TKSA-MC: A Web Server for rational mutation through the optimization of protein charge interactions

Vinícius G. Contessoto,[†,‡,¶] Vinícius M. de Oliveira,[‡,¶] Bruno R. Fernandes,[‡]

Gabriel G. Slade,[‡] and Vitor B. P. Leite[*,‡]

†*Brazilian Bioethanol Science and Technology Laboratory - CTBE, Campinas - SP,*

13083-100, Brazil

‡*Department of Physics, Institute of Biosciences, Letters and Exact Sciences*

São Paulo State University - UNESP, São José do Rio Preto, SP, 15054-000, Brazil

¶*These authors contributed equally to this work*

E-mail: vleite@sjrp.unesp.br

Phone: +55 (17) 3221-2240

Abstract

The TKSAMC is a web server which calculates protein charge-charge interactions via the Tanford-Kirkwood Surface Accessibility model with the Monte Carlo method for sampling different protein protonation states. The optimization of charge-charge interactions via directed mutations has successfully enhanced the thermal stability of different proteins and could be a key to protein engineering improvement. The server presents the electrostatic free energy contribution of each polar-charged residue to protein native state stability. The server also indicates which residues contribute to destabilizing the protein native state with positive energy and the side chain exposed to solvent. This residue is a candidate for mutation to increase protein thermostability as a function of the chosen pH condition. The web server is freely available at UNESP (São Paulo State University - DF/IBILCE): http://tksamc.df.ibilce.unesp.br.

1 Introduction

The thermal stability of proteins involves a balance of several interactions. A knowledge of the principals that govern these interactions is fundamental to develop biotechnological applications rationally.[1–3] Among these interactions, electrostatic effects are widely known to be crucial for thermal stability.[4–9] Solution conditions, such as salt concentration and pH tend to have a great influence on protein stability.[10–13]

Makhatadze and co-workers have explored the optimization of charge-charge interactions via directed mutations and have successfully enhanced the thermal stability of different proteins.[14–16] In the present paper, a Web Server that uses a similar method for predicting mutations for the enhancement of thermal stability is presented. This computational method consists in calculating the electrostatic free energy contribution of each ionizable residue of a protein using the Tanford-Kirkwood model with a correction that takes into account the solvent accessibility of these residues (TKSA).[17–19] All calculations are made for the protein

in its native state (see further details in *Material Methods*). The difference between the approach presented here and that used by Makhatadze and co-workers lies in the way the ionizable states of the residues are sampled. In the present study, the charges are sampled by the Monte Carlo method, which was found to be faster than the charge sample made using the Genetic Algorithm.

The TKSAMC is a web server which calculates the protein charge-charge interaction rapidly and consistently. The server provides the residues that are not electrostatically favorable to protein stability (possible target for mutations). The user can also select the temperature and the pH range for the calculations. The web server is freely available at UNESP (São Paulo State University - DF/IBILCE): http://tksamc.df.ibilce.unesp.br.

2 MATERIALS AND METHODS

2.1 Input files

The main input of this server is a file that contains the coordinates of the atoms of the protein's native structure. This file can be uploaded in the standard format of Protein Data Bank atom position. A sample of the PDB file can also be downloaded as an example in the server initial page. The most important parameter in TKSA-MC calculation is the pH of a solution; users must choose between a single pH value or a range of pH. If the users opt for a range of pH, the TKSA-MC is calculated for every 0.5 pH unit within the chosen pH range. Another input parameter is temperature, which can be specified or not; the default temperature in this server is 300K. Finally, users can opt to provide an email address to be notified when a given calculation is concluded.

2.2 Server calculation - TKSA-MC

In the Tanford-Kirkwood model (TK), the protein is treated as a spherical cavity with dielectric constant, ϵ_p and radius b, surrounded by an electrolyte solution modeled according to the Debye-Hückel theory.[17] Shire *et al* modified this model incorporating a solvent static accessibility correction for each ionizable residue (TKSA model).[18] This modification was mostly introduced to take into account the irregular protein-solvent interface. The present model is now referred to as the Tanford-Kirkwood model with a solvent accessibility (TKSA).[18,20,21] The interaction energy between two charged residues is given by,

$$U_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}),$$ (1)

where e is the elementary charge; b is the radius of the protein; a is the closest possible approach distance of an ion; A_{ij}, B_{ij} and C_{ij} are parameters obtained from the analytical solution of the Tanford and Kirkwood model, which are functions of the distance between ionizable residues, the dielectric constants, and the ionic strength;[17,22,23] and SA_{ij} is the average of the solvent-accessible surface area of residues i and j.[18,20,24]

