SMS in protein-protein communication

How **S**hort **M**otif **S**equences are read and understood

Ora Schueler-Furman

Faculty of Medicine, Hebrew University, Jerusalem, ISRAEL

Albert Szent-Györgyi Lecture, Eötvös Loránd University, Budapest 31/5/2016

Albert Szent-Györgyi

- So what the school has to do, in the first place, is to make us **learn how to learn**, to whet our appetites for knowledge, to teach us the delight of doing a job well and the excitement of creativity, to teach us to love what we do, and to help us to find what we love to do.

From: Teaching and expanding the knowledge (1964) Science ***3649****:*1278

Communication by SMS

- Short Message Service
- Short Motif Sequences = peptides
- enhance communication

Communication by SMS

Communication may be :

- short
- regulated

Herbie (ERBB4) 1/3/2016 15:00
Hey Chap, R U free?
Give me a ride? ✋

Herbie (ERBB4) 1/3/2016 20:00
… I'm w8ing … 🙌

Chap (YAP1) 1/3/2016 21:00 🚗
Yep, I got a space 💺

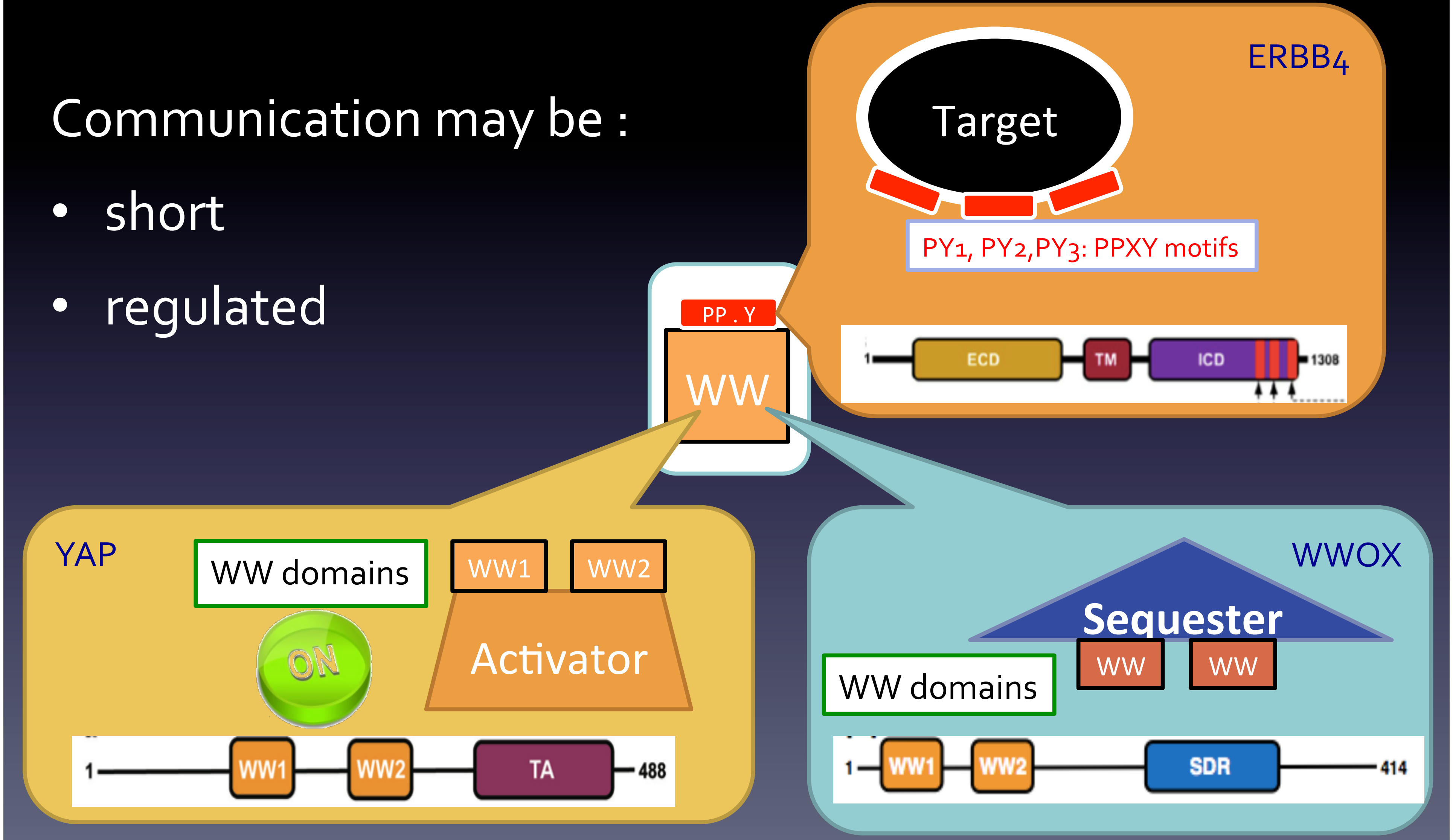
WOW (WWOX) 💺🍵
1/3/2016 21:00
Herbie, dinner!

Herbie (ERBB4) 1/3/2016 21:05
Sorry Chap - 2 l8 😟

Communication by SMS

Communication may be :

- short
- regulated



ERBB4

Target

PY1, PY2,PY3: PPXY motifs

PP . Y

WW

YAP

WW domains

ON

WW1

WW2

Activator

WWOX

Sequester

WW domains

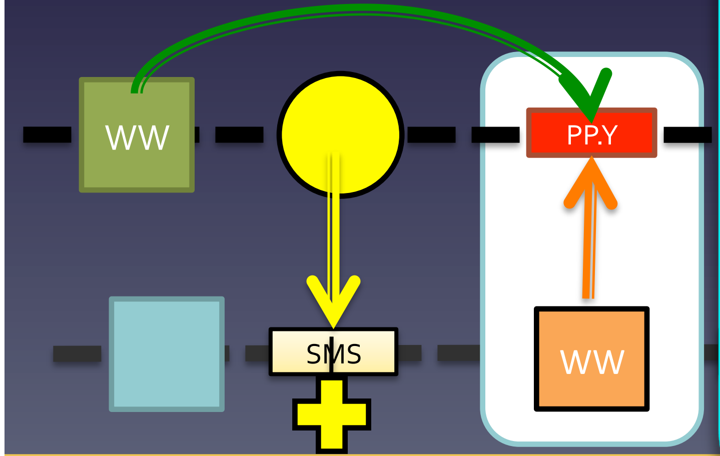
WW

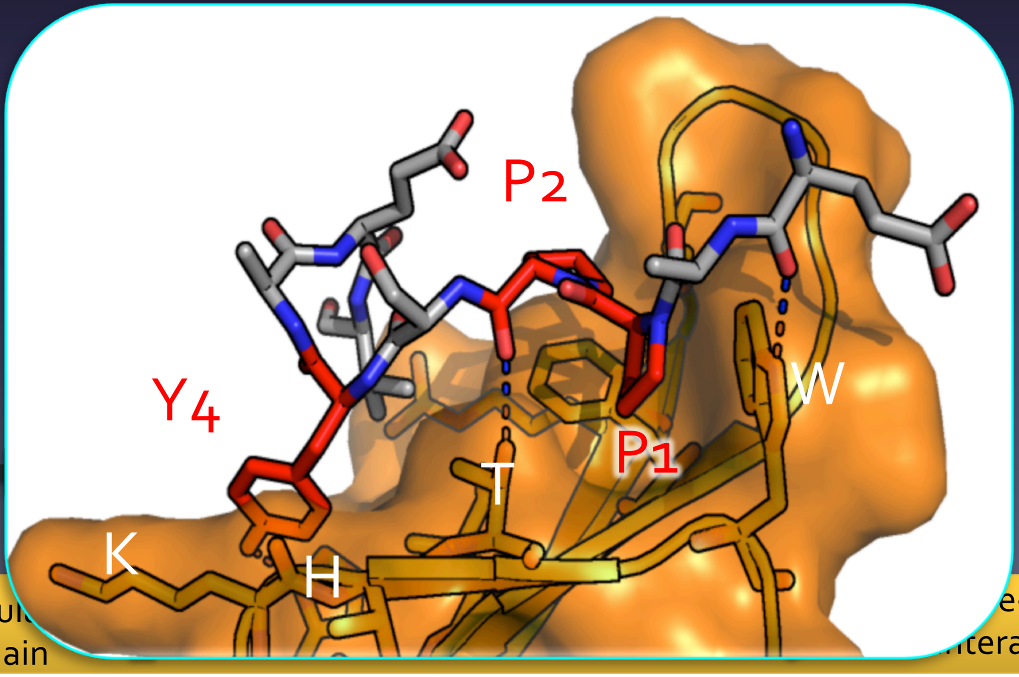
WW

SMS mediate many interactions

- transient
- regulation
- trans / cis
- PTM

- **sequence-motifs (SMS)**
- **structure :**
  - details of binding
  - drug design



