Analysis of pH-dependent elements in proteins: Geometry and properties of pairs of hydrogen-bonded carboxylic acid side-chains

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20328 · Source: PubMed

1 AUTHOR:

Gerd Wohlfahrt
Orion Pharma

49 PUBLICATIONS **1,275** CITATIONS

SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 58:396–406 (2005)

Analysis of pH-Dependent Elements in Proteins: Geometry and Properties of Pairs of Hydrogen-Bonded Carboxylic Acid Side-Chains

Gerd Wohlfahrt[*†]
University of Cologne, Institute of Biochemistry, Köln, Germany

ABSTRACT **A rather frequent but so far little discussed observation is that pairs of carboxylic acid side-chains in proteins can share a proton in a hydrogen bond. In the present article, quantum chemical calculations of simple model systems for carboxyl–carboxylate interactions are compared with structural observations from proteins. A detailed structural analysis of the proteins deposited in the PDB revealed that, in a subset of proteins sharing less than 90% sequence identity, 19% (314) contain at least one pair of carboxylic acids with their side-chain oxygen atoms within hydrogen-bonding distance. As the distance between those interacting oxygen atoms is frequently very short (~2.55 Å), many of these carboxylic acids are suggested to share a proton in a strong hydrogen bond. When situated in an appropriate structural environment (low dielectric constant), some might even form a low barrier hydrogen bond. The quantum chemical studies show that the most frequent geometric features of carboxyl–carboxylate pairs found in proteins, and no or symmetric ligation, are also the most stable arrangements at low dielectric constants, and they also suggest at medium and low pH a higher stability than for isosteric amide–carboxylate pairs. The presence of these pairs in 119 different enzymes found in the BRENDA database is set in relation to their properties and functions. This analysis shows that pH optima of enzymes with carboxyl–carboxylate pairs are shifted to lower than average values, whereas temperature optima seem to be increased. The described structural principles can be used as guidelines for rational protein design (e.g., in order to improve pH or temperature stability). Proteins 2005;58:396–406.** © 2004 Wiley-Liss, Inc.

Key words: carboxyl–carboxylate pairs; amide–carboxylate pairs; protein stability; strong hydrogen bonds; quantum mechanics; enzyme database

INTRODUCTION

The biotechnology industry has a long-standing interest in improving protein stability at high temperature and against chemical denaturants, oxidation, and proteolysis. There are two main approaches to accomplish this: directed evolution and rational (structure-based) protein design. Rules concerning protein stability are now better understood since site-directed mutagenesis studies have provided experimental data on the relative stability of many mutant proteins. This has confirmed the dominating influence of hydrophobic interactions in the core of a protein structure, but has also revealed the significant contribution of electrostatic interactions and hydrogen bonds.

Hydrogen bonding in many different forms plays an essential role in all type of biological systems[1] (e.g., for ligand binding, catalysis, protein folding, and stabilization). Frequently mentioned dependencies of the strength of a hydrogen bond are its length and linearity, its environment (ligands and solvation), and how well the pK_a values of the proton donor and acceptor match.[2] If the distance between donor and acceptor is shortened, the energy barrier can drop until the zero-point energy level is reached: This is defined as an LBHB.[3,4] The proposal that LBHBs might play a role in enzymatic catalysis was first put forward in 1993.[5,6] Although there is debate as to what extent hydrogen bonds found in proteins are actually LBHBs,[7,8] there are several examples for strong hydrogen bonding of either (ionic) SSHBs or potential LBHBs.[4,9] One type of strong hydrogen bond could be formed between two oxygens that share a proton at low pH (e.g., in two interacting carboxyl groups).[10,11]

In 1982, Sawyer and James[12] commented on the existence of carboxyl–carboxylate interactions in proteins and their possible importance in stabilizing protein–protein interactions. Since then, only one systematic study by Flocco and Mowbray[13] has analyzed the occurrence of these carboxyl–carboxylate pairs in proteins in a subset of the PDB,[14] consisting of proteins with less than 25% amino acid sequence identity. Of the 151 protein structures analyzed, the authors found some common tendencies in the spatial arrangement and environmental proper-

Abbreviations: 3D, three-dimensional; DFT, density functional theory; EC, Enzyme Commission; LBHB, low barrier hydrogen bond; PDB, Protein Data Bank; PSA, protein surface accessibility; QM, quantum mechanics; SCRF, self-consistent reaction field; SSHB, short strong hydrogen bond; ZPVE, zero point vibrational energy.

[†]Present address: Orion Pharma, P.O. Box 65, FIN-02101 Espoo, Finland. E-mail: gerd.wohlfahrt@orionpharma.com

*Correspondence to: Gerd Wohlfahrt, Orion Pharma, P.O. Box 65, FIN-02101 Espoo, Finland. E-mail: gerd.wohlfahrt@orionpharma.com

Received 28 June 2004; Accepted 12 August 2004

Published online 19 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20328

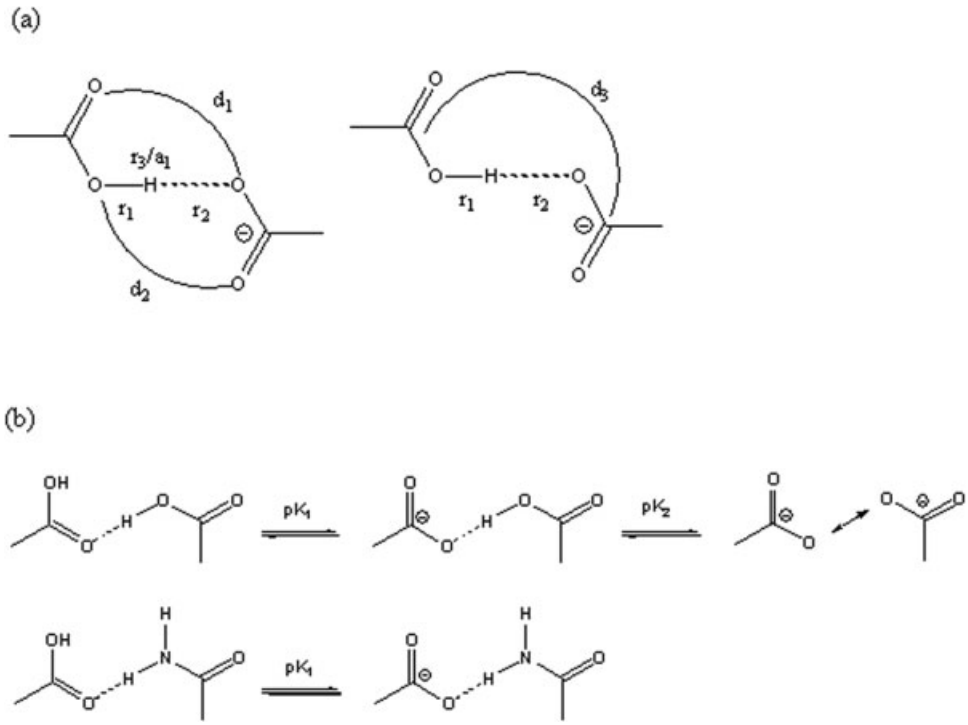

Fig. 1. (**a**) Definition of distances, angles and dihedrals used to describe the geometry of interacting pairs. Distances r_1 (O—H), r_2 (H···O), r_3 (O···O); angle a_1 (O—H···O); dihedrals d_1 and d_2 (O—C—O···O), and d_3, the angle between the planes of the carboxyl groups defined via the dihedral (C—O···O—C). (**b**) Schematic representation of different protonation states of carboxyl–carboxylate (top) and amide–carboxylate pairs (bottom) in the physiologically relevant pH range. The carboxyl–carboxylate pair is drawn so that the hydrogen bond is in the energetically favorable anti–syn geometry with respect to the noninteracting oxygen atoms.

ties of the 28 observed carboxyl–carboxylate pairs. They suggested that this type of carboxyl–carboxylate pairs might be important in enzyme catalysis and binding of substrates.

