

CAGI slide pack guidelines

I agree to my slides being used according to the following terms: **Alexey Strokach**

- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2016. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk 
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.

CAGI5 - Frataxin Challenge

Alexey Strokach, Carles Corbi-Verge, Philip M. Kim

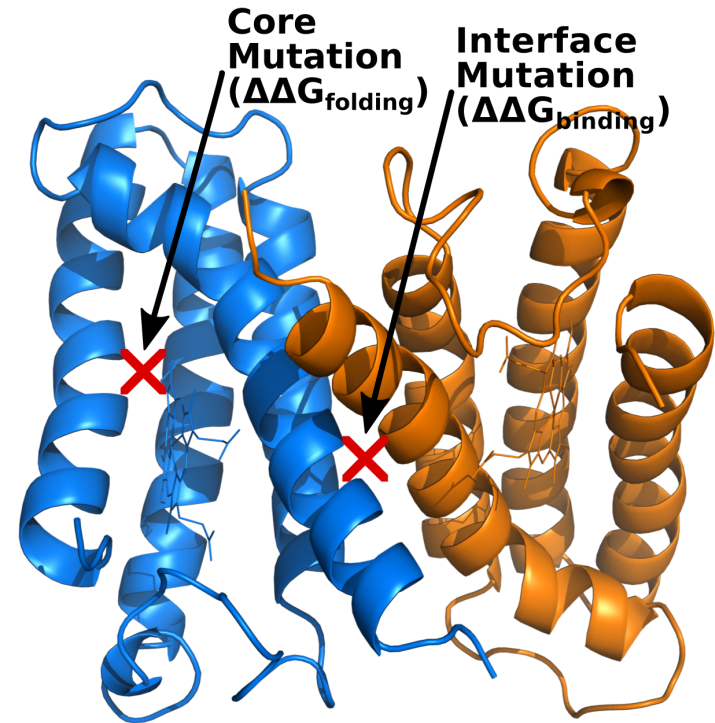
2018-07-06

Overview

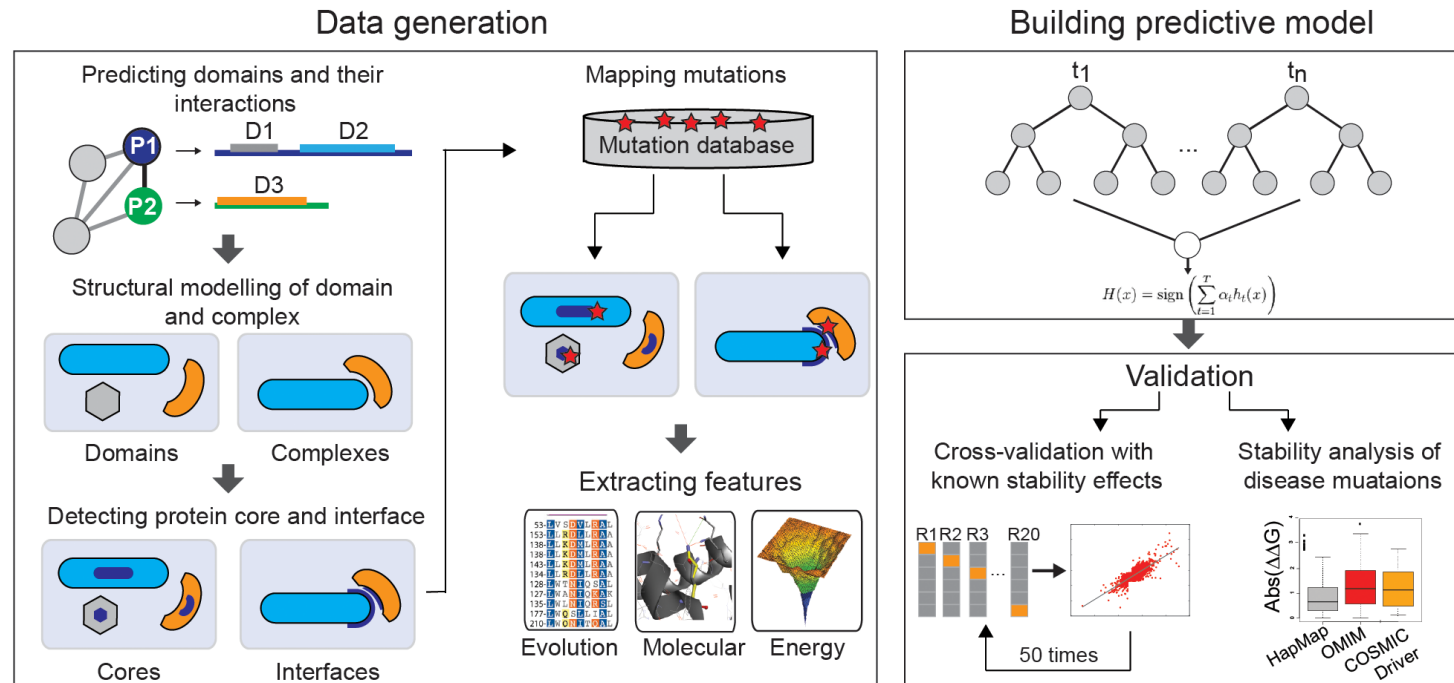
- Discuss the motivation for developing ELASPIC.
- Present the ELASPIC webserver.
- Discuss predictions made by ELASPIC and other methods for 8 mutations in Frataxin.