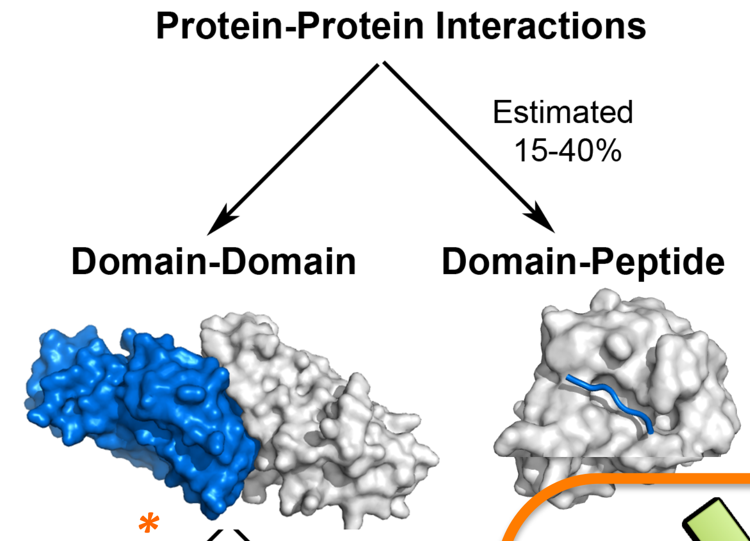


▬ ▬ unstructured region

☐ globul domain

e-protein nteraction

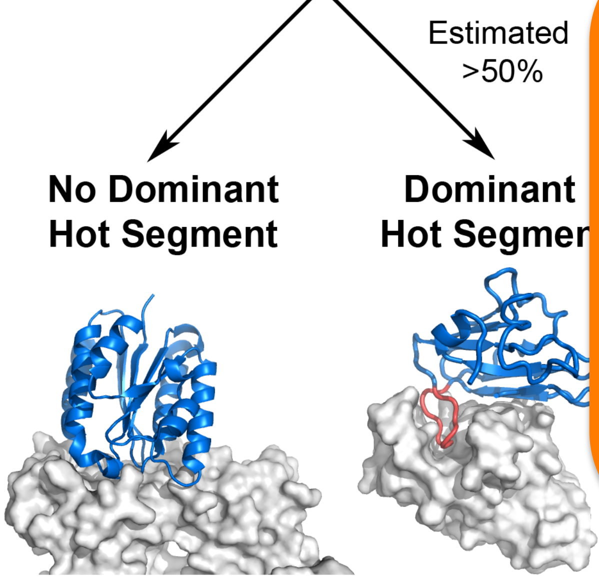
SMS mediate many interactions



Protein-Protein Interactions

Estimated 15-40%

Domain-Domain **Domain-Peptide**

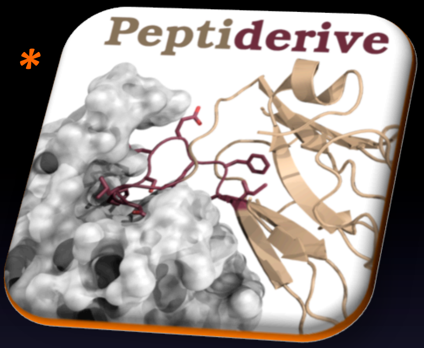


*

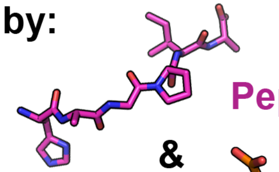
Estimated >50%

No Dominant Hot Segment

Dominant Hot Segmen

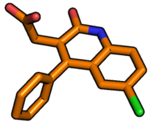


Amenable to inhibition by:

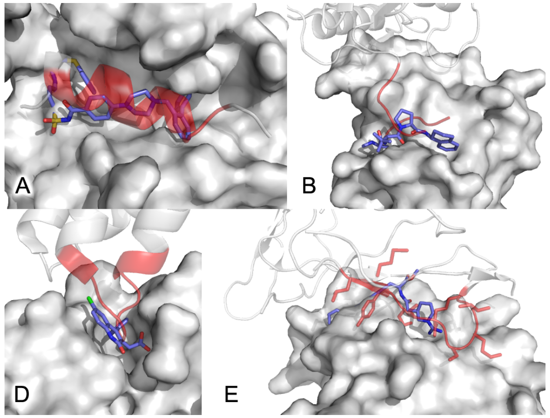


Peptides

&



Small molecules



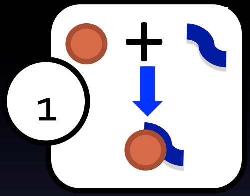
** London, Raveh, Movshovitz-Attias, Schueler-Furman. Proteins (2010); http://rosie.rosettacommons.org/peptiderive, Sedan & Marcu, NAR (2016)*

London, Barak and Schueler-Furman. Current Opinion Chemical Biology (2013)

Peptide-protein interactions - Aims

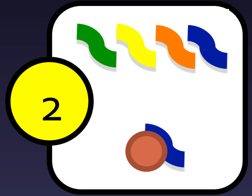
1. define structure of peptide-protein interaction

   Rosetta FlexPepDock & Cluspro PeptiDock

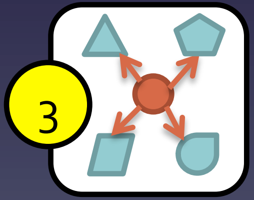


2. predict binding affinity & specificity

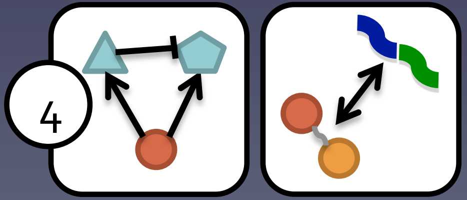
   Rosetta FlexPepBind



3. detect new interactions



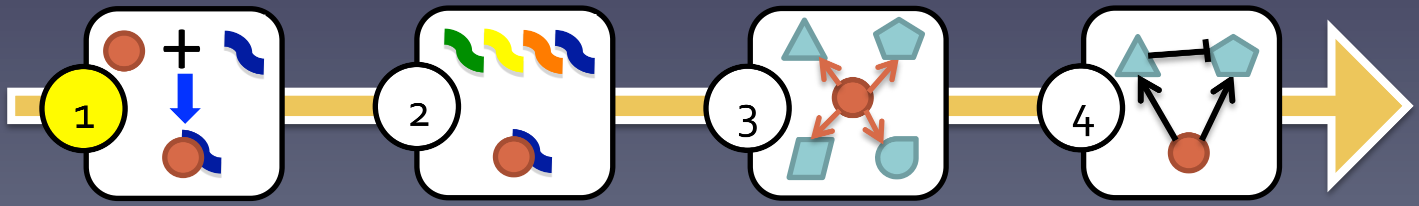
4. put things into context



Model structure: FlexPepdock

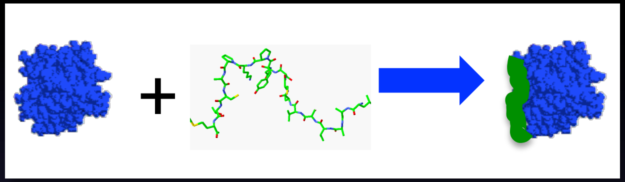
Nawsad Alam
Orly Marcu

Nir London
Barak Raveh
Lior Zimmerman



The challenge of peptide binding

Nature's challenge



$$\Delta G_{BIND} = \Delta H - T\Delta S$$

✧ **small** interface

✧ **flexible** peptide

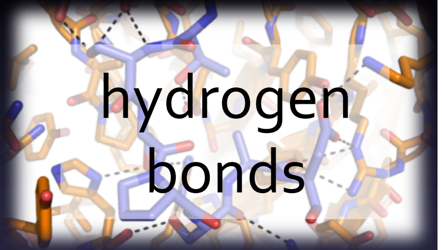
➢ how do peptides do it?

peptiDB

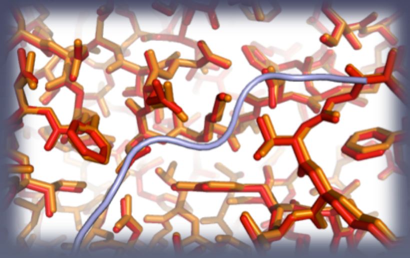
London, Movshovitz-Attias, Schueler-Furman (2010) Structure

Nature's solutions

① increase ΔH



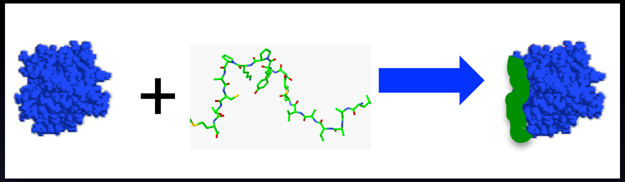
② decrease ΔS



prearranged receptor

The challenge of peptide binding

Our challenge



$\Delta G_{BIND} = \Delta H - T\Delta S$

✧ **small** interface

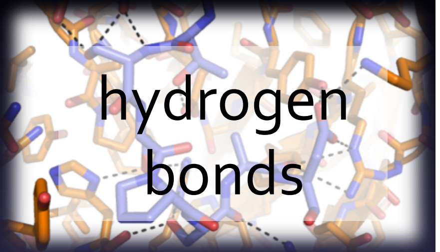
✧ **flexible** peptide

➢ how do we model these interactions?

