

1 Methods and Algorithms used by the H++

1.1 Pre-processing of submitted structures

Structures are pre-processed differently depending upon their input format, see the flowchart in Fig.1. There are two input formats supported by the H++ website, PDB and PQR. These files differ in that PQR files already have charges and radii assigned to each atom while PDB files do not.

If the input file is in PQR format, H++ makes minimal changes because it is assumed that the format is already suitable for electrostatic calculations. Changes made are as follows: atom names of all titratable amino-acids are brought in accordance with the format adopted by the AMBER¹ package and consistency checks are performed. These checks ensure that the total charge of the system is an integer (within a ± 0.05 unit charge tolerance per amino-acid) and the atomic radii are between 0.5 and 3 Å. If any of the above checks fail, the sequence of residues is discontinuous, or the atom names are different from the PDB standard and can not be recognized, execution terminates.

For a structure submitted in the conventional PDB format, H++ deletes all HETATM records; that is, only those atoms

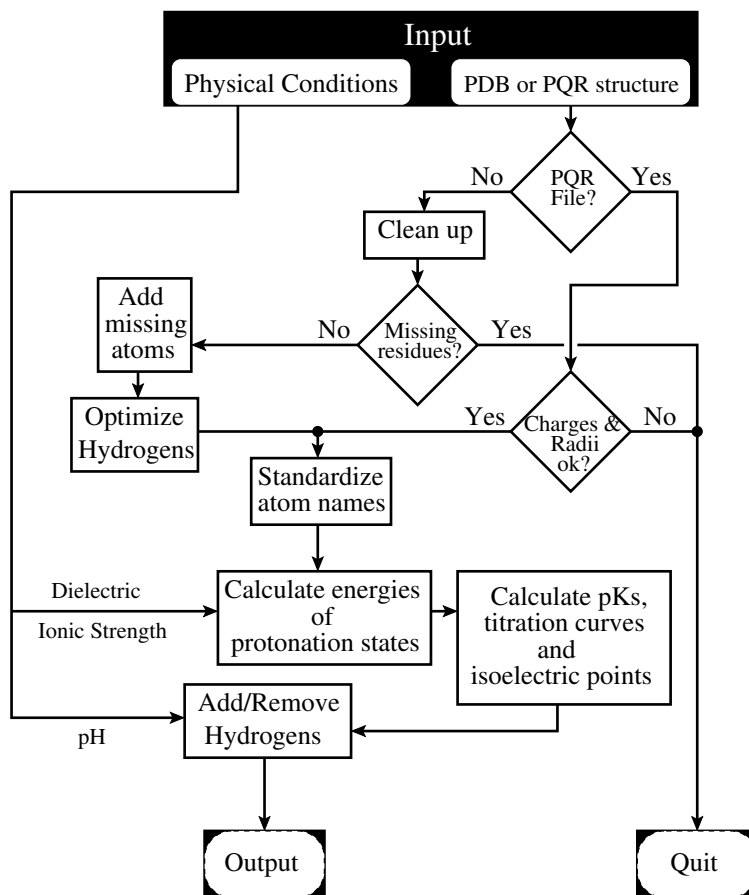


Fig. 1. Flowchart of computations performed by H++ server described here. The input is a structure file in either PDB or PQR (PDB + charges and radii) format. The output includes computed $pK_{1/2}$ values for all titratable groups, as well as titration curves, isoelectric point, and the original structure in which the protonation states of all ionization groups have been made consistent with the calculated pK s. The generated structure is available in several formats used by popular molecular modeling packages.

that belong to amino-acids or nucleotides are kept; this is the “clean-up” step in Figure 1. Removal of explicit water molecules and mobile counterions is generally consistent with the implicit solvent framework in which solvation effects are accounted for in the mean-field manner. If necessary, removed ligands can be included in the calculations by submitting the complete structure in the PQR format, avoiding the “clean-up” step. Sequence continuity is verified and all atom names are brought in accordance with the format adopted by the AMBER package. Deuterium atoms are replaced with equivalent hydrogens.

1.2 Addition of missing atoms and optimization of hydrogens

This section applies *only* to input structures in PDB format. Missing heavy atoms and protons (assuming standard protonation states of titratable groups) are added, atomic partial charges and radii (Bondi) are assigned using the PROTONATE and LEAP modules of AMBER. This is followed by a molecular dynamics based optimization of the positions of the added hydrogens. The protocol was tested earlier² in a similar context; it consists of 3 consecutive stages during which only hydrogens are allowed to move: first, 100 steps of conjugate gradient minimization; second, 500 steps of molecular dynamics at 300 K, with all torsional potentials involving hydrogens set to zero; and third, 100 steps of conjugate gradient minimization with the torsional potential returned to normal values. AMBER *parm99* force-field is used, where the integration time-step is 1 fs and the charge-charge interactions are computed in a uniform “vacuum” of dielectric $\epsilon_{out} = 4$.

