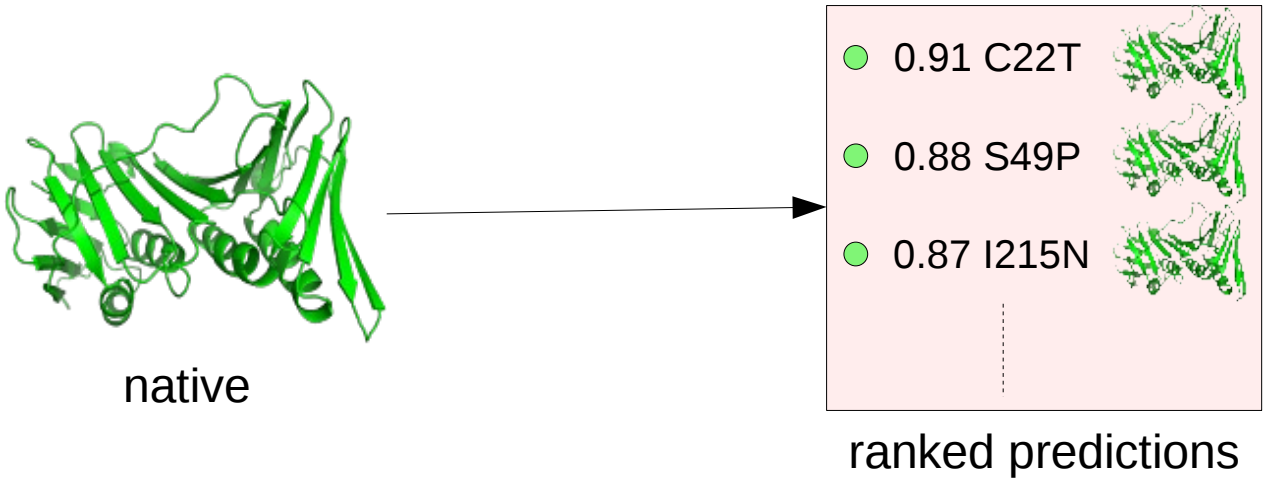
Predicting Temperature-Sensitive Mutations

Chris Poultney

Bonneau Lab

NYU

Motivation



- Find temperature-sensitive mutations
- Method to generate “top 5” list
- Use Rosetta to model mutations
- Machine learning for prediction

Conditional and “ts” Mutations

Wikipedia: Genetics

- Conditional Mutation
  - Wild-type (wt) phenotype under permissive conditions
  - Mutant phenotype under restrictive conditions
- Temperature-Sensitive (ts)
  - Restrictive condition is different temperature
- Context
  - Knock-out libraries (YKO)
  - Embryonic lethal phenotype

Lyons LA, Imes DL, Rah HC, Grahn RA (2005) Tyrosinase mutations associated with Siamese and Burmese patterns in the domestic cat (Felis catus). Anim Genet **36:** 119-126

Causes of ts Behavior

- What changes at restrictive temperature?
  - Drop in level or activity of gene product
- Implications for Rosetta-based method
  - Likely to detect
    - Decrease in stability
    - Failure to fold
  - Unlikely to detect
    - Reduced function (e.g., catalysis, DNA binding)
    - Aggregation

Chakshusmathi G, Mondal K, Lakshmi GS, Singh G, Roy A, Ch RB, Madhusudhanan S, Varadarajan R (2004) Design of temperature-sensitive mutants solely from amino acid sequence. Proc Natl Acad Sci U S A **101:** 7925-7930

