Aswad, D. W., & Deight, E. A. (1983) *J. Neurochem. 40*, 1718-1726.
Barber, J. R., & Clarke, S. (1983) *J. Biol. Chem. 258*, 1189-1196.
Bornstein, P. (1970) *Biochemistry 9*, 2408-2421.
Bornstein, P., & Balian, G. (1970) *J. Biol. Chem. 245*, 4854-4856.
Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132-145.
Canfield, R. E. (1963) *J. Biol. Chem. 238*, 2691-2697.
Cappugi, G., Nassi, P., Treves, C., & Ramponi, G. (1971) *Experientia 27*, 237-239.
Clarke, S., & O'Connor, C. M. (1983) *Trends Biochem. Sci. (Pers. Ed.) 8*, 391-394.
D'Alessio, G., Floridi, A., De Prisco, R., Pignero, A., & Leone, E. (1972) *Eur. J. Biochem. 26*, 153-161.
D'Alessio, G., Malorni, M. C., & Parente, A. (1975) *Biochemistry 14*, 1116-1121.
Delange, R. J., & Smith, E. L. (1971) *Enzymes (3rd Ed.) 3*, 81-118.
Di Donato, A., & D'Alessio, G. (1973) *Biochem. Biophys. Res. Commun. 55*, 919-928.
Di Donato, A., & D'Alessio, G. (1981) *Biochemistry 20*, 7232-7237.
Dostal, J., & Matousek, J. (1973) *J. Reprod. Fertil. 34*, 197-200.
Galletti, P., Ingrosso, D., Nappi, A., Gragnaniello, V., Iolascon, A., & Pinto, L. (1983) *Eur. J. Biochem. 135*, 25-31.
Groskopf, W. R., Holleman, J. W., Margoliash, E., & Klotz, I. M. (1966) *Biochemistry 5*, 3783-3796.
Jamalludin, M., Kim, S., & Paik, W. K. (1976) *Biochemistry 15*, 3077-3081.
Johnson, B. A., Freitag, N. E., & Aswad, D. W. (1985) *J. Biol. Chem. 260*, 10913-10916.
Kasper, C. B. (1975) in *Protein Sequence Determination* (Needleman, S. B., Ed.) pp 141-161, Springer-Verlag, Berlin.
Kim, S., & Paik, W. K. (1970) *J. Biol. Chem. 245*, 1806-1813.
Kim, S., & Li, C. H. (1979) *Int. J. Protein Res. 13*, 281-285.
Kim, S., Nochumson, S., Chin, W., & Paik, W. K. (1978) *Anal. Biochem. 84*, 415-422.
Manning, J. M. (1970) *J. Am. Chem. Soc. 92*, 7449-7454.
Matousek, J. (1973) *Experientia 29*, 858-859.
Matsuo, H., & Narita, K. (1975) in *Protein Sequence Determination* (Needleman, S. B., Ed.) pp 104-113, Springer-Verlag, Berlin.
Matsuo, H., Fujimoto, Y., & Tatsuno, T. (1966) *Biochem. Biophys. Res. Commun. 22*, 69-72.
Matsuo, H., Fujimoto, Y., & Tatsuno, T. (1967) *Chem. Pharm. Bull. 15*, 716-718.
McFadden, P. N., & Clarke, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2460-2464.
Murray, E. D., Jr., & Clarke, S. (1984) *J. Biol. Chem. 259*, 10714-10721.
Neurath, H., & Schwert, G. W. (1950) *Chem. Rev. 46*, 69-153.
Oliva, A., Galletti, P., Zappia, V., Paik, W. K., & Kim, S. (1980) *Eur. J. Biochem. 104*, 595-602.
Piccoli, R., & D'Alessio, G. (1984) *J. Biol. Chem. 259*, 693-695.
Pucci, P., Sannia, G., & Marino, G. (1983) *J. Chromatogr. 270*, 371-377.
Suzuki, H., Greco, L., Parente, A., Farina, B., La Montagna, R., & Leone, E. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 2, pp 93-94, N.B.R. Foundation, Washington, DC.
Tamburrini, M., Piccoli, R., De Prisco, R., Di Donato, A., & D'Alessio, G. (1986) *Ital. J. Biochem. 35*, 22-32.
Vescia, S., Tramontano, D., Augusti-Tocco, G., & D'Alessio, G. (1980) *Cancer Res. 40*, 3740-3744.
Waley, S. G., & Watson, J. (1954) *Biochem. J. 57*, 529-538.
Weber, K. L., & Konisberg, W. (1967) *J. Biol. Chem. 242*, 3563-3578.

