

CAGI slide pack guidelines

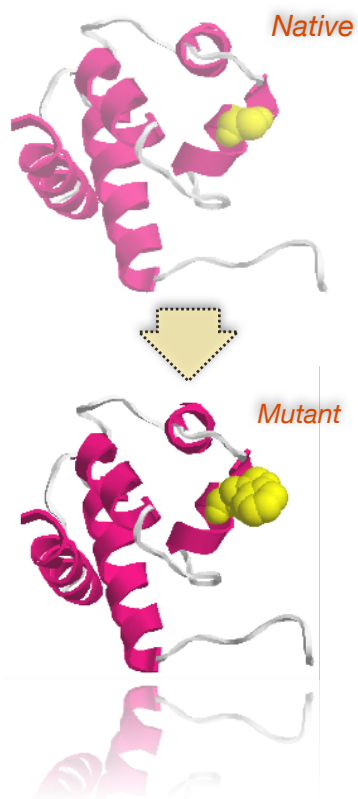
I agree to my slides being used according to the following terms:

Emidio Capriotti

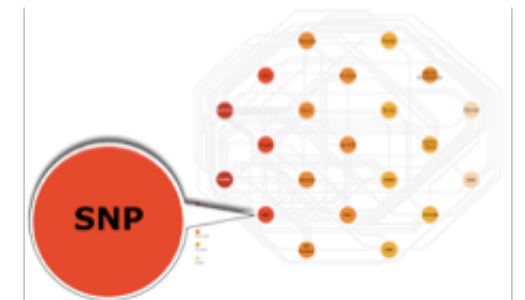
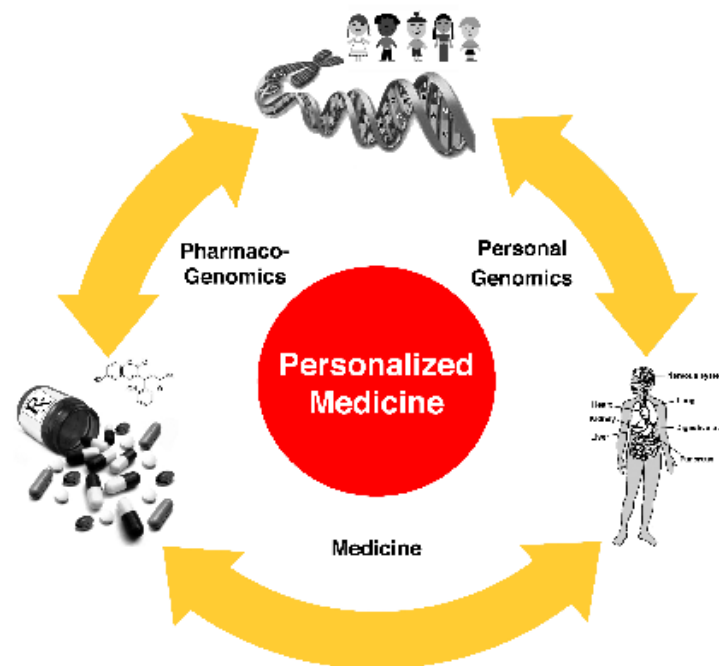
- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix 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 2013. 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.