Introduction

- We need tools for improving protein stability and increasing the affinity of proteins to their targets.
 - Design proteins that are easier to crystallize and that show higher yield and less aggregation when expressed in heterologous systems.
 - Design biocatalysts which remain active in inhospitable conditions, including high temperatures and pressures, acidic or basic pH, different salt types and concentrations, and admixture of organic solvents
 - Optimize protein and peptide “biologics” to improve their shelf life and increase their affinity and specificity to desired targets.
- We need tools which provide a mechanistic explanation as to why a variant may be deleterious or benign.
- ELASPIC (<http://elaspic.kimlab.org>) uses sequential and structural information to predict the effect of mutations on protein-folding and protein-protein interaction on a genome-wide scale.



Methods



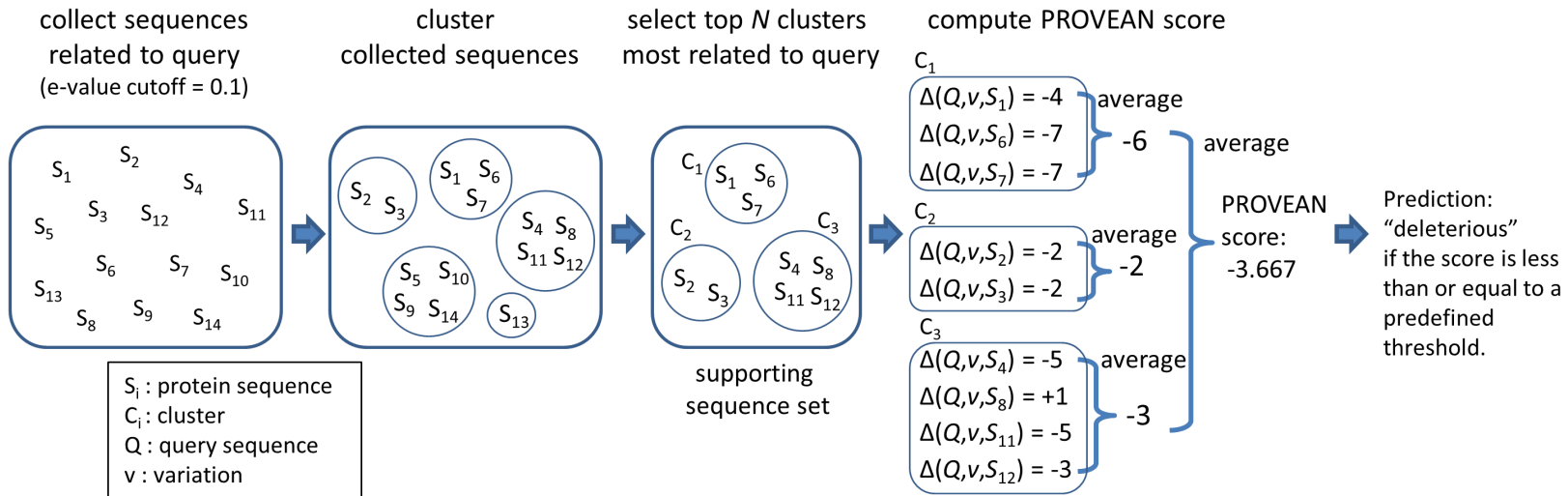
Berliner et. al. (2014) PLoS ONE 9(9): e107353.

FoldX energy function

$$\Delta G = w_1 \cdot \Delta G_{\text{vdw}} + w_2 \cdot \Delta G_{\text{solvH}} + w_3 \cdot \Delta G_{\text{solvP}} + w_4 \cdot \Delta G_{\text{hbond}} + w_5 \cdot \Delta G_{\text{wb}} + w_6 \cdot \Delta G_{\text{el}} + w_7 \cdot \Delta G_{\text{clash}} + w_8 \cdot T\Delta S_{\text{mc}} + w_9 \cdot T\Delta S_{\text{sc}} + w_{10} \cdot \Delta G_{\text{kon}}$$

