Application of Digital Signal Processing and proteochemometrics principles to protein engineering *.

Nicolas FONTAINE, Olivier SAVRIAMA and Xavier CADET

PEACCEL, Protein Engineering Accelerator, n°6 square Albin Cachot, boite 42, 75013 PARIS, France. Email: nicolas.fontaine@peaccel.com
*Patented by Peaccel (FTO) www.peaccel.com

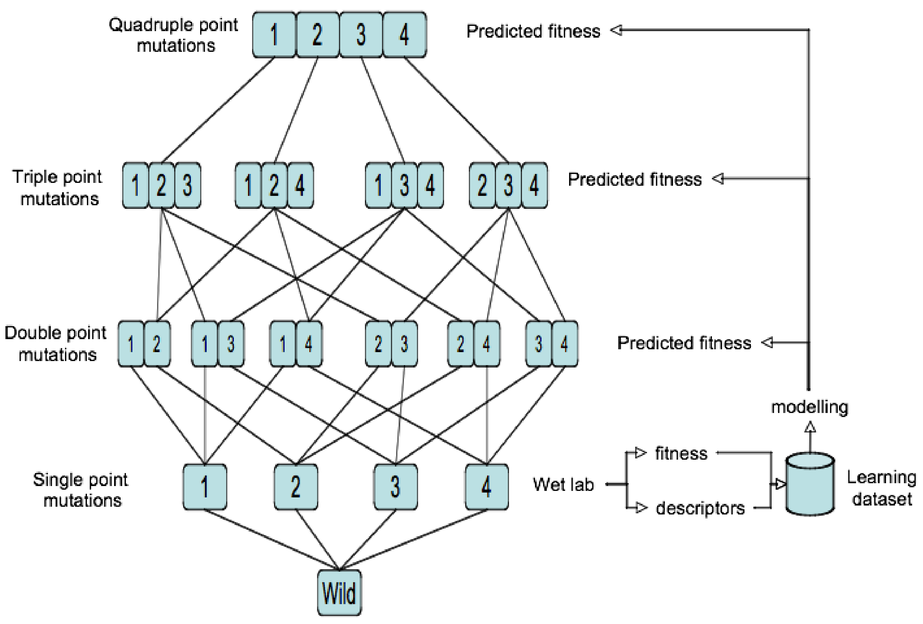
PEACCEL protein engineering accelerator

1 - The next generation approach for Protein Engineering

We present a strategy that combines wet-lab experimentation and computational protein design for engineering polypeptide chains. The enzyme/protein sequences were numerically coded and then processed using Fourier Transform (FT). Fourier coefficients were used to calculate the energy spectra called "protein spectrum". We use the protein spectrum to model the biological activity/fitness of enzyme/protein from sequence data. We assume that the protein fitness (catalytic efficacy, thermostabilty, binding affinity, aggregation, stability…) is not purely local, but globally distributed over the linear sequence of the protein. Our method (iSAR) finds patterns from primary sequences that correlate with changes in protein activity (or fitness) upon amino acids residue substitutions. A minimal wet lab data sampled from mutation libraries (single or multiple points mutations) were used as learning data sets in heuristic approaches that were applied to build predictive models. We show the performance of the approach on designed libraries for improvement of enantioselectivity of an epoxide hydrolase. We can screen up to 1 billion (10^9) protein candidates.

2 - Model for Fitness exploration

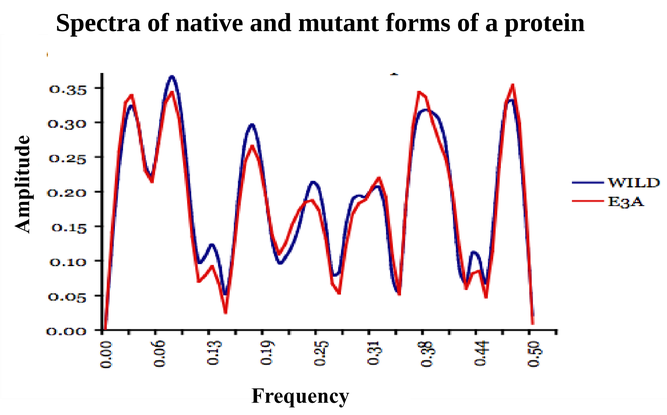
- Sequence based (no 3D structure is needed for modelling)
- Descriptors that capture global sequence information to avoid local fitness pitfalls of other methods
- Descriptors informative with respect to fitness/activity



3 - Mathematical treatment and Digital Signal Processing

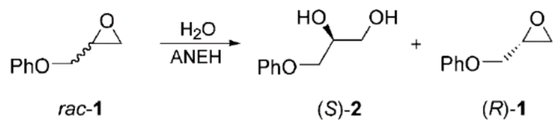
- Numerical encoding & choice of the best encoding using Aaindex database
- Spectral transform (FT)

