Comparative Modeling of Substrate Binding in the S_1' Subsite of Serine Carboxypeptidases from Yeast, Wheat, and Human†

Marc-André Elsliger,‡,§ Alexey V. Pshezhetsky,‡ Maya V. Vinogradova,|| Vytas K. Švedas,|| and Michel Potier*,‡

Service de Génétique Médicale, Hôpital Sainte-Justine, Département de Pédiatrie, Université de Montréal, Montréal (Québec) Canada, H3T 1C5, and A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia, 119899

Received November 29, 1995; Revised Manuscript Received April 24, 1996⊗

ABSTRACT: Human cathepsin A ("lysosomal protective protein"; E.C.3.4.16.5) is a multifunctional lysosomal protein which forms a high-molecular-weight complex with β-galactosidase and α-neuraminidase, protecting them against intralysosomal proteolysis. In addition to this protective function, cathepsin A is a serine carboxypeptidase and the understanding of its catalytic function requires a definition of its substrate specificity. For this purpose, we used a combined experimental [Pshezhetsky, A. V., Vinogradova, M. V., Elsliger, M.-A., El-Zein, F., Svedas, V. K., & Potier, M. (1995) *Anal. Biochem. 230*, 303−307] and theoretical approach comparing cathepsin A to two different homologous carboxypeptidases of the same family: yeast carboxypeptidase Y and wheat carboxypeptidase II. We computed the energies involved in substrate binding to the S_1' subsite (C-terminal) of cathepsin A using a structural model based on the X-ray structure of the homologous wheat carboxypeptidase II. The binding energies of N-blocked Phe-Xaa dipeptide substrates to the active sites of cathepsin A, wheat carboxypeptidase II, and yeast carboxypeptidase Y were estimated using a molecular mechanics force field supplemented with a solvation energy term. This theoretical analysis showed a good correlation with the experimentally determined free energies of substrate binding. This result validates the use of this approach to analyze the energetics of substrate binding to the S_1' subsite and provides a rational interpretation of serine carboxypeptidase−substrate interactions in molecular terms. We conclude that the three serine carboxypeptidases have similar affinities for substrates with hydrophobic P_1' amino acid residues but that the wheat enzyme has an additional capacity for binding positively charged P_1' residues. Finally, the substrate specificity of human cathepsin A is very similar to that of carboxypeptidase Y, with a high binding affinity for substrates with hydrophobic P_1' residues, but the affinity of cathepsin A for P_1' Phe residue is higher than for the Leu residue.

Serine carboxypeptidases catalyze the hydrolysis of C-terminal amino acids from peptides and proteins. These exopeptidases which also possess esterase and deamidase activities, are found in all eukaryotes studied (Breddam, 1986; Remington & Breddam, 1994; Jackman *et al*., 1990). Serine carboxypeptidases with high affinity for hydrophobic amino acid residues at the P_1' position are classified as C-type carboxypeptidases (EC 3.4.16.5) while those with higher affinity for basic P_1' residues are classified as D-type carboxypeptidases (EC 3.4.16.6) (Remington & Breddam, 1994). Among the best characterized serine carboxypeptidases are carboxypeptidase Y (CPD-Y)[1] from yeast, a C-type enzyme, and carboxypeptidase II (CPD-WII) from wheat, a D-type enzyme. Although these two carboxypeptidases have only about 25% overall sequence identity (Breddam *et al*., 1987), their three-dimensional folds are very similar (Remington et Breddam, 1994) and belong to the α/β-hydrolase family (Ollis *et al*., 1992). The tertiary structures of CPD-WII (Bullock et al., 1994) and CPD-Y (Endrizzi et al., 1994) have recently been determined at 2.0 and 2.8 Å resolution, respectively. The active site Ser-His-Asp "catalytic triad", the carboxylate binding site as well as the conserved residues which form the S_1' binding site[2] adopt a similar conformation in both enzymes. However, the S_1' subsite of CPD-WII contains four charged Asp/Glu residues while that of CPD-Y contains exclusively uncharged hydrophobic amino acid residues (Endrizzi *et al*., 1994).

The human lysosomal serine carboxypeptidase, cathepsin A (CATH-A, also known as the "lysosomal protective protein" and carboxypeptidase L), plays an important structural role in the lysosome by forming a high-molecular-weight complex with β-galactosidase (GAL, EC 3.2.1.23) and α-neuraminidase (NEUR, EC 3.2.1.18), protecting them against intra-lysosomal proteolysis (Hoogeveen *et al*., 1981; d'Azzo *et al*., 1982; Verheijen *et al*., 1982, 1985). Inherited deficiency of CATH-A caused by mutations in the CATH-A

† This research was supported by an operating grant (MT-10956) from the Medical Research Council of Canada to M.P.

* Address correspondence to this author at Service de Génétique Médicale, Hôpital Sainte-Justine,3175 Côte Sainte-Catherine, Montréal, Québec, Canada, H3T 1C5. Tel: (514) 345-4931/3585. FAX: (514) 345-4766. E-mail: potier@justine.umontreal.ca.

‡ Université de Montréal.

§ Present address: Institute of Molecular Biology, University of Oregon, Eugene, OR.

|| Moscow State Unversity.

⊗ Abstract published in *Advance ACS Abstracts*, November 1, 1996.

[1] Abbreviations: GAL, β-galactosidase; NEUR, α-neuraminidase; CATH-A, cathepsin A (also known as "lysosomal protective protein" and carboxypeptidase L); CPD-WII, wheat serine carboxypeptidase II; CPD-Y, yeast carboxypeptidase Y; CBZ, *N*-carbobenzoxy; BTEE, benzoyltyrosine ethyl ester; FA, furylacryloyl; BZS, L-benzylsuccinic acid, rms, root mean square.

[2] The binding site notation is that of Schechter and Berger (1967), where the binding site of the C-terminal amino acid of the substrate is denoted S_1' and the corresponding substrate residue is denoted P_1'.

S0006-2960(95)02833-9 CCC: $12.00

gene is responsible for a lysosomal storage disorder, named galactosialidosis, which is characterized by secondary deficiencies of both GAL and NEUR activities in patients' cells (Shimmoto *et al*., 1990, 1993; Zhou *et al*., 1991; Fukuhara *et al*., 1992). Site-directed mutagenesis of active site residues of CATH-A demonstrated that its catalytic function is not required to exert its lysosomal protective role on GAL and NEUR (Galjart *et al*., 1991). However, the conservation of its catalytic function throughout evolution suggests that CATH-A's carboxypeptidase activity has a physiological function. It has been proposed that CATH-A is involved in the maturation and/or degradation of certain regulatory peptides since human lysosomal and platelet CATH-A hydrolyze *in vitro* a wide spectrum of natural bioactive peptides including endothelin, Met-enkephalin-Arg-Phe, angiotensin I, bradykinin, and substance P (Jackman *et al*., 1990, 1992; Matsuda, 1976; Kawamura *et al*., 1976; Itoh *et al*., 1995; Marks *et al*., 1981; Miller *et al*., 1988).

Recent work from our laboratory confirmed that CATH-A is a C-type carboxypeptidase (Pshezhetsky *et al*., 1995). Homology modeling of CATH-A (Elsliger & Potier, 1994) based on the X-ray structure of CPD-WII (3sc2; Liao *et al*., 1990, 1992) and more recent crystalographic studies (Rudenko *et al*., 1995) indicated that CATH-A has the same α/β-hydrolase fold and active-site residues conformation as CPD-WII and CPD-Y but contains both polar and hydrophobic amino acid residues in its S_1' subsite.

The structural homology between active site residues of serine carboxypeptidases allowed us to rationalize the substrate specificity of CATH-A by comparing the energetics of substrate binding to CATH-A's S_1' subsite with those of CPD-WII and CPD-Y. The molecular mechanic based theoretical free energies of binding were computed using the CHARMm's force-field (Brooks *et al*., 1983) supplemented with a solvation energy term (Eisenberg & McLachlan, 1986) and compared with the experimental data. This approach was inspired by earlier work in which molecular mechanics based free-energy calculations were successfully used to approximate the Gibbs free energy involved in antigen−antibody (Novotny *et al*., 1989) and enzyme-substrate (Wilson *et al*., 1991) interactions.