Once the electrostatic energy between the titratable residues has been determined, it is possible to calculate the free energy, $\Delta G_N(\chi)$, for the protein native state in a given state of protonation χ:[25,26]

$$\Delta G_N(\chi) = -RT(\ln 10) \sum_{i=1}^{n} (q_i + x_i) pK_{a,ref,i} + \frac{1}{2} \sum_{i,j=1}^{n} U_{ij}(q_i + x_i)(q_j + x_j),$$ (2)

where R is the ideal gas constant; T is the temperature; q_i is the charge of the ionizable residue i in its deprotonated state; x_i is 0 or 1, according to the protonation state of the residue i; and $pK_{a,ref,i}$ is the reference pK_a. The probability that the protein will be in its native state with a particular state of protonation, $\rho_N(\chi)$, is given by,

$$\rho_N(\chi) = \frac{1}{Z_N} \exp \left[-\frac{\Delta G_N(\chi)}{RT} - \nu(\chi)(\ln 10)pH \right],$$ (3)

4

where $\nu(\chi)$ is the number of ionizable residues that are protonated in the state of protonation χ, and Z_N is the partition function for the protein in its native state. In the method, it is assumed that the unfolded state does not contribute to the electrostatic interactions.[27] The protonation states χ are calculated via Metropolis Monte Carlo. The mean total electrostatic energy, $\langle W_{qq} \rangle$ is computed by the average over these protonation states, taking into account Eq. 3:

$$\langle W_{qq} \rangle = \sum_{\chi} \left[\frac{1}{2} \sum_{i,j=1}^{n} U_{ij}(q_i + x_i)(q_j + x_j) \right] \rho_N(\chi). \tag{4}$$

Ibarra-Molero and Makhatadze have shown that the contribution of electrostatic interaction energy to free energy, ΔG_{qq}, can be described by $\Delta G_{qq} \approx -\langle W_{qq} \rangle$.[26]

2.3 Output files

For each submitted job a report is generated. The result page has a specific link which is sent to the e-mail provided by the user. All the data used in the method calculation and the plotted figures can be downloaded as a compressed folder in a link at the top of the page. An output file with the description of the charge positions, the reference pK_a, the normalized solvent-accessibility surface area, the energy contribution of each residue and the total energy is also available. If the submitted job is a single pH run, the result page presents a ΔG_{qq} bar graph, like that shown in Figure 1. If the submitted job uses the pH range calculation option, the result page gives the ΔG_{qq} bar graph at the top and the ΔG_{elec} as a function of pH at the bottom, as shown in Figure 2.