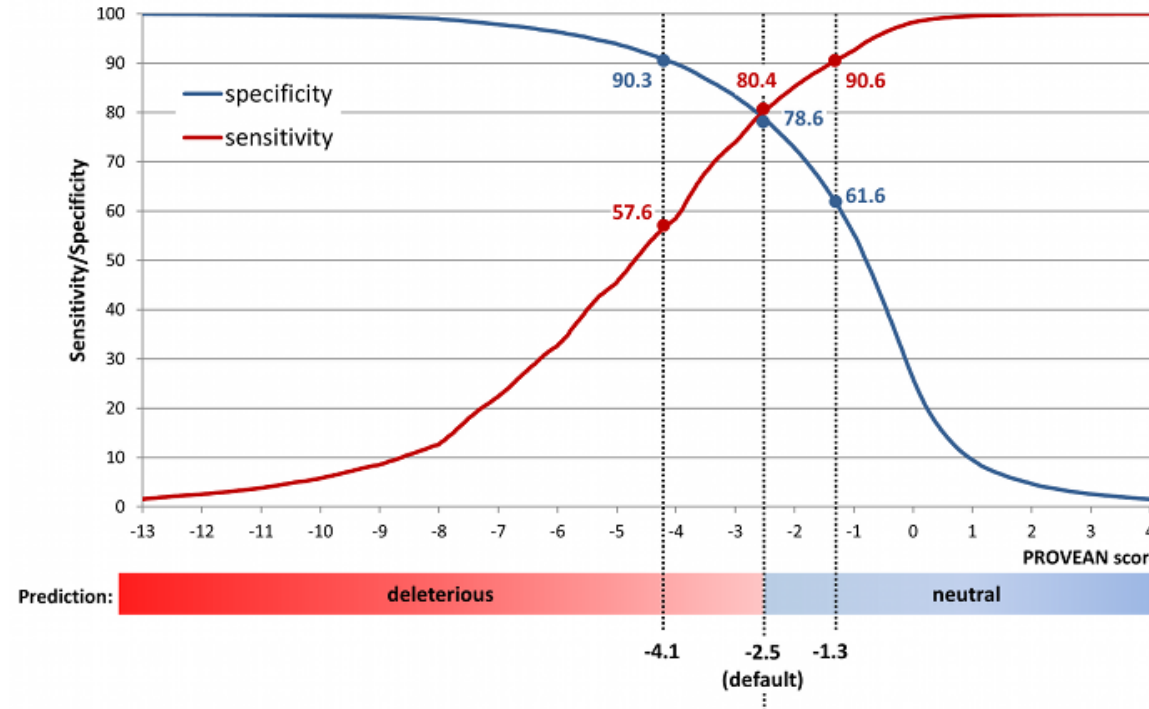
- $\{w_1 \dots w_{10}\}$ - tuned parameters.
- ΔG_{vdw} - experimental transfer energies from water to vapour.
- ΔG_{solvH} - hydrophobic desolvation score, scaled with the burial of the residue.
- ΔG_{solvP} - hydrophilic desolvation score, scaled with the burial of the residue.
- ΔG_{hbond} - hydrogen bonds, inferred through geometric constraints and empirical data.
- ΔG_{wb} - water molecules that have persistent interactions with the protein.
- ΔG_{el} - electrostatics, calculated using Coulomb's law.
- ΔG_{clash} - steric overlaps between atoms in the structure.
- $T\Delta S_{\text{mc}}$ - the entropic penalty for fixing the backbone in a given conformation.
- $T\Delta S_{\text{sc}}$ - the entropy cost of fixing a side chain in a particular conformation.
- ΔG_{kon} - electrostatic interactions between different chains.

Provean methodology



Source: <http://provean.jcvi.org/about.php> (<http://provean.jcvi.org/about.php>).

