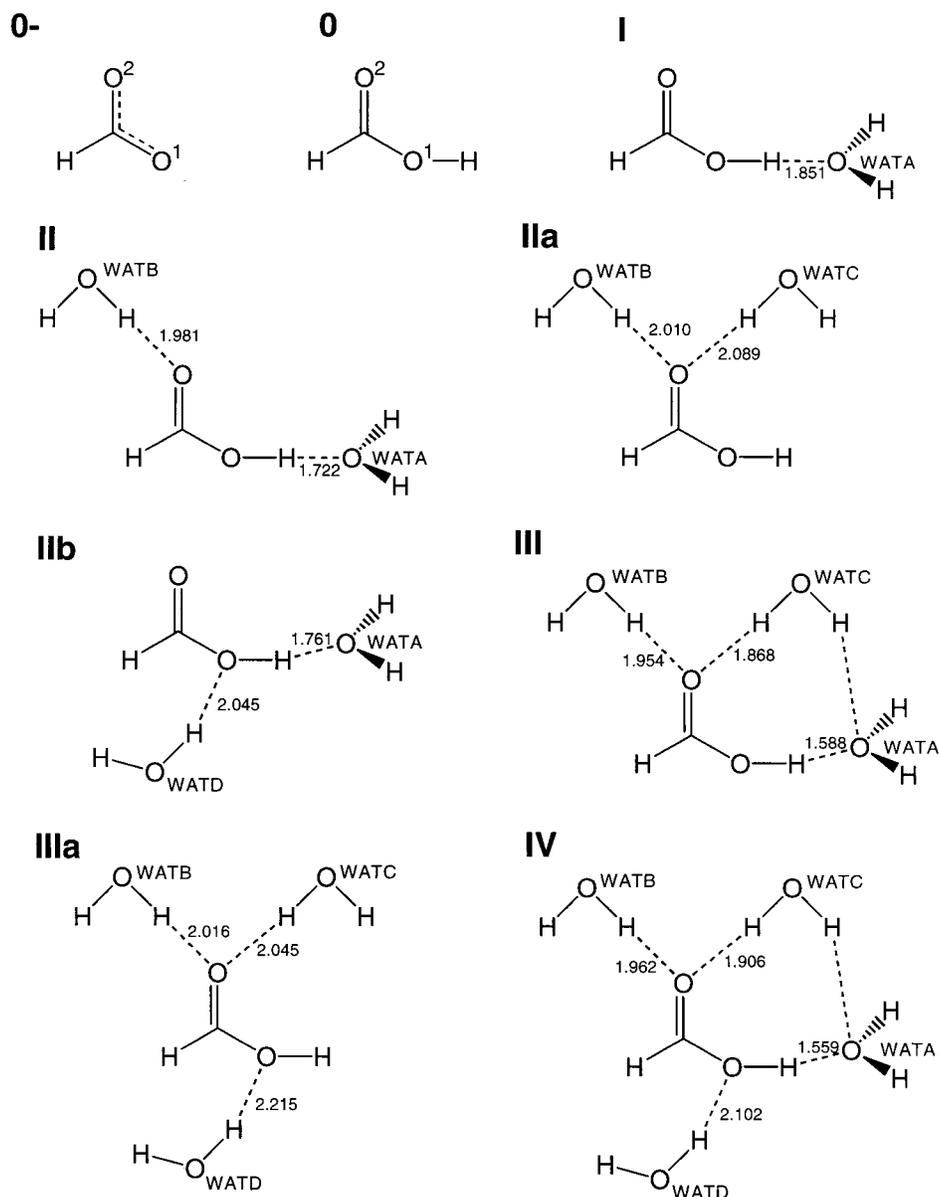
Figure 3. Hydrated complexes of formic acid considered in the quantum-mechanical calculations. Carboxyl oxygen atoms and water molecules are labeled as **0-** and **0**. The hydrogen bonding distances are also indicated.

Lee, Yang, and Parr[61] (BLYP), respectively. The Kohn−Sham orbitals were expanded in plane waves up to 70 Ry. Martins Troullier[62] pseudopotentials were used to describe the interactions between the ionic cores and the valence electrons. The systems were treated as isolated, as in ref 63.

