

nitude of the  $k_d$  for Ia. The inclusion of thermal spin-state equilibria would not seem to be necessary to account for the factor of  $10^3$  in dissociation rates with a factor of 50 due to the protein environment and another factor of 20 originating in small variation in base strength and  $\pi$  donor/acceptor properties of different ligands. There are solvation effects originating from different experimental media which were used to deduce these trends. The effects of solvation on the entropy of the reactions are able to contribute a factor of 10 in rates.<sup>36</sup>

Fe(IV) intermediates have been postulated in the reactivity of horse-radish peroxidase<sup>37</sup> and the cytochrome system.<sup>38</sup> Our observations also support a possible role for two-electron events for low-spin iron centers either by two-electron changes in the metal center itself or with concerted ligand oxidation as an accessible chemical pathway for biological activity.

**Acknowledgment.** We thank Maynard Metals Co., Inc., for funds that enabled us to purchase and assemble our cyclic voltammetry unit.

## References and Notes

- C. K. Chang and T. G. Traylor, *J. Am. Chem. Soc.*, **95**, 8477 (1973).
- J. H. Wang, *Acc. Chem. Res.*, **3**, 90 (1970), and references therein.
- R. J. Sundberg, R. E. Shepherd, and H. Taube, *J. Am. Chem. Soc.*, **94**, 6558 (1972).
- R. J. Sundberg, R. F. Bryan, I. V. Taylor, and H. Taube, *J. Am. Chem. Soc.*, **96**, 381 (1974).
- R. J. Sundberg and G. Gupta, *Bioinorg. Chem.*, **3**, 39 (1973).
- H. E. Tomas and J. M. Malin, *Inorg. Chem.*, **12**, 1039 (1973).
- NMR Spectral Service, Department of Chemistry, Purdue University.
- H. Bredereck and G. Theilig, *Chem. Ber.*, **86**, 88 (1953).
- Microanalysis Laboratory, Department of Chemistry, Purdue University.
- W. U. Malik and M. Aslam, *Ind. J. Chem.*, **8**, 736 (1970); W. Malik and R. Bembli, *Proc. Int. Conf. Coord. Chem.*, **16th**, 1 (1974).
- G. Enschwiller, *C. R. Acad. Sci., Ser. C*, **259**, 4281 (1964).
- The <sup>1</sup>H NMR spectra of (CN)<sub>5</sub>Fe(pz)<sup>3-</sup>, pz = pyrazine, exhibits a triplet split into doublets with resonances centered at  $\delta$  9.19, 8.72, and 8.26 with a doublet coupling constant of 0.3 Hz. The free ligand resonance occurs at  $\delta$  8.66 (s). No evidence for a dimer containing two (CN)<sub>5</sub>Fe<sup>3-</sup> units was observed in the presence of excess free ligand.
- R. E. Shepherd and H. Taube, *Inorg. Chem.*, **12**, 1392 (1973).
- H. E. Tomas and J. M. Malin, *Inorg. Chem.*, **12**, 2081 (1973).
- J. N. Murrell, *Q. Rev., Chem. Soc.*, **191** (1961).
- P. George, G. I. Hanania, D. H. Irvine, and I. Abu-Issa, *J. Chem. Soc.*, 5689 (1964); P. Mohr, W. Scheler, H. Schumann, and K. Muller, *Eur. J. Biochem.*, **3**, 158 (1967).
- See ref 6; H. E. Tomas and J. M. Malin, *Inorg. Chem.*, **13**, 1772 (1974).
- R. Ford, D. F. P. Rudd, R. Gaunder, and H. Taube, *J. Am. Chem. Soc.*, **90**, 1187 (1968).
- C. Creutz and A. Zwickel, *Inorg. Chem.*, **10**, 2395 (1971).
- P. Haake, L. Bausher, and W. Miller, *J. Am. Chem. Soc.*, **91**, 1113 (1969).
- L. G. Silén and A. E. Martell, *Chem. Soc., Spec. Publ.*, **No. 17**, 1 (1974).
- M. F. Perutz, *Nature (London)*, **228**, 726 (1970).
- A. Calabrese and R. G. Hayes, *J. Am. Chem. Soc.*, **96**, 5051 (1974).
- E. Antonini and M. Brunori, "Hemoglobin and Myoglobin in Their Reactions with Ligands", North-Holland Publishing Co., Amsterdam, 1971, p 114.
- Q. Gibson, *Prog. Reac. Kinet.*, **2**, 231 (1964).
- D. A. Baldwin, R. M. Pfeiffer, D. W. Reichgott, and N. Rose, *J. Am. Chem. Soc.*, **95**, 5152 (1973).
- L. Vaska and T. Yamaji, *J. Am. Chem. Soc.*, **93**, 6673 (1971).
- D. V. Stynes, H. C. Stynes, B. R. James, and J. A. Ibers, *J. Am. Chem. Soc.*, **95**, 4087 (1973).
- G. N. LaMar and F. A. Walker, *J. Am. Chem. Soc.*, **95**, 1782 (1973).
- G. N. LaMar, J. D. Satterlee, and R. V. Snyder, *J. Am. Chem. Soc.*, **96**, 7137 (1974).
- L. M. Epstein, D. K. Straub, and C. Maricondi, *Inorg. Chem.*, **6**, 1720 (1967).
- R. E. Shepherd, unpublished results; other dipositive ions exhibit values greater than 10.8: Co(II) (12.5), Cu(II) (11.7), Pd(II) (10.8). See R. Sundberg and R. Martin, *Chem. Rev.*, **74**, 471 (1974).
- R. F. Fenske and R. L. DeKock, *Inorg. Chem.*, **11**, 437 (1972).
- R. L. DeKock, A. C. Sarapu, and R. F. Fenske, *Inorg. Chem.*, **10**, 38 (1971).
- G. M. Brown, F. R. Hopf, T. J. Meyer, and D. G. Whitten, *J. Am. Chem. Soc.*, **97**, 5385 (1975).
- S. J. Cole, G. C. Curthoys, and E. A. Magnusson, *J. Am. Chem. Soc.*, **92**, 2991 (1970).
- L. E. Bennett, *Prog. Inorg. Chem.*, **18**, 1 (1973), and references therein.
- D. Dolphin and R. H. Felton, *Acc. Chem. Res.*, **7**, 26 (1974).
- H. C. Brown and B. Kanner, *J. Am. Chem. Soc.*, **88**, 986 (1966).

# Electrostatic Potentials of Proteins. 1. Carboxypeptidase A

David M. Hayes and Peter A. Kollman\*

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143. Received May 14, 1975

**Abstract:** We present a new approach to examining the qualitative features of protein structure and small molecule-protein interactions. The electrostatic potential, which is a useful guide for small molecule interactions, is examined for carboxypeptidase A (CPA), its apo derivative, and for the CPA-Gly-Tyr complex. Ab initio wave functions are used to represent the individual residue partial charges. An *o*-OH-Gly-Tyr analogue is predicted to bind more strongly to CPA than the natural Gly-Tyr substrate.

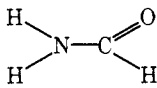
Since there are an ever increasing number of proteins for which high resolution x-ray data are becoming available, it is of interest to examine their electronic as well as conformational properties. In this report, we present a study of the electrostatic potential of carboxypeptidase A (CPA), one of the best resolved and characterized globular proteins. Quiocho and Lipscomb<sup>1</sup> have thoroughly discussed the structure and mechanism of action of CPA in light of x-ray structural studies.<sup>2</sup> One of our objectives in this study is to understand the important contributing factors to enzyme-small molecule binding. CPA is a good choice for these studies because its x-ray pattern is known to high resolution (2 Å) and because the structure of an enzyme-substrate complex (Gly-Tyr) of CPA has been determined.<sup>3</sup>

Briefly, CPA is an exopeptidase, i.e., it cleaves peptide bonds at the carboxyl terminal end of the peptide chain. It also has

catalytic activity toward esters and requires divalent metal for its catalytic activity (Zn<sup>2+</sup> in the native enzyme) and a pH for optimum catalytic activity of 7.5. In addition, upon binding of a substrate the enzyme undergoes a remarkable conformational change involving at least three side chains of CPA (Arg 145, Tyr 248, and Glu 270). This altered structure we will henceforth call "bound" CPA to distinguish it from the native or unbound enzyme. Many workers have probed the specificity of CPA by allowing it to react with a variety of different substrates.<sup>4</sup> These studies show that at least the five terminal peptides of the substrate influence the observed kinetic parameters.<sup>5</sup> This is consistent with the finding from the x-ray structure that the active site pocket (~18 Å in length) has room for at least five peptides.<sup>1,6</sup>

Since the publication of a paper by Bonaccorsi et al.,<sup>7</sup> there has been increasing interest in electrostatic potential maps as

Table I. Comparison of the Exactly Calculated Electrostatic Potential with the Monopole Approximation (4-31G Basis Set)

				
Mulliken populations		Electrostatic potential location	Monopole approximation	Exact
N	-0.888	2 Å from trans NH <sup>a</sup>	+0.038	+0.041
O	-0.633	2 Å from cis NH <sup>a</sup>	+0.015	+0.022
C	+0.570	2 Å from carbonyl O, trans to C-N <sup>b</sup>	-0.047	-0.059
H <sub>C</sub>	+0.191	2 Å from carbonyl O, cis to C-N <sup>b</sup>	-0.049	-0.044
H <sub>Nt</sub>	+0.370	2 Å from C-H <sup>a</sup>	+0.036	+0.020
H <sub>Nc</sub>	+0.400	2.1 Å above amide N <sup>c</sup>	-0.002	-0.001
		2.1 Å above carbonyl C <sup>c</sup>	+0.001	+0.004
		2.1 Å above carbonyl O <sup>c</sup>	-0.036	-0.032

<sup>a</sup> Along X-H line. <sup>b</sup> 60° from C=O line. <sup>c</sup> Above amide plane.