Mortensen and Breddam showed that the stability of a serine carboxypeptidase can be significantly decreased by removing a pair of glutamates in the active center of the enzyme.[15] The destabilizing effect was strongest at low pH values. At high pH values, this glutamic acid bridge tends to act as a destabilizing element due to charge repulsion. Also, the Cel6A enzyme from *Trichoderma reesei* contains carboxyl–carboxylate pairs, of which three have been subjected to site-directed mutagenesis.[16] The resulting mutant enzymes with amide–carboxylate pairs exhibit lower stability in acidic to medium pH and higher stability in the alkaline pH range compared to the wild-type.

As it is difficult to measure the strength of hydrogen bonds in proteins directly, theoretical methods are useful to improve the understanding of these systems. Quantum chemical studies have mainly focused on the formate–formic acid system in the gas phase, which could be considered as a simple model for the interactions of carboxyl side-chains in proteins.[10] Main determinants for the strength of similar hydrogen bonds are described in the literature as (1) the match of the pK_a values of proton donor and acceptor,[2,6,17,18] (2) geometry (bond length and angle),[19] and (3) polarity of the environment.[10,18] As the environment seems to play an important role for the nature of those hydrogen bonds, it should be included in quantum mechanical studies (e.g., by explicit treatment of interacting proximate residues and by simulating a "virtual solvent" using the SCRF method).[20,21]

The present study extends the quantum chemical calculations of simple model systems for hydrogen bonds between two carboxyl groups by using a solvation model,[21] compares the results with structural observations from proteins in the PDB,[14] and attempts to relate these to enzyme properties and functions, especially pH- and temperature-dependent activity and stability, found in the BRENDA database.[22]

METHODS

Ab Initio Calculations

Acetate anions and acetic acid molecules were used in quantum chemical calculations as model system for carboxyl–carboxylate pairs observed in proteins. The functionality of these molecules is equivalent to that of aspartate and glutamate side-chains in proteins. Acetamide was used to mimic asparagine and glutamine. Four different conformations are defined by the hydrogen bond being in syn or anti orientation relative to the oxygen not involved in the hydrogen bond [Figs. 1(a) and 2]. To study the influence of neutral hydrogen-bond donors from the protein environment and of microsolvation, one or two water molecules were added to the anti–syn system (Fig. 2).

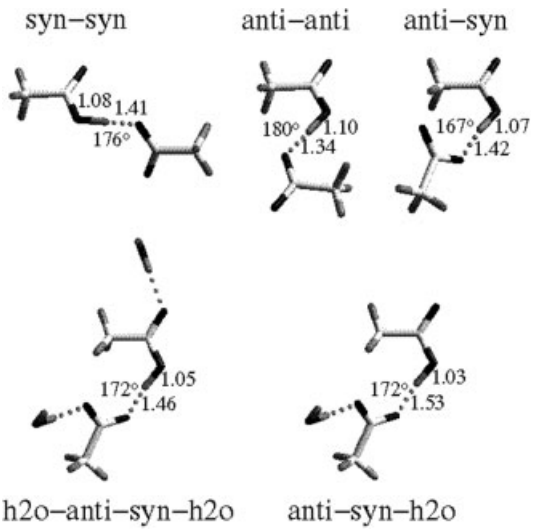

Fig. 2. The minimum geometry of three different configurations of acetate and acetic acid pairs resulting from optimization at the B3LYP/6-31+G** level at $\epsilon_{eff} = 4$ (**top**). The minimum geometry of the di- and monohydrated anti–syn system in the gas phase $\epsilon_{eff} = 1$ (**bottom**). Anti and syn refer to the conformation of the hydrogen bond, respectively, of the hydrogen atom relative to the oxygen atom, which is not involved in the hydrogen bond.

The ab initio calculations were performed with the JAGUAR 5.5 program package.[23] All structures were optimized using the 6-31+G** basis set at the B3LYP level of theory. B3LYP is a hybrid density functional incorporating Becke's exchange functional,[24] the Lee–Yang–Parr correlation functional,[25] and a Hartree–Fock exchange term. For comparison, selected structures were optimized using Moller–Plesset perturbation theory (MP2).

The continuum reaction field solvation model (SCRF),[21] implemented in JAGUAR, was used to simulate effects of the dielectrically heterogeneous protein–bulk water environment on the stability of the hydrogen bonds and on the shape of their potential profiles. SCRF represents the reaction field as a number of induced surface point charges at the solute–solvent interface. The SCRF solvation free energy was calculated at effective dielectric constants ϵ_{eff} of 4 and 80 using a water probe with a radius of 1.4 Å.

The shape of the potential energy surface of the hydrogen bond for some systems has been estimated by moving the proton in small increments of 0.025 Å from the hydrogen-donor oxygen in the acetic acid toward the interacting oxygen in the acetate.

Additional calculations were performed with the GAUSSIAN 98 program package.[26] Selected structures were further optimized in the gas phase at the B3LYP/6-311++G** level in order to study the effect of the basis set on the energy of different conformations. For those systems also the contribution of the ZPVEs obtained from frequency calculations on the stability of different conformations was compared.

Analysis of Protein Structures

The protein structures for this study were taken from the PDB.[14] A representative subset prepared by Hooft et al.[27] was used in the geometric analysis. This subset was extracted from the 06/22/1999 release of the PDB and contains 1619 protein with less than 90% sequence identity solved by X-ray analysis, with an *R*-factor of less than 0.25 and a resolution of better than 2.5 Å.

In the following, programs from the CCP4 suite[28] and Perl scripts were used to perform database searches and geometric analyses. An initial search for short distances (< 2.9 Å) between carboxyl oxygens from different side-chains was done with the CONTACT program (by Tadeusz Skarzynski) in order to identify all potential carboxyl–carboxylate pairs. The cutoff was chosen in order to reduce sampling of structures where, for example, carboxyl oxygens are only in indirect contact, because they bind to the same hydrogen-bond donor. In a second step, CONTACT was used to identify all carboxylic acid oxygens that are a short distance from each other because they complex metal ions. These pairs were excluded from further analysis.

The resulting subset of carboxyl–carboxylate pairs was analyzed with the GEOMCALC program (by Phil Evans) for their relative spatial arrangement. The angle a_1, which describes the linearity of a hydrogen bond [Fig. 1(a)], was calculated as the average of the two C—O···O angles, which are close to 120° for a linear hydrogen bond. Addition of 60° takes the geometry of an ideal linear hydrogen bond into account. The dihedral angles d_1 and d_2 between oxygen, carboxyl–carbon, and the oxygens involved in the hydrogen bond in the first and in the second residue and vice versa were determined [Fig. 1(a)]. The combination of these two dihedrals classifies the hydrogen bonds into the syn or anti configuration relative to the other oxygen. Pairs with d_1 or d_2 between 75° and 105° were excluded from certain further analysis, because their classification into one of the geometric groups is not well defined (Fig. 2). The dihedral angle d_3 between the carbons and oxygens forming the hydrogen bond [Fig. 1(a)] describes how much the plane of one carboxyl group deviates from the plane of the other (interplanar angle).