EXPERIMENTAL PROCEDURES

Enzymes. Human placental CATH-A was purified by affinity chromatography on agarose-Phe-Leu followed by FPLC gel-filtration as described by Pshezhetsky and Potier (1994). This preparation was subjected to an additional purification step in order to separate CATH-A from cathepsin D, which copurifies with CATH-A on the agarose-Phe-Leu column (Pshezhetsky & Potier, 1994). For this purpose, the preparation was dialyzed overnight against 10 mM Tris-HCl buffer, pH 7.5, centrifuged at 14 000*g* for 15 min, and applied to a FPLC Mono Q HR column (Pharmacia) equilibrated with the same buffer. CATH-A was eluted with 40 mL of a linear gradient of NaCl (0−0.4 M). The purified preparation of CATH-A was dialyzed against 20 mM sodium acetate buffer, pH 4.75, containing 0.15 M NaCl and 0.02% (w/v) NaN_3 (buffer A) and concentrated approximately to 0.5 mg of protein/mL in a Minicon concentration cell (Amicon). The preparation was more than 95% pure on SDS−PAGE (Laemmli 1970) and did not contain any detectable endopeptidase and prolyl-carboxypeptidase activities when assayed with Bz-Arg-Gly-Phe-Phe-Leu-4MβNA and Z-Pro-Ala substrates, respectively.

CPD-Y was purchased from Sigma Chemical Co. (St. Louis, MO) and used without additional purification.

Peptide Substrates. The N-terminal blocked CBZ-Phe-Xaa-OH and Fa-Phe-Xaa-OH dipeptides (where CBZ is carbobenzoxy, FA is furylacryloyl, and Xaa = Ala, Arg, Glu, Gly, Leu, or Phe) were purchased from Bachem Bioscience Inc. (U.S.A.) and Constanta Ltd (Moscow, Russia), respectively. All substrates were more than 98% pure based on reverse-phase HPLC analysis.

Enzyme Assay. CATH-A and CPD-Y enzyme activities with FA-Phe-Xaa-OH substrates were measured spectrophotometrically at 25 °C in 0.15 M sodium phosphate buffer, pH 6.0, as described by Pshezhetsky *et al*. (1995). The substrate concentration was varied from 0.0125 to 5.0 mM. In this range of substrate concentrations all reactions obeyed first-order kinetics. The K_M values were calculated from the dependency of the initial enzymatic reaction rate on substrate concentration using a nonlinear regression method.

Inhibition Analysis of Peptide Binding to CATH-A. The inhibition constants (K_i) of the CBZ-Phe-Xaa-OH dipeptides for CATH-A were determined with benzoyltyrosine ethyl ester (BTEE) as a substrate. The buffer and pH conditions were the same as for the FA-Phe-Xaa-OH dipeptide hydrolysis, but the concentration of the BTEE substrate was varied from 40 to 300 μM and that of the dipeptide inhibitors was varied from 0 to 5 mM. The initial rate of BTEE hydrolysis was measured spectrophotometrically at 256 nm using a $\Delta\epsilon = 1100\ M^{-1}\ cm^{-1}$ (Rick 1965). The K_i values were obtained with a Dixon plot of the dependencies of the initial enzymatic reaction rate on CBZ-Phe-Xaa-OH peptide concentrations using different concentrations of BTEE substrate. The competitive type of inhibition was verified for each dipeptide with a Lineweaver−Burk plot.

Experimental Substrate Binding Energy Calculations. The relative substrate binding energies ($\Delta\Delta G_{Bind}$) for CPD-Y and CATH-A, compared to a common reference state, were computed from the K_M by eq 1:

$$\Delta\Delta G_{Bind} = -RT \ln(K_M)_S + RT \ln(K_M)_{ref} \quad (1)$$

where $(K_M)_S$ and $(K_M)_{ref}$ are the experimentally determined binding constants of the FA-Phe-Xaa and FA-Phe-Gly substrates, respectively. R is the gas constant, and T is the temperature in K.

For CPD-WII, due to the unavailability of Michaelis−Menten constants, the published k_{cat}/K_M values (Breddam *et al*., 1987) were used to approximate relative binding energies with the following equation:

$$\Delta\Delta G_{Bind} = -RT \ln(k_{cat}/K_M) + RT \ln(k_{cat}/K_M)_{ref} \quad (2)$$

where the hydrolysis of CBZ-Phe-Ala was used as the reference state.

COMPUTATIONAL METHODS

In order to calculate relative binding energies for a variety of substrates, we must assume that the structure of the enzyme active site changes only slightly upon substrate binding. Indeed, comparison of the available serine carboxy-

Table 1: S_1' Residues Used in Substrate Modeling and Theoretical Binding Energy Calculations

CPD-Y	CATH-A	CPD-WII	function[a]
Asn 51	Asn 55	Asn 51	carboxylate binding[a]
Gly 52	Gly 56	Gly 52	carboxylate binding[a]
Gly 53	Gly 57	Gly 53	S'_1 subsite
Cys 56	Cys 60	Cys 56	S'_1 subsite
Ser 57	Ser 61	Ser 57	S'_1 subsite
Thr 60[b]	**Asp 64**	**Tyr 60**	S'_1 subsite[a]
Phe 64	**Thr 68**	**Glu 64**	S'_1 subsite[a]
Glu 65	Glu 69	Glu 65	S'_1 subsite[a]
Glu 145	Glu 149	Glu 145	carboxylate binding[a]
Ser 146	Ser 150	Ser 146	catalytic residue[a]
Asn 254	Asn 245	Asp 237	S'_1 subsite
Tyr 256	Tyr 247	Tyr 239	S'_1 subsite[a]
Tyr 269	**Pro 302**		S'_1 subsite[a]
Leu 272	**Asn 305**	**Glu 272**	S'_1 subsite[a]
Ser 297	**Met 333**	**Thr 302**	S'_1 subsite
His 397	His 429	His 397	catalytic residue[a]
Met 398	**Met 430**	**Glu 398**	S'_1 subsite[a]

[a] Taken from PDB 1ysc (Endrizzi et al., 1994). [b] Unconserved residues between the three enzymes are indicated in boldface type.

peptidase wild-type crystal structures (Brookhaven Protein Data Base entries 3sc2, 1whs, and 1ysc; Liao *et al*., 1992; Bullock *et al*., 1994; Endrizzi *et al*., 1994) with the same structures complexed with either Arg (leaving group; 3sc2/Arg; Liao *et al*., 1992) or L-benzylsuccinic acid (BZS, transition state analog; 1wht; Bullock *et al*., 1994) gives a root mean square (rms) deviation of 0.7 Å for the 10 conserved residues which form the S_1' binding pocket (including the oxyanion hole and active site triad). Thus, the conformation of the active site S_1' cavity seems to be relatively rigid, at least in terms of interaction with small molecules. Therefore, to a good approximation, the binding energy prediction requires that only the substrate flexibility be considered. Moreover, in order to keep the hydrogen-bonding contacts that bind the C-terminal carboxylate to the "oxyanion hole" and the scissile bond at an acceptable distance from the Oγ of the active site serine residue, the substrate backbone was kept fixed throughout the substrate modeling procedure. Within these constraints, the problem of substrate modeling (or docking) into the S_1' binding site thus becomes one of P_1' side chain modeling.

Substrate Modeling in the *S_1' Subsite*. The conformation of S_1' binding pocket residues (Table 1) observed in the X-ray crystal structures of CPD-Y (1ysc; Endrizzi *et al*., 1994) and CPD-WII (1whs; Bullock *et al*., 1994) as well as in the homology model of CATH-A (Elsliger & Potier, 1994) were used for modeling the enzyme−substrate complexes listed in Tables 2 and 3. All CPD-Y and CPD-WII coordinates were kindly provided by Dr. S. James Remington (University of Oregon).