Nature's solutions

① increase ΔH

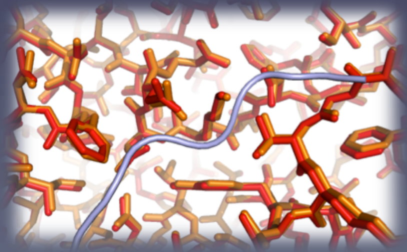

packing


hydrogen bonds


salt bridges

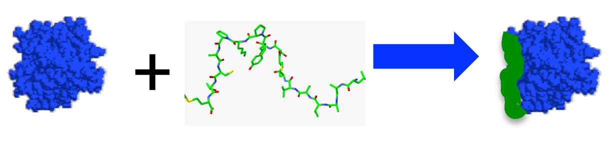

hot spots

② decrease ΔS


prearranged receptor

The challenge of peptide binding

Our challenge



$\Delta G_{BIND} = \Delta H - T\Delta S$

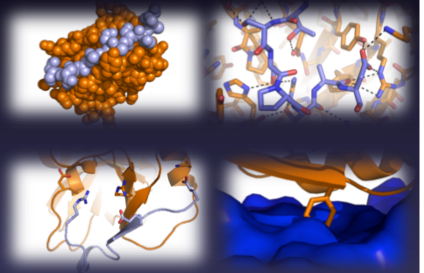
✧ **small** interface

✧ **flexible** peptide

➢ how do we model these interactions?

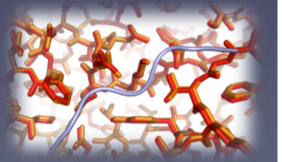
Our solution

① Extensive optimization of interface



② Restrict sampling to relevant space

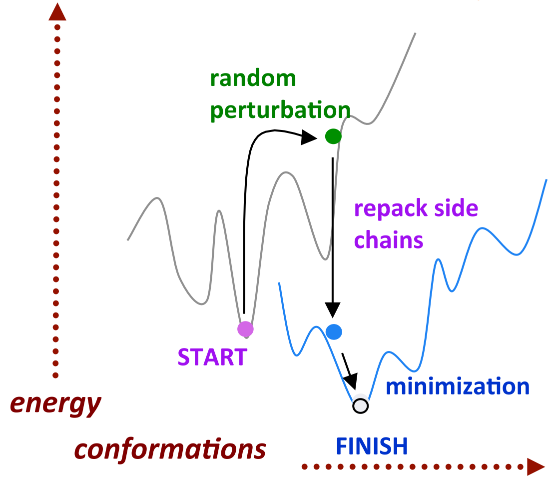
➢ focus on peptide



Rosetta infrastructure for structural modeling

Sampling and optimization

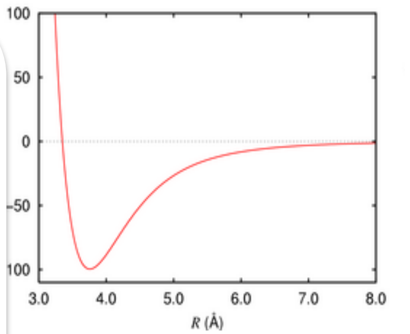
“**M**onte-**C**arlo search with energy **M**inimization” **MCM** (Liu *et al.*)



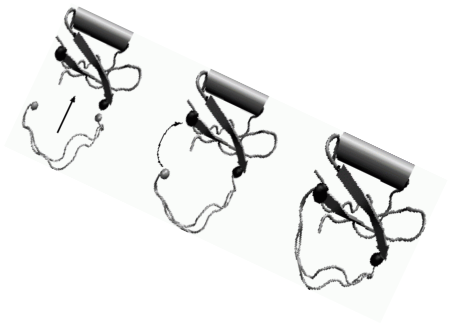
FlexPepDock protocol for flexible peptide docking and redesign

Energy function

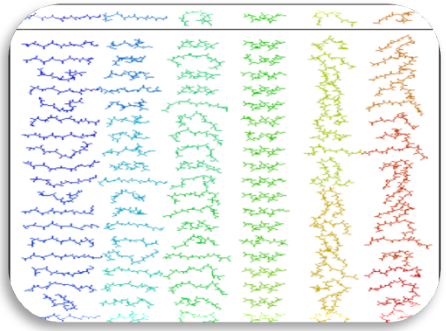
(full-atom/centroid side-chains)



Loop modeling

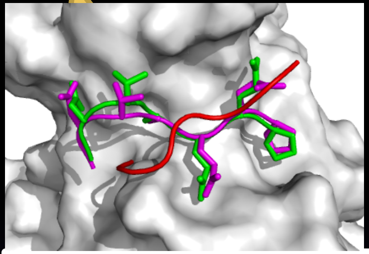


Fragment libraries



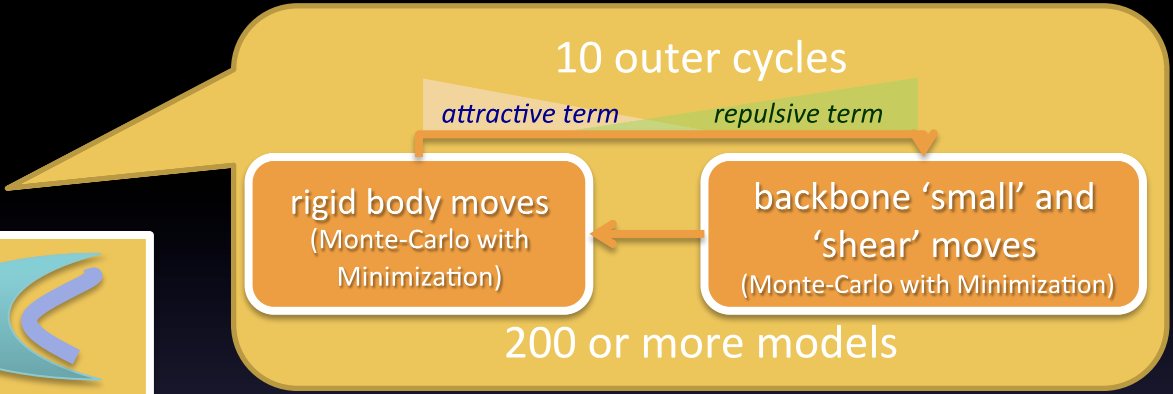
Raveh, et al. Proteins (2010)
Raveh et al. PLosOne (2011)
London et al. Current Opinion in Structural Biology (2013)

Rosetta FlexPepDock protocol



FlexPepDock

http://flexpepdock.furmanlab.cs.huji.ac.il



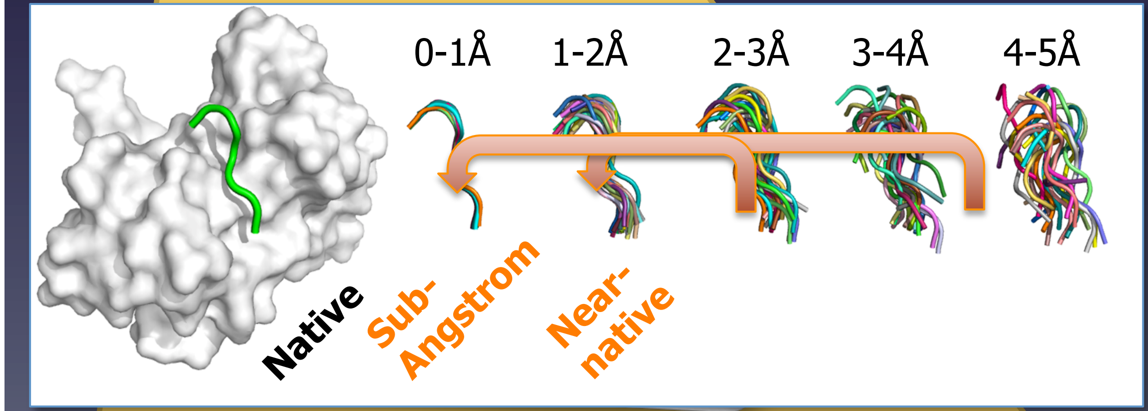
10 outer cycles

attractive term *repulsive term*

rigid body moves
(Monte-Carlo with Minimization)

backbone ‘small’ and ‘shear’ moves
(Monte-Carlo with Minimization)

200 or more models



General:

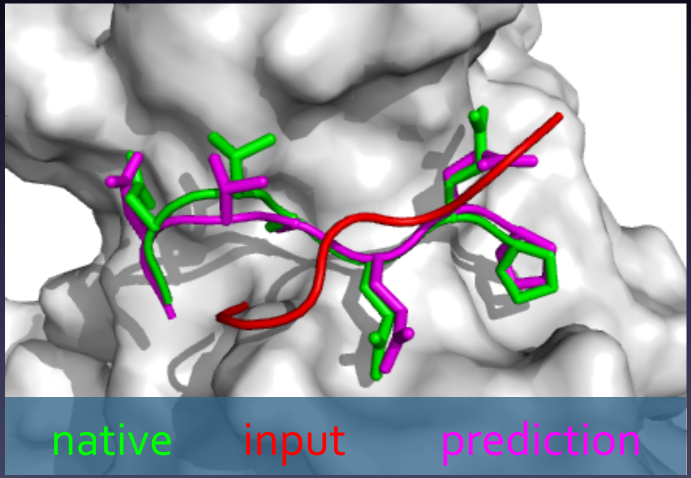
tested on large benchmark

peptiDB

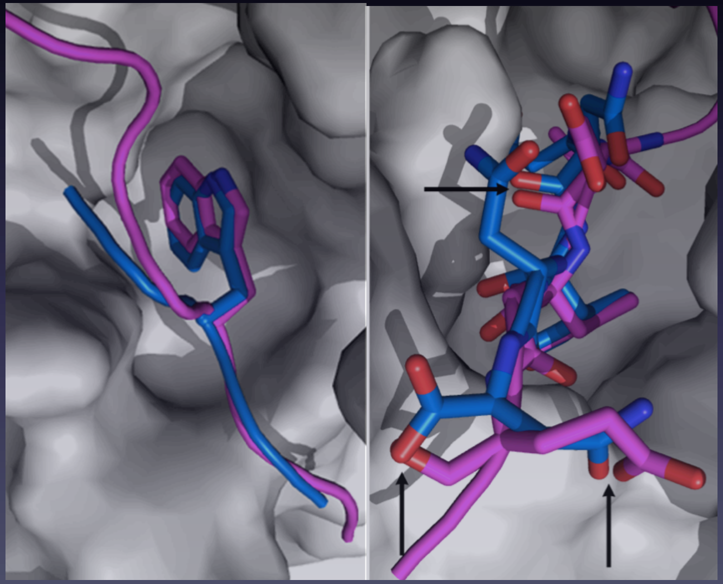
- Short peptides 5-15 aa
- High resolution < 2Å
- Non redundant < 70% seq id

Rosetta FlexPepDock

accurate prediction of motif residues



backbone rearrangements preserve critical interactions



- try our server :
http://flexpepdock.furmanlab.cs.huji.ac.il/

flex pep dock

London, Raveh et al. NAR (2011)

Impact of FlexPepDock

➜ first dedicated peptide-docking protocol

➜ motivated development of many new approaches

➜ CAPRI (docking competition) looks at peptides

➜ to appear soon: THE book on peptide docking

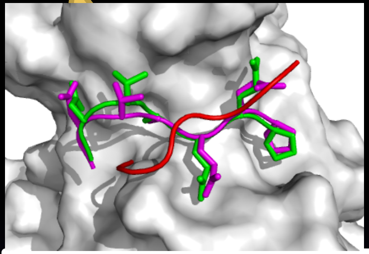
Springer Protocols

Methods in Molecular Biology

Humana Press

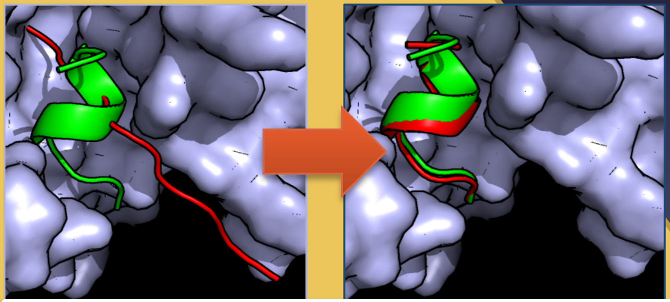
edited with Nir London

Extension of Rosetta FlexPepDock

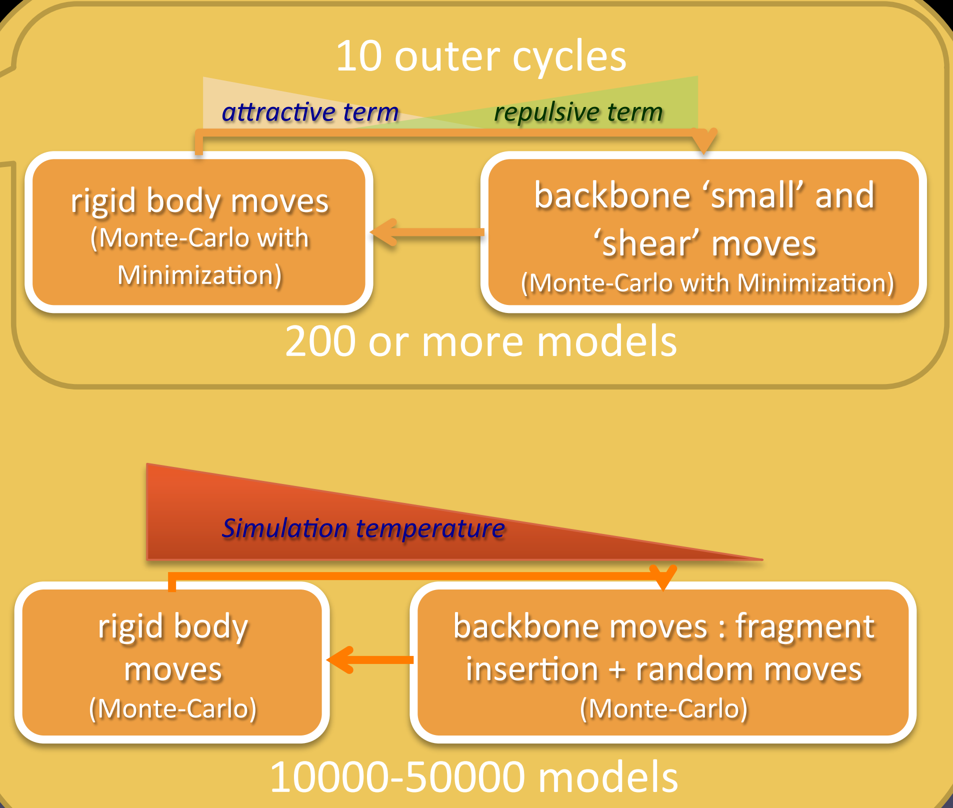


FlexPepDock

http://flexpepdock.furmanlab.cs.huji.ac.il



Ab initio FlexPepDock



10 outer cycles

attractive term *repulsive term*

rigid body moves
(Monte-Carlo with Minimization)

backbone 'small' and 'shear' moves
(Monte-Carlo with Minimization)

200 or more models

Simulation temperature

rigid body moves
(Monte-Carlo)

backbone moves : fragment insertion + random moves
(Monte-Carlo)

10000-50000 models

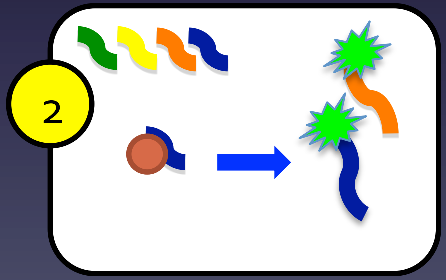
- **Full *ab initio***

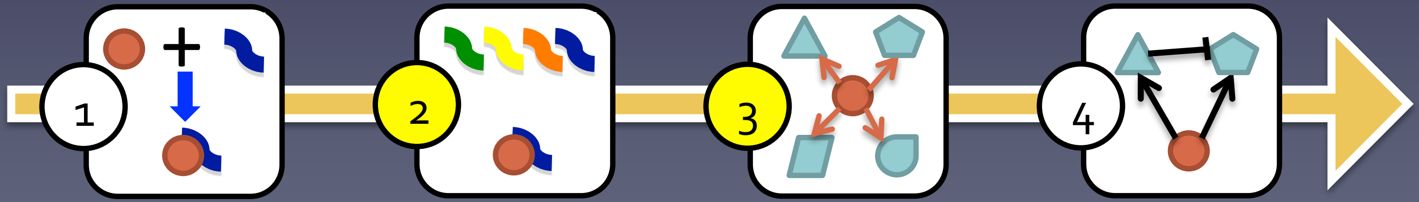
Model binding partners: FlexPepBind

Nawsad Alam

Nir London
Lior Zimmerman
Michal Sperber

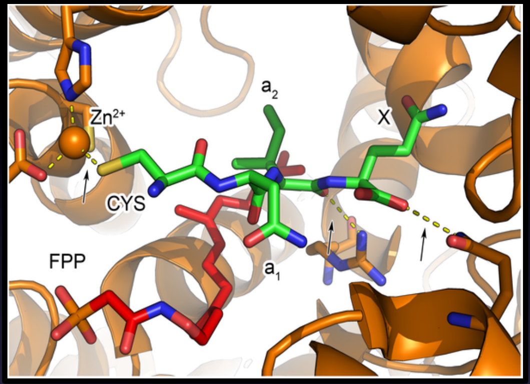
Experiments:
Noah Wolfson
Fierke lab, U Michigan





