Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 481–486
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Fast empirical pK_a prediction by Ewald summation

Elmar Krieger [a,*], Jens E. Nielsen [b], Chris A.E.M. Spronk [a], Gert Vriend [a]

[a] Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen, Toernooiveld 1, 6525ED Nijmegen, The Netherlands
[b] School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

Received 28 September 2005; received in revised form 21 February 2006; accepted 22 February 2006
Available online 27 April 2006

Abstract

pK_a calculations for macromolecules are normally performed by solving the Poisson–Boltzmann equation, accounting for the different dielectric constants of solvent and solute, as well as the ionic strength. Despite the large number of successful applications, there are some situations where the current algorithms are not suitable: (1) large scale, high-throughput analysis which requires calculations to be completed within a fraction of a second, e.g. when permanently monitoring pK_a shifts during a molecular dynamics simulation; (2) prediction of pK_as in periodic boundaries, e.g. when reconstructing entire protein crystal unit cells from PDB files, including the correct protonation patterns at experimental pH. Such in silico crystals are needed by 'self-parameterizing' molecular dynamics force fields like YASARA YAMBER, that optimize their parameters while energy-minimizing high-resolution protein crystals.

To address both problems, we define an empirical equation that expresses the pK_a as a function of electrostatic potential, hydrogen bonds and accessible surface area. The electrostatic potential is evaluated by Ewald summation, which captures periodic crystal environments and the uncertainty in atom positions using Gaussian charge densities. The empirical proportionality constants are derived from 217 experimentally determined pK_as, and despite its simplicity, this pK_a calculation method reaches a high overall jack-knifed accuracy, and is fast enough to be used during a molecular dynamics simulation. A reliable null-model to judge pK_a prediction accuracies is also presented.

Keywords: Protein pK_a prediction; Poisson–Boltzmann; Particle mesh Ewald; Crystal space; Force field optimization

1. Introduction

The prediction of pK_a values in proteins has made considerable progress over the last years [1,2]. The Poisson–Boltzmann equation (PBE) has become an important tool because it allows the calculation of the electrostatic potential in a heterogeneous solute–solvent system, taking into account dielectric boundaries and the ionic strength. Initial approaches to electrostatic calculations were based on rough approximations like spherical proteins [3]. The ability to solve the Poisson–Boltzmann equation for arbitrarily shaped proteins [4–6] cleared the path for a range of successful applications, such as studies of enzymatic activity [7], pH-dependent conformational changes [8] and protein stability [9–11]. These algorithms, however, are computationally expensive, and consequently led to the development of several simplified algorithms that avoid solving the PBE. Examples of these algorithms are the Debye–Hueckel approach [12] and the electrostatic screening functions [13,14].

pK_a calculations have always focused on proteins in their physiological environment, matching the experimental determination of pK_a values, which is also done in solution using NMR spectroscopy. However, the quality of pK_a calculations depends heavily on the availability of high resolution protein structures. NMR structures of sufficient resolution are often not available, and one is forced to predict solution pK_a values using X-ray structures. Much effort has been devoted to determining the regions of structural divergence, excluding residues involved in crystal contacts [15], optimizing X-ray structures [16] and incorporating information on protein flexibility [17].

1.1. The goal is pK_a prediction in protein crystals

The approach presented here has been developed due to a lack of solutions for a problem that appears paradoxical, given

Abbreviations: PME, particle mesh Ewald; PBE, Poisson–Boltzmann equation

* Corresponding author. Tel.: +31 24 3652248; fax: +31 24 3652977.
E-mail addresses: elmar@yasara.org, elmar@cmbi.ru.nl (E. Krieger).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.02.009