Modeling of the substrate P_1' side chain was accomplished by finding a conformation which approaches its lowest energy conformer in each of the three enzymes studied. The substrate backbone (N-Cα-COO) conformation as well as the Cβ and Cγ χ torsion angles were directly extracted from the CPD-WII/BZS complex X-ray crystal structure (1wht; Bullock et al., 1994). BZS, which forms a transition state-like complex with CPD-WII, resembles a C-terminal Phe residue with the difference that a CH_2 group replaces the amide of the peptide bond and a OH group replaces the Cα of the preceding amino acid residue (Bullock *et al*., 1994). Conformational space of the P_1' side chain within the active site was then explored using the following three-step procedure: (1) steric clashes of side chain atoms beyond Cγ were relieved by manually spinning the χ torsion angles. This first step was monitored with the continuous energy calculation feature in QUANTA-CHARMm (Molecular Simulations Inc., Burlington, MA) which allowed the empirical location of the lowest energy substrates side chain orientation in the S_1' subsite; (2) the resulting side chain configuration was then submitted to 100 steps of steepest descent gradient (SDG) followed by 100 steps of conjugate gradient (CG) energy minimization; and (3) the substrate side chain atoms were then submitted to 50 steps of SDG minimization limited to bonded (internal) energy terms, in order to reestablish correct bond lengths and angles. These last two steps were repeated until the rms energy gradient deviation was lower than 0.1 kcal mol^{-1} $Å^{-1}$. Energy minimization steps were performed with the QUANTA-CHARMm's molecular mechanics forcefield with a non-bond distance cutoff of 8.0 Å using a polar hydrogen atom model, where only hydrogens atoms attached to polar atoms were constructed explicitly by the program. Aliphatic hydrogens were represented as "extended" carbon atoms by increasing the carbon atom van der Waals radii. No explicit hydrogen-bonding term was used (i.e., hydrogen-bonding effects were described solely by a combination of electrostatic and van der Waals forces) during these substrate modeling steps. Major controversy persists with regards to the dielectric constant (ϵ) in proteins. Although low ϵ values of 2−5 have been most commonly used for internally buried groups in proteins, larger ϵ have been more recently used [reviewed by Warshel (1991)]. As a rather general rule of value of $\epsilon = 40$ gives a good approximation for most ionizable surface groups in proteins, even for internally buried groups (Warshel, 1991). In order to get some sense of the scaling effects of the dielectric constant on charge−charge interactions in our particular systems, ϵ was initially varied from 1 to 80. A minimum dielectric value of 20 was found necessary to scale strong polar interactions and was adopted in all subsequent calculations.

Theoretical Binding Energy Calculations. The free energy of side chain binding (ΔG_{Bind}) in the S_1' site of each enzyme−substrate complex was estimated as the sum of non-bonded (electrostatic, ΔE_{Elect}; hydrogen bond, ΔE_{HB}; and van der Waals, ΔE_{VDW}) interactions between the enzyme and substrate, the enzyme solvation term (ΔG^{E}_{Sol}), the substrate solvation term (ΔG^{S}_{Sol}), and the substrate internal energy (ΔE^{S}_{bond}):

$$\Delta G_{Bind} = \omega(\Delta E_{Elect} + \Delta E_{HB} + \Delta E_{VDW}) + \Delta G^{E}_{Sol} + \Delta G^{S}_{Sol} + \Delta E^{S}_{bond} \quad (3)$$

where ω is an empirically determined weight for the non-bond terms. These non-bonded energy terms were calculated between the substrate side chain atoms and enzyme residue atoms listed in Table 1 with the INTERACTION function of the CHARMm program as described above except that explicit hydrogen bonding and solvation energy terms were added. Hydrogen bonds were calculated between donor and acceptor atoms within a distance of 3.5 Å and an angle between the donor, the hydrogen atom, and the acceptor of >90°.

Table 2: Comparison of Experimental and Computed Binding Energies for CPD-Y and CATH-A[a]

	CPD-Y				CATH-A				
substrate[b]	K_M	ΔG	$\Delta\Delta G_{exp}$	$\Delta\Delta G_{comp}$	K_M	K_i	ΔG	$\Delta\Delta G_{exp}$	$\Delta\Delta G_{comp}$
Gly	5.80	−3.05	0.00	0.00	1.75	2.38	−3.75	0.00	0.00
Ala	1.23	−3.97	−0.92	−1.17	0.83	0.91	−4.20	−0.45	−0.53
Phe	0.40	−4.63	−1.58	−2.20	0.045	0.046	−5.93	−2.18	−3.20
Arg	0.49	−4.51	−1.46	−1.41	0.35	0.27	−4.71	−0.96	−1.63
Leu	0.11	−5.40	−2.35	−3.01	0.051	0.078	−5.85	−2.10	−1.65
Glu	11.57	−2.64	0.41	−0.61	14.00	5.44	−2.74	1.01	0.84
Asn					2.50		−3.55	0.20	1.13
Met					0.27		−4.86	−1.11	−1.00

[a] All K_M and K_i values in mM; all ΔG and $\Delta\Delta G$ values in kcal/mol. [b] Substrate, refers to Xaa in the following N-blocked dipeptides, N = FA-Phe-Xaa for CPD-Y; for CATH-A, N = FA-Phe-Xaa for K_M values, and N = CBZ-Phe-Xaa for K_i values.

Table 3: Comparison of Experimental and Computed Binding Energies for CDP-WII

substrate[a]	k_{cat}/K_m[b] (s^{-1} M^{-1})	ΔG[c]	$\Delta\Delta G_{exp}$	$\Delta\Delta G_{comp}$
Ala	1000	−4.09	0.00	0.00
Phe	4200	−4.94	−0.85	−0.82
Arg	59000	−6.50	−2.41	−2.73
Asn	200	−3.14	0.950	1.29
Met	7400	−5.27	−1.18	−1.72
His	3100	−4.76	−0.67	0.26
Lys	52000	−6.43	−2.34	−1.81
Ile	5900	−5.14	−1.05	−0.45
Ser	200	−3.14	0.95	0.61
Val	4100	−4.92	−0.83	−0.88

[a] Substrate refers to Xaa in CBZ-Ala-Xaa. [b] Data taken from Breddam et al. (1987). [c] All ΔG and $\Delta\Delta G$ values in kcal mol.

The solvation energy terms (ΔG_{ESol}, ΔG^{S}_{Sol}) used in eq 3 were evaluated using the following equation:

$$\Delta G_{Sol} = \Delta\sigma_i(A_i - A_i^r) \quad (4)$$

where ΔG_{Sol} is the free energy change associated with bringing atom *i* from an aqueous environment to the protein site, $\Delta\sigma$ is an atomic solvation parameter (ASP), A_i is the solvent-accessible surface area of atom *i* in the enzyme−substrate complex and A_i^r is the reference state (Eisenberg & MacLachlan 1986). The substrate reference state corresponded to that of the substrate side chain atoms completely exposed to an aqueous environment in the same conformation as in the enzyme−substrate complex model and the enzyme reference state was that of the enzyme without substrate. The ASP values used in this study were those of Eisenberg and MacLachlan (1986): C = 16, N/O = −6, O^- = −24, N^+ = −50, and S = 21 cal $Å^{-2}$ mol^{-1} except for Asp, Glu (O^- = −12 cal $Å^{-2}$ mol^{-1}), and Arg (N^+ = −24 cal $Å^{-2}$ mol^{-1}) where the charges were considered to be equally distributed between both terminal N^+ and O^-.

The relative binding energies were thus calculated with the equation

$$\Delta\Delta G_{Bind} = \Delta G^{i}_{Bind} - \Delta G^{ref}_{Bind} \quad (5)$$

where ΔG^{i}_{Bind} was calculated from eq 3 for each substrate listed in Tables 2 and 3. ΔG^{ref}_{Bind} is the reference state calculated with Gly P_1' residue for CPD-Y and CATH-A, and with Ala P_1' residue for CPD-WII.

Calibration and Parameterization. The CPD-Y−substrate complex models were used to calibrate the non-bond weight factor ω used in eq 3. Initial comparison of the calculated binding energies with the respective experimental values for the different CPD-Y−substrate complexes revealed that the non-bond terms (van der Waals, electrostatic, and hydrogen bond) were greatly overestimated for most enzyme−substrate complexes. Therefore, the van der Waals term was adjusted to maximize the correlation between calculated and experimental binding energies with the non polar substrates (Gly, Ala, Leu, and Phe) and the combined electrostatic and hydrogen bonding term with the polar substrates (Arg and Glu). A scaling factor of $\omega = 0.35$ yielded the best overall correlation between the experimental and computed binding energy values for CPD-Y and was subsequently used to calculate the CPD-WII and CATH-A substrate binding energies.

The substrate binding constants for CPD-Y and CATH-A were assayed at pH 6.0, where Glu side chains are likely in their deprotonated negatively charged form (Mortensen *et al.*, 1994). In contrast, the kinetic parameters (k_{cat}/K_M) for CPD-WII were determined at pH 4.2, where some or all of the S_1' Glu side chains (Glu64, Glu65, Glu145, Glu272, and Glu398) are likely protonated. Therefore, the CPD-WII S_1' Glu side chains were adjusted to a semi-protonated form by scaling the Oϵ partial atomic charge by 0.5.