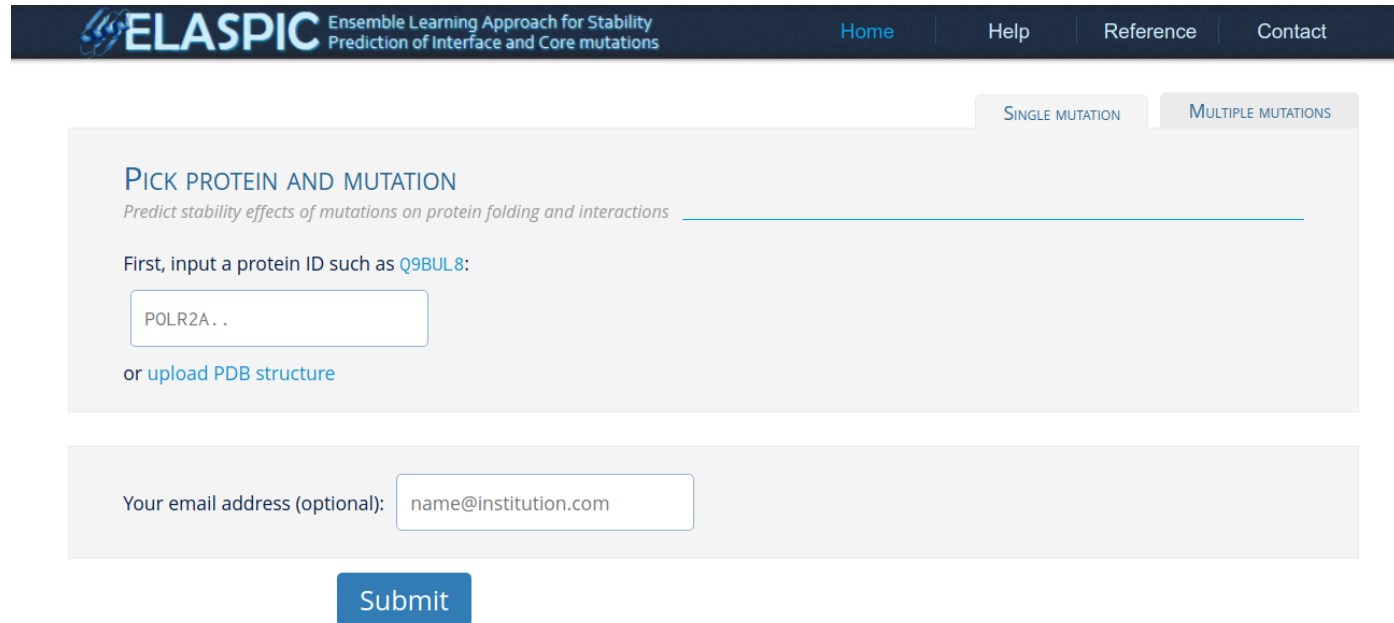
Provean score



Source: <http://provean.jcvi.org/about.php> (<http://provean.jcvi.org/about.php>).

ELASPIC webserver

<http://elaspic.kimlab.org/> (<http://elaspic.kimlab.org/>)



The screenshot shows the ELASPIC webserver interface. At the top is a dark blue header with the ELASPIC logo and the text "Ensemble Learning Approach for Stability Prediction of Interface and Core mutations". To the right of the logo are navigation links: "Home", "Help", "Reference", and "Contact". Below the header, there are two tabs: "SINGLE MUTATION" and "MULTIPLE MUTATIONS". The "SINGLE MUTATION" tab is active. The main content area is titled "PICK PROTEIN AND MUTATION" with the subtitle "Predict stability effects of mutations on protein folding and interactions". Below this, there is a text input field containing "POLR2A . .". Above the input field, there is a prompt: "First, input a protein ID such as Q9BUL8:". Below the input field, there is a link: "or upload PDB structure". At the bottom of the form, there is a text input field for "Your email address (optional):" containing "name@institution.com". Below the email field is a blue "Submit" button.

ELASPIC Ensemble Learning Approach for Stability Prediction of Interface and Core mutations

Home Help Reference Contact

SINGLE MUTATION MULTIPLE MUTATIONS

PICK PROTEIN AND MUTATION
Predict stability effects of mutations on protein folding and interactions

First, input a protein ID such as Q9BUL8:

POLR2A . .

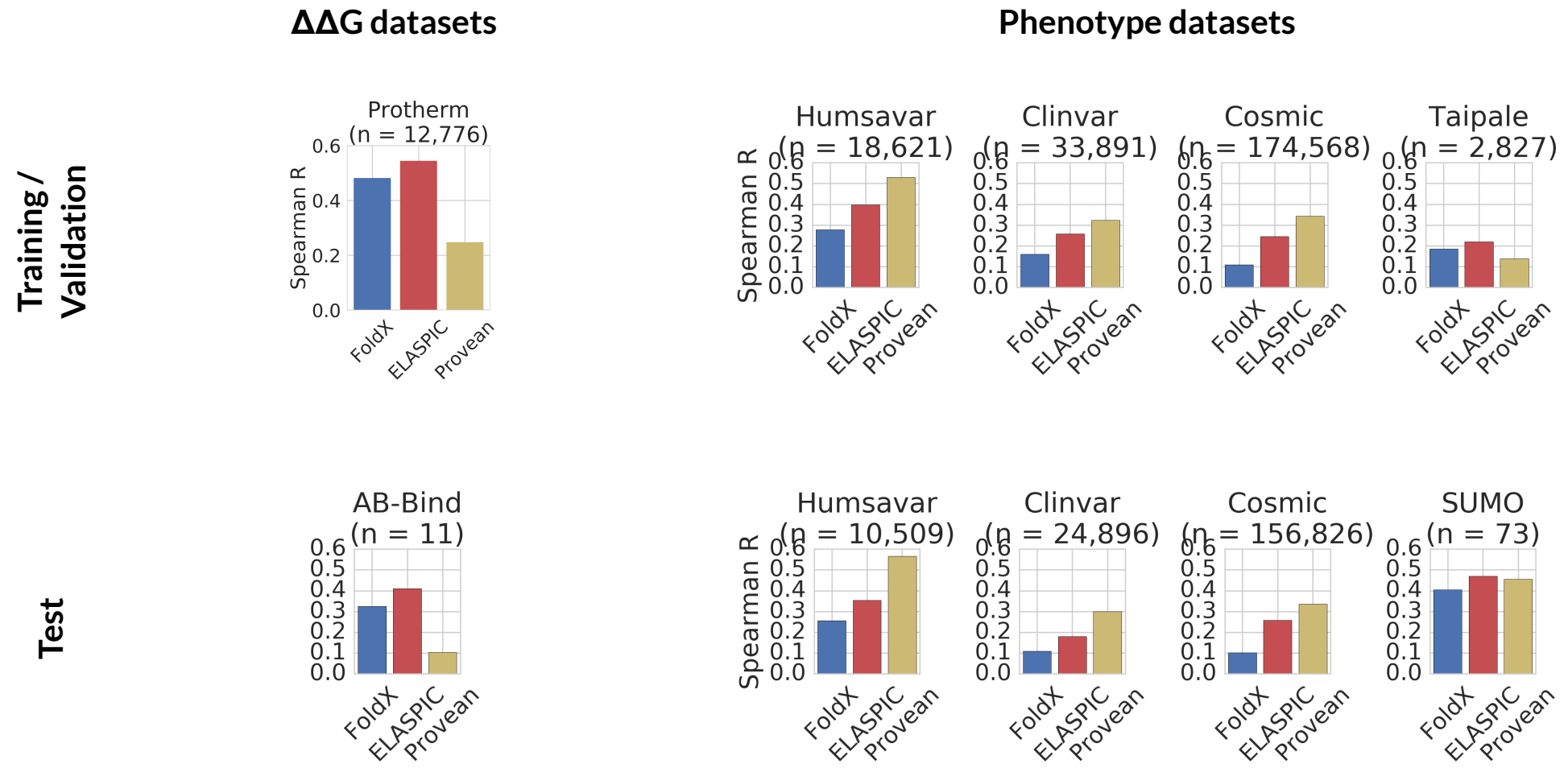
or [upload PDB structure](#)