For comparative purposes, certain analyses were repeated for amide–carboxylate and amide–amide pairs with the same PDB subset as above, but using a distance cutoff of 3.5 Å.

The solvent accessibility of side-chain oxygen atoms of the residues forming hydrogen bonds was calculated with the method of Lee and Richards,[29] implemented in the program PSA[30] using a probe radius of 1.4 Å.

Correlation With Enzyme Properties and Function

Three hundred fourteen different PDB codes of protein structures, wherein we identified carboxyl–carboxylate pairs, were used as query for the BRENDA database, which contains enzyme data and metabolic information extracted from primary literature.[22] For the matches, we compared the distribution of pH and temperature optima, and pH and temperature stability, with the average of all available enzyme data stored in BRENDA. In cases where several pH or temperature values have been found for a single enzyme, the values have been averaged.

TABLE I. Relative Energy (in kcal/mol) and Minimum Geometries for Different Spatial Arrangements of Carboxyl–Carboxylate and Amide–Carboxylate Pairs Using B3LYP/6-31 + G and a Dielectric Constant $\varepsilon_{eff} = 4$**

	syn–syn (COOH-COO)	syn–syn ($CONH_2$-COO)	anti–syn (COOH-COO)	anti–syn ($CONH_2$-COO)	anti–anti (COOH-COO)
ΔE_{HB}[a]	1.1	4.0	0.0	0.0	0.1
r_3 O–O/N	2.49	2.92	2.47	2.82	2.44
r_1 O/N–H	1.08	1.03	1.07	1.04	1.10
r_2 H. . .O/N	1.41	1.89	1.42	1.79	1.34
a_1 O–H–O/N	176.3°	174.5°	166.9°	173.2°	179.6°
d_1/d_2	3.1°/3.7°	4.7°/3.2°	179.2°/13.1°	179.9°/0.5°	179.8°/179.9°
d_3	165.0°	160.0°	56.4°	177.9°	177.5°

[a]The energy (kcal/mol) of different configurations is compared to the anti–syn system.
The anti–anti conformation for the amide–carboxylate pair is not stable without constraints; therefore, these values are excluded.

RESULTS AND DISCUSSION

Quantum Chemical Geometry and Energies of Carboxyl–Carboxylate and Amide–Carboxylate Pairs

An interesting type of hydrogen bond can be formed between the oxygens in two interacting carboxyl groups that share a proton at medium and low pH values. The pK_a values of such carboxyl–carboxylate pairs are expected to differ significantly from free or differently ligated aspartate or glutamate side-chains in proteins. At very low pH values, both of the carboxylic acid side-chains in a carboxyl–carboxylate pair are protonated, and a "normal strength" hydrogen bond is expected to be formed [Fig. 1(b)]. At intermediate pH (around pH 3–6), a proton is shared between both carboxyl groups, and this is suggested to lead to the formation of a strong hydrogen bond. This would result in a carboxyl–carboxylate pair with three microscopic pK_a values, which combine to the two macroscopic pK_a values shown in Figure 1(b). In contrast, if a "normal" hydrogen bond is formed between the two carboxyl groups, the ionization equilibria of the two interacting acid–base groups involve four microscopic pK_a values (i.e., the pK_a of each group in the presence of a neutral or charged neighbor). At high pH values, both carboxyl groups are negatively charged, and this leads to a repulsion of the side-chains [Fig. 1(b)]. In contrast to a carboxyl–carboxylate pair, a corresponding amide–carboxylate pair can exist in only two different protonation states at physiological pH, and in both states a hydrogen bond can be formed [Fig. 1(b)]. Below pH 4–5, the carboxylic acid side-chain is protonated, which probably destabilizes the hydrogen bond due to weaker electrostatic interactions. Above pH 4–5, the conjugate base is charged, and a "normal hydrogen bond" is formed between the carboxylate and amide groups.

To understand in more detail the structural properties of carboxyl–carboxylate pairs found in proteins, we built pairs with acetic acid interacting with acetate as models for different spatial arrangements of this system. The geometry of each different pair was optimized at the B3LYP/6-31+G** level of theory using ϵ_{eff} values of 1, 4, and 80. To study the effect of the basis set on the energy of different conformations, selected structures were optimized in the gas phase at the B3LYP/6-311++G** level. For those systems also, the contribution of the ZPVE was evaluated. The energy differences between different conformations are only slightly affected by using the ZPVE correction and the larger basis set (relative energies are increased for anti–anti to 0.4 and for syn–syn to 2.0 kcal/mol). As DFT methods and the B3LYP functional have been shown to be unreliable in some cases where dispersion forces are expected to play a dominant role,[31,32] the carboxyl–carboxylate pairs have been studied in gas phase at the MP2/6-31+G** level of theory. The relative energies of the conformations show the same trend, but the differences are even more pronounced (anti–syn, lowest; anti–anti, +2.3 kcal/mol; syn–syn, +4.8 kcal/mol; anti–anti barrier height, 0.6 kcal/mol), whereas geometries differ only slightly from the DFT calculations. These results suggests that the use of B3LYP/6-31+G** should be sufficient for the systems studied in the present article.

From all different start geometries, we obtained three different structures corresponding to anti–anti, anti–syn, and syn–syn minimum structures (Fig. 2). Table I shows their relative energies and selected geometric features at $\epsilon_{eff} = 4$. The gas phase calculations ($\epsilon_{eff} = 1$) represent a situation where the carboxyl–carboxylate pair is situated in a very low dielectric protein environment with no access to the solvent. Despite the fact that many carboxyl–carboxylate pairs found in protein structures are not accessible to the solvent, even for these cases, the $\epsilon_{eff} = 1$ model probably underestimates the polarity of the environment, because it does not account for reorganization of polar groups and water penetration.[33] An ϵ_{eff} of 4 could be a good assumption for the low dielectric environment in which many carboxyl–carboxylate pairs are situated, but this has to be estimated case by case, as the electrostatic properties of different protein sites can strongly deviate from one another.[33,34] The environment of pairs with higher solvent accessibility can be represented by the use of ϵ_{eff} between 4 and 80.