RESULTS AND DISCUSSION

The substrate specificity of enzymes arises from the balance between energetically favorable enzyme−substrate interactions (hydrogen bonding, hydrophobic, electrostatic, and van der Waals interactions) and energetically costly conformational strains induced by the proper positioning of the substrate into the active site. To understand how the favorable and unfavorable interactions influence the overall substrate specificity of an enzyme, we must define the energetic contributions of each participating functional group. This can be accomplished by the systematic variation of both the substrate and enzyme structures. Although the first procedure is more simple, the second involves site-directed mutagenesis which is time consuming. Alternatively, one should be able to estimate those contributions or at least guide mutagenesis studies through molecular modeling and theoretical energy calculations.

We present here the results of comparative molecular modeling of P_1' substrate side chain binding to serine carboxypeptidases from yeast, wheat, and human. This study was designed to analyze the energetics that govern substrate binding to the S_1' subsite of these carboxypeptidases in order to rationalize their substrate binding specificity.

Experimental Substrate Binding Energies. The experimentally determined free energy (ΔG) of substrate binding to CPD-Y and CATH-A are presented in Table 2. For CATH-A, the experimental binding energies of various

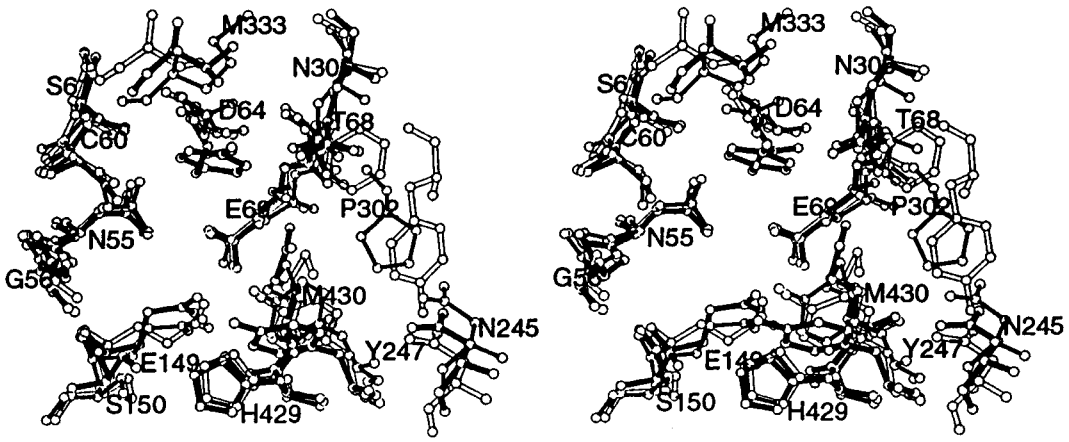

FIGURE 1: Superimposed S_1' binding sites of CPD-Y (open bonds), CPD-WII (filled bonds), and CATH-A (narrow bonds). Labels correspond to the CATH-A structure; see Table 1 for corresponding residues in CPD-Y and CPD-WII.

dipeptides were determined by two methods. First, they were calculated from the effective K_M values obtained from the direct kinetic measurement of the hydrolysis of chromogenic (FA) derivatives of the dipeptide substrates. Second, since it has been demonstrated that the same active site is responsible for both esterase and carboxypeptidase activities of serine carboxypeptidases (Breddam, 1986; Jackman *et al*., 1990), the binding constants of different CBZ-Phe-Xaa dipeptides in the CATH-A active site were measured from their capacity to competitively inhibit BTEE hydrolysis. Comparison of the obtained K_M and K_i values (Table 2) showed that the two independent methods gave similar values. For CPD-Y, we obtained K_M values of FA-Phe-Xaa substrate hydrolysis (Table 2) which were similar to those of Hayashi *et al*. (1974) for the corresponding CBZ-Phe-Xaa substrates. For CPD-WII the free energies of substrate binding (ΔG) were calculated from the published k_{cat}/K_M values of CBZ-Ala-Xaa dipeptide hydrolysis reported by Breddam *et al*. (1987) because of the unavailability of K_M values. For CPD-WII, the binding energies are considered as indicative (Table 3).

Although the nature of the amino acid residues which compose the S_1' binding sites (Table 1) differs between the carboxypeptidases, the magnitude of their binding energies for hydrophobic P_1' residues is remarkably similar (Tables 2 and 3). The major difference between the enzymes was observed for CPD-WII which has an additional capacity for binding positively charged (Arg and Lys) residues in the P_1' position.

The binding energies (ΔG) of the reference state substrate, FA-Phe-Gly, for CPD-Y and CATH-A (−3.05 and −3.76 kcal/mol, respectively, Table 2) represent the contributions of substrate P_1' residue main chains and carboxylates in addition to the P_1 (Phe) residue to the total binding energy. If we compare these values with the relative P_1' side chain binding energies ($\Delta\Delta G_{exp}$) which range from −2.35 to 0.41 kcal/mol for CPD-Y and from −2.22 to 0.86 kcal/mol for CATH-A, we can note the importance of the S_1' subsite contribution to the overall binding of dipeptide substrates. That is, for the best substrates, the P_1' side chain contributes to almost half of the overall dipeptide substrate binding energy.

Substrate Modeling. Although CPD-Y and CPD-WII have only about 25% identical amino acid residues, the overall fold of the polypeptide backbone chain is very similar and the active site Asp339/His397/Ser146 catalytic triad (including the amide nitrogen atoms of the oxyanion hole) superimpose within coordinate error with an rms deviation of only 0.29 Å (Endrizzi *et al*., 1994). When the conserved residues which form the S_1' binding pocket (Table 1) were added to the comparison, a root mean square deviation of 0.9 Å was obtained. Homologous modeling of CATH-A based on the X-ray structure of CPD-WII (Elsliger & Potier, 1994), which have about 32% sequence identity, indicated that these structural similarities are conserved in the human enzyme (Figure 1). Thus, the CATH-A model should be sufficiently accurate to allow a comparative study of substrate binding in the S_1' binding site of these three functionally related enzymes.

In order to evaluate the accuracy of our substrate modeling procedure, we compared our models of enzyme−substrate complexes with the available X-ray structures. Comparison of the CPD-WII/Phe complex model with the crystallographic CPD-WII/BZS structure showed that the minimization procedure did not substantially modify the configuration of the Phe side chain (Figure 2a). Moreover, although the Arg substrate was modeled on BZS, the final conformer obtained during substrate modeling is comparable to that found in the CPD-WII/Arg crystal structure (Figure 2b). These results, in addition to the good correlation observed between the experimentally and computed relative binding energies ($\Delta\Delta G$) values presented in Tables 2 and 3, validate the substrate modeling procedure. The structures of the lowest energy substrate conformation in the S_1' binding sites of CPD-Y, CPD-WII, and CATH-A used in the theoretical binding energy calculations are presented in Figure 3.

Analysis of the Computational Method. When modeling the energetic contributions of functional groups to the total binding energy, we assumed that no protein conformational changes occurred upon substrate binding. This assumption provides the simplest possible concept of non-covalent complex formation and allows the estimation of Gibbs free energy changes associated with substrate binding to a "rigid body" model where we considered only the flexibility of the P_1' side chain. Figure 4 shows a plot of the computed relative binding energies ($\Delta\Delta G$) *versus* the corresponding experimentally determined values for CPD-Y, CPD-WII, and CATH-A presented in Tables 2 and 3. Initial analysis of the calculated binding energies for the different CPD-Y-substrate complexes used as the calibration model, revealed

