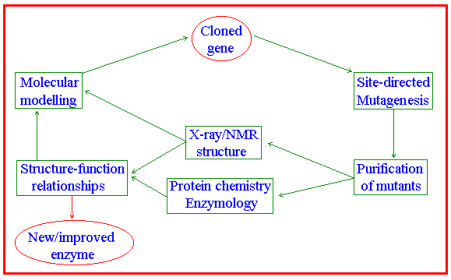
BINF 731

Protein Structure Analysis

Iosif Vaisman

2012

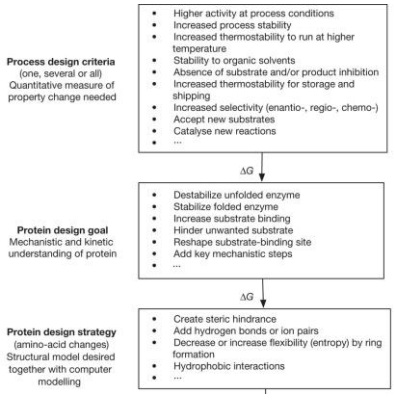
Protein Engineering



Protein Engineering

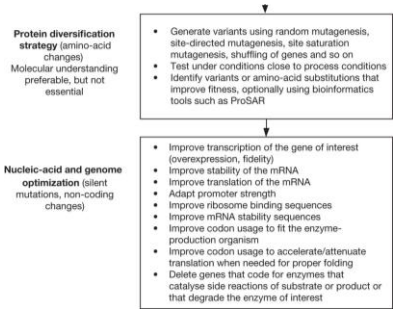
Increase catalytic activity
Change substrate binding site to increase specificity
Change the thermal stability
Increase proteins resistance to proteases
Change codon composition

Protein Engineering



Bornscheuer et al. Nature 485, 185-194 (2012)

Protein Engineering



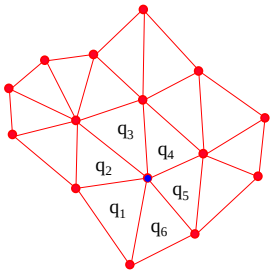
Bornscheuer et al. Nature 485, 185-194 (2012)

Computational Mutagenesis

Assumption: the structural differences between each mutant and the wild-type protein are usually minor and, therefore, their tessellations are similar

Approach: a single tessellation of either the wild-type or mutant protein structure can be used to develop environmental descriptors for quantitative evaluation of changes in mutant properties

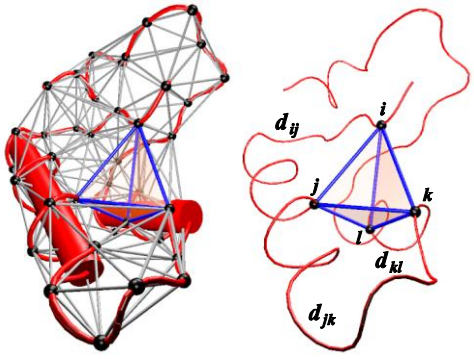
Residue and mutant score



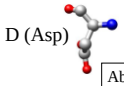
$q_{res} = \Sigma q_i$

$Q_{mut} = \Sigma(q_{res}^{wt} - q_{res}^{mut})$

Dealunay simplices classification

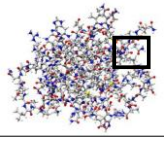


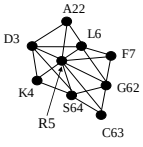
Delaunay Tessellation of Protein Structure



D (Asp) → C_α or center of mass

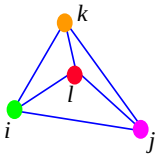
Abstract each amino acid to a point
Atomic coordinates – Protein Data Bank (PDB)





Delaunay tessellation: 3D "tiling" of space into non-overlapping, irregular tetrahedral simplices. Each simplex objectively defines a quadruplet of nearest-neighbor amino acids at its vertices.

