Predicting compensatory mutations in proteins using statistical models

Marion Chauveau, 3^{rd} year Phd student

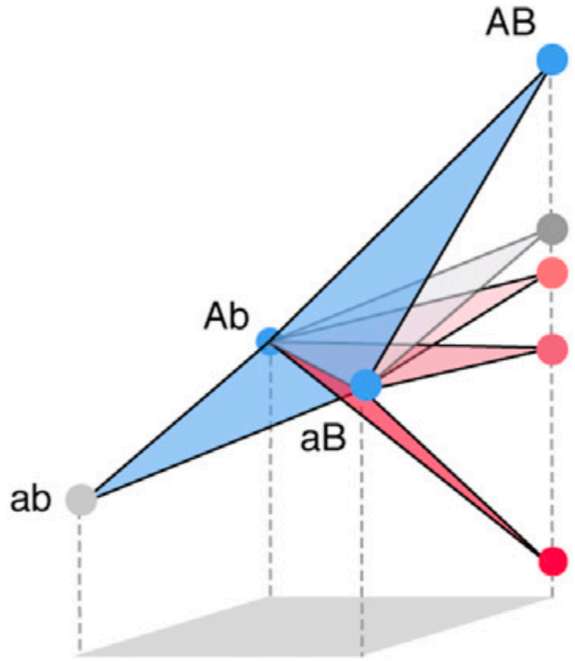
Supervisors: Ivan Junier & Olivier Rivoire

20/02/2025
BiGre Days

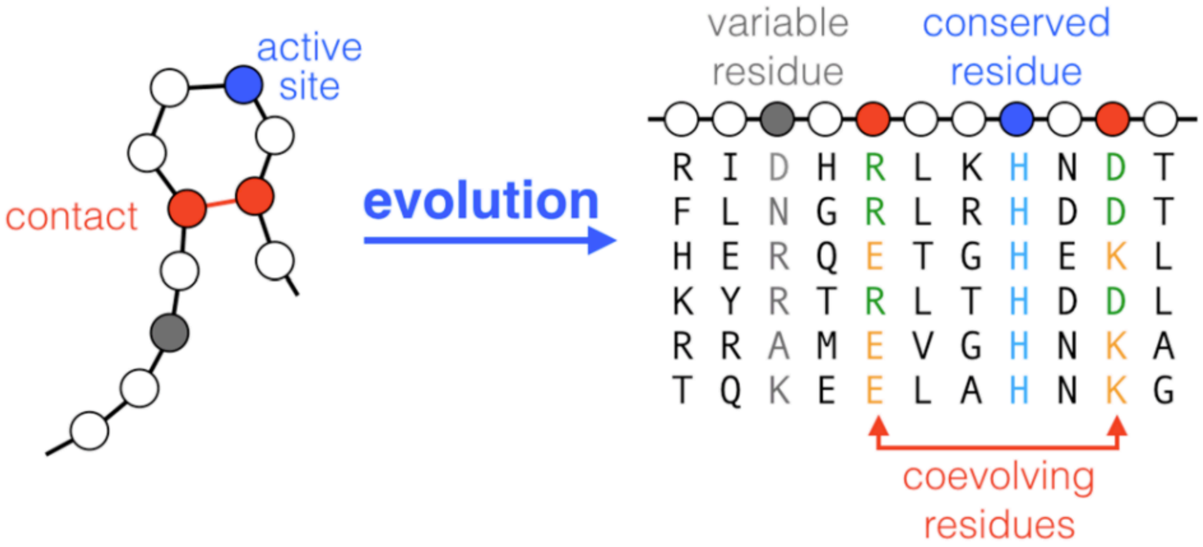
UMR 7083 Univ River ESPCI PARIS TIMC UGA Université Grenoble Alpes

Epistasis: Mutations are **context dependent** and can interact in **non-additive** ways

It can **change the tolerance of homologs to mutations** and induce **correlations in sequence alignments**



Miton, C. M., *Current opinion in structural biology*, **2021**

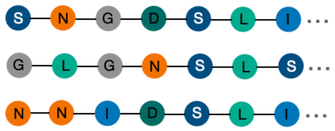


Cocco, S., *Reports on Progress in Physics*, **2018**

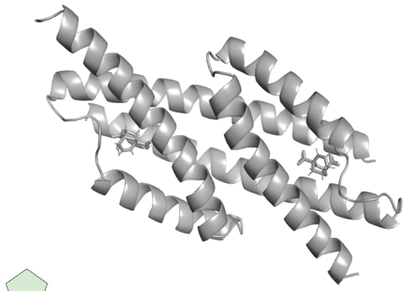
Weinreich, D. M., *Evolution*, **2005**
Cocco, S., *Reports on Progress in Physics*, **2018**
Miton, C. M., *Current opinion in structural biology*, **2021**

Homologous sequences

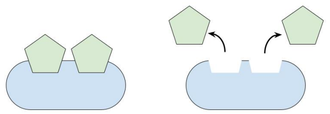
Amino acid chains



3d structure

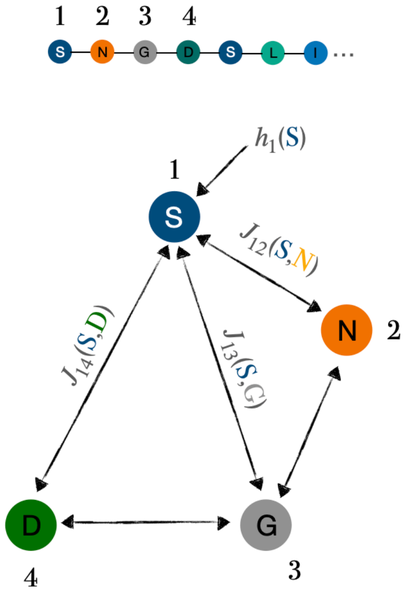


Function



Catalyse a specific reaction

Potts Model



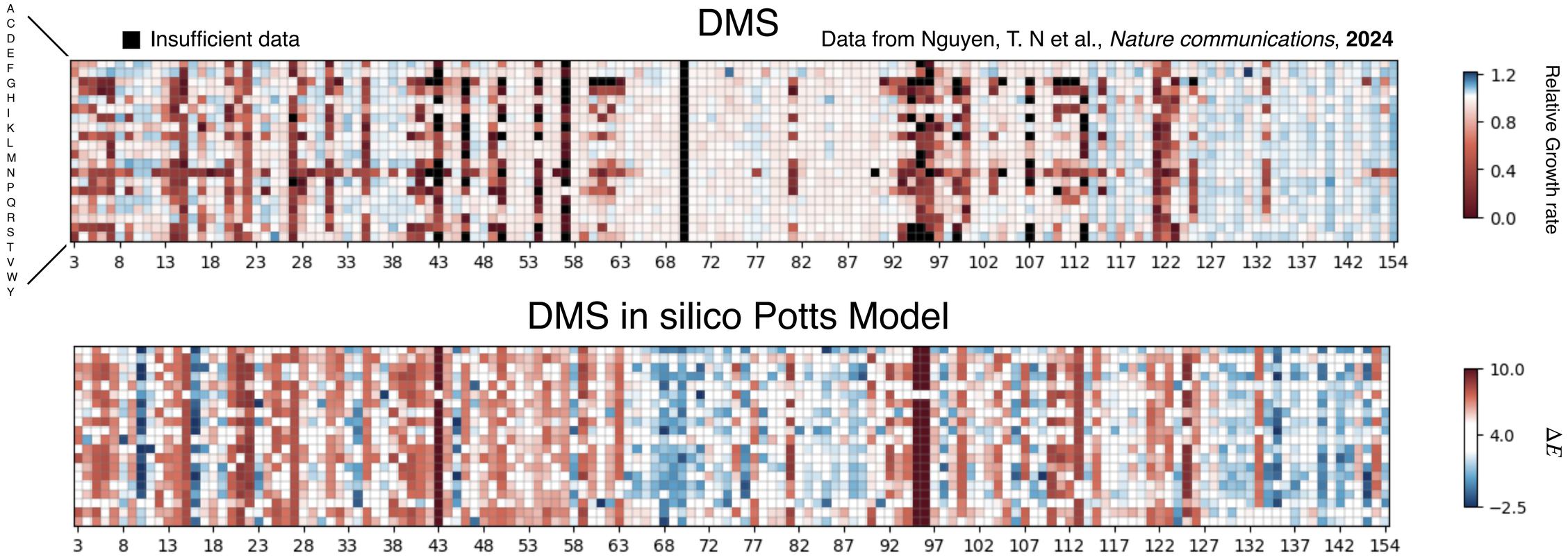
$$p(\{\sigma_i\}_{i=1,\dots,L}) = \frac{e^{-E(\boldsymbol{h},\boldsymbol{J})}}{Z(\boldsymbol{h},\boldsymbol{J})}$$

- Inferred from a MSA
- Trained to capture empirical **frequencies** & **pairwise frequencies**

By training a Potts Model on a protein family we obtain **Boltzmann distribution** with a **statistical energy**

Morcos F et al., *Proceedings of the National Academy of Sciences*, **2011**
Marks DS et al., *Nature biotechnology*, **2012**
Bitbol AF et al., *Proceedings of the National Academy of Sciences* **2016**
Hopf TA et al., *Nature biotechnology* **2017**
Russ WP et al., *Science* **2020**

DHFR is an essential enzyme involved in folate metabolism, which is **crucial for bacterial growth**



Bisardi M et al., *Molecular biology and evolution*, **2022**
RM Levy et al., *Current opinion in structural biology*, **2017**
TA Hopf et al., *Nature biotechnology*, **2017**
M Figliuzzi et al., *Molecular biology and evolution*, **2016**

Can we use this statistical energy to predict **compensatory mutations**?