Interestingly, the syn–anti structure with the hydroxyl proton bound in syn geometry to the oxygen converted to the anti–syn structure where the proton is closer (bound)

TABLE II. Quantum Chemical Energies (in Hartrees) for the Minimum Conformations of Different Model Systems on the B3LYP/6-31 + G Level of Theory**

Model system	ε = 1	ε = 4	ε = 80
Acetic acid (syn)	−229.105815	−229.117038	−229.119465
Acetic acid (anti)	−229.096247	−229.111594	−229.117118
Acetate	−228.543484	−228.634254	−228.663823
Acetamide	−209.237630	−209.252386	−209.257973
Acetic acid (anti)–H_2O	−305.540834	−305.560311	—
Acetic acid (syn)–H_2O	−305.550746	−305.565753	—
Acetate–H_2O	−305.004996	−305.088248	—
Acetic acid–acetate (syn–syn)	−457.689018	−457.767381	−457.796137
Acetic acid–acetate (anti–syn)	−457.694548	−457.769096	−457.793147
Acetic acid–acetate (anti–anti)	−457.692266	−457.768953	−457.793271
Acetamide–acetate (syn–syn)	−437.813253	−437.892331	—
Acetamide–acetate (anti–syn)	−437.820474	−437.898693	−437.923730
Acetic acid–acetate–H_2O	−534.150422	−534.220388	—
H_2O–acetic acid–acetate–H_2O	−610.601928	−610.673602	—

The SCRF solvation free energy is included for the calculations at ε_{eff} = 4 and 80.

to the other oxygen. This is in contrast to the observation that the anti conformation of the hydroxyl group in an isolated acetic acid molecule is 1.5–6.1 kcal/mol higher in energy than syn (Table II), but a stabilizing effect on syn conformations in carboxylic acids upon hydrogen bonding has also been described in the literature.[35]

The distances between hydrogen and oxygen are remarkably longer in all three minimum structures than in isolated acetic acid (0.97 Å) (Table I). In the anti–syn configuration, there is a further contact between the two molecules in addition to the hydrogen bond; the oxygen, which is not involved in hydrogen bonding, has a distance of 3.36 Å to the methyl carbon of the other acetate and of 2.35 Å to the closest methyl hydrogen (Fig. 2). This van der Waals contact might give some additional stabilization. The angle of the hydrogen bond is close to 180° for the anti–anti configuration, whereas it deviates slightly from linearity for the other systems (Table I). The interplanar angles of the carboxyl groups are close to 60° or 180° for the minimum structures (Table I), but rotational barriers around these hydrogen bonds are very low, which allows several different angles.

With ϵ_{eff} = 4, the anti–syn system shows the lowest energy, but anti–anti has an only 0.1 kcal/mol higher energy, and the syn–syn configuration lies 1.1 kcal/mol above this. At ϵ_{eff} = 80, syn–syn is the most stable conformation. The stabilizing energy of these hydrogen bonds drops strongly (~16 kcal/mol), going from ϵ_{eff} = 1 to 4, whereas going from ϵ_{eff} = 4 to 80, the stabilizing effect is reduced only by 2–5 kcal/mol.

The symmetric anti–anti system has the shortest O—O and the longest O—H distance. Its potential profile has the shape of a single well, with a small barrier of 0.1 kcal (ϵ_{eff} = 1) or 0.5 kcal (ϵ_{eff} = 4) (Figure 3). Therefore the hydrogen position is less fixed, as, for example, in amide–carboxylate hydrogen bonds. With ϵ_{eff} = 80, the profile has a more double-well character, with a barrier height of approximately 2.5 kcal (Fig. 3). In the anti–syn system at ϵ_{eff} = 4, the bottoms of the two wells differ only by about 0.8 kcal/mol.

The anti–anti system with its high symmetry and identical pK_a values of both partners fulfills, in principle, the environment-independent requirements for a LBHB, but the formation of a LBHB is strongly determined by the protein environment, and the shape of the potential does not necessarily correlate with the strength of the hydrogen bond.[7,36,37]

Upon microsolvation of the anti–syn complex (Fig. 2), by adding one water molecule, we observe a destabilization of the hydrogen bond by 3.8 kcal/mol at ϵ_{eff} = 1. In contrast, when two water molecules interact with the acid and the acetate, even a slight stabilization of the hydrogen bond by 0.6 kcal/mol is found. Similar results have been described for the syn–syn configuration of the formate–formic acid system.[10] A possible explanation is a stabilizing effect when proton donor and acceptor have matching pK_a values, which are less similar in the monohydrated system. The geometry of the microsolvated systems deviates only slightly from that of the nonhydrated ones by a small increase of the hydrogen-bond length (0.03 Å) in the monohydrated model, while this distance is shortened by 0.02 Å with two water molecules in the system. The simulation of the mono- and dihydrated anti–syn system with ϵ_{eff} = 4, which is also a simple model for a low dielectric protein environment with hydrogen donors bound to the carboxyl–carboxylate pair, also showed a destabilization (+2.4 kcal/mol) of the hydrogen bond by one water ligand and a stabilization (−1.1 kcal/mol) by symmetric ligation with two water molecules. A stabilizing effect of a single water molecule on a strong hydrogen bond between Asp32 and His64 in subtilisin has been suggested by Schiott.[38]

The stabilizing energy of acetate–acetic acid hydrogen bonds calculated with B3LYP/6-31++G* ranges from 24.9 to 28.4 kcal/mol in the gas phase, from 10.1 to 11.1 kcal/mol for ϵ_{eff} = 4, and 6.2 to 8.1 kcal/mol for ϵ_{eff} = 80

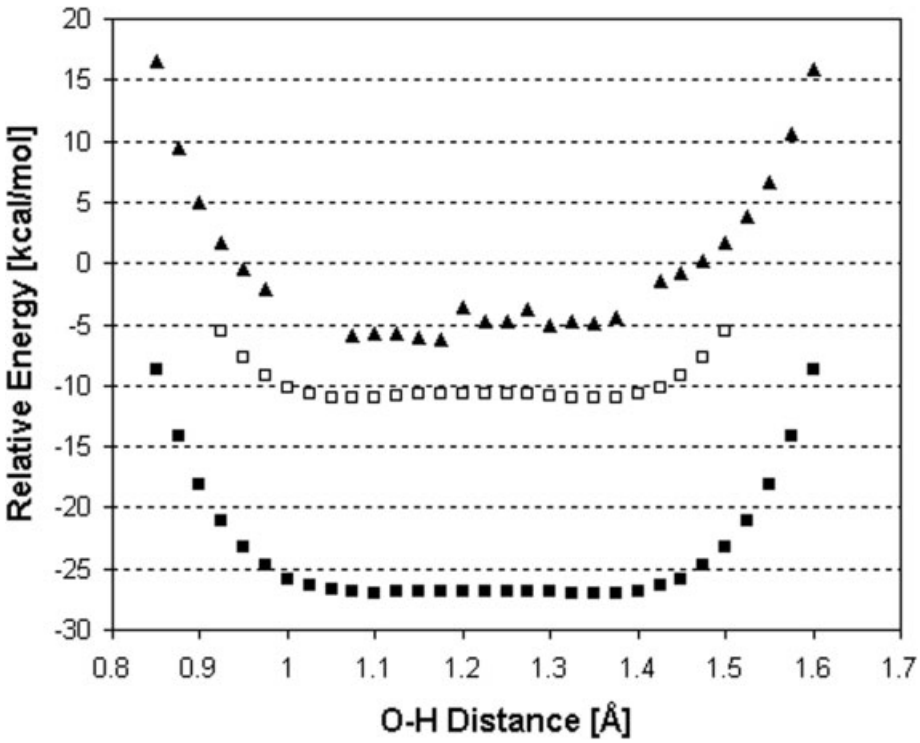

Fig. 3. Potential energy profiles of a hydrogen bond between acetate and acetic acid in the syn–syn conformation calculated with B3LYP/6-31+G** at $\epsilon_{eff} = 1$ (■), $\epsilon_{eff} = 4$ (□), and $\epsilon_{eff} = 80$ (▲).