Table II. Comparison of STO-3G and 4-31G for *N*-Methylacetamide

Total energy, au	-243.84050	-246.62748
Calcd dipole moment (evaluating $\langle \mu \rangle$ , D)	2.62	4.01
Dipole moment calcd with Mulliken populations (D)	1.99	3.55
Exptl dipole moment		3.73 D
Atomic populations of N	-0.378	-0.887
C	+0.310	+0.746
O	-0.303	-0.640
NH	+0.201	+0.381

useful in determining sites of protonation, H-bond complex geometries, and the directionality of "donor-acceptor" interactions. We have shown that the nature of the electrostatic potential allows one to predict the directionality of H-bond structure and to rationalize H-bond energies in related systems.<sup>8</sup> Morokuma<sup>9</sup> and co-workers have come to similar conclusions in their study of carbonyl cyanide-ROR and excited state acrolein H-bond complexes.

The ultimate goal of our studies in this area is to see how effective just the electrostatic term in the intermolecular potential can be in predicting qualitative features of enzyme-substrate and enzyme-inhibitor interactions. Principles derived in such a study could be useful in drug design.

The electrostatic energy is only one of the components of the intermolecular potential, the others being exchange repulsion, polarization, charge transfer, and dispersion (in the gas phase).<sup>10</sup> In aqueous solution, an important component in intermolecular interactions is the hydrophobic effect, which has its basis mainly in entropic changes in the surrounding water.<sup>11</sup> It is the basic premise of this study that the polarization, charge transfer, and dispersion are much less important than electrostatic effects in determining structural features of complex formation. The exchange repulsion is taken into account by not allowing the "substrate" to penetrate within the van der Waals radius of the protein atoms; the hydrophobic effect constrains us to rigorously compare only substrates of similar hydrophobicity.<sup>12</sup>

There have been a number of applications of electrostatic potential maps to properties of small biologically interesting molecules, where the structure of the macromolecular "receptor" is not known;<sup>13</sup> here we look at interactions where the structure of both the macro and small molecules are known. In this way we hope to derive useful principles for biological applications of the electrostatic potential approach.

**Computational Details.** To construct the electrostatic potentials of carboxypeptidase A, we used the results of ab initio molecular orbital calculations<sup>14</sup> on the individual amino acid residues. The Mulliken net atomic charges obtained from these calculations (carried out with both STO-3G and 4-31G basis

sets) were used in a "monopole" approximation in order to evaluate the electrostatic potential surrounding the protein. Since the use of Mulliken populations to evaluate the electrostatic potential is a rather drastic approximation, we compared the electrostatic potential surrounding formamide calculated exactly from the wave function with that approximated with Mulliken populations (Table I). At distances from the molecule reasonable for intermolecular interactions (e.g., H bonding to the N-H and C=O or  $\pi$  interactions) the agreement is surprisingly good, although not all the trends are reproduced. We have used both STO-3G and 4-31G results in this study. Because these basis sets differ so greatly in the degree of bond polarity they predict (see Table II)<sup>15</sup> any result we derive which is consistent with both point charge choices is more likely to be correct than if we were to rely only on one set of point charges. Since the individual amino acids in a protein are not isolated entities but are bound at either end to other amino acids, we carried out our calculations on model compounds which took these "boundary conditions" into consideration. Thus, most of the charges were obtained from calculations on the following model compound, where R is the appropriate amino acid side chain: OHC-NH-CHR-CO-NH<sub>2</sub>. Since our computer program is limited to molecules with 70 atomic orbitals or less, we were forced in a few cases to use a simpler model compound. Thus, for methionine, arginine, lysine, histidine, glutamine, and phenylalanine we used the following model compound: NH<sub>2</sub>-CHR-CHO; for tyrosine and tryptophan we simply used CH<sub>3</sub>-R. The charges for the phenylalanine side chain were taken directly from the first calculation; for tyrosine the charges for CH<sub>2</sub>-Ph-OH were taken from the latter calculation, except for C <sub>$\alpha$</sub>  and C <sub>$\alpha$</sub> H which came from the first calculation on Phe. For all the peptide linkages, the populations were those found in *N*-methylacetamide. The point charges for the individual amino acids were adjusted so that each amino acid (-NH-CHR-CO-) had an integer charge (0, +1, or -1); this involved adding a small amount of positive or negative charge (usually  $\sim 0.05 e^-$ ) to the  $\alpha$  carbon and  $\alpha$  hydrogen in the ratio of 2:1. The role of C <sub>$\alpha$</sub>  as an insulator between the peptide backbone and the side chain is clearly shown in Tables III and IV. Thus, except for the easily understood cases of proline and glycine, changing the side chain has very little effect on the net atomic charges of the backbone atoms. Similarly, changing the model peptide from OHC-NH-CHR-CO-NH<sub>2</sub> to H<sub>2</sub>N-CHR-CHO has little effect on the side-chain charges. Thus, for alanine R = -CH<sub>3</sub>, the carbon and hydrogen charges in the two cases are (-0.1788, 0.0749, 0.0571, 0.0680) vs. (-0.1784, 0.0674, 0.0657, 0.0672). These populations are also very insensitive to changes in conformation as is clear from Table V and the last two entries of Table IV. Thus, it seems reasonable to use a set of point charges for a given amino acid which is invariant to conformational changes.

Table III. Mulliken Populations of Backbone Atoms in the Model Compounds OHC-NH-CHR-CO-NH<sub>2</sub> (STO-3G)

Residue = R	N	NH	C <sub>α</sub>	C <sub>α</sub> H	C	O
Asn	-0.3875	0.2008	0.0496	0.0917	0.3120	-0.2837
Asp	-0.3726	0.2078	0.0447	0.0786	0.3084	-0.3174
Thr	-0.3867	0.1980	0.0347	0.0854	0.3191	-0.2847
Glu	-0.3751	0.1975	0.0419	0.1039	0.3099	-0.3046
Pro	-0.3127		0.0263	0.0758	0.3054	-0.2902
Ser	-0.3849	0.1994	0.0421	0.0879	0.3186	-0.2838
Gly	-0.3807	0.2011	-0.0282	0.0849	0.3157	-0.2879
Ile	-0.3851	0.1992	0.0440	0.0767	0.3114	-0.2909
Leu	-0.3988	0.2289	0.0421	0.0766	0.3114	-0.2912
Val	-0.3851	0.1990	0.0424	0.0764	0.3110	-0.2913
Ala	-0.3843	0.1989	0.0479	0.0793	0.3117	-0.2899
Range	0.0861 <sup>a</sup> 0.0262 <sup>b</sup>	0.0314	0.0778 <sup>c</sup> 0.0233 <sup>d</sup>	0.0281	0.0137	0.0337

<sup>a</sup>Including proline. <sup>b</sup>Excluding proline. <sup>c</sup>Including glycine.

<sup>d</sup>Excluding glycine.

Table IV. Mulliken Populations of C<sub>α</sub> in the Amino Aldehydes H<sub>2</sub>N-CHR-CHO (STO-3G)

Residue = R	C <sub>α</sub>	C <sub>α</sub> H	N	C
Ala	0.0496	0.0635	-0.4430	0.1330
Met	0.0478	0.0620	-0.4447	0.1323
Arg	0.0489	0.0705	-0.4431	0.1314
Lys	0.0474	0.0680	-0.4428	0.1313
His	0.0442	0.0620	-0.4516	0.1332
Gln	0.0429	0.0767	-0.4467	0.1339
Phe <sup>a</sup>	0.0483	0.0643	-0.4468	0.1323
Phe <sup>b</sup>	0.0479	0.0633	-0.4442	0.1289
Range	0.0067	0.0147	0.0088	0.0050

<sup>a</sup>χ<sub>1</sub> = 300°; χ<sub>1</sub> is the dihedral angle formed by the four atoms: N-C<sub>α</sub>-C<sub>β</sub>-C<sub>γ</sub>. See ref 21 for sign convention. <sup>b</sup>χ<sub>1</sub> = 60°.

To construct the 4-31G point charge choice, we redid the calculations on *N*-methylacetamide (NMA) at the 4-31G level. We used these charges for the peptide linkage. The point charges for the charged groups (carboxylate, ammonium, and guanidinium) were determined by calculation of the appropriate model compounds: HCOO<sup>-</sup>, CH<sub>3</sub>NH<sub>3</sub><sup>+</sup>, CH<sub>3</sub>NHC(NH<sub>2</sub>)<sub>2</sub><sup>+</sup>. For the neutral amino acids, we scaled the STO-3G charges of the α carbon and its hydrogen and the side chain by 0.400/0.170 (the ratio of the net negative charge of the peptide linkage of NMA as calculated at the 4-31G and STO-3G levels). This automatically retains charge neutrality of the entire residue. The actual point charges used for the STO-3G and 4-31G calculations are available from the authors on request.

Coordinates for CPA and the Gly-Tyr substrate were taken from the published work of Lipscomb et al.<sup>1,3</sup> Hydrogens were placed on the heavy atoms of CPA and the substrate using standard bond lengths and angles.<sup>16</sup> The state of ionization of the amino acids was that found predominantly at neutral pH. The x-ray structure is ambiguous on the orientation of the glutamine and asparagine side chains (i.e., the assignment of the side-chain carbonyl oxygen and NH<sub>2</sub> groups) so we attempted to determine these on the basis of optimum H-bond contacts and minimum steric repulsion. Roughly 7 of the glutamines and asparagines appeared to have a clear preferential orientation; the others were assigned in two different ways.<sup>17</sup> Both choices gave essentially the same electrostatic potential at the active site and similar overall appearance for the entire protein's electrostatic potential (vide infra).

