MUTANTELEC: AN *IN SILICO* MUTATION SIMULATION PLATFORM FOR COMPARATIVE ELECTROSTATIC POTENTIAL PROFILING OF PROTEINS.

Braulio Valdebenito-Maturana[1], Jose Antonio Reyes-Suarez[1], Jaime Henriquez[1], David S. Holmes[2,3], Raquel Quatrini[2,3], Ehmke Pohl[4,5,6], Mauricio Arenas-Salinas[1]

Correspondence to: Mauricio Arenas-Salinas (Email: mauricioa@gmail.com)

[1] Centro de Bioinformática y Simulación Molecular, Facultad de Ingeniería, Universidad de Talca, Talca, Chile, 346 5548.

[2] Fundación Ciencia & Vida, Santiago, Chile. 778 0272.

[3] Facultad de Ciencias Biologicas, Universidad Andres Bello, Santiago, Chile.

[4] Department of Chemistry, Durham University, Durham DH1 3LE, UK

[5] Department of Biosciences, Durham University, Durham DH1 3LE, UK

[6] Biophysical Sciences Institute, Durham University, Durham DH1 3LE, UK.

ABSTRACT:

The electrostatic potential plays a key role in many biological processes like determining the affinity of a ligand to a given protein target, and they are responsible for the catalytic activity of many enzymes. Understanding the effect that amino acid mutations will have on the electrostatic potential of a protein, will allow a thorough understanding of which residues are the most important in a protein. MutantElec, is a friendly web application for in silico generation of site-directed mutagenesis of proteins and the comparison of electrostatic potential between the wild type protein and the mutant(s), based on the three-dimensional structure of the protein. The effect of the mutation is evaluate using different approach to the traditional surface map. MutantElec provides a graphical display of the results that allows the visualization of changes occurring at close distance from the mutation and thus uncovers the local and global impact of a specific change.

Availability: http://structuralbio.utalca.cl/mutantelec/

INTRODUCTION

The electrostatic potential at the surface of biological macromolecules such as proteins and nucleic acids plays a key role in many biological processes. Electrostatic interactions (EIs) and hydrophobic interfaces guide substrates and ligands to their designated location, and govern all protein interactions with other macromolecules, [1], [2]. EIs play a key role in determining the affinity of a ligand to a given protein target, and they are responsible for the catalytic activity of many enzymes. One well-studied example is the *Saccharomyces cerevisiae* phosphoenolpyruvate carboxykinase (PEPCK) where the binding of the Mn^{+2} required for catalysis, is due to the EIs to the side chains of a Lys213 [3], [4]. The reversible interaction between protein and membrane is critical to many biological processes [5] and these associations have been shown to be partly mediated by electrostatic interactions [6], [7]. Also, electrostatic charge distribution of interacting protein surfaces determines the formation or stabilization of many protein complexes. Thereafter, the mutation of a few or even a single residue can induce the destabilization of the interface [8]. This is the case of the yeast mitochondrial malate dehydrogenase where mutation of His46, located in the α-C helix of the binding interface between monomers, results in the loss of polar interactions at the subunit interface and ultimately the dissociation of the dimer [9]. Substantial evidence of the effect of charge distributions on the functional characteristics of proteins can also be found in the literature. For example, the ability of the transcriptional regulators of the Ferric Uptake Regulator (Fur) family to bind and recognize specific DNA sequences at the promoter regions of their target genes is directly correlated with the electric charge distribution of the DNA binding site of these protein [10–12].

The electrostatic potential (EPs) of biological macromolecules can be estimated using Poisson-Boltzmann (PB) equations, and several software packages, such as APBS [13] and Delphi [14], have been developed to solve these equations. Many of these software take advantage of novel computational approaches, such as Grid computing for distributed calculations [15], and the use of graphic processor units (GPU) for electrostatic potential calculations [16], which enable calculations to be done in a matter of minutes. EPs are most conveniently displayed as color-coded surface representations using modern graphical programs such as

JSmol (http://www.jmol.org/), Pymol (http://www.pymol.org/), and VMD [17]. However, none of these tools reveals the net effect on the EP of the target proteins, nor on the global charge distribution of *in silico* mutant variants of the protein of interest.

Site-directed mutagenesis is a standard experimental technique used to generate site-specific mutations in known protein-coding genes. By deleting or substituting particular residues, the role of individual amino acids in the reaction mechanisms of enzymes or in the structural configuration of proteins can be studied [18–22]. However, this technique is very laborious and time consuming, and *in silico* tools to help direct the selection of candidate residues are scarce.

In spite of tremendous advances in molecular modeling and bioinformatics software, the *in-silico* design of single-site mutant variants and their evaluation requires significant expertise in various sophisticated bioinformatics tools designed for protein modeling and ligand analyses [24]. Molecular modelling software packages (e.g. VMD [17], ICM [25], SwissPDB [26], VegaZZ [27]) can generate the *in silico* mutants by replacement of residues, one at a time, but they do not calculate the new electrostatic surface automatically, and do not provide an automated pipeline to generate a series of mutations. Only recently, programs such as SAAMBE [28] and SAAFEC [29] have been developed, that allow the prediction of the changes in the free energy of binding and protein folding respectively, caused by amino acids mutations In addition, programs like BeAtMuSiC [30] and PoPMuSiC-2.0 [31] can predict the changes in protein–protein binding affinity and on protein stability as a consequence of *in silico* mutation. However, none of those programs allow the prediction of the changes that mutations will have in the EP of a protein. In turn, Delphi [14] performs versatile electrostatic potential calculations, but lacks the ability to generate *in silico* mutations. Importantly, most programs do not take into account non-natural amino acids, e.g., those are phosphorylated and dephosphorylated as part of the regulatory mechanism of the cell. Being such an important process in cellular regulatory networks [33,34], exclusion of phosphorylated residues is an important caveat that limits adequate study of the effect of mutations in EP calculations.

Despite the many virtues of the alluded software, most of them are not user friendly and require considerable expertise from the user. For example, some lack the option to input a PDB identifier to initiate the job. Others lack simplicity in their mutation interface: the users must specify manually the code, chain and number of the residue they wish to mutate, along with the new residue. In most applications this step is not automatically validated, and represents a source of errors to biologists with little or none computational knowledge. Furthermore, except for Delphi, none of the applications discussed has an online side-by-side comparison interface between the wild type and the mutant protein

To tackle the impairments alluded to above, we designed MutantElec, a web-based application for the study of the effects that mutations have in the EP of a protein of interest. It simulates *in silico* site-directed mutations and analyzes their effect on the EP distribution of the mutated protein. MutantElec provides a graphical display that allows the visualization of changes occurring at any distance from the introduced point mutation and, thus, uncovers both the local and global effects of a site-specific change. This application is user friend and easy to use, and has thus the potential to control frequent errors associated to misusage. We are confident that MutantElec will prove particularly valuable in the analysis of mutations involving amino acids with similar physicochemical properties and those that are target of posttranslational modifications (such as phosphorylation), where currently available graphical representations of the whole protein surface obscure small changes of the electrostatic potential.

METHODS

The MutantElect web application is based on an intuitive graphical user interface with three main component layers: 1) an INPUT layer, through which a protein structure in PDB format is selected and uploaded into the system, 2) an ANALYSIS layer comprising several subroutines that simulate the mutation(s) chosen by the user and estimates the electrostatic potential configuration of the wild type and mutated variants of the target protein, and 3) an OUTPUT layer, involving a set of subroutines for the display of the results in a user-friendly and intuitive graphical environment.

Multiple in-house scripts written in TCL and Perl programming languages were created to process and connect the input and output files of the various programs used along the pipeline, including the software used for mutation (Modeller, [32]), for EP calculations (APBS, [13]), as well as for automating the process (Figure 1).

INPUT layer: Consulting web interface

The first part of the system comprises the web interface that allows the user to choose the type of analysis and upload the target protein. The input data upload and consultation web interface was built using HTML, JavaScript and PHP languages, in a simple and intuitive web environment (Figure 2). The user can choose between two basic analyses: 1) site-specific mutagenesis and 2) mutagenesis of a specific region. The "site-specific mutagenesis" option enables the user to select the residue that is to be mutated and to specify the desired amino-acid exchange. Furthermore, through this option, any residue can be substituted by all the other 19 common amino acids using a scanning procedure (Figure 1S, Additional file 1), facilitating exploration of the most perturbing and least perturbing changes. In turn, the "mutagenesis of a specific region" option enables the user to mutate up to 10 amino acids in a row, in any selected region of the target protein (Figure 2S, Additional file 1). Alternatively, the user can upload a native and a mutant protein obtained experimentally, to compare the EPs of both using the tool without simulating mutations *in silico*.

