Theor Chem Acc (1999) 101:159–162
DOI 10.1007/s002149800m77

Theoretical
Chemistry Accounts

Regular article

Theoretical pK_a calculations of proteins; the tyrosine and lysine residues of β-elicitin*

Hans J. Vogel, André H. Juffer

Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4

Received: 24 April 1998 / Accepted: 4 August 1998 / Published online: 11 November 1998

Abstract. Elicitins are small proteins that are secreted by plant pathogenic fungi. In this work we have used a computer program that utilizes the boundary element method for heterogeneous dielectrics with ionic strength to calculate the pK_a of all titrating groups in the 98-residue protein β-cryptogein. Our results are in reasonable agreement with the experimentally determined pK_a values for the Tyr residues in the protein. We find that the functionally important Lys13 residue has a normal pK_a of 10.3. Our work also shows that there is no direct correlation between the exposure of an amino acid sidechain and its pK_a.

Keywords: β-Elicitin – pK_a Calculations – Continuum electrostatics, boundary element method

1 Introduction

Of the 20 common amino acids of which all proteins are made, six possess an ionizable group in their sidechains. These titratable groups can be protonated or deprotonated depending on pH (usually between 2 and 12 pH units). In addition, the free amino- and carboxy-terminal ends of the polypeptide chain can also be titrated. As a result, the behaviour of proteins resembles in many ways that of polyelectrolyte species. The charge of a protein's amino acid sidechains has a marked effect on it's stability, folding, interactions and activity. The apparent pK_a of each titratable site is influenced by the micro-environment provided by the protein's structure. The latter reflects inter-residue, residue-solvent (water) and long-range electrostatic interactions with other charged residues in the protein or salt ions in solution. While there are many experimental methods that allow one to determine the overall titration curve of a protein, only a few experimental spectroscopic methods have sufficient resolution (e.g. NMR spectroscopy) that they allow for the determination of pK_a values of individual titrating sites in a protein [1]. Yet in order to understand a protein's structural and functional properties, such detailed knowledge is required.

Recently, a number of computational approaches have been introduced, which allow for the calculation of the pH-dependant behaviour of individual titrating sites in protein molecules, e.g. [2–6]. These computer programs all require the input of a detailed three-dimensional crystal or solution structure. The various methods all make different assumptions, but many of them show impressive agreements with experiment. In particular, it has been found that better agreement with experimentally determined pK_a values is obtained when the dielectric constant of the protein is chosen as 20–40 [4, 6]. A similar value has also been deduced from molecular dynamics simulations of several proteins and it seems to give a good representation of the dielectric properties of the protein [7–9]. It should be realized that protein folding places the vast majority of the charged residues on the surface of the protein, an environment that is not well represented by a dielectric constant of 2–4, which is typical for organic non-polar solvents. In addition, the shielding of protein charges, that can arise from the presence of anions and cations in the solution around a protein, were also shown to have an important effect on the titration properties of protein groups. We have developed a computer program [6] that includes ionic strength, and hence the effect of screening can be assessed by performing calculations at different ionic concentrations. Our method incorporates the boundary element method (BEM [10, 11]) to describe all the electrostatic interactions accurately, rather than the finite difference method which is computationally less intensive, and has therefore been more frequently used.

In this contribution we have calculated the pK_a values for all titratable groups in the small protein β-cryptogein, a representative protein of the β-elicitin class. Elicitins are highly conserved 98-residue proteins that

*Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

Correspondence to: H.J. Vogel
e-mail: vogel@acs.ucalgary.ca

are secreted by the plant-pathogenic fungi *Phytophtora* and *Phytium*. They are a novel class of toxic proteins that elicit a strong defence response in the plant host. They cause leaf necrosis in many economically important crops [12]. Based on their amino acid sequences, elicitins have been divided into two classes. The two classes have a major difference in necrotic activity; β-elicitins are 100-fold more toxic than α-elicitins, yet these two proteins only differ in a few amino acid residues. Most notable, the α-elicitins have a Val residue at position 13, while the large majority of the β-elicitins have a Lys at this position [13, 14]. Site-directed mutagenesis studies show that this substitution plays a crucial role in determining the necrotic activity [15].