Compositional propensities of Delaunay simplices



$$q_{ijkl} = \log \frac{f_{ijkl}}{p_{ijkl}}$$

f- observed quadruplet frequency,

$p_{ijkl} = Ca_ia_ja_ka_l$, *a - residue frequency*

AAAA: $C = 4! / 4! = 1$
AAAV: $C = 4! / (3! \times 1!) = 4$
AAVV: $C = 4! / (2! \times 2!) = 6$
AAVR: $C = 4! / (2! \times 1! \times 1!) = 12$
AVRS: $C = 4! / (1! \times 1! \times 1! \times 1!)) = 24$

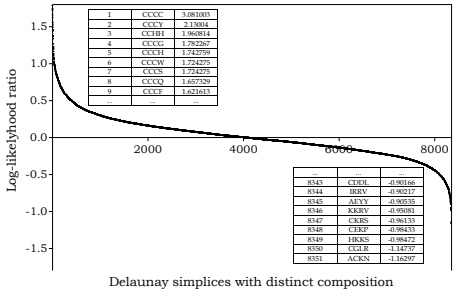
$$C = \frac{4!}{\prod_i^n (t_i!)}$$

Counting Quadruplets

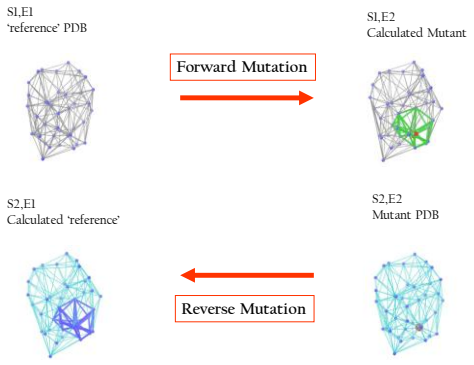
- assuming order independence among residues comprising Delaunay simplices, the maximum number of all possible combinations of quadruplets forming such simplices is 8855

C D E F	$\binom{20}{4}$	4845
C C D E	$20 \cdot \binom{19}{2}$	3420
C C D D	$\binom{20}{2}$	190
C C C D	$20 \cdot 19$	380
C C C C	20	20
		8855

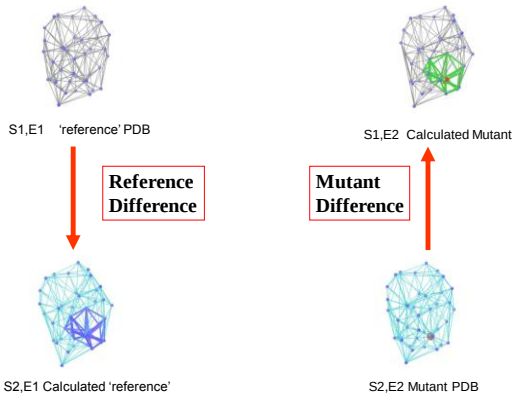
Log-likelihood of amino acid quadruplets with different compositions



Reversibility Analysis



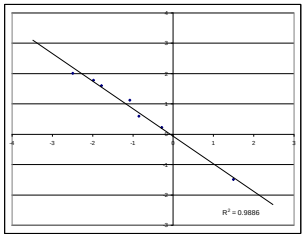
Structural Analysis



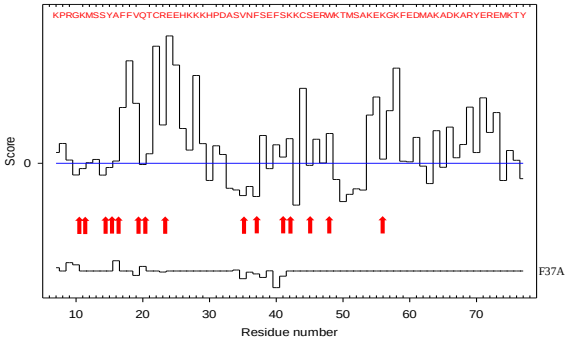
Computational mutagenesis of T4 lysozyme

