SOFTWARE TOOL ARTICLE

ABS–Scan: *In silico* alanine scanning mutagenesis for binding site residues in protein–ligand complex

[version 1; peer review: 1 approved with reservations, 1 not approved]

Praveen Anand[1], Deepesh Nagarajan[1], Sumanta Mukherjee[2], Nagasuma Chandra[1]

[1]Department of Biochemistry, Indian Institute of Science, Bangalore, Karnataka, 560012, India
[2]IISc Mathematics Initiative, Indian Institute of Science, Bangalore, Karnataka, 560012, India

v1 **First published:** 09 Sep 2014, **3**:214
https://doi.org/10.12688/f1000research.5165.1
Latest published: 01 Dec 2014, **3**:214
https://doi.org/10.12688/f1000research.5165.2

Abstract

Most physiological processes in living systems are fundamentally regulated by protein–ligand interactions. Understanding the process of ligand recognition by proteins is a vital activity in molecular biology and biochemistry. It is well known that the residues present at the binding site of the protein form pockets that provide a conducive environment for recognition of specific ligands. In many cases, the boundaries of these sites are not well defined. Here, we provide a web-server to systematically evaluate important residues in the binding site of the protein that contribute towards the ligand recognition through in silico alanine-scanning mutagenesis experiments. Each of the residues present at the binding site is computationally mutated to alanine. The ligand interaction energy is computed for each mutant and the corresponding ΔΔG values are computed by comparing it to the wild type protein, thus evaluating individual residue contributions towards ligand interaction. The server will thus provide clues to researchers about residues to obtain loss-of-function mutations and to understand drug resistant mutations. This web-tool can be freely accessed through the following address: http://proline.biochem.iisc.ernet.in/abscan/.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 01 Dec 2014	✓ view	✓ view	✓ view
version 1 09 Sep 2014	? view	✗ view	

1. **Sunando Datta**, Indian Institute of Science Education and Research Bhopal, Bhopal, India

2. **Bernard Offmann**, Université de Nantes, Nantes, France
Stéphane Téletchéa, Université de Nantes, Nantes, France

3. **Srikrishna Subramanian**, Université de Nantes, Nantes, France
Gurmeet Kaur, Université de Nantes, Nantes, France

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Nagasuma Chandra (nchandra@biochem.iisc.ernet.in)

Competing interests: No competing interests were disclosed.

Grant information: The authors(s) declare that no special grants were sanctioned for this project. PA was supported by Bristol-Myers Squibb fellowship while carrying out this work.

How to cite this article: Anand P, Nagarajan D, Mukherjee S and Chandra N. **ABS–Scan: *In silico* alanine scanning mutagenesis for binding site residues in protein–ligand complex [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2014, **3**:214 https://doi.org/10.12688/f1000research.5165.1

First published: 09 Sep 2014, **3**:214 https://doi.org/10.12688/f1000research.5165.1

Introduction

Currently (as of April 3, 2014)[1] there exist more than 72000 (as of April 3, 2014) experimentally determined protein structures complexed with small molecule ligands, providing an extensive data resource on protein binding sites. These binding sites vary in size ranging from six to thirty residues depending upon the size and the nature of the ligand. In most cases, the contribution of the individual amino acids towards the binding of a given ligand is not well understood. A well-established method of demonstrating the importance of a residue at the site is to create point mutants through site-directed mutagenesis[2]. Efforts towards characterization of entire functional site include tools such as alanine scanning mutagenesis (ASM)[3], where each residue is mutated to an alanine and its effect on the function is evaluated. ASM is indeed a well-used technique in experimental biology and has been successfully applied to the problems of protein folding and stability[4], protein-protein[5,6], and protein-ligand[7] interactions. The experimental success of this technique has resulted in further developments, including high-throughput and low-cost variants[8], greatly expanding its reach. Yet, given the time, cost and effort required for carrying out experimental biochemistry, a large majority of proteins are yet to be studied through this method.

Due to availability of a variety of structural bioinformatics tools, it is now feasible to carry out alanine scanning mutagenesis computationally[9]. Spurred by the successes and widespread adoption of the ASM technique, various computational resources now exist for *in-silico* alanine scanning. Prominent examples include Modeller[10] and the Rosetta software suite[11]. However, most packages are command-line oriented and are out of reach for researchers. Alanine scanning webservers with intuitive user interfaces such as Robetta webserver[12], the Rosetta Design web-server[13], ROSIE[14], FOLDX[15], BeATMuSiC[16], exist for the problems of protein folding, protein stability and protein-protein interactions. Although, there are workflows to evaluate ligand-binding energetics which require significant computational time and setup through free-energy calculations involving Molecular Mechanics/Generalized Born Surface Area method (MM-GBSA)[17,18], there is however, no intuitive web-tool available for analyzing alanine-scanning mutations of small-molecule binding site residues in real time. A common requirement for an experimental biochemist is to identify which amino acids to mutate in the protein to generate loss-of-function mutants. A web-tool to cater to that specific need will therefore be highly useful. The analysis will also provide deep insights into critical residues for interaction, residue pairs or sets that when mutated will abolish ligand binding and provide analytical insights for lead refinement in the process of drug discovery, as well understand drug resistance due to mutations.

We present a computational workflow and webserver, Alanine Binding Site-Scan (ABS-Scan), for automated alanine-scanning mutagenesis of protein-ligand interface residues. The workflow combines the libraries of widely used software packages including Modeller[10] for site-specific alanine mutagenesis and Autodock[19] for energetic evaluation of protein-ligand complexes.

Workflow

This workflow allows a user to submit a protein-ligand complex of their interest (Figure 1). The user is provided with an option of selecting a distance cut-off to define the binding site around a specific ligand for which, *in-silico* alanine scanning mutagenesis is carried out. Once the input parameters are obtained, the Modeller library is used to perform site-specific mutagenesis on all selected residues, coupled with steps of energy minimization. Each mutated structure, will then be scored by using Autodock 4.1 force field, to calculate the energetics of a protein-ligand complex. The essentiality of a residue can be determined by difference in interaction score of mutant and wild-type protein (ΔΔG value). These results are graphically presented to the user, along with a ranked list of residues in the given site that could be experimentally explored for site-directed mutagenesis. A Jmol applet displays protein-ligand interactions with residues colored according to the computed extents of contribution towards interaction, while a table simultaneously displays inter-molecular energy scores. We also provide a help-section explaining the results along with selected examples.