**Electrostatic Potential Maps of the Whole Protein.** We have evaluated the electrostatic potential of CPA in the *xy* plane (by constructing grids of 2 Å resolution) at different values of the *z* coordinate (see Figure 1 and the published coordinates<sup>1</sup>).

Table V. Mulliken Populations of Backbone Atoms in *N*-Methylacetamide as a Function of Methyl Torsion (STO-3G)

Conformation <sup>a</sup>	N	NH	C	O
Eclipsed	-0.3779	0.2010	0.3103	-0.3027
30°	-0.3779	0.2008	0.3087	-0.3005
Staggered	-0.3780	0.2006	0.3075	-0.2984

<sup>a</sup>The conformation refers to the dihedral angles that the *N*-methyl hydrogens make with the CH<sub>3</sub>CO-N bond.

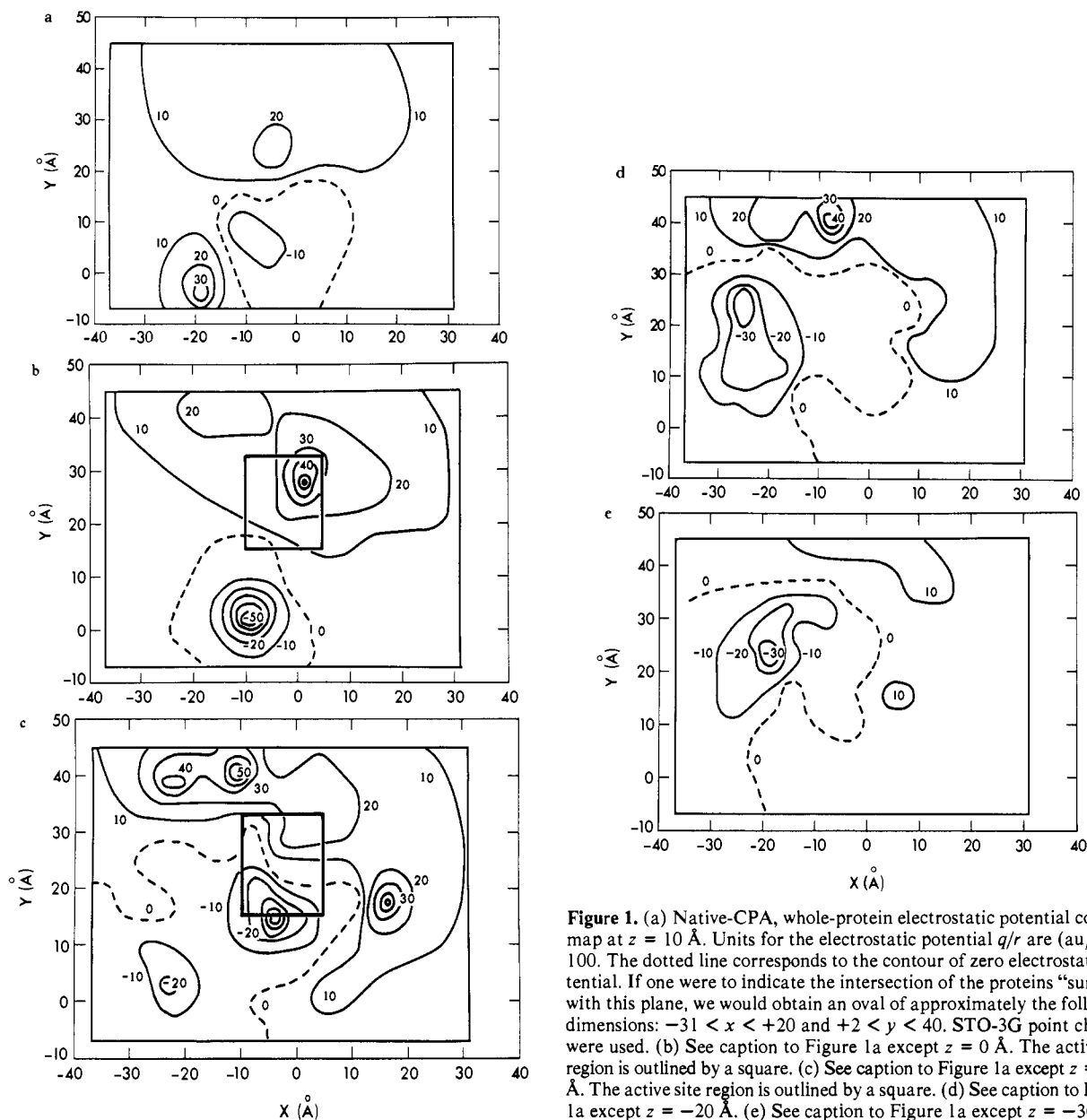
Table VI. Electrical Properties of CPA<sup>a</sup>

	STO-3G		4-31G	
	Native	Apo	Native	Apo
μ <sub>x</sub>	204	162	285	236
μ <sub>y</sub>	384	326	423	351
μ <sub>z</sub>	189	143	259	199
μ <sub>t</sub>	474	391	572	467
Q <sub>xx</sub>	4347	4167	5358	5068
Q <sub>yy</sub>	8307	7963	9061	8555
Q <sub>zz</sub>	1527	1307	2274	2042
Q <sub>xy</sub>	-920	-1168	-1134	-1542
Q <sub>xz</sub>	-3959	-4158	-4089	-4425
Q <sub>yz</sub>	1585	1310	1159	705

<sup>a</sup>Relative to center of mass = (-8.11, 20.48, -10.20) for native; (-8.12, 20.47, -10.20) of apo. Dipole moment in debyes and quadrupole moment in buckinghams.

We determined this electrostatic potential for *z* increments of 2.5 Å but the general shape and changes of the potential as one moves through the protein are evident from Figure 1, where the planar potential is described for *z* = 10, 0, -10, -20, and -30 Å. In order to help the reader orient the electrostatic potential contour maps relative to the residues in CPA, we present Figures 2-4. Figure 2 is a stereoview of the enzyme from the +*y* direction looking edge-on along the various plane corresponding to *z* = a constant. Note that the positive *z* direction points downward in this drawing. Figure 3 shows (in circles only) those residues which lie within 0.5 Å of each of the five planes (only the coordinates of the α carbon were used in deciding if a given residue lies near one of the planes). Thus, Figures 2 and 3 together should enable the reader to correctly place each of the contour maps of Figure 1. Figure 4 is a view of the enzyme from the +*z* direction. The protein has a net charge of +2 (including the zinc) which is reflected in the predominance of positive over negative regions. However, there are large regions of negative potential; for example, the lower left quadrant of Figure 1b reaches a minimum of -50.

It is reasonable to expect that the shape of the contour maps will be determined mainly by the number and kind (positive or negative) of charged residues nearby. Therefore, in Figure 3 we have shown (in boxes) all charged residues within 5 Å of each of the five planes (the coordinates of the charged portion of the residue were used in deciding if the residue lies near one of the planes). The position of these charged groups appears to account well for the gross features of the contour maps in Figure 1. Thus, the broad positive region at the top of Figure 1a can be attributed to Arg 124 and Arg 276. The narrow but more intensely positive region in the lower left corner is due to Arg 2. Similarly, Glu 17 and Glu 122 are responsible for the negative potential curves. The plane at *z* = 0 passes more nearly through the center of the enzyme and this is reflected in the greater number of entries in Figure 3b. The potential field shows greater variation (Figure 1b) than at *z* = 10 Å indicating a higher density of nearby charged groups. The highly positive region is due to Arg 71, Arg 127, and Lys 128 and the negative potential to aspartates 16, 20, and 23. The region of positive potential in Figure 1c is due to three lysines (231, 239,



**Figure 1.** (a) Native-CPA, whole-protein electrostatic potential contour map at  $z = 10 \text{ \AA}$ . Units for the electrostatic potential  $q/r$  are  $(\text{au}/\text{\AA}) \times 100$ . The dotted line corresponds to the contour of zero electrostatic potential. If one were to indicate the intersection of the proteins "surface" with this plane, we would obtain an oval of approximately the following dimensions:  $-31 < x < +20$  and  $+2 < y < 40$ . STO-3G point charges were used. (b) See caption to Figure 1a except  $z = 0 \text{ \AA}$ . The active site region is outlined by a square. (c) See caption to Figure 1a except  $z = -10 \text{ \AA}$ . The active site region is outlined by a square. (d) See caption to Figure 1a except  $z = -20 \text{ \AA}$ . (e) See caption to Figure 1a except  $z = -30 \text{ \AA}$ .

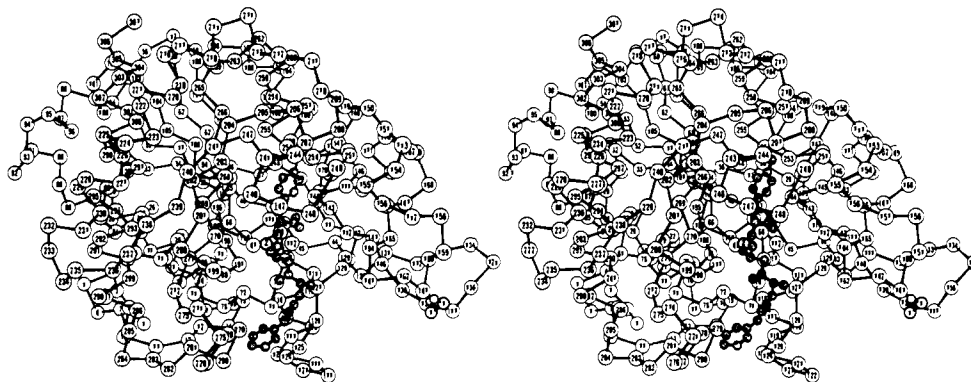
and 168) and to Arg 40 and 145 while the presence of Asp 65 and 142 and Glu 173, 175, 108, and 88 accounts nicely for the negative areas. Similarly, one can see that the positive regions of Figure 1d are the result of nearby lysines (224, 216, 153, 177, and 51) while the negative potential is due primarily to Glu 302 and aspartates 101, 104, 181, 148, and 256. Few charged residues lie near the  $z = -30$  plane and this is reflected in the shallowness of its contours.