Evaluating the Frataxin challenge predictions



CAGI 5 Conference
Chicago (IL), July 6, 2018



Emidio Capriotti
<http://biofold.org/>



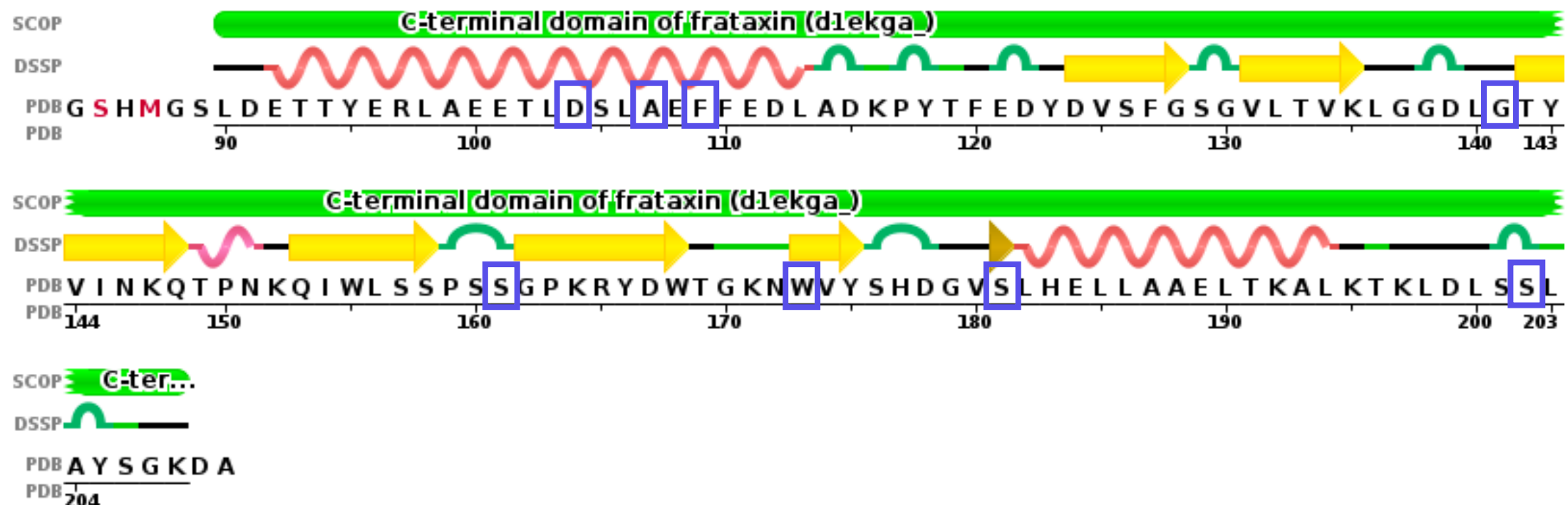
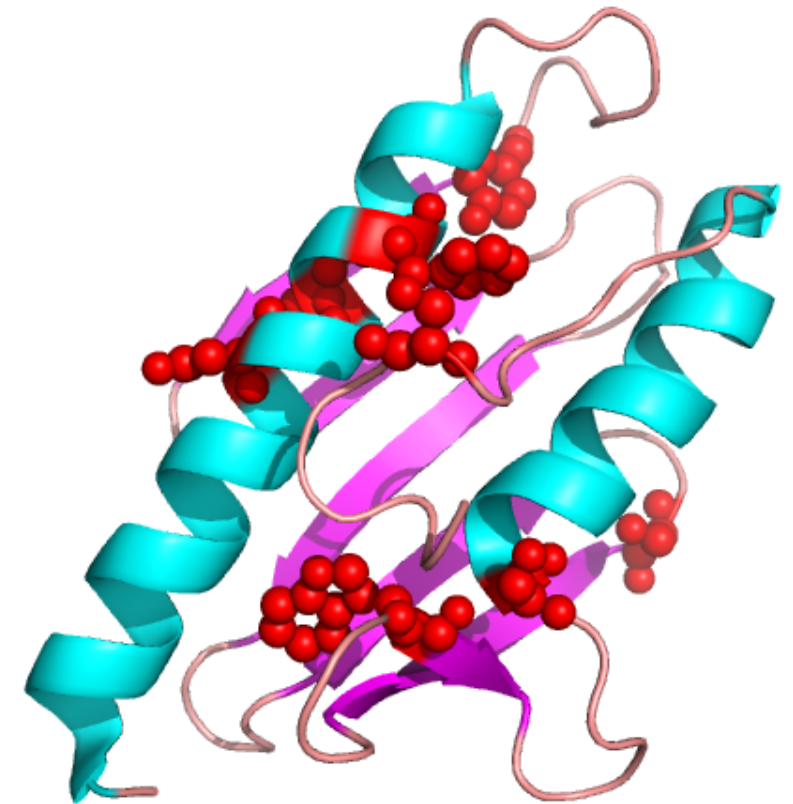
**Biomolecules
Folding and
Disease**