(Table II). The acetamide–acetate interactions provided up to 24.7 kcal/mol stabilization in the gas phase, 7.6 kcal/mol at $\epsilon_{eff} = 4$, and 1.2 kcal/mol at $\epsilon_{eff} = 80$.[39] This suggests a higher stability of hydrogen bonds between two carboxyl groups compared to amide–carboxylate interactions independent of the dielectric constant of the environment. This also correlates well with a 3–5°C decreased stability of a carboxyl–carboxylate to amide mutation in the *Trichoderma reesei* Cel6A enzyme.[16]

Analysis of Carboxyl–Carboxylate and Amide–Carboxylate Pairs in Protein Structures

In order to compare the observations from QM calculations of the simple model systems with corresponding structural elements in proteins, a nonredundant subset of structures from the PDB was analyzed. In 1619 protein structures with less than 90% sequence identity listed by Hooft et al.,[27] we found 314 different proteins (19.4% of the subset) containing 475 pairs of carboxylic acid oxygens that are in closer contact than 2.9 Å and do not concertedly complex a metal ion. Twenty-five percent of these carboxyl–carboxylate pairs consist of two aspartate, 48% of glutamate and aspartate, and 27% of two glutamate. The search for amide–carboxylate and amide–amide pairs in the same subset as described above, but now using a distance cutoff of 3.5 Å, resulted in 5750 pairs in 1236 proteins (76.3% of the subset). Two third of these potential hydrogen bonds are longer than 2.9 Å which therefore would have been no meaningful cutoff distance for this type of hydrogen bond. An analysis of geometric features [Fig. 1(a)] of these two types of hydrogen bonds is described below.

The distribution of O—O distances in carboxyl–carboxylate pairs has a maximum at 2.55 Å (Fig. 4, black bars). A second maximum has been observed around 3.1 Å,[13] which is not further analyzed here, as it partly results from indirect interactions promoted by a third partner. The maximum for amide–carboxylate and amide–amide pairs is around 3.0 Å (Fig. 4, white bars), and no significant difference is seen between these two groups (data not shown). The distributions of donor–acceptor distances should be seen in respect to the typical inaccuracy of hydrogen-bond lengths in X-ray structures with a resolution better than 2.5 Å, which is about 0.1–0.6 Å.[9]

The angles a_1 in carboxyl–carboxylate pairs are distributed around 180°, and the distribution is quite narrow [Fig. 5(a)]. It becomes even narrower if only O—O contacts of less than 2.65 Å are considered (data not shown). The a_1 angle distribution is much broader for amide–carboxylate and amide–amide pairs [Fig. 5(b)], suggesting a strong dependence of carboxyl–carboxylate hydrogen bonds on linear angles. This is in good agreement with the observation that already a 30° deviation from linearity destabilizes this type of hydrogen bond substantially.[19]

Three different groups of spatial arrangements were defined via the position of the hydrogen bond relative to the other oxygen of the carboxyl group (Fig. 2): (1) syn–syn (0–75°, 0–75°), (2) anti–syn, (0–75°, 105–180°), and (3) anti–anti (105–180°, 105–180°). Of the initial 475 pairs, 79 were excluded from further analysis because one of their dihedral angles lies between 75° and 105°, which makes classification into one of the geometric groups uncertain (Fig. 2). In the remaining 396 pairs, the anti–syn arrange-

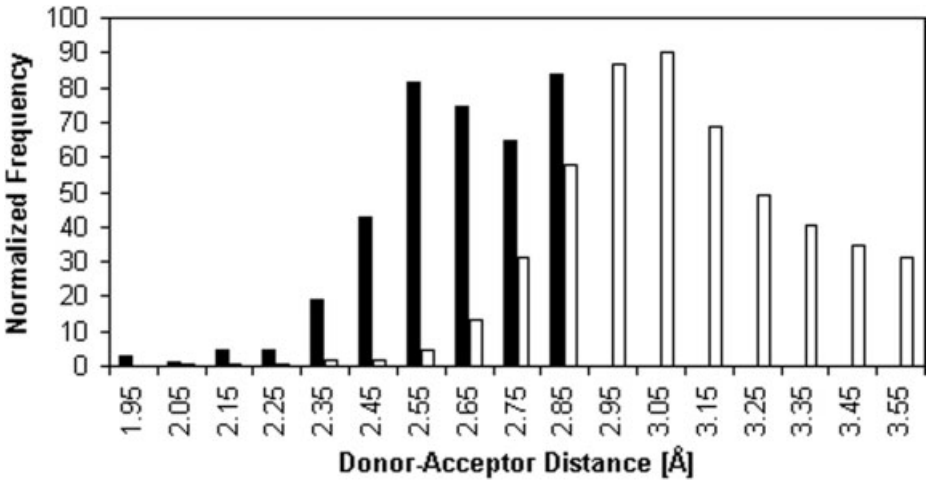

Fig. 4. Normalized distribution of the distances between potential hydrogen-bond donor and acceptor atoms in carboxyl–carboxylate (black bars), and amide–carboxylate and amide–amide pairs (white bars) in protein structures. The number of pairs in each 0.1-Å-wide distance interval was normalized by dividing it by the volume of the respective interval. In order to allow easier comparison, the normalized frequency of carboxyl–carboxylate pairs was multiplied by 7.

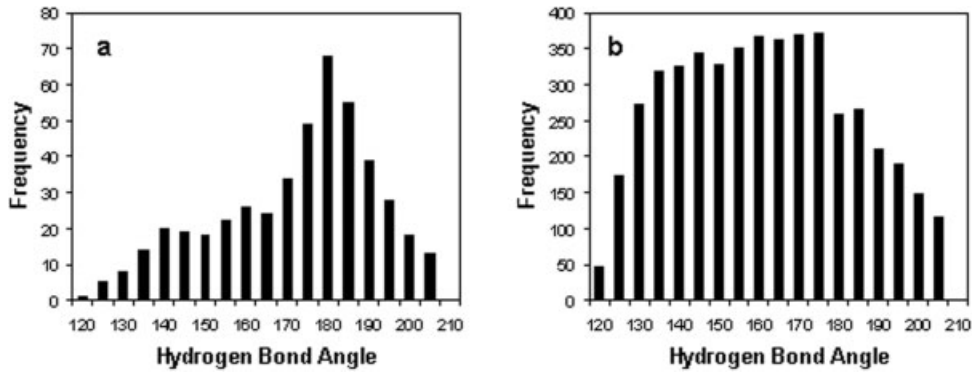

Fig. 5. Distribution of the hydrogen bond angle a_1 in (**a**) carboxyl–carboxylate, and (**b**) amide–carboxylate and amide–amide pairs in protein structures.

ment is the most common (62%) and anti–anti (14%), the least frequent. For amide containing pairs also, anti–syn is the most frequent conformation (57%), but here syn–syn occurs in 19%, the least frequent.

The relative spatial arrangement of the planes of the carboxyl groups (interplanar angle) in the three geometric groups has been determined (Fig. 6). The dihedrals are distributed over the whole range, but maximums are seen for anti–anti close to 180°, syn–syn at 70° and 130°, and for anti–syn at 100°. These correspond to energetic minimum conformations seen in the QM calculations (Table I). The distribution of O—O distances is rather similar in all three geometric groups (data not shown), with a slight but nonsignificant preference for longer distances in the syn–syn configuration, which might imply a lower stability of hydrogen bonds with this geometry, as also suggested by the QM calculations.