Your email address (optional): name@institution.com

Submit

Witvliet[†], Strokach[†], et. al. (2016) Bioinformatics 32(10): 1589-1591.

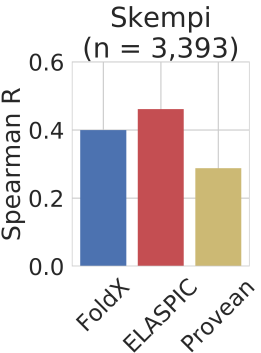
Performance on core mutations



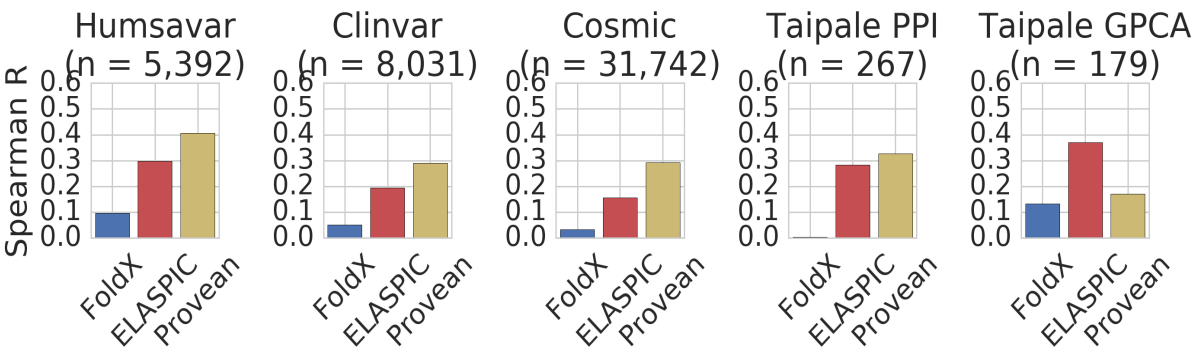
Performance on interface mutations

Training /
Validation

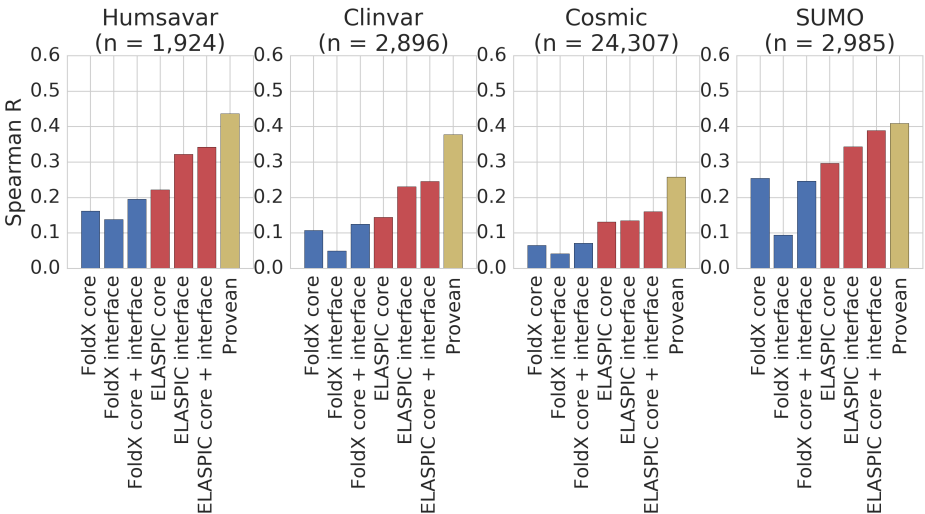
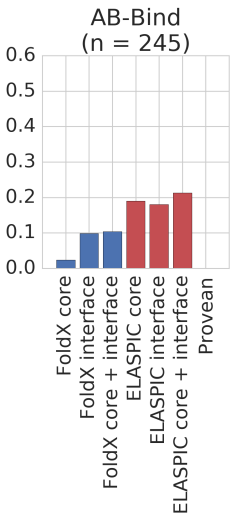
$\Delta\Delta G$ datasets