Reversibility of mutations

Protein	Mutation	Score change
1163	T26E	-2.49
1801	E26T	2.01
1163	A82S	1.49
1231	S82A	-1.49
1163	V87M	-0.28
1cu3	M87V	0.22
1163	A93C	-1.98
1381	C93A	1.78
1163	T152S	-1.08
1goj	S152T	1.12

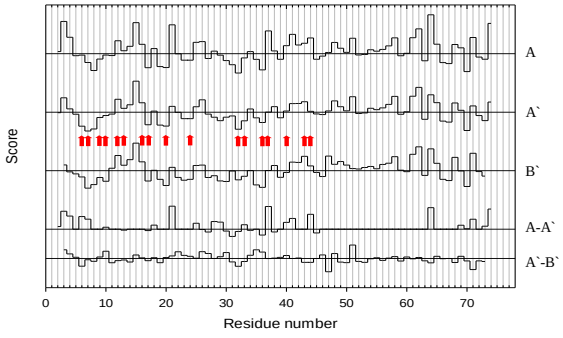


DNA binding residues in HMG1



Coordinate file 1ckt: Ohndorf U-M et al. *Nature* **399**:708

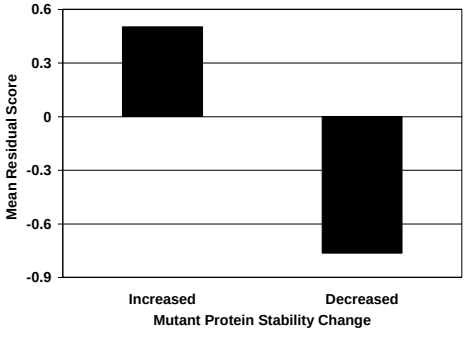
Protein-protein and protein-DNA interfaces (HMG-D)



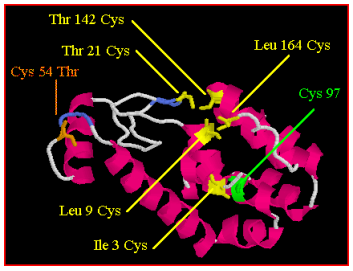
Coordinate file 1qrv: Murphy F V et al. *EMBO Journal* **18**:6610

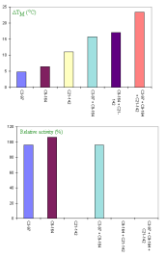
Universal Model Approach:

980 Experimental Mutants from 20 Proteins

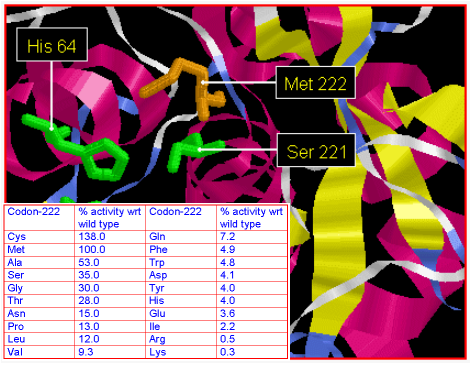


Protein Engineering



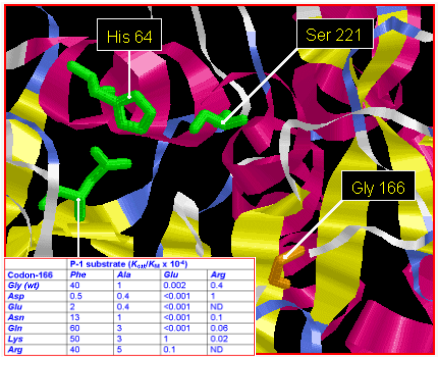


Protein Engineering



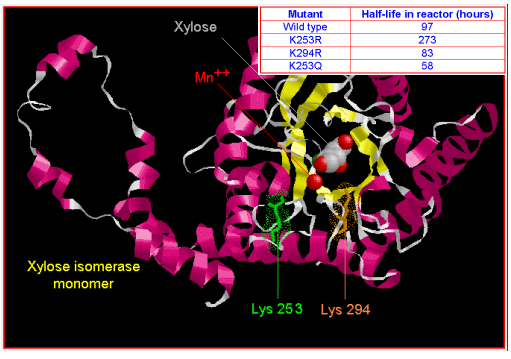
Codon-222	% activity wrt wild type	Codon-222	% activity wrt wild type
Cys	138.0	Gln	7.2
Met	100.0	Phe	4.9
Ala	53.0	Trp	4.8
Ser	35.0	Asp	4.1
Gly	30.0	Tyr	4.0
Thr	28.0	His	4.0
Asn	15.0	Glu	3.6
Pro	13.0	Ile	2.2
Leu	12.0	Arg	0.5
Val	9.3	Lys	0.3

Protein Engineering



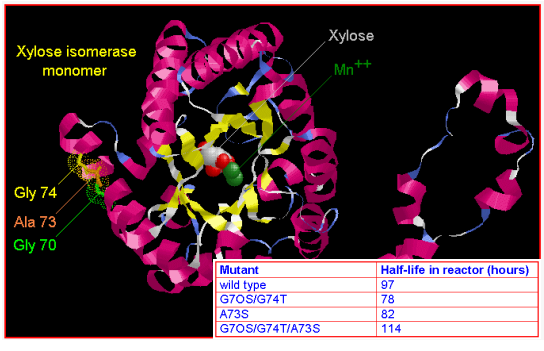
	P-1 substrate ($K_{cat}/K_M \times 10^4$)			
Codon-166	Phe	Ala	Glu	Arg
Gly (wt)	40	1	0.002	0.4
Asp	0.5	0.4	<0.001	1
Glu	2	0.4	<0.001	ND
Asn	13	1	<0.001	0.1
Gln	60	3	<0.001	0.06
Lys	50	3	1	0.02
Arg	40	5	0.1	ND

Protein Engineering



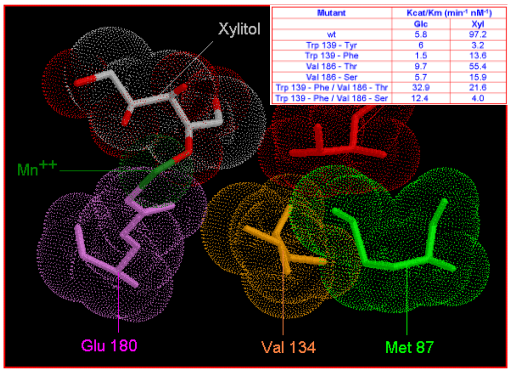
Mutant	Half-life in reactor (hours)
Wild type	97
K253R	273
K294R	83
K253Q	58

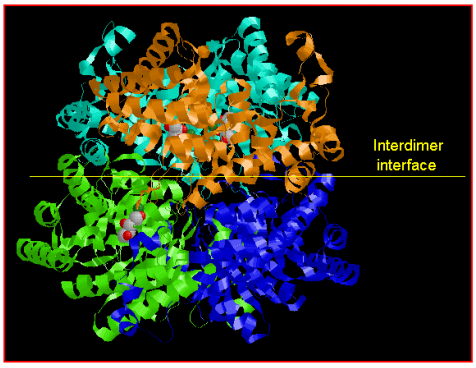
Protein Engineering



Mutant	Half-life in reactor (hours)
wild type	97
G7OS/G74T	78
A73S	82
G7OS/G74T/A73S	114

Protein Engineering



Mutant	Kcat/Km (min^{-1} nM^{-1})	
	Glc	Xyl
wt	5.8	97.2
Trp 139 - Tyr	6	3.2
Trp 139 - Phe	1.5	13.8
Val 186 - Thr	9.7	55.4
Val 186 - Ser	5.7	15.9
Trp 139 - Phe / Val 186 - Thr	32.9	21.6
Trp 139 - Phe / Val 186 - Ser	12.4	4.0

Protein Engineering