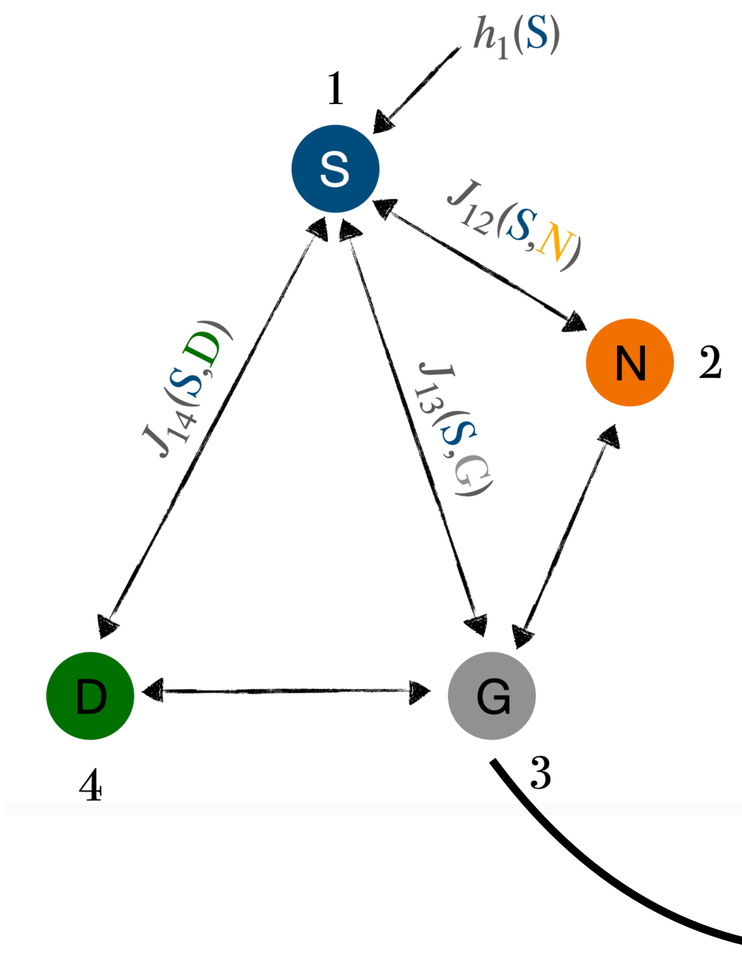
$$p(\{\sigma_i\}_{i=1,\ldots,L}) = \frac{e^{-E(\boldsymbol{h},\boldsymbol{J})}}{Z(\boldsymbol{h},\boldsymbol{J})}$$

$$\Delta\Delta E = \Delta E_{D3} - \Delta E_{G3}$$

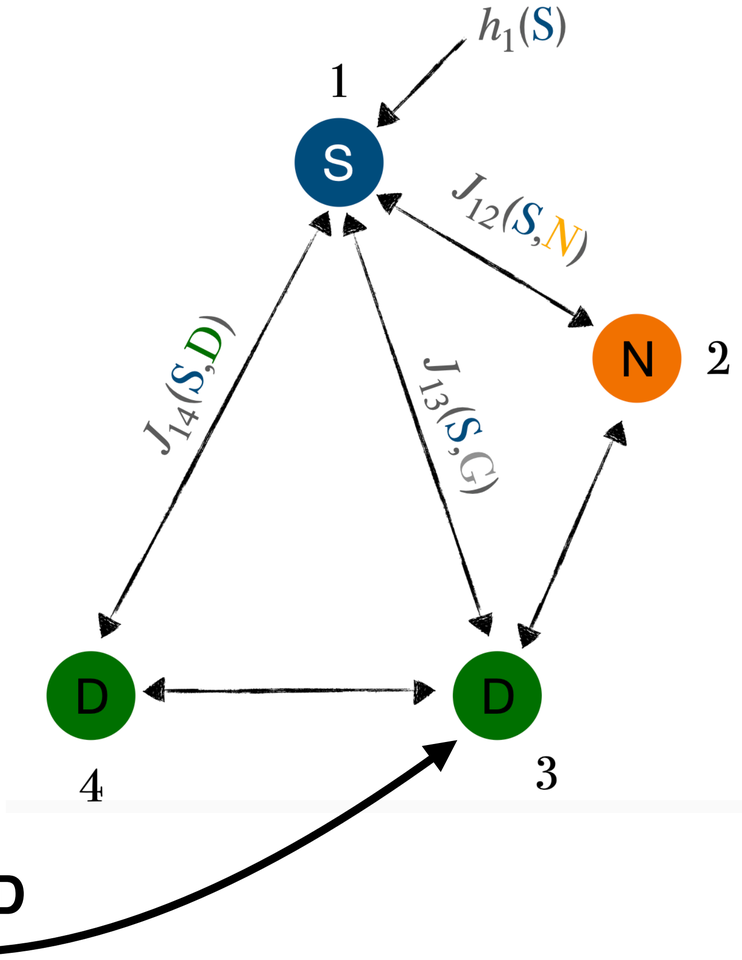
We want a mutation for which $\Delta\Delta E$ **is minimal**

→ *Epistasis*

Wild type sequence



Mutant G3D



G3D

Mutation on G121 decrease the hybride transfer rate constant in *E. Coli* **Dihydrofolate Reductase** (DHFR)