Free Energy of Charges in Solvated Proteins: Microscopic Calculations Using a Reversible Charging Process†

A. Warshel,* F. Sussman, and G. King
Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482
Received October 17, 1986

ABSTRACT: Evaluation of the free energy of ionization of acidic groups in proteins may be used as a powerful and general test case for determining the reliability of calculations of electrostatic energies in macromolecules. This work attacks this test case by using an adiabatic charging process that evaluates the *changes* in free energies associated with ionizing the acidic groups Asp-3 and Glu-7 in bovine pancreatic trypsin inhibitor and aspartic acid in solution. The results of these free energy calculations are very encouraging; the error range is about 1 kcal/mol for these free energy changes of about −70 kcal/mol. This indicates that we are finally approaching the stage of obtaining quantitative results in modeling the energetics of solvated proteins.

To correlate the structures of proteins with their functions, one must be able to express structures in terms of energies. Probably the most important requirement for such a correlation is the ability to evaluate the energies of charges in proteins. This is apparently the case in such key processes as proton transfer, electron transfer, subunit interaction, ion binding, and many enzymatic reactions. Thus, it is important to develop methods of calculation that yield reliable electrostatic energies in proteins (Warshel & Russell, 1984). In developing these methods it is essential to have experimental data capable of providing sufficiently discriminative test cases against which the calculations can be compared, so that incorrect or incomplete models may be discarded. Probably the

† This work was supported by NSF Grant CHE-8519194, by NIH Grant GM-24492, and by the NSF San Diego supercomputer center.

0006-2960/86/0425-8368$01.50/0

simplest yet the most powerful test case is provided by the energetics of ionized groups in proteins. These energies, which determine the intrinsic pK_a's of the given groups, include the self-energy term, whose evaluation is a prerequisite for consistent electrostatic calculations.

The only reported microscopic calculations of the energetics of ionized groups in proteins have been based on the simplified solvent model of the protein dipoles Langevin dipoles (PDLD) approach (Warshel & Russell, 1984; Russell & Warshel, 1985). The emergence of supercomputers makes it possible now to explore more rigorous and much more expensive models. Here we report an attempt to evaluate the free energies of some ionized acids in bovine pancreatic trypsin inhibitor (BPTI) and in solution, using an adiabatic charging approach based on the umbrella sampling method (Torrie & Valleau, 1977). This approach has already been used in our studies of reactions in solutions (Warshel, 1982) and proteins (Warshel & Russell, 1984; Warshel, 1984). Other versions of umbrella sampling and related perturbation methods were applied to various systems in solution (Beveridge et al., 1985; Jorgensen & Ravimohan, 1985). Here, however, our primary aim is not to elaborate on free energy methods (some of which were formulated quite early [e.g., Kirkwood (1935)]) but to determine the *error range* of adiabatic charging calculations in proteins.

A powerful and well-defined test case is provided by comparing the energetics of ionizing acidic groups in *different* sites of BPTI (Asp-3 and Glu-7) and in solution. The free energies of ionization are quite similar for these acids (less than 0.5 kcal/mol difference; Russell & Warshel, 1985). Obtaining a difference of this order between the calculated absolute values of the relevant energies ($\simeq$−70 kcal/mol) is a major challenge and a prerequisite for quantitative electrostatic calculations. This test case involves charged groups in drastically different environments, where the different energy contributions (permanent dipoles, induced dipoles, water dipoles) vary by as much as 20 kcal/mol. This point can be seen clearly in Figure 6 of Russell and Warshel (1985). The proper quantitative balance between the different energy contributions can be obtained only with theoretical approaches that capture the main physics of the system.

