

## Prediction of Secondary Ionization of the Phosphate Group in Phosphotyrosine Peptides

M. Wojciechowski,\* T. Grycuk,\* J. M. Antosiewicz,\* and B. Lesyng\*<sup>†</sup>

\*Department of Biophysics, Warsaw University, Żwirki i Wigury 93, 02-089 Warsaw, Poland, and the <sup>†</sup>Interdisciplinary Center for Mathematical and Computational Modelling, Warsaw University, Pawińskiego 5A, 02-106 Warsaw, Poland

**ABSTRACT** A computational approach, based on a continuum molecular electrostatics model, for the calculation of the  $pK_a$  values of secondary ionization of the phosphate group in phenyl phosphate derivatives is described. The method uses the ESP atomic charges of the mono-anionic and di-anionic forms of the ionizable phosphate group, computed with the use of the density functional method, and applies a new concept of the *model group*, being the reference state for the  $pK_a$  calculations. Both conformational flexibility and tautomeric degrees of freedom are taken into account in the calculations. The method was parameterized using experimentally available  $pK_a$  values of four derivatives of phenyl phosphates, and phosphotyrosine. Subsequently this parameterization was used to predict  $pK_a$  of the phosphate group in a short peptide Gly-Gly-Tyr(P)-Ala, and in a longer peptide consisting of 12 residues, the latter in water, and in a complex with a protein—phospholipase. The agreement between the computed and the experimental  $pK_a$  values is better than  $\pm 0.3$  pH units for the optimized solute dielectric constant of 11–13. This approach is promising and its extension to other phospho-amino acids is in progress.

### INTRODUCTION

Phosphate groups play a key role in the living world. They are present in proteins, nucleic acids, and other biomolecular systems, and no other residue appears to fulfill the multiple roles in molecular biology (Westheimer, 1987). Reversible phosphorylation of seryl, threonyl, tyrosyl, or histidyl residues is one of the most abundant covalent modifications of proteins which constitutes an ubiquitous mechanism for controlling and regulating intracellular processes (Greengard, 1978; Walaas and Greengard, 1991; Beltman et al., 1993; Kurosawa, 1994; Dekel, 1996; Patarca, 1996). A phosphate group bound to an amino acid side chain has either a single- or double-negative charge. The primary  $pK_a$  values of phosphoric acid and of phosphate mono-esters are  $\sim 2$ , which results in their at-least mono-anionic ionization state at physiological pH. The secondary  $pK_a$  of phosphoric acid is 7.2 according to one source (Weast, 1966) and 6.7 according to the other (Peacock and Nickless, 1969). Also, the secondary ionization constant of the phosphate group in well-solvated phosphate mono-esters is close to the value of 7. What is important, the total charge of the phosphate groups in phospho-amino acids in a complex molecular environment can be either  $-1$  or  $-2$ . A knowledge of whether the phosphate group is di-anionic or mono-anionic, in particular, if the ionization state is affected by the presence of ligands or other changes in the environment, is essential for understanding the role played by this group in catalytic events or in protein structure formation (Martinez-Liarte et al., 1992; Johnson and Barford, 1994; Yang et al., 1994;

Mavri and Vogel, 1996; Mestas and Lumb, 1999; Antosiewicz et al., 2000). Ionization state of the bound phosphate group in a protein or polypeptide can be determined by NMR spectroscopy methods (Kalbitzer and Rösch, 1981; Hoffmann et al., 1994) or a Fourier transform infrared spectroscopy (Sanchez-Ruiz and Martinez-Carrion, 1986). On the other hand, during the past several years a significant progress in the development of increasingly predictive theoretical models for protonation equilibria in proteins is observed (Bashford and Karplus, 1990; Yang et al., 1993; Antosiewicz et al., 1994, 1996; Sham et al., 1997; Mehler and Guarnieri, 1999). The previous theoretical approaches dealt rather exclusively with standard amino acids encountered in proteins. Inasmuch as phosphorylation is common in proteins, it would be desirable to also include phosphorylated side chains in the list of residues treated in standard modeling methods. In this study an approach for prediction of the secondary ionization constant of phosphotyrosine ( $pTyr$ ) in polypeptides and proteins is proposed.

### METHODS

#### Theoretical background

A number of methods for theoretical prediction of protonation equilibria in proteins use a concept of the *model compound*, which usually meant an isolated amino acid, and an assumption that the difference in protonation behavior of the model compound in an isolated state in solution, for which the ionization constant is assumed to be known, and the protonation behavior in the protein environment, is purely electrostatic in origin (Bashford and Karplus, 1990; Yang et al., 1993; Antosiewicz et al., 1994, 1996; Sham et al., 1997; Mehler and Guarnieri, 1999). Typically, the corresponding electrostatic energies are calculated for the model compound with frozen conformations as they appear in the protein structural file. A possibility of conformational movement of the protein was included in a different manner by some authors (You and Bashford, 1995; Zhou and Vijayakumar, 1997).

The present method makes one step further toward improvements of the prediction of proton equilibria in proteins. Namely, instead of the model

Submitted March 22, 2002, and accepted for publication August 29, 2002.