To verify our assumption that the qualitative aspects of the contours in Figure 1 are determined by the residues with net positive or negative charge and not by the many fractional charges, we have recalculated the contour maps using only the charged residues. Figure 5 shows such a map for  $z = -10$  and its close similarity to Figure 1c is evident.

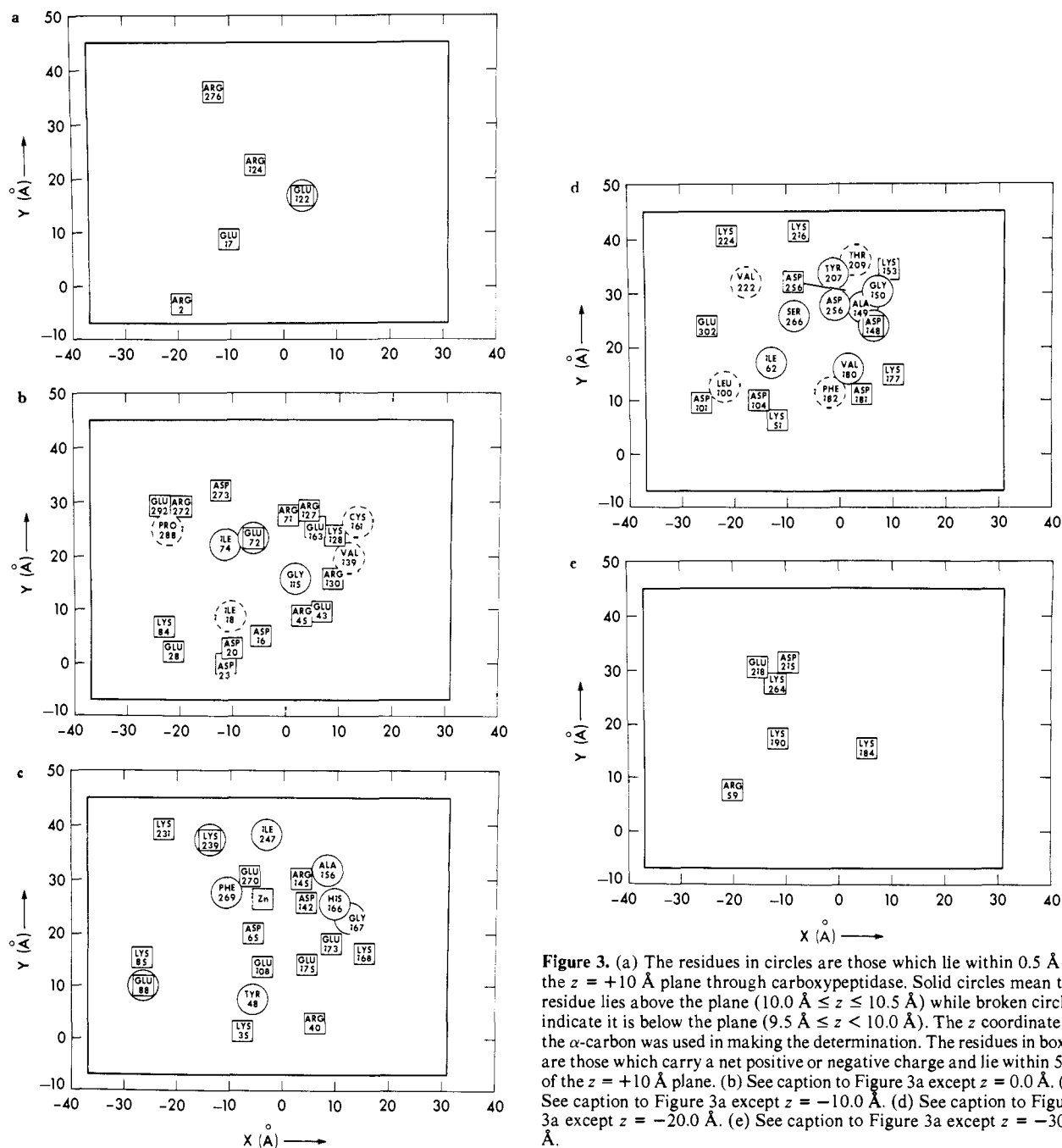
How does the point charge choice affect the electrostatic potential? We report the 4-31G electrostatic potential for  $z = -10 \text{ \AA}$  in Figure 6, which should be compared with Figure 1c. As one can see the shape of the potential and the boundaries between positive and negative regions are remarkably similar for the different point charge choices; this is also true at other  $z$  values. Thus, the electrostatic potential at  $2 \text{ \AA}$  resolution is surprisingly insensitive to point charge choice.

Since the apoenzyme ( $\text{Zn}^{2+}$  free) is found to bind substrate but to be catalytically inactive, we decided to compare its electrostatic potential to the native enzyme. As one might expect, the boundary between the positive and negative regions is shifted somewhat (of the order of  $5 \text{ \AA}$ ) to increase the regions of negative potential, but the general shape of the maps stays the same. Of interest is the fact that the site which the metal vacated is now a region of negative electrostatic potential; thus binding of the metal is favored from an electrostatic point of view.

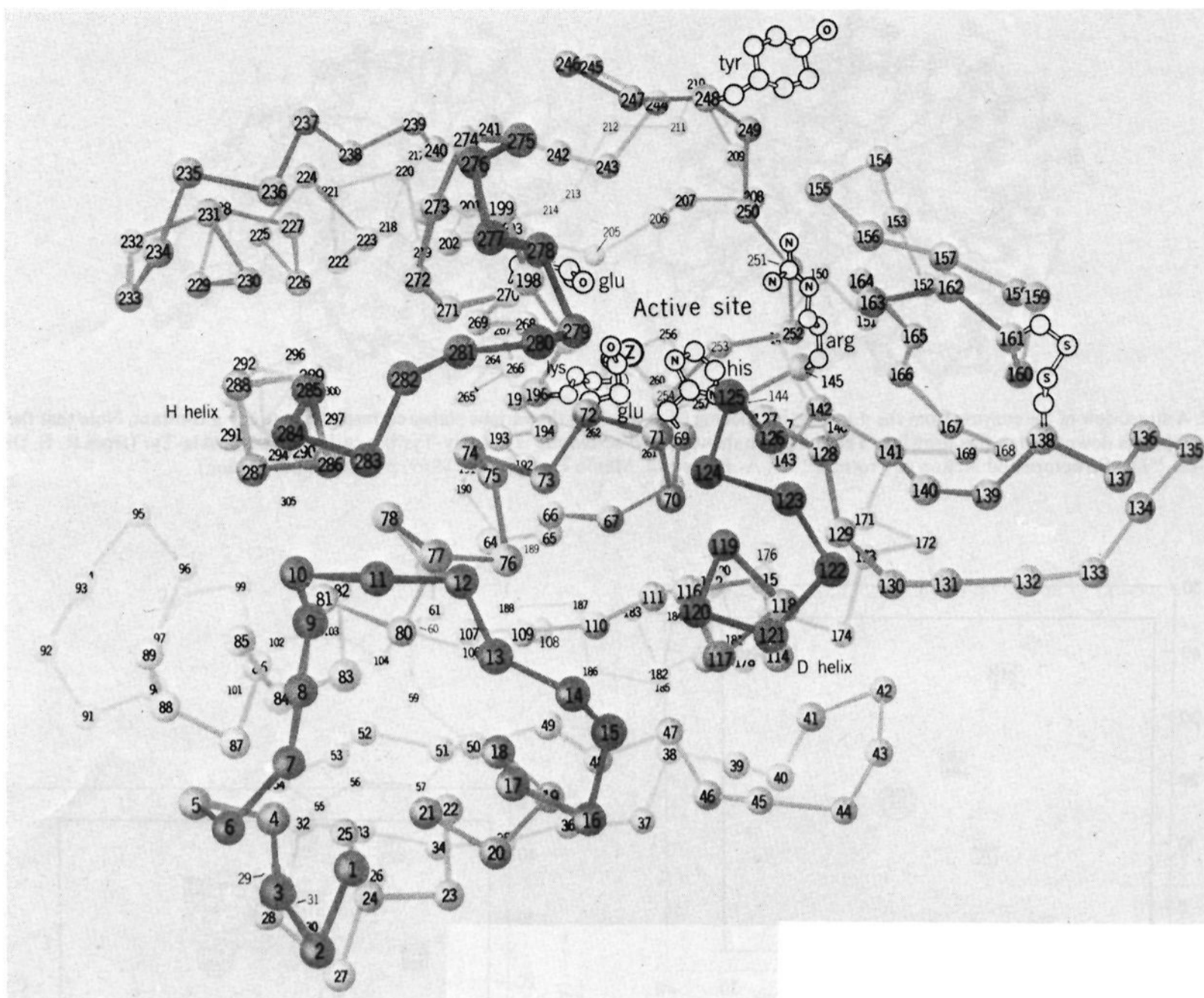
The strength of interactions between the protein and solvent and the ability of the protein to respond to external fields can be understood in terms of the molecule's electric moments. Therefore, we have computed the dipole and quadrupole moments of CPA and these are presented in Table VI for the native and apoenzyme. The moments are calculated relative to the center of mass of the protein (note that the dipole moment of the apoenzyme is origin independent because it has zero charge). The dipole moment of CPA has not yet been measured but that of horse methemoglobin is known to be  $500 \text{ D}$ ;<sup>18</sup> therefore our calculated dipole moment appears to be reasonable.