This study gives similar free energies (to within 1 kcal/mol) for the three ionization processes of our test case. This indicates that we are finally approaching the stage of obtaining quantitative results in calculating electrostatic free energies in solvated proteins.

Theoretical Model

In order to evaluate electrostatic energies in proteins, it is essential to represent explicitly both the protein and the surrounding solvent. This work uses an all-atom representation for the entire protein–water system. The protein is represented by the same force field used in our earlier studies, which includes the regular bonding and nonbonding terms, as well as the dielectric effect of the protein-induced dipoles (Warshel & Levitt, 1976). As described in detail in Russell and Warshel (1985), we represent the interaction between the permanent dipoles (μ) of the protein and the charges (Q) of the relevant "solute" (in this case the acidic group) and the permanent dipole–permanent dipole interaction by

$$V_{Q\mu} = 332\sum_{ij} Q_i q_j / r_{ij}$$

$$V_{\mu\mu} = 332\sum_{k<j} q_k q_j / r_{kj} \qquad (1)$$

in which Q_i and q_j are respectively the solute charges and the protein residual charges (in units of electron charge). The distances r_{ij} are in Å, and the potential energy is in kcal/mol. Note that these energy contributions are evaluated with the vacuum dieletric constant (ϵ = 1), since the dielectric contributions are included explicitly in the model. The interaction between the solute charges and the protein-induced dipoles is evaluated by (Russell & Warshel, 1985)

$$V_{\text{ind}} = -166\sum_i \xi_i \mu_i = -166\sum_i \alpha_i (\xi_i^0)^2 / d \qquad (2)$$

in which μ_i are the induced dipoles of the protein, ξ_i^0 is the vacuum field on the ith protein atom from the charges and residual charges of the system, α_i are the atomic polarizabilities, and d is an effective screening function (not to be confused with the dielectric constant) that reproduces the average results of self-consistent calculations. The parameters for the protein–solute force field are given in Table I.

Table I: Parameters Used in Calculations[a]

type	parameter	value (or source)
ionized acid charges (A^-)	Q_O, Q_C	−0.80, 0.60
protonated acid charges (AH)	Q_C, Q_{O_1}, Q_{O_2}, Q_H	+0.20, −0.20, −0.30, +0.30
water charges	q_O, q_H	−0.82, +0.41
water–acid van der Waals	A_{OO}, B_{OO}	710 000, −612
	A_{CO}, B_{CO}	450 000, −600
water–water van der Waals	A_{OO}, B_{OO}	629 000, −625
protein residual charges and polarizabilities		Table I of Russell and Warshel (1985)
protein force field		as in Warshel and Levitt (1976)

[a] A and B are the coefficients of the r^{-12} and r^{-6} terms, in units of Å^{12} kcal/mol and Å^6 kcal/mol, respectively. The van der Waals interactions involving the water hydrogens are taken as zero. Charges are given in units of electron charge.

The water molecules around the protein are represented by the surface-constrained all-atoms solvent (SCAAS) model (Warshel & King, 1985). This model uses a relatively small number of water molecules (80 in the present case) to create a sphere around the relevant group (see Figure 1). Related earlier surface-constrained models, which date back to the SCSSD model (Warshel, 1978), are reviewed in Warshel and Russell (1984). The SCAAS model incorporates angular and radial constraints in order to compensate for the artificial surface created as a result of using a finite number of water molecules.

The angular constraint forces the water molecules in the surface region to be polarized (on the average) in the direction expected of the corresponding infinite system. This polarization is estimated with the Langevin dipole relationship [see Russell and Warshel (1985)]. The polarization energy of the surface molecules is

$$V^s_{\text{pol}} = -166\sum_j {}^s\mu_{w_j}\xi_j^0 \qquad (3)$$

in which the summation runs only over the water molecules in the surface region, μ_w are the dipoles of the water molecules, and ξ^0 are the fields due to the protein residual charges and the solute charges. We require the average polarization energy of the surface dipoles to approach the Langevin estimate:

$$(V^s_{\text{pol}})_L = -166\sum_j {}^s\mu_{L,j}|\xi_j^0| \qquad (4)$$

In order to achieve this, we define a new field

$$(\xi^0)' = \xi^0(1 + \delta) \qquad (5)$$

with δ chosen to satisfy the relation

$$1 + \delta = (V^s_{\text{pol}})_L / V^s_{\text{pol}} \qquad (6)$$

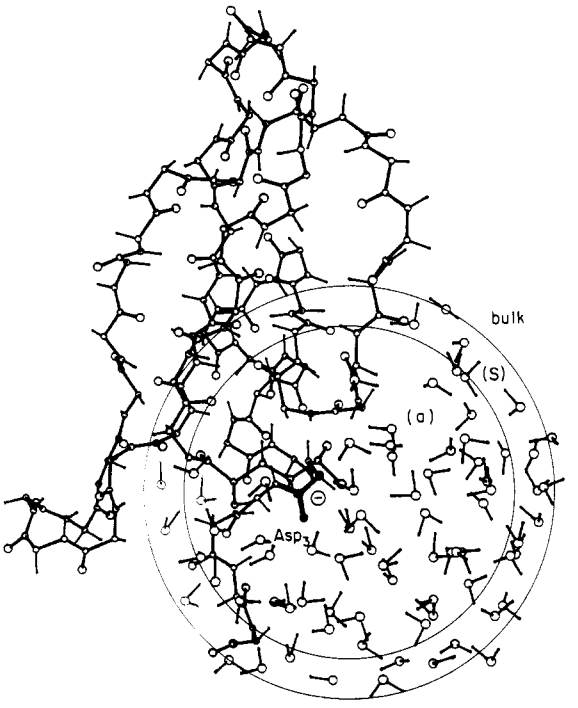

FIGURE 1: Model system used in this study. The figure describes the ionized Asp-3 of BPTI surrounded by the protein–water system. The water molecules are simulated by an all-atom model that is constrained at the surface (region s) to have the polarization expected in this region of an infinite system. The surface region is surrounded by bulk solvent with a high dielectric ($\epsilon = 80$).

Using $(\xi^0)'$ of eq 5 as the actual field on the surface molecules is a simple and effective means of satisfying our constraint requirement.

The radial constraint forces the water molecules in the protein–water system to maintain the correct density. This is accomplished by imposing the force

$$f_j = -k(R_j - R_j^0) \tag{7}$$

on water molecule j, in which R_j is the distance between molecule j and the center of the sphere. R_j^0 is determined by first defining the function $n(R)$:

$$n(R) = 1 - n_{\text{solute}} - n_{\text{protein}}(R) + 4\pi\rho\int_0^R g(r)r^2\,dr \tag{8}$$

which represents the number of water molecules within the radius R [$g(r)$ is the experimental radial distribution function for water]. The term n_{solute} is the effective number of water molecules excluded by the solute, and the term $n_{\text{protein}}(R)$ is the effective number of water molecules excluded by protein atoms within the radius R. The value of R for which $n(R) = j$ (with j an integer) is taken as R_j^0. The value used for the force constant k is 15 kcal mol^{-1} Å^{-2}.

Now our total protein–water–solute potential is represented by

$$V_{\text{tot}} = V_{\text{p}}^{\text{strain}} + V_{Q\mu}^{\text{p}} + V_{\mu\mu}^{\text{p}} + V_{\text{ind}}^{\text{p}} + V_{Q\text{w}} + V_{\text{pw}} + V_{\text{ww}} + V_{\text{constraints}} \tag{9}$$

in which $V_{\text{p}}^{\text{strain}}$ is the potential energy of the protein, excluding the electrostatic contributions of $V_{Q\mu}^{\text{p}}$, $V_{\mu\mu}^{\text{p}}$, and $V_{\text{ind}}^{\text{p}}$. $V_{Q\text{w}}$ is the interaction between the water and the solute; this term contains both the charge–charge and the van der Waals contributions. V_{pw} represents the interaction between the water molecules and the protein, V_{ww} is the water–water interaction, and $V_{\text{constraints}}$ is the potential of the angular and radial constraints. The parameters used are summarized in Table I. The parameters for the water–solute interactions were obtained by calibrating the SCAAS model to reproduce the experimental free energy of acid ionization in water.