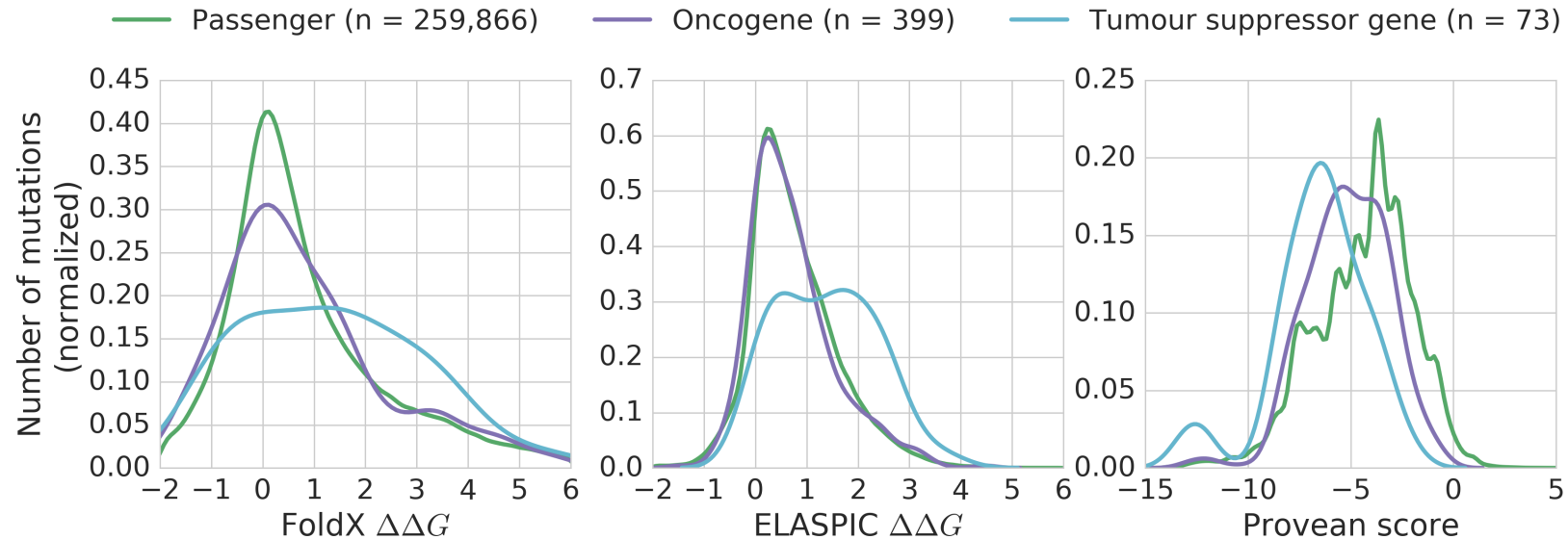
Phenotype datasets



Test



Performance on oncogenes / tumor suppressor genes in DoCM

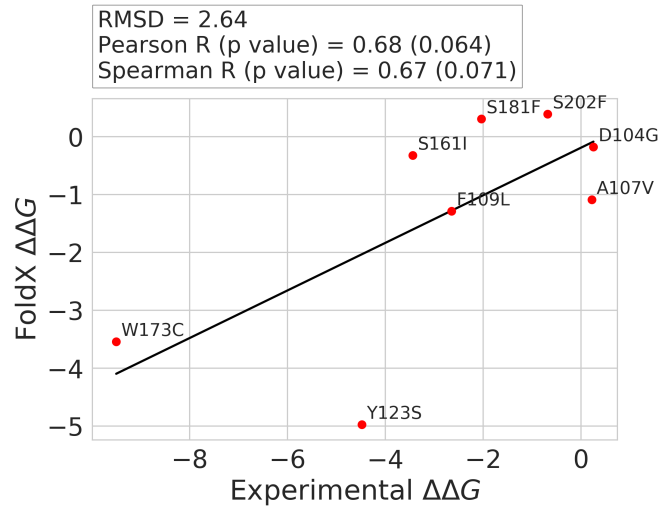


- Predictions for passenger mutations (green), mutations in tumour suppressors genes (purple), and mutations in oncogenes (cyan).