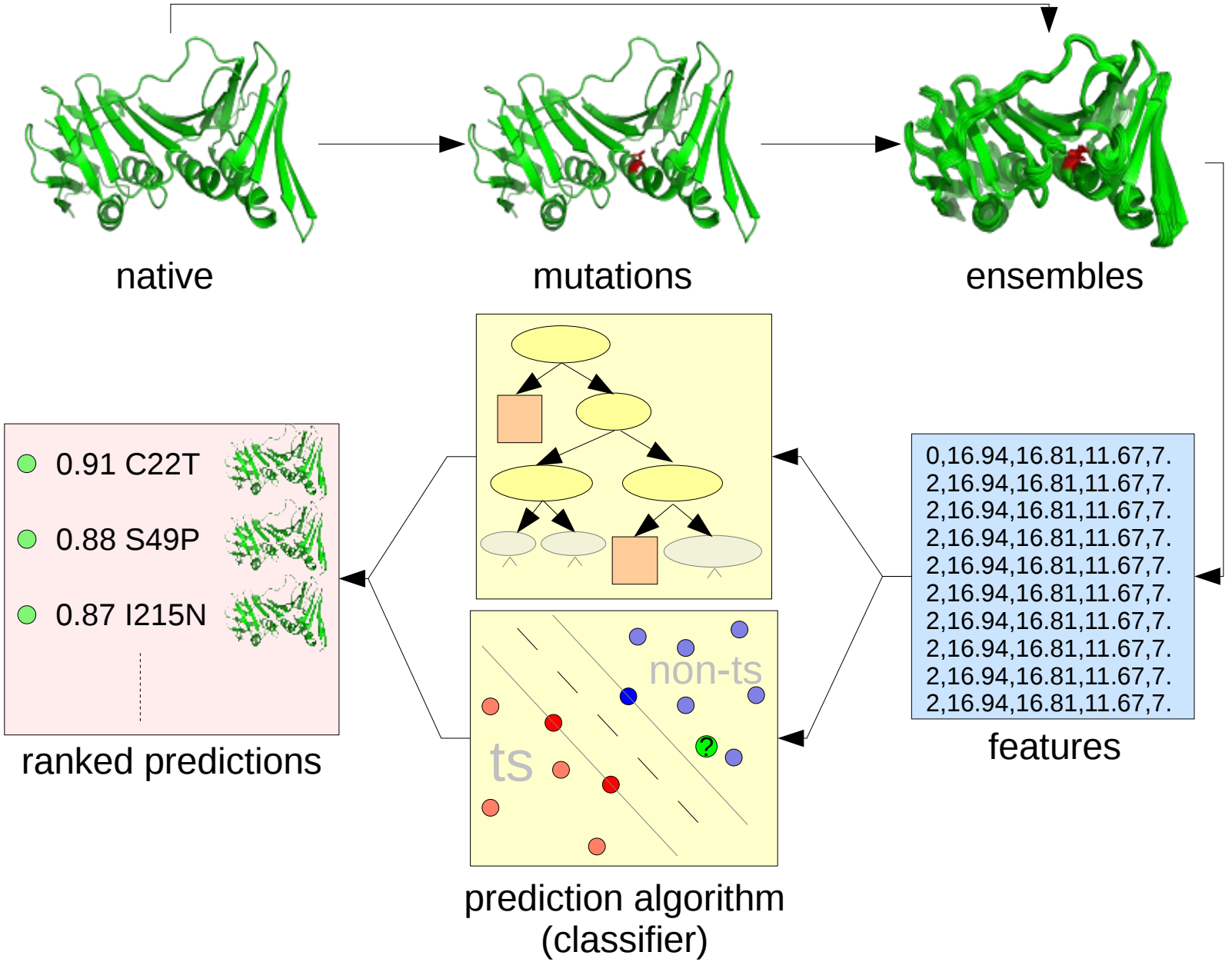
Sandberg WS, Schlunk PM, Zabin HB, Terwilliger TC (1995) Relationship between in vivo activity and in vitro measures of function and stability of a protein. Biochemistry **34:** 11970-11978

Generating ts Mutations

Category	Method(s)	Screening	Probability
Whole genome mutation	EMS, X-ray, ...	~100,000	~0.00001
Single gene mutation	PCR mutagenesis	~10,000	~0.0001
Single gene mutation	Diploid shuffle	~10,000	~0.0001*
Prediction from sequence	Burial from sequence	~65	~0.1
Prediction from structure	Top5	~5	~0.4