To evaluate free energy with this, or any other model, it is important to find a reaction coordinate that guarantees fast convergence. In cases of acid ionization, one can consider the reaction coordinate to be the transfer of the proton from the given acid to a water molecule. This reaction coordinate involves a concerted motion of the protein–solvent system and can best be described by the empirical valence bond (EVB) method (Warshel & Russell, 1984). However, since we are not interested here in activation free energies, but only in free energy differences, we can concentrate (see below) on the response of the protein to the reversible charging of the relevant groups. The corresponding free energy can be evaluated with the umbrella sampling equation (Torrie & Valleau, 1977):

$$\exp(-\beta\Delta G_{\text{ba}}) = \langle\exp(-\beta\Delta V_{\text{ba}})\rangle_{\text{a}} \tag{10}$$

in which $\beta = (k_{\text{b}}T)^{-1}$, k_{b} is the Boltzmann constant, $\Delta G_{\text{ba}} = G(Q_{\text{b}}) - G(Q_{\text{a}})$, and $\Delta V_{\text{ba}} = V(Q_{\text{b}}) - V(Q_{\text{a}})$. $G(Q_\lambda)$ and $V(Q_\lambda)$ are respectively the free energy and potential energy of the protein–solvent system when the solute (acid) charge is taken as Q_λ. The expression $\langle\ldots\rangle_{\text{a}}$ signifies an average of either a molecular dynamics or a Monte Carlo trajectory, propagated on the potential surface $V(Q_{\text{a}})$. The overall free energy of a charging process converges most efficiently when ΔG is evaluated in small steps, changing the charge in small increments from Q_{a} to Q_{b}. In this way the free energy is evaluated by a process that resembles the macroscopic integration described in standard physics textbooks. In fact, our umbrella sampling procedure (when done in small increments) is equivalent to the microscopic charging proposed in the statistical mechanics literature (Hill, 1956). The issue here, however, is not of formal justification but of *convergence*. It is entirely unclear at the outset how much computer time is required for evaluation of what is in *principle* a rigorous approach. If we can design an approach that converges within a reasonable computer time, then we have a way to study the energetics of charges in proteins. In our calculations we determine both ΔG_{ba} and ΔG_{ab}. The charging process is reversible when $\Delta G_{\text{ba}} = -\Delta G_{\text{ab}}$, which is only true for sufficiently long trajectories and/or sufficiently small charge increments. We use the reversibility of the charging process as a convergence criterion.

Calculations and Discussion

In this work we consider the free energy associated with the ionization of two acidic groups (Asp-3 and Glu-7) in BPTI and with the ionization of an aspartic acid in solution. The relevant free energies are given by (Russell & Warshel, 1985)

$$\Delta G_{\text{p}}(\text{AH}_{\text{p}} \rightarrow \text{A}_{\text{p}}^- + \text{H}_{\text{w}}^+) = \Delta G_{\text{sol}}^{\text{w}\rightarrow\text{p}}(\text{A}^-) - \Delta G_{\text{sol}}^{\text{w}\rightarrow\text{p}}(\text{AH}) + \Delta G_{\text{w}}(\text{AH}_{\text{w}} \rightarrow \text{A}_{\text{w}}^- + \text{H}_{\text{w}}^+) = \Delta G_{\text{sol}}^{\text{w}\rightarrow\text{p}}(\text{A}^-) - \Delta G_{\text{sol}}^{\text{w}\rightarrow\text{p}}(\text{AH}) + 2.3RT(\text{p}K_{\text{a}} - \text{pH}) \tag{11}$$

where $\Delta G_{\text{sol}}^{\text{w}\rightarrow\text{p}}$ is the difference in solvation energy of the indicated species in (p) protein and (w) water. As seen from this expression, one only has to calculate the solvation free energy of the given acid in its ionized and unionized forms. Furthermore, comparing different acids in proteins, one only has to calculate the solvation free energies of A^- and AH in their protein sites. That is, eq 11 can be rewritten as

$$\Delta G_{\text{p}}(\text{AH}_{\text{p}} \rightarrow \text{A}_{\text{p}}^- + \text{H}_{\text{w}}^+) = \Delta G_{\text{sol}}^{\text{p}}(\text{A}^-) - \Delta G_{\text{sol}}^{\text{p}}(\text{AH}) + \text{constant} = \Delta\Delta G_{\text{sol}}^{\text{p}} + \text{constant} \tag{12}$$