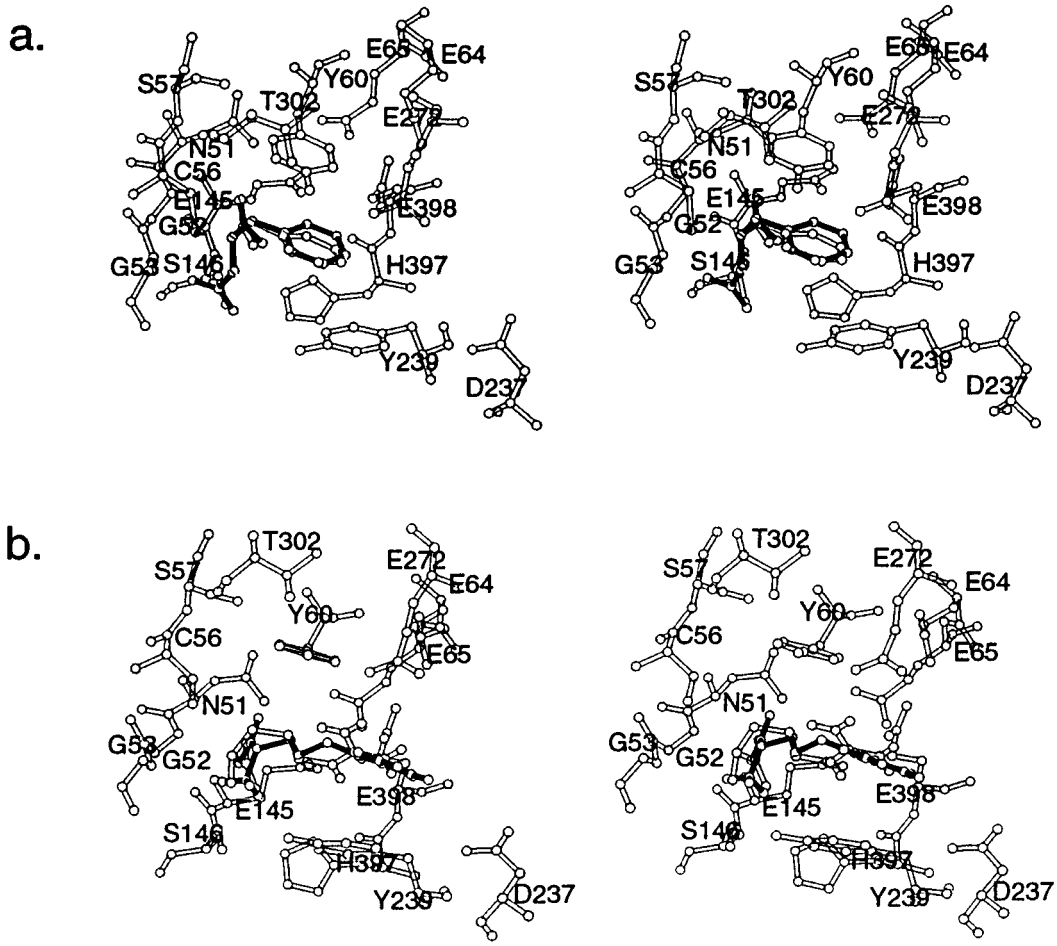

FIGURE 2: Comparison of the modeled P_1' Phe and Arg bound to CPD-WII (open bonds) with the respective crystallographic structures (filled bonds) of (a) the transition state analog BZS and (b) the reaction product Arg.

that the non-bond energy terms (van der Waals and electrostatic) were greatly overestimated for most enzyme−substrate complexes. This results mostly from the lack of explicit solvent representation and the use of extended carbon atoms for representing nonpolar hydrogens. The excess van der Waals energy probably originated from neglecting the desolvation energy (removing water molecules from the S_1' binding site) and/or overlooking the steric effects of nonpolar hydrogen atoms which were assimilated to extended carbon atoms. Similarly, for electrostatic interactions, the dielectric constant of 20 used did not entirely compensate for the shielding effects associated with solvation of polar residues indicating that it should be raised. Therefore, the van der Waals and electrostatic (including H-bonds) terms were independently adjusted to maximize the correlation between calculated and experimental binding energies for CPD-Y, the calibration model. We found that scaling both of these non-bond terms by a single weight factor $\omega = 0.35$ yielded an overall correlation of $r = 0.92$ between the experimental and computed binding energy values with a slope of 0.97 (Figure 4a). Note that for electrostatic interactions, this is equivalent to setting the dielectric constant to 57 or using a scaling factor of 1/57. When this value for the non-bond scaling factor ω was applied to CPD-WII and CATH-A, a correlation coefficient of $r = 0.93$ with slopes of 0.99 and 1.20 was obtained for these two enzymes, respectively (Figure 4b and 4c). The excellent correlation found for all three enzyme models indicate that the individual energy terms used in the calculations were properly scaled and validate the use of this computational method to analyze the energetics involved in substrate binding. Interestingly, a similar relative weight factor (i.e., $\Delta E_{total} = \Delta E_{NB} + 2.5\Delta E_{Sol}$) was used by Schiffer *et al.* (1993) for protein structure predictions using a similar combination of solvation free energy−molecular mechanics force field simulation.

Using a "rigid body" representation of the enzyme may be an oversimplification of the S_1' binding site, since a slight structural change may have large effects on ΔE_{VDW} value. We verified this by artificially moving the two CATH-A S_1' residues with the largest ΔE_{VDW} values (Asp64 and Tyr147) with the Phe substrate before substrate modeling and binding energy computation. This did not significantly alter the ΔE_{VDW} values which were changed by −0.02 when Asp64 or by −0.03 when Tyr147 was moved by 0.5 Å.

We must also note the special case of branched aliphatic residues Val and Ile (Figure 4b) binding in CPD-WII S_1' site for which steric clashes with the C56−C303 disulfide bond imposed large conformational strains on these substrates that greatly increased their internal bonded energy. This result indicated that a rotation around the N-Cα-C-C_β dihedral angle may be necessary for optimal binding of these substrates. Indeed, a 9° rotation of the N-Cα-C-C_β dihedral angle away from the C56−C303 disulfide bond fitted Val and Ile to the energy correlation curve. However, since the binding conformation seems to differ for Ile and Val as compared to that of BZS, they were excluded from the correlation presented in Figure 4b.

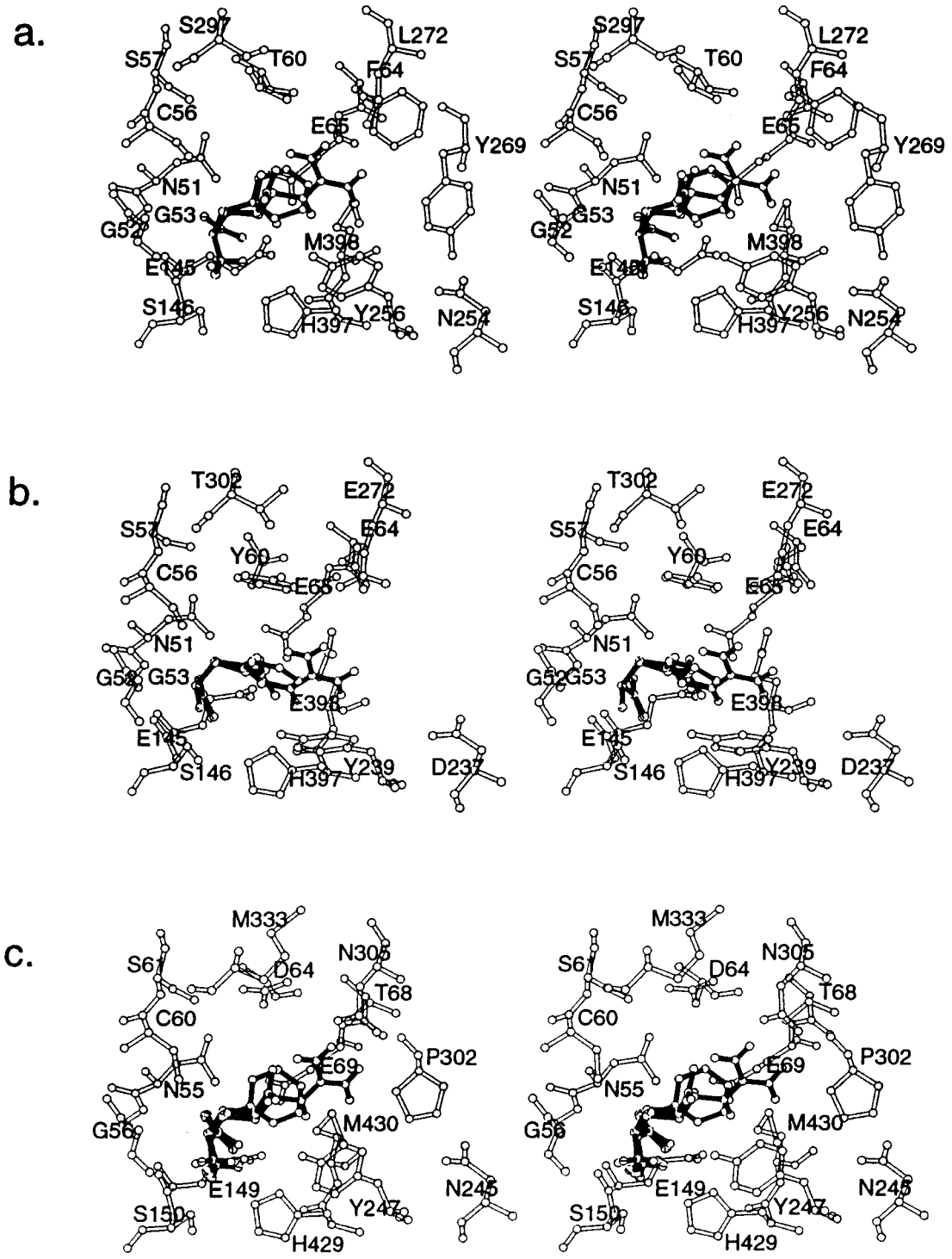

FIGURE 3: Stereoview of the modeled P_1' substrates (filled bonds) bound to the S_1' binding pockets (open bonds) of (a) CPD-Y, (b) CPD-WII, and (c) CATH-A.

