

Dynamism in protein evolution?



Projects

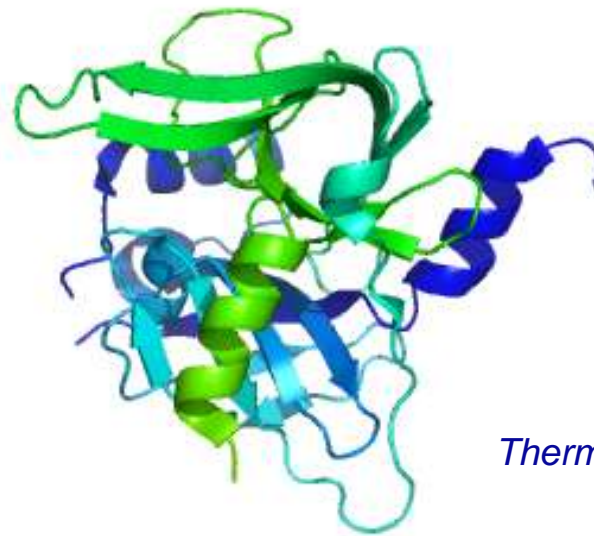
- ID in DNA methyl-transferase evolution
- Dynamic protein-DNA recognition
- Kemp evolution
- Ser -> Thr in AP



Problem



tightly packed proteins



Thermotoga maritima

- ~ 50% destabilizing mutation
- $\geq 80\%$ pathogenic mutations - stability

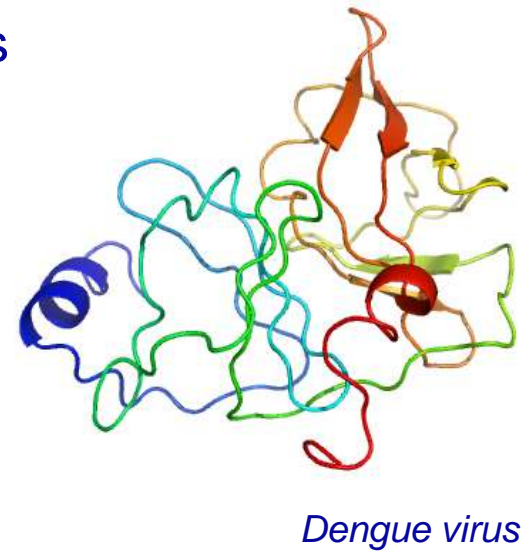
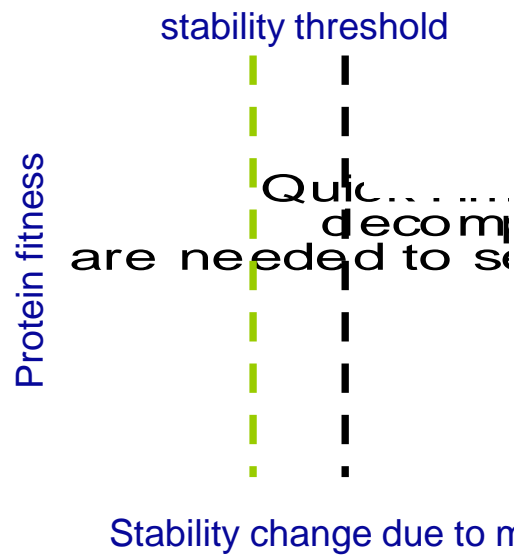
Bershtein (2006), Nature 444, 929.

Tokuriki (2008), PLoS Comp Biol 4, e1000002.

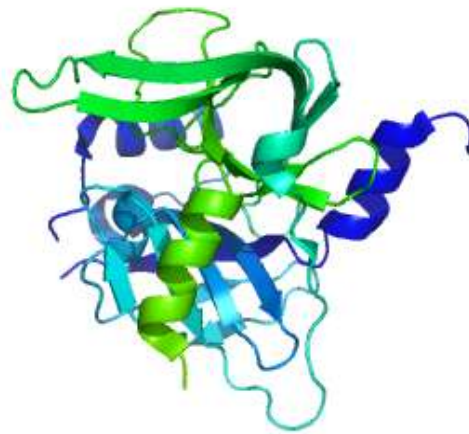
Problem



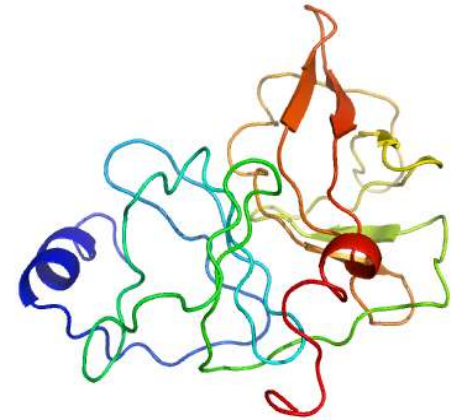
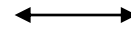
loosely packed proteins



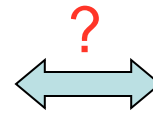
Problem



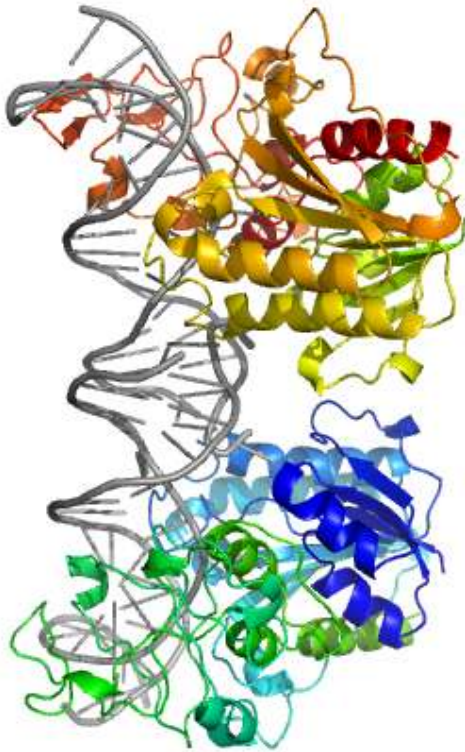
stability



evolvability



Question



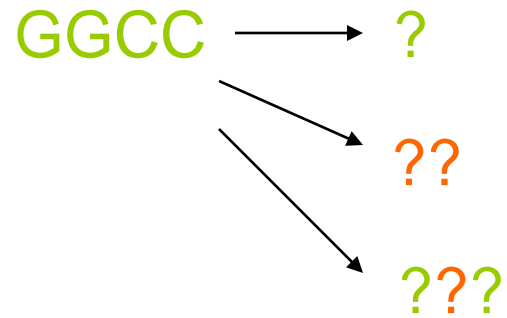
What structural properties facilitate evolution of new functions?



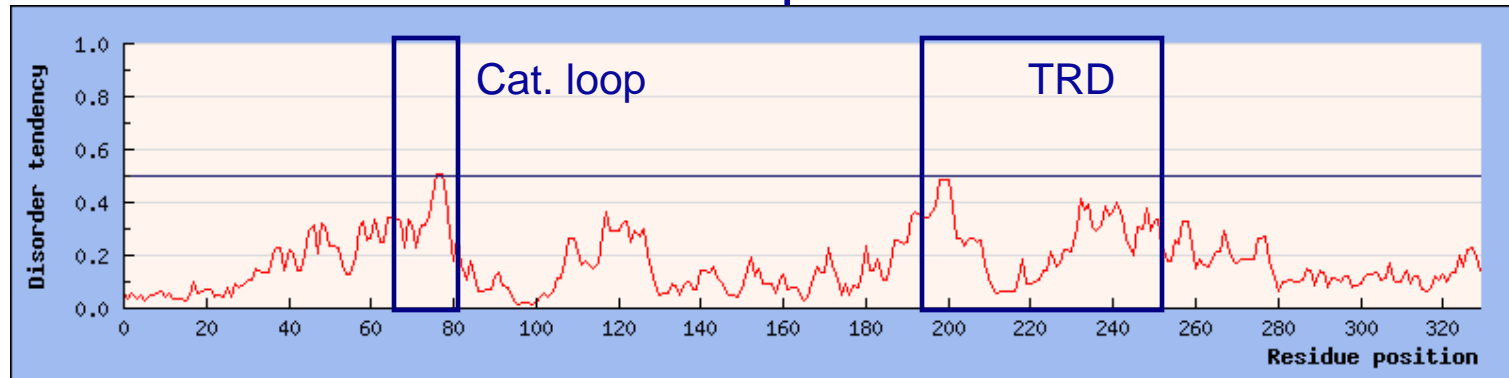
Can ID could be conserved?



MHaellI

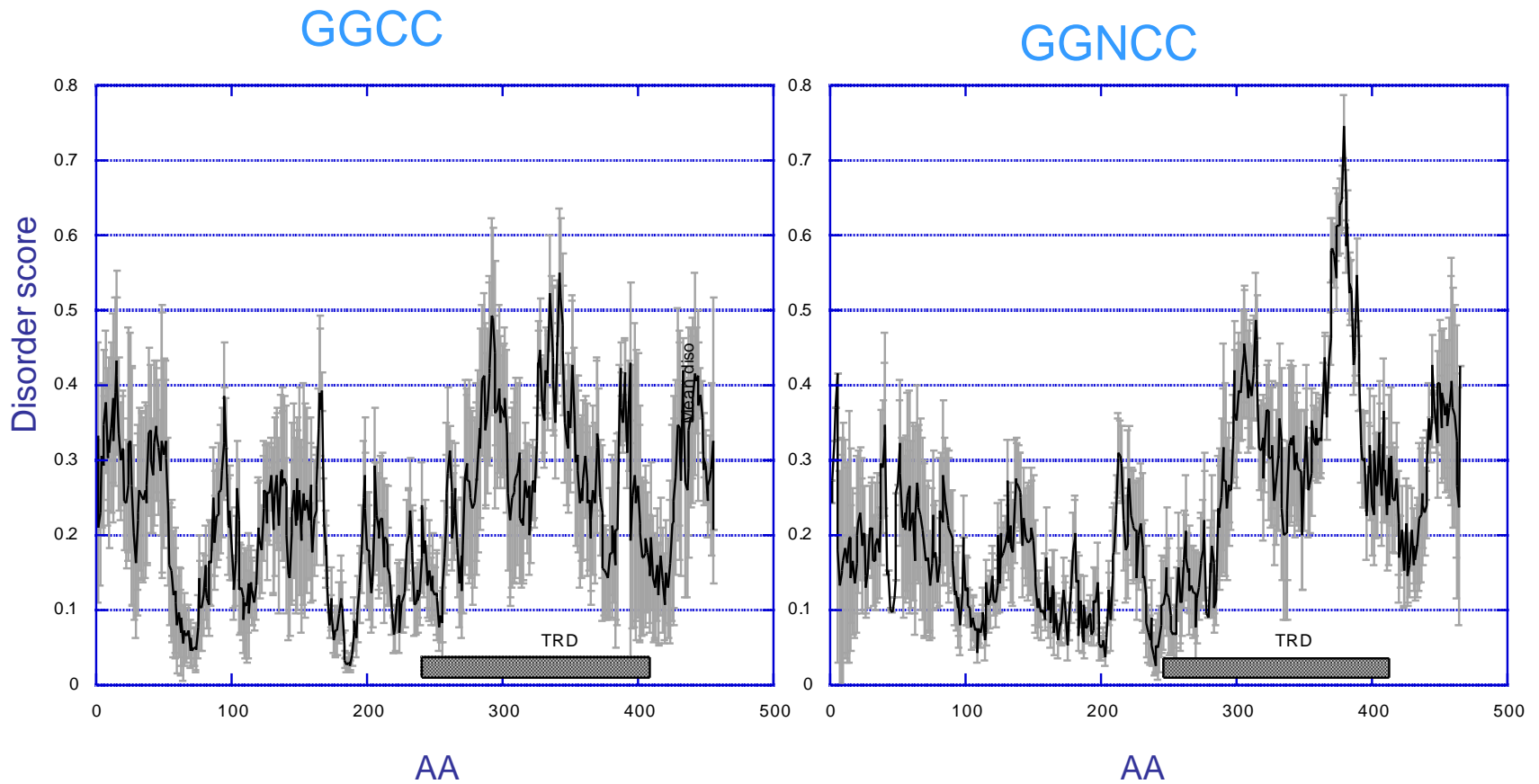


POD : pattern of order-disorder



IUPred

POD changes vs protein function



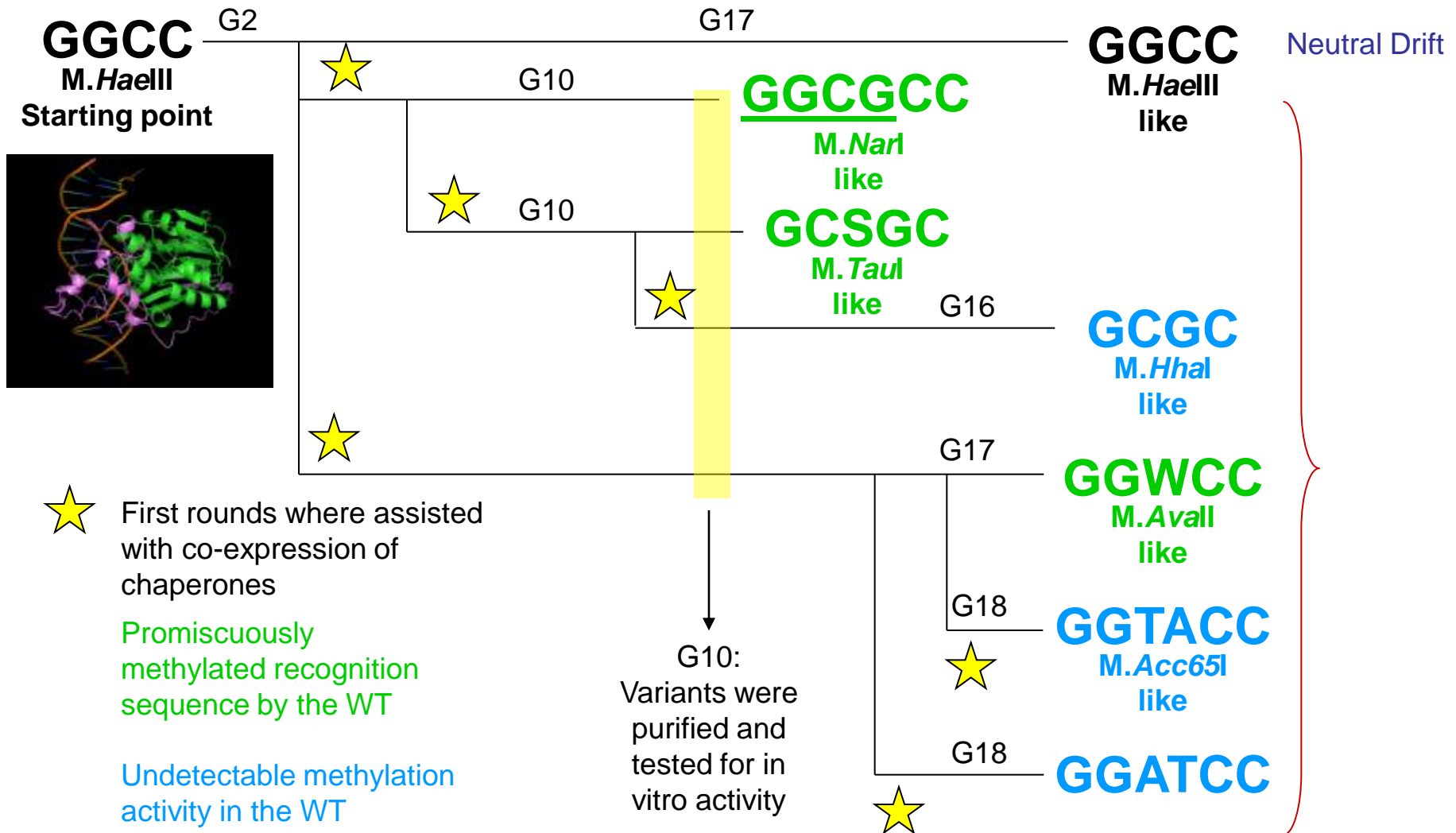
Can intrinsic disorder (ID) be conserved?