After choosing the type of analysis, the UPLOAD query form can be accessed for the user to upload a PDB file containing the 3D coordinates of the protein of interest. The coordinate file can be obtained from the PDB database or from a model generated by the user through comparative modelling, following the instructions provided. The file can be then processed to check the amino acid composition. If a non-standard residue is found, a warning is displayed indicating this residue was deleted of the analysis. The user must modify the PDB file by exchanging or removing the non-standard residues (eg: HETATM, ligand, etc) previously. This is important because the force field calculations do not generate appropriate parameters for non-standard residues.

Once the PDB file has been uploaded, the web interface displays all the amino acids in the target protein as a list, ordered numerically from the N- to the C-terminus. The user has then the option to select the residues to be mutated and the corresponding substitutions. Most commonly phosphorylated residues (phosphoserine, phosphothreonine, phosphotyrosine) available for the Force Field CHARMM [35] are also considered in the MutantElec pipeline. Additionally, the user can modify the parameters used to calculate the electrostatic potential through the APBS software (dielectric constant of the biomolecule, dielectric constant of the solvent, temperature). The default options used in the subsequent analyses are those of proteins under mesophilic conditions having water as solvent, i.e. temperature = 25°C and dielectric constant of water = 78.5. Once the job is submitted, calculations are sent to the job queue, where the request is processed.

ANALYSIS layer: Mutation and electrostatic properties calculation

Amino acid mutation in the 3D structure of the protein is performed by the *Mutate* module from the Modeller software [33]. The conformation of the mutant sidechain is optimized by energy minimization (conjugate gradient) and refined using a small number of steps of molecular dynamics as implemented in Modeller [34]. The generated 3D coordinates are then processed using the PDB2PQR software [35] to assign charge and radius parameters for each atom. This information is stored in the *pqr* file and used for further calculation of the electrostatic potential using the Poisson-Boltzmann (PB) equations as implemented in the APBS software [13]. Then, the ANALYSIS subroutines estimate the electrostatic potential of each amino acid residue in the input protein and the mutant variant, based on the known structural conformation of the input protein. Next, the variation in the electrostatic potential distribution caused by the mutation introduced is calculated. These changes are not only related to the point mutation introduced in the target protein (i.e. change of electrostatic charge associated to a single residue), but also to conformational changes of the mutant variant. These changes need to be assessed before the EP distribution of the mutant protein is recalculated. For this purpose the system does a fast optimization using the *Optimizers Module* from Modeller [36] through the conjugate gradients method and molecular dynamics simulation. After the spatial conformation of the mutant variant is estimated, its EP distribution is calculated. A detailed report of the observed changes in the EP associated

with the point mutation is then provided to the user. Multiple single mutations can be queried simultaneously without limitation and results are conveniently delivered as independent reports for each request processed.

OUTPUT layer: Information retrieval

The results are processed to generate files containing the EP per residue per protein. This is done because the APBS results contain the EP per atom, and having the information in this format would make interpreting the results a difficult task. For this, an in-house script was written, that takes the output from APBS, and calculates the "per residue EP" by summing the individual contributions of each atom of every residue of a given protein. Results are presented as numerical and graphical outputs, displayed online as a set of charts which plot the per residue EP distribution of the input protein (Figure 3Sa, Additional file 1) and the mutant variant (Figure 3Sb, Additional file 1), individually or combined (Figure 3Sc, Additional file 1). The difference in EP between the input protein and the mutant variant generated by MutantElec are also displayed (Figure 3Sd, Additional file 1). A summative chart is plotted, which integrates all the above results in a single figure (Figure 3A) which is accompanied by a close-up representation of the mutation site and its neighboring residues within a sphere of selection of 15Å (a default value, that can be manipulated at will by the user) This is exemplified in Figure 3B, using the Ferric uptake regulator from *P. aeruginosa* Fur_{PA} (PDB_ID:1MZB) [12], and a simulated substitution of residue Glu100 for Ile.

To evaluate the significance of the differences in the electrostatic potential uncovered between the input and the mutant protein, MutantElec performs the non-parametric Wilcoxon Signed-Rank Test, with a confidence level of 0.05 [37]. Additionally, the electrostatic potential maps of the surfaces are produced using JSmol (Figure 4S, Additional file 1). This facilitates the analysis of the results without the need to install and master additional programs, allowing the user to identify all potentially important changes in specific regions upon mutation, and to predict the impact of site directed mutations on protein structure and/or function. Finally, the application generates a compressed file (file *.tar.gz) with all the results, for the users to download and perform their own local analyses if deemed necessary. This file also contains the input and output files that were generated during the analysis run (Table 1). The user is notified via an email that includes links to

preview and download of the results. All results are also available in plain text format and can be exported into a spreadsheet for further analysis.

Website

MutantElec is freely available for non-commercial use at http://structuralbio.utalca.cl/mutantelec/. This server is supported by the Center of Bioinformatics at the University of Talca, and will be constantly updated and maintained to ensure reliable and continuous operation.

EXAMPLE: p53 protein.

The tumor suppressor protein p53 is a small transcription factor (393 amino acid) that binds to specific DNA sequences and regulates the expression of genes involved in DNA repair, cell cycle arrest, and apoptotic cell death [38]. This protein has been shown to play a key role in many human cancers and it is now estimated that approximately 50% of human tumors contain mutations in this gene [39]. It has several functional domains, including a transcriptional activation domain at its N-terminus, a sequence-specific DNA-binding domain (core domain), an oligomerization domain and a regulatory basic domain at the C-terminus. A wide range of structural and computational studies have contributed to unraveling the structural basis of activation and DNA binding [38,40,41].

Alterations in the human p53 protein have been shown to result in a partial or total loss of its ability to bind DNA and correlated with increased probabilities of developing tumors [42]. This has triggered growing interests in characterizing the structural effects that point mutations have on p53 function and cancer development [38,43–46]. However, evaluation of these mutations require complex analyses in the areas of biochemistry, molecular biology and biophysics that are typically time-consuming and are only possible with significant resources [39]. The results of such studies show that certain mutations in p53 can affect its structure or modify its non-covalent interactions, causing conformational changes that lead to non-functional proteins. [39]. In turn, tools like CellDesigner, have been used successfully to aid in the selection and

evaluation of drug targets for p53 [47]. However, this approach does not address the effects of the point mutations introduced on the proteins EP distribution or its potential effects.

As an example of the capabilities of MutantElec, the potential effect of mutating residue Arg249 in the DNA-binding domain of p53 was evaluated. This residue has one of the highest mutation rates present in patients that have developed cancer [48]. In the IARC TP53 database [49] the following mutations of residue 249 have been described R249S/G/I/K/M/N/T/W. Interestingly, although residue R249 does not directly mediate p53 binding to DNA [46], its mutation inactivates its function, suggesting that alterations of the neighboring regions caused by the point mutation.

Using MutantElec and the PDB_ID 1TUP [28] containing the 3D structure of the human p53 core domain as input protein, all the above mentioned mutations were generated *in silico*. The Arg249 residue was replaced with all of the other 19 natural amino-acids, using the "scanning" option of the program (Figure 1S, Additional file 1) and the EPs and the difference in EP between the wild type and each mutant versions of p53 were calculated and analyzed. The changes in the EP profiles for residues located less than 15Å apart from Arg249 are shown in Figure 4. The majority of mutations resulted in negative EP differences in the neighborhood of Arg249, with Glu249 and Asp249 being the two mutations that produced the most significant EP changes in the vicinity of the target residue. The profiles obtained indicate that these changes have an impact beyond the mutated residue, an effect that can also be observed in the electrostatic potential map (Figure 4S, Additional file 1). Further analyses on the effect of mutations of Arg249 with known inactivating effects on p53 (R249S/G/I/K/M/N/T/W) are displayed in Figure 5. The residues that produced major local changes in the EP were the negatively charged ones (Glu and Asp). The Lys249 mutation, in turn, was the only one to produce a positive change in the local EP.

The analysis with MutantElec also revealed which amino acids were most affected by the *in silico* mutation procedure. These residues were Tyr163, His168, Met246, Glu171, Ser166, Glu285 in decreasing order of magnitude, all of which have polar side chains with the exception of Met246 (Figure 6A). The spatial distribution of these residues (Figure 6B) reveals that they are less than 5Å away from the mutated Arg249

with the sole exception of residue Glu285 (<10Å). According to the IARC TP53 database, the most frequently encountered missense mutations that inactivate p53 are Met246, Glu171, His168 and Tyr163 (Table 2S, Supplementary material). These same four mutations, plus Val173 and Ser166, were uncovered by MutantElec as the ones with the most perturbing effects in the electrostatic potential of the DNA binding domain, and thus potentially affecting the capacity of p53 to bind its target DNA sequence. This provides an example of the value of MutantElec to predict and explain important biological changes due to single site variations or systematic scanning of perturbations in a protein sequence.