The three-dimensional structure of the β-elicitin, β-cryptogein, has recently been determined by X-ray crystallography [16] and by high resolution NMR methods [17]; these two independently determined structures are virtually identical (Fig. 1). Likewise, the secondary structure of a typical α-elicitin, α-capsicein, is very similar to that of β-cryptogein [18]. In order to help understand the difference in potency between α- and β-elicitins it was of interest to us to calculate the pK_a value of the critical β-elicitin Lys13 residue. In addition, the pK_a values of the five Tyr residues of β-cryptogein have been estimated by UV-difference spectroscopy [19], and these provided an interesting test case for comparing these experimental results with our computational approach.

2 Methods

The structure of β-cryptogein was taken from the Brookhaven Protein Data Bank [20] (entry 1beo [16]). pK_a calculations were performed as described in full detail elsewhere [6]. In short, the pK_a of a titrating site is computed from the shift in pK_a upon transfer of a model compound with a given pK_a from solution into the protein location. This is carried out in two steps. First, the intrinsic pK_a^{intr} of the sites is determined. The pK_a^{intr} is the pK_a of a titrating site while titrating sites have vanishing total charge. That is, it includes the effect of interactions with so-called "background charges" (pH-independent charges, e.g. backbone charges) and also a "self" energy term. Subsequently, an additional shift in pK_a is calculated through a Monte Carlo simulation procedure which samples all important protonation states of the protein at a specified pH. This step includes the site–site interactions of titratable groups. All interactions are assumed to be of electrostatic nature only and are computed by means of the BEM [10]. The BEM converts the linear Poisson-Boltzmann differential equation employed for the solvent and the Poisson differential equation used for the protein region into a set of linear integral equations valid on a surface enclosing the protein, which are then numerically solved. The surface is described as a triangulated surface and acts as a dielectric boundary such that mutual polarization effects between solvent and protein are included. The surface roughly follows the solvent-accessible surface of the protein. The triangulation algorithm employed in this work is described elsewhere [21]. Partial charges were taken from the molecular dynamics program Gromacs (version 1.3 [22]). Calculations were performed for several choices of the ionic strength (ranging from 0 to 0.15 M) and dielectric constant (20) of the protein. The dielectric constant of the solvent was always set to 78.5.

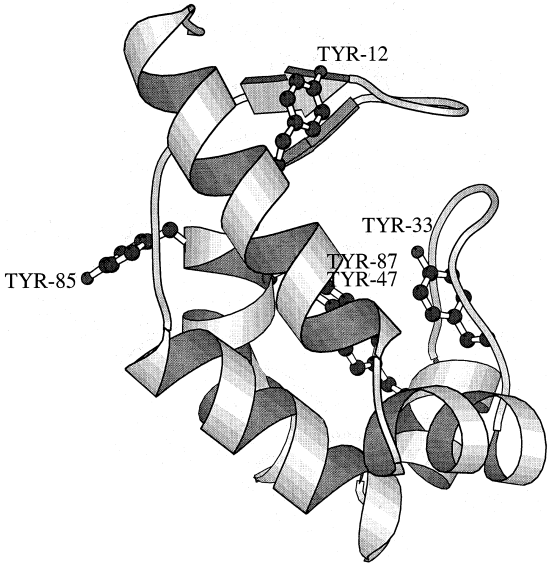

Fig. 1. Illustration of the structure of β-elicitin. The Tyr residues are explicitly displayed. The picture was prepared with Molscript [27]

3 Results

Table 1 presents the results of our calculations with the β-cryptogein protein. The experimental Tyr pK_a determinations [19] were carried out at 0.1 M ionic strength. The calculated pK_a values at 0.1 M ionic strength for the Tyr residues are in reasonable agreement with experimental data with the exception of Tyr87 and Tyr47. The intrinsic pK_a values for all Tyr residues calculated at 0.1 M ionic strength are all around 10.6 except for Tyr85 which is 9.8 (data not shown). This indicates that the calculated apparent pK_a values for the Tyr residues arise from interactions between the Tyr residues and other titratable groups and are not strongly influenced by interactions with background charges. The exception is Tyr85. This site is relatively solvent accessible (see Fig. 1 last two columns in Table 1) and is far away from other titrating sites. Its pK_a is apparently due to interactions with the background. In general, one can say that all calculated pK_a figures are close to pK_a values commonly observed for amino acids in proteins.