1.3 Electrostatic Calculations

The continuum electrostatics methodology widely used to calculate the energetics of proton transfer is described elsewhere;^{3,4} the model is available in the free software package MEAD.⁵ The *H++* calculations rely upon the single-conformer version of MEAD, although conformational variability is partially accounted for by the “smeared charge” representation of titratable groups (see below) and the simulated annealing of protons described above. While not the most systematic or exhaustive way of incorporating conformational variability, we believe, based on previous experiences,^{2,4,6} that this particular model is a reasonable balance between speed and accuracy, and is therefore a good choice for web-based calculations. In this model, the molecule is treated as a low dielectric medium ϵ_{in} , while the surrounding solvent is assigned a high dielectric constant ϵ_{out} . The electrostatic screening effects of (monovalent) salt enter via the Debye-Huckel screening parameter $\kappa[\text{\AA}^{-1}] \approx 0.316\sqrt{[\text{salt}][\text{mol}/L]}$. The salt concentration, ϵ_{in} and ϵ_{out} are accessible to the user, and reasonable defaults are provided. The difference between the $\text{p}K_a$ of a sidechain and the $\text{p}K_a$ of the corresponding model compound in free solution is determined by the combined effect of two distinct contributions to the total electrostatic (free) energy change. First is the “Born term” or desolvation penalty, which always penalizes burial of a charge inside a low dielectric medium. Second is the background term, which represents the electrostatic interactions

of the group in question with all other fixed charges in the molecule not belonging to any titratable groups. These energy terms, as well as the matrix of site–site interactions are estimated through a sequence of calculations in which sites in the protein and their corresponding model compounds have their charge distributions set to those of the protonated or deprotonated form, and suitable energy differences are taken. For the protonated states of ASP and GLU, in which the correct location of the proton is not known *a priori*, a “smeared charge” representation is employed, in which the neutralizing positive charge is symmetrically distributed: 0.45 on each carbonyl oxygen atom, and 0.1 on the carbon atom. The electrostatic calculations are based either on a Generalized Born or Poisson-Boltzmann models, as requested by the user. The particular GB model we are using was found earlier⁷ to work reasonably well in pK calculations on proteins; here its improved version⁸ is used. The set-up and finite-difference solution of the PB problems or analytical GB calculations are carried out using the MEAD program package. In the finite difference lattices, two levels of focusing are used. In the coarsest level the bounding box is set to twice the molecule’s maximum extent and the grid points are spaced 2 Å apart. The finest lattice is “focused” on the region of interest, and the grid points are 0.5 Å apart. The probe radius for defining the molecular surface, which is used as the boundary between the interior and exterior dielectric regions, is set to 1.4 Å.

1.4 Calculating titration curves, $pK_{1/2}$, and protonation states

The electrostatic calculations outlined above provide (free) energies of each of the 2^N protonation microstates² in the system where N is the number of ionizable sites. To make the subsequent calculation of the partition functions (and $pK_{1/2}$) manageable, a fast variant of a clustering approach is used.⁹ The approach subdivides the interacting sites into independent clusters based upon the strength of electrostatic site-site interactions between them. All electrostatic interactions for each ionizable site are sorted from highest to lowest, the top C_{max} sites are then selected to contribute to the calculation, and all others are ignored; the partition function for the site is then factored into computationally manageable components of maximum size C_{max} . Here, $C_{max} = 17$ is used: in tests on 600 representative proteins,¹⁰ we found [Jonathan Myers *et al.*, submitted] that $C_{max} = 17$ resulted in average errors of 0.2 pK units, compared to a standard treatment based upon a Monte-Carlo approach.¹¹

The probability of protonation is computed for every site over a range of pH values equally spaced by 0.1 pH units apart. Individual curves can be displayed for user-selected residues, and the total protonation curve is generated, showing the computed isoelectric point of the molecule. A diagram showing the 10 lowest protonation states and their relative free energies is also generated. These diagrams were found useful² for analysis of proton transfer events in biomolecular systems.

1.5 *Generating the PDB structure in its predicted protonation state*

The computed titration curves provide an estimate of the probability of protonation of each titratable site at the (user-specified) pH of the solvent. A simple scheme is used for assignment of protonation states: if the estimated probability is less than 0.5, the site is considered deprotonated; otherwise, the site is protonated. We follow AMBER conventions in placement of new hydrogen atoms and selection of hydrogen atoms to be removed. Deprotonated (neutral) states of ARG and TYR, not available in the AMBER databases, are obtained from the protonated forms by removal of the *HH22* and *HH* protons respectively; their partial charges are brought in accord with the appropriate model compound values supplied by the MEAD package. The structures submitted in PQR format do not undergo re-assignment of protonation states, allowing greater flexibility for this input format.

References

1. Pearlman, D., Case, D., Caldwell, J., Ross, W., Cheatham III, T., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. Amber, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.* 91:1–41, 1995.
2. Onufriev, A., Smondyrev, A., and Bashford, D. Proton affinity changes during unidirectional proton transport in the bacteriorhodopsin photocycle. *J. Mol. Biol.* 332:1183–1193, 2003.
3. Bashford, D. and Karplus, M. pka's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29(44):10219–10225, 1990.
4. Bashford, D. and Gerwert, K. Electrostatic calculations of the pka values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* 224:473–486, 1992.
5. Tishmack, P. A., Bashford, D., Harms, E., and Van Etten, R. L. Use of ¹h nmr spectroscopy and computer simulations to analyze histidine pka changes in a protein tyrosine phosphatase: Experimental and theoretical determination of electrostatic properties in a small protein. *Biochemistry* 36:11984, 1997.
6. Ullmann, G. M. and Knapp, E.-W. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.* 28:533–551, 1999.
7. Onufriev, A., Bashford, D., and Case, D. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* 104:3712–3720, 2000.
8. Onufriev, A., Bashford, D., and Case, D. Exploring native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55:383–394, 2004.
9. Gilson, M. K. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* 15:266–282, 1993.

10. Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks III, C. L. Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem* 25(2):265–84, 2004.
11. Beroza, P., Fredkin, D. R., Okamura, M. Y., and Feher, G. Protonation of interacting residues in a protein by monte carlo method. *Proc. Natl. Acad. Sci. U. S. A.* 88:5804–5808, 1991.