- in MHaeIII evolution
- in general?
- which features
- predictability



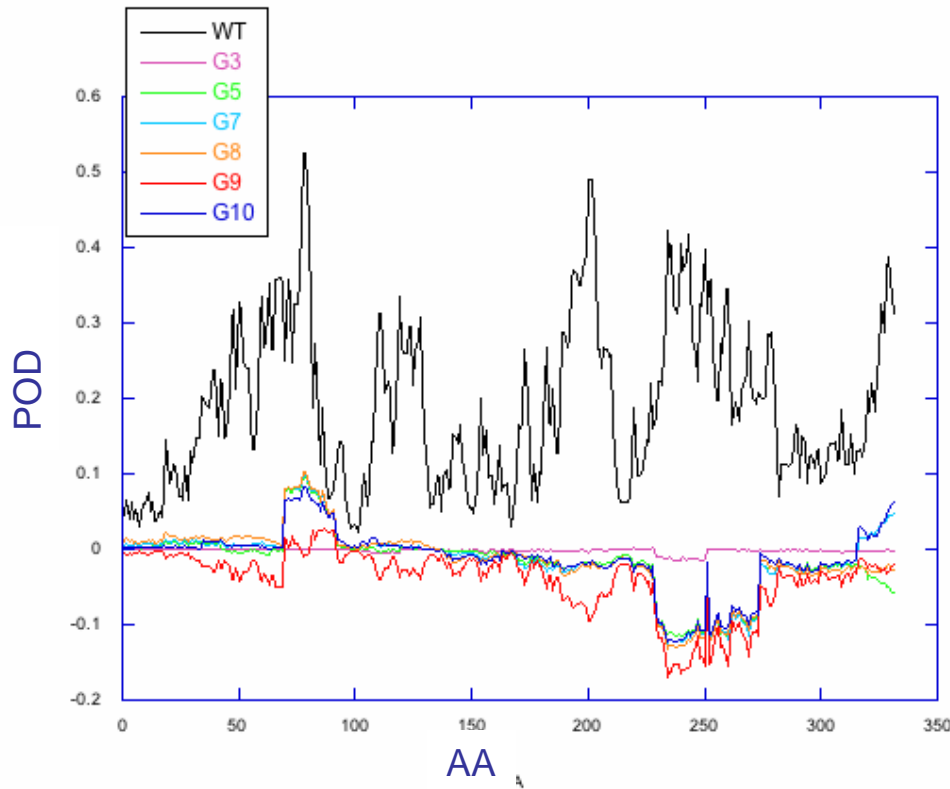
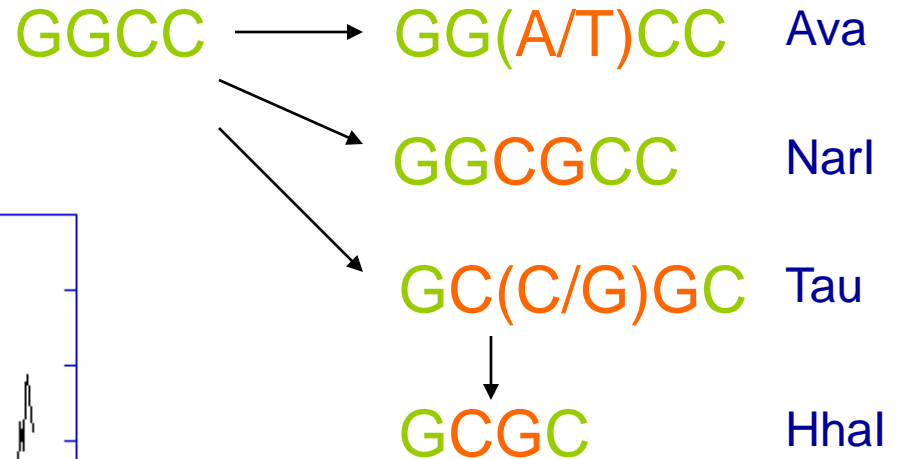
Can ID could be conserved?

in vitro evolution of MHaelll **Liat Rockah**



POD changes along specificity changes

Ava

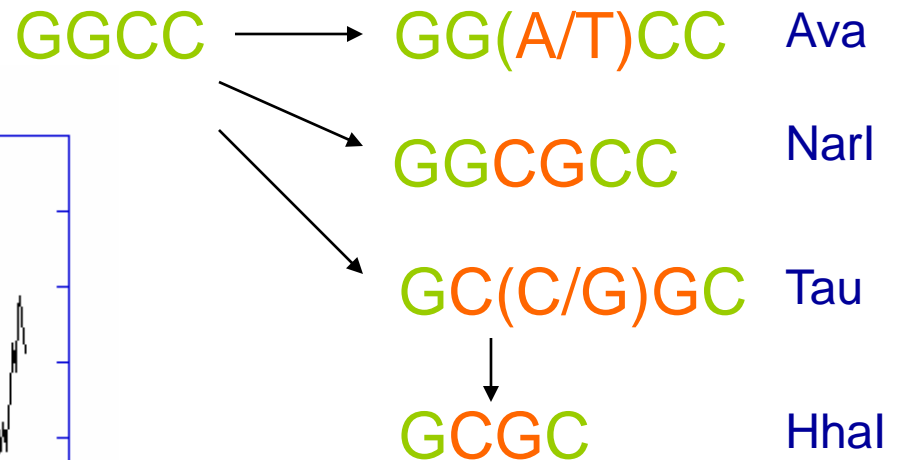
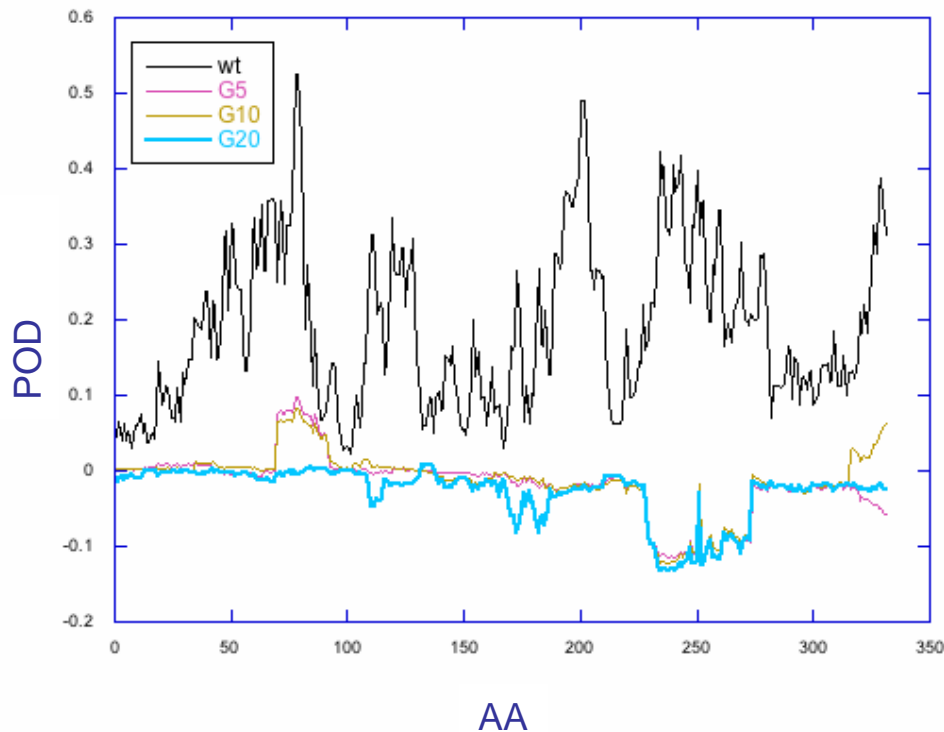


G1 -> G10



POD changes along specificity changes

Ava

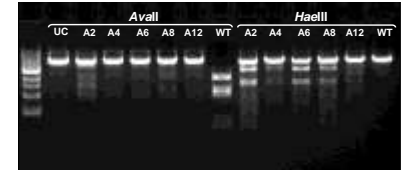


G5 → G20

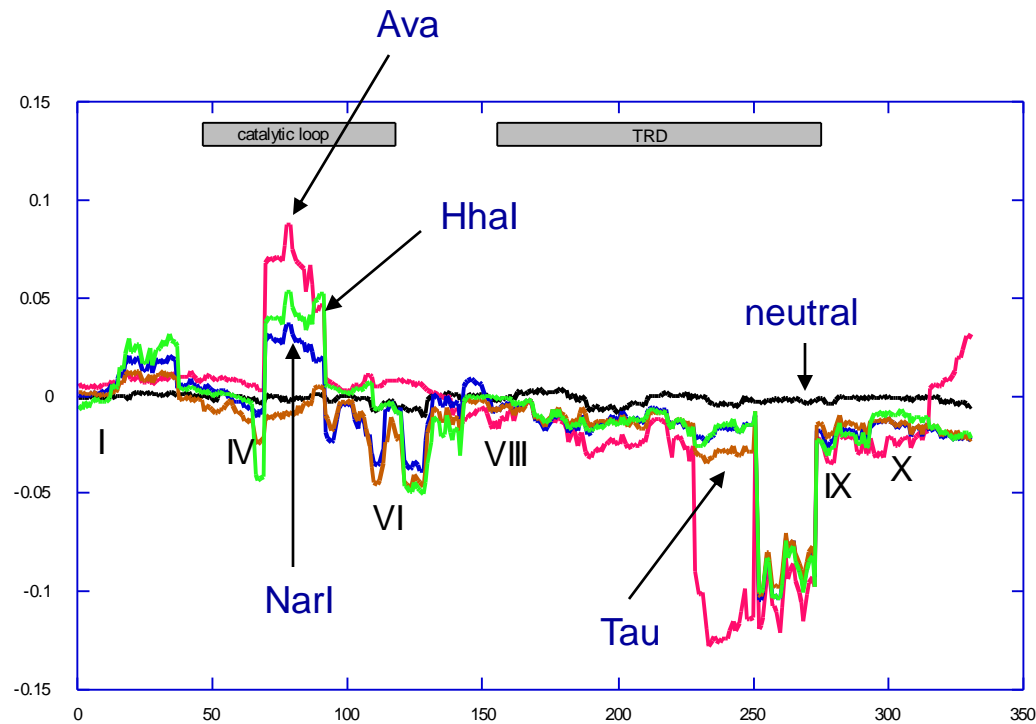


POD changes along specificity changes

Differences?



D

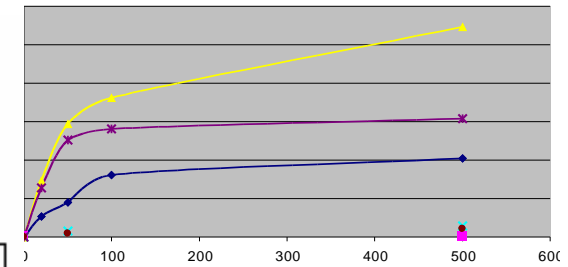
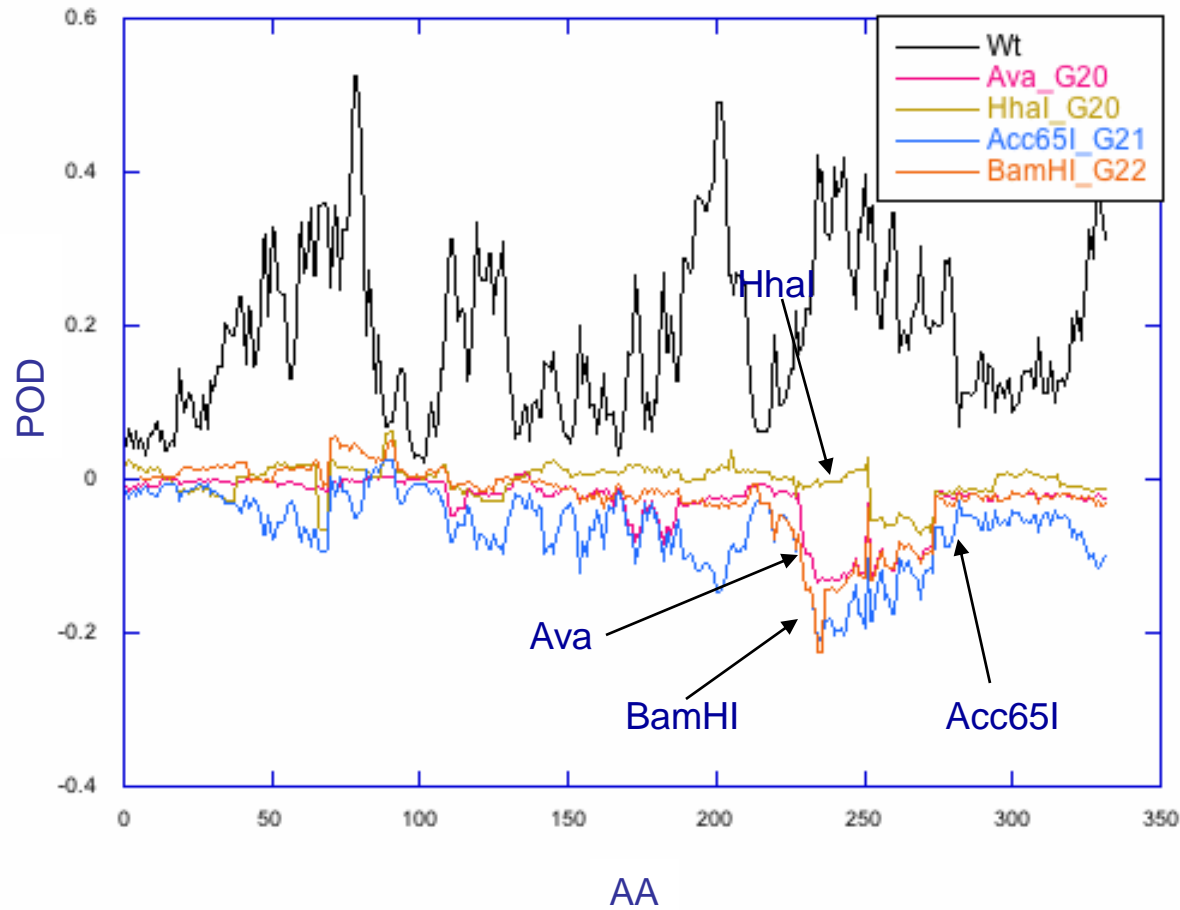


G10

new specificities ↔ neutral drift

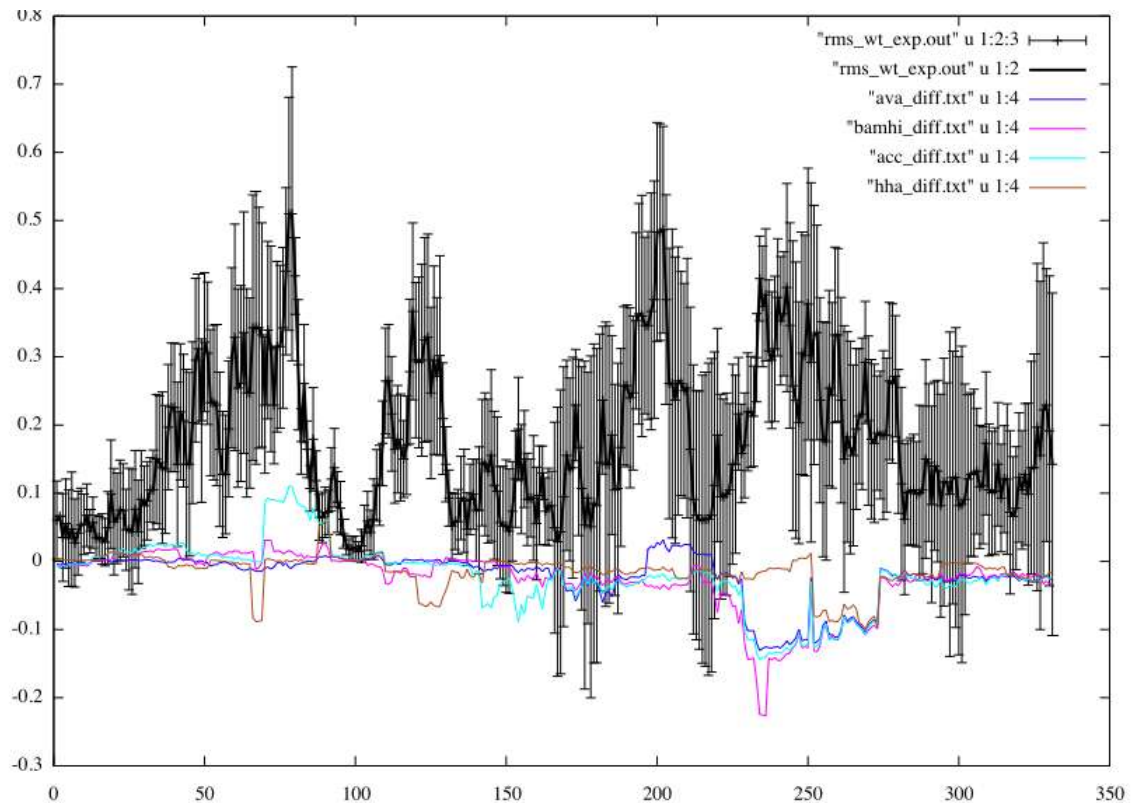
POD changes along specificity changes

Differences



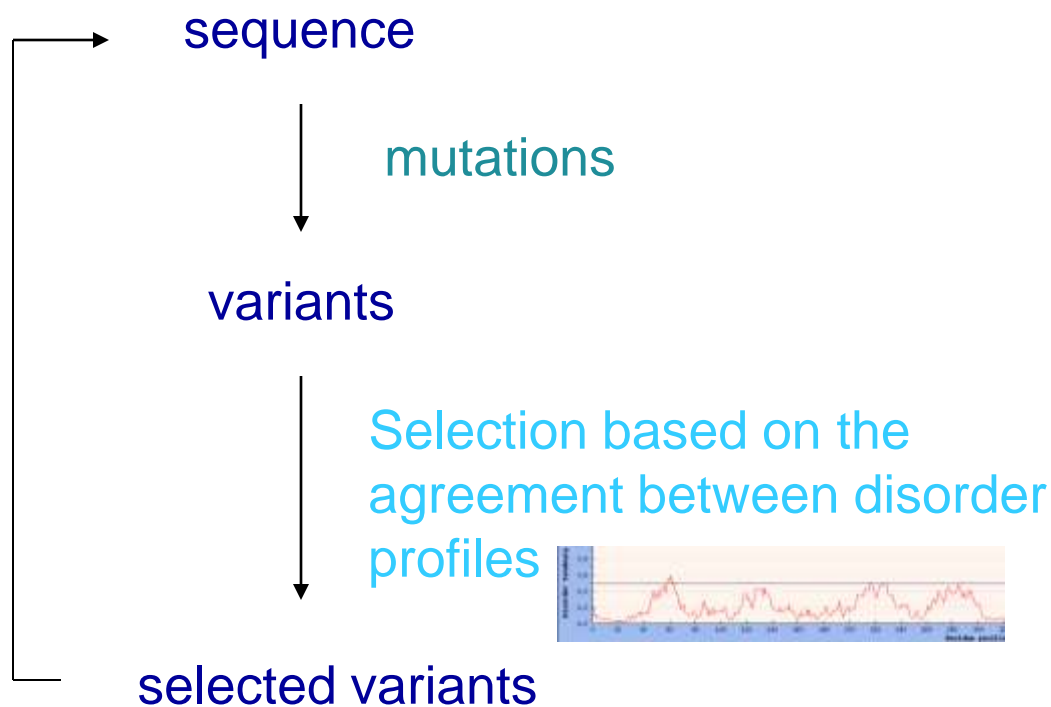
POD changes along specificity changes

Differences vs natural variations in GGCC family



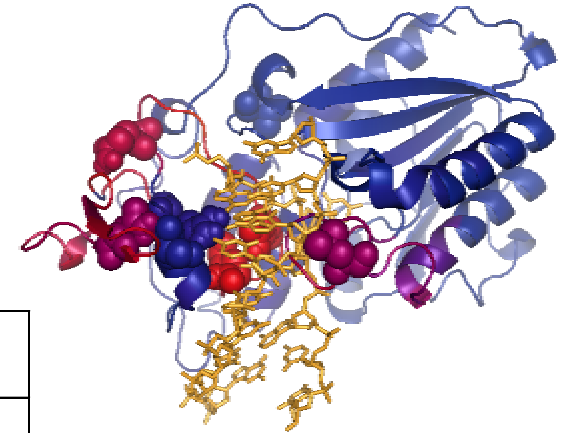
POD changes along specificity changes

How stochastic they are?