Address reprint requests to Bogdan Lesyng, Warsaw University, Pawińskiego 5A, Warsaw, Poland 02-106. Tel.: +48-22-874-9100; Fax: +48-22-874-9115; E-mail: lesyng@icm.edu.pl.

© 2003 by the Biophysical Society

0006-3495/03/02/750/07 \$2.00

compound, a model group is introduced with its  $pK_a(\text{model})$  which can be determined based on experimental  $pK_a$  values of small molecules containing the given group. The model group is characterized by the lowest possible conformational flexibility, and it is reduced only to those atoms which change their charge during deprotonation. A new understanding of  $pK_a(\text{model})$  was presented in full detail elsewhere (Gryczuk, 2002); therefore, only the main features of this approach are presented below.

The standard free enthalpy of proton dissociation can be expressed as follows:

$$\Delta G^0 = \Delta G^0(\text{model}) + \Delta \Delta G_e^0, \quad (1)$$

where the  $\Delta G^0(\text{model})$  term refers to the model system, and  $\Delta \Delta G_e^0$  represents the electrostatic correction to the standard free enthalpy  $\Delta G^0(\text{model})$ . The  $pK_a$  of the proton dissociation reaction is given as:

$$pK_a = pK_a(\text{model}) + \Delta pK_a, \quad (2)$$

where the contribution  $\Delta pK_a = \Delta \Delta G_e^0 / 2.303RT$  arises from the averaged interaction between the model system and the rest of the molecule, and the  $pK_a(\text{model}) = \Delta G^0(\text{model}) / 2.303RT$  refers to the model system. For a molecule with fixed conformation the following is valid:

$$\Delta \Delta G_e^0 = \Delta G_{e,u}^0 - \Delta G_{e,p}^0, \quad (3)$$

where  $\Delta G_{e,x}^0$  ( $x = p$  or  $u$  for the protonated and deprotonated state, subsequently) represents the electrostatic work of moving the model group from solvent to the molecule. A meaningful relation between  $pK_a$  and  $pK_a(\text{model})$  can be obtained if  $\Delta \Delta G_e^0$  is computed taking into account the representative ensemble of conformations of the molecule. The free enthalpy change,  $\Delta \Delta G_e^0$ , is defined as in Eq. 3, but now the contributions  $\Delta G_{e,x}^0$  represent the averaged electrostatic corrections for both protonation forms, and they are calculated as follows:

$$\Delta G_{e,x}^0 = -RT \ln \sum_k e^{-\Delta G_{e,x}^0(k)/RT}, \quad (4)$$

therefore, finally:

$$pK_a = pK_a(\text{model}) - \frac{1}{2.303} \ln \frac{\sum_k e^{-\Delta G_{e,u}^0(k)/RT}}{\sum_k e^{-\Delta G_{e,p}^0(k)/RT}}. \quad (5)$$

The summation extends over a representative ensemble of conformers sampled from the trajectory of the Molecular Dynamics simulation with suppressed electrostatics. The form of Eq. 5 indicates that contribution of each conformer to the relation between  $pK_a$  and  $pK_a(\text{model})$  is weighted by the electrostatic work of moving the model group from solvent to the molecule,  $\Delta G_{e,x}^0(k)$  for the  $k$ -th conformer. This equation allows one to estimate the  $pK_a(\text{model})$  based on the experimental  $pK_a$  of a small molecule containing the model group.

The calculated value of  $\Delta \Delta G_e^0$  depends on atomic partial charges and radii of the solute molecule, as well as on the solute and solvent dielectric constants and definition of the dielectric boundary between solute and solvent. For each set of these parameters and definition of the boundary, the  $pK_a(\text{model})$  reproducing the experimental  $pK_a$  values of a given molecule can be determined. In practice partial charges and radii of atoms are established independently. This leaves a solute dielectric constant and the  $pK_a(\text{model})$  as the only adjustable parameters in the model being applied for prediction of protonation equilibria based on the Poisson-Boltzmann model of solute-solvent systems.

Having experimental  $pK_a$  values of a number of small test molecules containing the same model group, one can choose the solute dielectric constant and corresponding  $pK_a(\text{model})$  according to Eq. 5 which give the best reproduction of the experimental data for all the test molecules. Such a procedure leads to solute dielectric constant substantially above values expected on the basis of only electronic polarization contribution. It should be stressed that the resulting  $pK_a(\text{model})$  and the solute dielectric constant are not direct, physical observables and their values are adjustable parameters.

## Atomic partial charges and force field parameters

Calculation of the electrostatic energy requires atomic charges and radii. The atomic radii were taken from the CHARMM force field. Charge parameterization for phosphotyrosine in two protonation states of the phosphate group is available (Feng et al., 1996), but the corresponding partial charges data for phenyl phosphate derivatives is not. Therefore, in this study a set of atomic charges that fit the electrostatic potential in the gas phase, so-called ESP charges, were computed using the Gaussian94 package (Frisch et al., 1995). The density functional method, with the B3LYP exchange and correlation functional and the 6-31<sup>2+</sup><sub>g(d,p)</sub> basis set, was chosen.