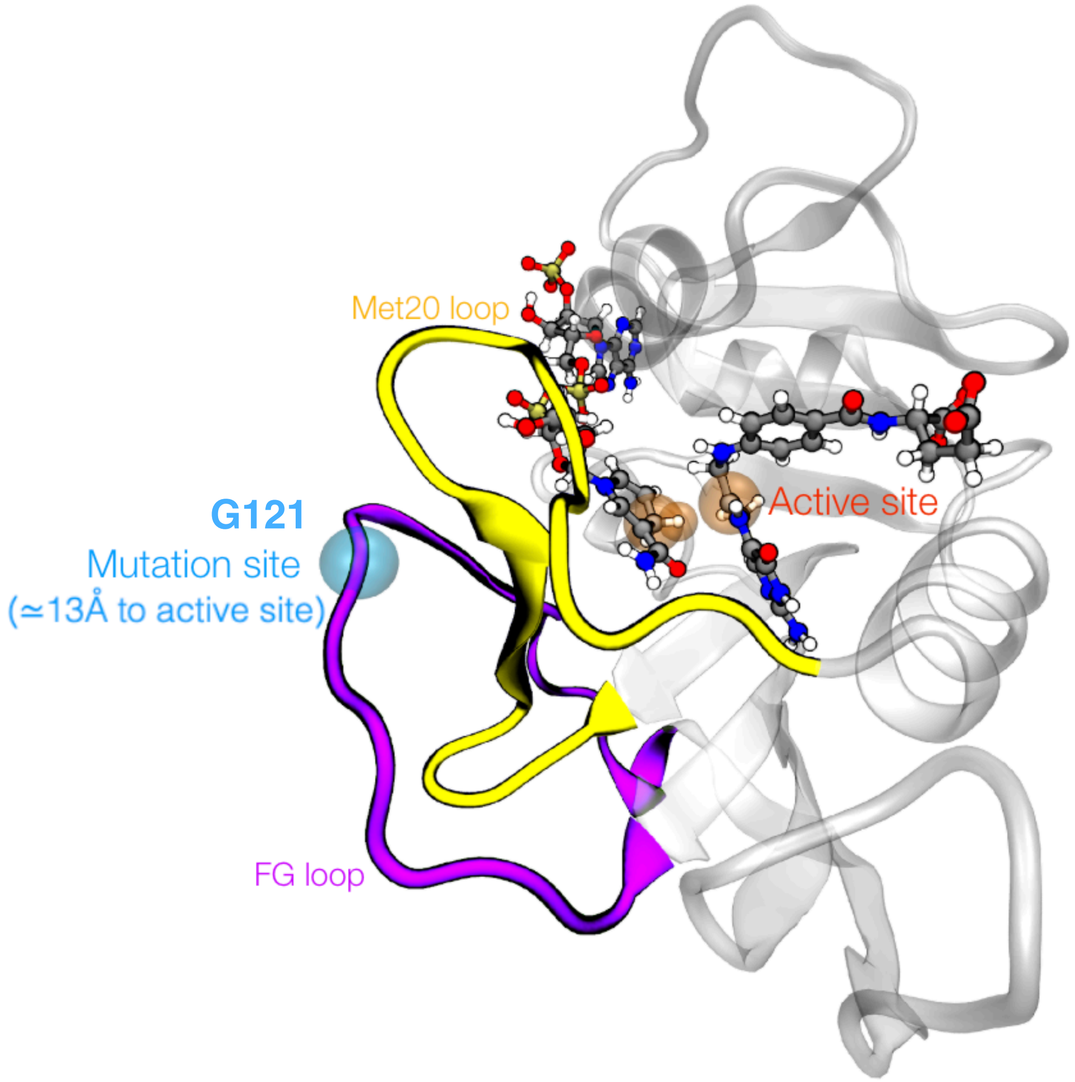


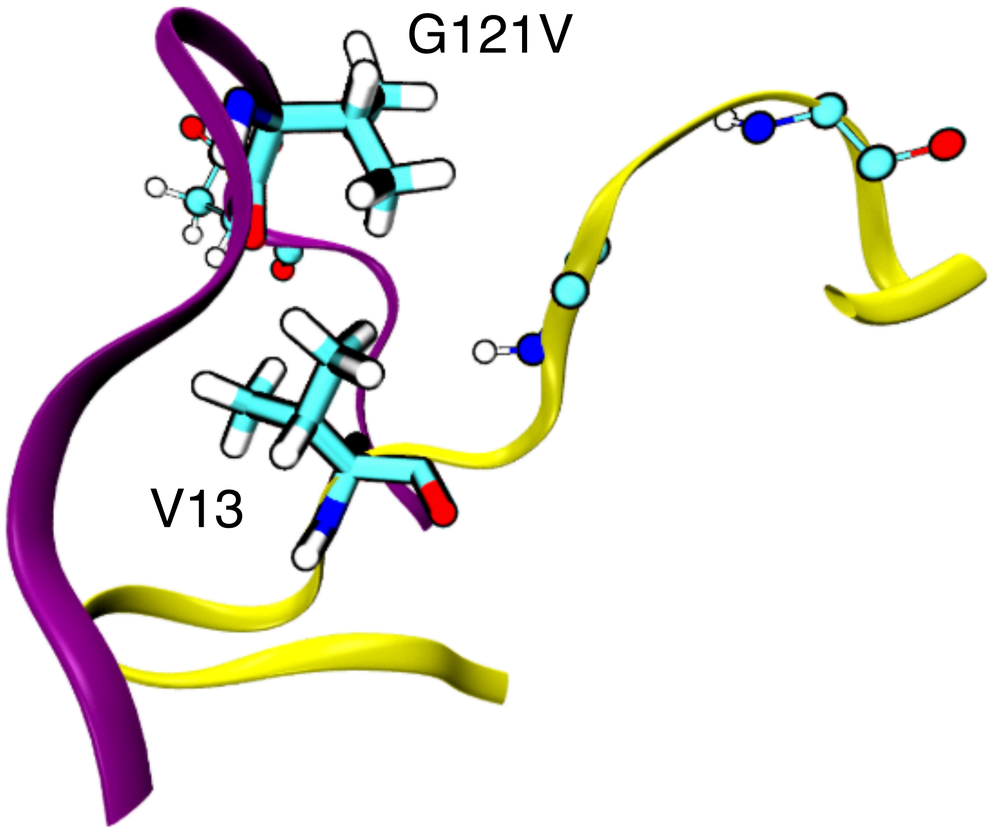
Table 3: Rate Constants (s^{-1}) for the Conformational Change and Hydride Transfer for Wild-Type and Mutant DHFRs at 25 °C in MTEN Buffer at pH 7.0

	enzyme	k_{conf}[a]	k_{hyd}[b]	
Size ↓	Gly-121 (wild-type)	nd[c]	220	↑ Rate
	Ala-121	nd	38.5	
	Val-121	3.5	1.4	

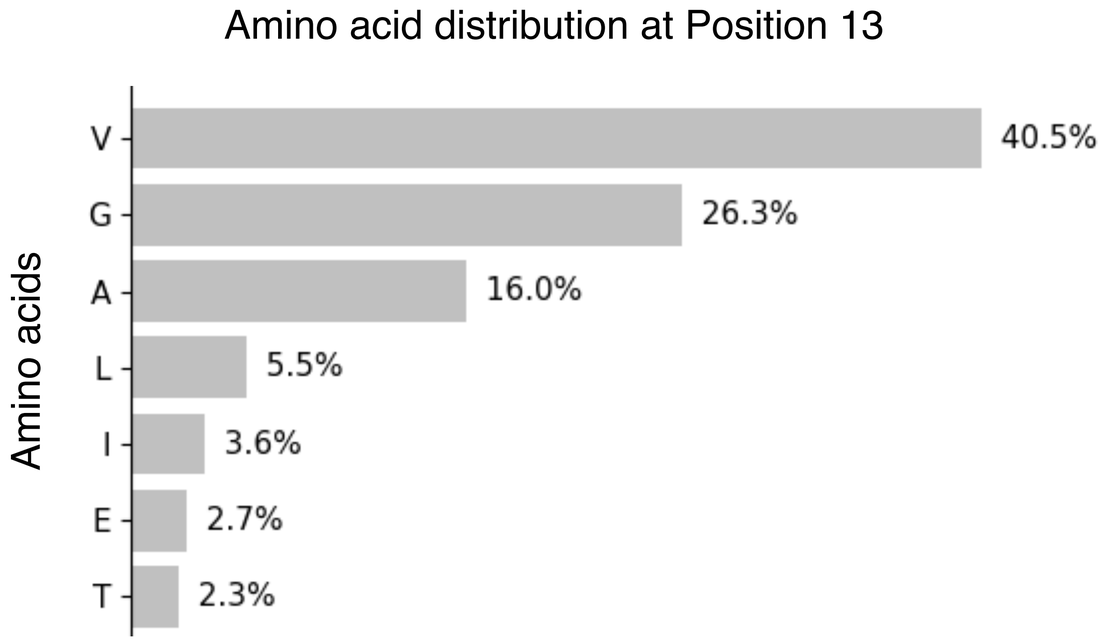
Cameron, Benkovic, *Biochemistry* **1997** 36(50):15792-800