POD changes along specificity changes

How stochastic they are?



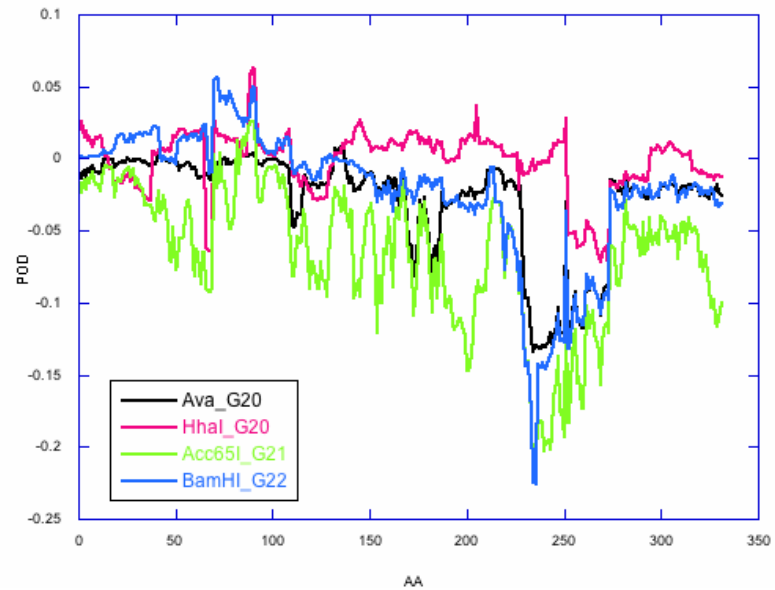
Ava

#position	mutation	pos.hit avg	pos.hit std	true hit avg	true hit std.			
L81	C	0.0158	0.0240	0.0000	0.0000	I 1.6%	-	-
H176	Y	1.0000	0.0000	0.8000	0.4000	Y 80.0%	I 20.0%	L 0.0%
S225	G	0.1334	0.2648	0.1334	0.2648	G 13.3%	-	-
R228	H	0.0000	0.0000	0.0000	0.0000	-	-	-
Q231	R	1.0000	0.0000	0.0000	0.0000	H 100.0%	-	-
Q240	L	1.0000	0.0000	1.0000	0.0000	L 100.0%	-	-
Q245	H	1.0000	0.0000	1.0000	0.0000	H 100.0%	-	-
N263	Y	1.0000	0.0000	0.8000	0.4000	Y 80.0%	I 20.0%	-

POD changes along specificity changes

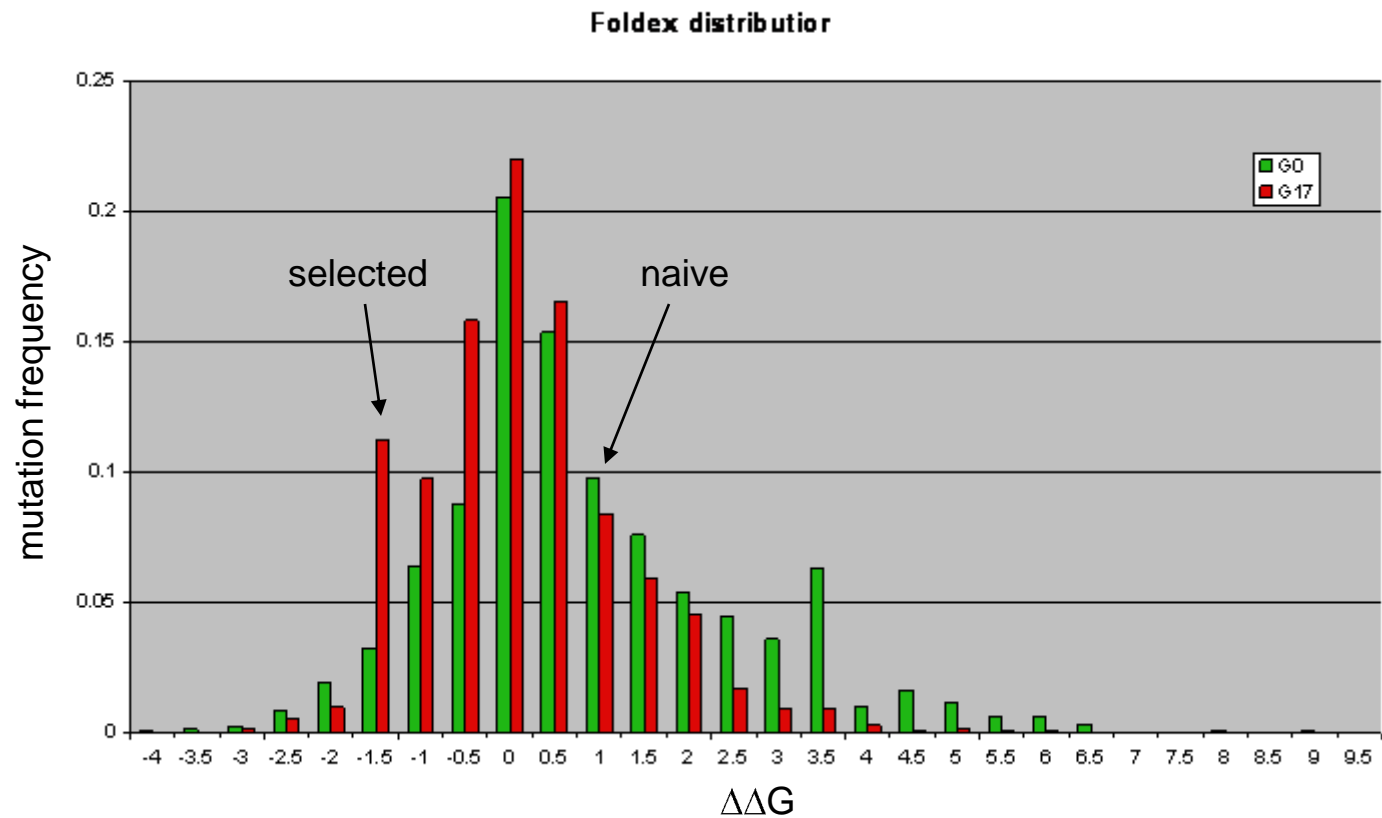
Copyright 2011 by the author(s)

What do they mean???



POD changes along specificity changes

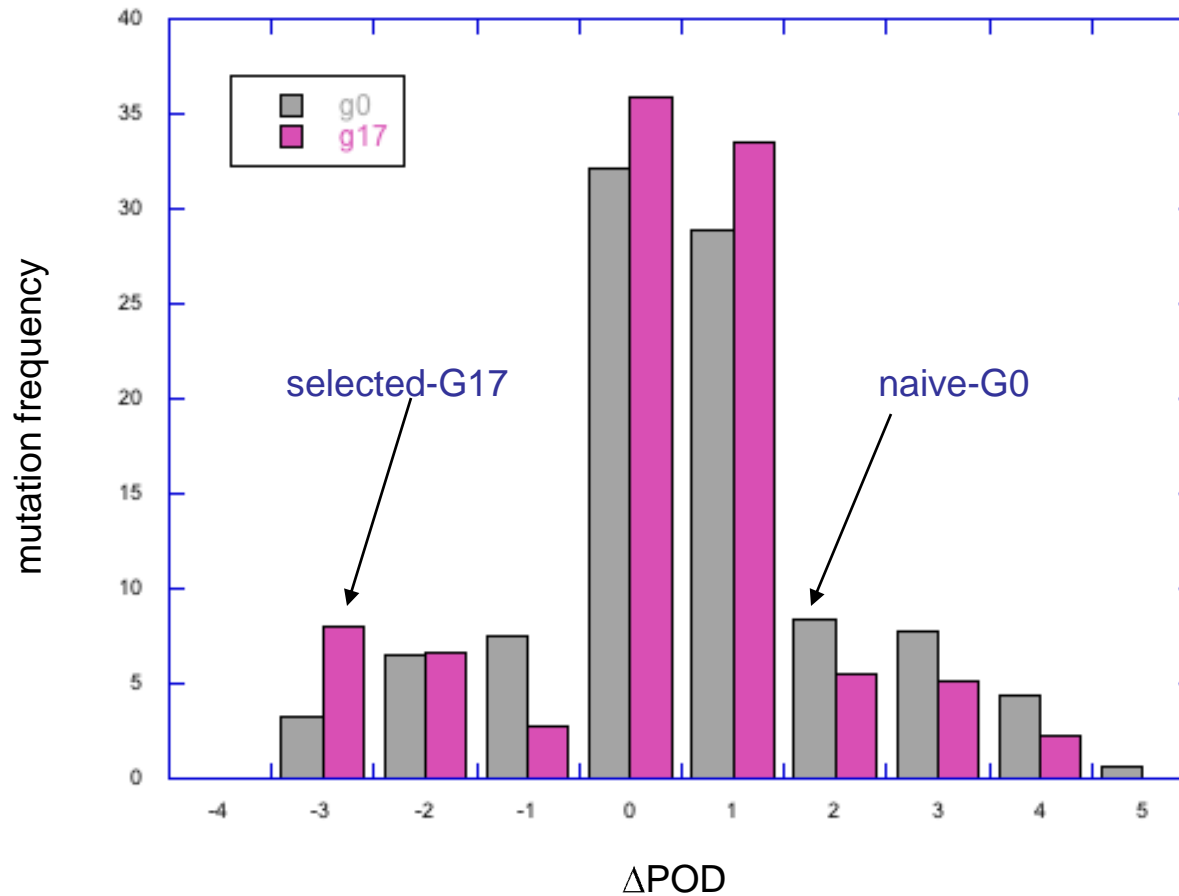
Liat Rockah's idea - stability



neutral drift - HTS

POD changes along specificity changes

following Liat's idea - PODs



neutral drift - HTS

POD changes along specificity changes

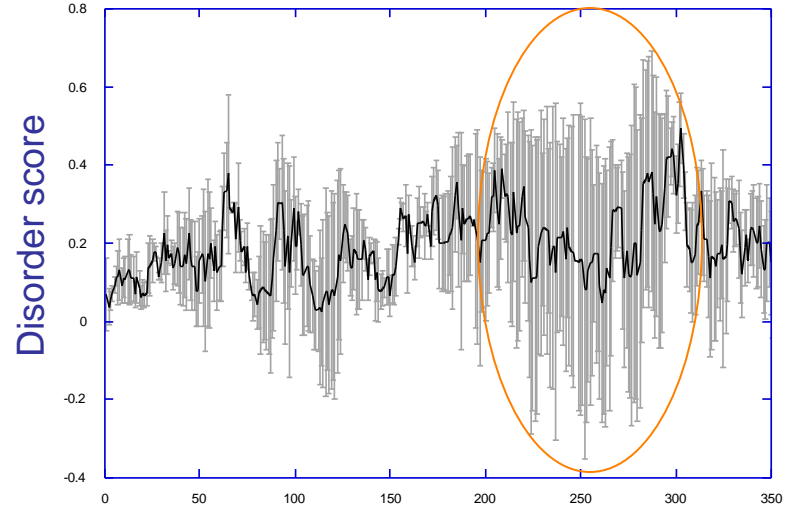
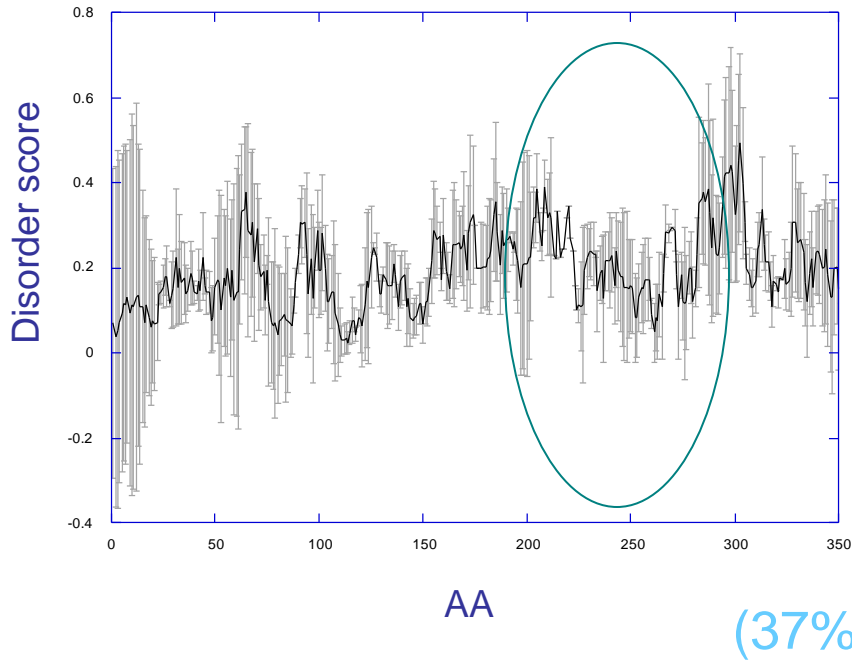
What do they mean???

global suppressors, stability

POD changes vs protein function

Disalign - POD alignment

NMael - CCGG



AA

GTCGAC (58%)

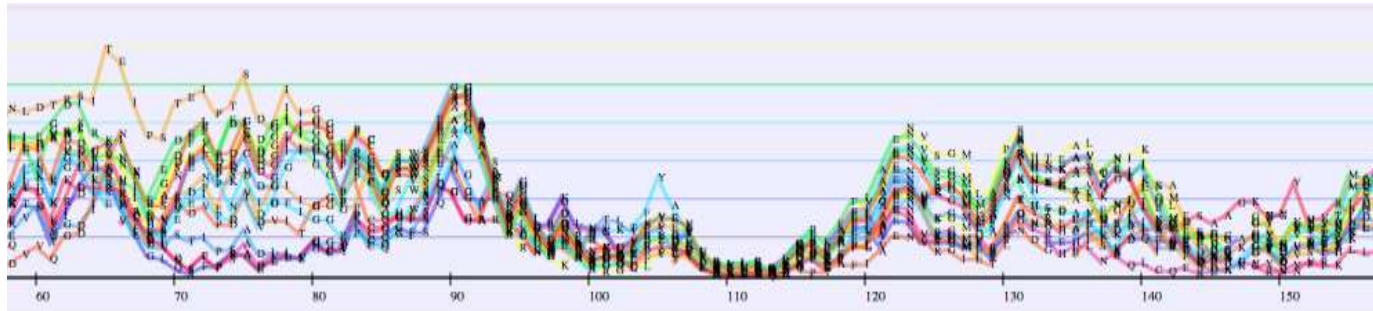


POD changes vs protein function

Can we exploit?