Geometry optimizations were carried out for complexes **I−IV**, **A**1−**A**4, and **B**1−**B**4 using the direct inversion in the iterative subspace method[64] with a convergence criterion of 5×10^{-4} for the largest component of the atomic forces. **A**3 and **B**1 turned out to be the most stable complexes (Figure 2). In contrast, protomers **A**4, **B**2, **B**3, and **B**4 were much higher in energy, and some of them (**A**4, **B**3, and **B**4) exhibited highly distorted conformations (Figure 2).

DFT-based molecular dynamics simulations were performed according to the Car−Parrinello scheme.[65] Simulations were

performed for the following complexes: (i) **A**1, which is the pattern proposed by Smith et al;[23] (ii) **A**2, which is obtained spontaneously after ≅ 0.5 ps of dynamics of **A**1; (iii) **A**3, which is the most stable complex for the monoprotonated form; and (iv) **B**1, which is the most stable complex for the diprotonated form. A time-step of 0.0096 fs and a fictitious electron mass of 400 au were used. The ion temperature was controlled by coupling the systems to a Nosé thermostat of 500 cm^{-1} frequency.[67] The run lengths were 2 ps.

In an established procedure,[66] the positions of terminal atoms were kept fixed during the MD simulations in order to mimic the relatively rigid protein frame.[32] This modeling has already been shown to be capable of reproducing the structural properties of the active site of the free enzyme.[66b] All calculations were carried out with the CPMD code.[68]

(60) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(61) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(62) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1943−2006.
(63) Barnett, R. N.; Landman, U. *Phys. Rev. B* **1993**, *48*, 2081−2097.
(64) Hutter, J.; Luthi, H. P.; Parrinello, M. *Comput. Mat. Sci.* **1994**, *2*, 244−248.
(65) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(66) (a) Piana, S.; Carloni, P. *Proteins*: *Struct. Funct. Genet.* **2000**, *39*, 26−36. (b) De Santis, L.; Carloni, P. *Proteins*: *Struct. Funct. Genet.* **1999**, *37*, 611−618.
(67) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.

Calculated Properties. The integrated electron density around the carbon atom was calculated according to the Hirshfeld method.[69] Mayer bond orders[70] were calculated by projecting the wave function into an atomic basis set. Centers of the maximally localized Boys orbitals[71,72] were calculated as in ref 73.

Calculation of ^{13}C NMR Shifts. The chemical shifts were obtained in the framework of density functional perturbation theory (DFPT),[53] which has recently been implemented in the CPMD code.[53,57,68] The so-called "$d(r) = r$" variant of the continuous set of gauge transformations (CSGT) method was used for the treatment of the gauge origin problem.[74]

In the CPMD code, use is made of pseudopotentials in order to describe the interactions between valence and core electrons. This implies that the core-electron contribution to the nuclear shielding is not included; however, it has been shown that the core electron contribution is approximately constant.[53] For the purpose of the present work, in which we focus on chemical shift differences, such a constant term can be ignored.

At finite temperature, because of the system fluctuations, the chemical shift varies with time; therefore, one must average the NMR signal over a sufficiently large number of configurations. In the present work, we have performed NMR calculations every 77.3 fs for complexes **A1**–**A3** and **B**1.

Because the configuration **A**1 (Figure 2) is unstable and evolves to model **A**2, the ^{13}C chemical shifts were calculated for **A**2 using the sampling interval of 77.3 fs. This corresponds to averaging over 20 configurations. For **A**3 and **B**1, 50 configurations were considered. To investigate the dependence of our results on the choice of pseudopotentials and basis set, test calculations were also carried out on 10 configurations taken from **B**1 with Goedecker pseudopotentials[75,76] and 100 Ry of cutoff. It turned out that the relative shifts change by as little as ∼0.1 ppm.

Calculation of ^{13}C NMR Isotope Shifts. Deuterium isotope effects on the ^{13}C chemical shift were calculated for models **A**3, **B**1, **II,** and **IIb** following the procedure of ref 37.

Results

Monoprotonated Form. The four possible protomers for the monoprotonated Asp dyad are depicted in Figure 2. Ab initio geometry optimization of the four patterns show that the H-bond pattern proposed based on ^{13}C NMR data[23] (**A**1) is *not* the most stable protomer. It is not surprising that the **A**1 optimized structure is higher in energy and more distorted than patterns **A**2 and **A**3, because there is no proton to stabilize the two Asp oxygen atoms that are in close proximity in the crystal structure.