$$\Delta\Delta E = \Delta E_{V121} - \Delta E_{G121}$$

Mutant G121V

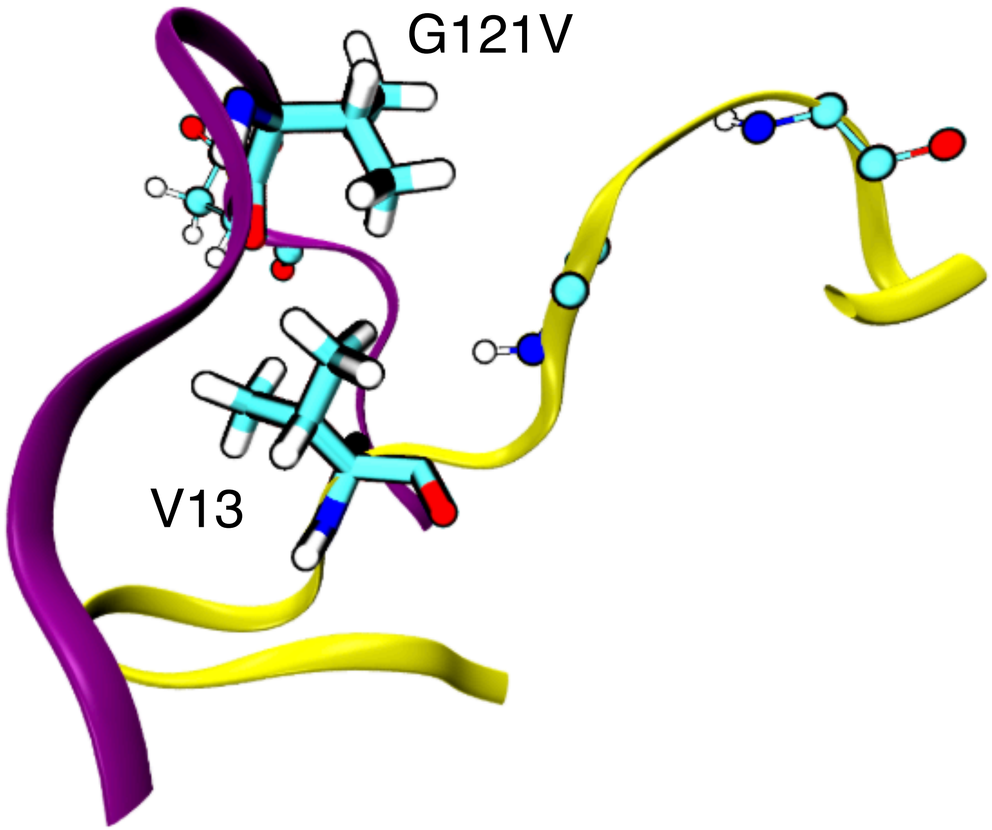


Mutation **V13G** has the smallest $\Delta\Delta E$

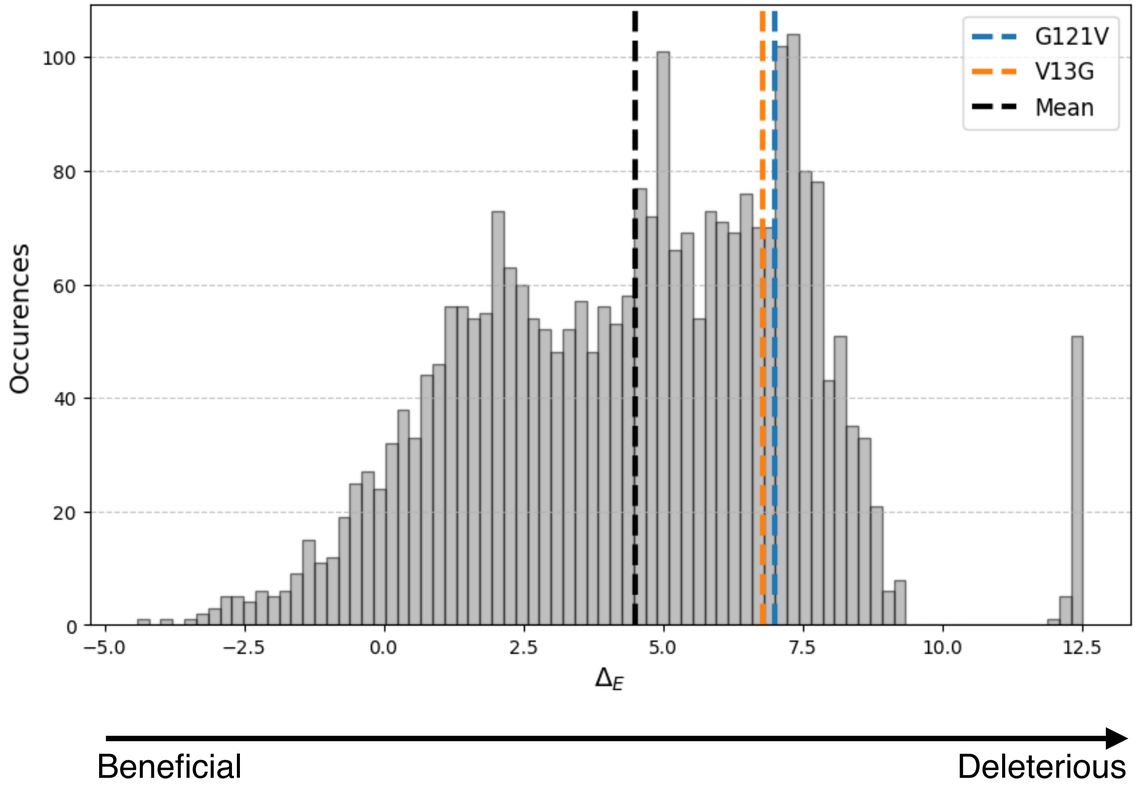


$$\Delta\Delta E = \Delta E_{V121} - \Delta E_{G121}$$

Mutant G121V

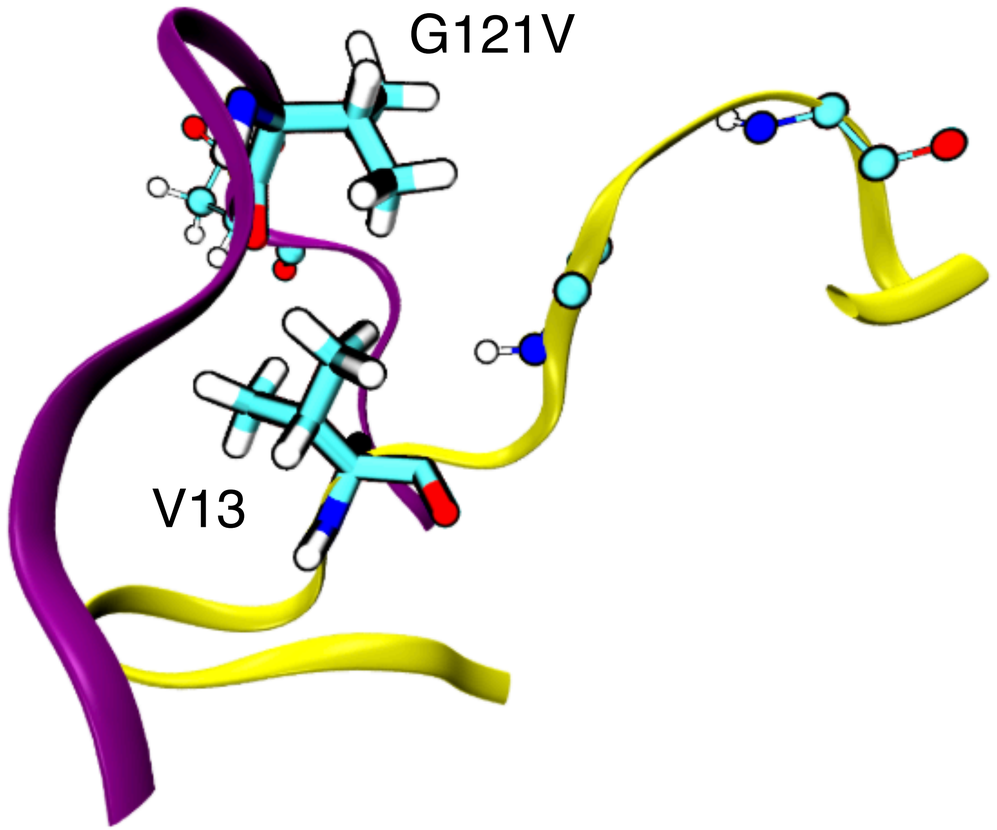


Mutation **V13G** has the smallest $\Delta\Delta E$



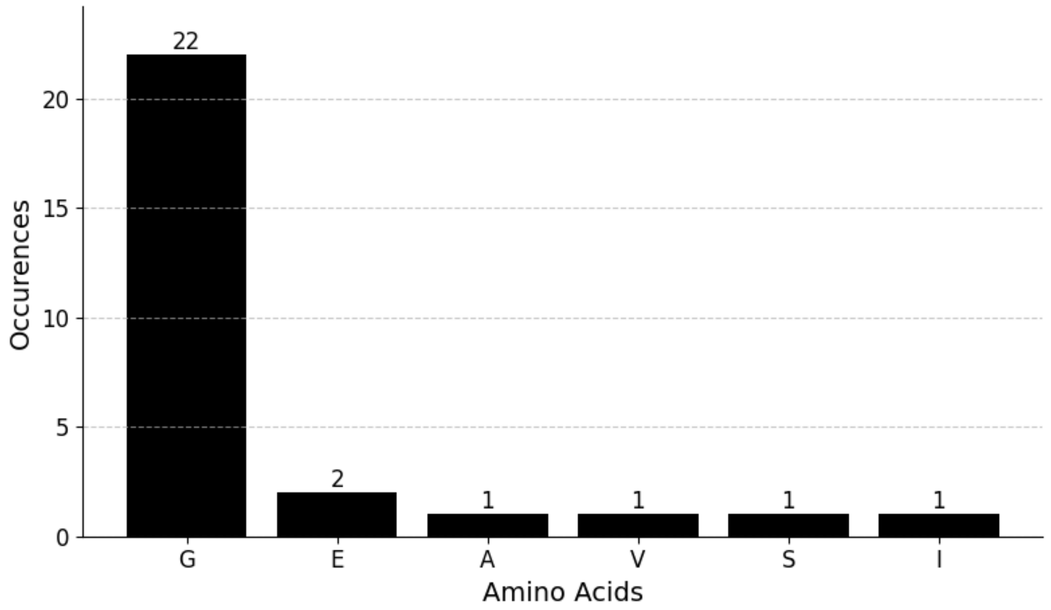
$$\Delta\Delta E = \Delta E_{V121} - \Delta E_{G121}$$