(also combine with MSA)

Disalign



coloring scale is absolute.
|.....70.....80.....90.....100.....110.....120.....130.....140.....150.....160.....170.....180....
--VID--VREIQSFPSADVLVGCYPCQGY**SQGGARQSSRGVNYLYR**-EFVRLAKRIKPKAF**VENV**S**GMQRSDFRHLLNSQLCQFRLAGYRVNVRVID**-VS-DFGV**PQERKRLIFVGVRS**DLGEV
'DYILGDISGLQSFPSAELLIGCYPCQGY**SQGGARKADRKINTLYL**-EFARALSKIKPKAF**VENV**S**SMVRRN**-FEHLLKDQPKVFEEA-GY-TVSS**QILNASHYGV**S**QDRKRIFIVGIRK**DYG-
'DYRLGSVADIKSFPAELLVGCYPCQGY**SQGGAREANRNINLYL**-EFLRALQIQPKAF**VENV**S**SMIRSTYRHLDDQISKFSAACKYGYHVRPAVMNAQDFGVAQ**ERKRI**IVGIRK**DLGV
'VQGDISKIDVSTIPSAIDILLGCYPCQGY**SLAGFRQLDDSRNTLYK**-YVVKYLEEKKPLAF**GENV**K**SMLTLGNG**-EIFKAI**ISDM**DKKGYT-LFYKLLNAKDYSVP**QNR**ERVI**IVGFR**NDLGIT
'VQGDIAKIDYSDVFD**TDVITGG**FPCQGY**SLAGFRKINDERNKLYR**-YFVKLVELKQPYAF**AENV**K**SILTLG**DG-EIEAI**IEDF**ASKGYD-VYPNLVNAADYGV**PQDR**WRVILY**GF**RKDLVK
'**NK**KPIEWLKPSEIPDVVGF**IGG**FPCQGS**SLAGSMCGADDPRGKTFYAYVDLVKEK**DPLFF**IAENV**P**SIVSR**THL-PEFKRLVNSFIDIGYN-VEYKVLNAKD**Y**-GV**PQDR**KRVF**IVG**YREDLNL
'**DT**RSI**IP**ST**EIP**TS**SD**GI**IGG**FPCQGS**SEAGAL**RGINDHRGQLFF**EYIRILK**KQKPF**IAENV**S**SIL**PERHREAFSEI**IKQF**NEI-GYNVSYLLNAHD-YSVP**QDR**KRV**IVG**YHVKT-EL
'**EC**DI**AKI**KEED**FP**EEID**GI**IGG**FPCQGS**SEAGALRGIDDARGQLFFDYIRILKSKQKPF**IAENV**S**SMLANR**HN**GA**VQNLKMF**DGC**-GYD**VTL**T**MANAKD**-Y**GVAQ**ERKRVF**YIG**FRKDL-EI
'**BR**RSIVNIH**ES**EIP**DC**CGI**IGG**FPCQGS**SEA**-CSLRGIKDKRQ**LF**DF**IRILKAKQ**PS**FLAENV**S**SMLAPR**HTEAL**TN**IK**Q**MFEDA-GYNLS**BFK**-LLNS-**SDF**KVP**QDR**Q**RV**FF**VG**IRKDLGF
'**DK**RSIVNIPADEV**PE**CD**GI**IGG**FPCQGS**SEAGAA**RC**IKDKRQ**LF**DF**IRILEAKQ**KPF**IAENV**S**SMLIS**KHTEALEG**IKEL**FRNA-GIGYEL**SFQ**MLNASDY**NVP**Q**DR**KRVF**YIG**FRKDLNF
'**KD**IREIASGD**FP**DD**CGI**IGG**FPCQGS**SEAGALRGIEDSRG**LF**Y**IRIL**LRDKQKPF**IAENV**S**SMMAKR**HS**EA**VQNI**ISH**FACA-GRGYN**VFI**Q**TL**NA**DF**Q**V**P**QDR**KRVF**YIG**FRKDLGI
'**EK**DIK**N**LQ**T**NEIN**F**SVD**GI**IGG**FPCQGS**SEAG**N**LKGIDDARGQLFY**EYLRLLKELK**KPF**IAENV**R**SMLAQR**HEK**SVK**NILNA**F**K**EC**-GYEV**N**THLVNAK-DY**GVAQ**ERL**RV**F**YIG**FR**DL**KIN
'**IE**GD**IR**HILEED**FP**NDID**GI**IGG**FPCCH**SEAGALRGINDARG**LF**Y**YIRILK**NKQK**PV**FT**AENV**S**SMLANR**HS**DA**VKSIL**N**MF**DDC**-GYD**V**T**V**N**M**VNAK-NY**GVAQ**ERKRVF**YIG**FR**DL**LAIN
'**IK**DDIT**N**VR**S**N**EL**PD**CGI**IGG**FPCQGS**SEAG**S**LKGIED**PR**Q**LF**Y**YIRIL**RD**K**SP**K**PF**IAENV**K**SMMAKR**HN**KAVQ**S**II**SQ**F**NKA-GYD**V**F**I**HLLNAS-DY**GVAQ**DR**KR**IF**Y**VG**FR**KDLN**IC**
'**IK**GD**IS**K**IS**DE**FP**K**CD**GI**IGG**FPCQGS**SEGG**S**LR**GIDD**PR**GLFY**YIRILK**QK**PI**FT**AENV**K**SMMAQ**RHN**KAVQ**E**PI**Q**E**FDNA-GYD**V**H**I**LLNAN-DY**GVAQ**DR**KR**IF**Y**IG**FR**KDLN**IN**
'**IK**KD**IR**E**IL**SE**EL**PK**SD**GI**IGG**FPCQGS**SEAG**S**LR**GIND**PR**GLFY**YIRILK**DIQ**PK**PF**IAENV**K**SMLSKR**NT**EA**V**K**DI**IK**EF**E**RA-GYN**V**F**I**KLLN**A**F-DY**GVAQ**DR**ER**RVF**Y**VG**FR**KDLN**IS**
|.....70.....80.....90.....100.....110.....120.....130.....140.....150.....160.....170.....180....
coloring scale is absolute.

Mtases - GGCC specificity

POD changes vs protein function

A coarse-grained structural feature,
related to dynamics, is preserved in evolution

Function? Mechanism? Significance?



(mere stability?)

POD changes along specificity changes

What does it mean???

given dynamical characteristics (even ID)

are essential for given functions

might seem trivial, but

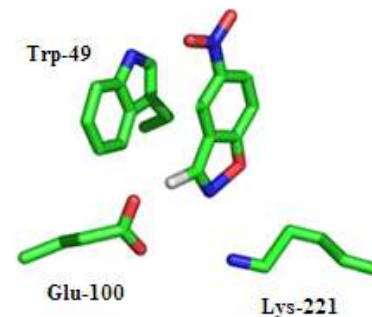
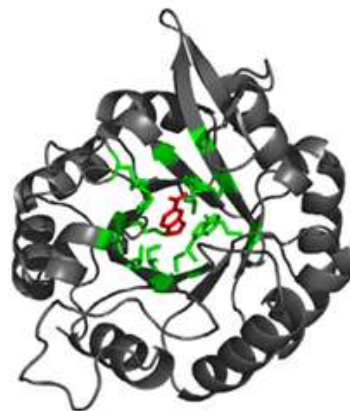
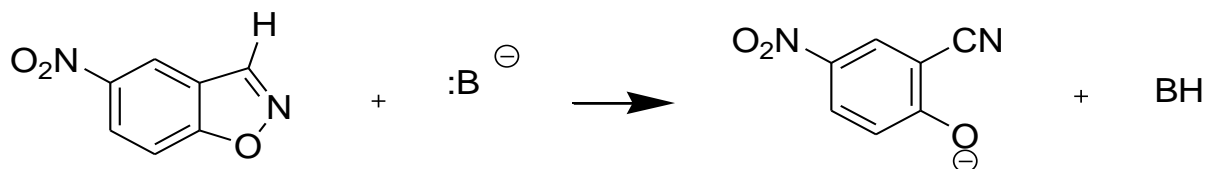


Dynamism in DNA recognition

- likely to be ubiquitous within the transcription machinery
- new approach to fine-tune DNA binding
- PT modifications
- multimerization



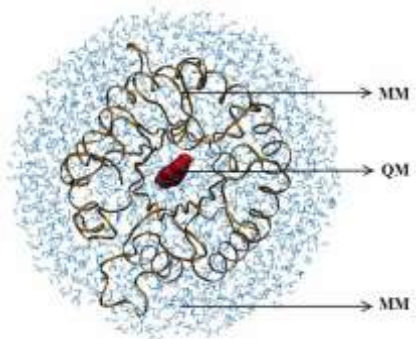
Dynamism in enzymatic catalysis



Olga Khersonsky PhD thesis

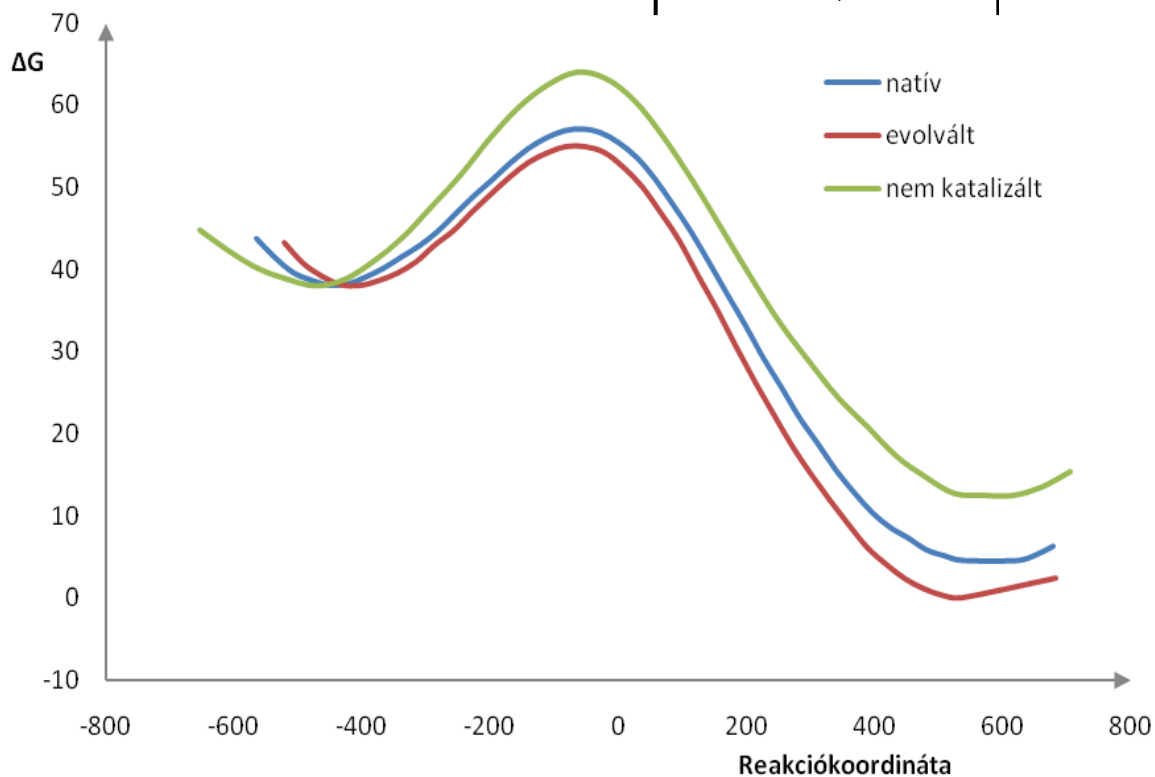
QuickTime™ and a decompressor are needed to see this picture.

Dynamism in enzymatic catalysis



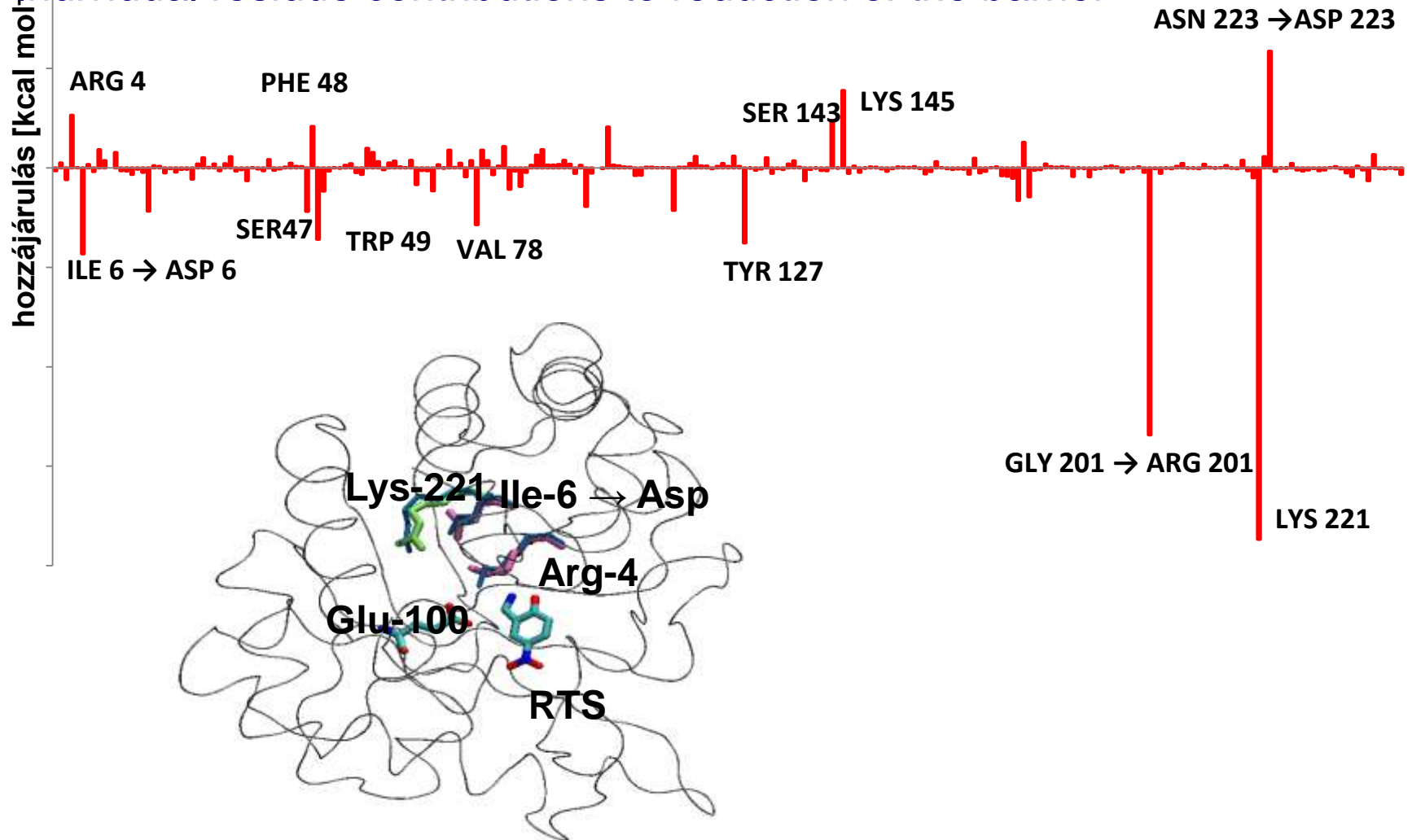
Enzyme	Δg^\ddagger simulation	Δg^\ddagger exp. (kcal/mol)
native	19,0	20,1
R7	17,0	17,8