Relative Importance of the Various Energy Terms Used in the Theoretical Binding Energy Calculations. We verified the relative importance of the individual energy terms used to calculate the binding energies by sequentially omitting them from the calculations (Table 4). No correlation was found between the experimental data and any of the individual (ΔG_{sol}, ΔE_{nb}, and ΔE_{bond}) or the coupled components ($\Delta E_{bond} + \Delta G_{Sol}$ and $\Delta E_{nb} + \Delta E_{bond}$) used in the theoretical calculations. Although the combined non-bonded and solvation ($\Delta E_{nb} + \Delta G_{Sol}$) contribution provided most of the information needed to estimate the substrate binding energy, when the ΔE_{int} term which reflects the internal molecular strain associated with fitting the substrate into the binding site was excluded, the correlation coefficient dropped below 0.9 for both CPD-Y and CPD-WII. This latter term did not have much influence on the CATH-A correlation coefficient, but it did greatly influence the slope of the correlation line. Clearly, all of the energy terms used in our calculations were necessary to correlate the computed values with the experimental data. Thus, the differences in the computed binding energies were not due to any one major energy term but were rather the consequence of the summation of a number of separate terms.

Intermolecular Forces Involved in Substrate Binding to the S_1' Subsites. One advantage of the binding energy calculation method presented here is its capacity to separate

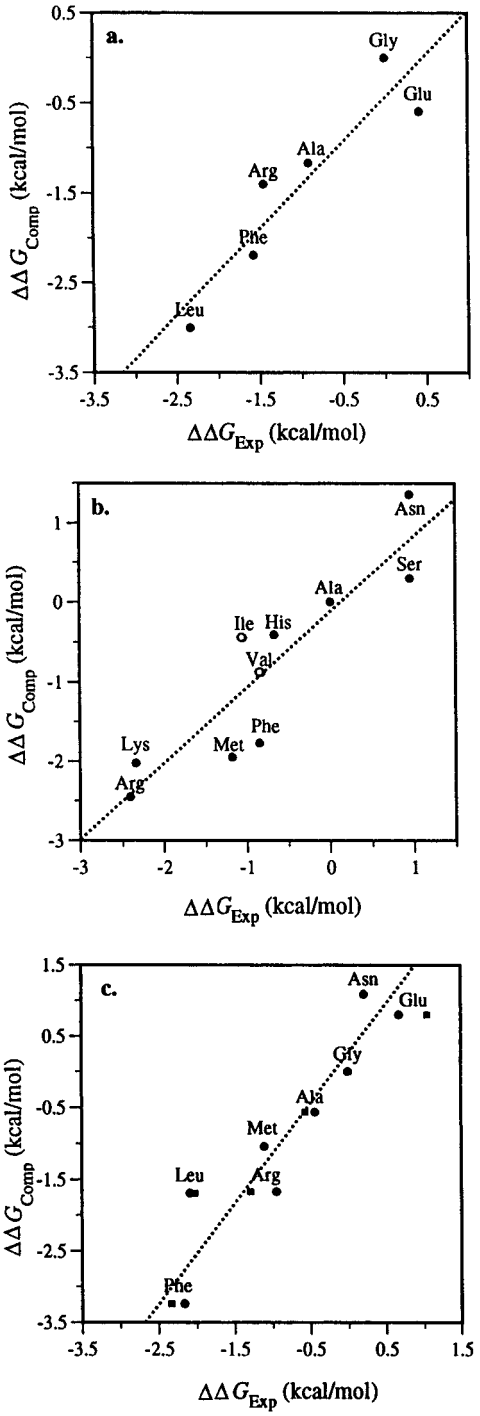

FIGURE 4: Experimental versus computed S_1' binding energies for the enzyme-substrate complexes listed in Tables 2 and 3. Computed binding energies are shown as a function of the experimentally determiend binding energies calculated for (a) CPD-Y (dashed line is the least-squares best fit line; computed = 0.96 × experimetal − 0.45; $r = 0.92$), (b) CPD-WII (computed = 0.99 × experimental − 0.20; $r = 0.93$) and (c) CATH-A (computed = 1.27 × experimental − 0.18; $r = 0.93$). For CATH-A (c) experimental values were calculated from both the K_i (●) and K_M (■) values.

the contributions of individual binding site residues giving further insight into the molecular properties of the enzyme−substrate interactions. The binding energy contributions for each enzyme−substrate complex are plotted on a residue by residue basis in Figure 5.

The P_1' substrates used in this analysis included Ala (small hydrophobic), Leu (large hydrophobic), Phe (aromatic, large hydrophobic), Arg (polar, positively charged), and Glu (polar, negatively charged), a series which probes a variety of potential interactions that might be favorable and unfavorable for binding in the S_1' subsite. The net energy difference is a balance between attractive (negative values) and repulsive (positive values) contributions. The residues that contribute to the binding energy can be subdivided into two groups: (1) those that make similar energy contributions for all enzyme−substrate complexes (indicated by an asterisk) and (2) those that discriminate between different P_1' substrates (indicated in bold in Figure 5). A breakdown of the individual components that contribute to the net binding energy calculations (Tables 2 and 3) for each enzyme−substrate complex is presented in Table 5.

Table 4: Importance of the Different Energy Terms Used in Equation 3

	CPD-Y		CATH-A		CPD-WII	
components	r	slope	r	slope	r	slope
ΔE_{comp}	0.92	0.96	0.93	1.27	0.93	0.99
$\Delta E_{nb}{}^a + \Delta E_{sol}$	0.81	1.68	0.93	1.85	0.88	1.39
$\Delta E_{bond} + \Delta E_{sol}$	0.54	0.17	0.40	0.46	0.40	−0.50
$\Delta E_{nb}{}^a + \Delta E_{bond}$	0.33	0.09	0.62	0.23	0.81	1.13
ΔE_{sol}	0.62	0.88	0.77	1.04	0.40	−1.18
$\Delta E_{nb}{}^a$	0.66	0.80	0.56	0.81	0.50	−0.87
ΔE_{bond}	0.57	−0.71	0.58	−0.57	0.50	−0.38

[a] $\Delta E_{nb} = (\Delta E_{Elect} + \Delta E_{HB} + \Delta E_{VDW})$.

One may argue that there is some overlap between the non-bonded energies computed with the CHARMm forcefield and the Eisenberg−McLachlan solvation free energy model. In fact some energy parameters may be computed twice with these different methods. This explains the rather small ω scaling factor needed in eq 3 to empirically balance these two complementary computational methods and obtain reasonable energy parameters. In fact both the CHARMm force field and the Eisenberg−McLachlan models alone were insufficient to predict enzyme−substrate interactions with all three carboxypeptidases studied. This is primarly due to the lack of explicit solvent and/or hydrophobic representation in the case of the molecular mechanic calculations done with the CHARMm force field. As for the Eisenberg−MacLachlan model it does not take into consideration the difference between transfert of a solute to an organic solvent and that of transfering a solute (substrate) to a protein site, which is a more rigid and heterogeneous environment. Therefore this model must be supplemented with additional energy parameters such as those of the CHARMm forcefield.