Forty-nine percent of the side-chain oxygens in carboxyl–carboxylate pairs have a solvent accessibility of less than 1 $Å^2$, which means that they probably do not interact with water. Only 7% of the oxygens have a solvent accessibility over 10 $Å^2$ (i.e., only in these few cases, two water molecules could interact simultaneously with the same oxygen in a carboxyl–carboxylate pair).[40] A direct correlation between the length as an indicator for the strength of hydrogen bonds and the solvent accessibility is not obvious (data not shown), but a large fraction of pairs with very low solvent accessibility and < 2.7 Å O—O distance exists (Fig. 7). Also, Flocco and Mowbray[13] observed that many pairs with short O—O distances have a low solvent accessibility. It is also remarkable that the solvent accessibility is in many cases nearly equal for both carboxyl groups (Fig. 7). This might reflect a preference for symmetric ligand binding to carboxyl–carboxylate pairs. Symmetric microsolvation has also a stabilizing effect on hydrogen bonds in QM calculations, whereas asymmetric ligation with water is destabilizing (Table II).[10]

Analyzing the potential hydrogen-bond donors close to carboxyl–carboxylate pairs, we mainly find water molecules and interactions with backbone nitrogens (Table III). There are 1490 interactions with crystal water and 1265 with backbone atoms. Forty-three pairs (9%) do not

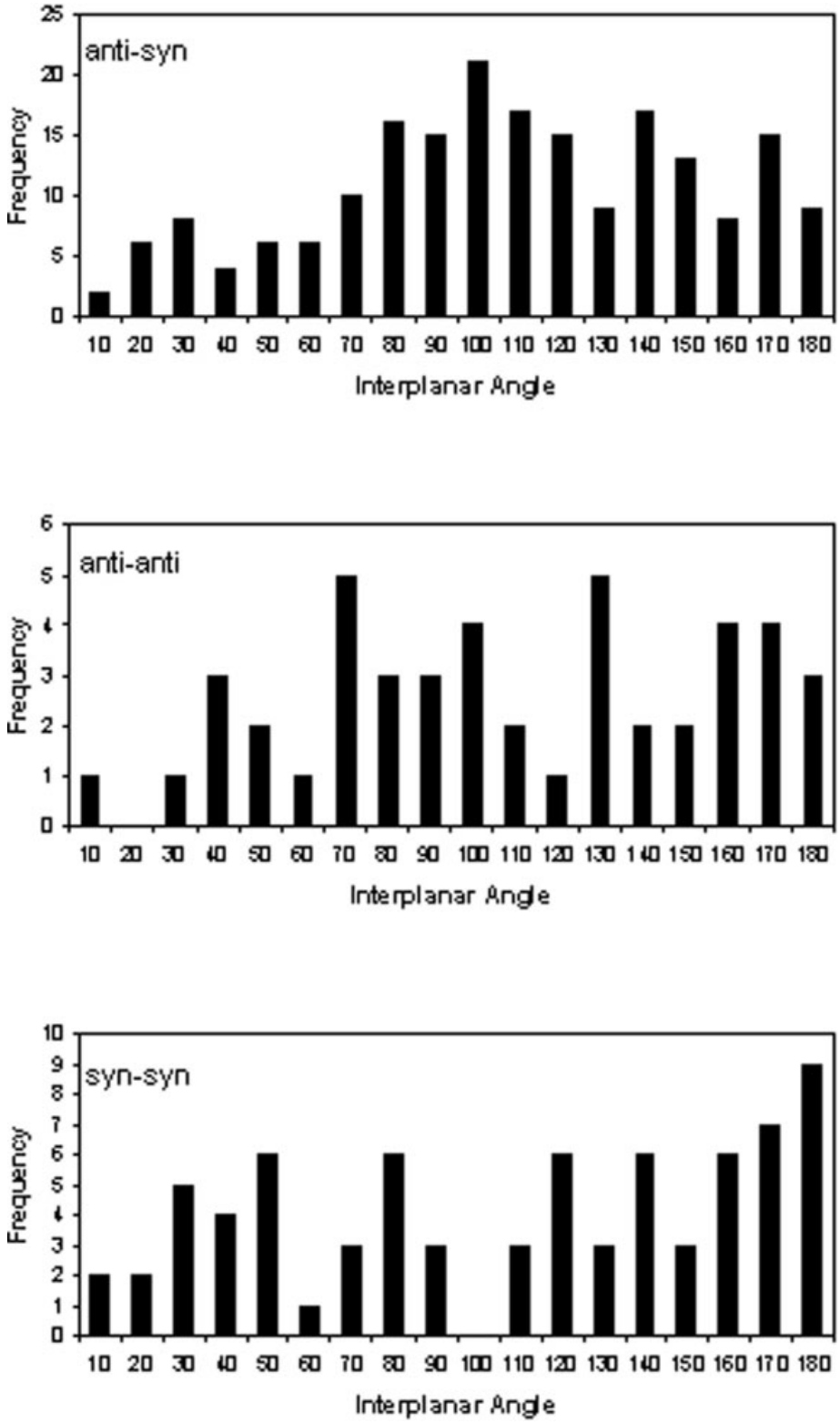

Fig. 6. Distribution of the interplanar angle d_3 in three different spatial arrangements of the carboxyl–carboxylate pairs in proteins (**top:** anti–syn; **middle:** anti–anti; **bottom:** syn–syn).

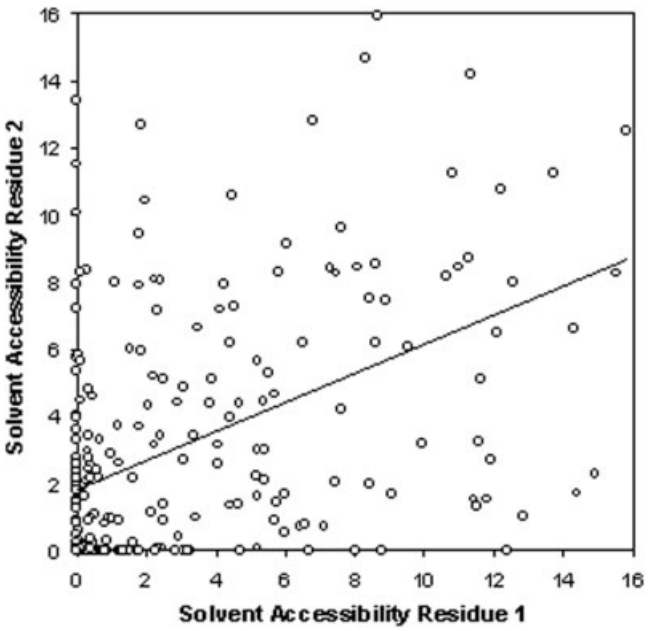

Fig. 7. Solvent accessibility of the side-chain oxygen atoms (in angstroms squared) in the first residue versus the second residue involved in the formation of carboxyl–carboxylate pairs in protein structures. The regression line indicates that the solvent accessibility of both residues forming a carboxyl–carboxylate pair is frequently very similar.

TABLE III. Distribution of Polar Groups That Are in Hydrogen-Bonding Distance (< 3.3 Å From Oxygens in Carboxyl–Carboxylate Pairs in Protein Structures

Interaction with	Number
Arg	195
Asn	53
Cys	4
Gln	45
His	117
Lys	125
Ser	67
Thr	54
Trp	36
Tyr	38
Non–amino acid	72
Crystal water	1490
Backbone N/O	1265
All	3129

show any clearly defined hydrogen-bond interactions. In 231 proteins, there are 734 polar atoms in side-chains that interact with carboxyl–carboxylate pairs. Of those, the positive residues Arg (27%), Lys (17%), and His (16%), which are also positively charged at slightly acidic pH values, are the most frequent (Table III).