CONCLUSIONS

MutantElec is a rapid and simple bioinformatic tool that can be used to predict and evaluate changes in the charge distribution of a protein after mutating, *in silico*, one or several amino acids. This information can be extremely useful for understanding the contribution and importance of specific amino acids to protein function. In addition, MutantElec can aid the understanding of how mutations can cause malfunctions of enzymes and the resulting physiological changes that cause diseases. It is also a useful tool for the rational design of site-directed mutagenesis experiments. MutantElec is user friendly and can be used by scientists who do not have extensive training in bioinformatics and structural biology. It is expected that MutantElec will be a useful tool for teaching and training in protein science and medicine.

ACKNOWLEDGEMENTS

MutantElec is supported by Center of Bioinformatics from the University of Talca and will be constantly updated and maintained to ensure reliable operation.

This work received financial support from FONDECYT N° 1140048 and 1130683, Basal CCTE PFB16, the Universidad de Talca funds for Research Initiation (Fondo de Proyectos de Investigación para Investigadores Iniciales) and a Santander Mobility Grant for collaboration with the Durham University, UK.

KEYWORDS: Electrostatic potential, site-directed mutagenesis, protein engineering, bioinformatics, rational protein design, p53, PDB.

REFERENCES

1. Jayaram B, Sharp K a, Honig B. Biopolymers. 1989;28:975–93.

2. Kahraman A, Morris RJ, Laskowski R a, Thornton JM. J. Mol. Biol. 2007;368:283–301.

3. Sepúlveda C, Poch A, Espinoza R, Cardemil E. Biochimie. 2010;92:814–9.

4. Yévenes A, Espinoza R, Rivas-Pardo J a, Villarreal JM, González-Nilo FD, Cardemil E. Biochimie. 2006;88:663–72.

5. Mulgrewnesbitt A, Diraviyam K, Wang J, Singh S, Murray P, Li Z, et al. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. 2006;1761:812–26.

6. Cho W, Stahelin R V. Annu. Rev. Biophys. Biomol. Struct. 2005;34:119–51.

7. Hurley JH, Misra S. Annu. Rev. Biophys. Biomol. Struct. 2000;29:49–79.

8. Jones S, Thornton JM. 1996;93:13–20.

9. Steffan JS, McAlister-Henn L. Arch. Biochem. Biophys. 1991;287:276–82.

10. Deng Z, Wang Q, Liu Z, Zhang M, Machado ACD, Chiu T-P, et al. Nat. Commun. 2015; 6:7642.

11. Sheikh A, Taylor GL. Mol. Microbiol. 2009;72:1208–20.

12. Pohl E, Haller JC, Mijovilovich A, Meyer-Klaucke W, Garman E, Vasil ML. Mol. Microbiol. 2003;47:903–15.

13. Baker NA, Sept D, Joseph S, Holst MJ, Mccammon JA. Proc. Natl. Acad. Sci. 2001; 98:10037–41.

14. Rocchia W, Alexov E, Honig B. J. Phys. Chem. B. 2001;105:6507–14.

15. Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, et al. J. Comput. Chem. 2011 ;32:1488–91.

16. Hardy DJ, Stone JE, Schulten K. Parallel Comput. 2009;35:164–77.

17. Humphrey W, Dalke A, Schulten K. J. Mol. Graph. 1996;14:33–8.

18. Lorentzen E, Siebers B, Hensel R, Pohl E. Biochemistry. 2005;44:4222–9.

19. Villarreal JM, Bueno C, Arenas F, Jabalquinto AM, González-Nilo FD, Encinas M V, et al. Int. J. Biochem. Cell Biol. 2006;38:576–88.

20. Caniuguir A, Cabrera R, Báez M, Vásquez CC, Babul J, Guixé V. FEBS Lett. 2005;579:2313–8.

21. Carvacho I, Gonzalez W, Torres YP, Brauchi S, Alvarez O, Gonzalez-Nilo FD, et al. J. Gen. Physiol.. 2008;131:147–61.

22. Abarca-Lagunas MJ, Rivas-Pardo JA, Ramírez-Sarmiento C a., Guixé V. FEBS Lett. 2015; 015;589:3271–6.

23. Saravanan P, Dubey VK, Patra S. J. Mol. Model. 2014;20:2501.

24. Arenas-Salinas M, Ortega-Salazar S, Gonzales-Nilo F, Pohl E, Holmes DS, Quatrini R. J. Comput. Aided. Mol. Des. 2014;28:1069–76.

25. Abagyan R, Totrov M, Kuznetsov D. J. Comput. Chem. 1994;15:488–506.

26. Guex N, Peitsch MC. Electrophoresis. 1997;18:2714–2723.

27. Pedretti A, Villa L, Vistoli G. Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta). 2003;109:229–32.

28. Petukh M, Dai L, Alexov E. Int. J. Mol. Sci. 2016;17:547.

29. Getov I, Petukh M, Alexov E. Int. J. Mol. Sci. 2016;17:512.

30. Dehouck Y, Kwasigroch JM, Rooman M, Gilis D. Nucleic Acids Res. 2013;41:333–9.

31. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M. Bioinformatics. 2009;25:2537–43.

32. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Methods Mol. Biol. 2008;426:145–59.

33. Sali A, Blundell TL. J. Mol. Biol. 1993;234:779–815.

34. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Annu. Rev. Biophys. Biomol. Struct. 2000;29:291–325.

35. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. Nucleic Acids Res. 2007];35:W522-5.

36. Fiser A, Do RKG, Šali A. Protein Sci. 2000;9:1753–73.

37. Wilcoxon F. J. Econ. Entomol. 1946;39:269.

38. Suad O, Rozenberg H, Brosh R, Diskin-Posner Y, Kessler N, Shimon LJW, et al. J. Mol. Biol. Elsevier Ltd; 2009;385:249–65.

39. Lu Q, Tan Y, Luo R. J. Phys. Chem. B. 2007;111:11538–45.

40. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Science. 1994;265:346–55.

41. May P, May E. Oncogene. 1999;18:7621–36.

42. Sidransky D, Mikkelsen T. Nature. 1992;355:846–7.

43. Smith P, Zhou B, Ho N, Yuan YC, Su L, Tsai SC, et al. Biochemistry. 2009;48:11134–41.

44. Emamzadah S, Tropia L, Vincenti I, Falquet B, Halazonetis TD. J. Mol. Biol. Elsevier B.V.; 2014;426:936–44.

45. Wallentine BD, Wang Y, Tretyachenko-Ladokhina V, Tan M, Senear DF, Luecke H. Acta Crystallogr. Sect. D Biol. Crystallogr. International Union of Crystallography; 2013;69:2146–56.

46. Eldar A, Rozenberg H, Diskin-Posner Y, Rohs R, Shakked Z. Nucleic Acids Res. 2013;41:8748–59.

47. Aguiló Llobet J, Freire Veiga AM, De la Iglesia Jiménez D, López Alonso V, Pazoz Sierra A. Eds., Ciencia y Tecnología para el Desarrollo (CYTED):Madrid, 2013, pp 75-82.

48. Golubovskaya VM, Finch R, Kweh F, Massoll N a., Campbell-Thompson M, Wallace MR, et al. Mol. Carcinog. 2008;47:373–82.

49. Mandey SHL, Schneiders MS, Koster JÃ, Waterham HR. Hum. Mutat. 2006;27:796–802.

50. Bode AM, Dong Z. Nat. Rev. Cancer. 2004;4:793–805.

FIGURE

Figure 1 workflow of the MutantElect web application. The three layers are color-coded and the components of each layer are represented accordingly. External software packages used in the pipeline are represented by the parallelogram while in-house developments (scripts and subroutines) are shown as rectangles. INPUT

layer (N° 1). Routine used to upload the input PDB file and to define the mutations to be performed. ANALYSIS layer (N° 2). Module employed in the generation of the *in silico* mutations and the calculation of the electrostatic potential. OUTPUT layer (N° 3). Subroutines in charge of performing the statistical analyses and comparisons and in generating the output.

Figure 2: Consulting web interface of MutantElec. A: Analysis options, namely "*site-specific"* or *"specific region"* of the target protein and entry form for input data upload. B: Calculation parameters and default values (temperature = 298.15 K (mesophilic conditions); dielectric constant of the water = 78.54; distance cutoff (radio of selection for the analysis) from the residue mutated = 15Å. C: Site-directed mutations selection scroll-down menu. The option "All (Scanning)" allows the calculation to be repeated 19 times for every other of the 19 possible residues.