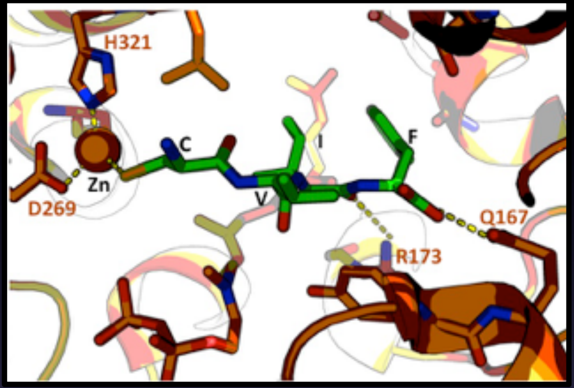
FlexPepBind applications

Farnesyl transferase



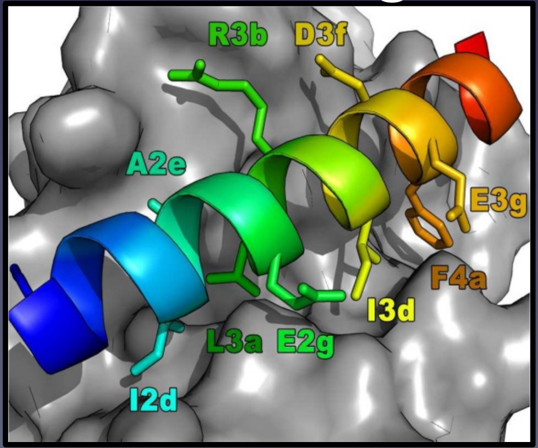
London *et al. PLoS Comput Biol* (2011)

Geranyl-geranyl transferase



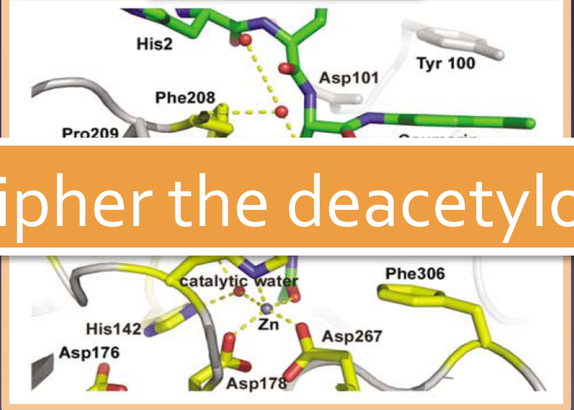
Sperber *et al., in preparation*

Bcl2-BH3



London *et al. Biochemistry* (2012)

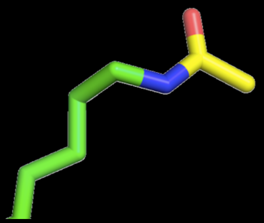
HDAC8



Decipher the deacetylome

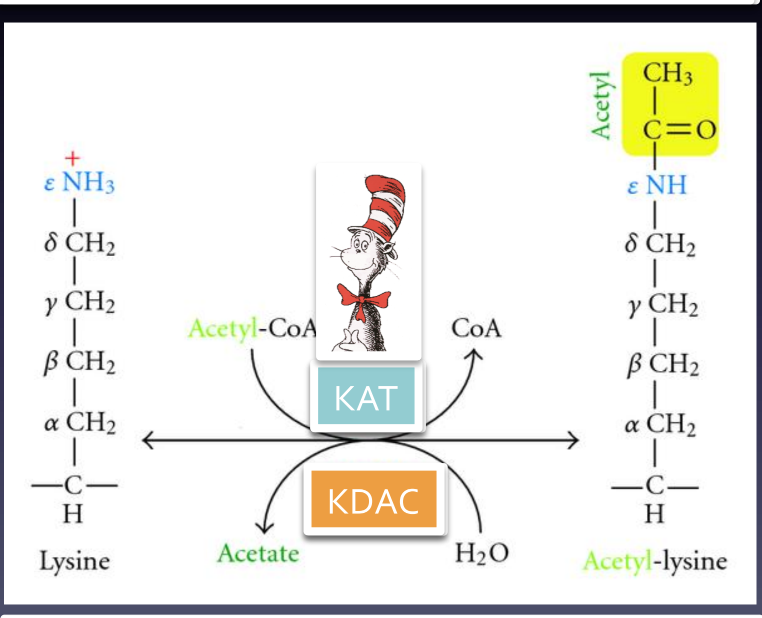
Alam *et al. Structure* (2016)

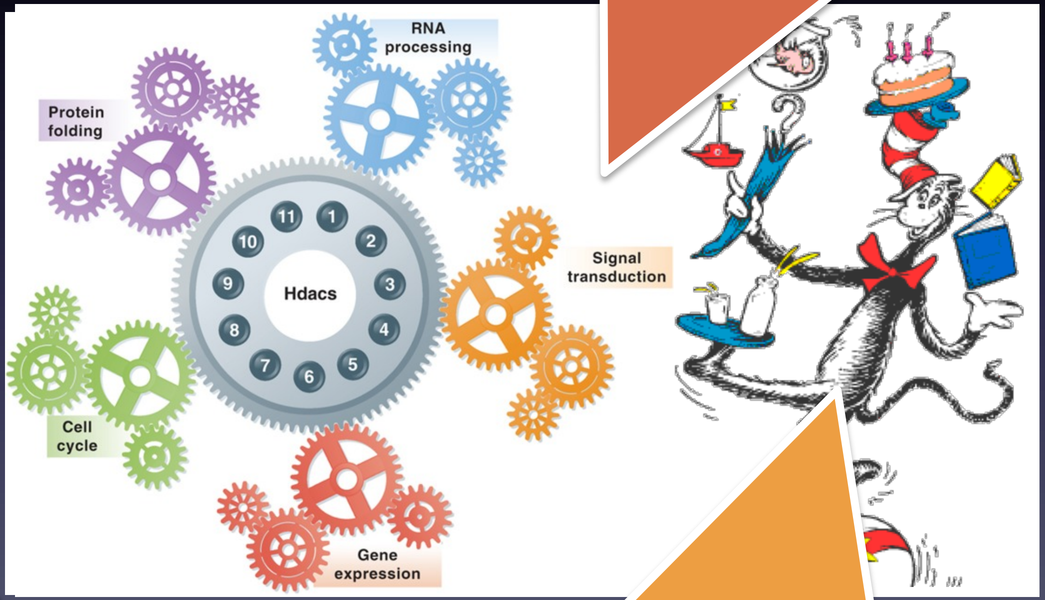
Lysine (de)acetylation



Lysine acetyl transferase

Can we find more substrates?