As many carboxyl–carboxylate pairs observed in proteins seem to have similar structural features as the systems resulting from QM optimizations, which represent at low dielectric constants LHBHs, these pairs fulfill at least the geometric criteria to form an LHBH. A high dielectric environment (e.g., at the protein surface) probably destabilizes those hydrogen bonds substantially, but there are many carboxyl–carboxylate pairs with very low solvent accessibility whose environment should have a sufficiently low dielectric constant to stabilize an LBHB. The strength of the delocalized LBHB is not necessarily much higher than of the localized double-well form, which exists in a high dielectric medium.[36,37] The strength probably increases continuously with decreasing ϵ, without any sudden jump when the barrier height drops under the ZPVE level.[20] The contribution of carboxyl–carboxylate pairs to protein stability compared to amide–carboxylate pairs might be, at neutral pH, rather small, but more pronounced differences are expected at acidic pH, where amide–carboxylate hydrogen bonds are weakened by protonation,[39] or in alkaline pH range, where repulsion between carboxylate groups leads to destabilization.

Function and Properties of Carboxyl–Carboxylate Pairs Containing Enzymes

In an attempt to ascertain if proteins with carboxyl–carboxylate pairs have properties that differ from most other proteins, functional data from the BRENDA enzyme database were analyzed. For 179 of the 314 PDB entries coding for structures with carboxyl–carboxylate pairs described above, an EC number and organism information could be retrieved from the BRENDA database version from April 2002.[22] There are 119 unique EC numbers, whose distribution over the six main enzyme classes is shown in Table IV. The distribution is similar to that in the whole database, and no enzyme class shows a clear preference for carboxyl–carboxylate pairs. The higher fraction of EC class 3 entries is related to the higher number of pH and temperature data available for this class (e.g., 47% of all pH and 36% of all temperature stability values in BRENDA are for hydrolases). The 179 enzymes in the subset are from 100 different organisms, and 16 proteins are of human origin.

Information about pH and temperature optima and stability was not available for all entries. When several pH or stability values were found for a single enzyme, the values have been averaged in order to avoid bias by certain enzymes with many database entries (Table V). But also using all explicit values, without averaging, leads to very similar results (data not shown). The average pH optimum in the carboxyl–carboxylate pair subset using all 156 explicit values for 60 unique enzymes is 6.5 ± 1.6, whereas this is 7.1 ± 1.5 for all 9819 values in BRENDA.[22] When looking separately at the EC class 3 (hydrolases), the average pH optimum for the carboxyl–carboxylate pair subset and the whole database are both around 6.4. As most hydrolases work in more acidic conditions, no significant shift of their pH optimum is expected by the presence of carboxyl–carboxylate pairs. A larger difference is seen for the pH stability, which is one unit lower than the average of the whole BRENDA database (Table V). This indicates that carboxyl–carboxylate pairs probably influence catalytic function only in a few cases, but they seem to be more important for protein stability at low pH values.

The average temperature optimum is about 6°C higher in the carboxyl subset, but two maxima at about 40°C and

TABLE IV. The Distribution of Unique EC Numbers Over the Main Enzyme Classes in the Carboxyl–Carboxylate Pairs–Containing Subset and in the Whole BRENDA Database

Main class	Oxidoreductases	Transferases	Hydrolases	Lyases	Isomerases	Ligases	All
COOH to subset	26	16	58	7	8	4	119
% of subset	22	14	49	6	7	3	100
Brenda	983	1023	1018	342	147	122	3637
% of Brenda	27	28	28	9	4	3	100

TABLE V. The Distribution of Temperature, pH Optima, and Stability in the Carboxyl–Carboxylate Pairs–Containing Subset of Enzymes and in the Whole BRENDA Database[15]

	pH opt. all	pH opt. (only EC 3)	pH opt. (not EC 3)	pH stabil. all	T opt. all	T opt. vertebrates	T opt. bacteria	T stabil. all
Subset	6.6 ± 1.4	6.3 ± 1.3	6.9 ± 1.4	4.8 ± 1.5	48 ± 17	36 ± 11	57 ± 16	53 ± 19
Data in subset	60	30	30	45	65	15	20	80
Data in BRENDA	9819	3272	6547	2330	663	93	125	6823
BRENDA	7.1 ± 1.5	6.4 ± 1.7	7.5 ± 1.1	5.8 ± 1.9	42 ± 19	38 ± 15	44 ± 16	50 ± 21

60°C are observed (data not shown), which suggests a subdivision of the data set for further analysis (e.g., according to different classes of organisms). The separate analysis of the temperature optima for vertebrates and bacteria, whose enzymes quite often operate at higher temperatures, shows for the vertebrates no significant difference, whereas bacterial enzymes with carboxyl pairs show a 13°C higher optimum than the average of all bacterial enzymes (Table V). There is less significant influence on the temperature stability seen in the presence of carboxyl–carboxylate pairs (Table V). Even if the differences between the data sets are statistically not significant because of the high standard deviation, it is obvious that the carboxyl–carboxylate pair-containing subset includes at least a higher fraction of proteins operating in more acidic conditions than the average enzymes in the BRENDA database. One problem of the data in BRENDA is the relatively large variation of experimental pH and temperature optima even for a single enzyme, which have been averaged in our study. Averaging all values of the data sets results in a high standard deviation. Grouping the enzymes into classes reduces the standard deviation, but also the number of examples in each group might become rather small for a meaningful interpretation, as for example, in the separate analysis of the temperature optima for vertebrates and bacteria (Table V).

The subset probably contains proteins where carboxyl–carboxylate pairs do not have any influence on the stability or function of the enzymes. In these instances the spatial arrangement of side-chains may be an artifact caused by, for example, acidic crystallization conditions. However, at least those proteins where carboxyl–carboxylate pairs have a quite ideal geometry or a nonpolar environment are interesting for further examination.

CONCLUSIONS

In about 19% of nonredundant protein structures from the PDB, one can find carboxyl–carboxylate pairs that have geometric features distinct from the isosteric amide–carboxylate pairs. Many carboxyl–carboxylate pairs in proteins have a low solvent accessibility and a geometry similar to the corresponding acetic acid–acetate systems resulting from QM optimizations, which represent LHBHs at least in low dielectric media. This suggests that some pairs in proteins could form LHBHs as well, but a conclusive answer to this question is beyond the scope of the present article and has to be determined for each individual case. Results from the QM calculations do not suggest that those hydrogen bonds are necessarily much stronger than "conventional" ones, but also the high-dielectric double-well form of carboxyl–carboxylate hydrogen bonds seems to be at least stronger than, for example, amide–carboxylate hydrogen bonds.

A related observation is that a substantial fraction of enzymes with carboxyl–carboxylate pairs is more stable than the average of enzymes found in the BRENDA database. This can be attributed either to higher stability due to more favorable protonation states at low pH values or greater strength of this type of hydrogen bond also at medium pH. The stability advantage should be especially prominent at low pH values, where other carboxylic acid–containing hydrogen bonds can be weakened by protonation of carboxylates.[39] Therefore, carboxyl–carboxylate pairs could be considered as an evolutionary strategy to stabilize proteins at acidic pH values.