Figure 3: MutantElec analyzes. A: Combinatorial chart displaying the electrostatic potential of the input and mutant proteins, and the difference in potential between both. B: Scheme of the environment surrounding the mutated residue (Glu100Ile) for the protein Fur_{PA} (PDB_ID:1MZB) [12]. In red are residues around the mutated residue (in blue) within the selected sphere of 15Å ratio used for the analysis. This distance can be modified by the user in the parameter setting web page with the option "Distance cutoff".

Figure 4: Difference in the calculated electrostatic potential between "wild type" p53 and 19 different mutants of residue R249. The change in the profile is shown for every residue located within 15Å of the mutated residue. These residues are ordered according to the distance to the R249 starting from the nearest one. The profiles indicate that the changes have an impact beyond the position of the mutation altering residues found in distinct functional domains of the p53 protein, this is shown in the highest peaks. In the bottom of the figure is shown a schema of the domain and residues present in p53 [50].

Figure 5: Calculated electrostatic potential for the residue Arg249 and the 19 variants generated with the scanning mode of the MutantElec system using the input PDB_ID:1TUP [40]. It is possible to observe the major change in the electrostatic potential for the residues with negative charge like Asp y Glu.

Figure 6: Analysis of the more significant change for the residue 249 for p53 protein. A: show amino acids that were most affected by the respective *in silico* mutations with respect to the wild type protein, this are the residues Tyr163, His168, Met246, Glu171, Ser166, Glu285 in decreasing order of magnitude. B: Spatial distribution of these residues, it is possible to observe some of these residues are located at less than 5Å to the mutated Arg249 except the residue Glu285 (>10Å).

Table 1: Description of the files that are sent to the user after the calculations have finished.

File Name	Description
protein.in	Configuration file for APBS calculations
protein.pdb	Coordinate file uploaded by the user
proteinWT-Mutant.pdb	Coordinate file for the mutant protein generated by Mutator.
proteinWT-Mutant.pqr	File containing information about atomic charge and radius information, necessary for the calculations with APBS software.
_aminopot.txt	Electrostatic potential for each amino acid expressed in mVolts.
_atompot.txt	Electrostatic potential for each atom expressed in mVolts.
_map.dx	Electrostatic maps to display in Pymol or VMD
_ distances.txt	File with the distances from the mutated residue and the remaining amino acids of the protein
_detailedValues.csv	Electrostatic potential differences between the input and mutant protein
_testResults.txt	File with the result of the Wilcoxon Signed-Rank Test

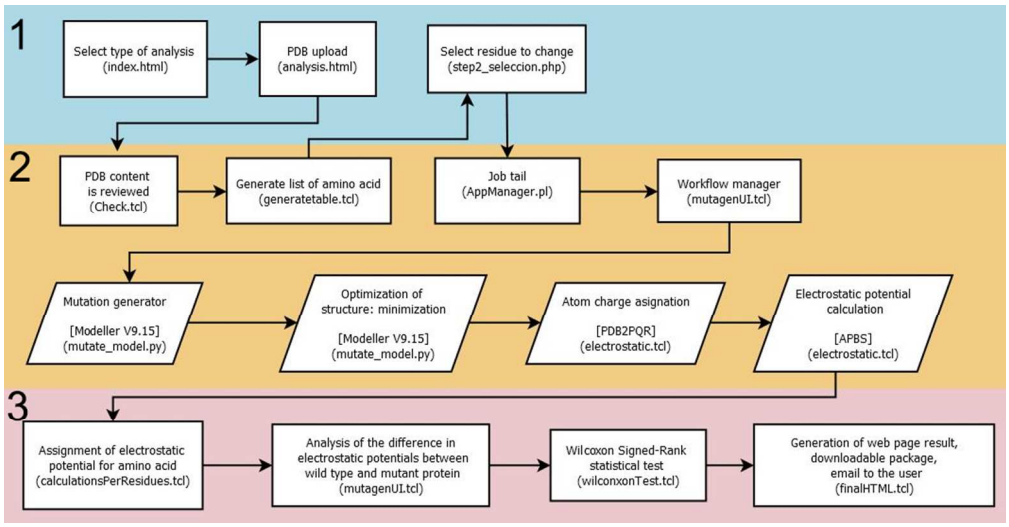

Figure 1 workflow of the MutantElect web application. The three layers are color-coded and the components of each layer are represented accordingly. External software packages used in the pipeline are represented by the parallelogram while in-house developments (scripts and subroutines) are shown as rectangles. INPUT layer (N° 1). Routine used to upload the input PDB file and to define the mutations to be performed. ANALYSIS layer (N° 2). Module employed in the generation of the in silico mutations and the calculation of the electrostatic potential. OUTPUT layer (N° 3). Subroutines in charge of performing the statistical analyses and comparisons and in generating the output.

Figure 1
373x192mm (72 x 72 DPI)

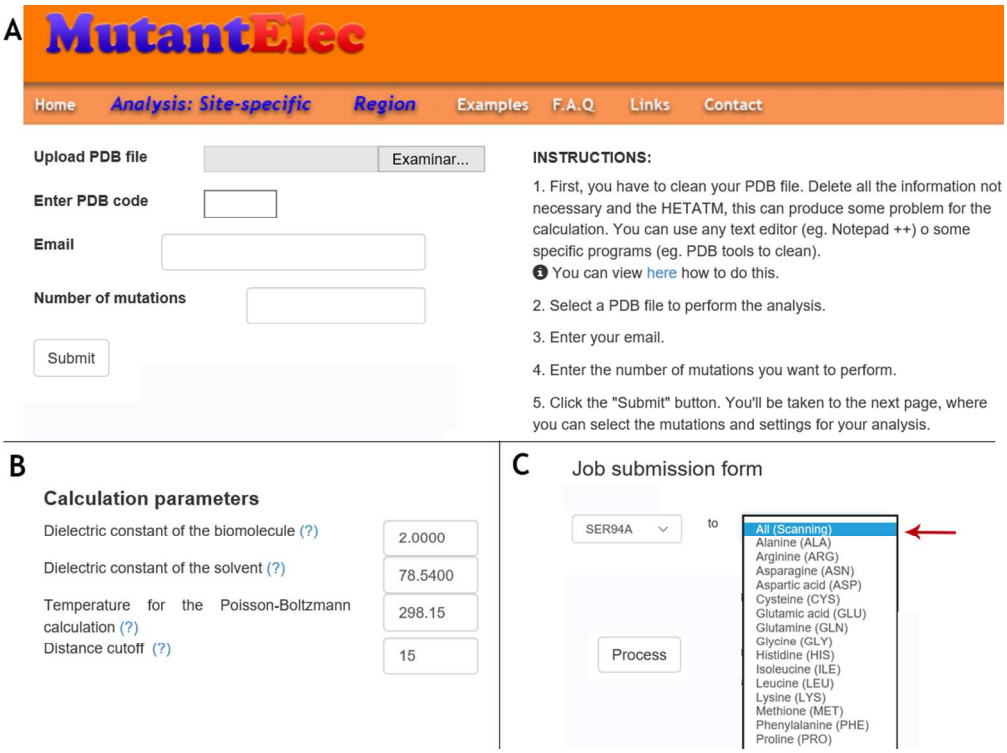

Figure 2: Consulting web interface of MutantElec. A: Analysis options, namely "site-specific" or "specific region" of the target protein and entry form for input data upload. B: Calculation parameters and default values (temperature = 298.15 K (mesophilic conditions); dielectric constant of the water = 78.54; distance cutoff (radio of selection for the analysis) from the residue mutated = 15Å. C: Site-directed mutations selection scroll-down menu. The option "All (Scanning)" allows the calculation to be repeated 19 times for every other of the 19 possible residues.