Thus, our key task is the evaluation of $\Delta\Delta G_{\text{sol}}^{\text{p}}$. This solvation

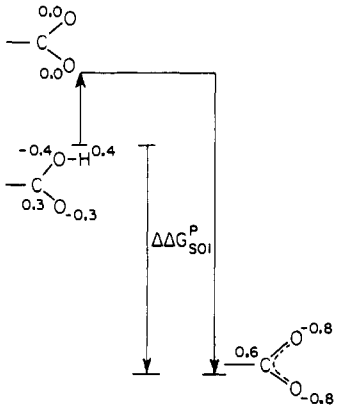

FIGURE 2: Charging cycle used in our adiabatic charging process.

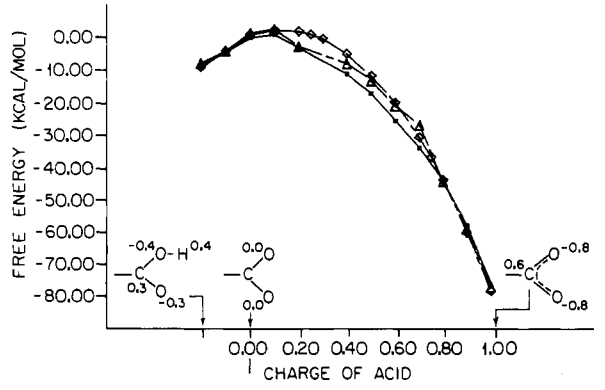

FIGURE 3: Calculated free energy of ionization for Asp-3 (△) and Glu-7 (■) in BPTI and Asp (◇) in solution.

Table II: Calculated Energetics of Ionizing Acidic Groups in BPTI and Solution[a]

acid	$\Delta G_{sol}(A^-)$	$\Delta G_{sol}(AH)$	$\Delta\Delta G(AH \rightarrow A^-)$
Asp-3	−79.8 ± 0.5	−9.5 ± 0.5	−70.3 ± 1.0
Glu-7	−79.0 ± 0.5	−9.8 ± 0.5	−69.2 ± 1.0
Asp in solution	−80.0 ± 0.5	−9.9 ± 0.5	−70.1 ± 1.0

[a] Energy values are given in kcal/mol.

free energy can be evaluated by the "charging" process described in Figure 2. In this process we start with the protonated acid and then remove the residual charges to form a neutral COO group. From here we perform a gradual charging process to form the fully ionized carboxylate group. The calculations consist of running trajectories at 300 K of 6 ps (after a 2-ps equilibration period) for each of the points of the charging process. About 2 h of Cray time is needed to complete a charging cycle containing 12 points.

Figure 3 and Table II display the results of our calculations of the free energies associated with charging Asp-3 and Glu-7 in BPTI, as well as of aspartic acid in water. The results are quite encouraging; the solvation free energies of these residues are around 70 kcal/mol, very close to the experimental estimate (Russell & Warshel, 1985). Maybe more important is the fact that the free energies of these three cases are similar to each other to within 1 kcal/mol. Since the observed difference in solvation free energies is about 0.3 kcal/mol [as estimated in Russell and Warshel (1985) from the corresponding intrinsic pK_a's], we conclude that the relative error in the calculation is at the 1 kcal/mol range. This is not to be confused with the difference between the absolute values of ΔG_{ba} and ΔG_{ab}, which is about 3 kcal/mol with the current simulation time and with increments of 0.1 charge units in the charging process. This difference becomes smaller upon reduction of the charge increments (i.e., inclusion of more points in the charging cycle). More importantly, the *average* of ΔG_{ba} and ΔG_{ab} (which is the value we report here) does not change by more than 1 kcal/mol when smaller charge increments are used. There is no rigorous way to determine the exact convergence error, since one cannot run infinitely long trajectories and explore the entire phase space. We believe that the best check for convergence is a test case (such as the one reported here) that compares free energy changes in very different environments, because the chance of accidental success is then quite small.