A consistent picture is provided by the ab initio MD simulation of **A**1. Indeed, during the dynamics, this protomer evolves spontaneously toward protomer **A**2 after ≅ 0.5 ps, Hδ2 (Asp25′) is transferred to Oh (pepstatin A) (Figure 1), and

(68) Hutter, J.; Ballone, P.; Bernasconi, M.; Focher, P.; Fois, E.; Goedecker, S.; Parrinello, M.; Tuckerman, M. *CPMD* (3.3); MPI für Festkörperforschung and IBM Zurich Research Laboratory, 1999.

(69) Hirshfeld, F. *Isr. J. Chem.* **1977**, *16*, 198–201.

(70) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270–274.

(71) Silvestrelli, P. L.; Marzari, N.; Vanderbilt, D.; Parrinello, M. *Solid State Commun.* **1998**, *107*, 7–11.

(72) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**, *56*, 12847–12865.

(73) Alber, F.; Folkers, G.; Carloni, P. *J. Phys. Chem.* **1999**, *103*, 3, 6121–6126.

(74) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1993**, *210*, 223–231.

(75) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641–3662.

(76) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

Table 1. Selected Calculated Properties of Formic Acid/Water Complexes[a]

complex	water molecules included	$N_{C(electrons)}$[b]	MBO ratio[c]	σ ppm[d]
0-		4.141	1.00	113.8
0		3.937	1.68	111.5
I	WATA	3.950	1.58	113.6
II	WATA, WATB	3.933	1.41	115.7
III	WATA, WATB, WATC	3.917	1.34	116.6
IV	WATA, WATB, WATC, WATD	3.909	1.38	116.0
IIa	WATB, WATC	3.903	1.55	113.6
IIb	WATA, WATD	3.938	1.62	112.8
IIc	WATA, WATC	3.928	1.48	116.3

[a] Labels from Figure 4. [b] N_C is the integrated number of electrons around C. [c] Ratio between the C–O^2 and C–O^1 Mayer bond orders.[70] [d] Carboxyl carbon ^{13}C NMR chemical shifts.

Table 2. Calculated ^{13}C Chemical Shifts (ppm) and Isotopic Shifts (ppm) of Complexes A2, A3, B1[a]

		Asp25			Asp25′	
model	charge	chemical shift	isotopic shift	charge	chemical shift	isotopic shift
A2	0	128.4(1.1)		−1	128.7(1.0)	
A3	−1	127.4(0.3)	0.00(0.01)	0	124(0.4)	0.10(0.03)
B1	0	127.5(0.5)	0.06(0.01)	0	129.8(0.6)	0.11(0.02)

[a] The values reported here have not been corrected to account for the core-electron contribution (see Methods).

simultaneously, H(Oh) (pepstatin A) is transferred to Oδ1 (Asp25). In the new structure (**A**2 in Figure 2), which is stable during the time scale of the simulation, both aspartyl groups interact strongly with the pepstatin A hydroxyl group, one as H-bond acceptor (Asp25′) and one as H-bond donor (Asp25).

The calculated ^{13}C chemical shifts of the carboxyl carbon atoms fluctuate considerably, as shown by their relatively large standard deviation, but they average out to very similar values (Table 2).

The protomer **A**3 (Figure 2), which is the lowest in energy, is also stable during the ab initio MD simulation. The calculated ^{13}C chemical shifts of the two carboxylate carbons differ by 3.4 ppm (Table 2), in qualitative agreement with experimental data. However, the calculated isotopic shifts (Table 2) disagree with the experimental data, which show that the low-field resonance exhibits an isotopic effect much larger than the high-field one.[23]

The discrepancy with experiment could arise in principle from the relatively small size of our models and the short time-scale. However, the calculated NMR properties are well-converged in the time-scale considered, as shown by their small standard deviation. Furthermore, our complexes appear to capture the relevant chemical interactions at the active site, because (i) they include all the groups interacting with the Asp dyad; (ii) the cleavage site is buried inside a hydrophobic region that is unlikely to generate a strong electric field;[32] and (iii) the protein electric field in the cleavage region is small, as compared to that of other enzymes.[77] We are, therefore, forced to abandon the monoprotonated state hypothesis and seek another explanation for the NMR signal.