Validation

Mainly two types of validation were carried out, first to find a correlation with experimentally determined binding affinities followed by sensitivity evaluation of the predicted ABScan ΔΔG scores. The first exercise involved systematically mining the available experimental information related to alanine-scanning mutagenesis of binding site residues. A methodical search was carried out to mine all the experimental results available in literature on alanine-scanning mutagenesis of residues at the binding site. Advanced search option in PDB was used for this purpose. All the PUBMED extracts were scanned for the term - “alanine scanning”. The results obtained were further filtered to contain only X-Ray experimental data, and abscence of any DNA, RNA or DNA/RNA hybrid in the PDB entity. The results were further restricted to only those entries that had ligands bound to them, and we expected that this would reduce the hits that contain alanine-scanning mutations for evaluating protein-protein interfaces. The above search criteria mentioned yielded 126 structure hits with 56 citations. The list of entries obtained, was further pruned to remove biologically irrelevant ligands, metal ions and modified residues. The list of 79 entries that we finally obtained can be accessed at http://proline.biochem.iisc.ernet.in/abscan/validation. Each of the above experiments involving alanine-scanning mutagenesis reports different mutant evaluation scores. The measures reported to test the fitness of the mutants include various attributes such as K_d, K_a, k_{cat}/K_M (for enzymes), specific substrate/product assays *etc*. These measures cannot be normalized to derive values having uniform units for direct comparison. We picked a few of the examples to see the correlation between experimentally reported mutant evaluation scores and the predicted ΔΔG values. One such example has been described here.

A study on testosterone binding site of rat 3-alpha-hydroxysteroid dehydrogenase (*PDBID: 1AFS*) by Heredia *et al.*[20] reports that binding site residue in direct contact with the ligand influences the rate determining step of the enzymatic reaction. Alanine scanning mutagenesis was performed on binding site residues of the hydroxysteroid dehydrogenase protein that could interact with the steroid ligand and K_d was experimentally determined for each mutant to prove this. The ABS-Scan analysis performed on this complex with both testosterone and progesterone also confirms this. A good correlation was observed between the reported K_d value and the corresponding ΔΔG score predicted by ABS-Scan (Figure 2). The details of the experimental values, predicted score and the web-server

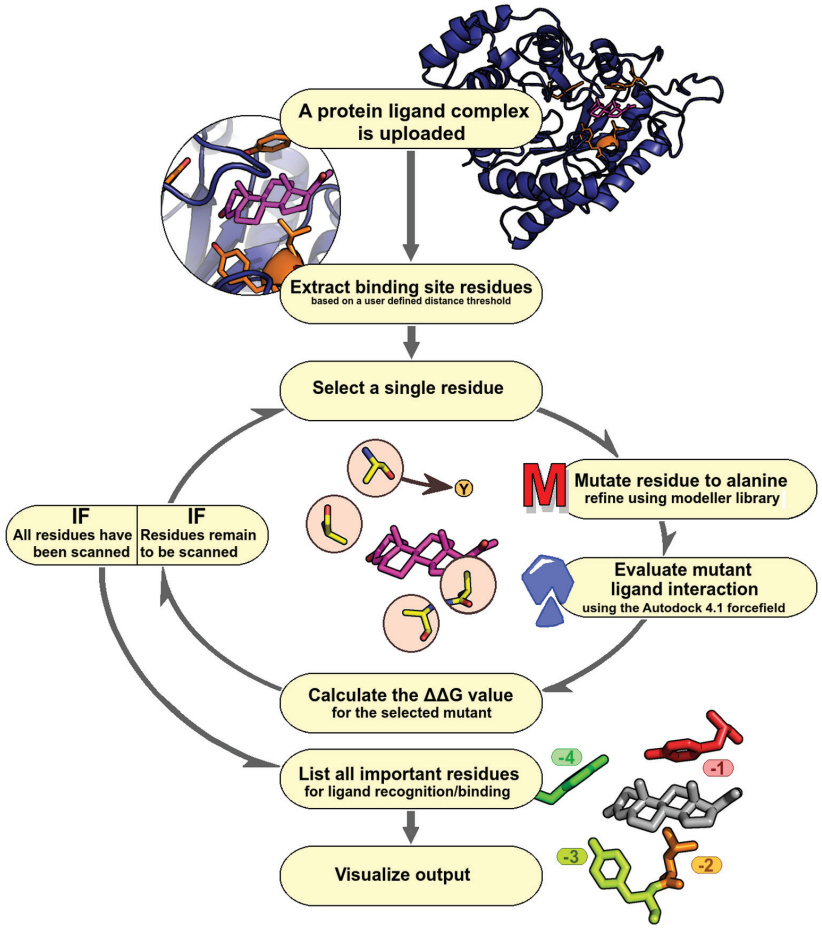

Figure 1. ABS-Scan workflow. Flowchart depicting various steps involved in ABS-Scan.