the facts mentioned above: the prediction of pK_a values in protein crystals. Because of the crystal packing interactions, these pK_as certainly differ from those measured in solution. The reason for addressing this problem becomes clear in view of recent developments in force field research. Thanks to the virtually unlimited resources provided by distributed computing systems like Models@Home [18], it became feasible to use complete proteins instead of small molecules as optimization targets when fitting the force field parameters [19]. This was done by randomly changing force field parameters and running simulations on a series of protein structures to see if the parameter changes would be beneficial. Obviously, the protein structures in the optimization set should be as realistic as possible, otherwise the force field might memorize features that are just structural artifacts. This can be achieved by taking high resolution X-ray structures and reconstructing the entire unit cell, including water molecules, counter ions and all solute hydrogens. The correct placement of polar hydrogens is especially important, and in addition to optimizing the hydrogen-bond network [20], this requires the pK_a values of all ionizable residues in the protein crystal and the pH at which the protein was crystallized. The force field parameters are then optimized in crystal space, so that all the interactions responsible for the experimentally observed structure can be considered, while converging at a force field like YAMBER [21]. Because crystal and solution environments obey the same laws of physics, the optimized force field can be used in both.

2. Ewald summation captures the periodic environment

Electrostatic calculations in periodic crystal systems are complicated by the infinite number of interactions. A clever way of making the problem tractable is Ewald summation [22], which allows the calculation of the potential due to the N particles in the unit cell and an infinite number of periodic replicas. The method combines a rapidly converging short-range term with a long-range component evaluated in reciprocal space [23]. If the reciprocal sum is calculated using a particle-mesh approximation, the resulting particle mesh Ewald (PME) algorithm [24] is considerably faster than the standard Ewald method. PME is part of almost every molecular dynamics program, and forms the basis for this work. However, we only use the reciprocal space portion, which provides the solution to Poisson's equation with periodic boundaries, Gaussian charge distributions and a single dielectric constant. By ignoring the short-range term and the associated damping of the reciprocal space term at short-range, we essentially remove the long-range attribute from the reciprocal space term: it now covers all distance ranges equally, and differs from Coulomb's law only by the use of Gaussian charge densities instead of localized point charges. Smeared-out Gaussians account for the uncertainty in atom positions (which also proved beneficial for the development of knowledge-based potentials [25]). Compared to the Poisson–Boltzmann equation, this approach however lacks two advantages: implicit counter ions and different dielectric constants for solvent and solute.

In an extensive optimization study, Demchuck and Wade [1] determined that the best dielectric constant for solvent exposed residues is close to the one of water (80), while the protein interior should be assigned a value in the range of 10–20. Since 20 differs from 80 only by a factor of 4, we hypothesized that a single global dielectric constant could suffice for accurate predictions, provided that some additional structural information was incorporated to account for the simplification.

3. The pK_a can be approximated as a function of electrostatic potential, hydrogen bonds and accessible surface

Using simplified physical considerations and some modeler's experience, we defined three rules of thumb for pK_a prediction. The first and partly the second rule have also been mentioned in a recent analysis of carboxyl pK_a values [26]:

- If an ionizable group is surrounded by negatively charged residues, corresponding to a negative electrostatic potential, protonation becomes easier, the pK_a increases. Similarly, if there are positively charged residues around, the pK_a decreases. As a first approximation, the pK_a shift is thus assumed to be proportional to the electrostatic potential.
- If an ionizable group accepts hydrogen bonds, the space to place a proton is reduced, protonation becomes harder, and the pK_a decreases. If after protonation, the group can donate a bond, protonation is favorable, the pK_a increases.
- If a group accepts hydrogen bonds and is buried, the pK_a is decreased even further, because the side-chain cannot facilitate protonation by moving to a different conformation where it does not receive hydrogen bonds. If a buried group can donate a hydrogen bond after protonation, the pK_a increases, because there is no space for water molecules that could ease the energetic cost of two hydrogen-bond acceptors facing each other.