1 mutation impacts the entire spectrum



Peaccel correlates variations caused by mutations in spectra with effects on biological properties

4 - Engineering enantioselectivity of enzymes

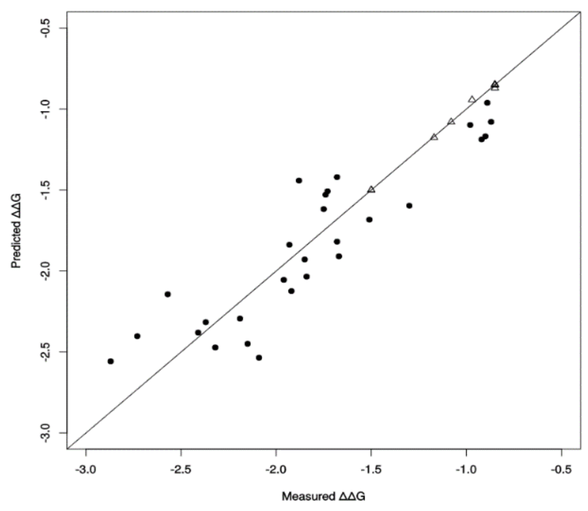


Reaction catalyzed by epoxide hydrolase from Aspergillus niger (ANEH)

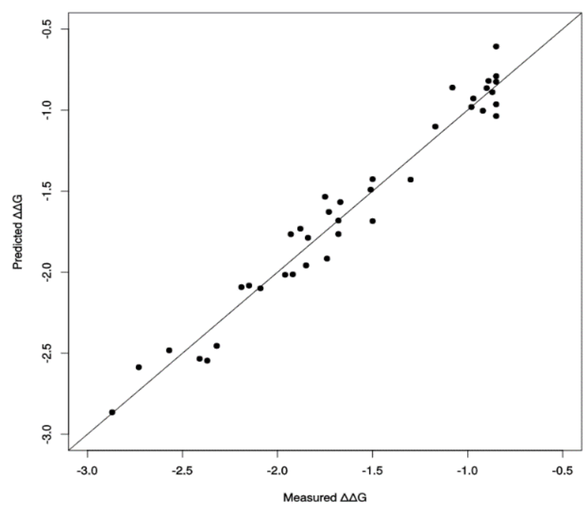
Iterative CASTing[1] resulted in an increase in enantioselectivity with an enantioselectivity factor (E) of 115 when compared to E=4.6 for the wild type (WT) ANEH, both in favor of the (S)-2 enantiomer. Five mutation sites were identified around the catalytic site of ANEH using CASTing. Each of these sites comprised of single, double or triple point mutations and were represented by the alphabets B, C, D, E and F (site A was excluded as it did not contribute much to improving the enantioselectivity)[1] . The CAST library amounts to a total of 9 single point mutations.

5 - Prediction using single or multiple point mutations

9 single point mutants of the enzyme were experimentally assessed for their enantioselectivity and used as a learning dataset to build a predictive model. Assessment of our heuristic on the prediction of the enantioselectivity indicates a very good predictability of our model using only single point mutations. Part of epistatic phenomenon is captured (data not shown).