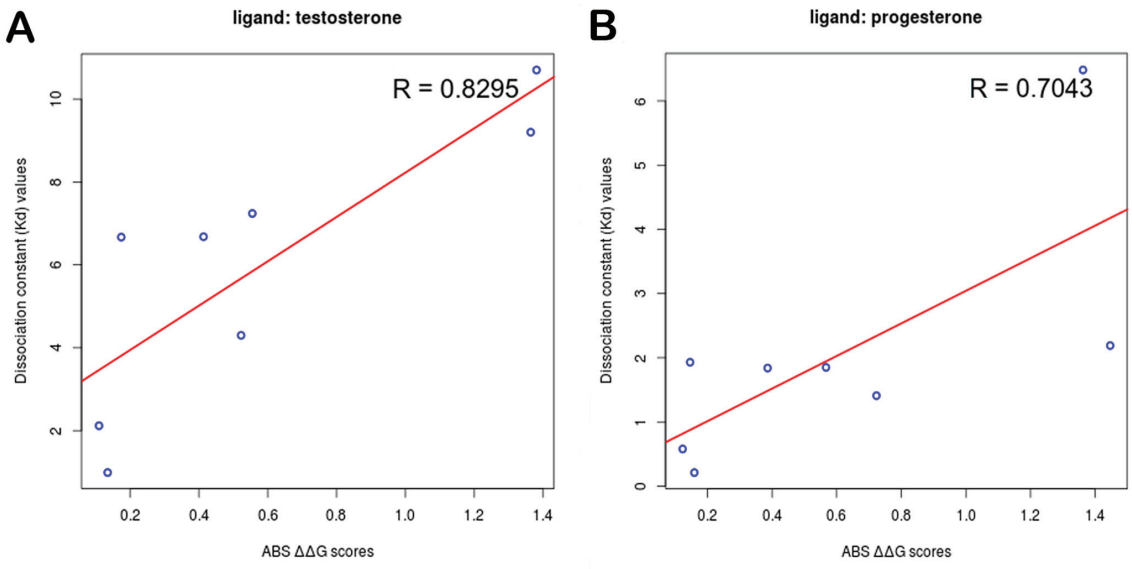

Figure 2. Experimental correlation. Good agreement is observed between experimental K_d values and predicted ΔΔG values determined for (**A**) testosterone & (**B**) progesterone binding site alanine scanning mutagenesis performed on rat 3-alpha-hydroxysteroid dehydrogenase.

output can be visualized along with other examples at http://proline.biochem.iisc.ernet.in/abscan/validation.

In order to determine the sensitivity of ABS-Scan, we compared predictions of essential residues through ABS-Scan in native complexes with corresponding decoy complexes. The complete dataset was obtained from the Community Structure-Activity Resource (CSAR - www.csardock.org/). Decoys in this dataset contain artificial docked complexes of protein with ligands having similar chemical properties to native ligands, but not known to interact with the protein. ABS-Scan is seen to effectively discriminate between the decoy and the native complexes (p-value ~0.004 calculated with Student's t-test) in ~67% of the cases ($\Delta\Delta G \geq 0.5$). This clearly indicates that residues important for ligand interaction can be identified through our approach (Figure 3). The details of validation protocol and results are accessible from the web-resource at *http://proline.biochem.iisc.ernet.in/abscan/validation*.

Implementation

The webserver was implemented using hypertext preprocessor (PHP). Autodock, Modeller and Pymol libraries have been used for modeling the mutation and evaluating the energetics. Integration of these back-end libraries for presentation as a functional and intuitive user interface is accomplished using Shell, Python, Java, HTML and PHP scripts. The web-server is platform independent and will run on any machine having internet access with browser installed. For the advanced users, a command-line interface in the form of a single python script can be accessed from github repository (*https://github.com/praveeniisc/ABS-Scan*). The script has been tested on Intel 2.83 GHz quad-core system running 32 bit linux OS(Ubuntu 12.04) with Modeller[10], MGL AutodockTools[19] & Pymol (http://pymol.org) installed. For the web-server d3.js library has been used for displaying the plots. Jmol Applet has been used to visualize the protein-ligand interaction.

Input

The input required for the server is the structure of protein-ligand complex in PDB format. Users can either provide the four-letter PDBID or upload the PDB structure file of the complex. An option is provided to define the cut-off distance and select the ligand to obtain binding site residues which would be mutated to alanine for evaluating the interaction energetics. A default distance cut-off of 4.5 Å is set to select all the residues within this distance from any atom of the ligand. In some the cases, metal ions[21] and water molecules are observed to play a crucial role in stabilizing the interactions[22]. Major problem involved in incorporating the ligand metal ion in ABS-Scan worflow is fixing the charge parameter as metal atoms can have different ionic states (Ex. Fe^{2+}, Fe^{3+} *etc.*) which is important for evaluating energetics. Enumerating all important structural water molecules involved in the ligand interaction is also highly dependent on the resolution of the crystal structure. Hence, an advanced option is provided to the user for uploading the PDBQT format of the ligand, to account for cases where the ligand contains unusual atom types, metal ions or uses bridge-water molecules for interaction.

Output

All the results produced by ABScan can be visualized interactively on the web-server. Jmol Applet is used to visualize the contribution of residues towards ligand interaction (Figure 4).

d3.js library has been utilized to plot the predicted $\Delta\Delta G$ values and subcomponents of the energetic scores reported by Autodock4 (Figure 5). An option is provided to download publication quality images in SVG/PDF/PNG formats. Twitter bootstrap java library is used for framework development on the webserver. An option is also provided to download the raw files containing individual mutants in PDB format, $\Delta\Delta G$ scores in the raw CSV format along with autodock energy scores.

Conclusions

ABS-Scan webserver can provide valuable insights on molecular recognition involving protein-ligand interactions. Experimentally determined protein-ligand structures can be studied to understand individual residue contributions towards ligand binding. Modeled complexes can also be submitted to infer the feasibility of the interaction. We believe that ABS-Scan would add one more dimension to the analysis of binding sites in proteins, comparison of various

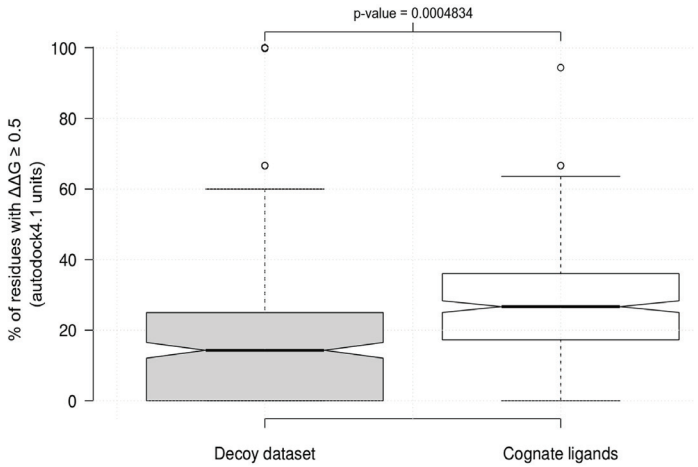

Figure 3. ABS-Scan sensitivity. Boxplot showing the difference in the % of the residues in the binding site of cognate and decoy dataset having a predicted $\Delta\Delta G$ score ≥0.5.