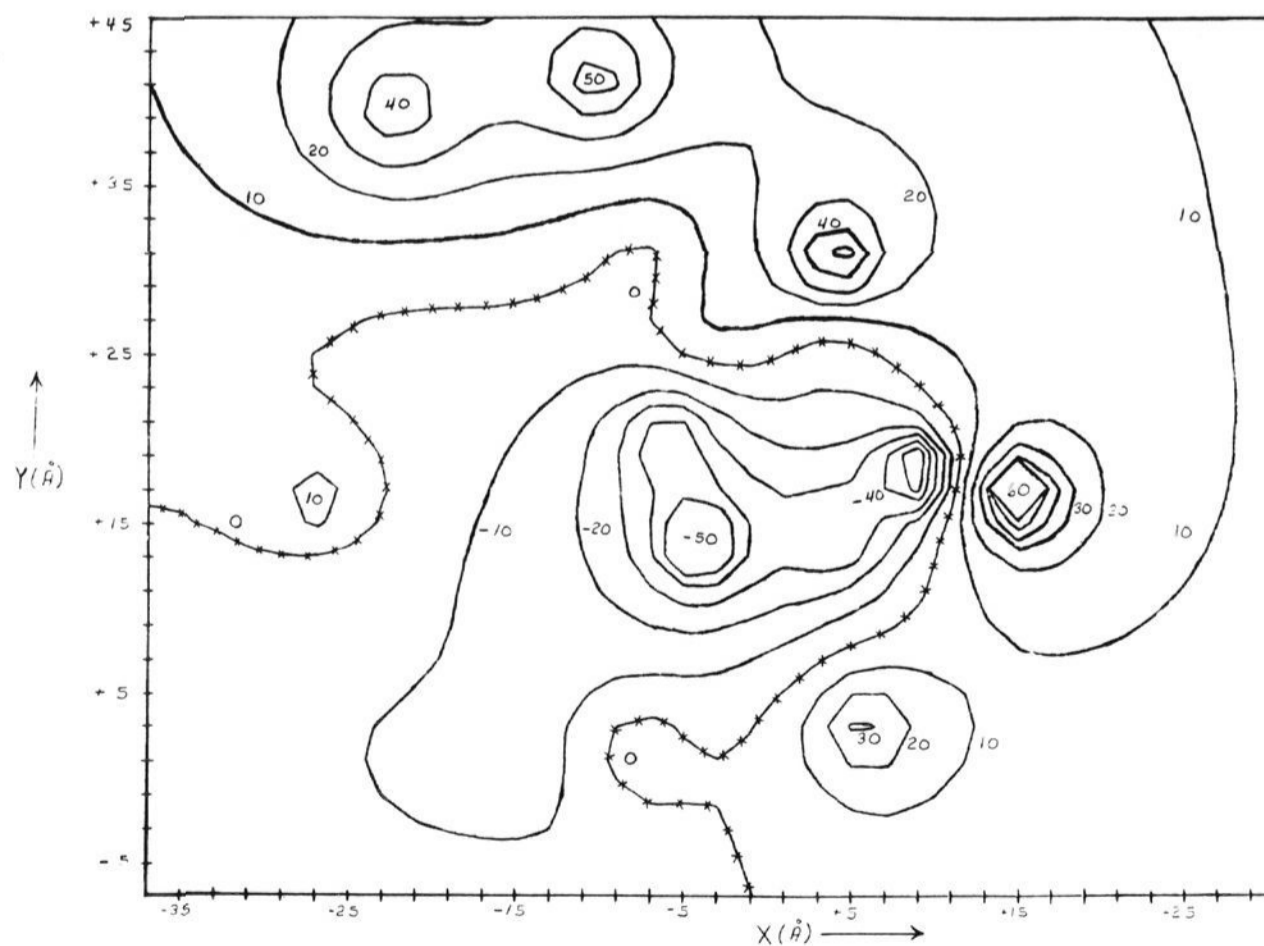
**Figure 2.** A stereoview of the enzyme from the  $+y$  direction looking edge-on along the various planes corresponding to  $z = a$  constant. Note that the positive  $z$  direction points downward in this drawing. The substrate shown in the active site is not Gly-Tyr but rather Cbz-Ala-Ala-Tyr (from R. E. Dickerson and I. Geis, "The Structure and Action of Proteins", W. A. Benjamin, Menlo Park, Calif., 1969, p 91, with permission).



**Figure 3.** (a) The residues in circles are those which lie within  $0.5 \text{ \AA}$  of the  $z = +10 \text{ \AA}$  plane through carboxypeptidase. Solid circles mean the residue lies above the plane ( $10.0 \text{ \AA} \leq z \leq 10.5 \text{ \AA}$ ) while broken circles indicate it is below the plane ( $9.5 \text{ \AA} \leq z < 10.0 \text{ \AA}$ ). The  $z$  coordinate of the  $\alpha$ -carbon was used in making the determination. The residues in boxes are those which carry a net positive or negative charge and lie within  $5 \text{ \AA}$  of the  $z = +10 \text{ \AA}$  plane. (b) See caption to Figure 3a except  $z = 0.0 \text{ \AA}$ . (c) See caption to Figure 3a except  $z = -10.0 \text{ \AA}$ . (d) See caption to Figure 3a except  $z = -20.0 \text{ \AA}$ . (e) See caption to Figure 3a except  $z = -30.0 \text{ \AA}$ .



**Figure 4.** A view of CPA looking toward the entrance to the active site from the  $+z$  direction (from R. E. Dickerson and I. Geis, "The Structure and Action of Proteins", W. A. Benjamin, Menlo Park, Calif., 1969, p 90, with permission).



**Figure 5.** Electrostatic potential contour map at  $z = -10.0 \text{ \AA}$  using only the carboxypeptidase residues with net positive or negative charge (Arg, Lys, Asp, Glu, and zinc).

Table VII. Coordinates of Gly-Tyr Substrate

Atom	x	y	z	
N	-4.43	29.94	-3.41	Fragment 1
H1	-5.08	30.30	-4.12	
H2	-4.63	30.38	-2.51	
H3	-4.53	28.93	-3.34	
CA	-2.66	30.36	-3.93	
CAH1	-2.49	31.43	-4.09	
CAH2	-1.90	29.97	-3.26	
C	-2.59	29.65	-5.24	Fragment 2
O	-2.66	28.45	-5.10	
N	-2.31	30.72	-6.22	
NH	-2.17	31.73	-6.25	
C	-0.73	29.94	-8.05	Fragment 3
O1	-0.41	30.96	-8.14	
O2	-0.02	28.57	-8.33	
CA	-2.22	29.83	-7.62	
CAH	-2.66	30.82	-7.54	Fragment 4
CB	-2.90	30.54	-8.70	
CBH1	-2.78	31.62	-8.55	
CBH2	-3.96	30.30	-8.69	
CG	-2.41	30.24	-10.06	
CD1	-2.04	31.14	-10.57	
CD1H	-2.00	32.10	-10.05	
CD2	-2.73	29.05	-10.76	
CD2H	-3.27	28.34	-10.13	
CE1	-1.66	30.84	-11.88	
CE1H	-1.01	31.53	-12.43	
CE2	-2.47	28.75	-11.98	
CE2H	-2.67	27.75	-12.35	
CZ	-1.89	29.77	-12.86	
O	-1.62	29.47	-14.08	
OH	-1.11	30.17	-14.48	

If one leaves out the 52 charged residues and zinc and calculates the dipole moment for the remaining 255 residues, one finds a great decrease in the dipole [ $\mu = (67, 3, 75)$  STO-3G and  $\mu = (124, -5, 97)$  4-31G] and this remaining dipole moment is almost exclusively due to the  $\alpha$  helices, which pack like cylinders<sup>6</sup> in a box and have a net polarity in the (+x, +z) direction. From the above it is clear that the overall electrical properties of the molecule are dominated by the charged groups. The apoenzyme is neutral, with 26 positive and an equal number of negative charges. The helices have a predominance of negative charge, with the first helix, from residues 14–28 having five anionic and no cationic residues. A local concentration of positive charge (four cationic and no anionic) is found in the random coil from residues 123–131. The positive region is near the entrance to the active site (see Figure 4) and one might speculate that this charge distribution helps to “steer” the negative end of the substrate into the catalytic site. The specificity of CPA for large hydrophobic terminal residues can be rationalized in the following way: the COO<sup>-</sup> termini of all peptides are attracted to the entrance of the active site but only those which have a sufficiently negative free energy of transfer make their way into the active site pocket. Note,

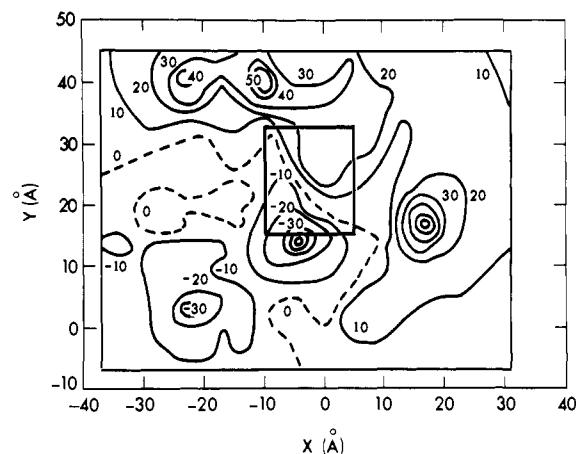


Figure 6. See caption to Figure 1c except that 4-31G point charges were used in constructing this map.

of course, that there are other secondary binding sites for anions on the surface of CPA (see, for example,  $z = -10$ , upper left-hand corner, Figure 1c, due to lysines 231 and 239) and these might also bind “hydrophilic” anions strongly.