As indicated, the pK_a values of Tyr47 and Tyr87 do not correspond so well with experiment. It appears that Tyr47 is rather close to Lys94 (about 0.4 nm). Both the Lys and Tyr display intrinsic pK_a values which are quite close to the pK_a figures of the corresponding free amino acids, but owing to a possible mutual interaction between Tyr47 and Lys94 the apparent pK_a values of these residues are shifted from what is normally expected (10 for Lys and 11 for Tyr). Tyr becomes negatively charged upon deprotonation, and is stabilized by the positive charge on the Lys. Therefore the pK_a of Tyr47 becomes more acid while Lys94 displays a more basic apparent pK_a. With increasing ionic strength this interaction is more effectively screened such that the pK_a values of these sites shift toward normal values. The experimental pK_a of Tyr33 has only been determined for a protein with high homology to β-cryptogein. This site is buried and has a p$K_a > 12$. This value is in close agreement with our computed value.

The influence of ionic strength on the calculated pK_a values is generally not high. The pK_a data for 0.05 M is virtually identical to the numbers at 0.15 M. For 0.0 M,

Table 1. pK_a values for β-cryptogein (β-elicitin), calculated with a dielectric constant of 20

Residue	0.0 M	0.05 M	0.1 M	0.15 M	Exp[a]	Residue[c] SAS	Site[d] SAS
N-terminal	7.8	8.0	8.0	8.0	–	148	34
Tyr12	11.0	10.8	10.8	10.7	11.5	76	17
Lys13	10.8	10.4	10.3	10.2	–	135	40
Asp21	1.0	1.7	1.8	1.9	–	45	26
Asp30	3.3	3.9	3.8	3.9	–	66	27
Tyr33	13.4	12.7	12.6	12.6	> 12.0 [b]	7	0
Lys39	11.0	10.6	10.6	10.6	–	182	57
Tyr47	8.9	9.1	9.1	9.2	> 12.0	14	9
Lys48	11.5	11.0	10.9	10.8	–	107	54
Lys61	11.1	10.6	10.6	10.6	–	129	30
Lys62	13.9	13.1	12.9	12.8	–	74	22
Asp72	1.8	2.5	2.7	2.8	–	62	42
Tyr85	9.6	9.6	9.6	9.6	10.0	85	41
Tyr87	12.9	12.3	12.2	12.1	10.5	7	0
Lys94	13.9	12.9	12.8	12.7	–	98	48
C-terminal	3.4	4.2	4.3	4.3	–	99	23

[a] Experimental values reported by Nespoulous and Pernollet [19]
[b] This value was not determined for β-cryptogein, but in the analogous protein cinnamonin it was estimated at >12
[c] Residue solvent-accessible surface (SAS) area in $Å^2$ calculated by the DSSP program of Kabsch and Sander [24]
[d] The site SAS (e.g. for the Tyr, one employs the OH group) was determined by the ASC (analytical surface calculation) program of Eisenhaber et al. [25, 26]

the pK_a values are occasionally off by one pK_a unit with respect to the non-zero ionic strength case. However, the use of zero ionic strength is artificial since there are always counter ions present in solutions of proteins which give rise to some screening of electrostatic interactions.

4 Discussion

Four amino acid residues in the elicitins have been pinpointed as being involved in the markedly different necrotic activities of α- and β-elicitins; these are Ala2, Lys13, Asp72 and Lys94 [16]. Of these four, the substitution of Lys13 by a hydrophobic residue has the most drastic effect on the activity [15]. Our results (see Table 1) show that Lys13 has a pK_a value that is typical for an exposed Lys residue, suggesting that this residue is probable crucial for the correct binding of β-elicitin to its receptor, rather than for forming an ion-pair in the protein. The $\in$-amino group of Lys94 has an increased pK_a, which results from its interaction with Tyr47. The pK_a of the Asp72 residue is also as expected for an exposed carboxylate group.