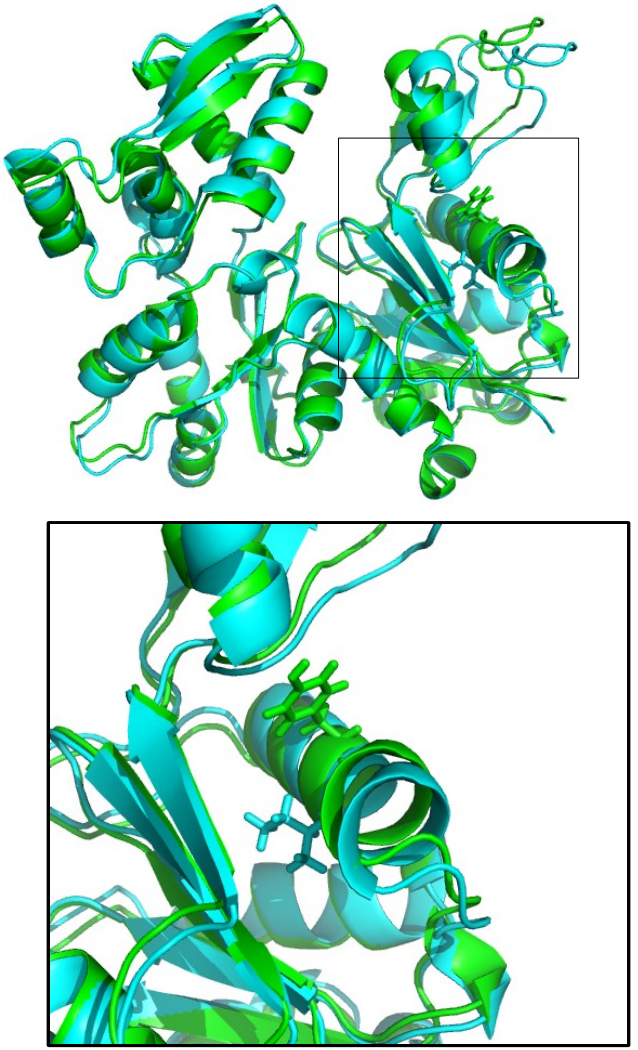
Suzuki DT, Grigliatti T, Williamson R (1971) Temperature-sensitive mutations in Drosophila melanogaster. VII. A mutation (para-ts) causing reversible adult paralysis. Proc Natl Acad Sci U S A **68:** 890-893
Dohmen RJ, Wu P, Varshavsky A (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science **263:** 1273-1276
Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N (2004) Temperature-sensitive control of protein activity by conditionally splicing inteins. Nat Biotechnol **22:** 871-876
Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, Hieter P (2008) Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol Cell **30:** 248-258
Chakshusmathi G, Mondal K, Lakshmi GS, Singh G, Roy A, Ch RB, Madhusudhanan S, Varadarajan R (2004) Design of temperature-sensitive mutants solely from amino acid sequence. Proc Natl Acad Sci U S A **101:** 7925-7930

Method Overview



Rosetta & Finding the “Sweet Spot”

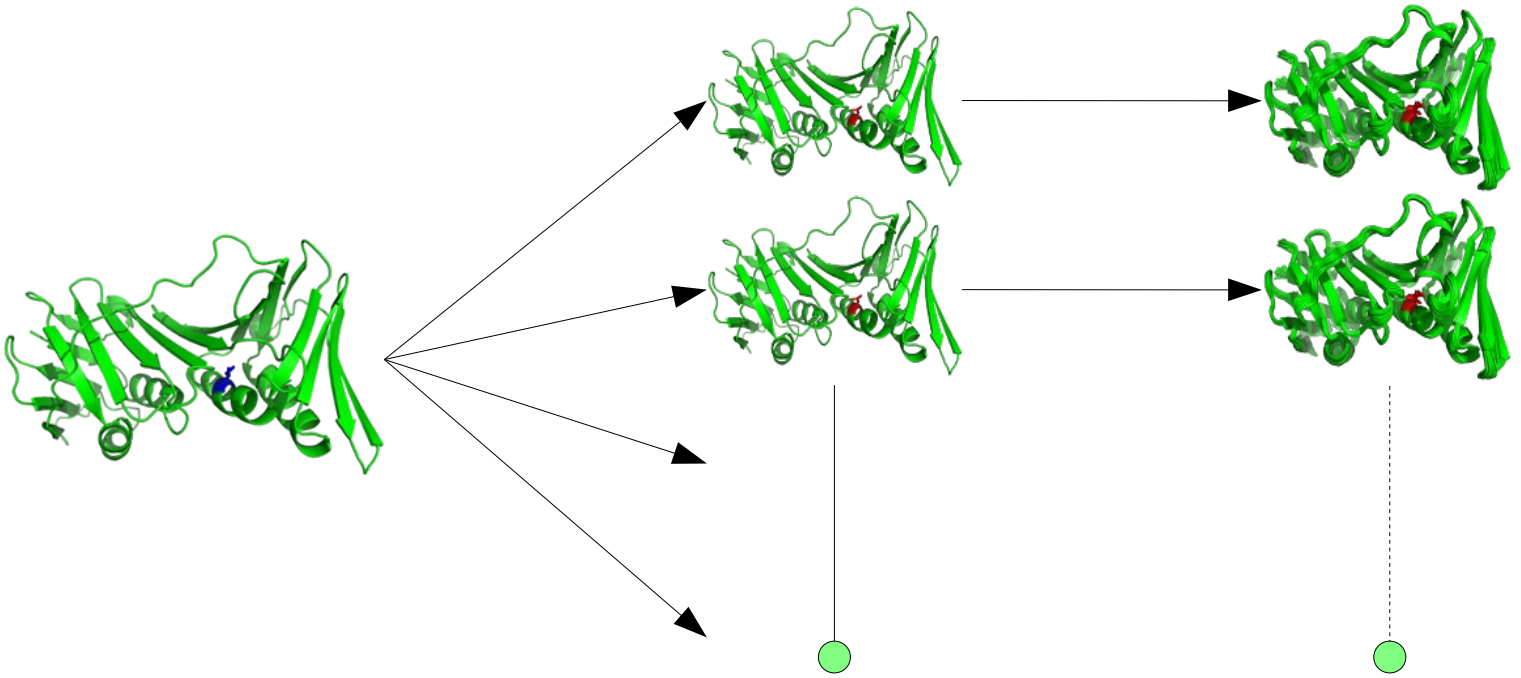
- “Sweet Spot”: intermediate degree of destabilization
  - Moderate increase in energy (e.g. fa_rep)
- Proteins vary in starting energy and properties (e.g. stability)
- Rosetta score function
- Allow structure to adjust to mutation