**Electrostatic Potential Maps of the Active Site.** On the whole protein potential maps for  $z = 0$  and  $-10$  Å, the active site region is enclosed by solid lines and we now take a “high resolution” look at the electrostatic potential in this region. Figures 7 and 8 contain electrostatic potentials of the active site region of native and apo-CPA at 0.5 Å resolution, with van der Waals forbidden regions blocked out. Some of the key enzyme residues, such as Glu 270, Arg 145, Tyr 248, and Zn<sup>2+</sup>, as well as certain substrate atoms are indicated in the figures. In addition, Figure 9 contains a representation of the Gly-Tyr substrate atoms and their location in the active site and Table VII contains the Gly-Tyr coordinates. Thus, by interpolating between different  $z$  values, one can determine the electrostatic potential at the individual substrate atoms.

Figure 10 is a representation of the same region as Figure 7c, but now using 4-31G charges. The close similarity between the regions is obvious.

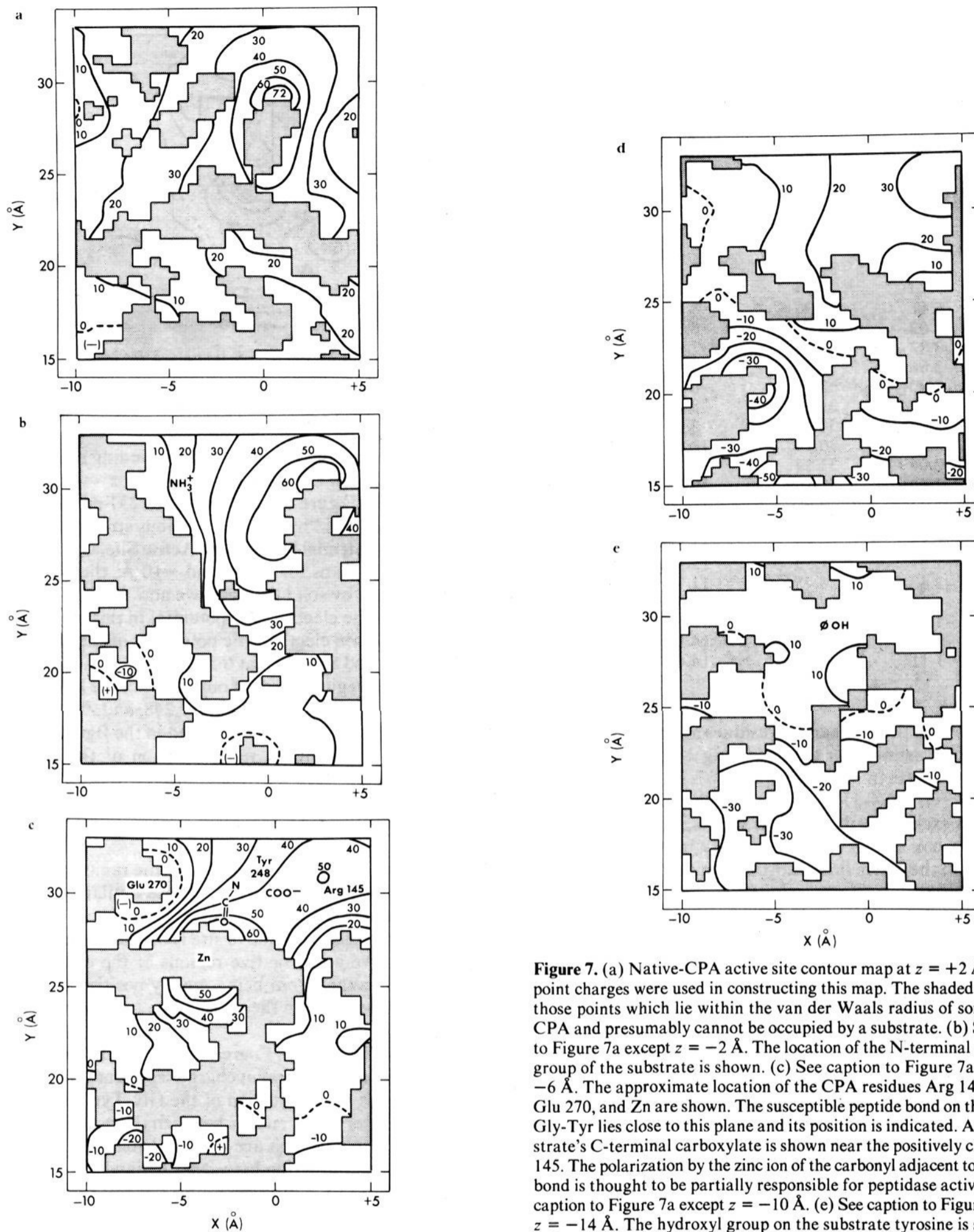
It is interesting that the active site is on the borderline between the positive and negative regions of the electrostatic potential and changes from being mostly positive for native CPA to mostly negative in the apoenzyme (see the  $z = -6$  Å maps of Figures 7 and 8).

**Point Charge Interaction Energies of CPA with the Gly-Tyr Substrate.** One can use our point charge representation of CPA and an analogous representation of the Gly-Tyr “substrate” to evaluate an interaction energy between the protein and the substrate. The results for this are given in Table VIII. We have broken down the interaction between CPA and the substrate into four parts, (1) CPA-ammonium, (2) CPA-amide, (3) CPA-carboxyl, and (4) CPA-tyrosyl, and have evaluated this

Table VIII. Electrostatic Interaction Energies for Gly-Tyr with CPA (kcal/mol)

Fragment	STO					4-31G				
	Native, Zn = 2	Bound, Zn = 2	Native, Zn = 1.4	Bound, Zn = 1.4	Apo	Native, Zn = 2	Bound, Zn = 2	Native, Zn = 1.4	Bound, Zn = 1.4	Apo
(1) Ammonium	60.15	72.07	61.04	64.26	-81.63	61.14	76.83	60.92	69.38	-79.64
(2) Amide	-42.63	-44.13	-37.91	-37.92	15.80	-92.65	-93.82	-77.57	-81.12	32.84
(3) Carboxyl	-106.83	-118.85	-116.66	-118.31	-7.29	-150.32	-141.44	-138.56	-141.37	-27.15
(4) Tyrosyl	15.22	14.50	12.12	13.45	-10.53	42.60	40.61	30.81	38.15	-18.26
Total	-74.09	-76.41	-81.41	-78.52	-83.65	-139.23	-117.82	-124.40	-114.96	-92.21
(1) <sup>a</sup> Amine	2.31	-3.75			6.74	10.31	-6.08			18.39
Total	-131.93	-152.23			4.72	-190.06	-200.73			5.82

<sup>a</sup> See ref 1.