EVB QM/MM

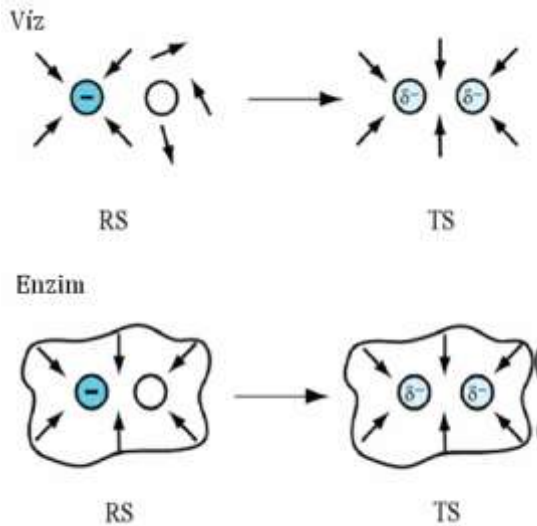


Dynamism in enzymatic catalysis

Individual residue contributions to reduction of the barrier

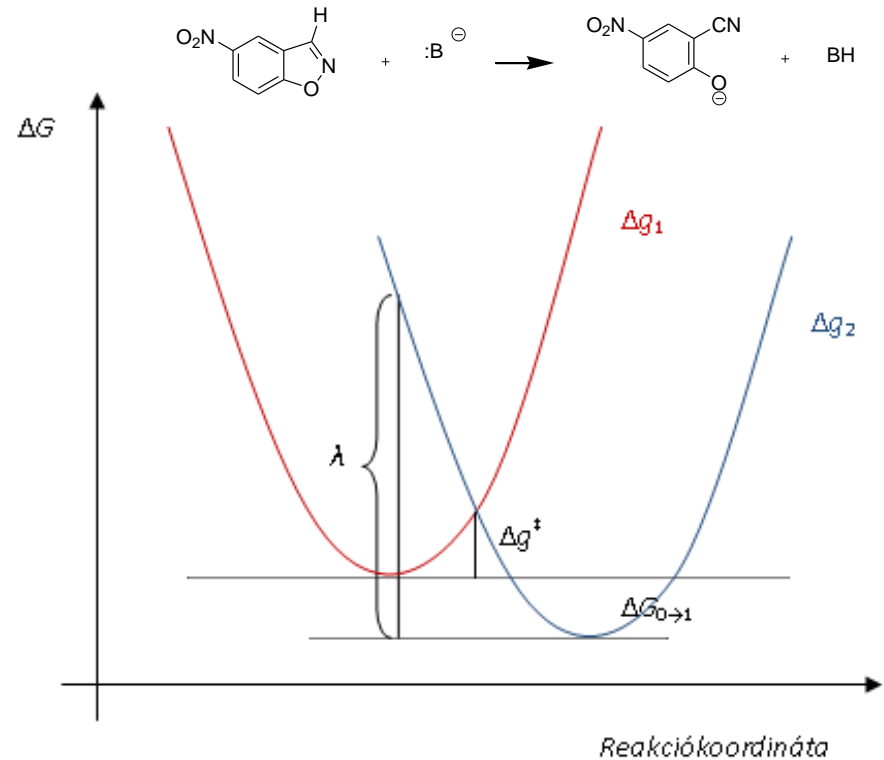


Dynamism in enzymatic catalysis



$$\lambda = \Delta G_{0 \rightarrow *} - \Delta G_{0 \rightarrow 1}$$

$$\lambda = \Delta g_2(RS) - \Delta g_2(PS)$$

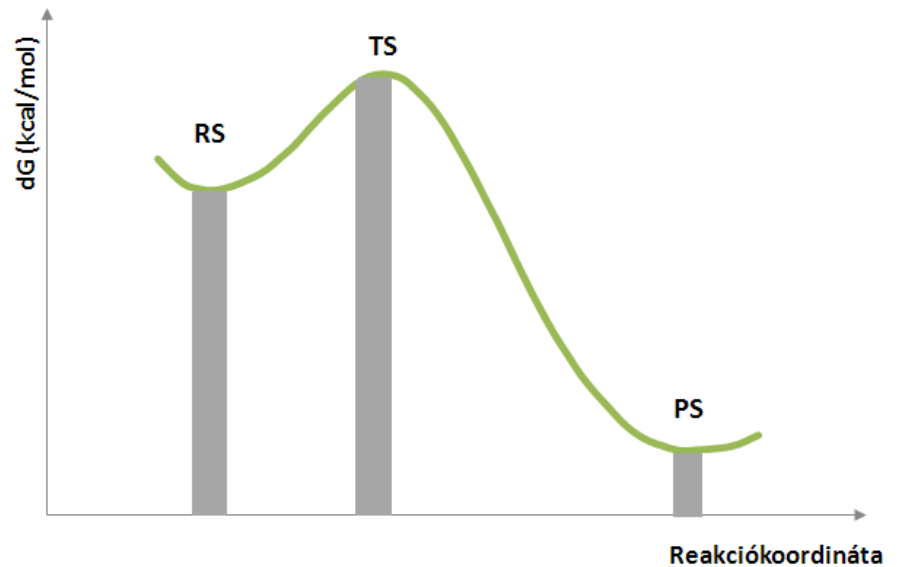
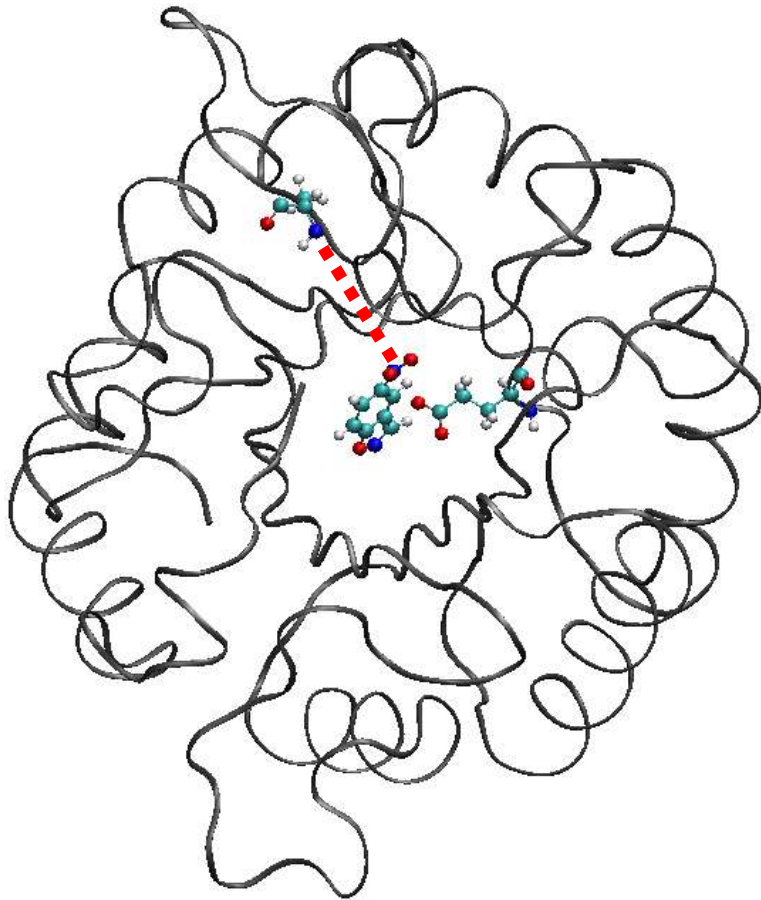


Improvement in reorganisation energy ~ 40 kcal/mol

closely related to protein dynamics

Dynamism in enzymatic catalysis

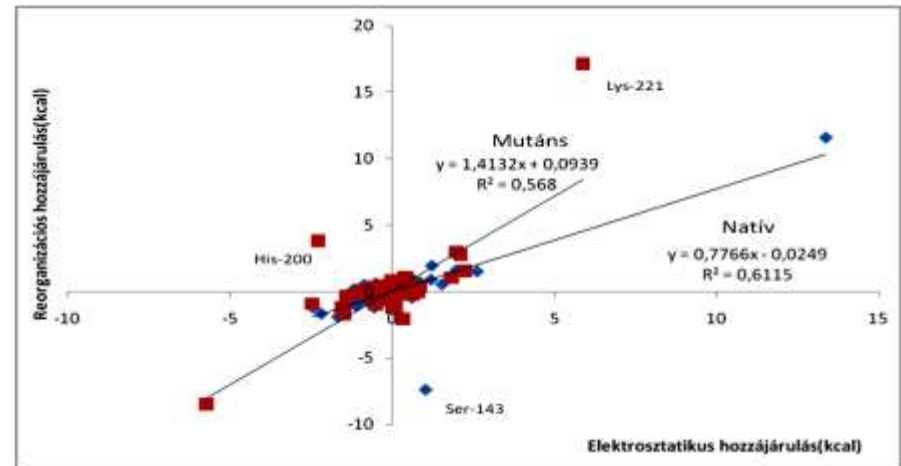
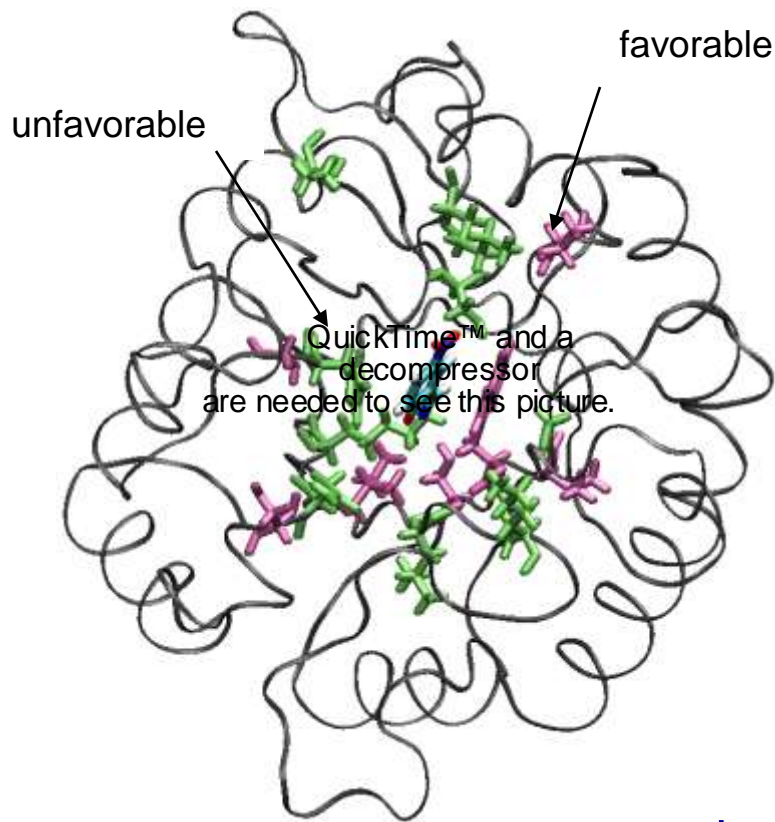
Individual residue contributions to reorganisation energy



$$\left\langle E_i^{el} \left[\mathbf{R} \right]_{\lambda_j} \right\rangle = \left\langle 332 \sum_k^{N_Q} \sum_l^{N_R} \frac{q_k^i q_l}{r_{kl}} \right\rangle_{\lambda_j}$$

Dynamism in enzymatic catalysis

Individual residue contributions to reorganisation energy



in R7 the reorganisation energy is optimized

Dynamism in enzymatic catalysis

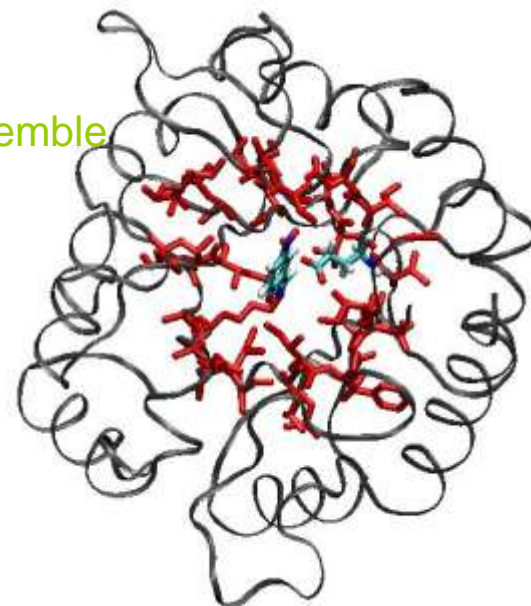
Can we exploit ?

Idea: screening for improving reorganisation energy

screening algorithm

1. generation of mutants on RS and PS (TS) conformational ensemble
2. optimization
3. calculation of REORG energy and average

screening using all residues within 12 Å from substrate



Dynamism in enzymatic catalysis

Comparison to Olga's data

$$\Delta\Delta G_{\text{reorg}} \sim 0 \text{ or } \ll 0$$

All mutation are either **neutral**
for the **reorganization energy**
or **improve** it

QuickTime™ and a
decompressor
are needed to see this picture.

K221A

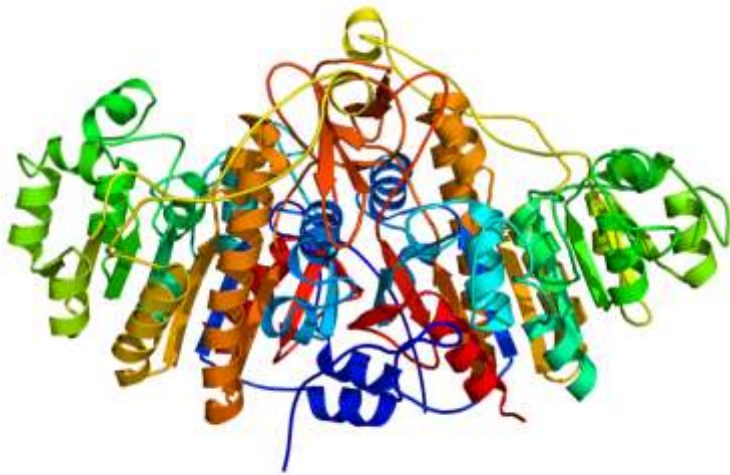
$$\frac{k_{\text{kat}}(\text{mut})}{k_{\text{kat}}(\text{nat})} \cong 3$$

$$\Delta\Delta g^{\ddagger}_{\text{reorg}} = -2,3 \text{ kcal mol}^{-1}$$

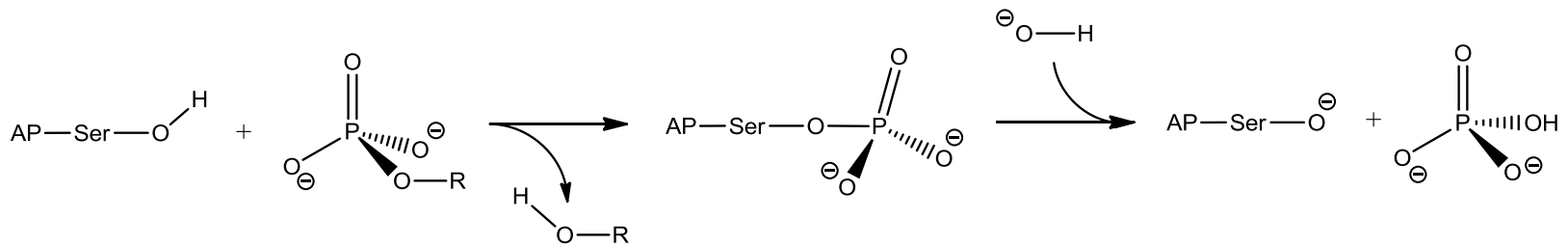
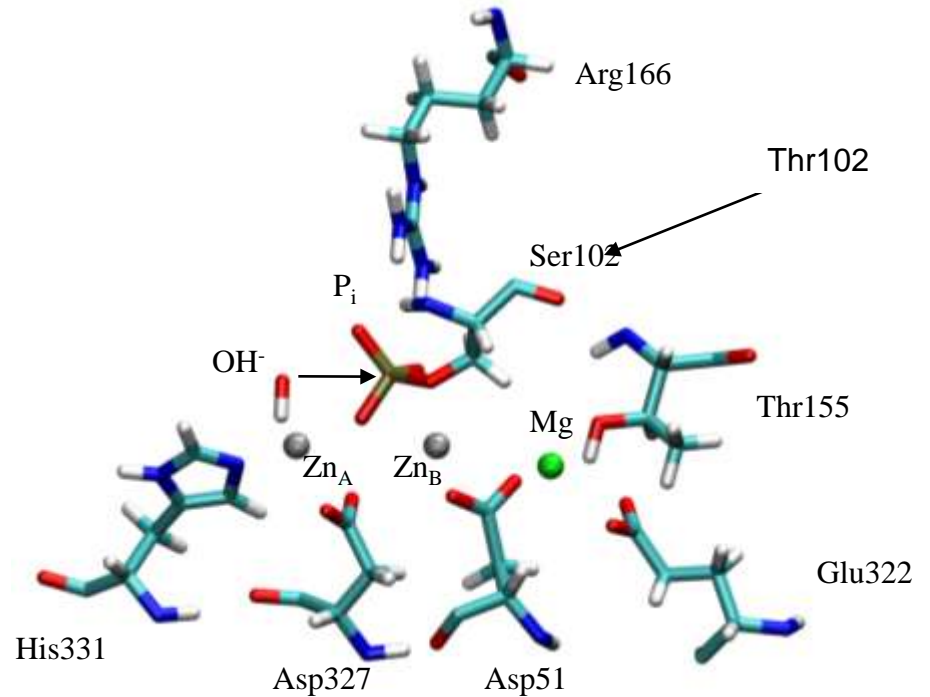
Dynamism in enzymatic catalysis

reorganisation energy seems
to be an important factor in
enzymatic evolution

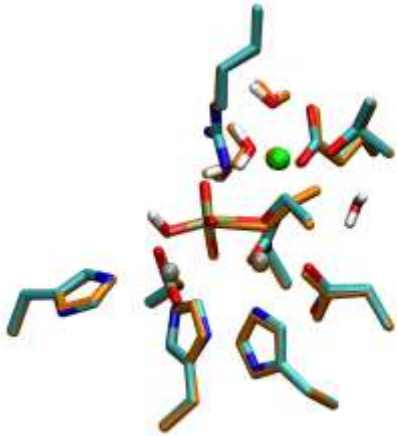
Dynamism in enzymatic catalysis



alkaline phosphatase



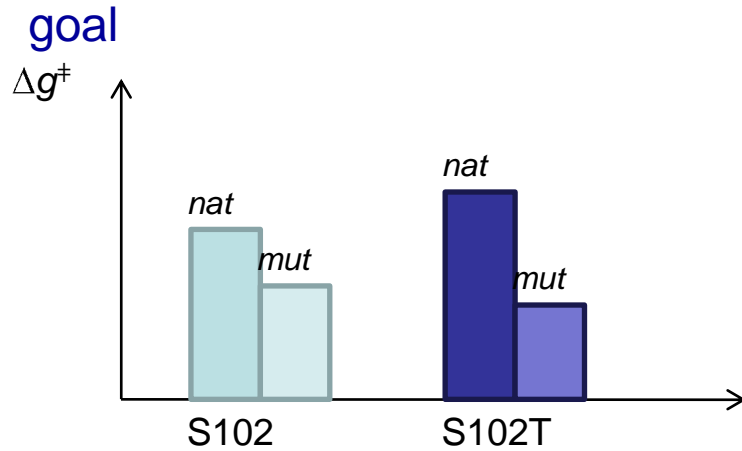
Dynamism in enzymatic catalysis



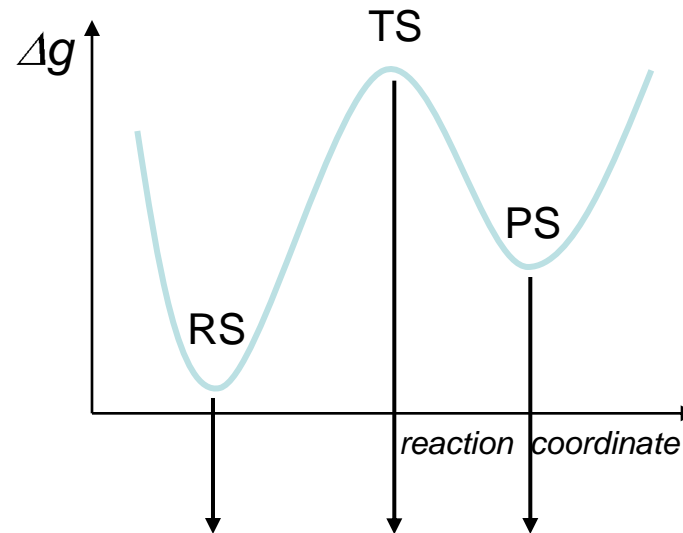
S102/T102 transition state
structure (superimposed)