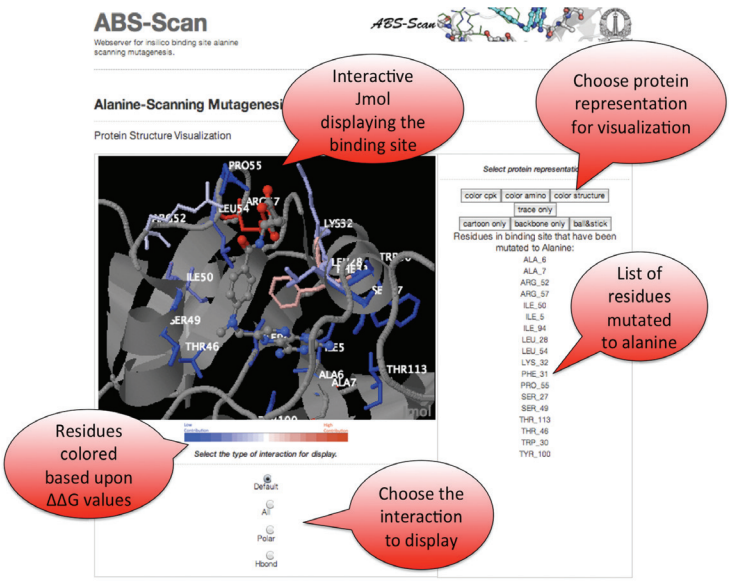

Figure 4. ABS-Scan interactive display. Snapshot explaining the Jmol applet output on the ABScan server. The individual residues are colored in red to blue gradient depending upon the contribution towards the ligand interaction as predicted by ABScan ΔΔG score. Options to visualize the different kinds of interaction - polar, hbonds *etc.* is also provided.

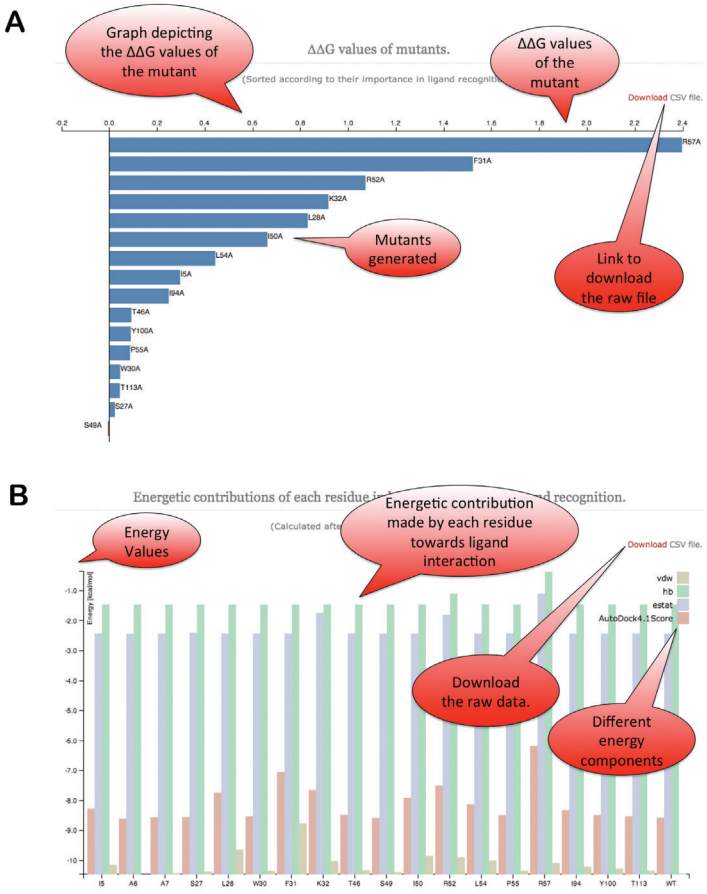

Figure 5. ABS-Scan energy plots. (**A**) ΔΔG values reported for each of the alanine mutation performed for the residues present at the binding site. The residues are ordered according to their contribution/ΔΔG values. (**B**) The different energy component of autodock interaction score plotted for each of the alanine mutant produced at the binding site.

ligand interactions and be of importance to researchers performing ASM studies.

Software availability

Software access

http://proline.biochem.iisc.ernet.in/abscan/

Latest source code

https://github.com/praveeniisc/ABS-Scan

Source code as at the time of publication

https://github.com/F1000Research/ABS-Scan/releases/tag/V1.0

Archived source code as at the time of publication

http://dx.doi.org/10.5281/zenodo.11423[24]

Software license

Author contributions

Conceived and designed the experiments: NSC. Performed the experiments: PA,DN,SM. Analyzed the data: PA,DN,SM,NSC. Wrote the paper: PA,DN,NSC. Website design and implementation: PA.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

We acknowledge all the members of the NSC lab for useful suggestions during the development of the web-server and visualization of the results.