Figure 2

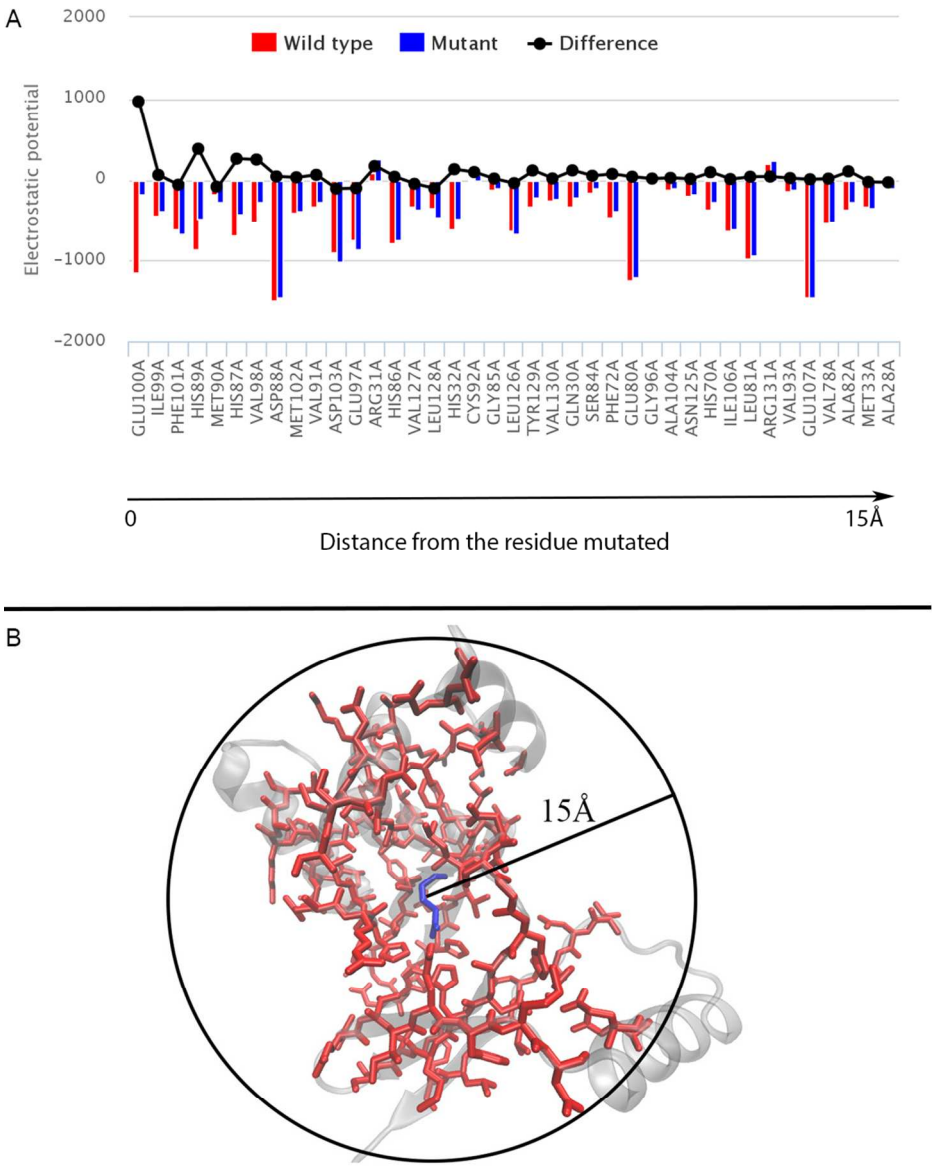

Figure 3: MutantElec analyzes. A: Combinatorial chart displaying the electrostatic potential of the input and mutant proteins, and the difference in potential between both. B: Scheme of the environment surrounding the mutated residue (Glu100Ile) for the protein FurPA (PDB_ID:1MZB) [12]. In red are residues around the mutated residue (in blue) within the selected sphere of 15Å ratio used for the analysis. This distance can be modified by the user in the parameter setting web page with the option "Distance cutoff".
Figure 3

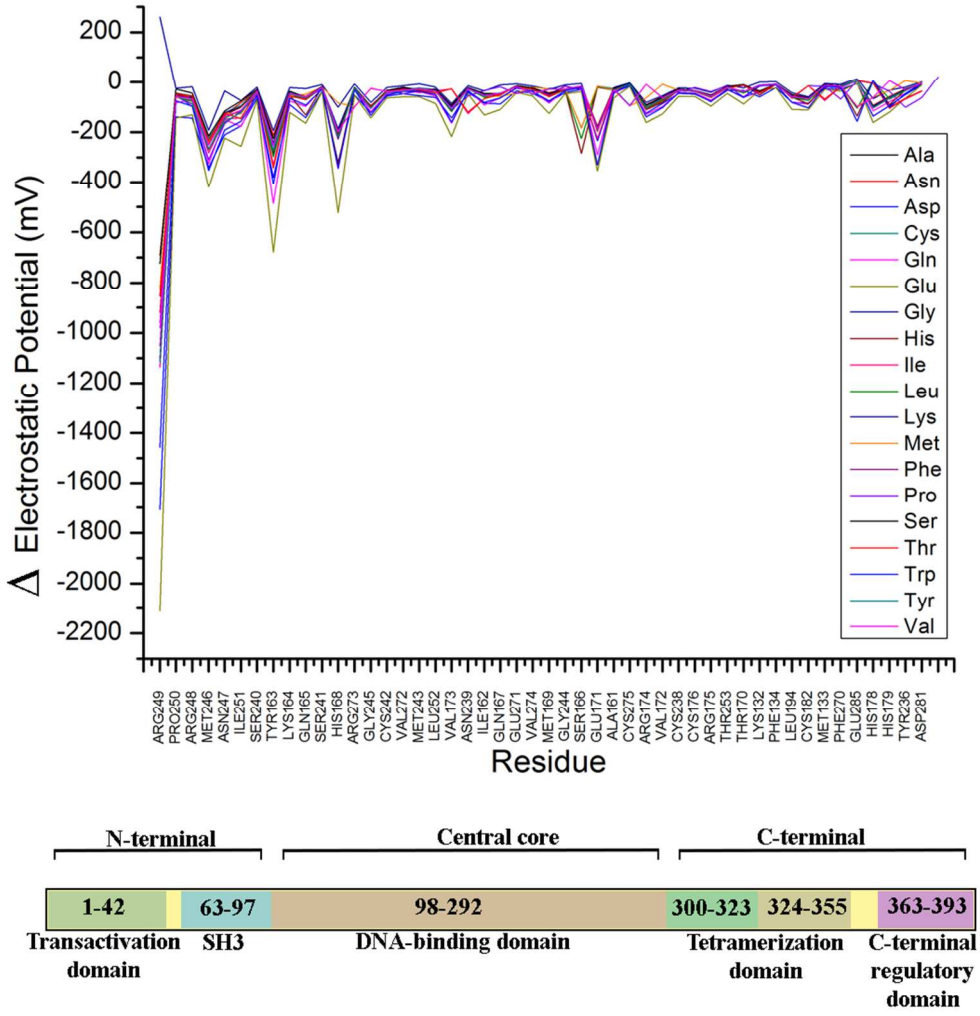

Figure 4: Difference in the calculated electrostatic potential between "wild type" p53 and 19 different mutants of residue R249. The change in the profile is shown for every residue located within 15Å of the mutated residue. These residues are ordered according to the distance to the R249 starting from the nearest one. The profiles indicate that the changes have an impact beyond the position of the mutation altering residues found in distinct functional domains of the p53 protein, this is shown in the highest peaks. In the bottom of the figure is shown a schema of the domain and residues present in p53 [51].
Figure 4

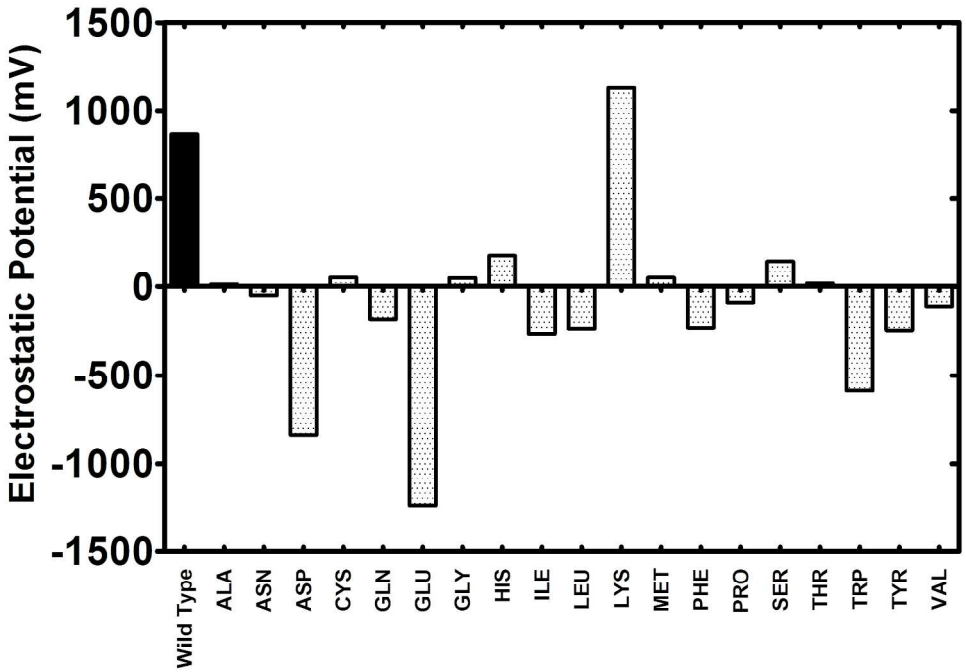

Figure 5: Calculated electrostatic potential for the residue Arg249 and the 19 variants generated with the scanning mode of the MutantElec system using the input PDB_ID:1TUP [39]. It is possible to observe the major change in the electrostatic potential for the residues with negative charge like Asp y Glu.
Figure 5
200x145mm (300 x 300 DPI)