These three assumptions were fused into an empirical equation relating the pK_a of a residue with the electrostatic potential, the number of hydrogen bonds and the accessible surface area:

$$\mathrm{p}K_a = \mathrm{Model}\,\mathrm{p}K_a + \sum_{\text{Ionizable atoms}} [-A \times \mathrm{Ewald}E_i + B \times \mathrm{HB}_i] + \mathrm{Sign}(\mathrm{HBSum}) \times C \times \mathrm{SurfaceLoss} \quad (1)$$

In this equation, *Model* pK_a is the standard pK_a value of a certain residue type, EwaldE_i is the reciprocal space portion of the Ewald energy of a charge +1 at the location of the ith ionizable atom in the residue (in kcal/mol), HB$_i$ is the difference between (potentially) donated and accepted hydrogen bonds at the ith atom, HBSum is the sum over all HB$_i$, and SurfaceLoss is the loss of accessible surface area of the side-chain with respect to a fully exposed state. A, B and C are empirical proportionality constants. The four unknown parameters *Model* pK_a, A, B and C are globally optimized for each amino acid type so that the RMSD between predicted and

observed pK_a values is minimal. More details about the equation can be found in Section 4.

The RMSD was chosen as optimization target because it is ideally suited for analyzing pK_a prediction accuracy: one is not so much interested in the small shifts of isolated surface residues, but in the large shifts that significantly influence the protonation states and dominate the RMSD when mispredicted. As the main goal of this work is to develop a method for pK_a prediction in protein crystals used in force field parameterization, all residues are equally important. No matter if a wrong protonation state is assigned to an active site residue or to a surface residue involved in crystal contacts – the influence on the optimized force field is equally bad. Hence the RMSD is calculated for all residues with experimentally determined pK_a values, and our goal is to obtain a low overall RMSD.

When fitting the parameters to reproduce experimental pK_a values, it is important to note that these pK_a values were measured in solution. The Ewald energy in Eq. (1) must therefore also be calculated in a solution environment. This is achieved by placing an isolated protein in a very large cell, so that the periodic boundaries implied by the Ewald summation have no significant influence. Validation of prediction accuracies is thus also only possible in solution, since there is currently no experimental method to directly measure pK_as in protein crystals. The resulting parameters can however be used directly for pK_a prediction in protein crystals, just like the same molecular dynamics force field parameters can be used to simulate proteins in solution and crystals.

4. Materials and methods

4.1. Datasets of experimental pK_a values

A total of 227 experimentally measured pK_a values were compiled for this study. They consisted of the Asp/Glu specific dataset collected by Forsyth et al. [26], from which we removed double occurrences of the same protein to prevent compromising the jack-knife test, and four cases where it was uncertain if the structure deposited in the PDB was close to the one used for the pK_a measurements: CD2 because it undergoes domain swapping (see PDB IDs 1CDC and 1HNG), chymotrypsin inhibitor 2, because 19 important N-terminal residues were disordered in the structure, subunit C of F0F1-ATPase because its structure was determined in a chloroform–methanol mixture and HIV-1 protease/KNI-272 complex because there were no force field parameters available for the essential ligand. Then the histidine specific dataset from Edgcomb and Murphy [27] and our own previously described collection [15] were included, which added mainly pK_a values for lysines and tyrosines. Too few pK_a values of carboxyl-termini were available to be included in this analysis, which lead to slightly different RMSDs in Table 3 compared to our previous work [15].

4.2. Hydrogen-bond counting

The numbers of accepted and donated hydrogen bonds contributing to HBSum in Eq. (1) were determined after an optimization of the hydrogen-bond network with WHAT IF [20], which was previously shown to significantly improve pK_a prediction accuracy [15]. A hydrogen bond was allowed to contribute to HBSum if the distance between the hydrogen and the acceptor was below 2.5 Å and donor and acceptor were separated by more than three covalent bonds. For carboxyl oxygens and histidine nitrogens that were not protonated in the optimized network, a (potentially) donated bond was counted if there was an acceptor separated by more than three bonds within 3.5 Å.