Mutant G121V



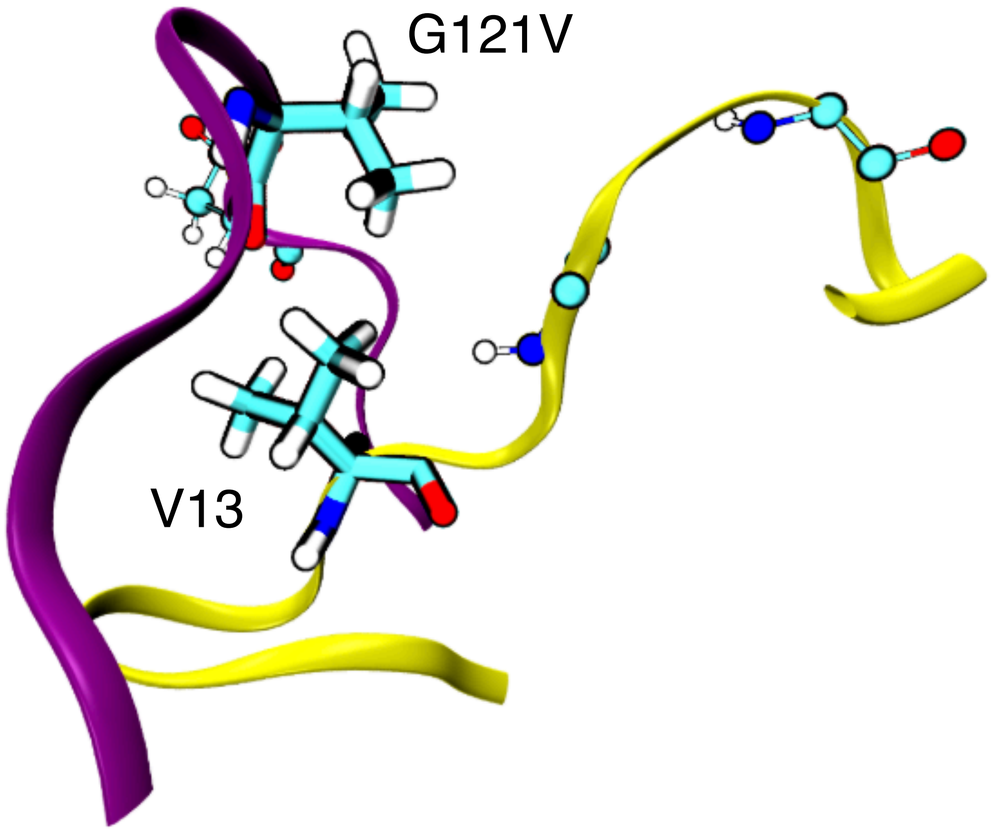
Mutation **V13G** has the smallest $\Delta\Delta E$

Amino acid at position 13 for sequences with V121

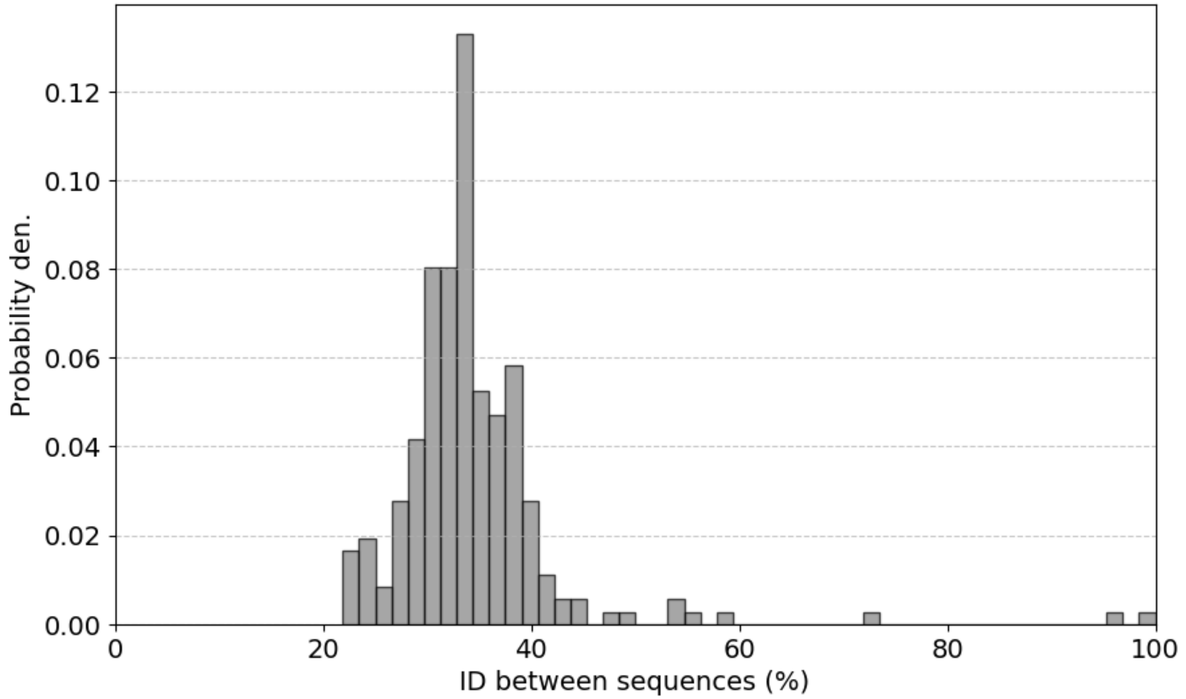


$$\Delta\Delta E = \Delta E_{V121} - \Delta E_{G121}$$

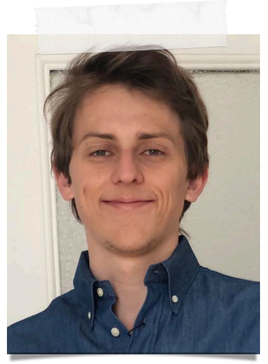
Mutant G121V



Mutation **V13G** has the smallest $\Delta\Delta E$

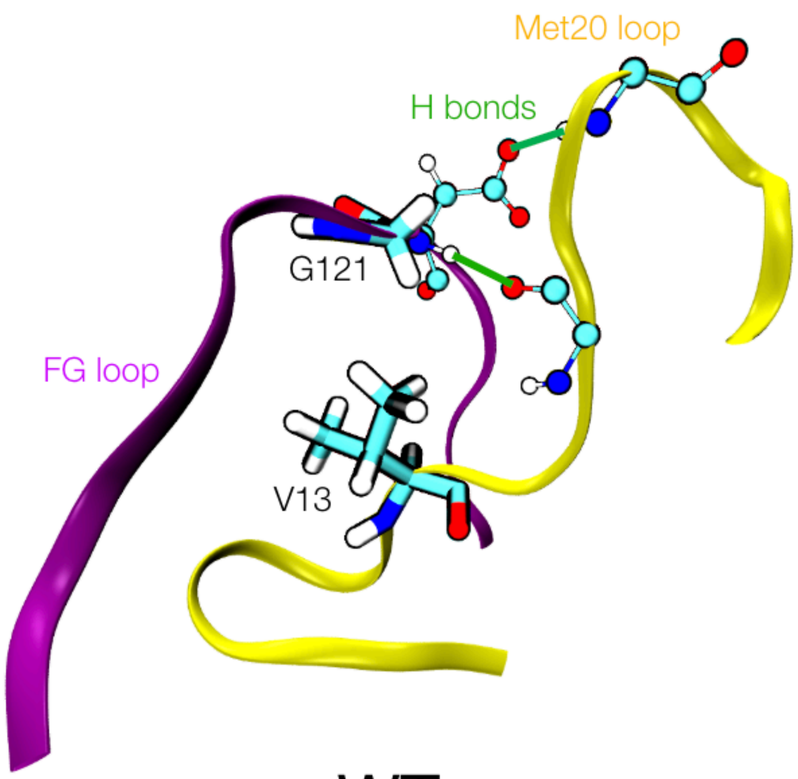


Mutation V13G seems to **compensate for the steric hindrance** introduced by the G121V mutation between the FG and Met20 loops