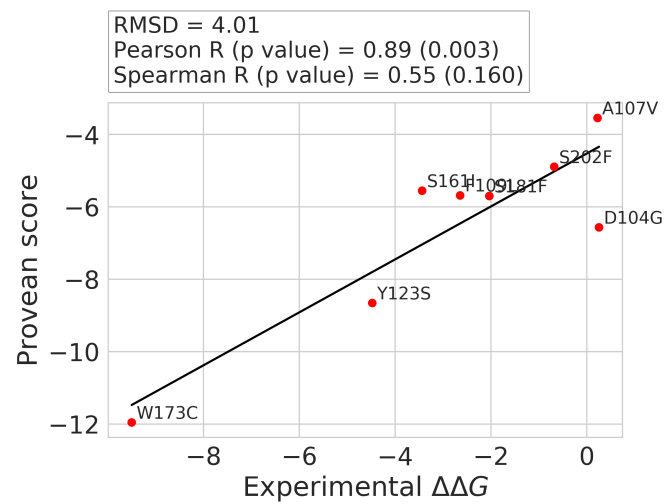
Results

ELASPIC results

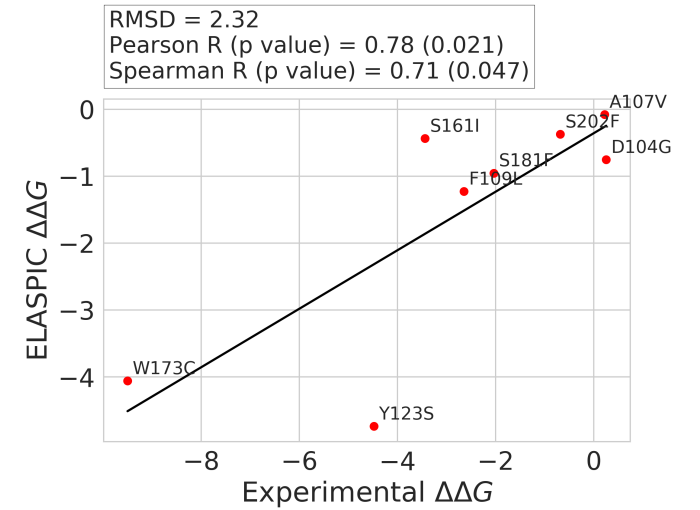
FoldX



Provean



ELASPIC



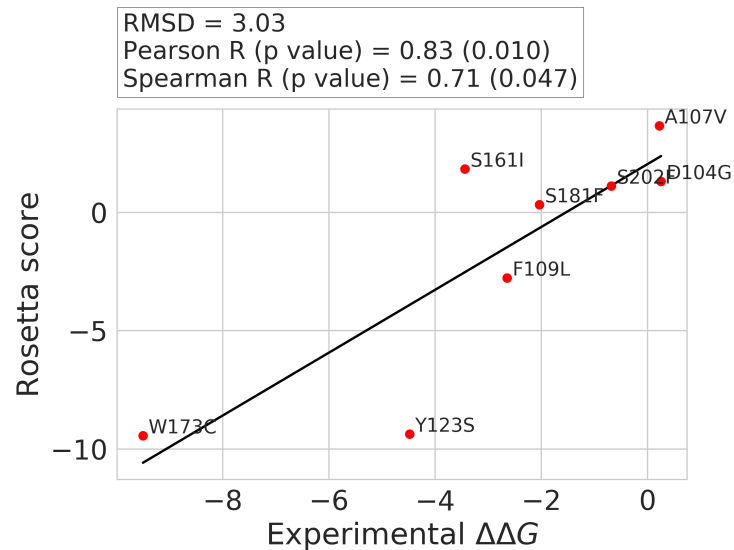
Note: $\Delta\Delta G$ values are changes in the Gibbs free energy of *unfolding* in kcal / mol.

ELASPIC results

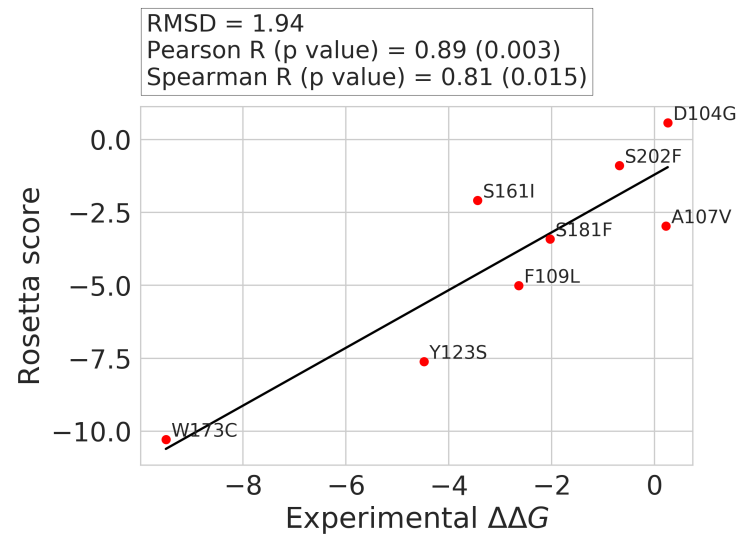
<http://elaspic.kimlab.org/result/a5393e/>
(<http://elaspic.kimlab.org/result/a5393e/>).