CPD-Y S_1'. The energy profile presented in Figure 5a shows that Cys56, Thr60, Tyr256, Leu272, and Met398 contribute indiscriminately to the bulk of the binding energy for all substrates, with the exception of Leu272 which does not contribute to the binding of Ala. Discrimination between positively charged (Arg) and negatively charged (Glu) substrates comes from the two acidic active site residues Glu65 and Glu145. These residues are primarily involved in securing the substrate C-terminal carboxylate through hydrogen bonding, which indicates that one or both of these residues must be protonated (Mortensen *et al.*, 1994). This was not taken into consideration in our calculations and could account for some of the excess electrostatic energy. With the exception of Met398 sulfur atom, there are no available hydrogen bond donors or acceptors within reach of the substrate. Moreover, in CPD-Y the presence of Thr60 in place of Tyr residue in CPD-WII (compare Figure 3a with

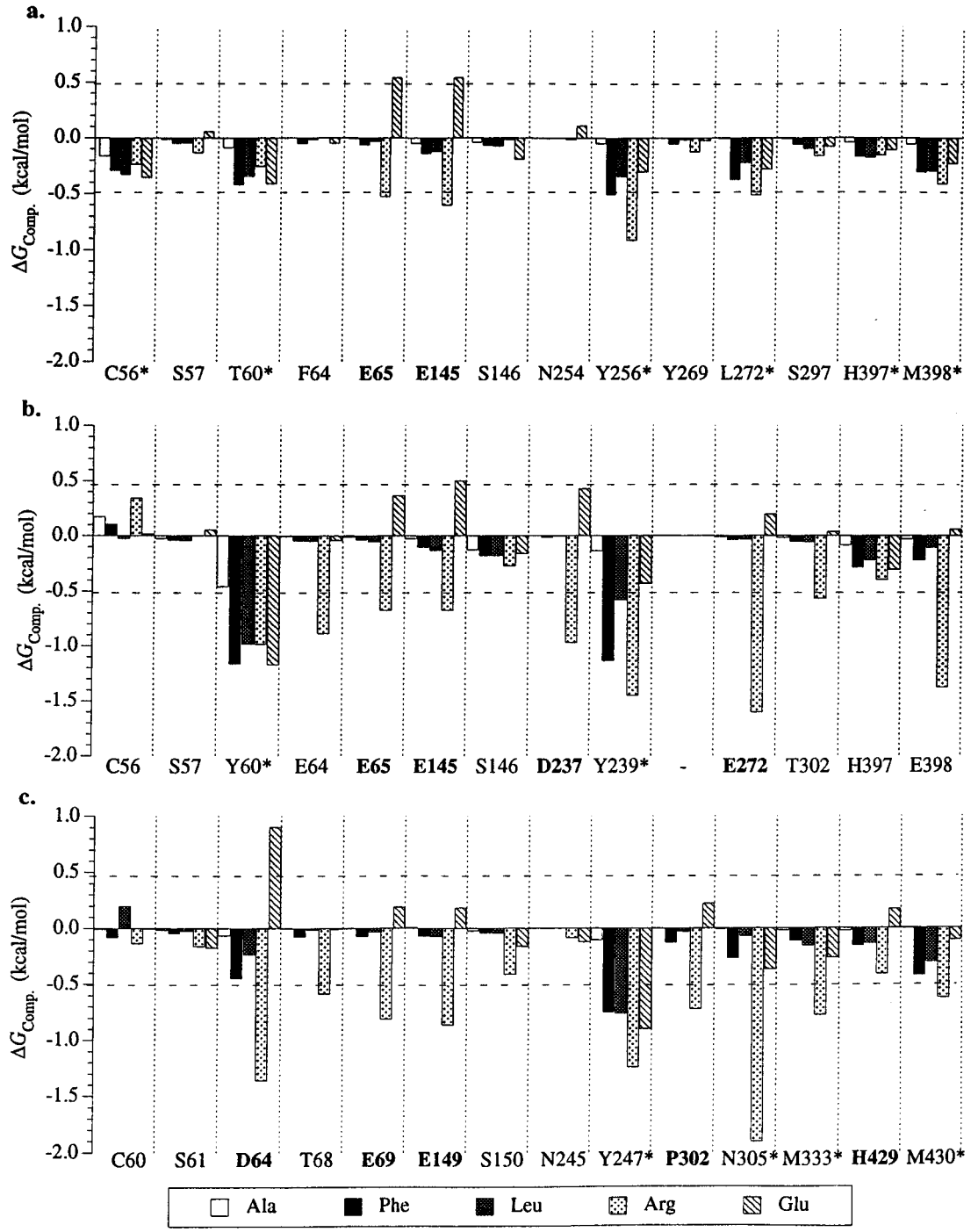

FIGURE 5: Energy profiles showing the residue contributions to the binding energy for the enzyme−substrate complexes with Ala (□), Phe (■), Leu (shaded bars), Arg (dotted bars) and Glu (cross-hatched bars) for (a) CPD-Y, (b) CPD-WII, and (c) CATH-A.

3b) results in a comparatively larger S_1' binding site (Endrizzi *et al*., 1994). This is reflected by the lower overall energy profile (Figure 5a) due to the greater intermolecular distance between the substrate and the enzyme active site residues. Solvation and van der Waals energies account for most of the substrate P_1' side chain binding energy (Table 5). Thus, the hydrophobic nature of the CPD-Y S_1' subsite provides a "greasy pit" that preferentially binds hydrophobic side chains.

Our results are consistent with site-directed mutagenesis studies of selected CPD-Y S_1' residues (Stennicke *et al*., 1994). These authors showed that the substitution of Thr60, Leu272, and Met398 residues by acidic Asp and/or Glu residues resulted in a general drop of enzymatic activity, confirming the implication of these residues in substrate binding and/or catalysis. Moreover, this study also showed that substitution of Phe64 and Ser297 had little or no effect on enzyme activity, and that substitution of Leu272 had no effect on FA-Phe-Ala hydrolysis. These latter findings can be clearly predicted from the CPD-Y energy profile presented in Figure 5a. Recent studies also indicated some mobility in the Met398 side chain, this was not taken into consideration in our calculation (Sørensen *et al*., 1995).

CPD-WII S_1'. The S_1' binding site of CPD-WII is made of a hydrophobic groove formed by Tyr60 and Tyr239 which contributes indiscriminately to the binding of all substrates, and a negatively charged pocket, formed by Glu272, Asp237, and Glu398, which are responsible for discriminating between charged P_1' residues (Figure 3b and 5b). The

Table 5: Contribution of the Individual Energy Terms to the Computed S_1' Binding Energy Calculations (in kcal/mol)

	CPD-Y					CATH-A					CPD-WII				
P_1'	$\Delta E_{Elect} + \Delta E_{HB}$	ΔE_{WDV}	ΔG_{Sol}	ΔE_{bind}	ΔG_{tot}	$\Delta E_{Elect} + \Delta E_{HB}$	ΔE_{WDV}	ΔG_{Sol}	ΔE_{bind}	ΔG_{tot}	$\Delta E_{Elect} + \Delta E_{HB}$	ΔE_{WDV}	ΔG_{Sol}	ΔE_{bind}	ΔG_{tot}
Gly	0.00	0.00	−1.30	0.00	−1.30	0.00	0.00	−1.70	0.00	−1.70					
Ala	0.00	−0.74	−1.99	0.25	−2.47	0.00	−0.52	−2.33	0.61	−2.23	0.00	−0.65	−2.41	0.32	−2.74
Phe	0.00	−2.58	−4.38	3.46	−3.50	0.00	−3.02	−4.57	2.68	−4.90	0.00	−2.75	−4.04	3.23	−3.56
Arg	−0.33	−3.09	−0.82	1.54	−2.71	−1.55	−2.78	−0.80	1.79	−3.33	−4.50	−3.42	−0.64	3.09	−5.47
Leu	0.00	−2.07	−3.94	1.70	−4.31	0.00	−1.77	−4.01	2.43	−3.35					
Glu	0.63	−1.83	−1.64	0.94	−1.90	1.55	−1.75	−1.59	0.93	−0.86					
Asn						−0.17	−1.35	−1.98	2.63	−0.57	−0.57	−2.01	−1.60	2.72	−1.46
Met						−0.04	−0.89	−4.02	2.17	−2.70	0.04	−2.09	−4.06	2.09	−4.02
His											−0.05	−2.52	−2.76	2.85	−2.48
Lys											−3.36	−2.90	−1.53	3.25	−4.54
Ile											0.00	−1.99	−3.82	2.62	−3.19
Ser											0.04	−0.83	−2.01	0.33	−2.30
Val											0.00	−1.34	−3.37	1.09	−3.62

hydrophobic cleft formed by Tyr239 and Tyr60 accommodates the alkyl part of the P_1' side chains, while the acidic pocket interacts with the terminal charges of longer side chains of basic structures but not with the shorter side chains of hydrophobic substrates (Liao *et al*., 1992). This is clearly visible in Figure 5b. Therefore, the bipartite subsite of CPD-WII S_1' is better adapted for binding substrates with basic P_1' residues than that of CPD-Y.