4.3. Calculation of the electrostatic potential

Ewald$E$ in Eq. (1) was calculated using the PME algorithm [24] implemented in YASARA (available from http://www.YASARA.org, including the pK_a prediction module described here). To avoid singularities and short-range noise, all calculations were done with the reciprocal part of the Ewald sum only, with a grid spacing < 1 Å, sixth order B-splines and a tolerance of 1e−5 for the direct space sum (which was used to determine the convergence parameter for the reciprocal sum). The simulation cell was 30 Å larger than the protein along each axis. Charges were assigned to all atoms based on the Amber 99 force field [28]. All ionizable groups were in their standard protonation states (i.e. D, E and H deprotonated, K and Y protonated), no iterations to sample different protonation patterns were done. For lysine and tyrosine, the potential was calculated at the protonated NZ and OH atoms, for histidine at the deprotonated NE2 atom, and for aspartate and glutamate at both deprotonated oxygens (for the latter two residues, the sum in Eq. (1) therefore runs over two atoms). As the histidine ring can flip, the NE2 atom was assigned based on the following rules: (1) if the histidine accepts a hydrogen bond, the acceptor is the NE2 atom. (2) If the histidine donates a bond, the other nitrogen is labeled NE2. (3) If neither the first nor the second is true, both nitrogens are temporarily protonated and the one with the higher electrostatic potential is assumed to be NE2. The resulting energies to be used in Eq. (1) typically fall in the range −100 to +100 kcal/mol, the common scaling factor 0.5 accounting for the bi-directionality of electrostatic interactions is omitted.

4.4. Accessibility calculations

The loss of accessible surface area was calculated by subtracting the side-chain accessibilities calculated with WHAT IF's standard parameters [29] from the following values corresponding to a fully exposed state: Asp 34 Å^2, Glu 40 Å^2, Tyr 60 Å^2, His 51 Å^2, Lys 55 Å^2.

4.5. Performance details

The performance of the method was evaluated on a 3 GHz Pentium IV machine using a typical TIM barrel (PDB ID 5TIM chain A). Evaluation of the electrostatic potential with the PME algorithm [24] takes 0.25 s, followed by calculation of the accessible surface (0.10 s) and assignment of hydrogen bonds

Table 1
Prediction accuracy for 227 experimentally determined pK_as

Residue type	No.	Null-model	Ewald, three parameters	Ewald, four parameters
Asp	83	0.948	0.739	0.710
Glu	81	0.717	0.673	0.678
His	35	1.563	1.636	1.592
Tyr	6	0.837	0.837	0.837
Lys	22	0.502	0.399	0.399
All	227	0.965	0.899	0.879

The second column lists the number of predictions per residue type. The RMSDs obtained with the optimized null-model are shown in the third column. The fourth column lists the results obtained if only three parameters are used (parameter C in Eq. (1) set to zero). Only six tyrosine residues were present, which was not enough to reliably fit more than one parameter (the model pK_a). RMSDs for tyrosine are therefore identical in all cases. For the 22 lysine residues, only three parameters could be fit, giving the same results in the fourth and fifth column.

(0.004 s), summing up to 0.354 s in total. For comparison, one step of a 5TIM molecular dynamics simulation (29 000 atoms including water, simulation parameters as described previously [21]) takes 0.63 s. Hence it costs only ∼2% performance to monitor all pK_a values ever 25th simulation step.

The pK_a predictions together with the user-specified pH can then be used to reassign the protonation states, or preferably (since the influence on the simulation is smaller) to assign new target values for the fractional protonation λ, which defines a mixture of force field parameters for the protonated and deprotonated states [30].

5. Results and discussion

To evaluate the accuracy of the Ewald summation approach, pK_a predictions were made for 227 aspartate, glutamate, histidine, lysine and tyrosine residues in a set of 27 structures. For the remaining ionizable side-chain types, we did not have enough experimental data to fit the parameters in Eq. (1).

The RMSDs of the predicted pK_a values from the experimental ones are listed in Table 1. All RMSDs, including the one for the null-model, have been obtained with a Jack-knife approach, i.e. the parameters were separately determined for each of the 27 structures using the remaining 26 structures.