References

1. Rose PW, Bi C, Bluhm WF, *et al.*: **The RCSB Protein Data Bank: new resources for research and education.** *Nucleic Acids Res.* 2013; **41**(Database issue): D475–82.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Morrison KL, Weiss GA: **Combinatorial alanine-scanning.** *Curr Opin Chem Biol.* 2001; **5**(3): 302–7.
PubMed Abstract | Publisher Full Text
3. Weiss GA, Watanabe CK, Zhong A, *et al.*: **Rapid mapping of protein functional epitopes by combinatorial alanine scanning.** *Proc Natl Acad Sci U S A.* 2000; **97**(16): 8950–4.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Williams AD, Shivaprasad S, Wetzel R: **Alanine scanning mutagenesis of Abeta(1-40) amyloid fibril stability.** *J Mol Biol.* 2006; **357**(4): 1283–94.
PubMed Abstract | Publisher Full Text
5. Ashkenazi A, Presta LG, Marsters SA, *et al.*: **Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis.** *Proc Natl Acad Sci U S A.* 1990; **87**(18): 7150–4.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Kristensen C, Kjeldsen T, Wiberg FC, *et al.*: **Alanine scanning mutagenesis of insulin.** *J Biol Chem.* 1997; **272**(20): 12978–83.
PubMed Abstract | Publisher Full Text
7. Tang WJ, Stanzel M, Gilman AG: **Truncation and alanine-scanning mutants of type I adenylyl cyclase.** *Biochemistry.* 1995; **34**(44): 14563–72.
PubMed Abstract | Publisher Full Text
8. Jain PC, Varadarajan R: **A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library.** *Anal Biochem.* 2014; **449**: 90–8.
PubMed Abstract | Publisher Full Text
9. Bromberg Y, Rost B: **Comprehensive *in silico* mutagenesis highlights functionally important residues in proteins.** *Bioinformatics.* 2008; **24**(16): i207–12.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Eswar N, Eramian D, Webb B, *et al.*: **Protein structure modeling with MODELLER.** *Methods Mol Biol.* 2008; **426**: 145–59.
PubMed Abstract | Publisher Full Text
11. Kaufmann KW, Lemmon GH, Deluca SL, *et al.*: **Practically useful: what the Rosetta protein modeling suite can do for you.** *Biochemistry.* 2010; **49**(14): 2987–98.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Kim DE, Chivian D, Baker D: **Protein structure prediction and analysis using the Robetta server.** *Nucleic Acids Res.* 2004; **32**(Web Server issue): W526–31.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Liu Y, Kuhlman B: **RosettaDesign server for protein design.** *Nucleic Acids Res.* 2006; **34**(Web Server issue): W235–8.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Lyskov S, Chou FC, Conchúir SÓ, *et al.*: **Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE).** *PLoS One.* 2013; **8**(5): e63906.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Schymkowitz J, Borg J, Stricher F, *et al.*: **The FoldX web server: an online force field.** *Nucleic Acids Res.* 2005; **33**(Web Server issue): W382–8.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Dehouck Y, Kwasigroch JM, Rooman M, *et al.*: **BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations.** *Nucleic Acids Res.* 2013; **41**(Web Server issue): W333–9.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Homeyer N, Gohlke H: **FEW: a workflow tool for free energy calculations of ligand binding.** *J Comput Chem.* 2013; **34**(11): 965–73.
PubMed Abstract | Publisher Full Text
18. Greenidge PA, Kramer C, Mozziconacci JC, *et al.*: **MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement.** *J Chem Inf Model.* 2013; **53**(1): 201–9.
PubMed Abstract | Publisher Full Text
19. Trott O, Olson AJ: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.** *J Comput Chem.* 2010; **31**(2): 455–61.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Bennett MJ, Albert RH, Jez JM, *et al.*: **Steroid recognition and regulation of hormone action: crystal structure of testosterone and NADP+ bound to 3 alpha-hydroxysteroid/dihydrodiol dehydrogenase.** *Structure.* 1997; **5**(6): 799–812.
PubMed Abstract | Publisher Full Text
21. Andreini C, Bertini I, Cavallaro G, *et al.*: **Structural analysis of metal sites in proteins: non-heme iron sites as a case study.** *J Mol Biol.* 2009; **388**(2): 356–80.
PubMed Abstract | Publisher Full Text
22. Mobley DL, Dill KA: **Binding of small-molecule ligands to proteins: "what you see" is not always "what you get".** *Structure.* 2009; **17**(4): 489–98.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Krüger DM, Gohlke H: **DrugScorePPI webserver: fast and accurate *in silico* alanine scanning for scoring protein-protein interactions.** *Nucleic Acids Res.* 2010; **38**(Web Server issue): W480–6.
PubMed Abstract | Publisher Full Text | Free Full Text
24. Anand P, Nagarajan D, Mukherjee S, *et al.*: **ABS-Scan.** *Zenodo.* 2014.
Data Source

Open Peer Review

Current Peer Review Status: ? ✗

Version 1

Reviewer Report 17 October 2014

https://doi.org/10.5256/f1000research.5509.r6276

✗ **Bernard Offmann**
Unité Fonctionnalité et Ingénierie des Protéines, Université de Nantes, Nantes, France
Stéphane Téletchéa
Unité Fonctionnalité et Ingénierie des Protéines, Université de Nantes, Nantes, France

This paper reports an attempt to develop an original tool that simulates alanine scanning mutagenesis to probe residues involved in the process of ligand recognition in proteins.

More precisely, the work describes the development of a work flow that implements known methodologies for homology modeling of alanine single-point mutants of a protein and for molecular docking. Even though, this can be viewed as a methodological paper.

We have some serious concerns regarding this work.

1. The authors claim that they performed a "validation" of their tool on a dataset that comprises "79 entries" carefully selected from PDB (also cf point 2 below). Their evaluation is based on finding a correlation between docking scores with experimentally determined binding affinities. In their paper, the authors provide evidence of this validation by providing results of "Experimental correlation" for only one example (Figure 2) which relates to binding of rat 3-alpha-hydroxysteroid dehydrogenase (PDB: 1AFS) to testosterone and progesterone. Since they must have it, clearly, the authors should provide their evaluation of this correlation on all "79 entries". I would expect at least that they provide a new Figure 2 that comprises all data points coming from these "79 entries" to sustain their claim and help readers to evaluate the global performance of their tool. They attempted to provide few additional results on their website (http://proline.biochem.iisc.ernet.in/abscan/validation). It is more confusing because the results provided for the vitamin D receptor (PDB: 1IE9) is not about binding affinities but "translational activity". I'm here suggesting that detailed data for all mutations taken from all "79" entries are provided to the community in the form of a table or downloadable flat or excel-type file.