The agreement obtained between our calculated pK_a values for the Tyr residues of β-cryptogein and those experimentally determined [19] is reasonable. This is perhaps somewhat surprising, since the ultraviolet difference spectroscopy method used is not very accurate for proteins containing multiple aromatic groups: the absorbance maxima of the five Tyr residues are all rather broad and overlap with each other. Moreover, contributions of the two Phe residues were ignored, and the possibility of structural changes in the protein upon substitution of a Tyr residue was not considered. It will be interesting to see if these pK_a values can be experimentally verified by high resolution NMR methods (P. Gooley, personal communication).

In their structural analysis of the Tyr residues, Pernollet's group [16, 19] distinguished them in "buried" and "exposed" residues, and they suggested a correlation between the extent of burial of a Tyr and its pK_a [although the experimental apparent pK_a of the buried Tyr87 (10.5) does not support such a conclusion]. Our recent calculations of pK_a values in a number of proteins allow us to test whether such a relationship does in fact exist. In Fig. 2 we have therefore plotted the individual Tyr and Lys pK_a values calculated in a number of small well-defined proteins versus the residues' surface exposure. These diagrams clearly show that there is no obvious correlation between these two parameters. This result highlights once again that, in addition to solvent exposure, other electrostatic contributions can have a determining effect on the pK_a of residues in proteins. However, our theoretical work does support the notion that fully buried Tyr residues normally do have very high pK_a values that are experimentally difficult to assess because the protein often denatures while these groups titrate.

As others [4] and we [1] have noted before, the quality of the calculations depends on the choice of the dielectric constant for the protein. A higher value usually results in a better agreement with experiment. This is in agreement with some molecular dynamics simulations of solvated proteins and peptides, which have indicated that the dielectric constant of a protein could be around 30 [7–9]. It has been suggested that a high dielectric constant could account for conformational flexibility and specific ion binding [4]. The latter suggestion was recently tested for the Ca^{2+} binding protein calbindin [23]. It was concluded that a high dielectric constant indeed would predict reasonable estimates of pK_a shifts upon Ca^{2+} binding. Most probably, however, this is because of possible errors in force fields for proteins concerning the protein-ion interaction which are "hidden" by a high dielectric constant.