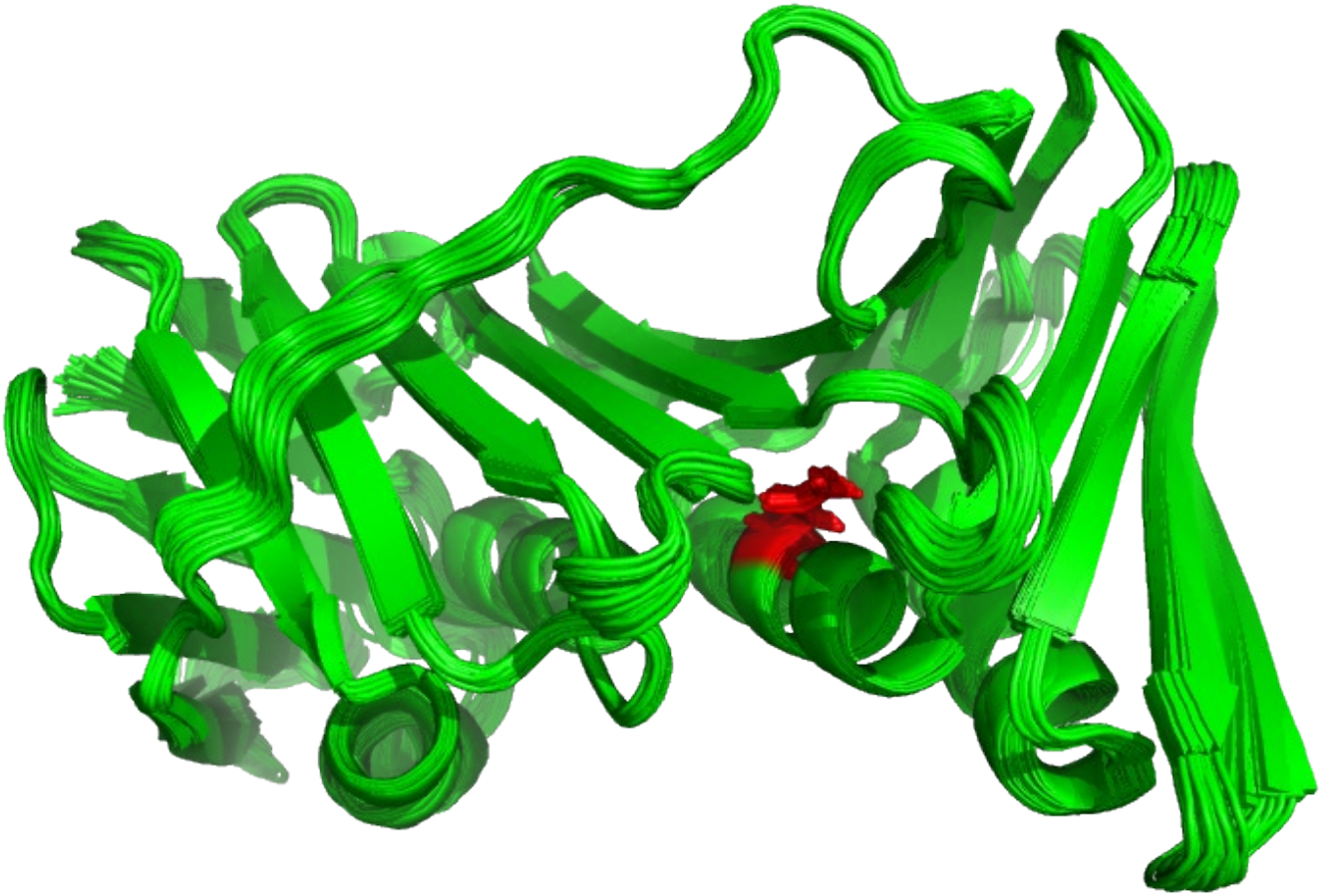
green: ts blue: non-ts

Rosetta Protocol

- Start with native structure
- Model mutations at buried sites (<10% accessible)
- Perform 50 relax runs
  - Generates model ensembles and score files