One important question is whether or not a pK_a prediction method performs better than the null-model, which trivially assumes a constant pK_a value for all residues of a certain type, and can be surprisingly difficult to beat [15]. As can be seen from Table 1, our empirical equation gives better overall results (0.879) than the null-model (0.965). The best results are achieved for aspartate (0.71 compared to 0.95), while histidine turns out to be the most difficult residue (1.59 compared to 1.56).

It has been noted before that there is virtually no correlation between accessible surface area and pK_a shift [26,27]. Indeed, the surface term of our equation makes the weakest contribution and can be left out without significantly compromising the accuracy. This is shown in the fourth column of Table 1 (overall RMSD 0.899).

To estimate the relative importance of the Ewald- and hydrogen-bonding terms, we also tested a reduced model considering only the two parameters 'model pK_a' and 'hydrogen bonding'. The resulting RMSD of 0.935 was roughly half-way between the null-model (0.965 in Table 1) and the three parameter model (0.899), indicating that Ewald energy and hydrogen bonding contribute about equally to the prediction.

Table 2 lists three different parameter sets: the null-model alone, the parameters for Eq. (1) without the surface term (C set to zero), and the parameters for the complete Eq. (1). It can be seen that without the surface term, parameters have a clear physical meaning. For example, accepting a hydrogen bond lowers the pK_a of aspartate by 0.3 units, and the one of glutamate by 0.17 units. This matches the previous finding that pK_a values of glutamates are less influenced by hydrogen bonds [26]. One simple explanation might be that the glutamate side-chain is more flexible, so rather than being protonated while accepting hydrogen bonds, it adapts a different conformation.

As soon as the surface term is included, we find parameter dependencies that make a physical interpretation hardly fruitful (e.g. the sign of one parameter changes unexpectedly, while another one compensates). This finding does not indicate that desolvation effects are unimportant, it just shows that either the surface term cannot truly capture their physical basis, or that the number of residues with significant desolvation effects is small. Nevertheless, the surface term passed the Jack-knife test, indicating that the increase of accuracy is not just due to the additional optimization parameter, but that there is indeed a small signal present.

Comparison of these results with other prediction methods is difficult, because they are very dataset dependent (e.g. the relatively high RMSD for aspartate is mainly caused by Asp-26

Table 2
Empirical parameters for pK_a prediction

Residue type	Model-1	Model-3	A	B	Model-4	A	B	C
Asp	3.220	3.280	0.00264	0.3032	3.270	0.00254	0.4725	−0.01663
Glu	4.090	3.949	0.00209	0.1670	3.904	0.00224	0.2883	−0.01145
His	6.200	5.942	0.01112	0.7447	5.871	0.01002	−1.1772	0.05323
Tyr	10.800	10.800	–	–	10.800	–	–	–
Lys	10.760	10.938	0.00408	0.0924	10.941	0.00424	−0.0042	0.00479

The table shows three different sets with an increasing number of parameters. In the first case, only the model pK_a is optimized, resulting in the null-model (pK_a values in column 2). If additionally the electrostatic potential and the hydrogen bonds are considered, three parameters are required (columns 3–5). Inclusion of the surface term requires four parameters (columns 6–9). A, B and C are the parameters used in Eq. (1). The lack of data for tyrosin allowed to fit just one parameter, the model pK_a.

Table 3
Comparison of pK_a prediction accuracy for four different methods

Residue type	No.	Null-model	Ewald, three parameters	Ewald, four parameters	Poisson–Boltzmann
Asp	45	0.847	0.592	0.573	0.733
Glu	41	0.802	0.777	0.788	0.970
His	8	1.591	1.318	1.206	1.588
Tyr	6	0.837	0.837	0.837	0.972
Lys	22	0.502	0.399	0.399	0.533
All	122	0.853	0.714	0.699	0.882

The null-model (column 3), the empirical equation described here, without (column 4) and with (column 5) the surface term, and finally the Poisson–Boltzmann equation based approach described previously [15]. Listed is the RMSD between predicted and experimentally measured pK_a.

in Thioredoxin, with a predicted pK_a of 3.5 and a measured pK_a of 8.1 [31]). We therefore calculated RMSDs for a subset of residues, matching the dataset used in our previous analysis based on the Poisson–Boltzmann equation [15], which compared favorably to other commonly used pK_a calculation methods. Table 3 shows that the empirical approach works surprisingly well, giving lower RMSDs in all cases. The optimized null-model also performs better than expected.