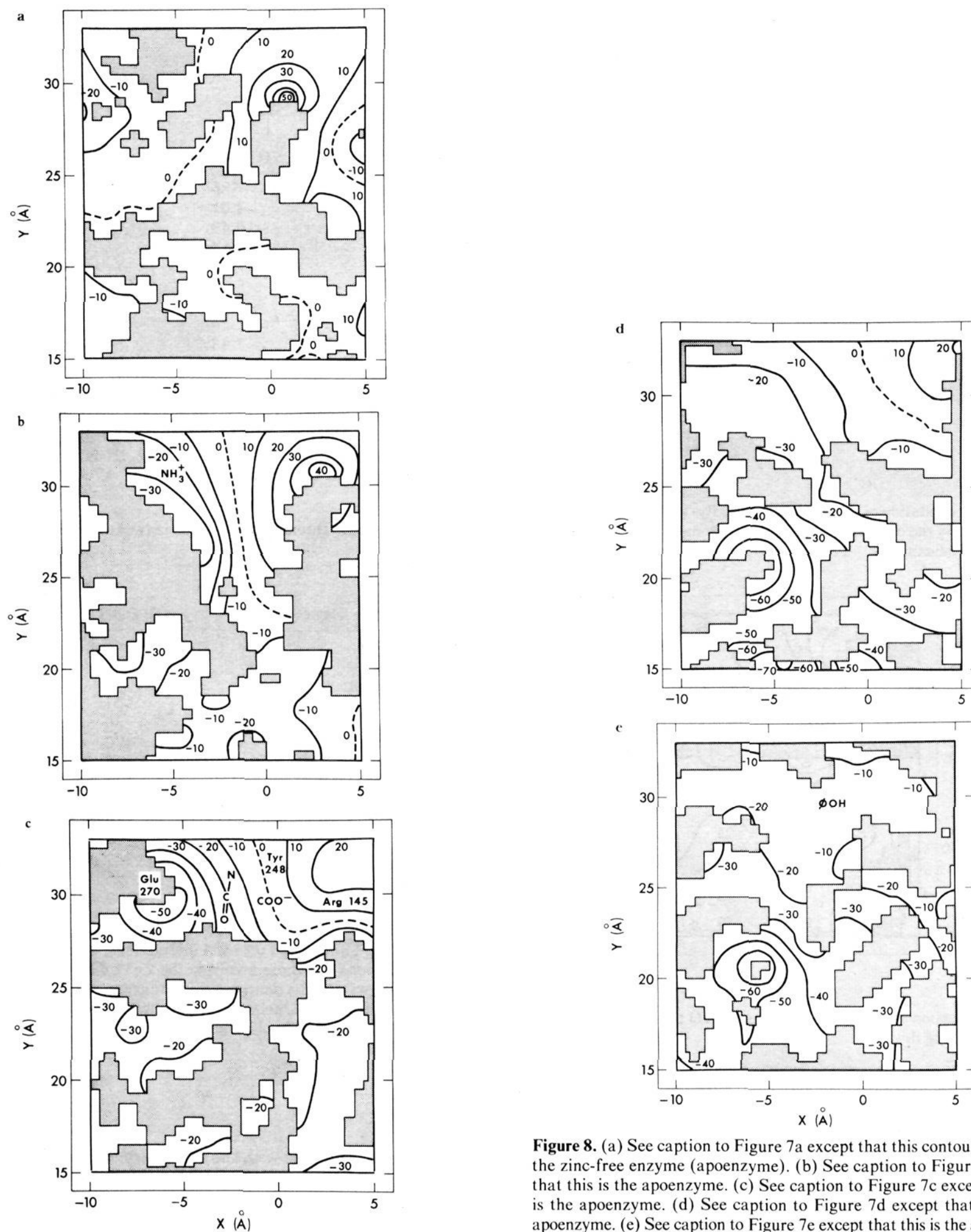


**Figure 7.** (a) Native-CPA active site contour map at  $z = +2 \text{ \AA}$ . STO-3G point charges were used in constructing this map. The shaded regions are those points which lie within the van der Waals radius of some atom of CPA and presumably cannot be occupied by a substrate. (b) See caption to Figure 7a except  $z = -2 \text{ \AA}$ . The location of the N-terminal ammonium group of the substrate is shown. (c) See caption to Figure 7a except  $z = -6 \text{ \AA}$ . The approximate location of the CPA residues Arg 145, Tyr 248, Glu 270, and Zn are shown. The susceptible peptide bond on the substrate Gly-Tyr lies close to this plane and its position is indicated. Also the substrate's C-terminal carboxylate is shown near the positively charged Arg 145. The polarization by the zinc ion of the carbonyl adjacent to the peptide bond is thought to be partially responsible for peptidase activity. (d) See caption to Figure 7a except  $z = -10 \text{ \AA}$ . (e) See caption to Figure 7a except  $z = -14 \text{ \AA}$ . The hydroxyl group on the substrate tyrosine is shown.

interaction energy using (a) the native and bound enzyme conformations, (b) Zn charges of +2, +1.4, and 0.8, and (c) STO-3G and 4-31G point charges. In addition, we have compared the interaction energy of an anionic (rather than zwitterionic) form of Gly-Tyr, since the positive potential in the binding region would be expected to favor this form. By setting the atomic charge on zinc equal to zero, we hoped to mimic the situation in the apoenzyme except for any structural changes which might occur upon removal of the metal ion. The case with the Zn charge equal to +1.4 (and +0.8) allowed for the possibility that the ion's positive charge is delocalized onto its neighboring ligands. In this case, we adjusted the charge on these ligands upward by a total of +0.6 (1.2).

When the substrate is bound in the active site of CPA, its positive ammonium group interacts repulsively with the enzyme. On the other hand, the substrate amide linkage, whose carbonyl binds to  $\text{Zn}^{2+}$ , and the carboxyl group, which forms a salt link to Arg 145, are strongly attracted to the enzyme. The tyrosyl group has a smaller, repulsive interaction with the enzyme. In the apoenzyme, the interaction energy of each of these fragments is significantly changed, but the total attraction energy remains almost the same. It is interesting that the apoenzyme appears to bind peptides and esters strongly but not to catalyze their hydrolysis. It may be that the apoenzyme binds the substrate at a different location within the active site or possibly even outside the active site region. If one considers





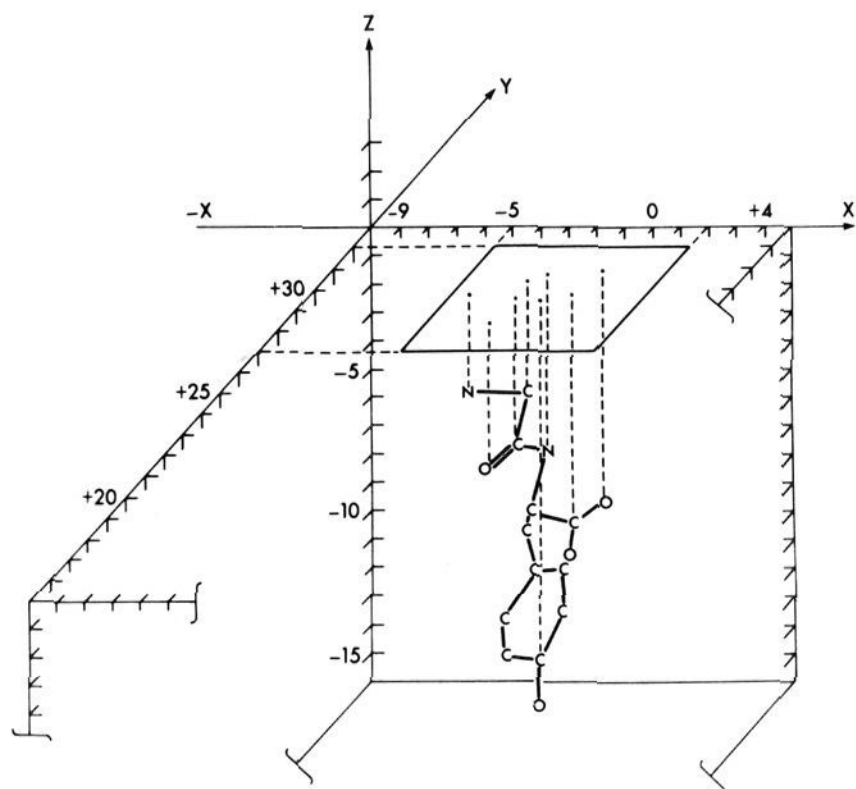
**Figure 8.** (a) See caption to Figure 7a except that this contour map is for the zinc-free enzyme (apoenzyme). (b) See caption to Figure 7b except that this is the apoenzyme. (c) See caption to Figure 7c except that this is the apoenzyme. (d) See caption to Figure 7d except that this is the apoenzyme. (e) See caption to Figure 7e except that this is the apoenzyme.

an anionic substrate (with a N-terminal amine instead of ammonium), the binding interaction with the enzyme is greatly enhanced; Quioco and Lipscomb<sup>1</sup> point out that the state of ionization of the amine has not yet been determined.

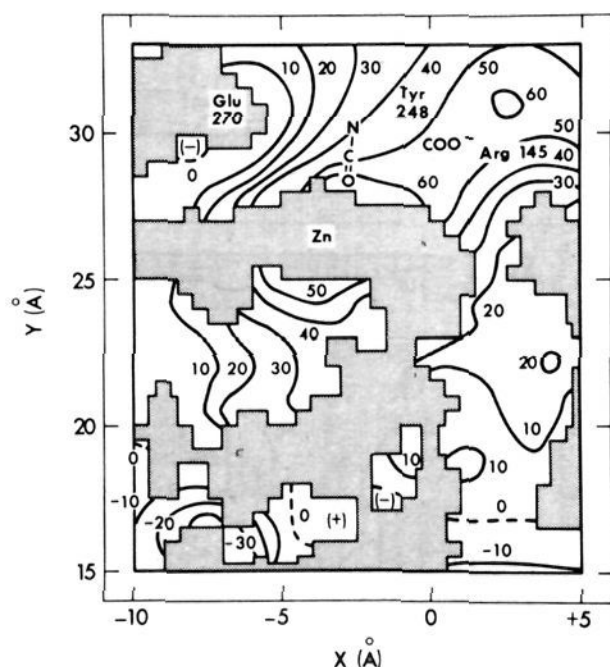
One must realize that the magnitude of our calculated interaction energy is certainly too large, since we are comparing the energies of the enzyme-substrate complex with the separated gas-phase (unsolvated) enzyme and substrate. If one places waters on the ammonium, amide, carboxyl, and tyrosine OH of the substrate the calculated solvation energy (based on electrostatics) is of the same magnitude as the enzyme-substrate interaction energy.

One of the first interesting questions one might ask is

whether the substrate is bound more tightly in the native or "bound" CPA conformation? Unfortunately, the answer to this question is equivocal and depends on our point charge choice, the STO-3G charges predicting the "bound" geometry of CPA to hold the substrate more tightly and the 4-31G charges predicting the reverse. When the Zn charge is +1.4,<sup>19</sup> one finds the order of binding reversed for the STO-3G charges. When the substrate's N-terminal group is an amine, both STO-3G and 4-31G predict the bound geometry to hold the substrate more tightly than the native. Thus, we cannot give a definitive answer to the question of whether electrostatics can satisfactorily describe the process in which an incoming substrate induces a conformation change in the enzyme ("induced



**Figure 9.** A representation of the location of the Gly-Tyr substrate when bound in the active site of CPA. The projection of some of the atoms onto the  $xy$  plane is indicated.



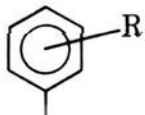
**Figure 10.** See caption to Figure 7c except that 4-31G point charges were used in constructing this map.

fit"). One needs more precise x-ray data for the "bound" enzyme and a better potential function to answer such a question.

Another area in which these point charge calculations might be useful is in the design of improved substrate or inhibitor molecules. Since solvation and desolvation play such a large role in the binding of substrates and we have no simple way to treat hydrophobic effects in this model, it makes sense to compare substrates of similar hydrophobicity. In Table IX we list the interaction energies for different -OH substitutions on the tyrosine ring (including only the interaction energies of the -PhOH moiety). Although chemically there are only three different substitution positions (ortho, meta, and para) the protein bound Gly-Tyr has five different locations for the oxygen and ten for the hydrogens. At both the STO-3G and 4-31G level, one predicts relative binding affinities of ortho > meta > para. No experimental data are available on the relative binding strength of "artificial" Gly-Tyr substrates to CPA.

**Protein-Protein Point Charge Interaction Energies.** In an attempt to gain some insight into the role of electrostatics in holding the protein together, we evaluated the interaction energy between different secondary structural units<sup>4,20</sup> and these results are summarized in Table X.