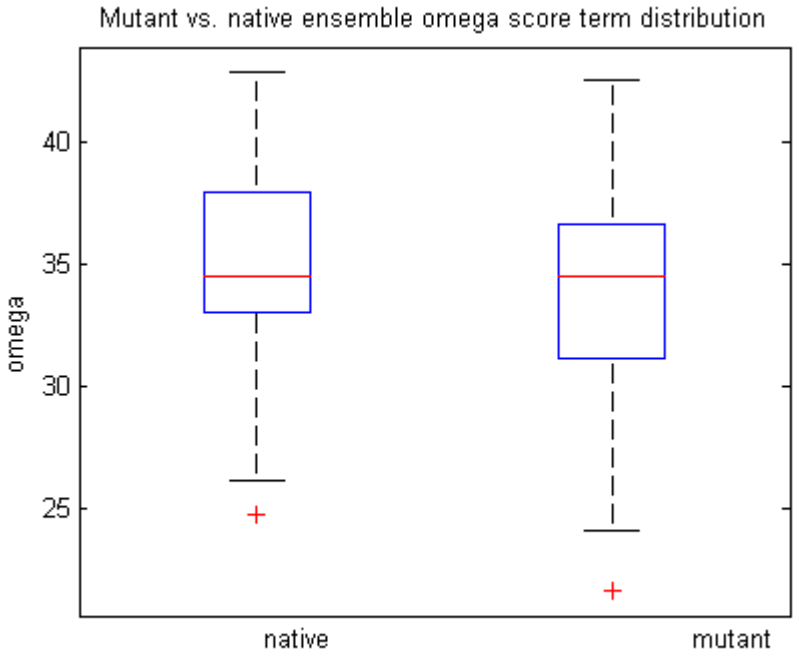
Relax Ensemble

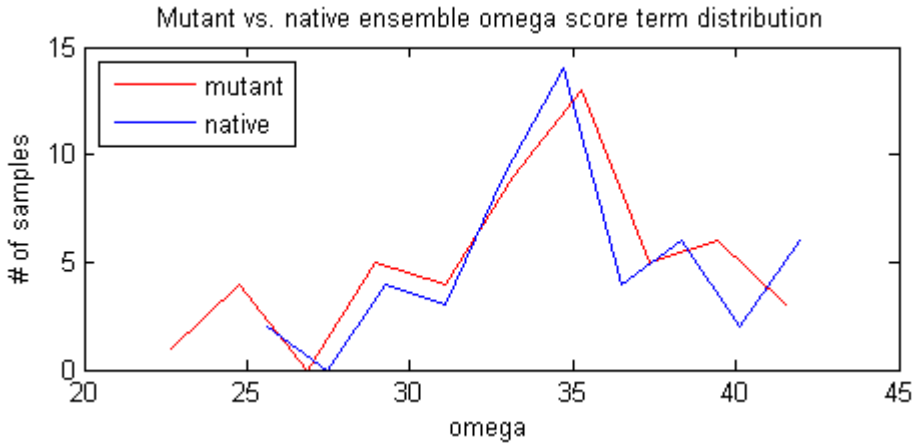


relax.linuxgccrelease -database $MINI_DB -s YPL228W-W251A.pdb -native YPL228W.pdb -nstruct 50 -relax:fast -out:file:scorefile YPL228W-W251A.sc -out:pdb_gz