3   RESULTS

3.1 ΔG_{qq} bar graphs analysis

The TKSA-MC method indicates the residues which contribute to destabilizing the protein native state. The algorithm calculates the protein electrostatic energy, taking into account the contribution of each residue with a polar-charged side chain. Figure 1 presents an example of the ΔG_{qq} bar graph for ubiquitin.[27]
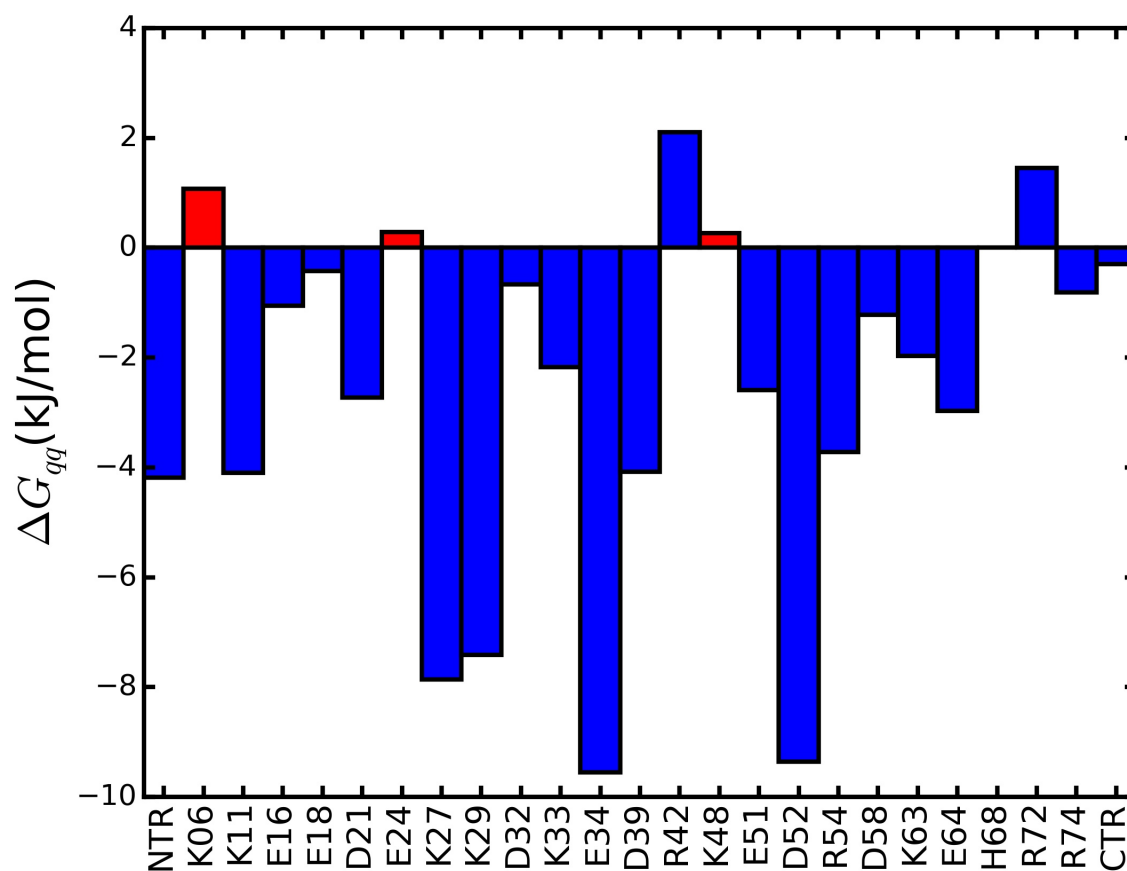

Figure 1: Charge–charge interaction energy ΔG_{qq} calculated by the TKSA-MC model for each ionizable residue, in this example, for ubiquitin. The simulation parameter values were pH 5.5 and temperature 300K. The ubiquitin structure used was PDB ID: 1UBQ. It has 25 ionizable groups, in which NTR and CTR are the N and C terminals, respectively.

The bars indicate the charge-charge energy contribution of each ionizable residue to protein native state stability. The red bars indicate the candidate residues to be mutated to increase protein thermostability. The selection criteria follow the methodology introduced by Ibarra-Molero et al.,[26] which indicates the residues that present unfavorable energy values $\Delta G_{qq} \geq 0$ and are exposed to solvent with SASA $\geq 50\%$. In the ubiquitin example, the highlighted residues are K06, E24 and K48, and this information is also printed on the result screen page with its respective ΔG_{qq} and SASA values. All the data for plotting the ΔG_{qq} bar graph figure is available for download in the output files. The total energy contribution to protein native state stability is given as a sum of the ΔG_{qq} for each ionizable group, named as ΔG_{elec}. The ΔG_{elec} the ubiquitin in pH 5.5 is -61.99kJ/mol.