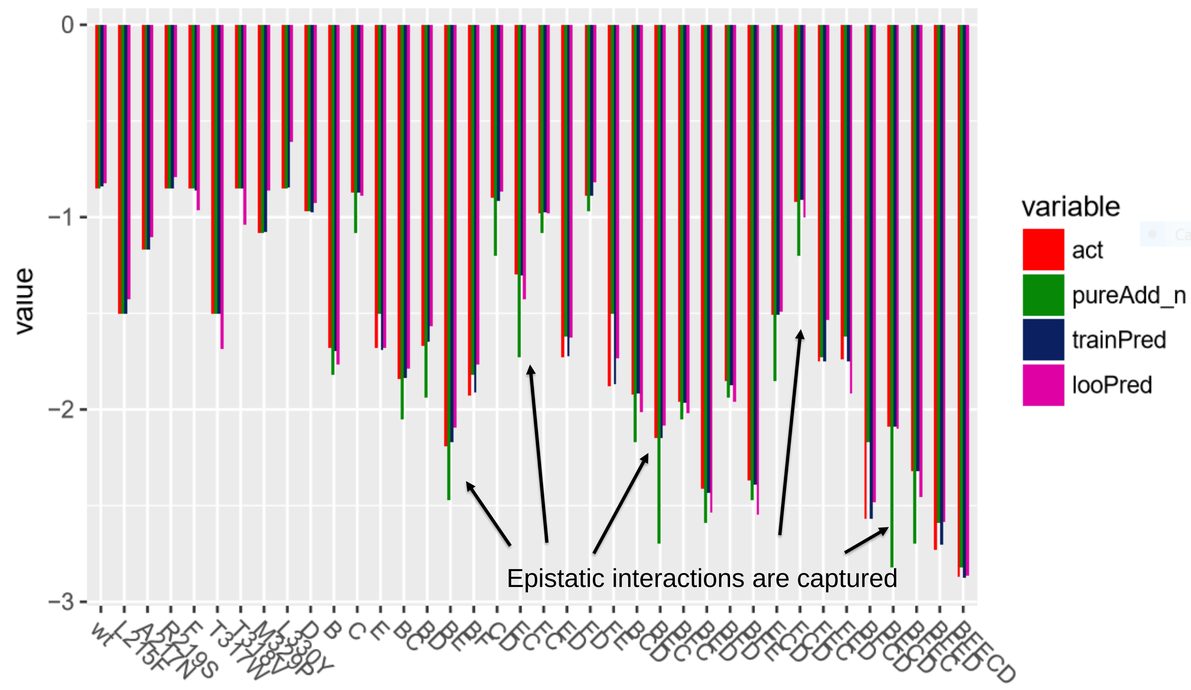


Model 1
Train set: wt+9 single point mutants, RMSE:0.01, R2:0.99
Test set: 28 multiple point mutants, RMSE: 0.24, R2: 0.81

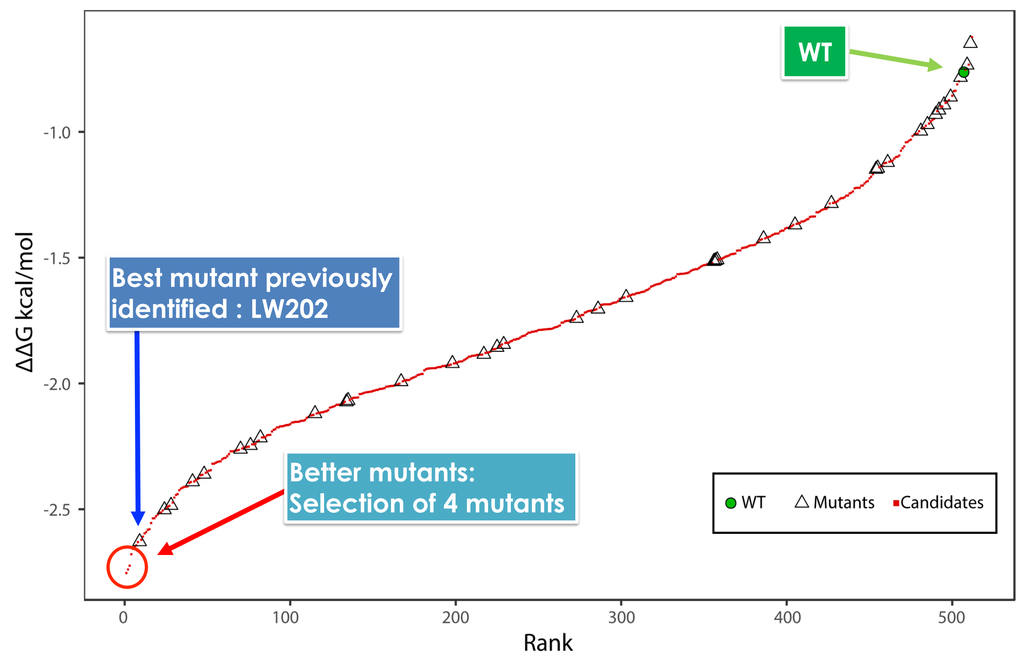


Model 2
Train set: WT + 37 mutants
LOOCV: cvRMSE: 0.117, cv R2: 0.96

Incorporation of ΔΔG values of multiple point mutations into the iSAR model generation brought improvements to the prediction of epistatic interactions.



6 - Ranking of 2^9 possible mutants & experimental validation



Combinatorial predictions: 2^9 = 512 possible mutants
References: WT: 4.6 LW202: 115
Experimental E-value for best selected mutants:
Mutant 1: 253 Mutant 2: 240
Mutant 3: 228 Mutant 4: 160

7 - Conclusion

- iSAR: Disruptive approach for Protein Engineering (based on sequence information)
- Good prediction using only single point mutants
- Epistatic interactions captured using multiple point mutants
- 55x enantioselectivity improvement compared to WT

Reference:
1 Reetz, M. T., Wang, L. W., & Bocola, M. (2006). Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. Angewandte Chemie, 118(8), 1258-1263