While the focus in this work is on accurate pK_a prediction for all residues, catalytic active site residues are often the most interesting ones. Based on the articles describing the 27 proteins used here, 20 catalytic aspartates, glutamates and histidines could be identified. Not surprisingly, prediction accuracy is worse (1.36 pK_a units RMSD) but still compares well to the optimal null-model (1.51). The same holds for the comparison to the Poisson–Boltzmann method in the common subset (1.46–1.61 units RMSD).

6. Conclusions

While the initial motivation for this work was the need to predict pK_a values in protein crystals, we ended up with two interesting findings.

First, our empirical approach based on a global dielectric constant and hydrogen bond counting resulted in a lower RMSD than the PBE based method. Partly, this can be attributed to the parameter fitting procedure, which allows us to find optimum values for variables that are difficult to determine, both theoretically and experimentally. PBE calculations also provide room for improvement by parameter optimization. This has already been shown for the dielectric constants [1] and the hydrogen-bonding network [15], while the next obvious candidates are the model pK_a values. These are crucial parameters, and estimating them from compounds that only resemble amino acids carries the inherent danger of a systematic error. No physical meaning is lost if they are optimized instead. Another important reason for the lower RMSD is fewer mispredictions. Due to its quadratic nature, the RMSD is dominated by the large mispredictions. We found that the PBE method occasionally predicts large pK_a shifts that are not observed experimentally. While it is sometimes suggested that pK_a prediction should only look at residues that exhibit large shifts and not bother about the rest, this finding shows that all residues are important: a lot can be learned from analyzing why theory predicts a shift if none is found experimentally.

Second, the null-model is still hard to beat, as long as it is optimized and not assumed to be equal to pK_a values of model compounds used in PBE calculations. For example, the optimum null-model pK_a for aspartate is 3.22 (Table 2). This differs by $\sim$0.8 pK_a units from the value used in PBE calculations [15]. When comparing prediction accuracies with the null-model, it is essential to use the optimized values, so that the null-model gets a fair chance, since results obtained with the classic null-model [15] are much worse (1.069 instead of 0.853 pK_a units RMSD in Table 3).

As a conclusion, pK_a prediction remains a difficult topic, hampered by uncertainties in the experimental data and the underlying protein structures. It is likely that even a perfect method would still be facing unexplainable outliers and a lot of noise, making it hard to objectively decide between competing theories. Even though our approach can in principle not reproduce titration curves of tightly coupled residues, its focus on high overall accuracy for real-life residue distributions and support for periodic boundaries and non-orthorhombic cells makes it well suited for the intended purpose: the rapid large scale prediction of pK_a values and assignment of protonation states in force field parameterization, molecular dynamics simulations and homology model refinement.

Acknowledgements

This work was supported by the EU (fifth Framework program, contract number QLG2-CT-2000-01313) and the users of the molecular modeling program YASARA.

References

[1] E. Demchuk, R.C. Wade, Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins, J. Phys. Chem. 100 (1996) 17373–17387.
[2] A. Warshel, Calculations of enzymatic reactions: calculations of pK_a, proton transfer reactions, and general acid catalysis reactions in enzymes, Biochemistry 20 (1981) 3177.
[3] C. Tanford, R. Roxby, Interpretation of protein titration curves. Application to lysozyme, Biochemistry 11 (1972) 2193–2198.
[4] J. Warwicker, H.C. Watson, Calculation of the electric potential in the active site cleft due to alpha-helix dipoles, J. Mol. Biol. 157 (1982) 671–679.
[5] A.S. Yang, M.R. Gunner, R. Sampogna, K. Sharp, B. Honig, On the calculation of pK_as in proteins. 15 (1993) 252–265.
[6] J. Antosiewicz, J.A. McCammon, M.K. Gilson, Prediction of pH-dependent properties of proteins, J. Mol. Biol. 238 (1994) 415–436.