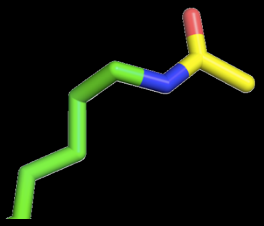


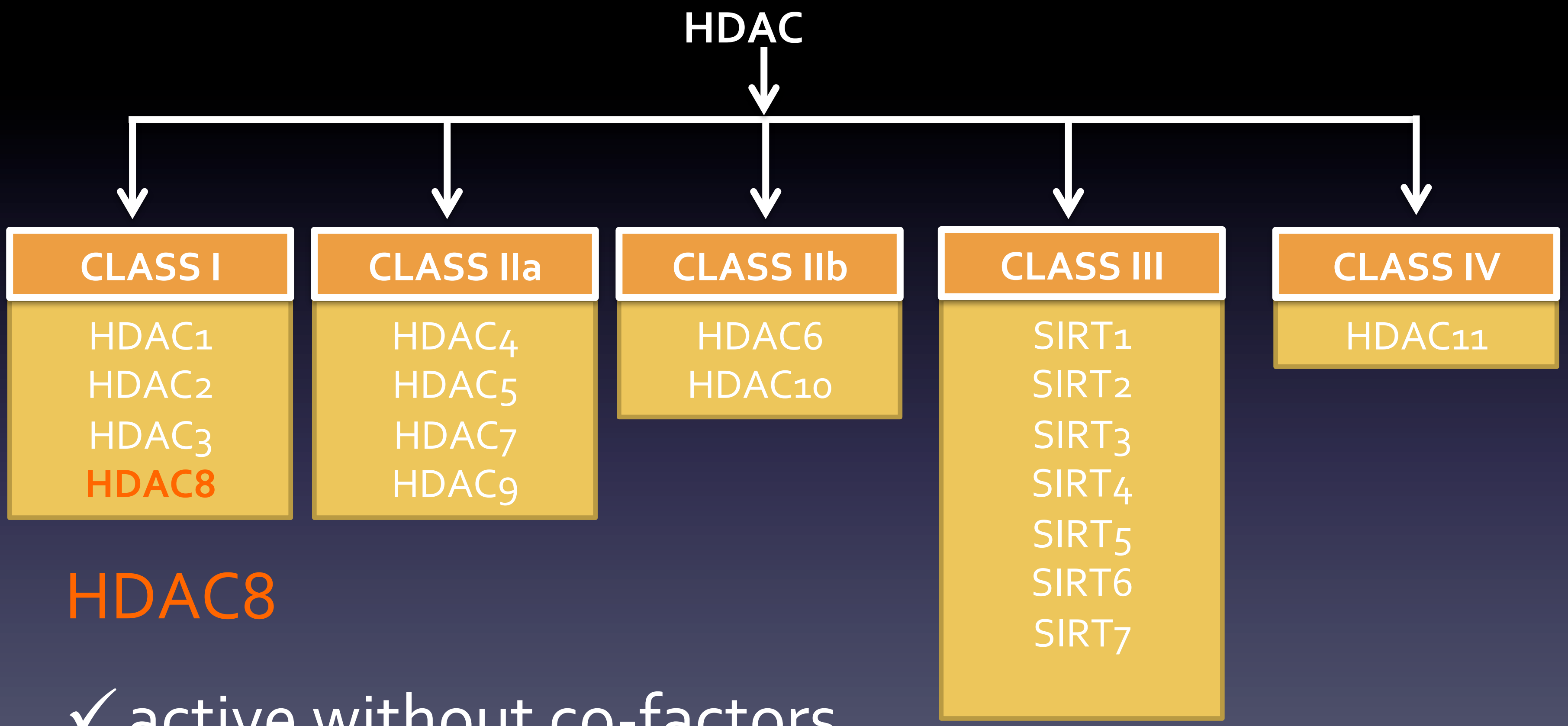


Lysine deacetylase

“Oh no, that is not all I can do” said the Kat in the Hat…

Lysine deacetylation





HDAC

CLASS I	CLASS IIa	CLASS IIb	CLASS III	CLASS IV
HDAC1	HDAC4	HDAC6	SIRT1	HDAC11
HDAC2	HDAC5	HDAC10	SIRT2	
HDAC3	HDAC7		SIRT3	
HDAC8	HDAC9		SIRT4	
			SIRT5	
			SIRT6	
			SIRT7	

HDAC8

✓ active without co-factors

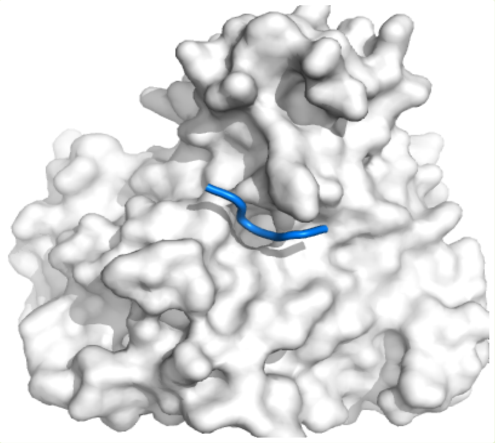
✓ related to important diseases

Known substrates:
P53, SMC3, ERRα, ARID1A, Cortactin 9, Inv(16)

FlexPepBind Framework

1 INPUT

- peptide-protein complex structure

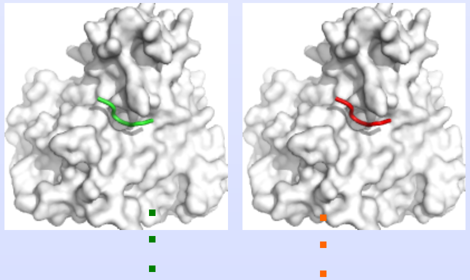


- training set

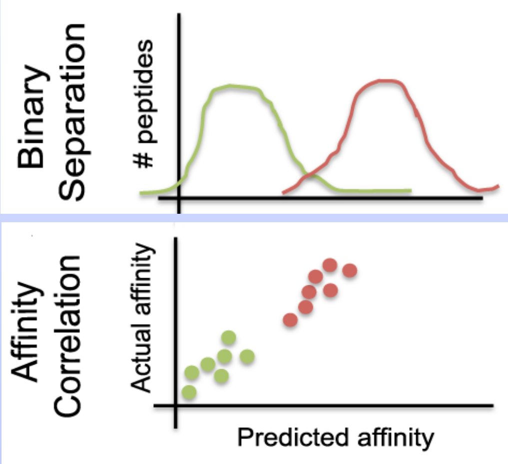
Binders	Non-Binders
GYKacFGC	GWKacMGC
GIKacWGC	GMKacGGC
GRKacFGC	GKKacKGC
⋮	⋮

2 CALIBRATION

- model set
- adjust protocol



- optimize for:



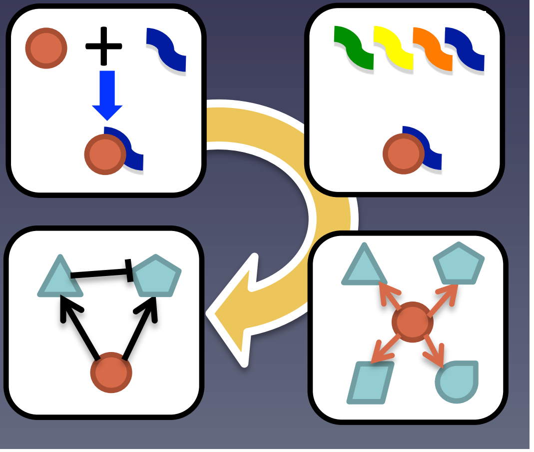
3 PREDICTION

- large-scale application

RESULTS

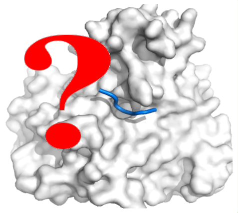
- identify new protein targets from discovered peptide substrates

**...ERFEMFRELNEALELKDAQ
AGKEPGGSRAHSSHLKSKKG
QSTSRHKKLMFKTEGP...**



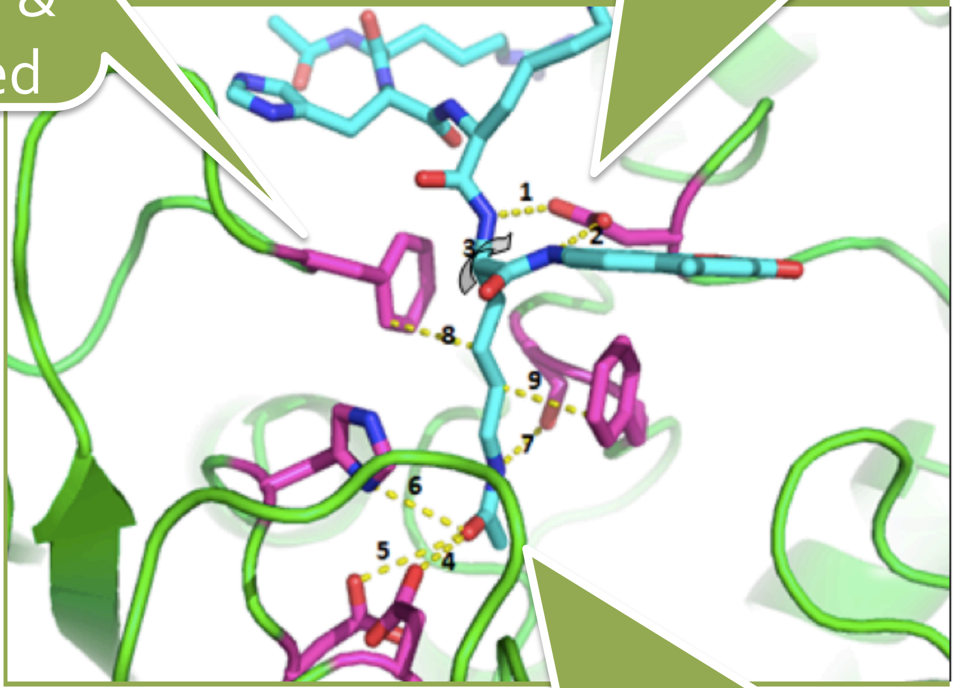
HDAC8 substrate hunt: input

1 **INPUT**



K_{Ac} side chain planar & stacked

K_{Ac} backbone fixed by 2 hydrogen bonds to D101



K_{Ac} acetyl coordinated by Zn ion

2v5w: HDAC8 - K-acetylated p53-derived peptide with coumarine tail

binding site surrounded by loops

HDAC8 substrate hunt: input (II)