version: 3.0 release

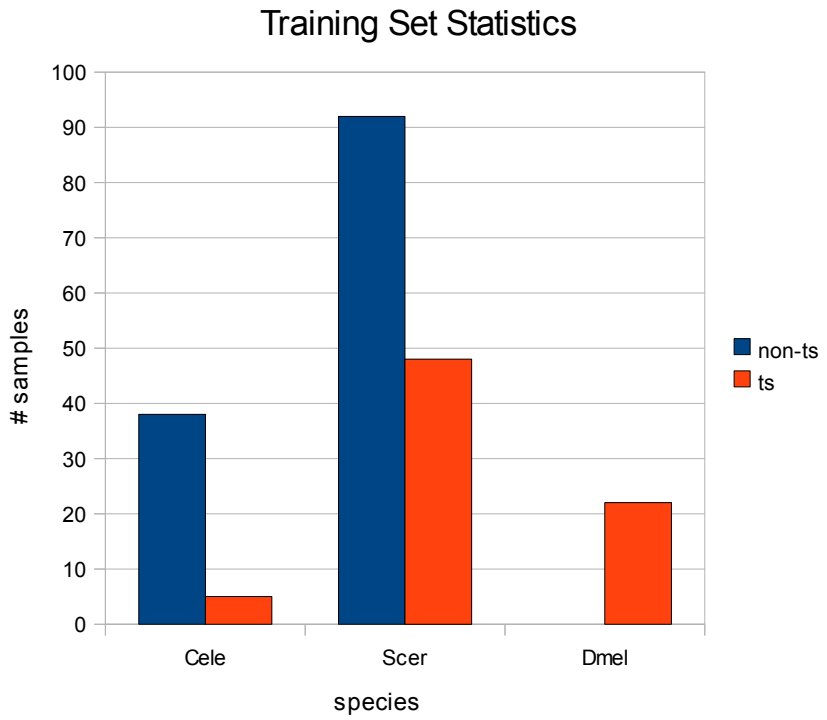
Scores to Features

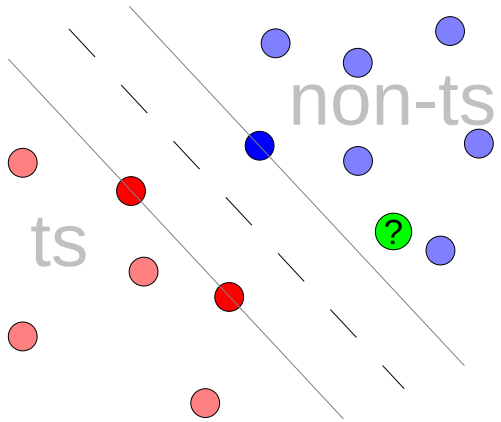




- How to quantify effect of mutation?
  - Different starting energy
  - Different native qualities
- Requirements
  - Compare score term distributions
  - Normalize across proteins
- Compare quartiles of mutant and native ensembles