2. The amount of independent PDB entries in their dataset is not 79. In fact, in some of PDB

entries, multiple ligands were observed. Surprisingly, they consider these as separate entries. So their data is redundant with respect to the proteins.

3. When generating homology models for protein variants, even if these are single point mutants, assessment of the quality of the models is a critical step. Selecting best models may not be that trivial. The authors need to clarify how they implement in their work flow the assessment of the quality of the models and consequently, what criteria they used for selecting the best models (and how many of them) that will be subjected to molecular docking.

4. Regarding the alanine scanning procedure, there are issues regarding the treatment of alanine and proline. They should both be discarded from the alanine scanning protocol: alanine is already present in the structure while proline is not suitable for mutations because of the major protein backbone rearrangements that should be performed to properly mutate it.

5. For such a tool, it is at stake to evaluate its performance using different homology modeling and molecular docking methods. The rational behind the choice of Modeler over other methods like Rosetta is not indicated. Likewise, the reason why Autodock and not Dock etc or even Autodock Vina is not explained.

6. The efficiency of molecular docking using AutoDock is also dependent on the docking protocol used. In such an automated "screen", care should be taken about the preparation of the receptor, the ligand and the grid. For example, are the ligands kept flexible ? In the manuscript, there are no indications about how the authors dealt with this central issue. The authors are encouraged to describe precisely and discuss their docking protocol.

7. According to the AutoDock 4.0 article, the median error range in energy estimation for any protein-ligand evaluation is 1.5-2.0 kcal/mol. In their study, the $\Delta\Delta G$ differences for ligand binding between mutant and native forms of the proteins are far below 2.0 kcal/mol. Thus, it is difficult to rank the mutants. Also, how the authors chose the 0.5 kcal/mol $\Delta\Delta G$ threshold is not clear. There is no discussion how this threshold compares with the intrinsic limits in precision of AutoDock.

8. The definition of ligand in the tool is problematic. In case of oligo or polysaccharides, the carbohydrate residues are erroneously considered separately. For example, in the 1J84 entry from PDB, the carbohydrate-binding module (CBM) is bound to cellotretraose, a 1,4-β-D-glucan composed of four ß-D-glucose residues linked by ß-1,4 osidic linkages. When this PDB entry is submitted to ABS-Scan, it erroneously splits the oligomer into smaller entities that correspond to the chemical IDs of its constituents (BGC 401, 402, 403, 404). This is a serious flaw in their software.

9. While it is common to see people to reuse available codes, the authors do not properly cite the source of their codes they posted on Github and used for providing a complete service to the community: at least 80% of the “alanine_scanning.py” code comes from either MODELLER examples (http://salilab.org/MODELLER/wiki/Mutate_model) or AutoDock code (http://mgltools.scripps.edu/api/AutoDockTools/AutoDockTools.Utilities24.compute_AutoDock41_score-pysrc.html).

Competing Interests: B. Offmann is founder and hold shares in the private company PEACCEL Inc (USA) and affiliated PEACCEL SAS (France). It's business purpose is to propose predictive services for protein engineering. S. Téletchéa declares no competing interest.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 18 Nov 2014

Nagasuma Chandra

We thank the reviewers for their time and effort. There were some useful suggestions, which we have incorporated but do not agree with all the points raised. A detailed point-by-point response is given below.

1. *The authors claim that they performed a "validation" of their tool on a dataset that comprises "79 entries" carefully selected from PDB (also cf point 2 below). Their evaluation is based on finding a correlation between docking scores with experimentally determined binding affinities. In their paper, the authors provide evidence of this validation by providing results of "Experimental correlation" for only one example (Figure 2)which relates to binding of rat 3-alpha-hydroxysteroid dehydrogenase (PDB: 1AFS) to testosterone and progesterone. Since they must have it, clearly, the authors should provide their evaluation of this correlation on all "79 entries". I would expect at least that they provide a new Figure 2 that comprises all data points coming from these "79 entries" to sustain their claim and help readers to evaluate the global performance of their tool. They attempted to provide few additional results on their website (http://proline.biochem.iisc.ernet.in/abscan/validation). It is more confusing because the results provided for the vitamin D receptor (PDB: 1IE9) is not about binding affinities but "translational activity". I'm here suggesting that detailed data for all mutations taken from all "79" entries are provided to the community in the form of a table or downloadable flat or excel-type file.*

   A suitable dataset for validation would be one that reports binding affinities for both wild-type and mutant proteins with same ligand, performed in a uniform experimental environment, for large number of proteins. Although such a dataset exists for protein-protein alanine scanning mutagenesis for eg., Rosetta alanine scanning), there are none reported for protein-ligand interactions.

   Since no such dataset was available to us, we systematically extracted PDB entries of ligand bound complexes and the corresponding binding sites in them that contained information about experimental alanine-scanning mutagenesis. However, the manner in which the effects of mutagenesis are reported in these differ significantly. While differences in ligand binding strengths (K_a or K_d values) are reported for some, changes in catalytic efficiencies are reported for some others. For some others, reporter assays are given which indicate capability of the downstream process more qualitatively. Hence it is difficult to perform a systematic comparison from these with

the Δ ΔG values calculated from our tool in this study. Nevertheless from this dataset, some examples were hand-picked, corresponding primary literature were read and known residue importances obtained, which were then compared with the predicted ones from our tool. In any case, ABS-Scan analysis has been successfully performed on 54 (the remaining 25 cases were not processed by default steps due to unusual atom types in proteins/ligands) complexes, which provide the extent of contribution to ligand binding of each residue in each site, in the form of a ranked list of residue-wide Δ ΔG values. All this information has been made available to the community, through our webserver - (http://proline.biochem.iisc.ernet.in/abscan/validation).