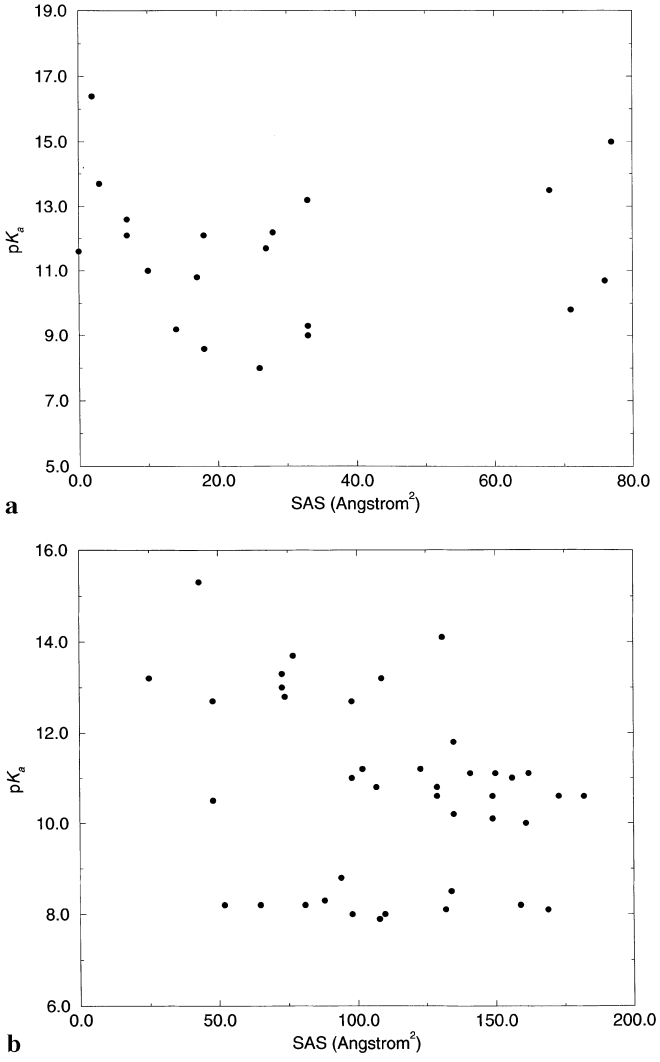

Fig. 2. A Calculated pK_a of Tyr residues in several proteins versus the residue solvent-accessible surface (SAS) area as calculated by the DSSP program [24]. The pK_a values for the Tyr residues were compiled for the proteins chitosanase [28], β-cryptogein (this work), thioredoxin (J. Kemmink, A.H. Juffer, unpublished), ovomucoid and BPTI [6]. **B** Calculated pK_a of Lys residues versus the residue SAS area. The pK_a values for Lys residues were compiled for the proteins chitosanase [28], β-cryptogein (this work), thioredoxin (J. Kemmink, A.H. Juffer, unpublished), ovomucoid and BPTI [6]

In addition, conformational changes in the protein upon ion binding, as well as coupling between calcium and proton binding are generally ignored in these computations. In general, the protein cannot be seen as a uniform dielectric medium: the rigid "core" could be viewed as a low dielectric medium while the outer regions of the protein (containing many mobile Lys residues) is much more polarizable [8].

Acknowledgements. The work was supported by grants for A.H.J. from the Alberta Heritage Foundation for Medical Research (AHFMR) and the Human Frontiers Science Program (HFSP). H.J.V is the recipient of a Scientist Award of the AHFMR. We are indebted to Ms. Margaret Tew for typing the manuscript.

References

1. Juffer AH (1998) Biochem Cell Biol (in press)
2. Bashford D, Karplus M (1990) Biochemistry 29:10219
3. Honig B, Sharp K, Yang AS (1993) J Phys Chem 97:1101
4. Antosiewicz J, McCammon JA, Gilson MK (1996) Biochemistry 35:7819
5. Kesvatera T, Jönsson B, Thulin E, Linse S (1996) J Mol Biol 259:828
6. Juffer AH, Argos P, Vogel HJ (1997) J Phys Chem B 101:7664
7. Smith PE, Brunnen RM, Mark AE, van Gunsteren WF (1993) J Phys Chem 97:2009
8. Simonson T, Brooks CL III (1996) J Am Chem Soc 118:8452
9. Löffler G, Schreiber H, Steinhauser O (1997) J Mol Biol 270:520
10. Juffer AH, Botta EFF, van Keulen BAM, van der Ploeg A, Berendsen HJC (1991) J Comput Phys 97:144
11. Zauhar RJ, Morgan RSJ (1985) J Mol Biol 186:815
12. Yu LM (1995) Proc Natl Acad Sci USA 92:4088
13. Huet JC, Nespoulous C, Pernollet JC (1992) Phytochemistry 31:1471
14. Huet JC, Pernollet JC (1993) Phytochemistry 33:797
15. O'Donohue MJ, Gousseau H, Huet JC, Tepfer D, Pernollet JC (1995) Plant Mol Biol 27:577
16. Boissy G, de la Fortelle E, Kahn R, Huet JC, Bricogne G, Pernollet JC, Brunie S (1996) Structure 4:1429
17. Fefeu S, Bouaziz S, Huet JC, Pernollet JC, Guittet E (1997) Protein Sci 6:2279
18. Bouaziz S, van Heijenoort C, Huet JC, Pernollet JC, Guittet I (1994) Biochemistry 33:8188
19. Nespoulous C, Pernollet JC (1994) Int J Pept Protein Res 43:154
20. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MP, Rodgers JR, Kennard T, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535
21. Juffer AH, Vogel HJ (1998) J Comput-Aided Mol Design 12:289
22. van der Spoel D, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, van Druunen R, Berendsen HJC (1996) Gromacs users manual. University of Groningen
23. Juffer AH, Vogel HJ (1998) (in press)
24. Kabsch W, Sander C (1983) Biopolymers 22:2577
25. Eisenhaber F, Argos P (1993) J Comput Chem 14:1272
26. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) J Comput Chem 16:273
27. Kraulis P (1991) J Appl Crystallogr 24:946
28. Fukamizo T, Juffer AH, Honda Y, Brzezinski R, Vogel HJ (1998) (in press)