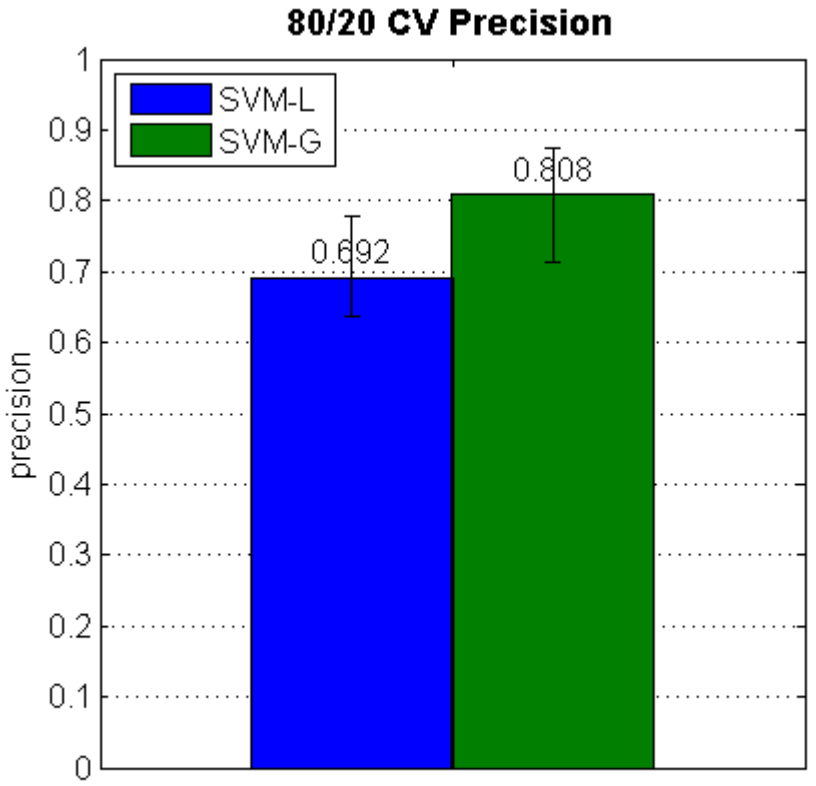
Machine Learning



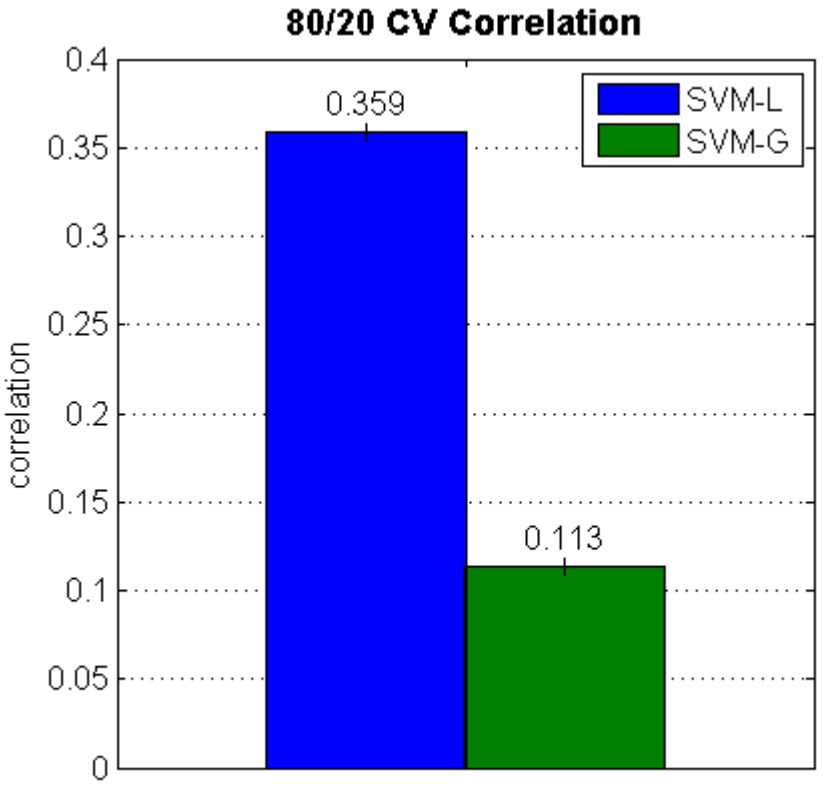


- Input: 81 features
  - 75 from from Rosetta score terms
  - 6 other (accessibility, etc.)
- Training set: yeast, worm, and fly from literature
  - 382 samples: 75 ts, 207 non-ts
- Algorithms
  - SVM-L: SVM with linear kernel
  - SVM-G: SVM with gaussian kernel (RBF)

Cross-Validation Results



SVM-L: 0.692±.074
SVM-G: 0.808±.082
rnd: 0.366



SVM-L: 0.359
SVM-G: 0.113

Top Features

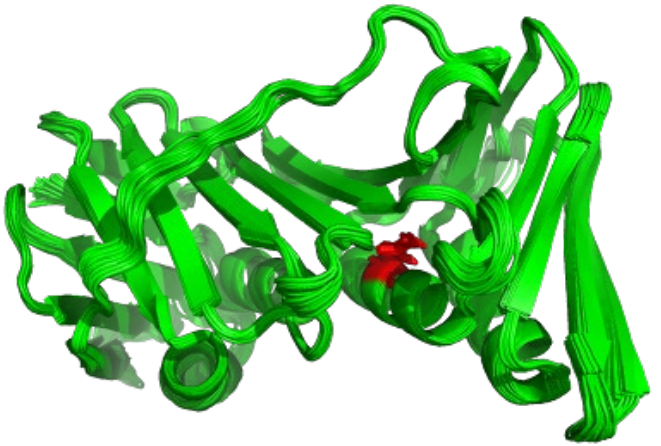
- Most important features
  - 'Qn' suffix score-term-to-feature conversion
  - Residue change
    - aminochange, p_aa_ppQ3, ramaQ2
  - Local structure change
    - fa_repQ2, hbond_sr_bbQ2, hbond_bb_scQ1
  - Global structure change
    - gdtmm2_2Q3, gdtmm1_1Q3
  - Changes within relax run
    - Repack_stdev_scoreQ2, Repack_average_scoreQ2

Experimental Validation

- Initial validation on three species
  - Worm: Kris Gunsalus & Fabio Piano
  - Fly: Claude Desplan
  - Yeast: David Gresham
- Current yeast validation
  - Made predictions on yeast actin
    - 375 residues, well-characterized
    - Difficult to find ts mutations at random
  - Chose 7 candidates from SMO-L, SVM-G top 5
  - Literature search: all mutations uncharacterized
  - [insert cool results here]