Besides this, given the lack of systematic reports of experimental data, validation can only be performed to understand the significance of the Δ ΔG scores calculated from our tool. For this, we have taken two large datasets (a) protein complexes with native ligands versus decoy ligands from and (b) list of well curated with precise binding site definitions for known protein-ligand complexes used for benchmarking docking algorithms. From both of these, Δ ΔG scores are in the range of 0.5 was significant.

a) A fresh dataset derived from PDB-Bind core dataset consisting of 195 protein-ligand complexes, which has been developed for the purposes of benchmarking docking algorithms (Kim *et al*., 2004, Huang *et al.*, 2008). Of the 195, 135 could be processed successfully for preparation of the protein-ligand complexes for analysis. (The others that could not be included, are likely to contain either unusual atom names or types or missing protein/ligand atoms or unusual convention for ligand atoms and hence could not be processed).

b) A dataset of 343 protein-ligand complexes, each with a native and a decoy ligand. 288 structures out of 343 could be successfully evaluated. (Here again the others were omitted due to difficulties in automatic protein/ligand preparation).

In the process, since ABSscan has been run for all these complexes, information about key contributing residues is generated for each of them. This has been made available through the webserver. Residue-wise contribution is obtained and presented in a ranked order for each complex, thus providing a ready resource of important residues for ligand binding.

The results of these can be accessed from the validation section on the webserver – http://proline.biochem.iisc.ernet.in/abscan/validation

2. *The amount of independent PDB entries in their dataset is not 79. In fact, in some of PDB entries, multiple ligands were observed. Surprisingly, they consider these as separate entries. So their data is redundant with respect to the proteins.*

These reflect independent binding sites (with bound ligands). As can be expected, some proteins have multiple sites with different ligands, making it necessary to consider them separately. Hence 79 sites are unique and come from 46 PDB entries. In the original manuscript, the dataset of 79 was never meant to reflect '**unique PDB entries'**. In any case we refer to them now as 'binding site entities' to reflect this more

clearly.

3. *When generating homology models for protein variants, even if these are single point mutants, assessment of the quality of the models is a critical step. Selecting best models may not be that trivial. The authors need to clarify how they implement in their work flow the assessment of the quality of the models and consequently, what criteria they used for selecting the best models (and how many of them) that will be subjected to molecular docking.*

   Model quality has been considered as part of the modelling pipeline itself. Given the scale of the study, it is practical to generate one model for each mutant, but care is taken to ensure that it is optimal and free of errors in terms of bad contacts or atomic clashes. The optimization protocol used consists of 200 iterations of conjugate gradient, followed by molecular dynamic simulation for 4fs and simulated annealing with 200 iterations at different temperatures (This is the default protocol suggested in Model_mutate.py of Modeller - http://salilab.org/modeller/wiki/Mutate%20model). The initial restraints for generation of the model is derived from the wild-type structure itself. Assumptions necessary for modelling point mutations introduced through alanine-scanning mutagenesis protocol at the binding sites are that (a) they are unlikely to change the overall structure of the protein drastically and (b) the ligand moiety roughly retains the same conformation in comparison with the wild-type complex to interact with the mutated structure.

   Since modelling protocols have been well established for a long time now, we did not see the need for adding this information explicitly in the original MS. In any case, based on the reviewers suggestion, this information has been added to the revised version. Normalized DOPE scores are reported for both the native and mutant structures. DOPE refers to 'Discrete Optimized Protein Energy' and is a statistical potential which checks for the feasibility of the observed interactions. Protein structures with lower DOPE scores (typically in negative range -1.5 to -2.5 for experimentally solved structures) can be considered to be of good quality (Shen and Sali., 2006).

4. *Regarding the alanine scanning procedure, there are issues regarding the treatment of alanine and proline. They should both be discarded from the alanine scanning protocol: alanine is already present in the structure while proline is not suitable for mutations because of the major protein backbone rearrangements that should be performed to properly mutate it.*

   This required addition of simple screens to filter out these residues from consideration for alanine scanning, which has been done. Changes have been made to both the source code and the web-tool now. Glycine mutations are also filtered out.

5. *For such a tool, it is at stake to evaluate its performance using different homology modeling and molecular docking methods. The rational behind the choice of Modeler over other methods like Rosetta is not indicated. Likewise, the reason why Autodock and not*

Dock etc or even Autodock Vina is not explained.

The goal of our study is not to develop a modelling algorithm or a new parameter for building models. The most widely used tool for homology modelling – Modeller, which we have currently included in the workflow, has about 1500 citations. Currently there are more than 50 tools for homology modeling -(http://en.wikipedia.org/wiki/List_of_protein_structure_prediction_software) and roughly the same number of tools for protein-ligand docking (http://en.wikipedia.org/wiki/Docking_%28molecular%29). The precise reason for choosing 'Modeller' or 'Autodock' is perhaps because of our own experience in using these tools along with availability of extensive documentation, tutorials and ease of implementation. Moreover, both these libraries had python bindings available and hence could be merged into a single script using python. In future, we plan to develop a pymol plugin for the same.
A simple bash script for processing the protein-ligand complex to determine the interaction energy using ROSETTA force fields has also been included in the github repository. This again, is only for the advanced users and we might incorporate it in the future versions of the pipeline.

6. *The efficiency of molecular docking using AutoDock is also dependent on the docking protocol used. In such an automated "screen", care should be taken about the preparation of the receptor, the ligand and the grid. For example, are the ligands kept flexible? In the manuscript, there are no indications about how the authors dealt with this central issue. The authors are encouraged to describe precisely and discuss their docking protocol.*

   We would like to clarify here that there is no docking performed in the whole exercise. We only score the complex in the given conformation using the force fields. By default, through prepare_receptor4.py and prepare_ligand4.py Gasteiger charges and polar hydrogens are added while evaluating the interaction energy. This has been mentioned in the manuscript:

   "Each mutated structure, will then be scored by using Autodock 4.1 force field, to calculate the energetics of a protein-ligand complex. The contribution from the residue is then determined by calculating the difference in interaction score of the mutant and the wild-type protein ($\Delta\Delta G$ value)."