The ESP charges were computed for all phenyl phosphate derivatives and for phosphotyrosine, for both ionization states of the phosphate group, i.e., the single- and double-ionized states. The calculations were carried out for a set of molecular conformations of each molecule covering the available conformational space of the system. The resulting charges were averaged over all conformations. Some charges were additionally averaged due to the local symmetry, e.g., CD1 and CD2 carbon atoms of the benzene ring. Such approximation removes problems of predefined asymmetry.

One should note that the ESP charges were used exclusively in the Poisson-Boltzmann calculations. The Molecular Dynamics simulations were carried out with electrostatics switched off (see the next paragraph), and the remaining parameters for these molecules were taken from the CHARMM force field. Based on the computed ESP charges, the model group for phenyl phosphate and its parameterization was proposed (see ESP charges in Results).

## Generation of an ensemble of structures

For all molecules considered in this study but the phospholipase-dodecapeptide complex, an ensemble of structures was generated by Langevin Dynamics method (Widmalm and Pastor, 1992) using the CHARMM v2.26 software (Brunger and Karplus, 1988). The electrostatic interactions were switched off, therefore the sampled structures were not weighted by electrostatic energies. The simulation procedure consists of a short minimization, a 6-ps equilibration, and the 200-ns dynamics. The Langevin Dynamics was carried out with a 2-fs time step and the friction coefficient  $FBETA = 50 \text{ ps}^{-1}$ . To avoid translational movement of the whole system, a harmonic constraint was applied to the central atom. After each 100 ps of dynamics, the current structure was saved, leading finally to an ensemble of 2000 conformers. Adding new conformers did not change the value of the  $pK_a(\text{model})$ . In the case of phospholipase-dodecapeptide complex, a set of 18 NMR-based structures was used (Pascal et al., 1994), available in the Protein Data Bank (PDB; Bernstein et al., 1977) under access code 2PLE.

## Parameterization of the $pK_a(\text{model})$ versus solute dielectric constant

For each structure sampled from the Langevin trajectory, the electrostatic contribution to the free enthalpy was calculated using the UHBD software (Davis et al., 1991). The method is based on a continuum dielectric model, and the linear form of the Poisson-Boltzmann equation. The equation was solved using a finite difference method (FDPB). We used standard parameters for the solvent dielectric constant (78.0 for water), solvent probe radius of 1.4 Å, ion strength of 150.0 mM, and ion exclusion radius of 2.0 Å. There is no standard value of the solute dielectric constant for proteins; its value changes depending on the local environment, the approximations made, and the atom charges applied (Gilson and Honig, 1986; Nakamura et al., 1988; Sham et al., 1997; Schutz and Warshel, 2001). To find the most appropriate solute dielectric constant for the phosphate group in phenyl phosphates and related compounds, the four phenyl phosphate derivatives and phosphotyrosine were parameterized as follows. The  $\Delta G_{e,x}^0$  contributions

(see Eq. 4) for all these molecules, for the protonated and deprotonated states of their titratable sites, were computed for the solute dielectric constant ranging from 5 to 20. These energies allow parameterization of the  $pK_a(\text{model})$  as a function of the solute dielectric constant, using Eq. 5.

Experimental data for all the molecules are presented in Table 1. The experimental data for the first five molecules were used for parameterization, and the data for the last three molecules were used for comparison between the experiment and our theoretical prediction.

Two sets of parameters were determined. First, four phenyl phosphate derivatives were used as reference data. For each molecule of the four,  $pK_a(\text{model})$  was calculated. Subsequently the four dependencies were averaged. The purpose was to obtain  $pK_a(\text{model})$ , which is independent of small changes in parameterization of the model group. By assumption, the model group in each of both protonation states has the same characteristics in any environment. Second, the  $pK_a$  values of phosphotyrosine were used for parameterization. There are two published values, 5.80 (Vogel, 1989) and 5.55 (Robitaille et al., 1991), for the experimental  $pK_a$  values for phosphotyrosine. Parameterizations were performed for both of them, independently.

### Application to the tetra- and dodecapeptide and to a complex of the dodecapeptide with phospholipase C- $\gamma_1$

The methodology described above was applied to predict protonation equilibria for phosphotyrosine in two polypeptide chains, the tetrapeptide (Gly-Gly-pTyr-Ala) and the dodecapeptide (Asp-Asn-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Asp-Pro-Lys), and for a complex of the latter with the C-terminal SH2 domain of phospholipase C- $\gamma_1$ . The  $pK_a$  value for phosphotyrosine in the tetrapeptide was determined as 5.90 (Hoffmann et al., 1994). The  $pK_a$  of 6.1 was determined for phosphotyrosine in the free dodecapeptide, whereas in the complex with the SH2 domain the value of 4.0 was measured (Singer and Forman-Kay, 1997).