Future Work

- Improving speed
  - Current algorithm: global (runs over every residue)
  - In development: local (runs only on residues near mutation)
  - Estimated speedup ~10-fold
- ts prediction for the masses
  - Public web server
  - Submit structure of interest
  - Receive ranked list of candidate ts mutations

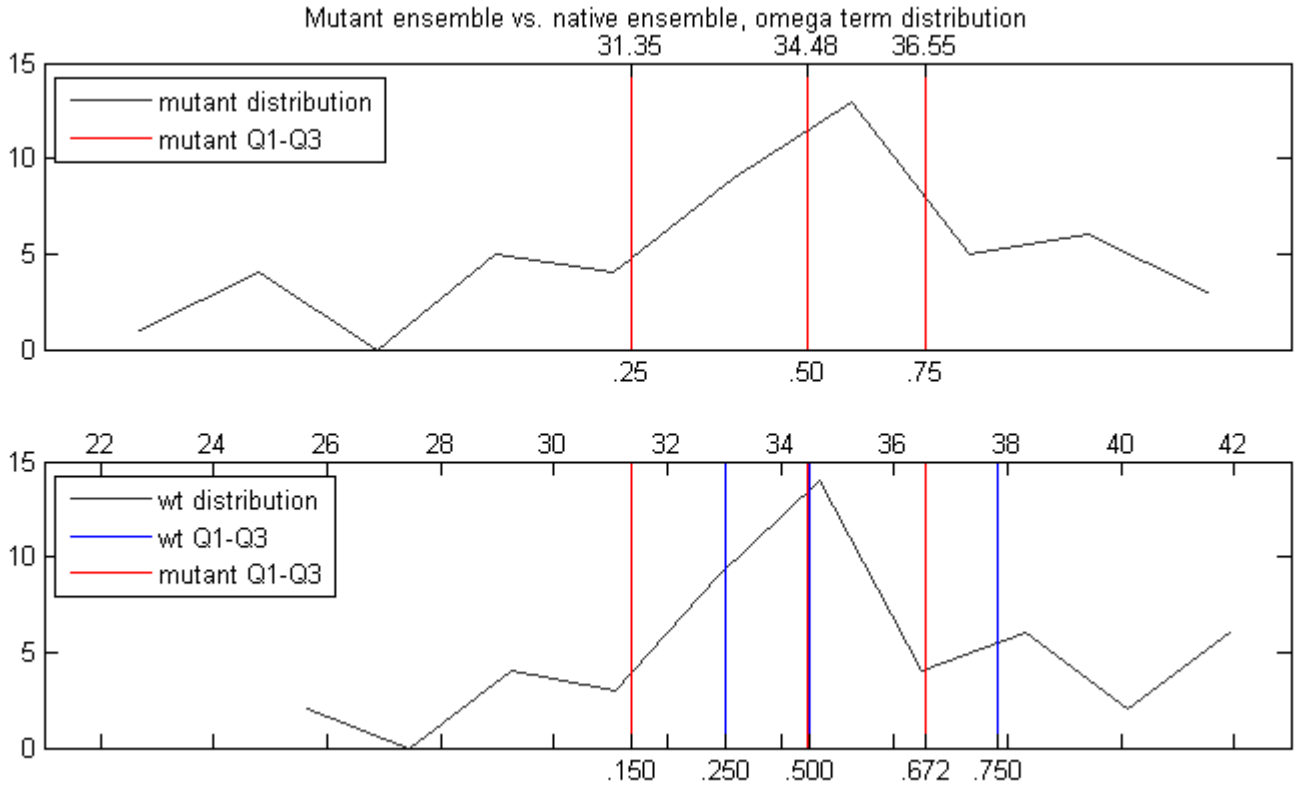


Thanks

Dennis Shasha
Rich Bonneau
Glenn Butterfoss
Kevin Drew
Kris Gunsalus
Michelle Gutwein
David Jukam
David Gresham

Processing Scores: Quartile Method

- Calculate percentile of mutation ensemble quartiles Q1, Q2, Q3 w.r.t. native ensemble



$$Output: (omegaQ1, omegaQ2, omegaQ3) = (0.150, 0.499, 0.672)$$

Training Method

- Training Set
  - Split into 80% / 20%
  - Parameter selection: 10-fold CV on 80%
  - Testing: train on 80%, test on 20%
  - Repeat 5x