Enzyme	Attacking step		Departure of leaving group		$\Delta g^\ddagger(\text{calc.})$	$\Delta g^\ddagger(\text{exp.})$
	ΔG	Δg^\ddagger	ΔG	Δg^\ddagger		
native	-11.2	11.1	13.9	26.0	14.8	15.2
S102T	-13.7	11.5	33.9	38.1	24.4	20.2
D369N	-4.0	18.4	-1.0	22.3	18.3	18.0
D369A	2.2	25.7	6.0	23.5	25.7	20.2
E322D	-10.0	12.1	14.5	27.2	17.2	18.9
R166K	-8.6	12.8	12.6	26.1	17.5	17.6

Dynamism in enzymatic catalysis

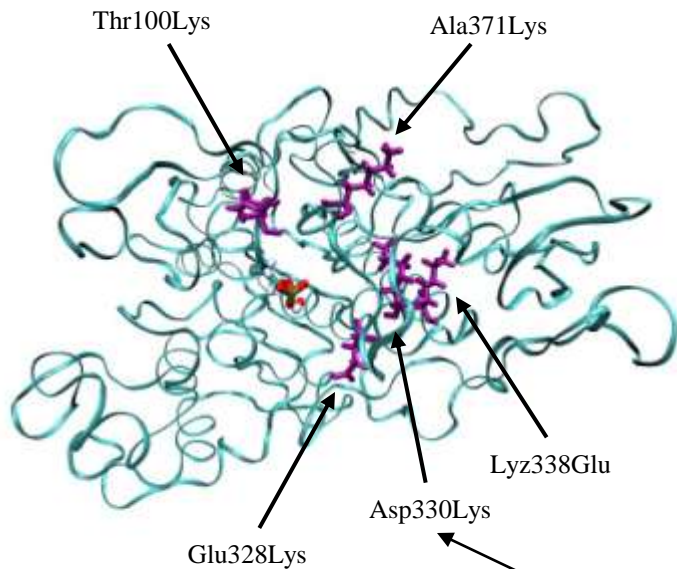


conformational ensembles



collecting configurations at the stationary regions

Dynamism in enzymatic catalysis



Enzyme	$\Delta\Delta g^\ddagger$ (S102)	$\Delta\Delta g^\ddagger$ (S102T)	$\Delta\Delta g^\ddagger$ (S102T-S102)
“native”	-	-	9.6
multimut1: T100K, K328E, Q338K, A371K	-4.3	-4.4	9.5
multimut2: T100K, K328E, D330K, Q338K, A371K	-3.2	-4.8	8.0
multimut3: S56K, T100K, K328E, D330K, Q338K, A371K	-2.3	-4.1	7.8

cannot be improved further

Dynamism in enzymatic catalysis



compensatory mechanisms related
to protein dynamics

Dynamism in protein evolution

- specific POD features related to function
- conserved ID features in DNA recognition
- reorganisation energy opt. along enzyme evolution
- compensatory mechanisms related to dynamics



Perspectives

sequence → structure → function



alternative
sequences



many
structures



alternative
functions



thanks

Institute of Enzymology

Rita Pancsa

Grzegorz Papaj

Anikó Lábás

Eszter Szabó

Letif Mones

István Simon



Weizmann Institute of Science

Liat Rockah

Olga Khersonsky

Misha Shoskine

Eynat Dellus

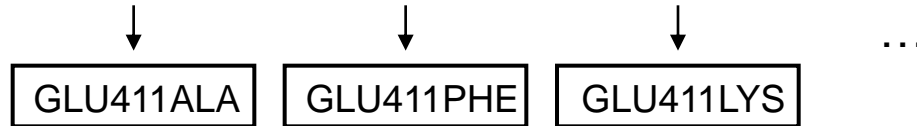
Ágnes Tóth-Petróczy

Nobuhiro Tokuriki

Dan Tawfik

AP: Protocol of changing residues

1. Mutations at the stationary configurations
(Ala, Phe, Lys, Glu, Ser)



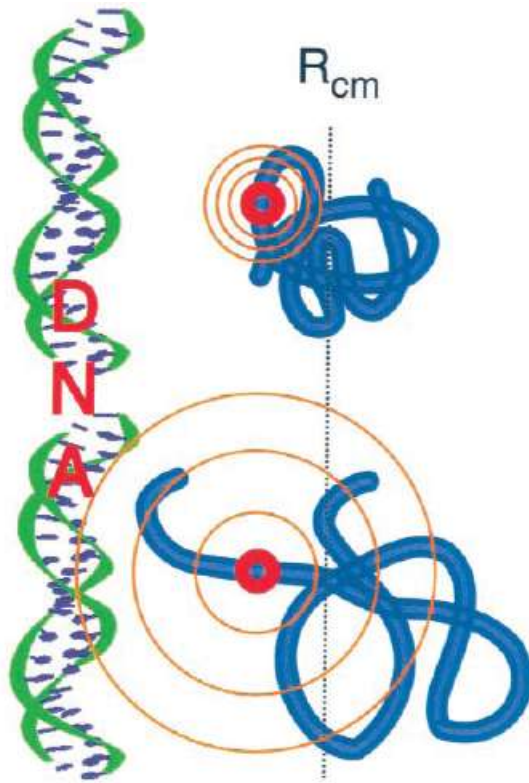
2. Optimization of the geometry of each mutant
(*second approx.: static geometry of the protein during the optimization*)

3. Calculate the new interaction energy of each mutant with the QM zone
(*third approx.: calculation of the electrostatic energy on the mapping potential*)

4. Selection criterion:

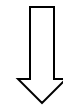
$$\Delta\Delta E_{S102 \rightarrow S102T}^{el} \mathbf{R} \stackrel{-}{=} \Delta E_{S102T, RS \rightarrow TS}^{el} \mathbf{R} \stackrel{-}{=} \Delta E_{S102, RS \rightarrow TS}^{el} \mathbf{R} \stackrel{-}{=} \left\langle E_{S102T}^{el} \mathbf{R} \right\rangle_{TS} - \left\langle E_{S102T}^{el} \mathbf{R} \right\rangle_{RS} - \left\langle E_{S102}^{el} \mathbf{R} \right\rangle_{TS} - \left\langle E_{S102}^{el} \mathbf{R} \right\rangle_{RS} .$$

IDP recognition - facile binding



fly-casting

- distinguished short motifs
- secondary structure elements
- large capture radius
- multiple contact sites

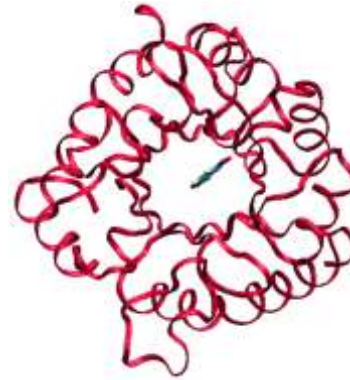
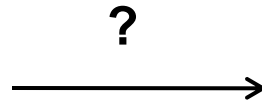


kinetics

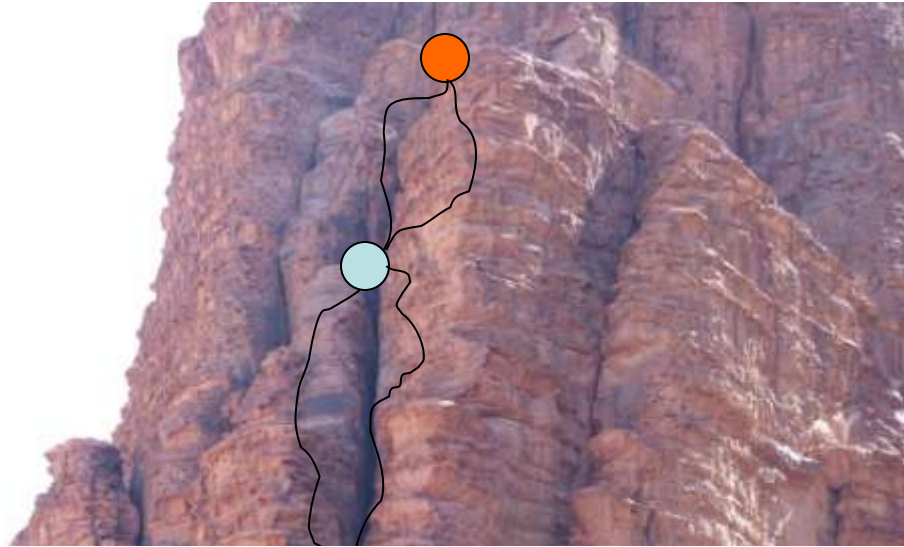
Dynamism in enzymatic catalysis



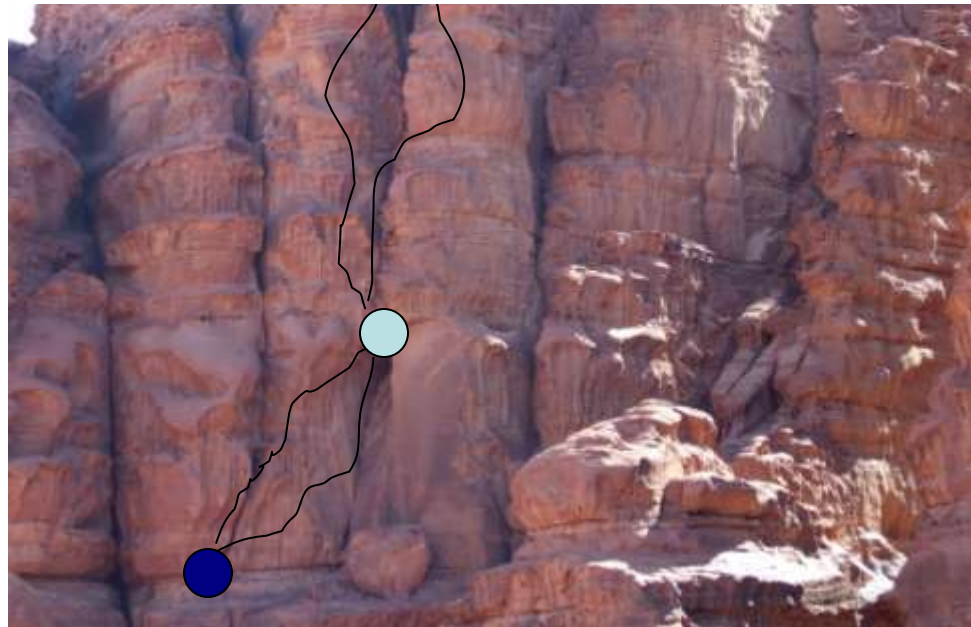
$k_{cat} = 0,018 \text{ s}^{-1}$



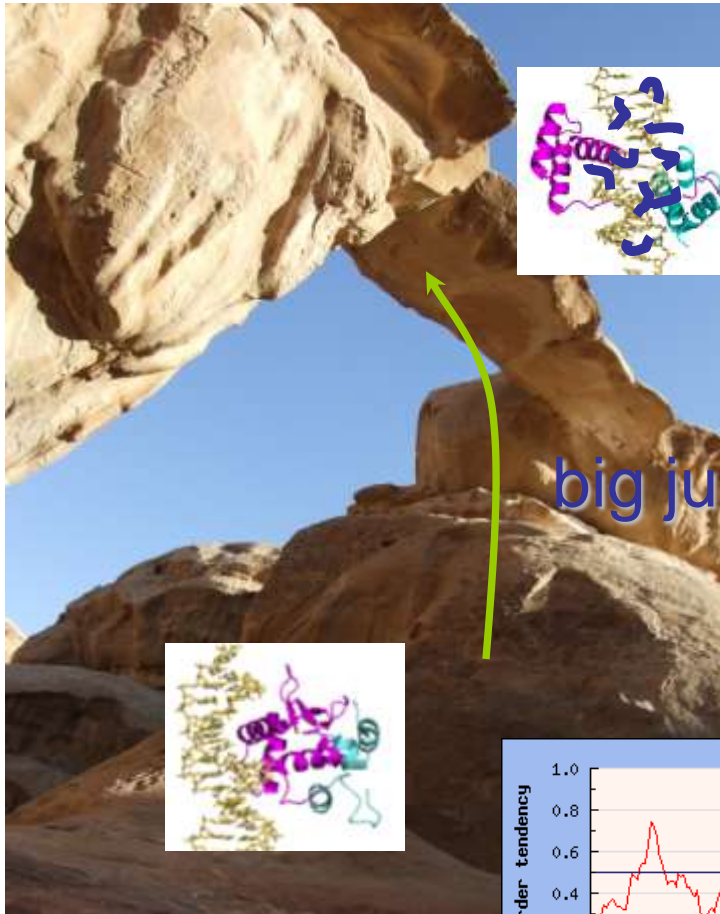
$k_{cat} = 0,76 \text{ s}^{-1}$



Function and evolution

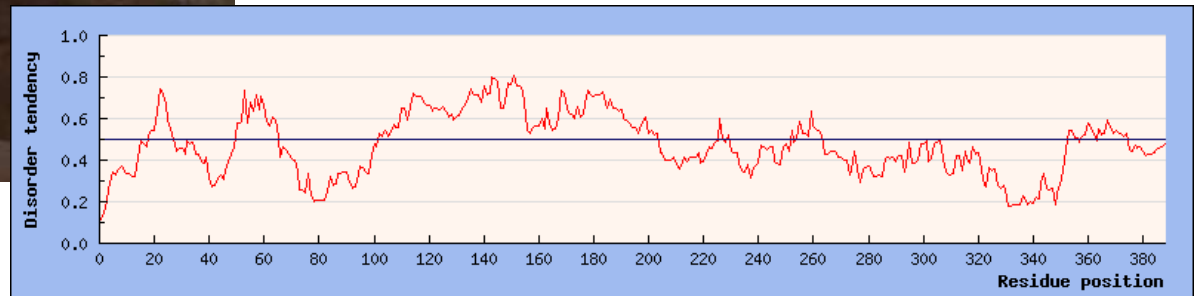


IDPs - from an evolutionary perspective



- development of new functions

big jump?



Fuzziness - perspectives



- ◆ functional complexity
- ♠ multiple binding sites
- ♠ sequence insensitivity
- ♥ evolution of new functions

thanks

Peter Tompa

Agnes Toth-Petroczy



POD changes vs protein function

QuickTime™ and a decompressor are needed to see this picture.