A potential application of the results from the present analysis is the rational design of strong hydrogen bonds in a protein. Since the strength of a hydrogen bond depends on several factors, rules that predict the position and type of amino acid exchanges are required to increase the stability of a protein. Based on geometric features observed from QM calculations and from the PDB, it is possible to identify pairs of amino acids in 3D structures of proteins that could be altered to carboxyl–carboxylate pairs in order to increase the protein stability at low or medium pH. Important criteria for stable carboxyl–carboxylate pairs are (1) a short distance (< 2.7 Å) between hydrogen donor and acceptor atoms, (2) an angle close to 180° for the hydrogen bond, (3) preferably an anti–syn arrangement of the carboxyl groups, (4) very low solvent

accessibility of the carboxyl oxygens, or (5) similar or no solvent–ligand interactions for both carboxyl groups involved in the hydrogen bond.

Carboxyl–carboxylate pairs can also be removed in order to stabilize a protein molecule at higher pH, where they have a destabilizing effect.[16] Furthermore, since carboxyl–carboxylate pairs have a repulsive interaction at high pH values (pH > 7), introducing carboxyl–carboxylate pairs could be used to introduce pH-dependent switches that would alter the protein properties, such as ligand binding, at elevated pH values.

ACKNOWLEDGMENTS

The help of Svena Lüdke and Thorsten Fallisch with the analysis of structural data and Dr. Christian Ebeling with the BRENDA database is gratefully acknowledged. Drs. Harry Boer, Anu Koivula, Vijay Kumar, and Lars-Olof Pietilä are thanked for helpful comments on the manuscript.

REFERENCES

1. Jeffrey GA, Saenger W. Hydrogen bonding in biological structures. Berlin: Springer-Verlag; 1991. 569 p.
2. Remer LC, Jensen JH. Toward a general theory of hydrogen bonding: the short, strong hydrogen bond $[HOH^{\cdots}OH]^{-}$. J Phys Chem A 2000;104:9266–9275.
3. Perrin CL, Nielson JB. "Strong" hydrogen bonds in chemistry and biology. Annu Rev Phys Chem 1997;48:511–544.
4. Cleland WW, Frey PA, Gerlt JA. The low barrier hydrogen bond in enzymatic catalysis. J Biol Chem 1998;273:25529–25532.
5. Gerlt JA, Gassman PG. Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters. Biochemistry 1993;32:11943–11952.
6. Cleland WW, Kreevoy MM. Low-barrier hydrogen bonds and enzymic catalysis. Science 1994;264:1887–1890.
7. Warshel A, Papazyan A. Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. Proc Natl Acad Sci USA 1996;93:13665–13670.
8. Schutz CN, Warshel A. The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp…His pair in serine proteases. Proteins 2004;55,711–723.
9. Harris TK, Mildvan AS. High-precision measurement of hydrogen bond lengths in proteins by nuclear magnetic resonance methods. Proteins 1999;35:275–282.
10. Pan Y, McAllister MA. Characterization of low-barrier hydrogen bonds: 1. Microsolvation effects: an ab initio and DFT investigation. J Am Chem Soc 1997;119:7561–7566.
11. Tolstoy PM, Schah-Mohammedi P, Smirnov SN, Golubev NS, Denisov GS, Limbach HH. Characterization of fluxional hydrogen-bonded complexes of acetic acid and acetate by NMR: geometries and isotope and solvent effects. J Am Chem Soc 2004;126:5621–5634.
12. Sawyer L, James MNG. Carboxyl–carboxylate interactions in proteins. Nature 1982;295:79–80.
13. Flocco MM, Mowbray SL. Strange bedfellows: interactions between acidic side-chains in proteins. J Mol Biol 1995;254:96–105.
14. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
15. Mortensen UH, Breddam K. A conserved glutamic acid bridge in serine carboxypeptidases, belonging to the α/β hydrolase fold, acts as a pH-dependent protein-stabilizing element. Protein Sci 1994;3:838–842.
16. Wohlfahrt G, Pellikka T, Boer H, Teeri TT, Koivula, A. Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. Biochemistry 2003;42:10095–10103.
17. Kumar AG, McAllister MA. Characterization of low-barrier hydrogen bonds: 8. Substituent effects on the strength and geometry of the formic acid–formate anion model system: an ab initio and DFT investigation. J Am Chem Soc 1998;120:3159–3165.
18. Shan SO, Herschlag D. The change in hydrogen bond strength accompanying charge rearrangement: implications for enzymatic catalysis. Proc Natl Acad Sci USA 1996;93:14474–14479.
19. Smallwood CJ, McAllister MA. Characterization of low-barrier hydrogen bonds: 7. Relationship between strength and geometry of short-strong hydrogen bonds: the formic acid–formate anion model system: an ab initio and DFT investigation. J Am Chem Soc 1997;119:11277–11281.
20. Scheiner S, Kar T. The nonexistence of specially stabilized hydrogen bonds in enzymes. J Am Chem Soc 1995;117:6970–6975.
21. Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitcoff D, Honig B. A new model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short range hydrogen bonding effects. J Phys Chem 1996;100:11775–11788.
22. Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res 2002;30:47–49.
23. JAGUAR 5.5. Portland, OR: Schrödinger, Inc.; 2003.
24. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 1988;38:3098–3100.
25. Lee C, Yang G, Parr RG. Development of the Colle–Salvetti correlation energy formula into a functional of the electron density. Physical Review B 1988;37:785–789.
26. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski, VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98, revision A5. Pittsburgh: Gaussian, Inc.; 1998.
27. Hooft RWW, Sander, C, Vriend G. Verification of protein structures: side-chain planarity. J Appl Crystallogr 1996;29:714–716.
28. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
29. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;14:379–400.
30. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington, JP. JOY: protein sequence–structure representation and analysis. Bioinformatics 1998;14:617–623.
31. Kristyan S, Pulay P. Can (semi)local density functional theory account for the London dispersion forces? Chem Phys Lett 1994;229:175–180.
32. Valdes H, Sordo JA. Ab initio and DFT studies on van der Waals trimers: the OCS·(CO2)2 complexes. J Comput Chem 2002;23:444–455.
33. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 2001;44,400–417.
34. Shokhen M, Albeck A. Is there a weak H-bond→LBHB transition on tetrahedral complex formation in serine proteases? Proteins 2004;54:468–477.
35. Li Y, Houk KN. Theoretical assessments of the basicity and nucleophilicity of carboxylate syn and anti lone pairs. J Am Chem Soc 1989;111:4505–4507.
36. Garcia-Viloca M, Gonzalez-Lafont A, Lluch JM. On pK_a matching as a requirement to form a low-barrier hydrogen bond: a theoretical study in gas phase. J Phys Chem A 1997;101:3880–3886.
37. Overgaard J, Schiott B, Larsen FK, Iversen BB. The charge density distribution in a model compound of the catalytic triad in serine proteases. Chemistry 2001;7:3756–3767.
38. Schiott B. The influence of solvation on short strong hydrogen bonds: a density functional theory study of the Asp–His interaction in subtilisins. Chem Commun 2004;5:498–499.
39. Huyghues-Despointes BM, Klingler TM, Baldwin RL. Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln.Asp (i, i + 4) interaction. Biochemistry 1995;34:13267–13271.
40. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;12:2577–2637.