## RESULTS

### ESP charges

All model compounds chosen for parameterization contain the same model group. It is composed of the whole phosphate group and the three carbons from the tyrosine ring, i.e., the CZ, CE1, and CE2 atoms. Table 2 presents the computed, ESP, and B3LYP//6-31<sup>2+</sup> $g(d,p)$  partial charges

**TABLE 2** Atomic charges in the model group

Atom name	Mono-anionic form	Di-anionic form
H	0.42	—
O3	−0.73	−1.06
P	1.50	1.85
O1	−0.92	−1.06
O2	−0.92	−1.06
OZ	−0.50	−0.72
CZ	0.35	0.45
CE1	−0.10	−0.20
CE2	−0.10	−0.20

of the atoms forming the model group in the phenyl phosphate derivatives and related compounds. These charges were common for all the derivatives used for parameterization, and several charges of atoms, like phosphate oxygens, were averaged, assuming the local symmetry. The computed charges, using the 6-31<sup>2+</sup> $g(d,p)$  basis set, are quite large; e.g., at the phosphorus they are +1.50 and +1.85 for the protonated and deprotonated state, respectively. Note that the more negative phosphate group has the more positive phosphorus atom.

### Dependence of $pK_a(\text{model})$ on the solute dielectric constant $\epsilon$

Table 3 presents the computed  $pK_a(\text{model})$  values as functions of the solute dielectric constant, together with the average value over the four phenyl phosphate derivatives used in the parameterization. The differences between the computed dependencies for the phenyl phosphate derivatives can be considered as a measure of errors of our method resulting from the uncertainty of experimental data and approximations of the model. One should note that such approximations are inherent to all molecular modeling methods.

Table 4 presents dependence of the computed  $pK_a(\text{model})$  as a function of the solute dielectric constant, based on two

**TABLE 1** Experimental  $pK_a$  values for the examined molecules

Molecular system	$pK_a$	Ref.*
Phosphoric acid monophenyl ester	5.83	[1–5]
Phosphoric acid mono-(4-bromo-phenyl)	5.44	[1]
p-Nitrophenyl phosphate	5.20	[4]
Phosphoric acid mono-(4-chloro-phenyl ester)	5.84	[6,7]
Phosphotyrosine	5.55	[8]
Phosphotyrosine	5.80	[9]
Short peptide, Gly-Gly-Tyr(P)-Ala	5.90	[10]
Dodecapeptide	6.10	[11]
Dodecapeptide-phospholipase complex	4.0	[11]

\*References cited in the Table: [1], King and Delory, 1939; [2], Chanley and Feageson, 1955; [3], Bourne and Williams, 1984; [4], Kimura et al., 1997; [5], Gupta et al., 1987; [6], Maguire and Shaw, 1953; [7], Phillips, 1958; [8], Robitaille et al., 1991; [9], Vogel, 1989; [10], Hoffmann et al., 1994; [11], Singer and Forman-Kay, 1997.

**TABLE 3** The  $pK_a(\text{model})$  as a function of the solute dielectric constant for the phenyl phosphate derivatives used in this work, and the average value of the  $pK_a(\text{model})$

Dielectric constant	Phenyl phosphate derivatives				Average $pK_a(\text{model})$
	Br	Cl	H	NO <sub>2</sub>	
5	5.41	5.78	5.53	5.68	5.60
6	5.34	5.71	5.50	5.53	5.52
7	5.30	5.67	5.48	5.43	5.47
8	5.26	5.64	5.47	5.35	5.43
9	5.24	5.62	5.46	5.29	5.40
10	5.22	5.60	5.45	5.24	5.38
12	5.19	5.57	5.44	5.17	5.34
14	5.17	5.55	5.43	5.12	5.32
15	5.16	5.54	5.43	5.09	5.31
17	5.14	5.52	5.42	5.04	5.28
20	5.11	5.50	5.41	4.98	5.25

**TABLE 4** The  $pK_a(\text{model})$  as a function of the solute dielectric constant for phosphotyrosine. Two available experimental values are taken into account

Dielectric constant	Phosphotyrosine $pK_{a(\text{exp})} = 5.80^*$	$pK_a(\text{model})$ $pK_{a(\text{exp})} = 5.55^\dagger$
4	5.03	4.78
5	5.08	4.83
6	5.13	4.88
7	5.16	4.91
8	5.18	4.93
9	5.19	4.94
10	5.19	4.94
12	5.21	4.96
14	5.22	4.97
17	5.24	4.99
20	5.27	5.02

\*Robitaille et al., 1991;  $^\dagger$ Vogel, 1989

available experimental  $pK_a$  values for phosphotyrosine, in the  $\epsilon$  range from 5 to 20. It can be noted that both parameterizations are different from the data obtained for the four derivatives of phenyl phosphate. However, the value of the  $pK_a(\text{model})$  for the phosphotyrosine data obtained for the measured  $pK_a$  value of 5.80 and the average dependence derived from the four phenyl phosphate derivatives are close, starting from  $\epsilon = 14$ .

It can also be noted that  $pK_a(\text{model})$  is the decreasing function of  $\epsilon$  for the phenyl phosphate derivatives and it is the increasing function for phosphotyrosine. Different dependence of  $pK_a(\text{model})$  in the phenyl phosphate derivatives and in phosphotyrosine can be explained by interactions of the model group with the charges in its neighborhood, i.e., CD1, CD2, HD1, and HD2. For the phenyl phosphate derivatives these atoms have the net positive charge close to 0. For phosphotyrosine the net charge is  $\sim -0.1$ . The negative charge at the nearest neighborhood of the protonation site makes the  $pK_a(\text{model})$  lower than a positive charge would, and that effect should gradually disappear with the increase of  $\epsilon$ .