Rosetta results

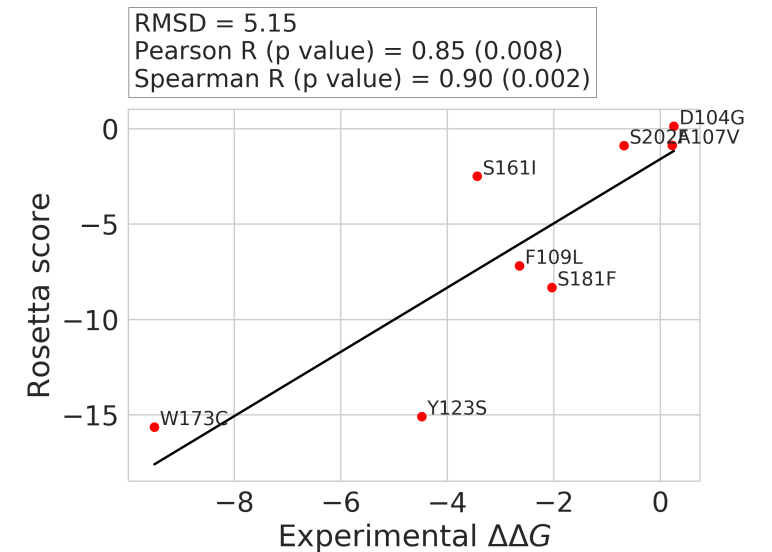
Protocol: **ddg_monomer**
Weights: **soft_rep_design**



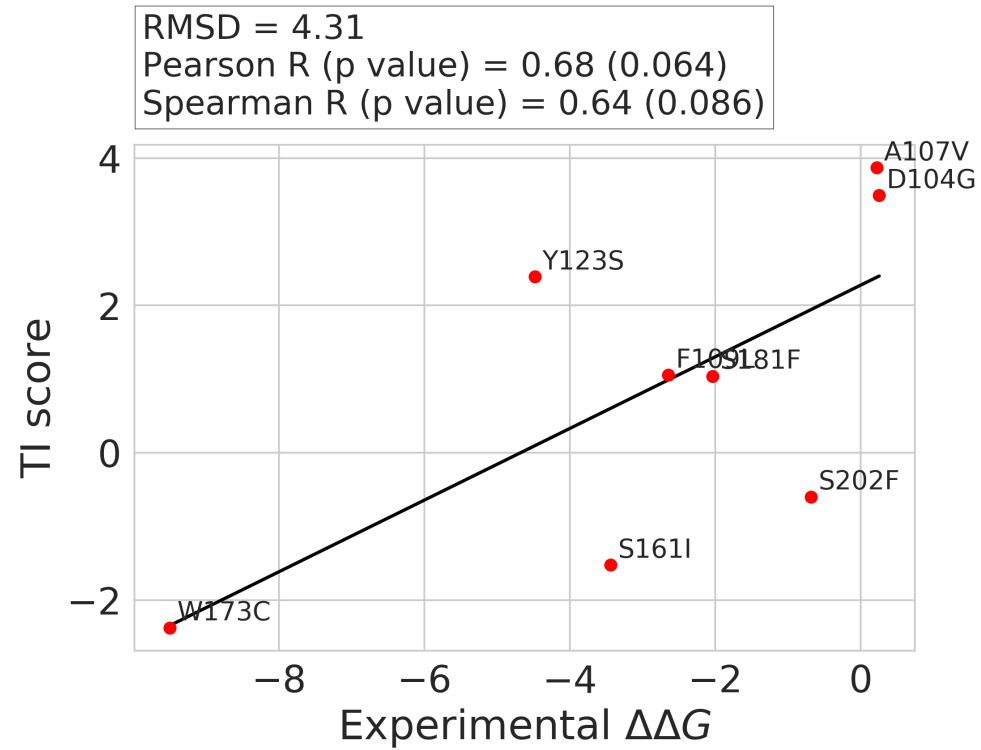
Protocol: **cartesian_ddg**
Weights: **talaris2013**



Protocol: **cartesian_ddg**
Weights: **beta_nov15**



Amber (thermodynamic integration) results



Conclusions

- Evolutionary information can be very useful for protein structure prediction and design.
- The ELASPIC webserver provides a user-friendly interface for evaluating the structural impact of mutations.
- Rosetta cartesian_ddg protocol with the talaris2013 or beta_nov15 weights produces the most accurate results.

Additional resources

- <http://kimlab.org/> (<http://kimlab.org/>).
- <http://elaspic.kimlab.org/> (<http://elaspic.kimlab.org/>).
- <https://gitlab.com/kimlab/elaspic/> (<https://gitlab.com/kimlab/elaspic/>).
- https://gitlab.com/ostrokach/msc_thesis/ (https://gitlab.com/ostrokach/msc_thesis/).
- <https://ostrokach-presentations.gitlab.io/cagi5/> (<https://ostrokach-presentations.gitlab.io/cagi5/>).

Acknowledgements

Supervisor

- Philip M. Kim

Members of kimlab

- David Becerra
- Recep Colak
- Carles Corbi
- Michael Garton
- Clare Juhyun Jeon
- Mark Sun
- Joan Teyra

Project students

- Daniel Witvliet