[7] X. Raquet, V. Lounnas, J. Lamotte-Brasseur, J.M. Frere, R.C. Wade, pK_a calculations for class A beta-lactamases: methodological and mechanistic implications, Biophys. J. 73 (1997) 2416–2426.
[8] D. Morikis, A.H. Elcock, P.A. Jennings, J.A. McCammon, Native-state conformational dynamics of GART: A regulatory pH-dependent coil-helix transition examined by electrostatic calculations, Protein Sci. 10 (2001) 2363–2378.
[9] D.O. Alonso, K.A. Dill, D. Stigter, The three states of globular proteins: acid denaturation, Biopolymers 31 (1991) 1631–1649.
[10] A.M. Lambeir, J. Backmann, J. Ruiz-Sanz, V. Filimonov, J.E. Nielsen, I. Kursula, B.V. Norledge, R.K. Wierenga, The ionization of a buried glutamic acid is thermodynamically linked to the stability of Leishmania mexicana triose phosphate isomerase, Eur. J. Biochem. 267 (2000) 2516–2524.
[11] A.S. Yang, B. Honig, Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin, J. Mol. Biol. 237 (1993) 602–614.
[12] J. Warwicker, Simplified methods for pK_a and acid pH-dependent stability estimation in proteins: removing dielectric and counterion boundaries, Protein Sci. 8 (1999) 418–425.
[13] L. Sandberg, O. Edholm, A fast and simple method to calculate protonation states in proteins, Proteins 36 (1999) 474–483.
[14] E.L. Mehler, F. Guarnieri, A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins, Biophys. J. 77 (1999) 3–22.
[15] J.E. Nielsen, G. Vriend, Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pK_a calculations, Proteins 43 (2001) 403–412.
[16] J.E. Nielsen, J.A. McCammon, On the evaluation and optimization of protein X-ray structures for pK_a calculations, Protein Sci. 12 (2003) 313–326.
[17] R.E. Georgescu, E.G. Alexov, M.R. Gunner, Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins, Biophys. J. 83 (2002) 1731–1748.
[18] E. Krieger, G. Vriend, Models@Home: distributed computing in bioinformatics using a screensaver based approach, Bioinformatics 18 (2002) 315–318.
[19] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field, Proteins 47 (2002) 393–402.
[20] R.W.W. Hooft, C. Sander, G. Vriend, Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures, Proteins 26 (1996) 363–376.
[21] E. Krieger, T. Darden, S.B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: force field parameterization in crystal space, Proteins 57 (2004) 678–683.
[22] P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys. 64 (1921) 253–287.
[23] J.W. Perram, H.G. Petersen, S.W. de Leeuw, An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles, Mol. Phys. 65 (1988) 875–889.
[24] U. Essman, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[25] G. Vriend, C. Sander, Quality control of protein models: directional atomic contact analysis, J. Appl. Cryst. 26 (1993) 47–60.
[26] W.R. Forsyth, J.M. Antosiewicz, A.D. Robertson, Empirical relationships between protein structure and carboxyl pK_a values in proteins, Proteins 48 (2002) 388–403.
[27] S.P. Edgcomb, P.M. Murphy, Variability in the pK_a of histidine side-chains correlates with burial within proteins, Proteins 49 (2002) 1–6.
[28] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comp. Chem. 21 (2000) 1049–1074.
[29] G. Vriend, WHAT IF—a molecular modeling and drug design program, J. Mol. Graph. 8 (1990) 52–56.
[30] U. Börjesson, P.H. Hünenberger, Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines, J. Chem. Phys. 114 (2001) 9706–9719.
[31] J. Qin, G.M. Clore, A.M. Gronenborn, Ionization equilibria for side-chain carboxyl groups in oxidized and reduced human thioredoxin and in the complex with its target peptide from the transcription factor NF kappa B, Biochemistry 35 (1996) 7–13.