CATH-A S_1'. The energy profile of substrate binding to CATH-A S_1' subsite (Figure 3c and 5c) shows that Tyr247 provides the majority of the indiscriminate binding energy, with some contributions from Met430 and Asn305. Tyr247, which is conserved in all known serine carboxypeptidase sequences (Tyr256 in CPD-Y and Tyr239 in CPD-WII), not only shields the active site His residue from solvent (Liao *et al*., 1992) but also greatly contributes to substrate binding for the three enzymes studied. Discrimination between acidic and basic substrates comes mostly from Asp64 which reduces the binding capacity of the negatively charged Glu residues in P_1' and promotes binding of substrates with positively charged residues in the P_1' position such as Arg. These results also demonstrate the importance of Asn305 which greatly increases the binding energy of Arg by the formation of a hydrogen bond.

Finally, our particular interest was to provide a rational interpretation of the experimental results on the substrate specificity of the human lysosomal CATH-A, in order to gain further insight into the physiological role of its catalytic activity. Indeed, our understanding of CATH-A substrate specificity was greatly enhanced by incorporating information from the theoretical approach. We determined the different energy contributions to substrate binding within an enzyme active site, on a residue per residue basis. This provided a better picture of the molecular interactions involved in enzyme−substrate complex formation which helps to interpret the discriminating mechanism of substrate specificity. Comparison of energy profiles of CATH-A with those of the yeast and wheat enzymes showed a close similarity between the CPD-Y and CATH-A S_1' binding sites which rely mostly on hydrophobic interactions for substrate binding, as opposed to CPD-WII which also relies on polar interactions. However, the S_1' binding site of CATH-A has some of the polar characteristics of CPD-WII. We can therefore conclude that the S_1' binding specificity of CATH-A is quite broad at pH 6.0, at least toward dipeptide substrates, but that this enzyme does preferentially bind substrates with large hydrophobic (Phe and Leu) and positively charged (Arg) residues in the P_1' position, which closely resembles the substrate specificity of CPD-Y, with the exception that CATH-A binds P_1' Phe residues more strongly than Leu residues.

Concluding Remarks

Although "microscopic" approaches for protein−ligand binding free energy calculations, where solvent and atomic architecture of protein and ligand are explicitly used in the simulation, have also been proposed (Lee *et al*., 1992; Åqvist *et al*., 1994), the results of our study indicate that macroscopic approaches which are less computer intensive could yield reasonable results. The detailed analysis of substrate−active site residue interactions could help to design the site-directed mutagenesis experiments aimed at modifying either binding affinity or the substrate specificity of carboxypeptidases.

ACKNOWLEDGMENT

The authors thank Dr. S. James Remington and his team (Institute of Molecular Biology, University of Oregon, Eugene, OR) for graciously providing the CPD-WII and CPD-Y coordinates before publication. The assistance of M. Patenaude in the preparation of this manuscript is greatly appreciated.

REFERENCES

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

Bullock, T. L., Branchaud, B., & Remington, S. J. (1994) *Biochemistry 33*, 11127−11134.

Breddam, K. (1986) *Carlsberg Res. Commun. 51*, 83−128.

Breddam, K., Sørensen, S. B., & Svendsen, I. (1987) *Carlsberg Res. Commun. 52*, 297−311.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

d'Azzo, A., Hoogeveen, A. T., Reuser, A. J. J., Robinson, D., & Galjaard, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4535−4539.

Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199−203.

Elsliger, M.-A., & Potier, M. (1994) *Proteins 18*, 81−93.

Endrizzi, J. A., Breddam, K., & Remington, S. J. (1994) *Biochemistry 33*, 11106−11120.

Fukuhara, Y., Takano, T., Shimmoto, M., Oshima, A., Takeda, E., Kuroda, Y., Sakuraba, H., & Suzuki, Y. (1992) *Brain Dysfunct. 5*, 319−325.

Galjart, N. J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E. J., & d'Azzo, A. (1991), *J. Biol. Chem. 266*, 14754−14762.
Hayashi, R., Bai, Y., & Hata, T. (1975) *J. Biochem. 77*, 69−79.
Hoogeveen, A., d'Azzo, A., Brossmer, R., & Galjaard, H. (1981) *Biochem. Biophys. Res. Commun. 103*, 292−299.
Itoh, K., Kase, R., Shimmoto, M., Satake, H., & Suzuki, Y. (1995) *J. Biol. Chem. 270*, 515−518.
Jackman, H. L., Tan, F. L., Tamei, H., Beurling-Harbury, C., Li, X. Y., Skidgel, R. A., & Erdos, E. G. (1990) *J. Biol. Chem. 265*, 11265−11272.
Jackman, H. L., Morris, P. W., Deddish, P. A., Skidgel, R. A., & Erdos, E. G. (1992) *J. Biol. Chem. 267*, 2872−2875.
Kawamura, Y., Matoba, T., Hata, T., & Doi, E. (1976) *J. Biochem. 81*, 435−44.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Lee, F. S., Chu, Z.-T., Bolger, M. B., & Washel, A. (1992) *Protein Eng. 5*, 215−228.
Liao, D. I., & Remington, S. J. (1990) *J. Biol. Chem. 265*, 6528−6531.
Liao, D. I., Breddam, K., Sweet, R. M., Bullock, T., & Remington, S. J. (1992) *Biochemsitry 31*, 9796−9812.
Marks, N., Sachs, L., & Stern, F. (1981) *Peptides 2*, 159−164.
Matsuda, K. (1976) *J. Biochem. 80*, 659−669.
Miller, J. J., Changaris, D. G., & Levy, R. S. (1988) *Biochem. Biophys. Res. Commun. 154*, 1122−1129.
Mortensen, U. H., Remington, S. J., & Breddam, K. (1994) *Biochemistry 33*, 508−517.
Novotny, J., Bruccoleri, R. E., & Saul, F. A. (1989) *Biochemistry 28*, 4735−4749.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. L., Silman, J., Sussman, J. L., Verschueren, K. H. G., & Goldman, A. (1992) *Prot. Eng. 5*, 197−211.
Pshezhetsky, A. V., & Potier, M. (1994) *Arch. Biochem. Biophys. 313*, 64−70.
Pshezhetsky, A. V., Vinogradova, M. V., Elsliger, M.-A., El-Zein, F., Svedas, V. K., & Potier, M. (1995) *Anal. Biochem. 230*, 303−307.
Remington, S. J., & Breddam, K. (1994) *Methods Enzymol. 244*, 231−248.
Rick W. (1965) In *Methods of Enzymatic Analysis* (Bergmeyer, H. U., Ed.) pp 800−806, Academic Press, New York.
Rudenko, G., Bonten, E., d'sAzzo, A., & Hol, W. G. J. (1995) *Structure 3*, 1249−1259.
Schecter, I., & Berger, B. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
Schiffer, C. A., Caldweel, J. W., Stroud, R. M., & Kollman, P. A. (1992) *Protein Sci. 1*, 396−400.
Shimmoto, M., Takano, T., Fukuhara, Y., Oshima, A., Sakuraba, H., & Suzuki, Y. (1990) *Proc. Jpn. Acad. 66B*, 217−222.
Shimmoto, M., Fukuhara, Y., Itoh, K., Oshima, A., Sakuraba, H., & Suzuki, Y. (1993) *J. Clin. Invest. 91*, 2393−2398.
Sørensen, S. B., Raaschou-Nielsen, M., Mortensen, U. H., Remington, S. J., & Breddam, K. (1995) *J. Am. Chem. Soc. 117*, 5944−5950.
Stennicke, H. R., Mortensen, U. H., Christensen, U., Remington, S. J., & Breddam, K. (1994) *Protein Eng. 7*, 911−916.
Verheijen, F. W., Brossmer, R., & Galjaard, H. (1982) *Biochem. Biophys. Res. Commun. 108*, 868−875.
Verheijen, F. W., Palmeri, S., Hoogeveen, A. T., & Galjaard, H. (1985) *Eur. J. Biochem. 149*, 315−321.
Warshel, A. (1991) *Annu. Rev. Biophys. Chem. 20*, 267−298.
Wilson, C., Mace, J. E., & Agard, D. A. (1991) *J. Mol. Biol. 220*, 495−506.
Zhou, X. Y., Galjart, N. J., Willemsen, R., Gillemans, N., Galjaard, H., & d'Azzo, A. (1991) *EMBO J. 10*, 4041−4048.

BI952833L