### Predicted $pK_a$ values for tetra- and dodecapeptide

The tetrapeptide and the uncomplexed dodecapeptide were used to check applicability of our method for prediction of  $pK_a$  values. Regarding conformational flexibility, those molecules were treated in the same way as described above. For each of them, 2000 structures were sampled from the Langevin Dynamics trajectories with the electrostatic interactions switched off. The  $pK_a(\text{model})$  value, based on the four phenyl phosphate derivatives or phosphotyrosine, with the corresponding solute dielectric constant, were used to calculate  $pK_a$  of the phosphate group in these peptides (see Eq. 4).

Calculations for the dodecapeptide complex with phospholipase were carried out for several sets of structures out of

18 deposited in the PDB file (Pascal et al., 1994), beginning from the first six structures up to all 18 structures. This procedure was applied because proper averaging would require knowledge of the probabilities of structures and this information was not available. Nevertheless, starting from the first six structures up to all 18 from the PDB file, the best solute dielectric constant is always in the range of 11–14. All parameterizations of the  $pK_a(\text{model})$ , i.e., two for phosphotyrosine and the third for phenyl phosphate derivatives, give similar accuracy in prediction of the  $pK_a$  values.

The  $pK_a(\text{model})$  parameterization derived from the  $pK_a$  value of 5.55 for phosphotyrosine gives the best agreement with the experimental  $pK_a$  for the dodecapeptide-phospholipase complex, regardless of the number of structures used in the calculations. However, the same parameterization gives worse agreement with the experiment for tetrapeptide and dodecapeptide in water. The remaining two parameterizations give values of  $pK_a$  close to each other. These parameterizations give a similar level of accuracy for dodecapeptide complexed with protein as the first parameterization, and moreover, give better agreement with the experimental  $pK_a$  values of the peptides in aqueous environment.

Tables 5–7 show the best results together with the corresponding solute dielectric constant, along with the results for the most optimal solute dielectric constant, both for each set of parameters. The results of the calculations based on the first eight structures from the PDB file for the peptide-protein complex are presented in Tables, as well as in Fig. 1.

Note that the values of the optimal solvent dielectric constant for each of the molecules differ, and spread in range from 7 to 20. The  $pK_a$  values obtained for the solute dielectric constant from this range differ from that obtained for the solute dielectric constant of 12 not more than 0.25 pH unit. That difference could be considered as an error arising from the applied approximations and it is close to the experimental error. Inclusion of the conformational flexibility does not lower the solute dielectric constant to the value of  $\sim 2$ , which could be interpreted as arising from the effect of electronic polarization. High value of the solute dielectric constant correlates also with previous results (Gilson and

**TABLE 5** The  $pK_a$  determined experimentally and computed, based on the parameterization for the model compounds

Molecule	Experimental $pK_a$	Computed $pK_a$ for $\epsilon = 12$	Optimal $pK_a(\epsilon)$
Tetrapeptide	5.9	5.75	5.90 ( $\epsilon = 7$ )
Dodecapeptide:			
Uncomplexed	6.1	6.35	6.12 ( $\epsilon = 20$ )
Complexed	4.0	3.93	3.93 ( $\epsilon = 12$ )

The results for the optimal  $\epsilon = 12$  are presented in column 3. The best results for particular calculations are presented in column 4. The solute dielectric constants used in the calculations are given in the parentheses.

**TABLE 6** The  $pK_a$  determined experimentally and computed, based on the parameterization for phosphotyrosine with the experimental  $pK_a = 5.55$

Molecule	Experimental $pK_a$	Computed $pK_a$ for $\epsilon = 12$	Optimal $pK_a(\epsilon)$
Tetrapeptide	5.9	5.36	5.37 ( $\epsilon = 20$ )
Dodecapeptide:			
uncomplexed	6.1	5.66	5.80 ( $\epsilon = 20$ )
complexed	4.0	3.71	4.16 ( $\epsilon = 13$ )

The results for the optimal  $\epsilon = 12$  are presented in column 3. The best results for particular calculations are presented in column 4. The solute dielectric constants used in the calculations are given in the parentheses.

**TABLE 7** The  $pK_a$  determined experimentally and computed, based on the parameterization for phosphotyrosine with the experimental  $pK_a = 5.80$

Molecule	Experimental $pK_a$	Computed $pK_a$ for $\epsilon = 12$	Optimal $pK_a(\epsilon)$
Tetrapeptide	5.9	5.61	5.63 ( $\epsilon = 20$ )
Dodecapeptide:			
uncomplexed	6.1	6.14	6.11 ( $\epsilon = 20$ )
complexed	4.0	3.74	4.10 ( $\epsilon = 10$ )