3.2 ΔG_{elec} and pH dependence

The total energy contribution ΔG_{elec} as well as the ionization degree of each polar side chain residue and its ΔG_{qq} are pH dependent. Figure 2 presents an example of ΔG_{elec} as a function of pH for ubiquitin. The pH range in the example shown in Figure 2 was from 2 to 12 in steps of 0.5 in a total of 21 TKSA-MC runs. All the data used to plot the ΔG_{elec} curve and the respective pH bar graph figures are available for download in the output files.
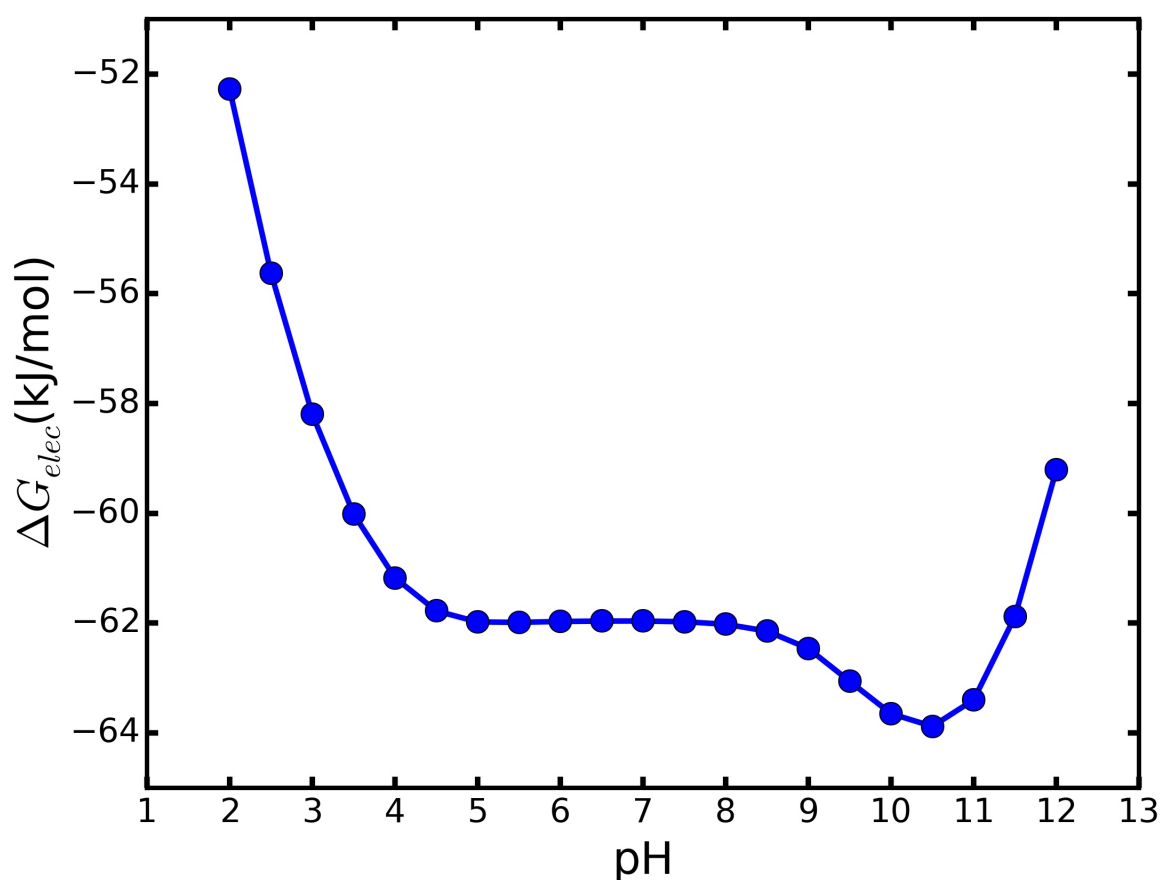

Figure 2: ΔG_{elec} as a function of pH calculated from pH 2 to 12 for ubiquitin in a total of 21 TKSA-MC runs at a temperature of 300K.

4 CONCLUDING REMARKS

The TKSA-MC method is a Monte Carlo/Metropolis based approach to sampling the ion-izable protein groups. This method is an adaptation of the genetic algorithm sampling approach TKSA-GA method introduce by Ibarra-Molero et al..[26] The web server provides a fast calculation of the charge-charge interaction using the Tanford-Kirkwood model protein approach with pre-selected pH and temperature values.

As a result, the TKSA-MC indicates the residues which contribute to destabilizing the protein native state. The charge-charge interaction optimization by mutation is applied to protein engineering.[9,16,26–29] The indicated residue is a candidate for mutation for increasing the protein thermostability in the chosen pH and temperature conditions. It is recommended that the TKSA-MC also be run for the mutated protein, and the pH range option be run to check the pH-dependence of the protein stability due to electrostatic contribution ΔG_{elec}. The server also has a section with a discussion of the expected results and how to interpret them, as well a section containing the theory and methodology details. A personalized support via e-mail is also provided.

5 FUNDING

5.0.1 Conflict of interest statement.

None declared.

References

(1) Roca, M.; Liu, H.; Messer, B.; Warshel, A. On the Relationship between Thermal Stability and Catalytic Power of Enzymes. *Biochemistry* **2007**, *46*, 15076–15088.

(2) Bye, J.; Platts, L.; Falconer, R. Biopharmaceutical liquid formulation: a review of the science of protein stability and solubility in aqueous environments. *Biotech. Lett.* **2014**, *36*, 869–875.

(3) Huber, K.; Olek, K.; Müller, A.; Tan, C.; Bennett, K.; Colinge, J.; Superti-Furga, G. Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. *Nature Methods* **2015**, *12*, 1055–1057.

(4) Tripathi, S.; Garcìa, A. E.; Makhatadze, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J. Phys. Chem. B* **2015**, *119*, 13103–13112.