We suggest that both aspartyl side chains are protonated (protomers **B**1–**B**4 in Figure 2), but because of the asymmetric nature of the inhibitor, one of the two aspartic acids has a weaker interaction with its surroundings, leading to a different ^{13}C chemical shift signal and to a different isotopic effect (Figure 2).

Diprotonated Form. Ab initio energy minimization of the four possible protomers shows that **B**1 is by far the most stable

(77) Piana, S.; Carloni, P.; Parrinello, M. Submitted.
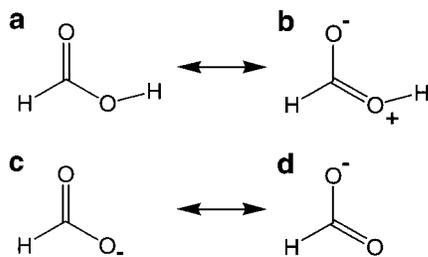
Figure 4. Resonance structures representing the π system in formic acid (**a** and **b**) and its conjugated base (**c** and **d**). While **c** and **d** are equivalent, **a** is greater than **b**, because it involves a lower degree of charge separation.

protomer (Figure 2), because it optimizes the H-bond connectivity. Interestingly, it is not dissimilar from what is known about other transition-state analog/HIV-1 PR complexes.[17,19,29] During the ab initio MD, the H-bond pattern is maintained, and the system is stable during the entire simulation time. The asymmetric inhibitor hydroxyl group forms a very strong H-bond with Asp25′ (dOO = 2.61(0.05) Å, $\angle$O$-$H...O Θ = 172(4)°) and weaker H-bond interactions with Asp25 (dOO = 2.71(0.09) Å, $\angle$O$-$H...O Θ = 152(8)°) (Figure 2). Furthermore, during the simulation, each Asp is stabilized by two additional hydrogen bonds: one with the other Asp group (Oδ1$-$H$\cdots$Oδ1) and the other with the Gly27 and Gly 27′ amide groups (N$-$H$\cdots$Oδ1 in Figure 1). The calculated chemical and isotopic shifts of the two carboxyl carbons differ significantly (Table 2), in agreement with experimental evidence.[27]

It may seem somewhat surprising that the two Asp groups in the same protonation state (model **B**1) resonate at markedly different frequencies and exhibit very different isotopic shifts. In an attempt to rationalize these facts, we now investigate more systematically the effects of H-bonding on protonated carboxylic acids in a low-dielectric medium. To this end, we calculate the ^{13}C NMR properties of formic acid/water H-bond complexes in vacuo (Figure 3) and use the result to interpret the ^{13}C NMR signals of the Asp dyad in complex **B**1.

We first perform calculations on formiate and formic acid (**0** and **0-**, Figure 3). These show a carbon deshielding of $\approx$2 ppm for formiate (Table 1). A low-field shift can only be due to a change in magnetic anisotropy (MA), and it has to be larger than the two effects producing high-field shifts, namely the increase of diamagnetic contributions due to an increase of electronic density and a decrease in the polarity of the C$-$O bond.

The origin of MA lies in the carboxylate π-electron system.[78] In the deprotonated form, the π-electrons are more delocalized, and the valence bond structures more important in HCOO$^-$ (Figure 4). This is consistent with the calculation of Mayer bond orders[70] (MBOs), which give a lower C$-$O^2 and C$-$O^1 ratio in the deprotonated form (Table 1).

These results for the isolated molecules give the key to interpreting a series of calculations performed on formic acid/water complexes (Figure 3). The results are summarized in Table 1. Clearly, the complexes can be classified into those that stabilize the π-resonance (complexes **I**, **II**, **III**) and those that do not (complexes **IIa** and **IIb**). In **IV**, both stabilizing and destabilizing interactions are present, although the latter predominate (Table 1). The MBO analysis confirms this picture. We also find that an H-bond gives a roughly additive contribution to the shielding of about 1$-$2 ppm, leading to an average value of 1.5(1) ppm. We can use the results of these model

calculations to interpret the HIV-1 PR results. In fact, the Asp side chains are in a low-dielectric-medium environment[9] and their behavior is expected to be similar to that in the gas phase.