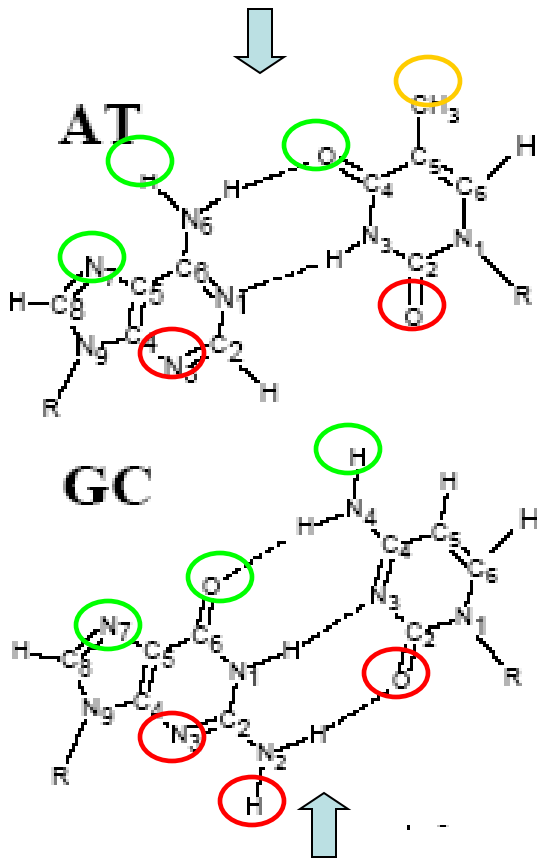
Disalign - POD alignment
in combination with sequence constraints



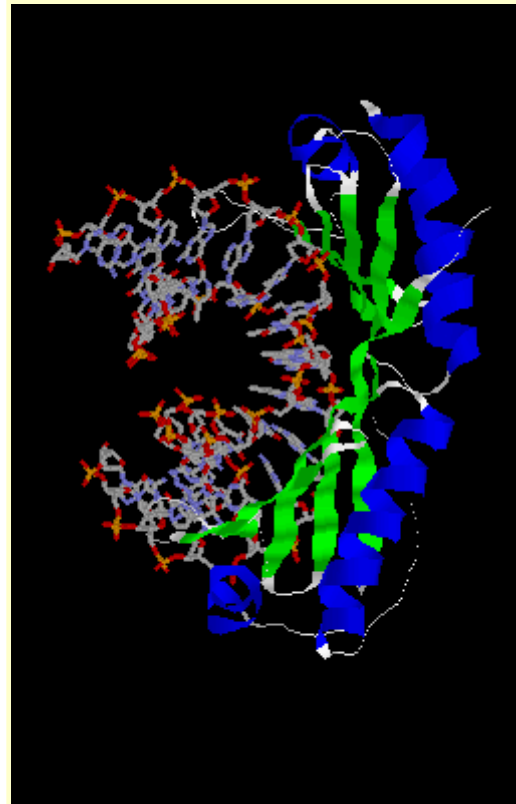
yeast Ure2

Dynamism in DNA recognition

specific H-bonds

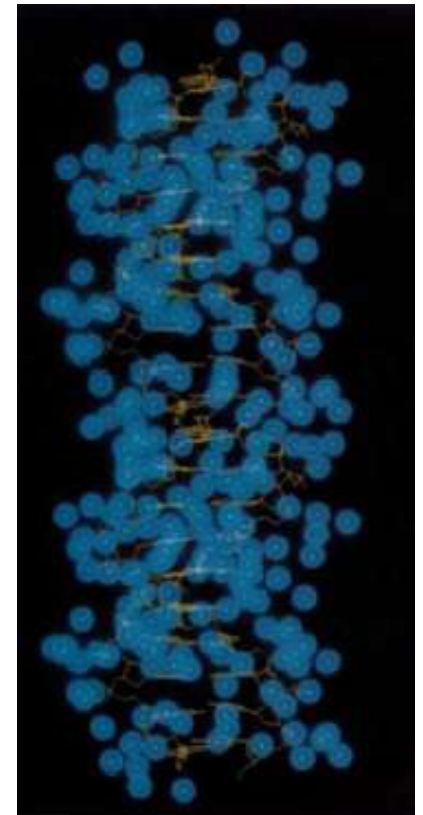


phosphate contacts



TBP -DNA

water



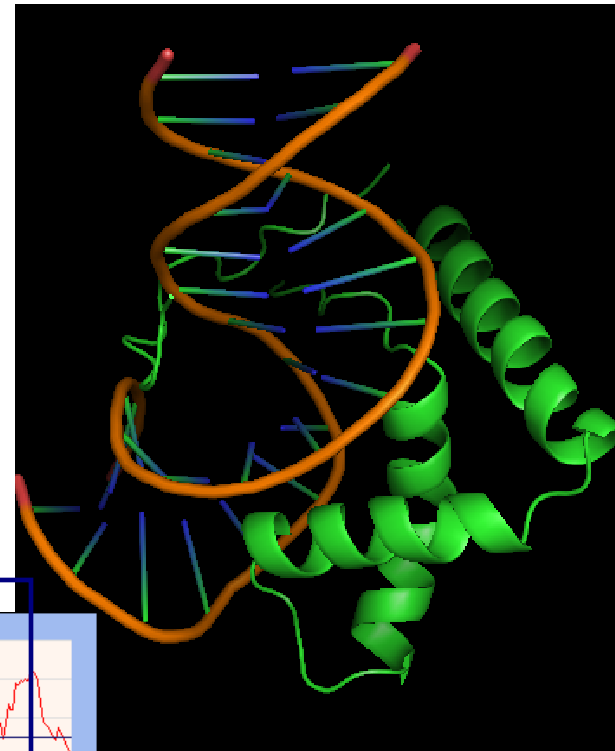
Trp repressor -DNA

Dynamism in DNA recognition

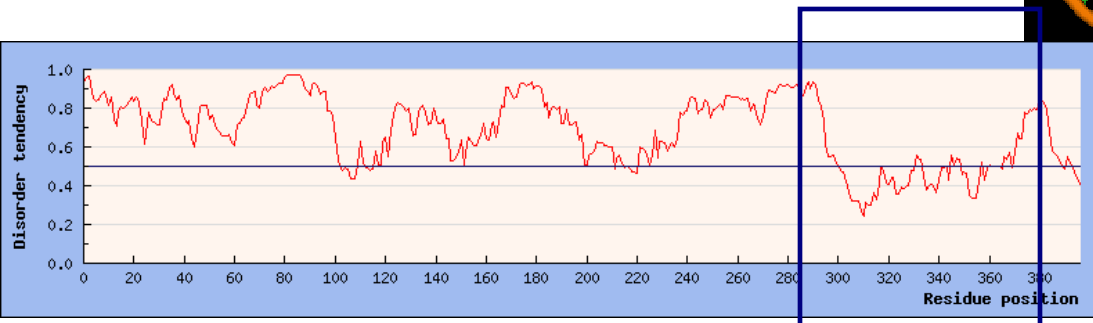
Classical theme:

ID folding upon binding

Can the ID contribute?

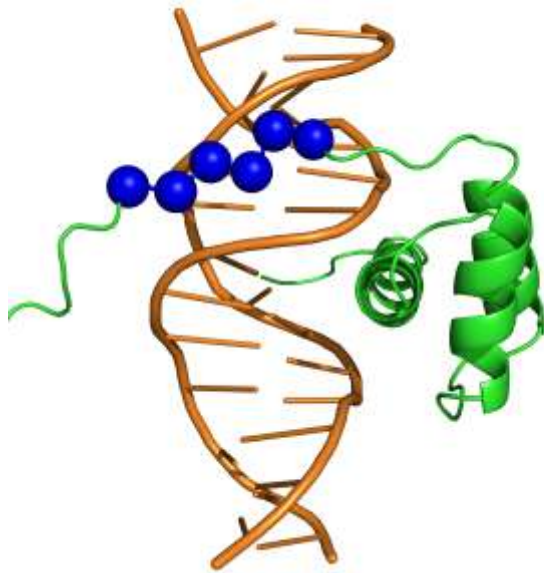


LEF-1



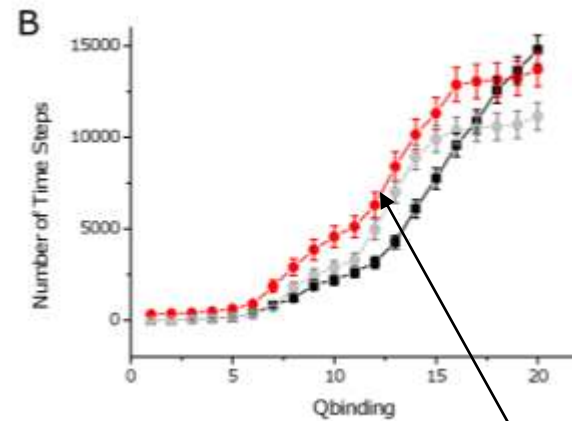
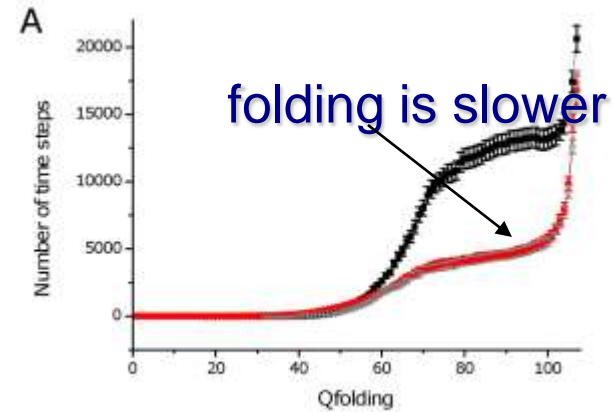
IDP recognition - facile binding

Kinetic benefits



NK2 - DNA

- coarse-grained simulations

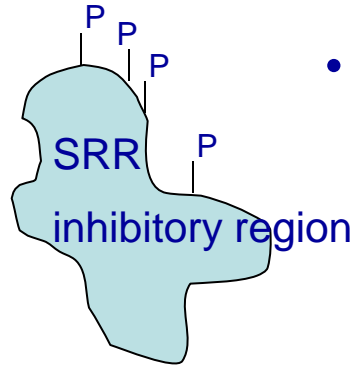


binding is faster

Dynamism in DNA recognition

Ets-1 transcription factor

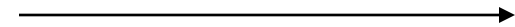
QuickTime™ and a decompressor are needed to see this picture.



- 100-1000 fold affinity decrease

- does not gain structure in complex
- no sec. str. increase upon phosphorylation

QuickTime™ and a decompressor are needed to see this picture.



decreasing flexibility

Fuzzy protein-DNA complexes

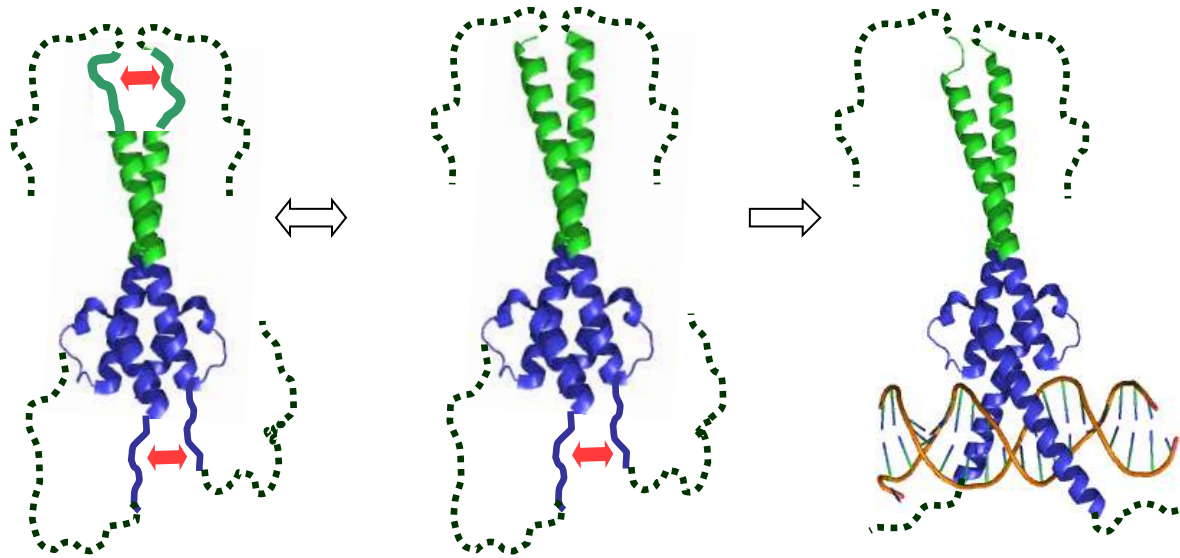
bound ID affects affinity/specificity

structural + biochemical evidence

Model	IDP	ID region	Conservation	Posttranslational modification
<i>Conformational selection</i>				
	Max	NTD CTD	*	phosphorylation
	MeCP2	NTD ID CTD		
	TDG	CTD		acetylation
	Neurogenin	Basic motif		
	ApLLP	NTD,CTD	*	
<i>Flexibility modulation</i>				
	Ets-1	SRR		phosphorylation
	SSB	CTD	*	
<i>Competitive binding</i>				
	PC4	NTD		phosphorylation, acetylation
	FACT	NTD CTD	*	phosphorylation
	HMGB1	CTD	*	phosphorylation
	Ubx	I1, I2, R	*	
	DSS1/Brh2	1-70		
	NKX3.1	AD, SI	*	phosphorylation
	PPAR-g	NTD		
	UvrD	CTD	*	
	b-telomere	CTD		phosphorylation
<i>Tethering</i>				
	Oct-1	Linker	*	
	RPA	IULD	*	phosphorylation
	KorB	NTD, linker		

Dynamism in DNA recognition

Conformational selection

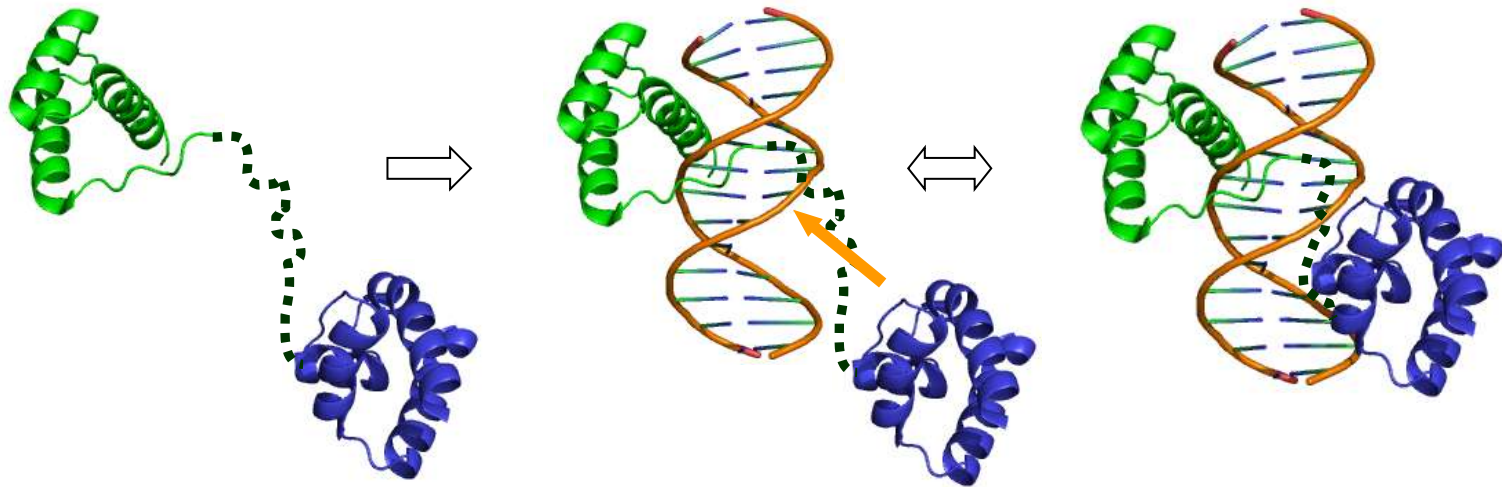


ID NTD, CTD: promote formation of recognition helices

Max transcription factor

Dynamism in DNA recognition

Tethering

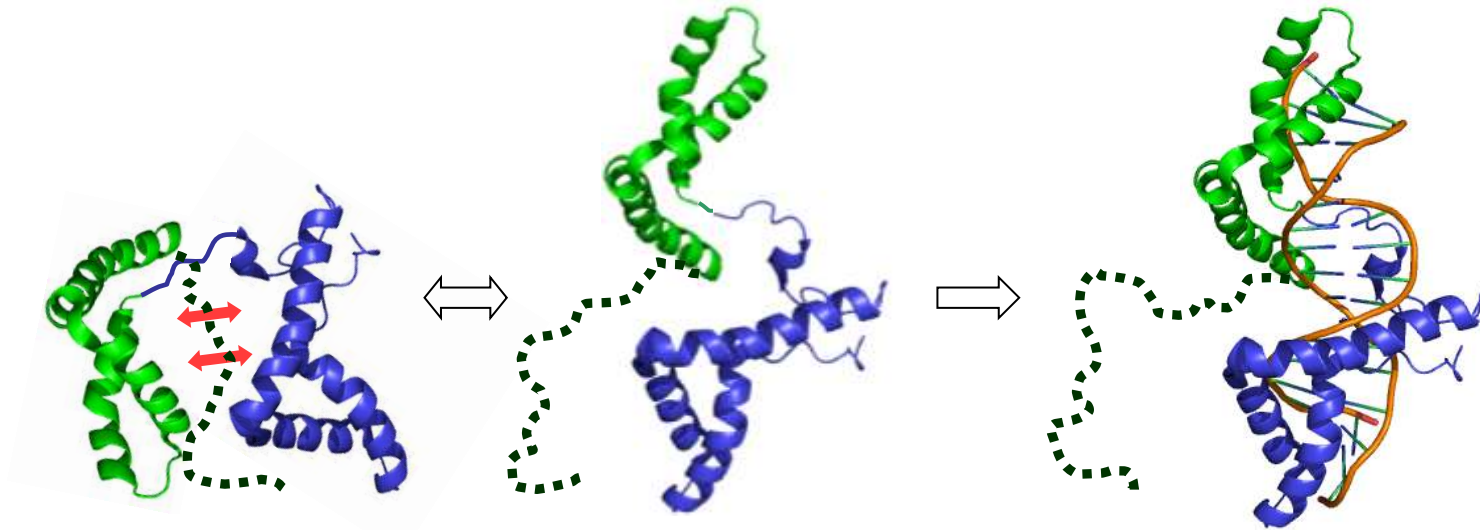


ID linker: tether globular domains to DNA

Oct-1

Dynamism in DNA recognition

Competitive binding

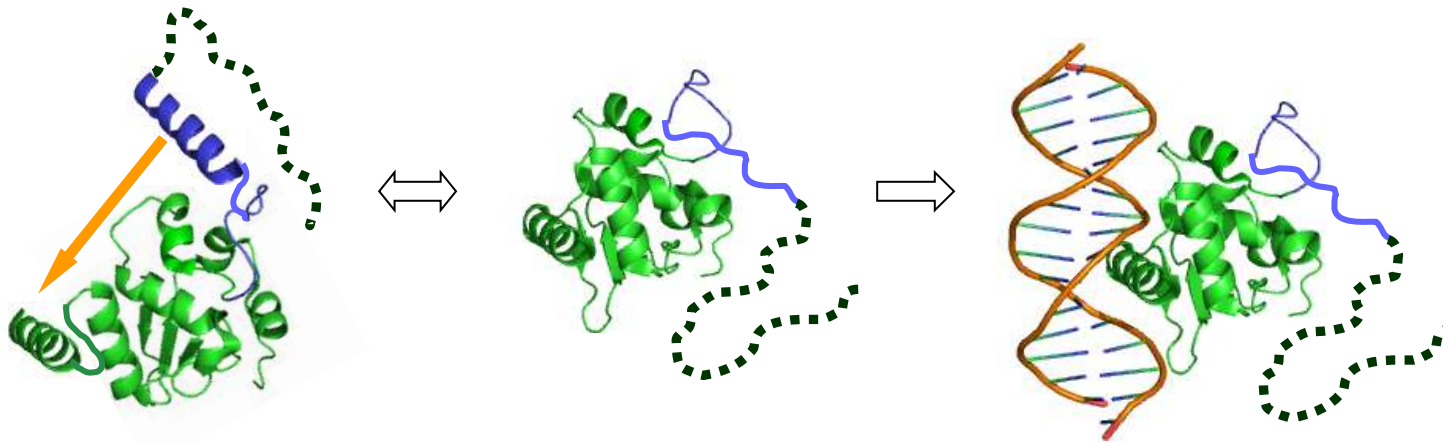


ID region: competition between protein-DNA and intra- or intermolecular protein-protein interactions