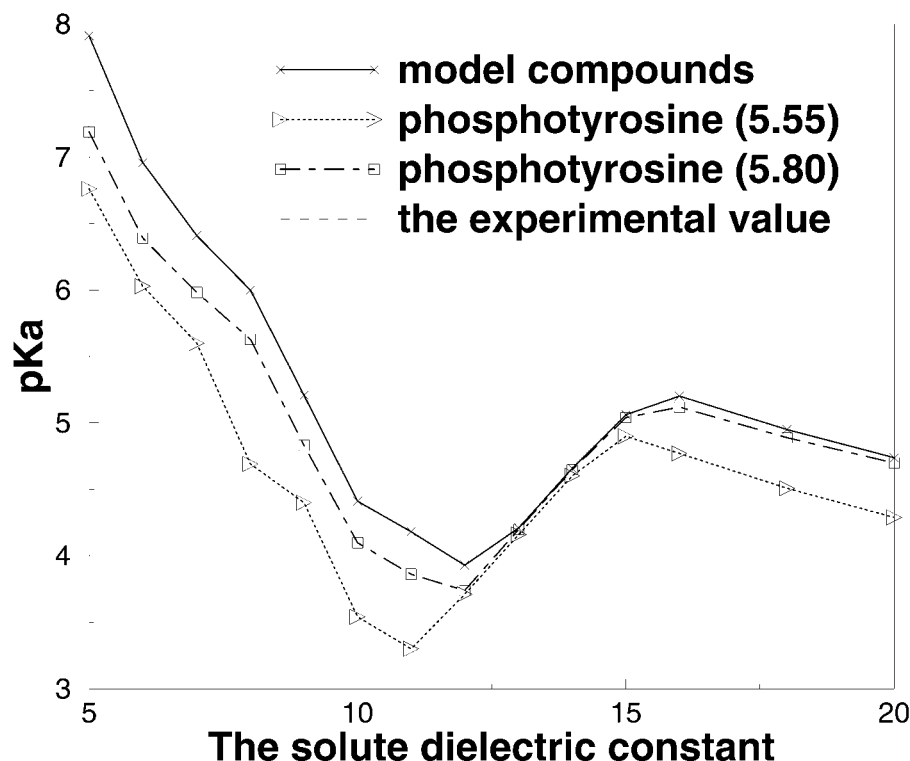
The results for the optimal  $\epsilon = 12$  are presented in column 3. The best results for particular calculations are presented in column 4. The solute dielectric constants used in the calculations are given in the parentheses.

Honig, 1986; Nakamura et al., 1988; Sham et al., 1997; Schutz and Warshel, 2001). One of the present authors (T.G.) also obtained the high solute dielectric constant (12.0) using the CHARMM charges exclusively (Grycuk, 2002). Therefore, it should be noted that arriving at the optimal solute dielectric constant of  $\sim 12$  is not a result of using the ESP charges for the titrated phospho residues and the CHARMM charges for other standard amino acids; it is the property of the model. This is quite an important observation because the ESP charges can be easily computed, using ab initio methods, whereas further optimization of the charges (as well as other force field parameters) is not a straightforward procedure.

Fig. 1 presents results for the dodecapeptide complexed with phospholipase for each set of parameters, as functions of the solute dielectric constant.

## DISCUSSION

Answering the question whether a phosphate group is single or doubly ionized in a real molecular environment, for example being bound to a protein molecule, is of key importance for studying and understanding the role of phosphorylation processes in activation of biomolecules. In particular, in our previous studies (Błachut-Okrańska et al., 1999; Antosiewicz et al., 2000) we have shown that the ionization state of phosphate groups may depend, or be coupled to, conformational changes of phosphorylated proteins. Information on ionization states of the phosphate groups is also required when analyzing mechanisms of the phosphorylation processes by kinases and phosphatases at the atomic level. The method proposed in this study allows us to predict  $pK_a$  values of the secondary ionization of phosphate groups bound to tyrosines in peptides, with



**FIGURE 1** Dependence of the predicted  $pK_a$  value for phosphotyrosine in the dodecapeptide complexed with phospholipase, on the solute dielectric constant for the three parameterizations described in the text.

accuracy of  $\sim \pm 0.3$  pH units, when the solute dielectric constant of 11–13 is assumed. This promising approach is now being extended to other phosphorylated peptides; in particular, those which contain phosphorylated serine and/or threonine.

It should be noted that although we applied the ESP charges for the phosphorylated molecules, rather than the CHARMM charges, not available for these compounds, it seems to have a little influence on the results of the present work, i.e., on the achieved agreement with the experimental  $pK_a$  values and the estimated best solute dielectric constant. A very similar solute dielectric constant was obtained based exclusively on molecules fully parameterized with the CHARMM force field (Grycuk, 2002). This suggests that the procedure described here is of quite general applicability.

Summarizing, the main steps of the applied procedure are the following:

1. Search for experimental  $pK_a$  values for a model compound (e.g., amino acid), possibly closest to your titratable subunit.
2. Calculate the ESP charges for all possible protonation states of any new titratable subunit present in your biopolymer. If parameterization for all subunits is available, skip this point.
3. Define the model group within the titratable subunit which is described by the lowest possible conformational flexibility, and reduced exclusively to those atoms which change their charge during deprotonation.
4. Perform the Langevin Dynamics simulation for the model compound with the suppressed electrostatic interactions, to generate the representative ensemble of structures (here, 2000 structures).
5. Perform the Poisson-Boltzmann calculations to determine the electrostatic contributions  $\Delta G_{e,x}^0(k)$  for the ensemble of structures in both protonation states and for a selected dielectric constant of the solute molecule. The solvent dielectric constant assumes its experimental value (78 for water).
6. Apply Eq. 5 to find the  $pK_a(\text{model})$  for the selected solute dielectric constant.
7. If you want to optimize the solute dielectric constant (here  $\sim 12$ ), repeat points 5 and 6 for different dielectric constants.