7. *According to the AutoDock 4.0 article, the median error range in energy estimation for any protein-ligand evaluation is 1.5-2.0 kcal/mol. In their study, the* $\Delta\Delta G$ differences for ligand binding between mutant and native forms of the proteins are far below 2.0 kcal/mol. Thus, it is difficult to rank the mutants. Also, how the authors chose the 0.5 kcal/mol $\Delta\Delta G$ threshold is not clear. There is no discussion how this threshold compares with the intrinsic limits in precision of AutoDock.

   The median error range of the energy estimation reported in AutoDock 4.0 article is for the total ΔG score between the experimental and predicted values, whereas in this case it is for individual residue contributions. The distribution of the $\Delta\Delta G$ values obtained for the decoy and cognate ligands from the CSAR dataset (

http://www.csardock.org/) was used to define a cut-off of 0.5. This has also been validated on PDBbind core dataset (http://www.pdbbind-cn.org/). Figures 3A and 3B have been added along with explanations in the manuscript.

We believe that intrinsic limits on precision of Autodock scoring would not be a major concern as both the wild type and the mutant are evaluated using the same scoring scheme and the cut-off has been chosen on basis of native protein-ligand complexes in CSAR and PDBbind datasets.

8. *The definition of ligand in the tool is problematic. In case of oligo or polysaccharides, the carbohydrate residues are erroneously considered separately. For example, in the 1J84 entry from PDB, the carbohydrate-binding module (CBM) is bound to cellotretraose, a 1,4-ß-D-glucan composed of four ß-D-glucose residues linked by ß-1,4 osidic linkages. When this PDB entry is submitted to ABS-Scan, it erroneously splits the oligomer into smaller entities that correspond to the chemical IDs of its constituents (BGC 401, 402, 403, 404). This is a serious flaw in their software.*

   This is not really a 'problem' and is an established work-around to avoid long computation and hence long waiting time for the user. All this does is to split peptides or oligosaccharides into individual moieties (typically for a peptide, each amino acid is considered as a moiety and for an oligosaccharide, each monosaccharide is considered as a moiety), as per the convention currently followed by PDB. How can this be a 'serious flaw'? It does not, in any manner, influence the results. Many other tools for protein-ligand interaction analysis such as LPC (Ligand-protein contacts, Ligplot+, Ligplus *etc.)* also track ligands through such residue identifiers.

   However, an advanced option has now been added to provide the range of the ligand residue numbers to be considered as a single moiety during the entire protocol. For example, now a residue range 401-404 can be provided for 1J84 instead of a single residue number to consider the whole oligocomplex as single ligand. The script has also been accordingly modified in github.

9. *While it is common to see people to reuse available codes, the authors do not properly cite the source of their codes they posted on Github and used for providing a complete service to the community: at least 80% of the "alanine_scanning.py" code comes from either MODELLER examples (http://salilab.org/MODELLER/wiki/Mutate_model) or AutoDock code (http://mgltools.scripps.edu/api/AutoDockTools/AutoDockTools.Utilities24.compute_AutoDock41_score-pysrc.html).*

   We have indeed already cited all the tools used in the manuscript to which source codes are linked. In any case, these references are now highlighted in the source code also. Both Autodock and Modeller are released under the GNU public license, making their source code freely usable to all interested parties. Moreover, these are the primary source and codes are not extracted from any third-party tools. The purpose of putting it on Github is to be completely open about the details of the protocol and make our work fully accessible to anyone interested.

We would again like to remind the reviewer here that source-code is used only by an advanced user. The reviewer may be aware of the time and effort involved in producing a web-application interface that is embedded with visualization features. This has been done with the belief that it will save precious time for researchers who do not have the expertise or the interest in installation and handling command-line interfaces for such tools. We initially proposed this as a web-tool, but since that section is no longer available in *F1000Research*, we submitted it as a software tool.

Competing Interests: No competing interests were disclosed.

Reviewer Report 30 September 2014

https://doi.org/10.5256/f1000research.5509.r6262

?

Sunando Datta
Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, India

The manuscript 'ABS–Scan: *In silico* alanine scanning mutagenesis for binding site residues in protein–ligand complex' reports development of a web server for performing *in silico* alanine scanning mutations for studying protein-small molecule interactions. It further validates the tool by taking a list of already published Alanine scanning data along with the X-ray crystallographic structures of the relevant protein-ligand complexes. ABS-Scan provides a user-friendly web interface and will be very much useful for experimentalists to assess the outcome of mutations designed to study protein-ligand (small molecule) interactions. Overall the web tool is well explained in the manuscript.

My main concern is the method for energy calculations, authors used to predict the individual contribution upon mutation. It does not include waters and metals. As it is well established that many of the protein-ligand interactions are water mediated and therefore water plays very important role in determining the specificity as well as affinity. Authors could use energy function which includes waters and metals as well otherwise the current version could only be used for protein-ligand complexes in which water/metal atoms have been shown to play any role in stabilizing the ligand in the binding pocket.

Thus in my opinion, I think it could be indexed with inclusion of an updated energy function. Alternatively, its sole applicability for protein-ligand interaction without involvement of solvent molecules should be mentioned in the conclusion.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 18 Nov 2014

Nagasuma Chandra

We thank the reviewer for going through our manuscript and finding the work useful. These are indeed valid points and have been addressed in the revised version.
Bridge water molecules do play an important role in the protein-ligand interactions. One has to take into account the resolution of the protein structure to determine the confidence of the placed water molecules. Hence an advanced option is provided wherein these water molecules when present at the site can be considered to be a part of the corresponding ligand moiety. The user can upload his/her own pdbqt file for the ligand with the appropriate water molecules added to it. An example of protein lysine methyltransferases complexed with S-adenosyl methionine has been described in the manuscript. The corresponding pqbqt file of the ligand can be downloaded from the example section. The results of these can also be accessed from example section of the web-server.

Competing Interests: No competing interests were disclosed.