**Table IX.** Interaction Energies of Substituted Phenylalanine Side Chains (kcal/mol)

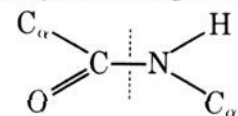
			
Substitution R <sup>a</sup>		STO energy	4-31G energy
1	H	-7.86	-22.20
2	CE1OH, H1	-4.01	-9.94
3	CE1OH, H2	-10.49	-22.17
4	CD1OH, H1	-6.62	-23.13
5	CD1OH, H2	-7.32	-18.65
6	CZOH, H1	-5.05	-14.91
7	CZOH, H2	-7.63	-24.10
8	CD2OH, H1	-11.24	-30.72
9	CD2OH, H2	-7.48	-24.26
10	CE2OH, H1	-9.99	-26.56
11	CE2OH, H2	-10.11	-27.75

<sup>a</sup> See Table VII for coordinates of carbon atoms; CD1OH, H1 means OH substituted on CD1 with OH cis to CG-CD bond. Then, as one proceeds through CE1, CZ, CE2, and DC2, H1 is always the proton position with the OH cis to the preceding C-C bond.

**Table X.** Electrostatic Interaction Energies between Secondary Structure Elements

Interaction between	E, kcal/mol	
	STO-3G	4-31G
Zn <sup>2+</sup> ··· α helices	-547	-501
Zn <sup>2+</sup> ··· β sheet	-150	-178
Zn <sup>2+</sup> ··· random coil	226	228
Zn <sup>2+</sup> ··· turns	107	60
Zn <sup>2+</sup> ··· loops	30	44
α helices ··· α helices	470	293
α helices ··· β sheet	-44	-85
β sheet ··· β sheet	-224	-333
Whole protein (native) <sup>a</sup>	-1178	-2318
Whole protein (apo) <sup>b</sup>	-844	-1971
Whole protein (apo, excluding neighbor) <sup>c</sup>	-347	+230
Whole protein (apo, excluding peptide connection) <sup>d</sup>	-497	+451

<sup>a</sup> Including all unit-unit interactions. There are 38 units (including Zn<sup>2+</sup>); thus 38 \* (38 - 1)/2 total interactions. <sup>b</sup> Now including all interactions but those involving the Zn<sup>2+</sup>. <sup>c</sup> Excluding the 36 interactions between units contiguous in sequence. <sup>d</sup> Subtracting out only the directly bonded, geminal, and vicinal interactions in the peptide linkages connecting the 37 segments.



The Zn<sup>2+</sup> atom interacts with the secondary elements as one would predict from their net charges. Helix-helix interactions are strongly repulsive, helix-β slightly attractive, and β-β interactions clearly attractive, as one would expect from their H-bonding alignment.

The whole protein results are ambiguous because when one considers the apoenzyme (which is thought to have a structure similar to native CPA) and excludes interactions between neighboring secondary elements, the resulting electrostatic interaction energy is calculated to be attractive when STO-3G charges are used but repulsive for the 4-31G charges (including the zinc ion in our calculations lowers the internal interaction energy of the native structure by approximately another 340 kcal/mol for both STO-3G and 4-31G point charge sets). Thus, we are unable to make a prediction as to the sign of the internal interaction energy of both the apo and native enzymes. A definitive statement as to the sign and approximate magnitude of this energy and its effect upon protein folding would require a more accurate choice of point charges and a knowl-

edge of the electrostatic interaction energy for the denatured or unfolded structures.

The possibility that the net internal interaction energy of the protein in its native state is repulsive is not as absurd as it might seem at first glance. The role of water and counterions in solvating the charged residues and the favorable free energy for burying hydrophobic groups could certainly be more important than the internal interaction energy in determining the minimum energy structure.

### Summary and Future Directions

It appears that the electrostatic potential and point charges of proteins can be useful in understanding macroscopic electrical properties of proteins and in designing substrates and inhibitors. A quantitative understanding of subtle effects like changes of binding energy due to "induced fit" conformation changes of the protein residues probably requires a more sophisticated potential function.<sup>22</sup>

Future areas of interest are the examination of the electrostatic potentials of a number of other globular proteins to see if the maps are similar for similar classes of proteins. Immunoglobulins may have very large fluctuations in their electrostatic potential which might help account for their high binding constants. Protein subunit interactions are also of interest, since these may contain an important directionality due to electrostatic effects.

Intercalation of small molecules into nucleic acids is now being studied in our laboratory making use of the electrostatic potential of the nucleotide phosphates.

Ultimately, one would like to be able to predict the nature of the "active site" or "receptor pocket" without complete three-dimensional x-ray data on the protein. By examining the structure-activity relationships in a number of "substrates", attempts have been made by Petrongolo et al.<sup>13</sup> and by Weinstein et al.<sup>13</sup> to do just that. We hope that a knowledge of the electrostatic potential of a number of "receptor" proteins will provide general principles that make such a task easier.

**Acknowledgments.** We are grateful for financial support from General Medical Sciences (Grant GM-20564 and Career Development Award GM-70718 to P.A.K.). Comments by Dr. Tom Steitz and many discussions with our colleague Dr. Irwin Kuntz were invaluable to this study. We also express our thanks to the referees of this paper for several suggestions on how to improve the figures, and Richard Tuck for his help in computer graphics.

### References and Notes

- (1) F. A. Quiocho and W. N. Lipscomb, *Adv. Protein Chem.*, **25**, 1-78 (1971).
- (2) See also W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 3797 (1973), and references cited therein.
- (3) W. N. Lipscomb, J. A. Hartsuck, F. Quiocho, P. H. Bethge, M. L. Ludwig, T. A. Steitz, H. Muirhead, and J. C. Coppola, *Brookhaven Symp. Biol.*, **21**, 24 (1968).
- (4) (a) M. A. Stahmann, J. S. Fruton, and M. Bergmann, *J. Biol. Chem.*, **164**, 753 (1946); (b) K. Hofmann and M. Bergmann, *ibid.*, **134**, 225 (1940); (c) I. Schechter and A. Berger, *Biochemistry*, **5**, 3371 (1966); (d) M. Bergmann and J. S. Fruton, *J. Biol. Chem.*, **117**, 189 (1937); (e) E. L. Smith and H. T. Hanson, *ibid.*, **179**, 803 (1949); (f) C. A. Dekker, S. P. Taylor, and J. S. Fruton, *ibid.*, **180**, 155 (1949).
- (5) N. Abramowitz, I. Schechter, and A. Berger, *Biochem. Biophys. Res. Commun.*, **29**, 862 (1967).
- (6) See R. E. Dickerson and I. Geis, "The Structure and Action of Proteins", W. A. Benjamin, New York, N.Y., 1969, for a stereoview of the structure of CPA.
- (7) R. Bonaccorsi, A. Pullman, E. Scrocco, and J. Tomasi, *Chem. Phys. Lett.*, **12**, 622 (1972).
- (8) P. Kollman, J. McKelvey, A. Johansson, and S. Rothenberg, *J. Am. Chem. Soc.*, **97**, 955 (1975).
- (9) W. A. Lathan and K. Morokuma, *J. Am. Chem. Soc.*, **97**, 3615 (1975); S. Iwata and K. Morokuma, *ibid.*, **97**, 966 (1975).
- (10) See, for example, H. Margenau and N. Kestner, "Theory of Intermolecular Forces", Pergamon Press, Oxford, 1973.
- (11) C. Tanford, "The Hydrophobic Effect", Wiley, New York, N.Y., 1973.
- (12) Using linear regression techniques and octanol-water partition coefficients such as is extensively applied by Hansch and co-workers [see, for example, C. Hansch and C. Silipo, *J. Med. Chem.*, **17**, 661 (1974)] one might be able to quantitatively factor out hydrophobic effects in order to examine the electrostatic effect alone.
- (13) See, for example, C. Petrongolo, J. Tomasi, B. Macchia, and F. Macchia, *J. Med. Chem.*, **17**, 501 (1974), and H. Weinstein, S. Srebrenik, R. Pauncz, S. Maayani, S. Cohen, and M. Sokolovsky, *Jerusalem Symp. Quantum Chem. Biochem.*, in press.
- (14) Using Gaussian 70, a computer program developed for molecular orbital calculations at Carnegie Mellon University in the group of J. Pople (QCPE no. 236). We thank W. J. Hehre for the use of the CDC 7600 version of the program.
- (15) One of us (P.A.K.) has found that STO-3G calculations generally underestimate dipole moments while 4-31G overestimates them.
- (16) The following standard bond lengths were used: C-H = 1.09 Å; O-H = 0.96 Å; N-H = 1.02 Å. All X-Y-H bond angles were either approximately tetrahedral or 120° depending on the state of hybridization of the Y atom.
- (17) The Asn and Gln assignments are in the order in which they occur in CPA starting from the N-terminal ammonium end of the protein. A "0" means that the second atom in the list of atomic coordinates given in ref 1 for each Asn and Gln residue is the side-chain nitrogen and the third is the oxygen. When the flag is 1, the assignments are reversed. Thus our assignments are Asn, 1101?001110?11101; Gln, ?11?11?1100 the question marks indicate those residues for which there was no clear preference based on H-bond contacts and minimum steric repulsion. In our calculations, we considered two cases: (1) all ? = 0 and (2) all ? = 1.
- (18) W. Ortung, *Biochemistry*, **9**, 2394 (1970).
- (19) We thank Dr. Gilda Loew for pointing out to us the possible sensitivity of the results to delocalization of the positive charge on zinc onto neighboring ligands.
- (20) See ref 6, I. D. Kuntz, *J. Am. Chem. Soc.*, **94**, 8568 (1972), and P. Kollman, D. Hayes, and I. D. Kuntz, *Jerusalem Symp. Quantum Chem. Biochem.*, in press, for descriptions of the different secondary structure elements and their frequency of occurrence in CPA.
- (21) J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *J. Mol. Biol.*, **15**, 399 (1966).
- (22) See, for example, H. A. Scheraga, *Chem. Rev.*, **71**, 195 (1971).