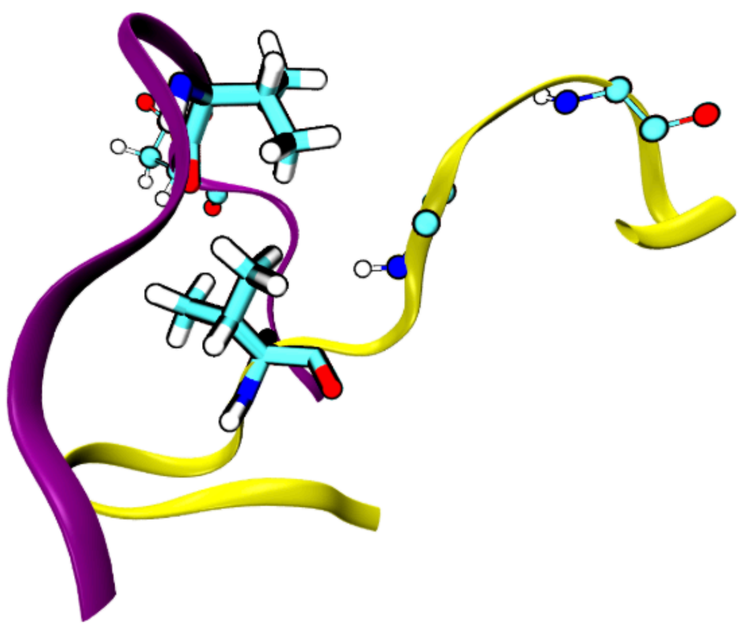


Paul Guenon

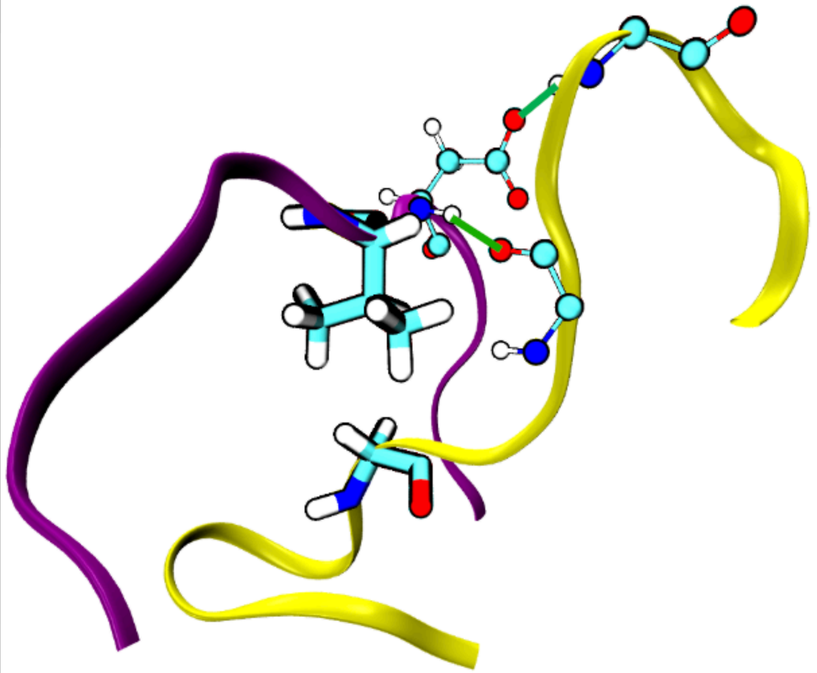
Supervised by Damien Laage & Guillaume Stirnemann



WT



G121V



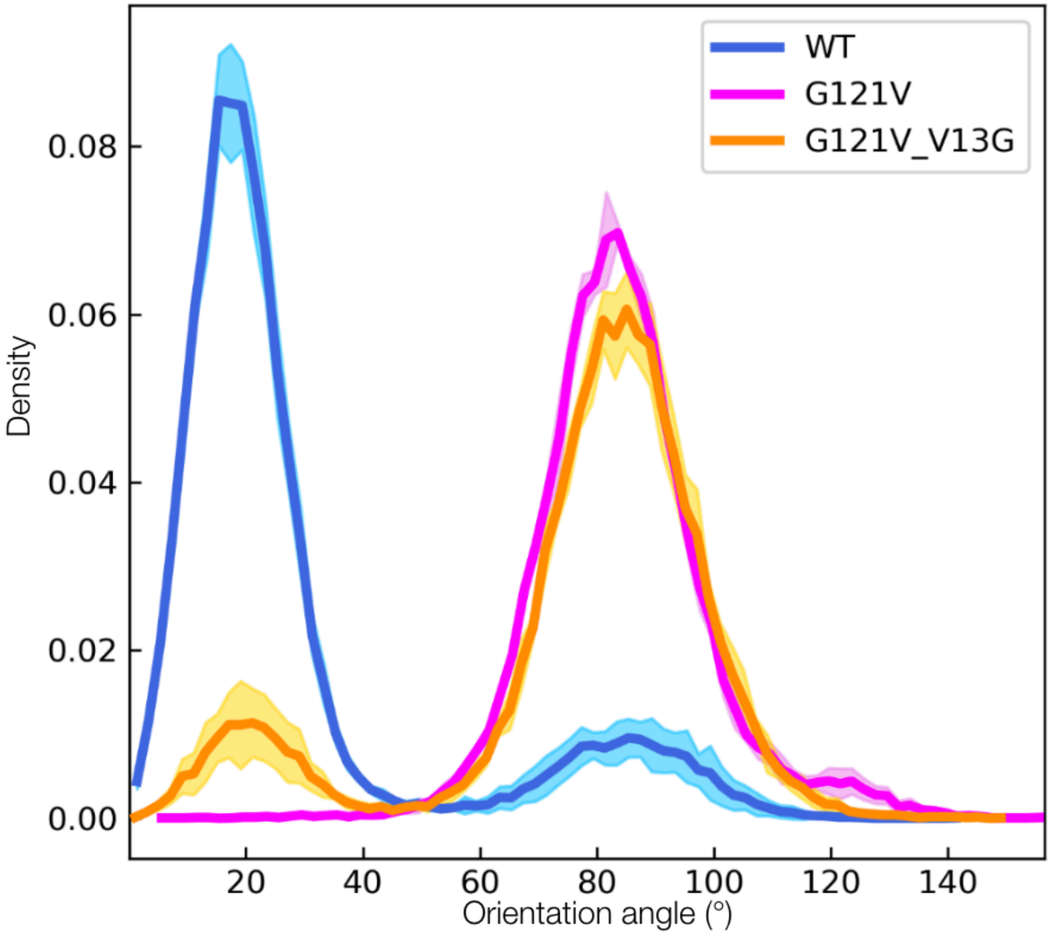
G121V_V13G

Molecular dynamics simulations predicts a **tenfold increase in the hybrid transfer rate** for the mutant **G121V_V13G**

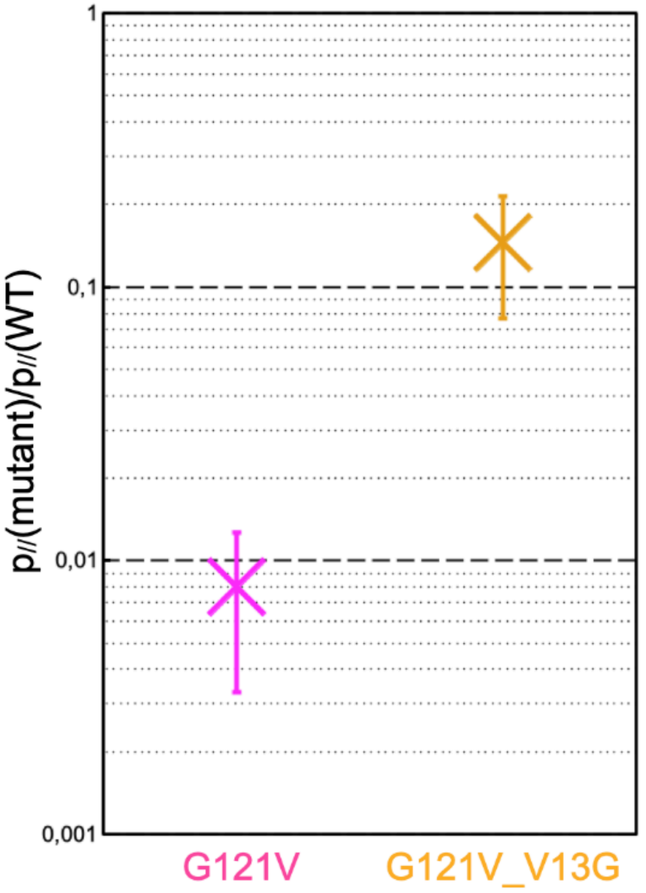
Paul Guenon

Supervised by Damien Laage & Guillaume Stirnemann

Distribution of the orientation of the substrates, from 400ns REMD simulations

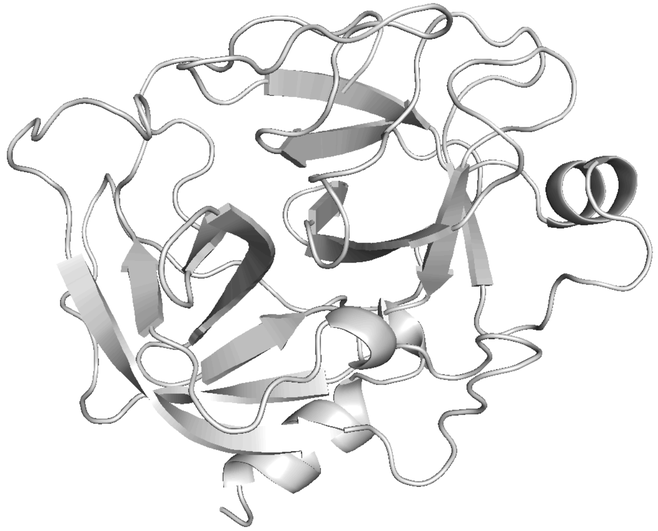


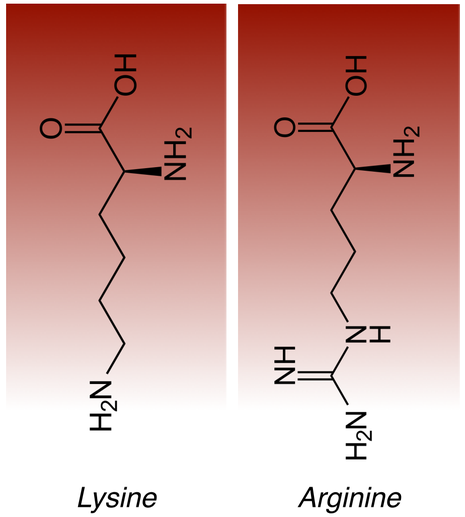
$$\frac{p_{\parallel}(\text{mutant})}{p_{\parallel}(\text{WT})} \approx \frac{k_{hyd}(\text{mutant})}{k_{hyd}(\text{WT})}$$