To make a semiquantitative estimate, let us enumerate the shielding and deshielding contributions to **B**1: (i) the Asp$-$Asp H-bond, which deshields both aspartyl groups; (ii) the H-bonds of Asp25′ with Gly27′ and with the inhibitor ((Oh-O$\delta2_{ASP25}$ = 2.61(5) Å), which deshield Asp25′; (iii) the Asp25$-$Gly27 H-bond, which shields Asp25; and (iv) the Asp25-inhibitor H-bond, which deshields Asp25.

Assuming that the H-bonding contributions are additive and similar to that of water, we arrive at an estimate of 3.0 ppm, in semiquantitative agreement with the calculated difference in chemical shifts (Table 2). This increases our confidence in our results, and our simplified model can be used to obtain rough estimates in situations in which a full quantum-chemical calculation is not possible.

We now turn our attention to isotopic effects. Our calculations for two representative formiate/water complexes (**II and IIb**) are fully consistent with the experimental[79] and theoretical[37] evidence that isotope shifts depend strongly on the strength of their H-bond pattern formed by the carboxylic acids. Indeed, the calculated isotopic shift of complex **II** (0.05 ppm; H-bond distance, 1.721 Å) is larger than that of **IIb** (0.03 ppm; H-bond distance, 1.766 Å). These findings help rationalize the difference of isotopic shifts between Asp25 and Asp25′ in HIV-1 protease.[80,81] Asp25′ (Table 2), which forms strong H-bond interactions with the environment, exhibits a large isotopic effect, which can be compared with the experimental value of 0.18(0.04) ppm. In contrast, Asp25, which forms weak H-bond interactions, exhibits a small isotopic shift, which can be compared with the experimental value of 0.00(0.04) ppm.[82] Thus, the lack of isotopic shift may not always be characteristic of a deprotonated carboxylic group, as previously proposed,[23,27,29] but might be associated with a protonated carboxylic acid in a low polar medium.

Discussion

Protonation State of HIV-1 PR/Pepstatin A Complex. The interpretation of ^{13}C NMR isotopic shifts and chemical shifts at the cleavage site of HIV-1 PR is highly nontrivial, because the signals of Asp carboxyl groups may partially overlap. The Cγ signals of solvent exposed Asp groups in proteins are usually found in the range 175$-$178 ppm (protonated form) and 178$-$181 ppm (deprotonated form). Furthermore, the effect of local interactions may be very difficult to estimate.

Within the limitations of the time-scale investigated and the relatively small size of the model complex used, our ab initio calculations suggest that the protonation state proposed for the pepstatin A/HIV-1 PR complex on the basis of ^{13}C NMR data (**A**1 in Figure 2) is not the most stable pattern. Consistent results are provided by the ab initio MD simulations. Indeed, through a double proton-transfer process, **A**1 evolves in the subpico-second time-scale to a different protonation pattern (**A**2 in Figure 2). **A**2 is stable for the rest of the dynamics and exhibits almost equivalent, MD-averaged chemical shifts of the carboxyl carbons, in contrast to experiment.[23] The chemical equivalence

(78) Streitwieser, A.; Heatcock, C. H.; Kosower, E. M. *Introduction to Organic Chemistry*; MacMillian Publishing Company: New York, 1992.

(79) Reuben, J. *J. Am. Chem. Soc.* **1986**, *109*, 316$-$321.
(80) Reuben, J. *J. Am. Chem. Soc.* **1985**, *108*, 1735$-$1738.
(81) Bordner, J.; Hammen, P. D.; Whipple, E. B. *J. Am. Chem. Soc.* **1989**, *111*, 6572$-$6578.
(82) The difference in the isotopic shift between the two aspartyl groups arises mostly from the high asymmetry of the Morse potential experienced by Asp 25′, which is caused by the strong H-bond between Asp 25′ and the inhibitor.

of the Asp groups is explainable in terms of similar H-bond stabilization of the two Asp groups, leading to a similar amount of deshielding.