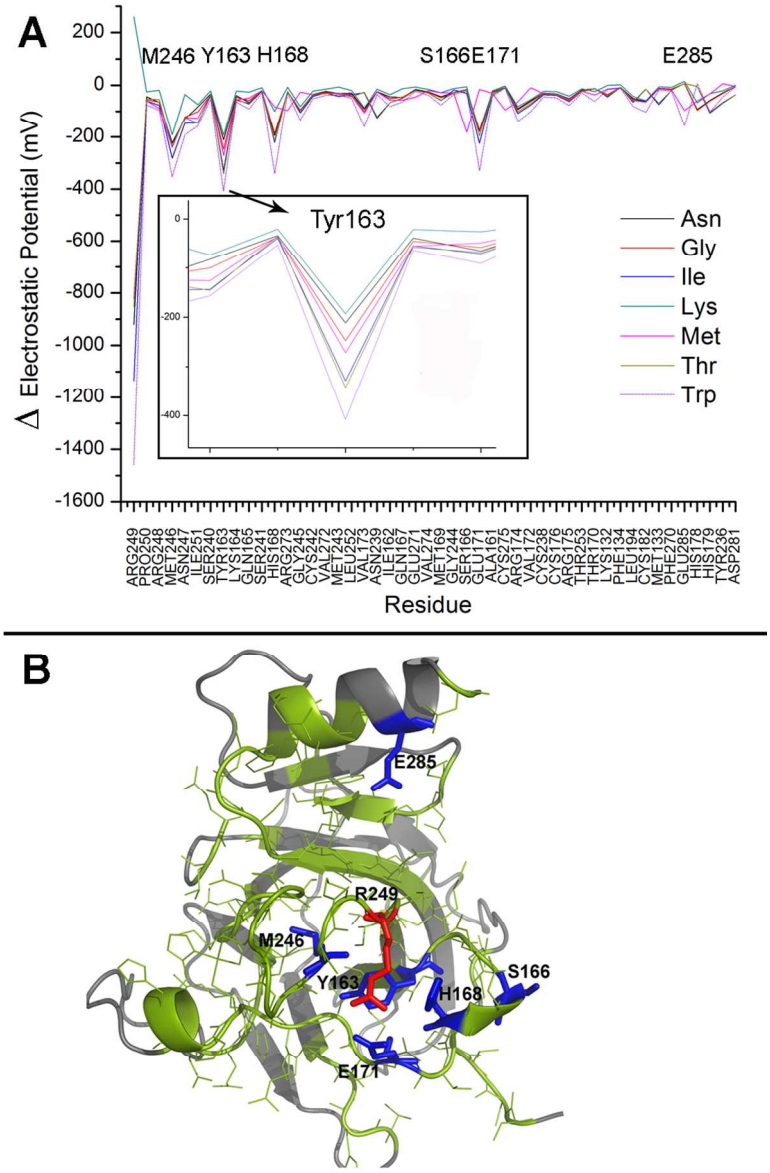

Figure 6: Analysis of the more significant change for the residue 249 for p53 protein. A: show amino acids that were most affected by the respective in silico mutations with respect to the wild type protein, this are the residues Tyr163, His168, Met246, Glu171, Ser166, Glu285 in decreasing order of magnitude. B: Spatial distribution of these residues, it is possible to observe some of these residues are located at less than 5Å to the mutated Arg249 except the residue Glu285 (>10Å).

Figure 6

Supplementary material

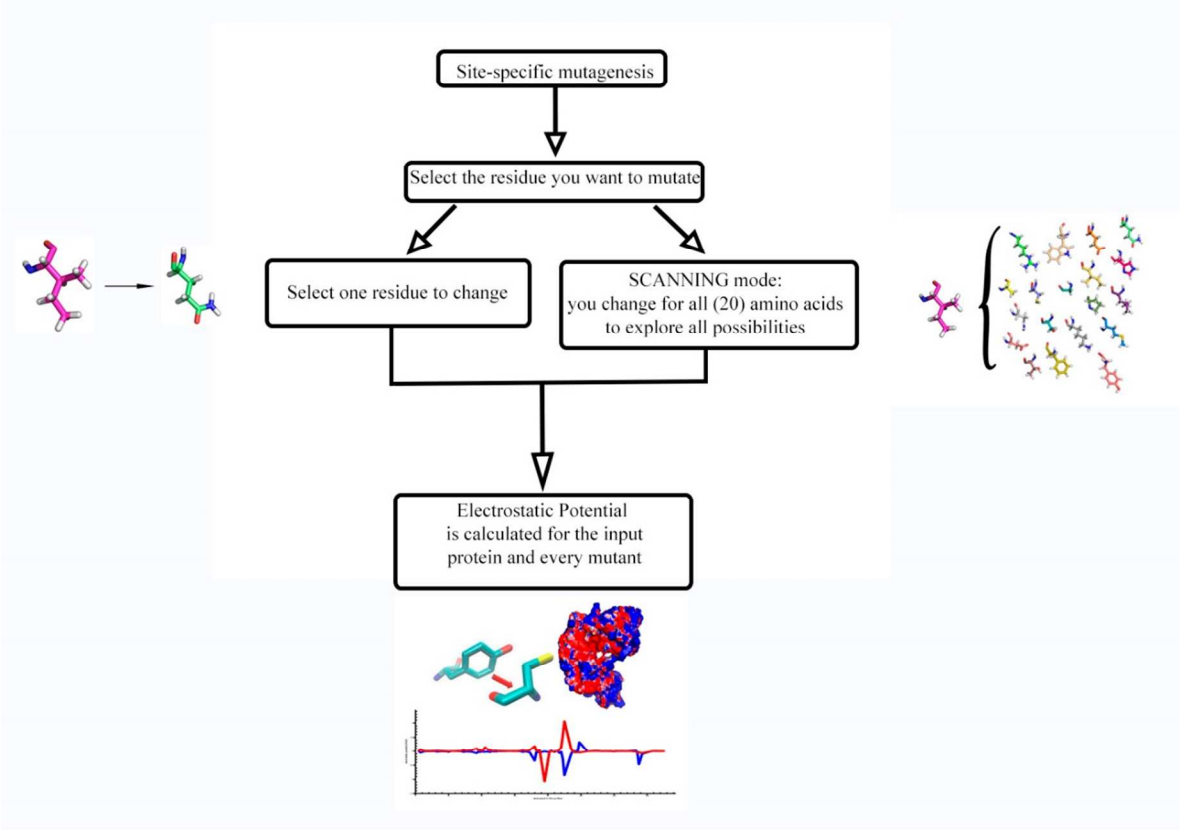

Figure 1S: "*Site-specific mutagenesis"* workflow. The user can choose between two options for generating the mutation: (i) Select one specific amino acid or (ii) select scanning mode. The latter repeats the calculation 19 time for every other of the 19 amino acid. The comparison between the electrostatic potential of the input protein and each mutant is calculated and the link to check the results are sent to the user by email.