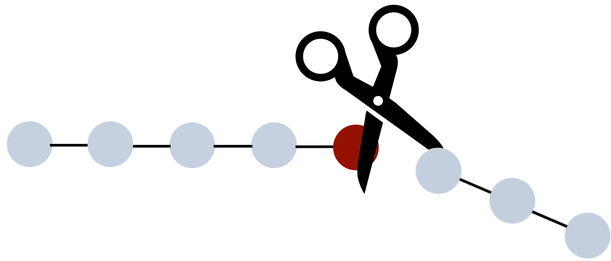
Serine Proteases are enzymes which catalysis the hydrolysis of peptide bonds with varying specificity profiles

Trypsin

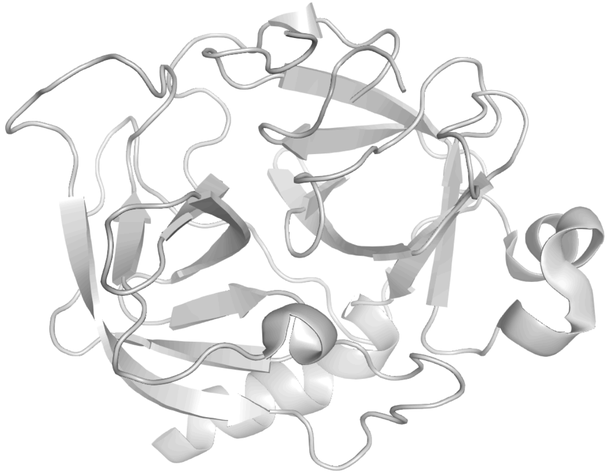


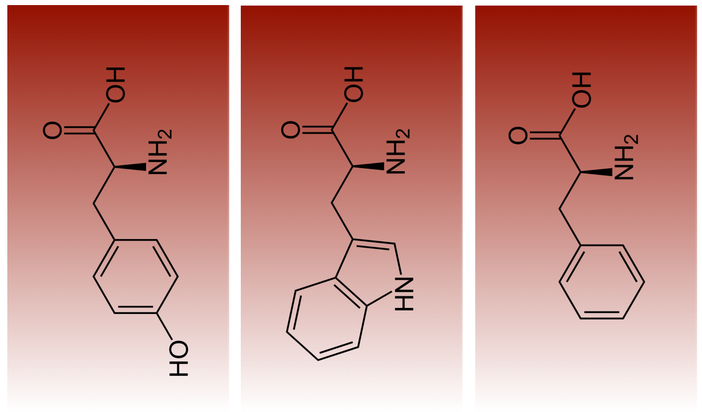


Lysine *Arginine*

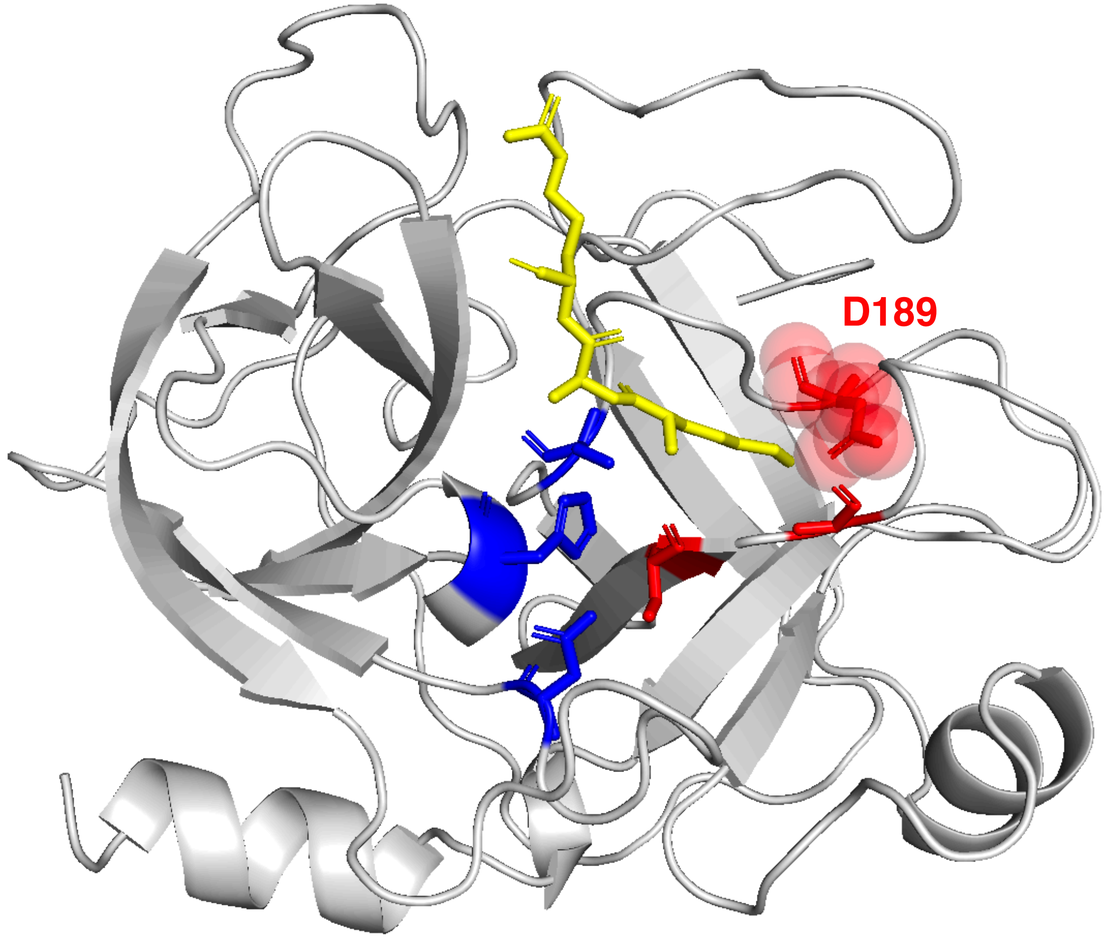


Chymotrypsin





Tyrosine *Tryptophan* *Phenylananine*



Inhibitor protein

Catalytic triad

Linking pocket

$$\Delta\Delta E = \Delta E_{S189} - \Delta E_{D189}$$

Enzyme	k_{cat}/K_m (M^{-1} s^{-1})					Specificity
	Phe	Tyr	Trp	Leu	Lys	
Ch	1.6×10^6	4.5×10^6	6.8×10^6	1.2×10^5	850	10^2 to 10^3
Tr	4.5*	1.8*	0.2*	0.2*	1.2×10^6†	10^{-6} to 10^{-7}
D189S	33*	150*	2.3*	4.7*	16†	0.1 to 10

Hedstrom et al., *Science* **1992**

Residue 189 is crucial for the specificity of the serine protease but swapping this residue is **not sufficient** to get a specificity conversion

Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

D189 Hedstrom et al.