This point completes the parameterization procedure. Now, the computed  $pK_a(\text{model})$  may be used along with  $pK_a(\text{model})$  of other groups when performing multiple titration procedure for the biopolymer.

This work was supported by the Ministry of National Education, Poland (BW 1482/BF) and by the State Committee for Scientific Research, Poland (8T11F01616).

## REFERENCES

Antosiewicz, J., E. Błachut-Okrasińska, T. Grycuk, and B. Lesyng. 2000. A correlation between protonation equilibria in biomolecular systems and

their shapes: studies using a Poisson-Boltzmann model. In GAKUTO International Series, Mathematical Science and Applications, vol. 14. N. Kenmochi, editor. Gakkotosho Co., Ltd., Tokyo, Japan. 11–17.

Antosiewicz, J., J. M. Briggs, A. E. Elcock, M. K. Gilson, and J. A. McCammon. 1996. Computing the ionization states of proteins with a detailed charge model. *J. Comp. Chem.* 17:1633–1644.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Bashford, D., and M. Karplus. 1990.  $pK_a$ 's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry.* 29:10219–10225.

Beltman, J., W. K. Sonnenburg, and J. A. Beavo. 1993. The role of protein phosphorylation in the regulation of cyclic nucleotide phosphodiesterases. *Mol. and Cellular Biochem.* 128:239–253.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. J. Tasumi. 1977. The protein data bank: a computer-based archival file for molecular structures. *J. Mol. Biol.* 123:557–594.

Błachut-Okrasińska, E., B. Lesyng, J. M. Briggs, J. A. McCammon, and J. M. Antosiewicz. 1999. Poisson-Boltzmann model studies of molecular electrostatic properties of the cAMP-dependent protein kinase. *Eur. Biophys. J.* 28:457–467.

Bourne, N., and A. Williams. 1984. Effective charge on oxygen in phosphoryl (-PO) group transfer from an oxygen donor. *J. Org. Chem.* 49:1200–1204.

Brunger, A. T., and M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins: Struct. Func. Gen.* 4:148–156.

Chanley, J. D., and E. Feagenson. 1955. The mechanism of the hydrolysis of organic phosphates. III. Aromatic phosphates. *J. Am. Chem. Soc.* 77:4002–4004.

Davis, M. E., J. D. Madura, B. A. Luty, and J. A. McCammon. 1991. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp. Phys. Commun.* 62:187–197.

Dekel, N. 1996. Protein phosphorylation/dephosphorylation in the meiotic cell cycle of mammalian oocytes. *Rev. Reprod.* 1:82–88.

Feng, M., M. Philippopoulos, A. D. MacKerell, Jr., and C. Lim. 1996. Structural characterization of the phosphotyrosine binding region of a high-affinity domain-phosphopeptide complex by molecular dynamics simulation and chemical shift calculations. *J. Am. Chem. Soc.* 118: 11265–11277.

Frisch, M. J., G. W. Trucks, H. B. Schegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople. 1995. Gaussian 94, Rev. E2. Gaussian, Inc., Pittsburgh, Pennsylvania.

Gilson, M. K., and B. H. Honig. 1986. The dielectric constant of a folded protein. *Biopolymers.* 25:2097–2119.

Greengard, P. 1978. Phosphorylated proteins as physiological effectors. *Science.* 199:146–152.

Grycuk, T. 2002. Revision of the model system concept for the prediction of  $pK_a$ 's in proteins. *J. Phys. Chem. B.* 106:1434–1445.

Gupta, S. C., N. B. Islam, D. L. Whalen, H. Yagi, and D. M. Jerina. 1987. Bifunctional catalysis in the nucleotide-catalysed hydrolysis of ( $\pm$ ) -7  $\beta$ ,8 $\alpha$ -dihydroxy-9 $\alpha$ ,10 $\alpha$ -epoxy-7,8,9,10- tetrahydrobenzo[a]pyrene. *J. Org. Chem.* 52:3812–3815.

Hoffmann, R., I. Reichert, W. O. Wachs, M. Zeppezauer, and H. R. Kalbitzer. 1994.  $^1\text{H}$  and  $^{31}\text{P}$  NMR spectroscopy of phosphorylated model peptides. *Int. J. Pept. Protein Res.* 44:193–198.

Johnson, L. N., and D. Barford. 1994. Electrostatic effects in the control of glycogen phosphorylase by phosphorylation. *Protein Sci.* 3:1726–1730.