(5) Wunderlich, M.; Martin†, A.; Schmid, F. Stabilization of the Cold Shock Protein CspB from Bacillus subtilis by Evolutionary Optimization of Coulombic Interactions. *J. Mol. Biol.* **2005**,

(6) Papaleoa, E.; Olufsena, M.; De Gioiab, L.; Brandsdala, B. Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. *J. Mol. Graph. Mod.* **2007**, *26*, 93–103.

(7) Contessoto, V. G.; de Oliveira, V. M.; de Carvalho, S. J.; Oliveira, L. C.; Leite, V. B. P. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.* **2016**, *12*, 3270–3277.

(8) Coronado, M.; Caruso, I.; De Oliveira, V.; Contessoto, V.; Leite, V.; Arni, R.; Eberle, R. Cold shock protein A from Corynebacterium pseudotuberculosis: Role of electrostatic forces in the stability of the secondary structure. *Protein Pept. Lett.* **2017**,

(9) de Oliveira, V. M.; Contessoto, V. G.; Fernandes, B. R.; Slade, G. G.; Leite, V. B. P. Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models. *Biophys J. - In Press*

(10) Lindman, S.; Xue, W.-F.; Szczepankiewicz, O.; Bauer, M. C.; Nilsson, H.; Linse, S. Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophys. J.* **2006**, *90*, 2911–2921.

(11) Yamasaki, M.; Yano, H. Differential scanning calorimetric studies on bovine serum albumin: I. Effects of pH and ionic strength. *Int. J. Biol. Macr.* **1990**, *12*, 263–268.

(12) Kim, K.; Kim, S.; Yang, H.; Kwon, D. Changes of glycinin conformation due to pH, heat and salt determined by differential scanning calorimetry and circular dichroism. *Int. J. Food Sci. Tech.* **2004**, *39*, 385–393.

(13) Bye, J.; Falconer, R. Thermal stability of lysozyme as a function of ion concentration: A reappraisal of the relationship between the Hofmeister series and protein stability. *Prot. Sci.* **2013**, *22*, 1563–1570.

(14) Makhatadze, G. I. Linking computation and experiments to study the role of charge–charge interactions in protein folding and stability. *Physical Biology* **2017**, *14*, 013002.

(15) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C.; Makhatadze, G. I. Rational stabilization of enzymes by computational redesign of surface charge–charge interactions. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2601–2606.

(16) Permyakov, S. E.; Makhatadze, G. I.; Owenius, R.; Uversky, V. N.; Brooks, C. L.; Permyakov, E. A.; Berliner, L. J. How to improve nature: study of the electrostatic properties of the surface of alpha-lactalbumin. *Protein Eng., Des. Sel.* **2005**, *18*, 425–433.

(17) Tanford, C.; Kirkwood, J. G. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.

(18) Shire, S.; Hanania, G.; Gurd, F. N. Electrostatic effects in myoglobin. Hydrogen ion equilibriums in sperm whale ferrimyoglobin. *Biochemistry* **1974**, *13*, 2967–2974.

(19) Gopi, S.; Devanshu, D.; Krishna, P.; Naganathan, A. N. pStab: prediction of stable mutants, unfolding curves, stability maps and protein electrostatic frustration. *Bioinformatics*

(20) Tanford, C.; Roxby, R. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry* **1972**, *11*, 2192–2198.

(21) Orttung, W. H. Proton binding and dipole moment of hemoglobin. Refined calculations. *Biochemistry* **1970**, *9*, 2394–2402.

(22) Kirkwood, J. G. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **1934**, *2*, 351–361.

(23) Da Silva, F. L. B.; Jönsson, B.; Penfold, R. A critical investigation of the Tanford-Kirkwood shceme by means of Monte Carlo simulations. *Protein Sci.* **2001**, *10*, 1415–1425.

(24) Havranek, J. J.; Harbury, P. B. Tanford–Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11145–11150.

(25) Bashford, D.; Karplus, M. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* **1991**, *95*, 9556–9561.

(26) Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I.; Sanchez-Ruiz, J. M. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **1999**, *38*, 8138–8149.

(27) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. Protein stability and surface electro- statics: a charged relationship. *Biochemistry* **2006**, *45*, 2761–2766.

(28) Ibarra-Molero, B.; Sanchez-Ruiz, J. M. Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins. *J. Phys. Chem. B* **2002**, *106*, 6609–6613.

(29) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. Mechanism of Ther- mostabilization in a Designed Cold Shock Protein with Optimized Surface Electrostatic Interactions. *J. Mol. Biol.* **2004**, *336*, 929–942.