Department of Pharmacy
and Biotechnology (FaBiT)
University of Bologna



Frataxin challenge

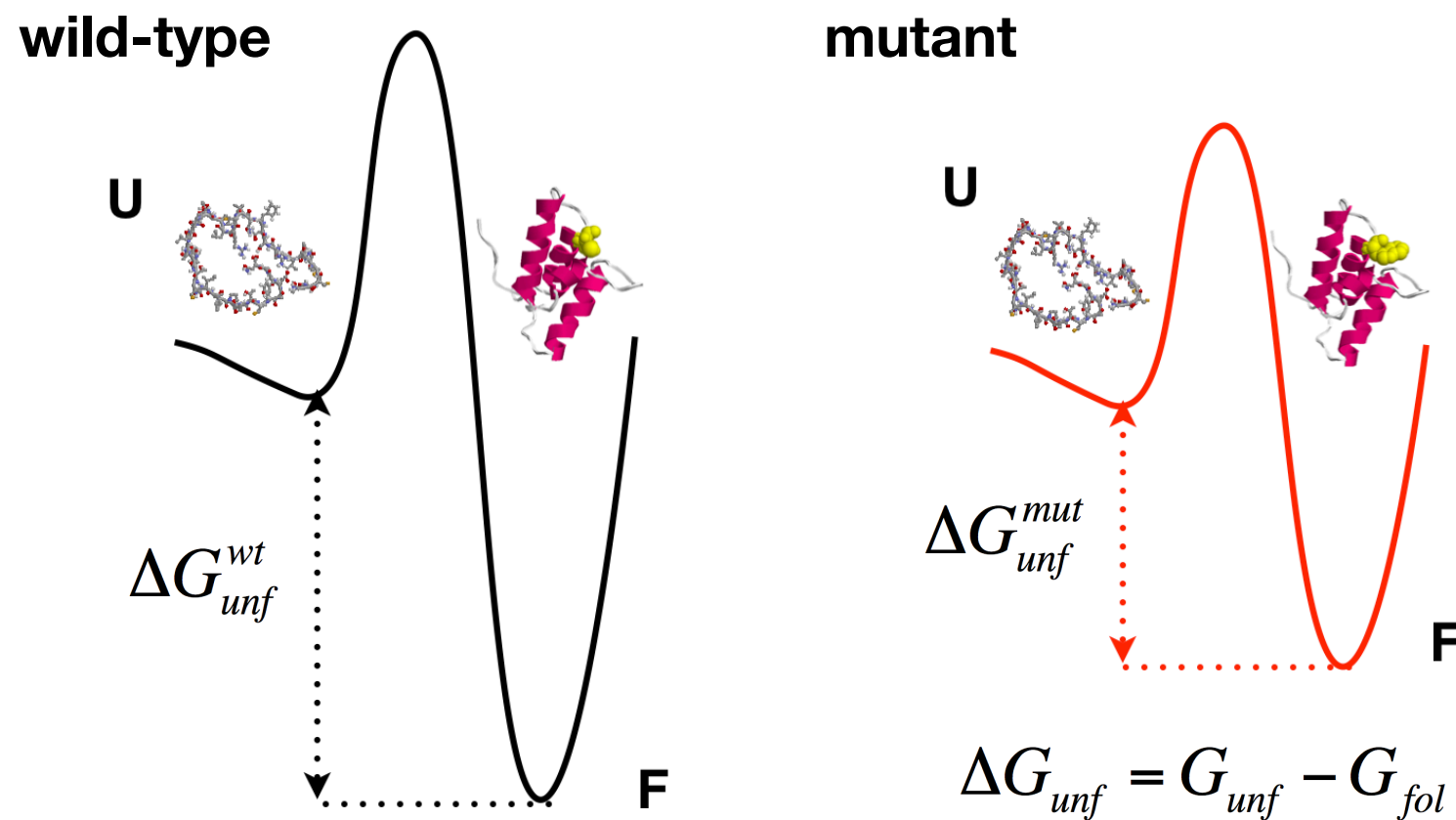
- Participants were asked to submit **predictions of the variation of the unfolding free energy change** upon mutation at concentration 0 of denaturant ($\Delta\Delta G_{H2O}$).
- Data providers at the University of Roma experimentally determined the **unfolding free energy of the wild-type and 8 mutants using CD and Fluorescence**. The average value between the two $\Delta\Delta G_{H2O}$ is considered for the assessment