HMG-B

Dynamism in DNA recognition

Flexibility modulation



ID region: tunes dynamic properties of the recognition region

Ets-1

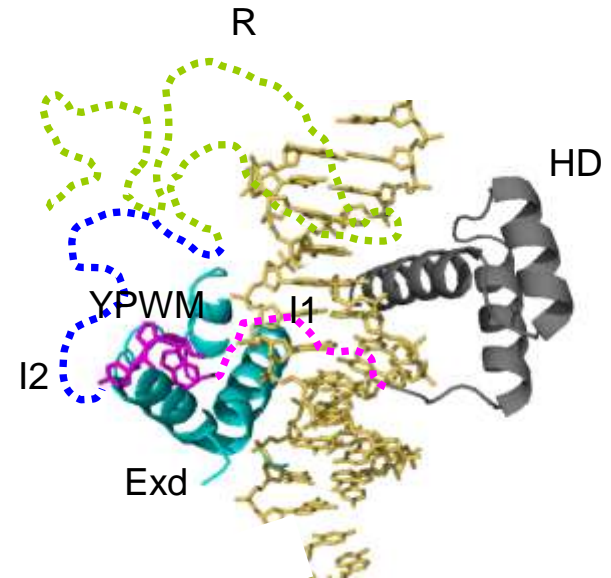
Dynamism in DNA recognition

Dynamic DNA readout

- Conformational selection
- Flexibility modulation
- Tethering
- Competitive binding

Regulatory tools

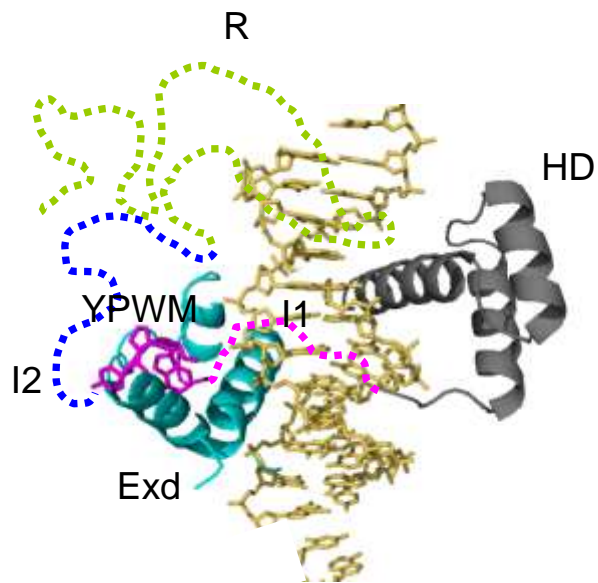
- protein-protein interactions
- post-translational modifications
- alternative splicing



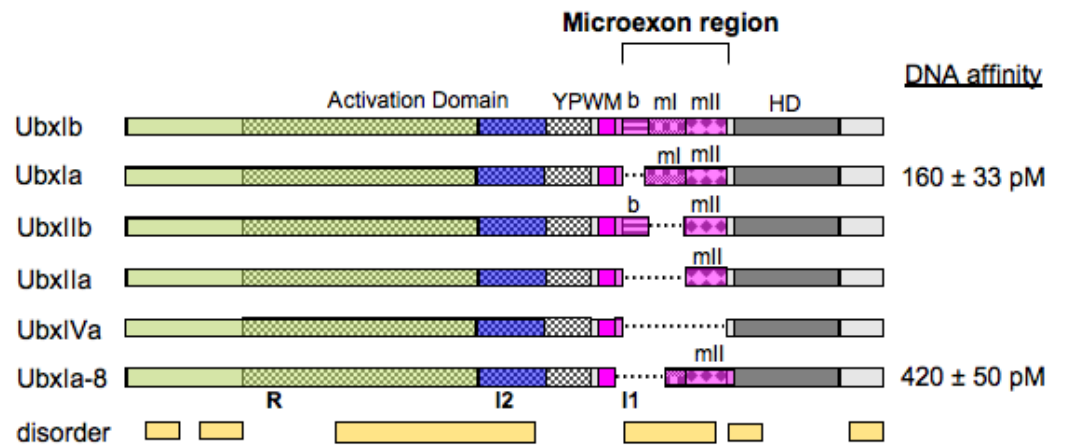
Ultrabithorax homeotic transcription factor

Dynamism in DNA recognition

conserved ID character



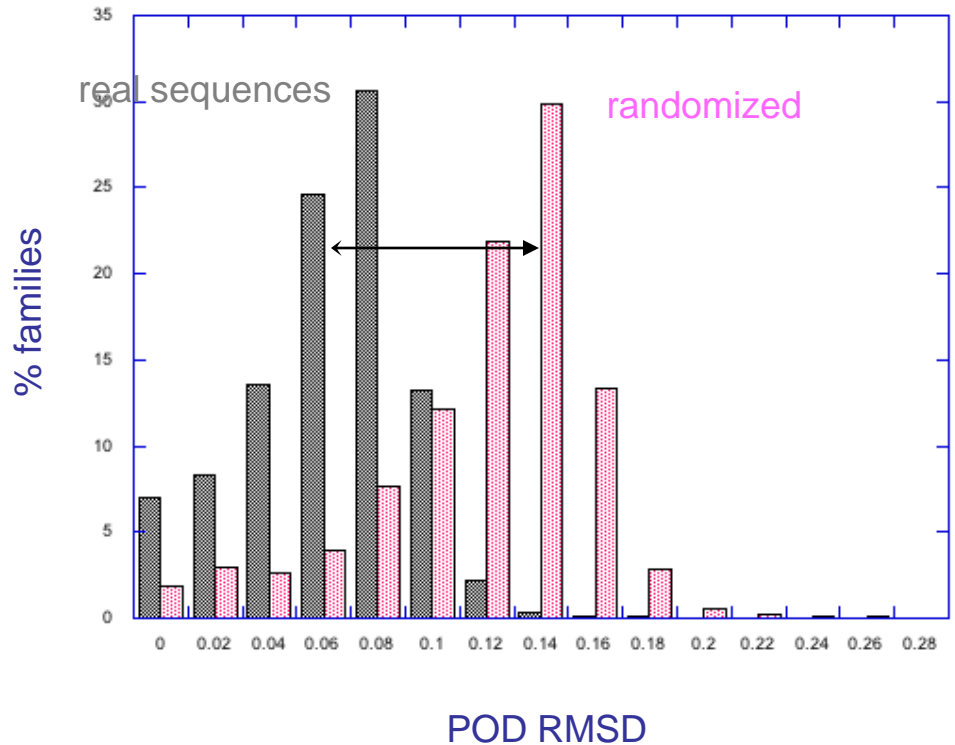
Ultrabithorax homeotic transcription factor



context-specific regulation

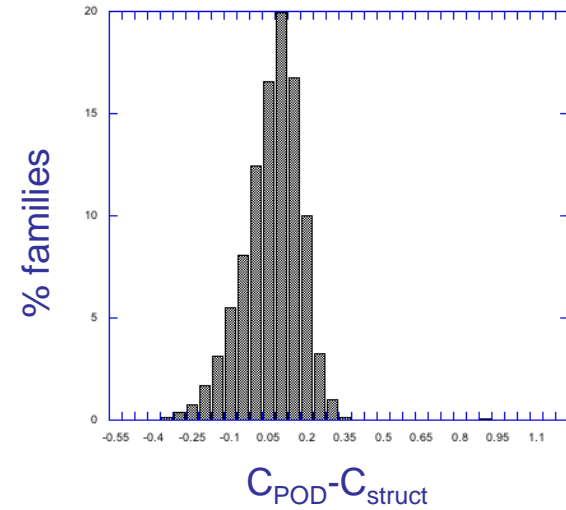
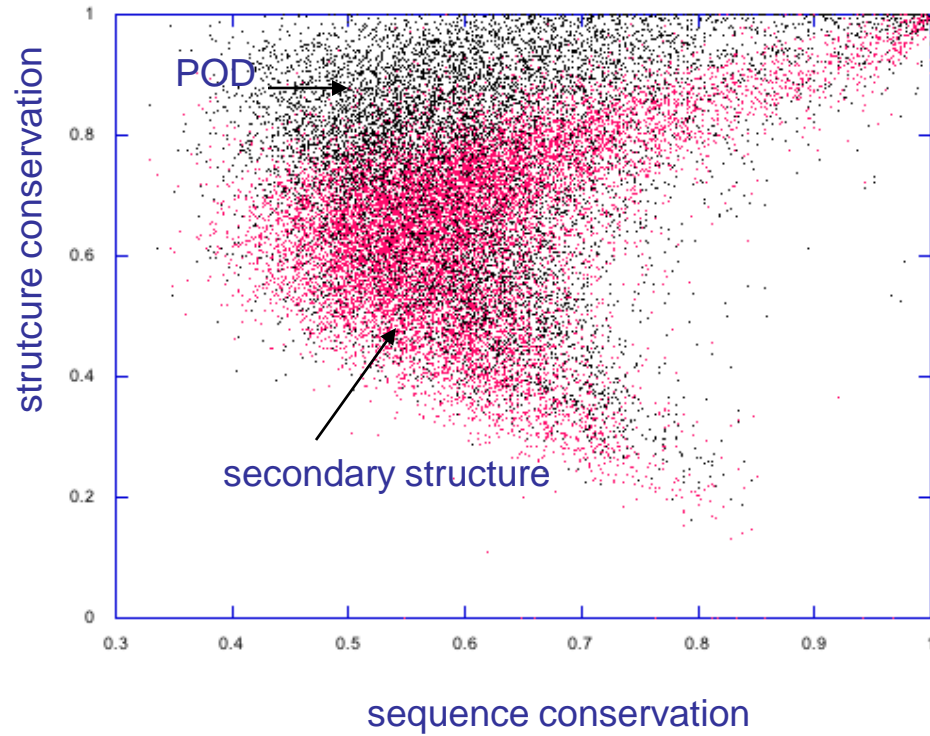
POD changes vs protein function

general - Pfam database



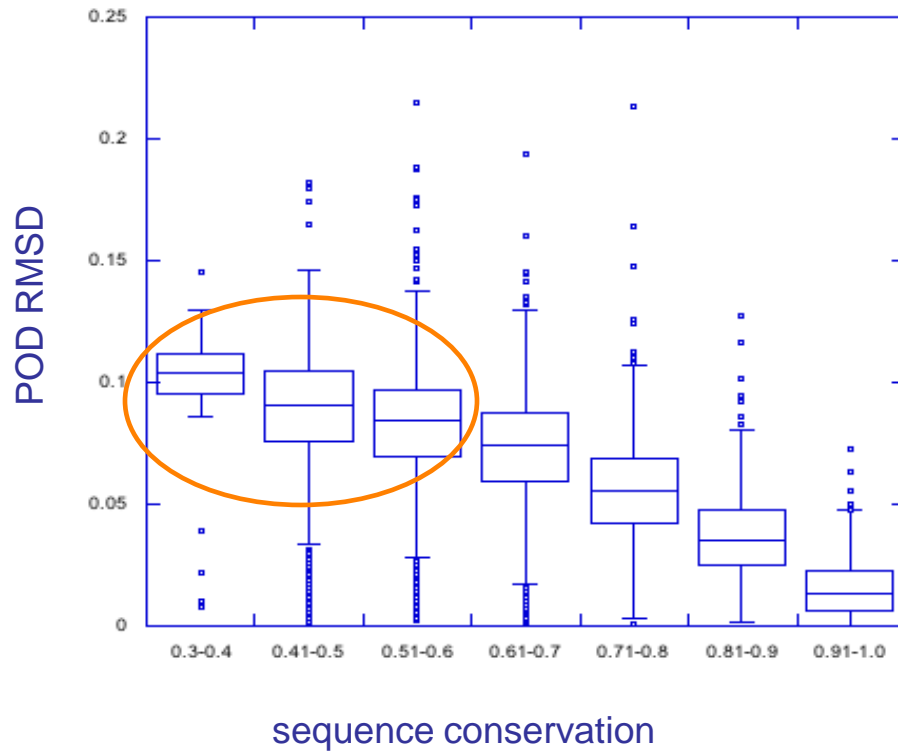
POD changes vs protein function

conservation of topology



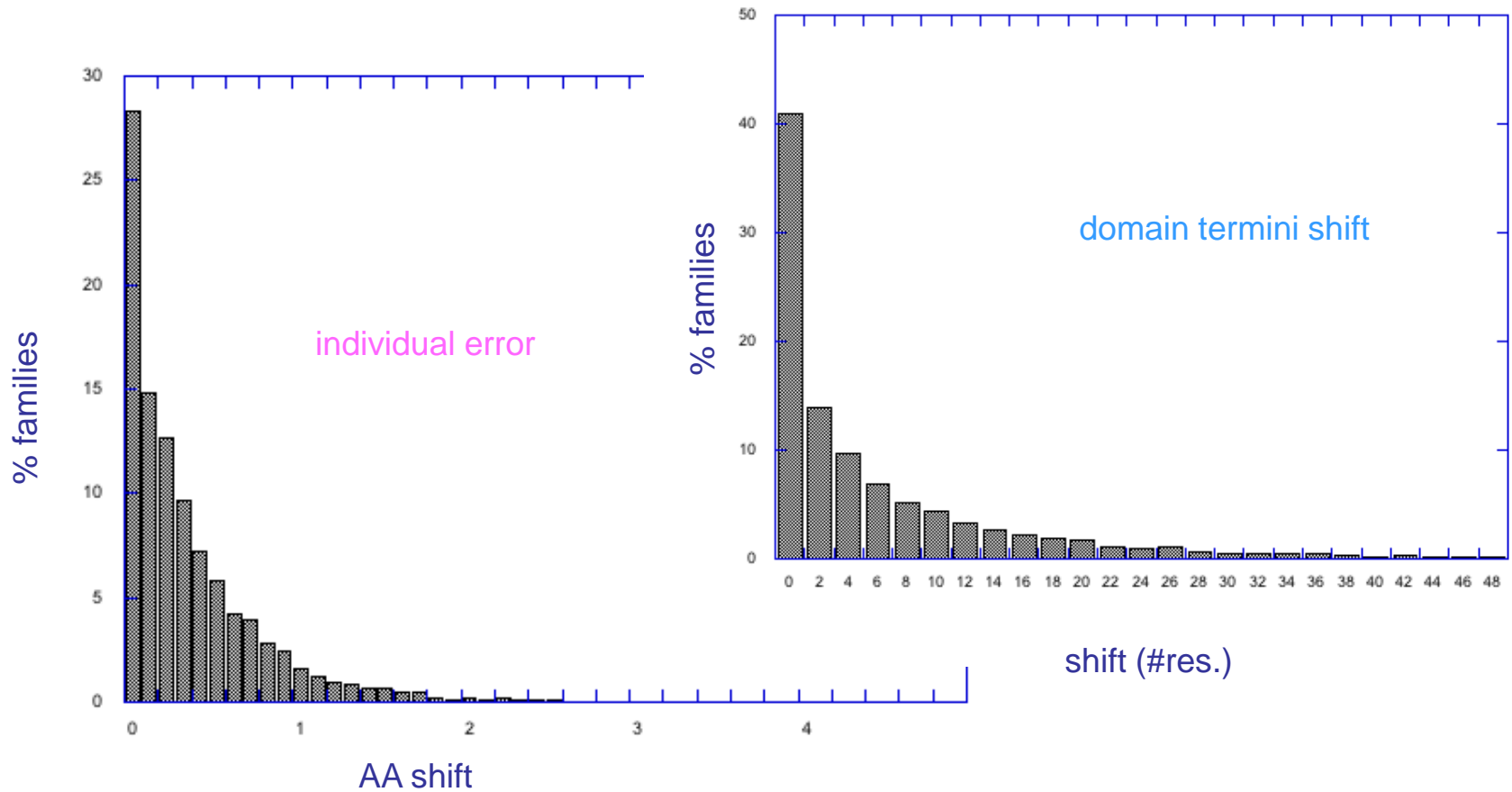
POD changes vs protein function

conservation of topology



POD changes vs protein function

Pfam - POD realignment



(Balibase benchmarks)