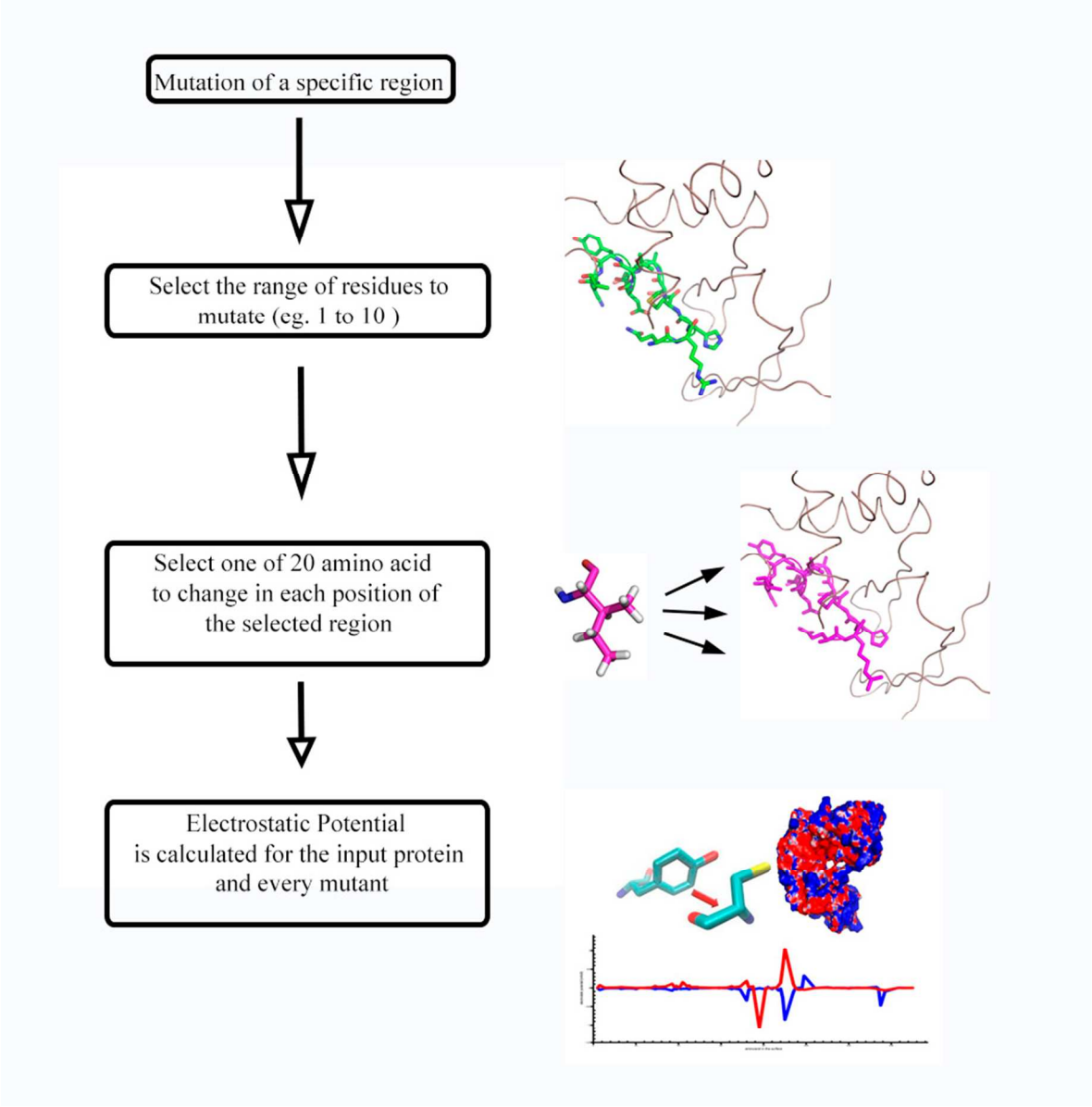

Figure 2S: "*Mutation of a specific region"* workflow. The user has to choose the zone in which to carry out the mutagenesis. The system permits the analysis of a region of 10 residues changing each of the specific amino acid selected for mutation. The comparison between the electrostatic potential of the input protein and each mutant is calculated and the results are sent to the user.

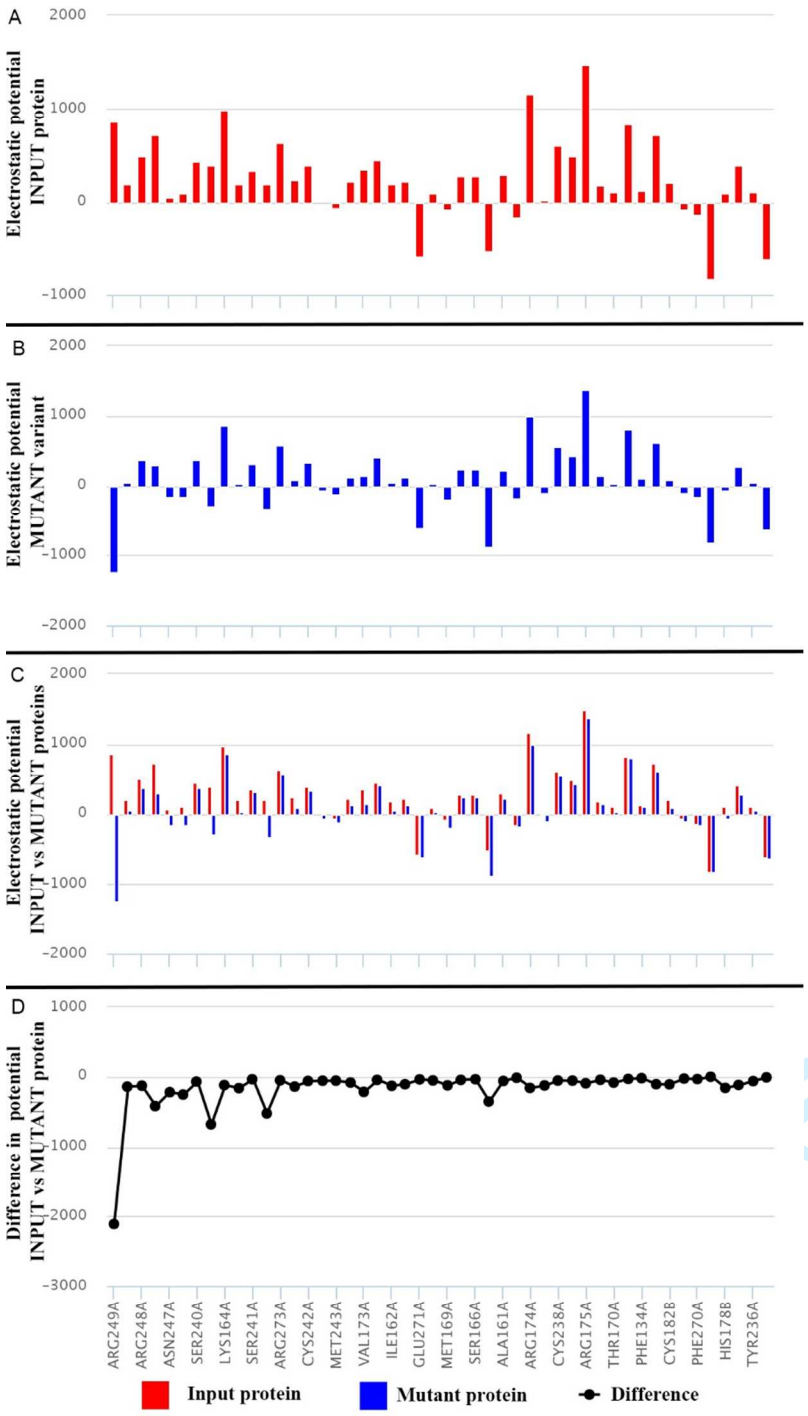

Figure 3S: Examples of graphs generated and provided on the MutantElec website. A: Electrostatic potential of the input protein. B: Electrostatic potential of the mutant protein. C: Electrostatic potential of the input and mutant protein. D: Difference in electrostatic potential.

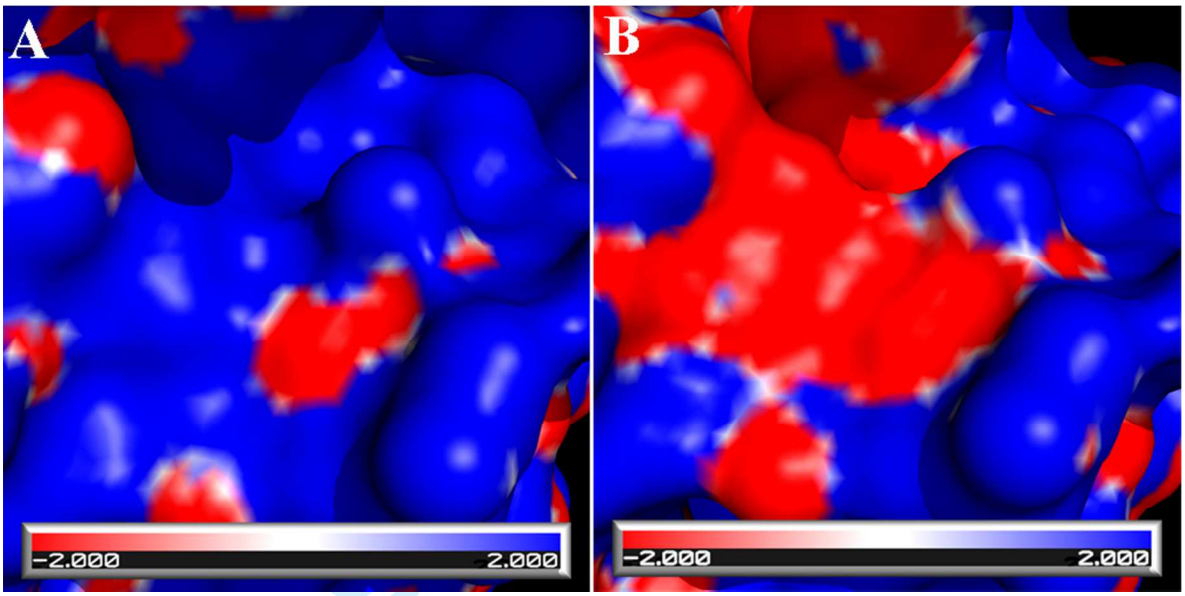

Figure 4S: Representation of the electrostatic potential maps for residue 249 for the p53 protein (PDB_ID:1TUP) [35]. Negative charges are shown in red and positive charges in blue. A: Representation of the electrostatic potential map for the region neighboring residue Arg249 for the wildtype protein. B: Representation of the electrostatic potential map for the region neighboring residue 249 for the mutant protein Arg249Glu. Is possible observe the region near to the residue 249 is negatively charged due to the change of Arg for Glu.