Our earlier attempt to calculate the energetics of the acidic groups of BPTI was based on the PDLD method. This study gave a similar trend but with a larger error range. The explanation for the improved agreement seems to reside in the fact that the X-ray structure of BPTI (on which the PDLD calculation was based) was obtained at a pH range where both Asp-3 and Glu-7 are ionized. On the other hand, the calculations reported in Figure 7 of Russell and Warshel (1985) correspond to the case in which one acid is ionized. In fact, the PDLD method gives better results when we calculate the ionization of one acid when the other is also ionized [Figure 9 of Russell and Warshel (1985)]. As opposed to the PDLD calculation, the present calculation allows the protein to relax from the initial structure (i.e., the X-ray structure) to one in which only one of the residues is ionized. Thus, the ionized Glu-7 (which is not stabilized sufficiently by the surrounding water molecules in the unrelaxed X-ray structure) gains significant stabilization from protein configurations that are polarized toward Glu-7 and allow water molecules to penetrate closer to the ionized acid.

The adiabatic charging method should not be viewed as the automatic method of choice in all cases. The method requires a very large amount of computer time, and the PDLD calculations are a factor of about 500 faster. Furthermore, the PDLD gives in a straightforward way the different contributions to the total electrostatic energy (induced dipoles, permanent dipoles, etc.). Our attempts to obtain these individual contributions with the method used here do not yet give converging results. It appears that the energy distribution (and other distribution functions) converges much more slowly than the total free energy.

The error range of our adiabatic charging process should be examined with longer simulation times and with exploration of more ionized groups. Nevertheless, it seems from this study that we are moving toward an error range of less than 2 kcal/mol. Such an error range might allow one to reach semiquantitative conclusions in studying enzymatic reactions where the catalytic energy is on the order of 7 kcal/mol. Furthermore, the error range in studies of site-directed mutagenesis (Warshel & Sussman, 1986) are expected to involve a smaller error range; the environmental changes induced by mutation are usually smaller than those studied in this work.

REFERENCES

Beveridge, D. L., Mezei, M., Ravishanker, G., & Jayaram, B. (1985) *J. Biosci. 8*, 167–178.

Deisenhofer, J., & Steigemann, W. (1975) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. 31*, 238–250.

Hill, T. L. (1956) *Statistical Mechanics*, pp 191–193, McGraw-Hill, New York.

Jorgensen, W. L., & Ravimohan, C. (1985) *J. Chem. Phys. 83*, 3050–3054.

Kirkwood, J. G. (1935) *J. Chem. Phys. 3*, 300–313.

Russell, S. T., & Warshel, A. (1985) *J. Mol. Biol. 185*, 389–404.

Torrie, G. M., & Valleau, J. P. (1977) *J. Comput. Phys. 23*, 187-199.
Warshel, A. (1978) *Chem. Phys. Lett. 55*, 454-458.
Warshel, A. (1982) *J. Phys. Chem. 86*, 2218-2224.
Warshel, A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 444-448.
Warshel, A., & Levitt, M. (1976) *J. Mol. Biol. 103*, 227-249.
Warshel, A., & Russell, S. T. (1984) *Q. Rev. Biophys. 17*, 283-422.
Warshel, A., & King, G. (1985) *Chem. Phys. Lett. 121*, 124-129.
Warshel, A., & Sussman, F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3806-3810.