Stability data

The variation of the unfolding free energy change ($\Delta\Delta G_{H2O}$) was calculated as:

$$\Delta\Delta G_{unf} = \Delta G_{unf}^{mut} - \Delta G_{unf}^{wt}$$



$$\Delta\Delta G_{unf} > 0 \text{ stabilizing}$$

$$\Delta\Delta G_{unf} < 0 \text{ destabilizing}$$

Evaluation scores

All predictions have been evaluated calculating the following regression scores

- Pearson Correlation Coefficient
- Spearman Rank Correlation Coefficient
- Kendall Tau Rank Correlation Coefficient
- RMSE (Root Mean Square Error)
- MAE (Mean Absolute Error)

Alternative versions of RMSE and MAE are calculated assigning 0 to the pairs of experimental and predicted values which distance is below 6 standard deviation

For the binary classification scores we considered a $\Delta\Delta G$ threshold of -1.0 kcal/mole obtaining a set of 5 destabilizing and 3 non destabilizing variants. Thus we calculated:

- Weighted Accuracy
- Matthews Correlation Coefficient
- AUC

The methods are ranked according to the average rank over 10 scores

Submitted predictions

To this challenge participated 6 Groups for a total of 12 submissions

- Group 1, 1 submission
- Group 2, 1 submission
- Group 3, 3 submissions
- Group 4, 2 late submissions
- Group 5, 2 submissions
- Group 6, 3 submissions

In the evaluation process we included the predictions from I-Mutant and Fold-X

Conclusions

- Top ranking methods for predicting the impact of missense variants on protein stability reached **good performances in terms of correlation coefficients** but the predictions still result in relatively high RMSE and MAE values.
- The selection of appropriate thresholds for defining a **binary classification task is important for the evaluation of the predictors** and useful to mitigate possible problems related to the distribution of the experimental data.
- The **detection of possible outliers in the experimental data** is important for selecting **more accurate datasets and estimate lower and upper bound** levels of performance.
- **Comparison with state-of-the-art and/or baseline methods** is important to estimate the **complexity of the challenge and progress made**

Acknowledgments

Data Providers: **Roberta Chiaraluce and Valerio Consalvi** (University of Roma)

CAGI Organizer: Steven Brenner, John Moulton, Gaia Andreoletti

Challenge Participants: 6 Groups

Lichtarge Lab - Baylor College of Medicine, United States

Bologna Biocomputing - University of Bologna, Italy

EASE - Griffith University, Australia

AIBI - Texas A&M University, United States

Prediction 0 Indian Institute of Science, India

ostrokach - University of Toronto, Canada

Biocomputing Group at the University of Bologna:

Rita Casadio, Pier Luigi Martelli and Castrense Savojardo

Biocomputing Group at the University of Padova:

Silvio Tosatto, Marco Carraro, Francesco Reggiani