Table 1S. **Comparison of web applications for mutation analysis**. Comparison of existing web applications that study the effect of mutations in parameters such as binding affinity or binding free energy. The applications were evaluated in 5 aspects with regards to usability: "File submission" (whether it requires the user uploads the PDB file or if he can just input a code), "Job submission" (how friendly and intuitive is the interface to submit a job), "Mutation interface" (does it require the user to manually input residue number, names, or are they shown, so the user only has to click them?), "Results" (are they easy to understand?), "Response time" (how long does a job take to finish). Also, there is a column with other observations that didn't fit the above criteria, and a score from 0 to 10, where **0** correspond to a bad usability/bad response time and results, and **10** an intuitive and user-friendly interface with good usability and results easy to understand. In all cases, the platform where tested with the 1CSE pdb, and 3 mutations (if they were allowed).

Name	Application	File submission	Job submission	Mutation interface	Results	Other observations	Response time	Score
MutaBind http://www.ncbi.nlm.nih.gov/projects/mutabind	It evaluates the effects of variations and disease mutations on protein-protein interactions. It predicts if a mutation disrupts an interaction and calculates the change in binding affinity. The structure of a protein-protein complex is required for this method.	It provides the option to enter the PDB identifier or upload a custom PDB file	There's a drag-and-drop interface, which allows for an intuitive use in selecting partners of interaction. A representation of the PDB file is also shown, highlighting the different chains.	Doesn't require knowledge of the exact information in the PDB, because it is shown dynamically on the web page. It also provides the option of doing several mutations.	A table is shown, in which, by each mutation, the change of binding affinity is shown, whether is deleterious or not, and its confidence. The mutated PDB files can be downloaded.	The web page design is highly modern. It's usability is user-friendly, and intuitive.	29 estimated	8/10
BeatMusic http://babylone.ulb.ac.be/beatmusic/	Prediction of binding affinity changes upon mutations.	It provides the option to enter the PDB identifier or upload a custom PDB file	A list with all chains of the PDB file is shown, and the user must select the first and the second partner of the protein-protein interaction.	Doesn't require knowledge of the exact information in the PDB, because it is shown dynamically on the web page. However, only up to 10 mutations can be performed.	A table is shown, in which, by each mutation, the change of binding affinity is shown. If it increases or decreases the binding affinity, the result is highlighted.	The web page design is highly influenced by music. The menu is shown in musical terms, such as "Play", "Listen" or "Learn". This may distract an user, and it also makes the platform hard to take seriously.	~1 minute	9/10
SAAMBE http://compbio.clemson.edu/saambe_webserver/	Predicting the effect of single amino acid substitution on the binding free energy of protein complexes.	It doesn't allow the basic option of submitting a PDB file with just its identifier. The PDB file must be uploaded.	There's no validation in the selection of the partners of the protein complexes. They should be known *a priori*. This could present a difficulty in biologists that don't know much about a	The exact position, chain, and aminoacid to be mutated must be known and manually entered, which may lead to errors, as is isn't validated. This is	The original and the mutated PDB files can be downloaded. The calculation results include the change in several energy parameteres. This is delivered in a plain text file.	Only one mutation can be done.	~15 minutes	4/10

			PDB file. Since the partners can't be selected on the web page, if one inputs the wrong code, the job will end with no results. Even more, the error file just says "Wild type", which doesn't help at all.	also cumbersome for people with less computational knowledge, as it may impede the correct usage of the platform. Also, "Valine" can't be selected as a residue (nor original or as a mutation).				
SAAFEC http://compbio.clemson.edu/SAAFEC/	Calculating folding free energy changes in proteins caused by missense mutations.	It doesn't allow the basic option of submitting a PDB file with just its identifier. The PDB file must be uploaded.	There's only one interface to enter the mutations, and then the job can be submitted.	The exact position, chain, and aminoacid to be mutated must be known and manually entered, which may lead to errors, as is isn't validated. This is also cumbersome for people with less computational knowledge, as it may impede the correct usage of the platform.	A table with the change of several energy parameters is shown. Also, the original and the mutated PDB files can be downloaded. The calculation can also be downloaded as a plain text file.	Only one mutation can be done.	~15 minutes	5/10
Delphi http://compbio.clemson.edu/sapp/delphi_webserver/	Online Poisson-Boltzmann solver for calculating electrostatic energies and potential in biological macromolecules.	It doesn't allow the basic option of submitting a PDB file with just its identifier. The PDB file must be uploaded.	The interface guides the user through several pages, in which several parameters of the electrostatic potential calculations can be specified. In this regard, Delphi is highly customizable. However, this may deter users that just want to see a potential map, as they'll see the web page as something complex.	Mutations can't be performed. The platform is only directed at the calculation of electrostatic potentials.	The results are shown in a web page, with a Jmol applet, in which the user may watch the electrostatic potential map. The results are also available for download.	The results page uses Jmol. This may present difficulties for the users, since that requires the Java Runtime Environment. There're several security issues with Java, and in some computers is not allowed to use it, making it impossible for the user to see the results in a quick an efficient way.	~20 minutes	4/10
PoPMuSiC https://soft.dezyme.com/query/create/pop	Predicts the change in folding free energy upon mutation.	It provides the option to enter the PDB identifier or upload a custom PDB file	There's only one interface to enter the mutations, and then the job can be submitted.	The exact position, chain, and aminoacid to be mutated must be known and manually entered, which may lead to errors. Even though it is validated, it doesn't offer any	The results are delivered in a plain text file, and also in a very detailed way, in a colored and user-friendly page.	It requires registration. Also, the user's guide is outdated.	<1 minute	7/10

				help to solve the issue. This is also cumbersome for people with less computational knowledge, as it may impede the correct usage of the platform.				

Table 2S: Relationship between amino acid position in p53 and the number of mutants available in p53 (information obtained from the IARC TP53 database, [45]). In highlight are shown the amino acids that undergo significant changes of electrostatic potential in the mutation MutantElecR249X as predicted by MutantElec.

Amino acid position	Number of mutants	Amino acid position	Number of mutants	Amino acid position	Number of mutants
246	9	156	4	42	1
281	9	159	4	46	1
113	8	175	4	49	1
242	8	178	4	57	1
245	8	194	4	62	1
248	8	195	4	75	1
249	8	208	4	89	1
279	8	213	4	92	1
286	8	214	4	108	1
132	7	236	4	116	1
134	7	237	4	118	1
152	7	243	4	119	1
238	7	247	4	121	1
239	7	250	4	123	1
241	7	254	4	124	1
244	7	262	4	144	1
270	7	266	4	146	1
273	7	267	4	154	1
275	7	285	4	161	1
120	6	305	4	164	1
127	6	98	3	171	1
130	6	111	3	174	1
135	6	133	3	181	1
151	6	145	3	196	1
157	6	155	3	202	1
158	6	168	3	209	1
176	6	172	3	211	1
193	6	177	3	212	1
205	6	179	3	217	1
215	6	276	3	223	1
219	6	282	3	231	1
258	6	297	3	233	1
259	6	306	3	235	1
272	6	330	3	252	1
278	6	332	3	256	1
280	6	341	3	263	1
109	5	27	2	264	1
110	5	34	2	271	1
126	5	97	2	284	1

143	5		107	2		295	1
162	5		117	2		296	1
163	5		131	2		300	1
173	5		138	2		308	1
216	5		139	2		316	1
220	5		140	2		342	1
232	5		147	2		348	1
234	5		197	2		350	1
251	5		199	2		351	1
255	5		200	2		361	1
257	5		218	2		364	1
265	5		240	2		375	1
274	5		253	2			
277	5		260	2			
337	5		283	2			
105	4		293	2			
122	4		304	2			
125	4		344	2			
136	4		347	2			
141	4		36	1			

Graphical Table of Contents

The electrostatic potential plays a key role in many biological processes like determining the affinity of a ligand to a given protein target, and they are responsible for the catalytic activity of many enzymes. Understanding the effect that amino acid mutations will have on the electrostatic potential of a protein, will allow a thorough understanding of which residues are the most important in a protein this is essential for the drug design.