Domain Organization of Chicken Gizzard Myosin Light Chain Kinase Deduced from a Cloned cDNA†

Vince Guerriero, Jr.,‡ Mario A. Russo,§ Norma J. Olson, John A. Putkey,∥ and Anthony R. Means*
Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030
Received April 23, 1986; Revised Manuscript Received August 14, 1986

ABSTRACT: Myosin light chain kinases (MLCK) are the most studied of the calmodulin-activated enzymes; however, minimal sequence information is available for the smooth muscle form of the enzyme. The production of an antibody against the enzyme and the use of expression vectors for constructing cDNA libraries have facilitated the isolation of a cDNA for this kinase. The derived amino sequence was found to contain a region of high homology (54%) to the rabbit skeletal muscle enzyme and also very significant homology (35%) to the catalytic subunit of phosphorylase *b* kinase and cGMP-dependent protein kinase. All of these homologies were found in the known catalytic domains of these enzyme, thus enabling us to predict the location of the catalytic domain for the chicken gizzard myosin light chain kinase. Within the catalytic domain a consensus sequence for an ATP-binding site was located. Subcloning and expression of different regions of the cDNA defined a 192 base pair fragment coding for the calmodulin-binding domain of MLCK. Both of the cAMP-dependent protein kinase phosphorylation sites were identified by sequence homology. A linear model for MLCK is presented placing the various domains in relative position. Northern blot analysis and S_1 protection and mapping experiments have revealed that the mRNA for MLCK is 5.5 kilobases in length, but there also exists a second mRNA of 2.7 kilobases that shares a high degree of homology with about 520 base pairs at the 3′ end of the cDNA for MLCK.

Myosin light chain kinases (MLCK) are Ca^{2+}-calmodulin-dependent enzymes that phosphorylate the regulatory light chain of myosin. In smooth muscle and nonmuscle tissues this phosphorylation is obligatory for the stimulation of myosin ATPase, which precedes tension development (Adelstein & Klee, 1980). Myosin light chain kinases have been isolated from a variety of sources including smooth muscle, skeletal muscle, cardiac muscle, platelets, and macrophages. The enzymes are substrate-specific and preferentially phophorylate light chains from homologous tissues. One common characteristic these enzymes share is the requirement of calmodulin for activity.

These enzymes are also phosphorylated by cAMP-dependent protein kinase. The smooth muscle (Conti & Adelstein, 1981) and nonmuscle forms (Nishikawa et al., 1984) of the enzyme are inhibited by phosphorylation due to a decreased affinity for calmodulin. The cardiac (Wolf & Hofman, 1980) and skeletal muscle (Edelman & Krebs, 1982) forms are also phosphorylated, but this has no effect on either enzyme activity or calmodulin affinity. Finally, the enzyme from the striated muscle of *Limulus* (Sellers & Harvey, 1984) is not phosphorylated by the catalytic subunit of cAMP-dependent protein kinase.

These differences in various myosin light chain kinase enzymes must arise from alterations in the primary structure of the proteins. To date, the only MLCK to be extensively sequenced is from rabbit skeletal muscle (Takio et al., 1985). The organization of the functional domains in this enzyme are similar to those predicted for the chicken smooth muscle MLCK from the proteolysis mapping studies of Foyt et al. (1985). As a further step in understanding the relationship between the primary structure and location of functional domains of these enzymes, we have cloned and sequenced a cDNA encoding the carboxy-terminal 60% of chicken gizzard MLCK. Comparison of this sequence with that of the rabbit skeletal enzyme reveals a high degree of homology in the predicted catalytic site. Indeed the similarity of this region with identical sized portions of other protein kinases suggests a common evolutionary history.

Materials and Methods

Antibody to MLCK was produced and purified as described by Guerriero et al. (1981). Protein A was iodinated by the Bolton-Hunter procedure (1973) as modified by Chafouleas et al. (1979).

Preparation of Poly(A+) RNA. The muscular portion of chicken gizzards was cut into small pieces (approximately 3-5

† This work was supported by National Institutes of Health Grant HD-07503 (A.R.M.), American Cancer Society Grant BC-326 (A.R.M.), Robert A. Welch Foundation Grant Q-611 (A.R.M.), and USPHS Grant RR-05425 (V.G.).
* Address correspondence to this author.
‡ Present address: Department of Animal Sciences, University of Arizona, Tucson, AZ 85721.
§ Present address: Instituto di Isologia ed Embriologia Generale, Universita di Roma, Rome, Italy.
∥ Present address: Department of Biochemistry and Molecular Biology, University of Texas Medical School, Houston, TX 77225.

0006-2960/86/0425-8372$01.50/0