- Kalbitzer, H. R., and P. Rösch. 1981. The effect of phosphorylation of the histidyl residue in the tetrapeptide Gly-Gly-His-Ala. Changes of chemical shift and pK values in  $^1\text{H}$ - and  $^{31}\text{P}$ -NMR spectra. *Organic Magnet. Reson.* 17:88–91.
- Kimura, E., S. Aoki, T. Koike, and M. Shiro. 1997. A Tris(ZnII-1,4,7,10-tetraazacyclododecane) complex as a new receptor for phosphate dianions in aqueous solution. *J. Am. Chem. Soc.* 119:3068–3076.
- King, E. J., and G. E. Delory. 1939. The rates of enzymatic hydrolysis of phosphoric esters. *Biochem. J.* 33:1185–1188.
- Kurosawa, M. 1994. Phosphorylation and dephosphorylation of protein in regulating cellular function. *J. Pharmacol. Toxicol. Rev.* 31:135–139.
- Maguire, M. H., and G. Shaw. 1953. Synthetic plant hormones. I. Some esters of phosphoric acid. *J. Chem. Soc.* 1479–1482.
- Martinez-Liarte, J. H., A. Iriarte, and M. Martinez-Carrion. 1992. Inorganic phosphate binding and electrostatic effects in the active center of aspartate aminotransferase apoenzyme. *Biochemistry.* 31:2712–2719.
- Mavri, J., and H. J. Vogel. 1996. Ion pair formation of phosphorylated amino acids and lysine and arginine side chains: a theoretical study. *Proteins: Struct. Funct. Genet.* 24:495–501.
- Mehler, E. L., and F. Guarnieri. 1999. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.* 77:3–22.
- Mestas, S. P., and K. J. Lumb. 1999. Electrostatic contribution of phosphorylation to the stability of the CREB-CBP activator-coactivator complex. *Nat. Struct. Biol.* 6:613–614.
- Nakamura, H., T. Sakamoto, and A. Wada. 1988. A theoretical study of the dielectric constant of protein. *Protein Eng.* 2:177–183.
- Pascal, S. M., A. U. Singer, G. Gish, T. Yamazaki, S. E. Shoelson, T. Pawson, L. E. Kay, and J. D. Forman-Kay. 1994. Nuclear magnetic resonance structure of an sh2 domain of phospholipase c-gamma 1 complexed with a high affinity binding peptide. *Cell.* 77:461–472.
- Patarca, R. 1996. Protein phosphorylation and dephosphorylation in physiologic and oncologic processes. *Crit. Rev. Oncog.* 7:343–432.
- Peacock, C. J., and G. Nickless. 1969. The dissociation constants of some phosphorus(V) acids. *Z. Naturforsch.* 24a:245–247.
- Phillips, J. N. 1958. Strength of chloro-substituted phenoxyacetic and related phosphorus-containing acids. *J. Chem. Soc.* 4271–4276.
- Robitaille, P. L., P. A. Robitaille, G. G. Brown, Jr., and G. G. Brown. 1991. An analysis of the pH-dependent chemical-shift behavior of phosphorus-containing metabolites. *J. Mag. Res.* 92:73–84.
- Sanchez-Ruiz, J. M., and M. Martinez-Carrion. 1986. Ionization state of the coenzyme 5'-phosphate ester in cytosolic aspartate aminotransferase. A Fourier transform infrared spectroscopic study. *Biochemistry.* 25:2915–2920.
- Schutz, C. N., and A. Warshel. 2001. What are the dielectric “constants” of proteins and how to validate electrostatic models. *Proteins: Struct. Funct. Genet.* 44:400–417.
- Sham, Y. Y., Z. T. Chu, and A. Warshel. 1997. Consistent calculations of pK(a)'s of ionizable residues in proteins: Semi-microscopic and microscopic approaches. *J. Phys. Chem.* 101:4458–4472.
- Singer, A. U., and J. D. Forman-Kay. 1997. pH titration studies of an SH2 domain-phosphopeptide complex: Unusual histidine and phosphate  $pK_a$  values. *Protein Sci.* 6:1910–1919.
- Vogel, H. J. 1989. Phosphorus-31 NMR of phosphoproteins. *Methods Enzymol.* 177:263–282.
- Walaas, S. I., and P. Greengard. 1991. Protein phosphorylation and neuronal function. *Pharmacol. Rev.* 43:299–349.
- Weast, R. C. editor. 1966. CRC Handbook of Physics and Chemistry, 47th ed. CRC Press, Cleveland, Ohio.
- Westheimer, F. H. 1987. Why nature chose phosphates. *Science.* 235:1173–1178.
- Widmalm, G., and R. W. Pastor. 1992. Comparison of Langevin and molecular dynamics simulations—equilibrium and dynamics of ethylene glycol in water. *J. Chem. Soc. Faraday Trans.* 88:1747–1754.
- Yang, A. S., M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. 1993. On the calculation of  $pK_a$ s in proteins. *Proteins: Struct. Func. Gen.* 15:252–265.
- Yang, H. L. S. Z., G. Zhi, A. J. T. Stull, and K. M. Trybus. 1994. Charge replacement near the phosphorylatable serine of the myosin regulatory light chain mimics aspects of phosphorylation. *Proc. Natl. Acad. Sci. USA.* 91:1490–1494.
- You, T. J., and D. Bashford. 1995. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* 69:1721–1733.
- Zhou, H., and M. Vijayakumar. 1997. Modeling of protein conformational fluctuations in  $pK_a$  predictions. *J. Mol. Biol.* 267:1002–1011.