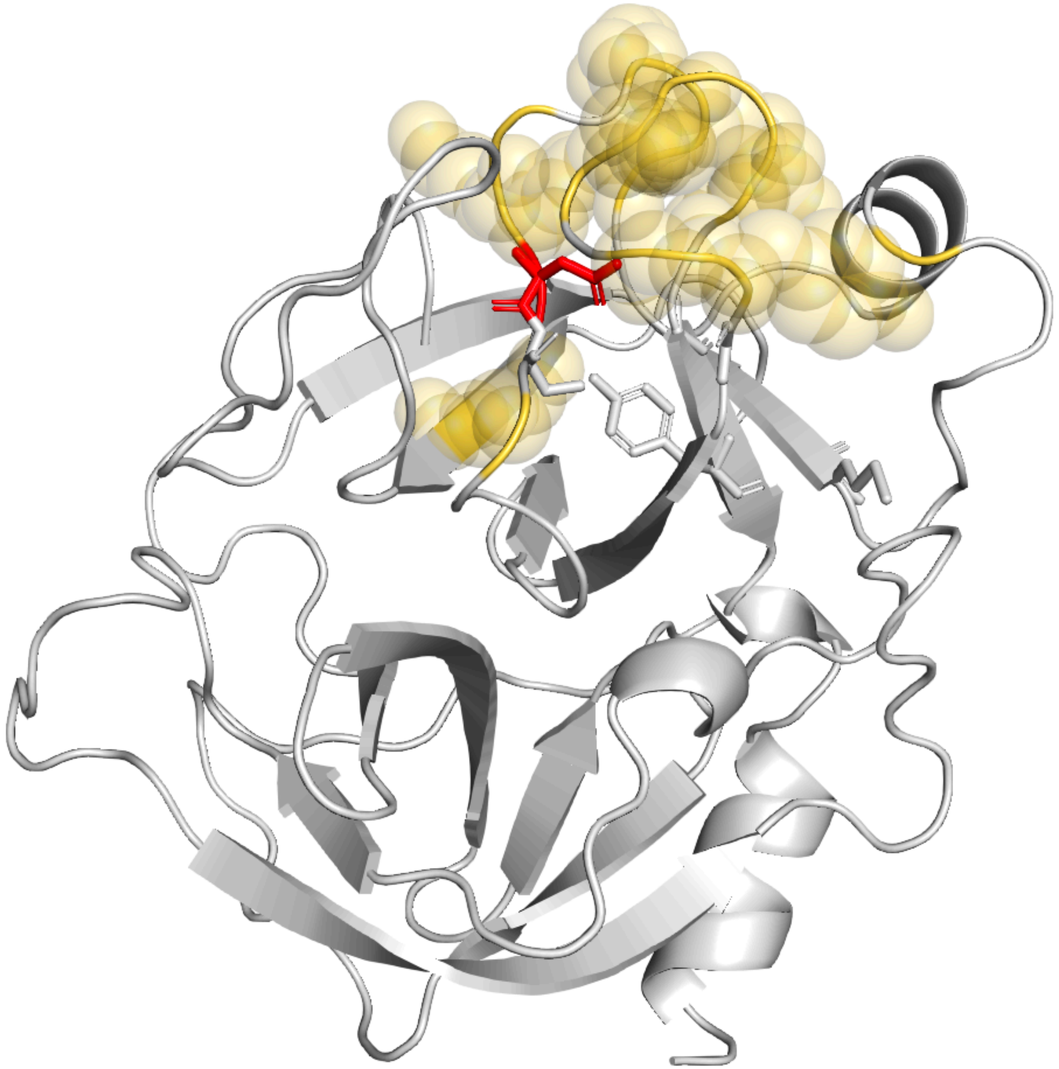


Table 1: Specificity of Chymotrypsin, Trypsin, and Trypsin Mutants[a]

enzyme	Suc-Ala-Ala-Pro-X-AMC k_{cat}/K_m (M^{-1} s^{-1}), X = Phe	Tyr	Trp	Lys	Phe:Tyr:Trp:Lys
chymotrypsin	$(4.9 \pm 0.5) \times 10^5$	$(8.6 \pm 0.2) \times 10^5$	$(1.3 \pm 0.1) \times 10^6$	69 ± 1	$1:2:3:10^{-4}$
trypsin	6.0 ± 0.2	8.9 ± 0.3	1.48 ± 0.01	$(1.9 \pm 0.1) \times 10^5$	$1:1.5:0.3:10^4$
D189S	38 ± 1	38 ± 8	4.8 ± 3	7.3 ± 0.5	1:1:0.1:0.2
Tr→Ch[S1+L1+L2]	800 ± 20	$(6.1 \pm 0.6) \times 10^3$	$(1.0 \pm 0.1) \times 10^3$	10 ± 1	1:7:1:0.01
Tr→Ch[S1+L1+L2+D189]	1.24 ± 0.01	nd	nd	15.3 ± 0.2	1:nd:nd:13
Tr→Ch[S1 + L1* + L2]	240 ± 10	nd	nd	nd	nd
Tr→Ch[S1 + L1 + L2 + Y172W]	$(9.3 \pm 0.7) \times 10^3$	$(6.1 \pm 0.2) \times 10^4$	$(3.2 \pm 0.3) \times 10^4$	9.0 ± 0.7	$1:6:3:10^{-3}$
Y172W	5.4 ± 0.2	10.1 ± 0.4	≤ 2.7	$(2.5 \pm 0.1) \times 10^3$	1:2:0.5:500

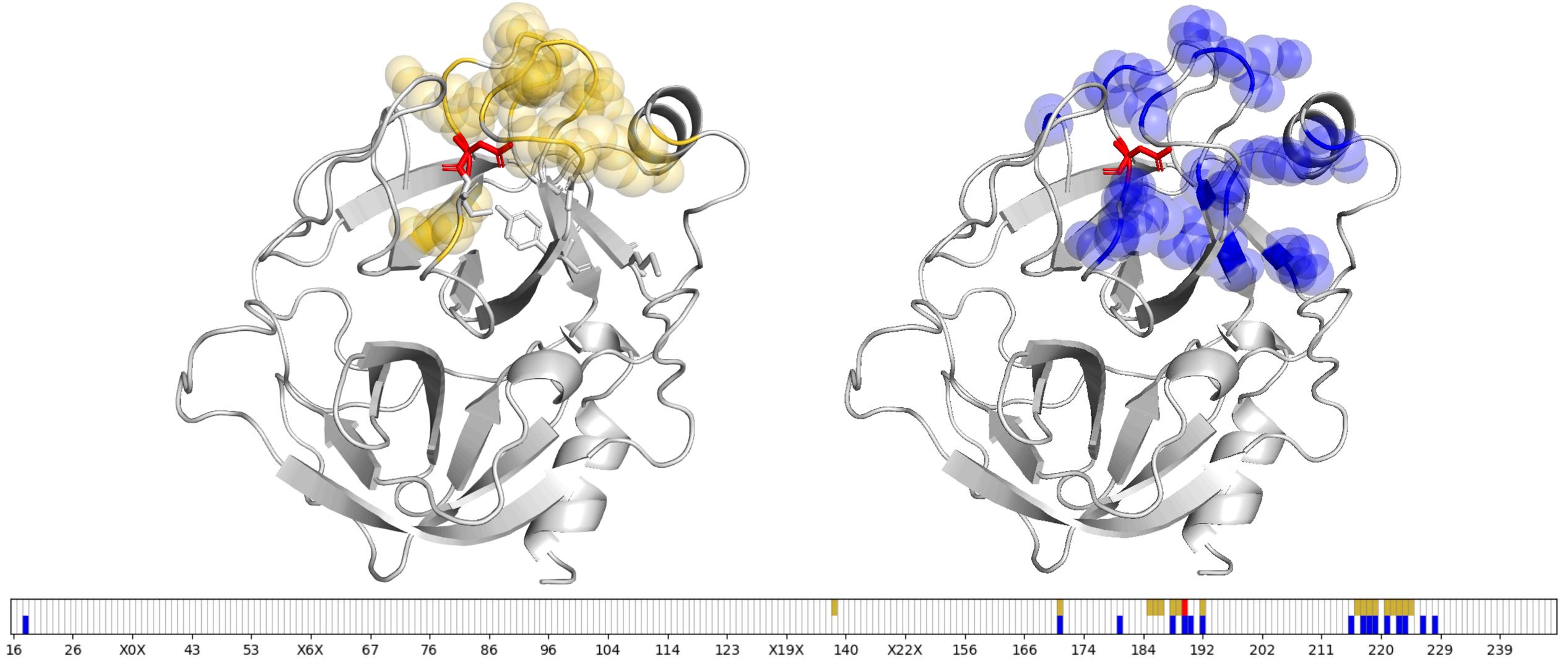
[a] Conditions: 50 mM Hepes, pH 8.0, 100 mM NaCl, and 10 mM $CaCl_2$, 37 °C. Assays were performed at $[S] \ll K_m$, as described in Materials and Methods. Note that the values for chymotrypsin and Tr→Ch[S1 + L1 + L2] are slightly lower (~3-fold) than our previous reports. This discrepancy most likely results from an improvement in instrumentation, resulting in better sensitivity and thermostating. The relative activities of chymotrypsin and Tr→Ch[S1 + L1 + L2] are unchanged. The values for D189S and wild-type trypsin agree with previous reports from the Graf laboratory (Graf et al., 1988). nd, not determined.

Hedstrom et al., *Biochemistry* **1994**

Starting from **D189S** can we use our approach to propose mutations for **specificity conversion** in **Serine Proteases?**

Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

D189 Hedstrom et al. Model predictions



Predicted mutations are consistent with **experimentally tested conversion** from the literature

Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

Conclusion

- Simple procedure that rely on a statistical model trained to capture empirical frequencies & Pairwise frequencies
- Results on **DHFR** are consistent with **Molecular Dynamics** simulations
- Results on **S1A** family are consistent with **experimentally tested conversion** from the literature (Hedstrom et al.)

Upcoming experimental results

- Application to other systems?

Contact: marion.chauveau@univ-grenoble-alpes.fr

Thanks for your attention

Contact: marion.chauveau@univ-grenoble-alpes.fr

Supervisors: Olivier Rivoire & Ivan Junier

CompBio@TrEE