Ab initio molecular dynamics calculations of the most stable protomer (**A**3 in Figure 2) provide evidence that carboxylate carbons are not chemically equivalent (Table 2). However, in this pattern only, the high field resonance undergoes an isotopic shift, in contrast with the experimental data.[23]

Therefore, we propose that a diprotonated, neutral form of the Asp dyad is present in the active site of the enzyme. We focus on pattern **B**1, which is by far the most stable of the four possible protomers (Figure 2). The diprotonated state turns out to be stable in the time-scale investigated; furthermore, the two calculated ^{13}C chemical and isotopic shifts are markedly different, in qualitative agreement with experiment.[23]

We notice that the calculated difference in NMR parameters is smaller than that obtained experimentally. Several effects may be responsible for this discrepancy, from the use of the BLYP approximation[60,61] to the relatively small size of the quantum-mechanical models used (for instance, use of acetic acid for the Asp groups and 2-propanol for pepstatin) and the short time-scale; nevertheless, we would like to stress that our model captures the difference in ^{13}C NMR properties of two identical groups in an anisotropic chemical environment.

To rationalize the inequivalence of the two Asp groups, we have developed simple a valence bond model based on calculations of formic acid/water complexes. Our simple model suggests that the modification of the chemical environment of the two C sites is caused by the interactions of the Asp dyad with the asymmetric hydroxyl group of the inhibitor.

^{13}C NMR Properties of HIV-1 PR/Inhibitor Complexes.
Our calculations both on model complexes and on the HIV-1 PR/pepstatin A complex can be summarized using the following simple rules:

1. In the *protonated form* of the Asp group, both isotopic effects and deshielding of the ^{13}C NMR signal increase with the strength of H-bond interactions between the carboxylic OH group and the inhibitor. Specifically, the magnitude of isotopic shifts turns out to be negligible for O–H···O distances larger than 1.7 Å.

2. The *deprotonated form* exhibits larger deshielding than the protonated form and negligible isotopic shifts, in agreement with experiment.[34]

We test these two simple rules on the experimental evidences available on HIV-1 PR/inhibitor complexes,[23] namely the adducts with DMP323,[27,29,83] KNI-272[84] and pepstatin A.[32]

The adduct with the symmetric DMP323 inhibitor[29] exhibits one signal at 176.4 ppm, with a sizable isotopic shift. Our rules predict that the ^{13}C NMR signal should arise from two chemically equivalent protonated Asp groups, in agreement with the previous assignment.[29] The presence of a sizable isotopic shift is totally consistent with the available X-ray structural data of the complex.[83] In the structure, both of the Asp O–H groups form strong hydrogen bond interactions with the hydroxyl group belonging to the inhibitor ($d(O_{Asp}-O_{DMP323}) = 2.5$ Å).

In the complex with the asymmetric drug KNI-272,[27] two signals are present. The low-field signal (at 176.0 ppm) exhibits a large isotopic shift. Thus, rule 1 predicts that this signal should arise from a protonated Asp group. The high-field signal (177.4

ppm) could in principle be due either to a protonated carboxyl group with strong H-bond interactions or to a deprotonated Asp. Because this signal is characterized by a very small isotopic shift, however, it must arise from a deprotonated Asp (Rule 2). Inspection of the X-ray structure of the complex[84] allows the protonated Asp to be readily identified as Asp25, in agreement with the previous assignment.[27]

Of course, the complex with pepstatin A[23] also obeys this rule. Indeed, the high-field signal exhibits a large isotopic shift, which is readily assigned to a protonated Asp (Rule 1). The extremely low-field (172.40 ppm) signal does not exhibit an isotopic shift. Because the deprotonated form is *always* more deshielded than the protonated form (rule 2), the signal cannot arise from a deprotonated carboxyl group. Furthermore, the absence of isotopic shifts suggests that the Asp group forms weak H-bond interactions. Thus, we conclude that the second Asp group also is protonated and that it forms weak H-bond interactions (rule 1) (Figure 1), in contrast to the previous suggestion.[23]

In conclusion, the validity of such simple rules can be of great value in the future interpretation of ^{13}C NMR properties of HIV-1 PR aspartyl dyad, because these rules can replace rather elaborate quantum-chemical calculations. However, we notice that in this first investigation, the full calculation was required to establish the validity of the scheme.

Protonation State of Unbound and Bound HIV-1 PR. The finding that the Asp dyad in the complex with pepstatin A is doubly protonated is fully consistent with the recent isothermal titration experiments on pepstatin A's binding to HIV-1 PR, which pointed to an uptake of 0.4 protons upon binding at pH 6.5.[85] Indeed, because the free enzyme is most probably monoprotonated,[19,66,86] we may postulate that the Asp pair charge changes from −1 to 0 upon pepstatin binding.

It is rather intriguing that some inhibitors, such as pepstatin A,[29] DMP323,[29] XL-075,[19] and U85548e[17] stabilize this neutral Asp dyad, but other inhibitors[17,19,27,86,87] stabilize the negatively and di-negatively charged forms. In this section, we attempt to provide a general hypothesis about the protonation state of all of the inhibitor/HIV-1 protease complexes.

Generally speaking, the presence of an inhibitor bound to the protein is expected to *decrease* the polarity of the medium at the cleavage site relative to the free enzyme, where the Asp dyad is accessible to the solvent. Thus, we may expect that inhibitor-binding causes a stabilization of the diprotonated, neutral Asp dyad. This is, indeed, the case with some complexes,[17,19,28] among which is pepstatin A (as shown in this work).

Why then do inhibitors chemically related to pepstatin A (such as KNI-272) stabilize the monoprotonated forms? Inspection of the adducts' X-ray structures for which the protonation state is known[32,83,84,87,88] may provide an answer to this question. It turns out that the monoprotonated state is always accompanied by the presence of two strong hydrogen bonds between the Asp dyad and H-bond donors belonging either to the inhibitor or to an ordered water molecule ($d(O-O) < 3.0$ Å). These highly stabilizing interactions are not present in the diprotonated form. We, therefore, suggest that these H-bond interactions can

(83) Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N. *Science* **1994**, *263*, 380−384.
(84) Baldwin, E.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. *Structure* **1995**, *3*, 581−590.

(85) Xie, D.; Gulnik, S.; Collins, L.; Gustchina, E.; Bhat, T. N.; Erickson, J. W. *Adv. Exp. Med. Biol.* **1998**, *436*, 381−386.
(86) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454−8463.
(87) Silva, A. M.; Cachau, R. E.; Sham, H. L.; Erickson, J. W. *J. Mol. Biol.* **1996**, *255*, 321−346.
(88) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600−1609.

stabilize the negatively charged Asp dyad by providing a rather polar medium that is not very dissimilar from that of the free enzyme.

Finally, in the cases in which the inhibitor is positively charged,[17,86] also the *nonprotonated*, highly charged Asp dyad may be stabilized because of favorable electrostatic interactions with the inhibitor.

Implications for the Catalytic Mechanism of Aspartic Proteases. Our finding that the Asp dyad is doubly protonated may provide the key for understanding the kinetic mechanism of aspartyl proteases that was proposed by Northrop and co-workers on the basis of solvent isotopic effects of pepsin.[21] In this proposal, the substrate binds to the free enzyme (E form) and products dissociate from a different form of the enzyme (F form). The difference between E and F was not known. Pepstatin was also found to bind in the F form[22] with a puzzling inverse solvent isotope effect on the slow-binding kinetics. Our previous work on HIV-1 PR suggested that the E form is locked in a coplanar conformation by a low-barrier hydrogen bond (LBHB) connecting the two Asp groups.[66a] The LBHB, which highly stabilizes the Asp dyad, might account for the inverse solvent isotope effect, because a LBHB has a very low fractionation factor;[89] on the other hand, our present calculations indicate that such an interaction is not present when the enzyme is bound to pepstatin. We, therefore, suggest that the transition between the F and E forms of the enzyme is associated with the formation of the LBHB.

Conclusions

Our calculations suggest that both Asp groups in the complex with pepstatin A are protonated, in contrast to previous hypotheses. The chemical inequivalence of the carboxyl carbons arises from strong interactions with nonsymmetric neighboring groups. Our calculations are fully consistent with the experimental data available on HIV-1/pepstatin complexes, and they provide the key for interpreting the NMR properties on other adducts. The calculations also allow the key step of the kinetic isomechanism of proteases to be identified.

(89) Cleland, W. W.; Frey, P. A.; Gertl, J. A. *J. Biol. Chem.* **1998**, *273*, 25529−25532.