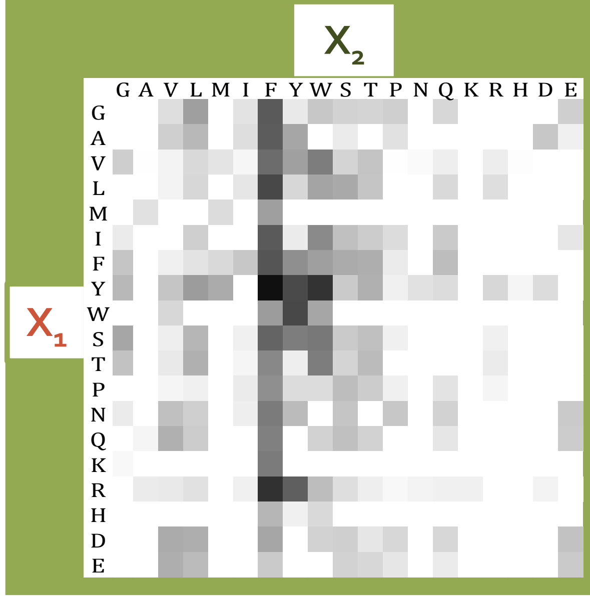
1

INPUT

Binders	?	Non-Binders
GYKacFGC		GWKacMGC
GIKacWGC		GMKacGGC
GRKacFGC		GKKacKGC

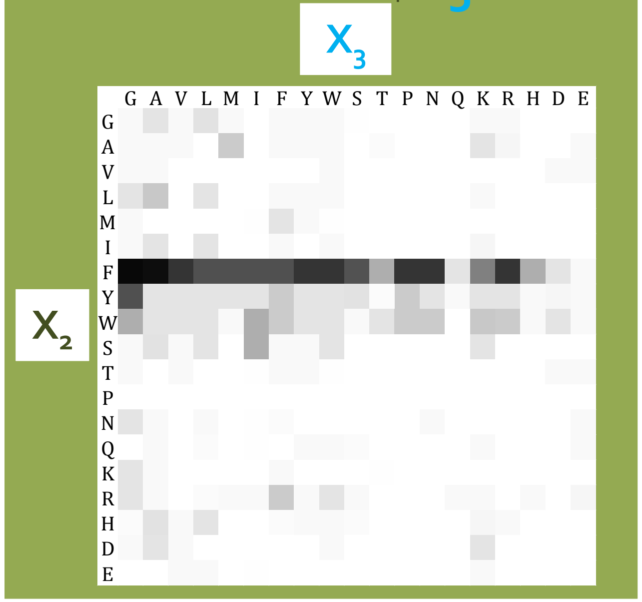
training set

GX_1KacX_2GC



test set

$GRKacX_2X_3GC$



SAMDI*

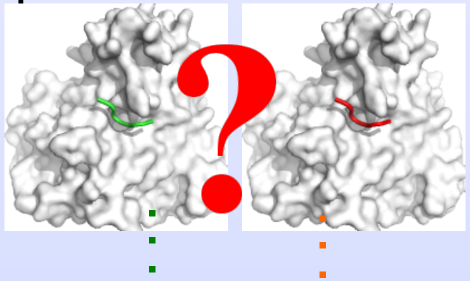
* **S**elf-**A**ssembled monolayers for **M**atrix assisted laser **D**esorption / **I**onization time-of-flight mass spectrometry

Gurard-Levin *et al. ACS Chem Biol.* (2010).

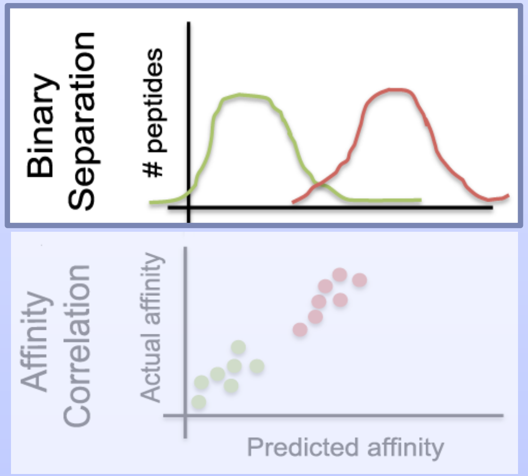
HDAC8 substrates: critical features

2 CALIBRATION

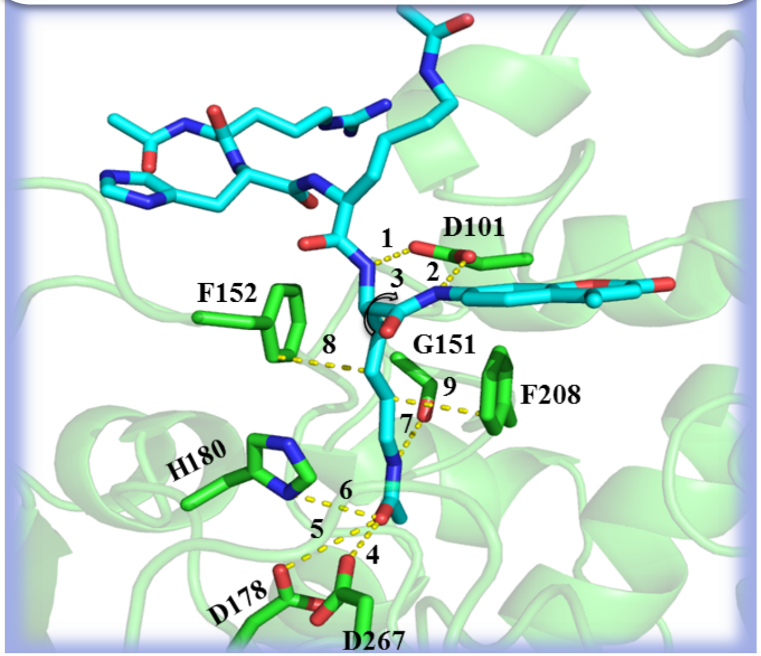
- model training set
- adjust protocol
- predict test set



optimize for:



reinforce critical features of interaction

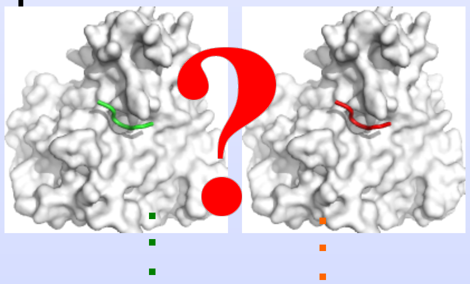


favorable binding ⇧ == ⇧ favorable catalysis

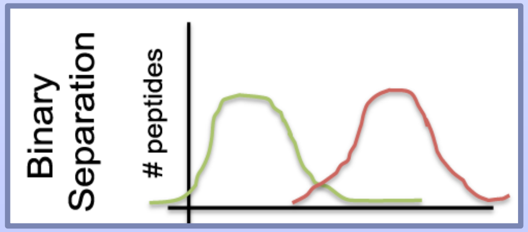
HDAC8 substrate hunt: optimize

② **CALIBRATION**

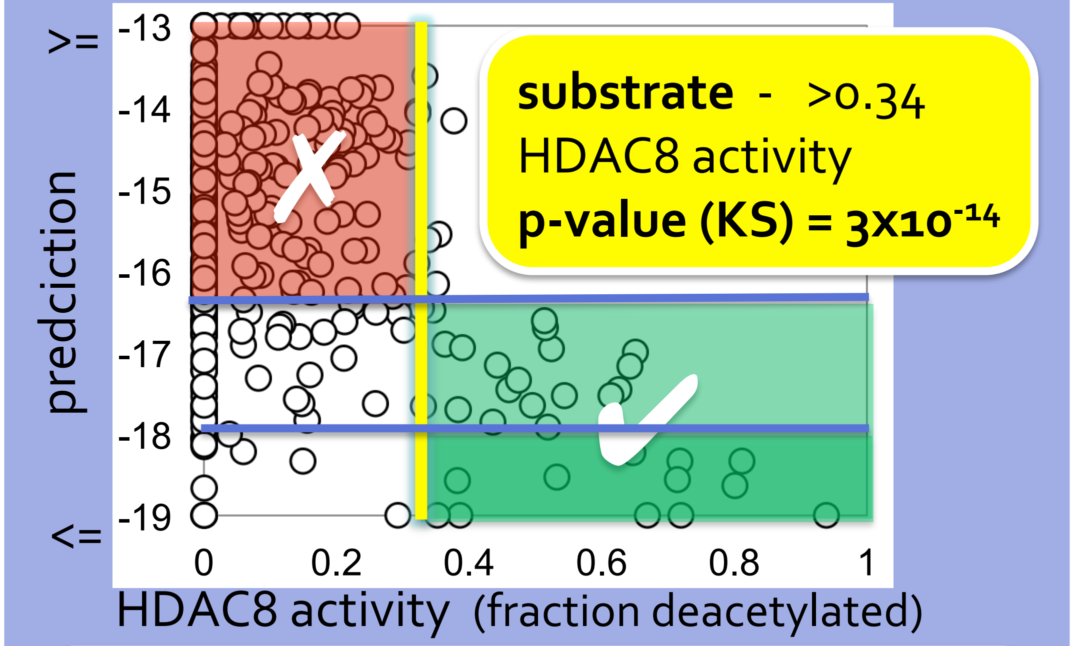
- model training set
- adjust protocol
- predict test set



optimize for binary separation



training set GX$_1$KacX$_2$GC



substrate - >0.34
HDAC8 activity
p-value (KS) = 3x10^{-14}

thresholds:

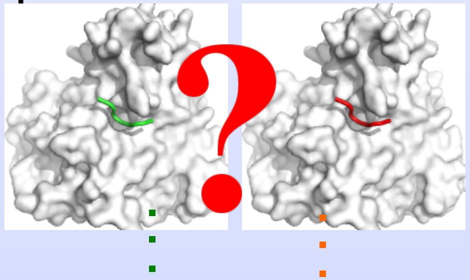
stringent <-18.0 **TP: 35% FP: 4%**

lenient <-16.5 **TP: 91% FP: 19%**

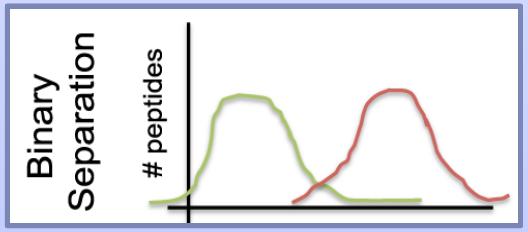
HDAC8 substrate hunt: validate

2 CALIBRATION

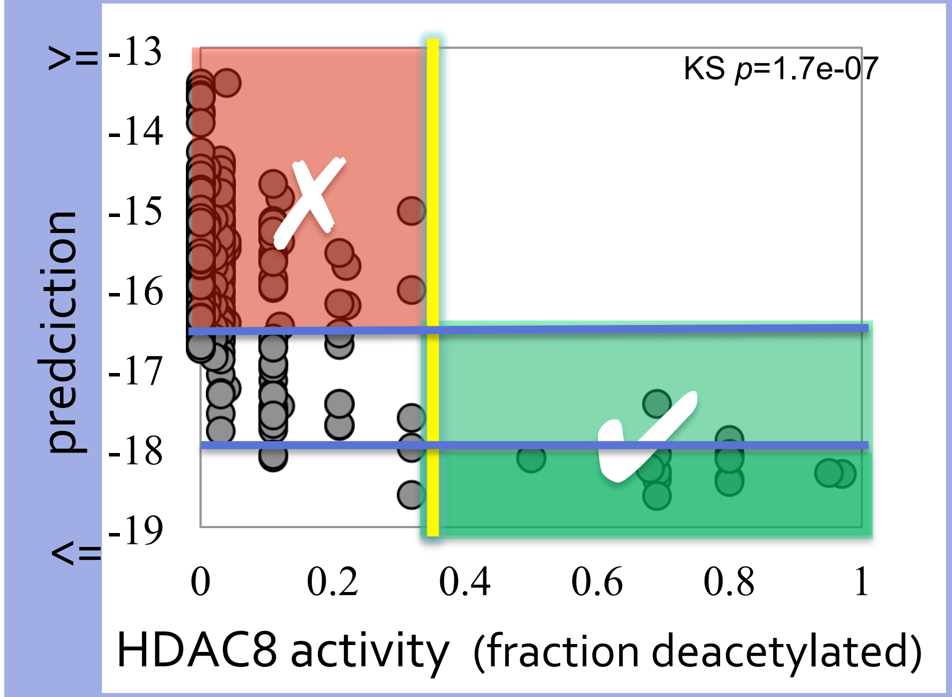
- model training set
- adjust protocol
- predict test set



optimize for binary separation



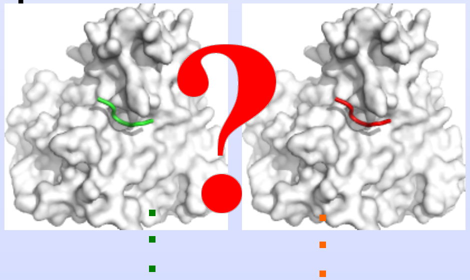
Test set GRKacX_2X_3GC



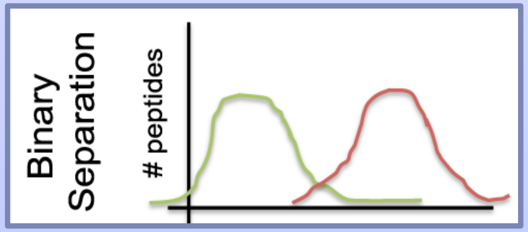
Summary of validation

2 CALIBRATION

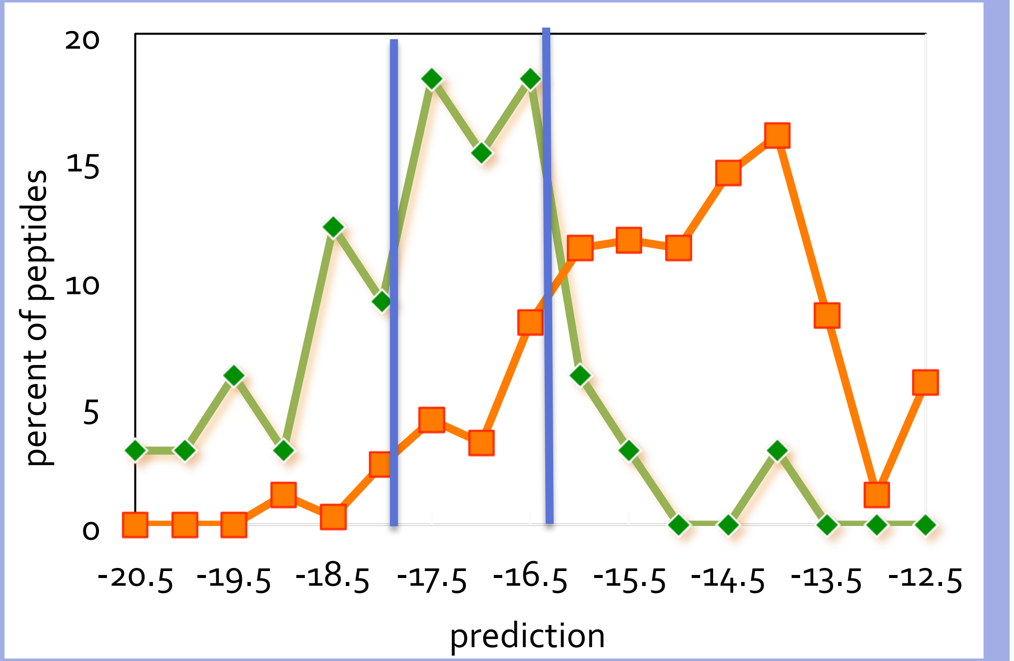
- model training set
- adjust protocol
- predict test set



optimize for binary separation



same, as histogram



Prediction of new HDAC8 substrates

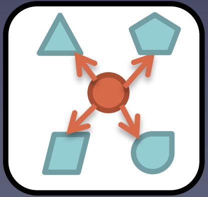
3 PREDICTION

- apply protocol to set of candidate sequences

RESULTS

- identify new protein targets from discovered peptide substrates

…ERFEMFRELNEALELKDAQ AGKEPGGSRAHSSHLKSKKG QSTSRHKKLMFKTEGP…



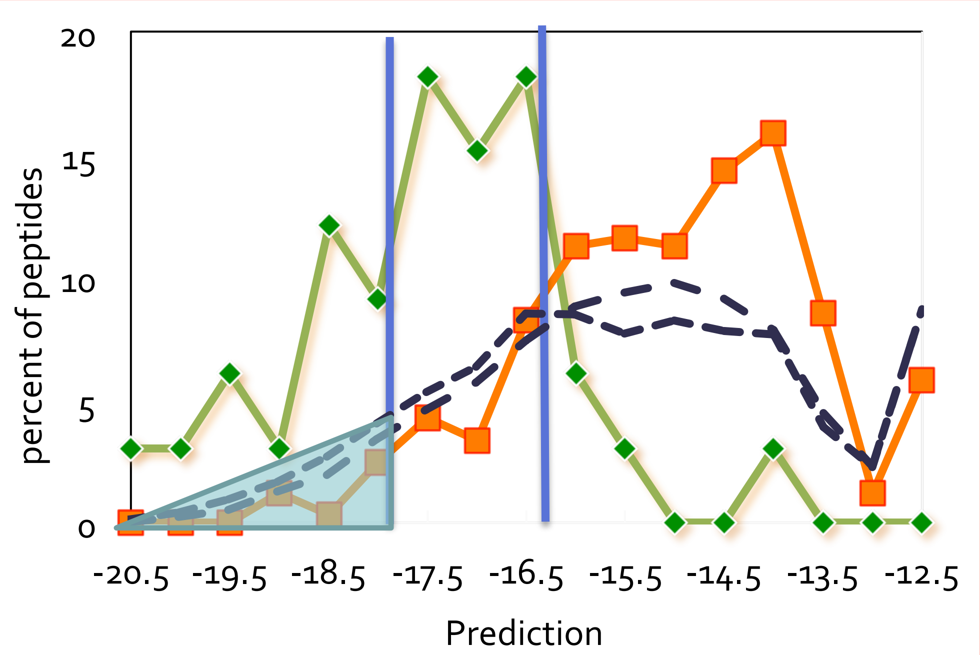
PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, ubiquitination, acetylation and methylation. See About PhosphoSite above for more information.

Please cite the following reference for this resource: Hornbeck PV, et al (2015) *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. Nucleic Acids Res. 43:D512-20. [reprint]

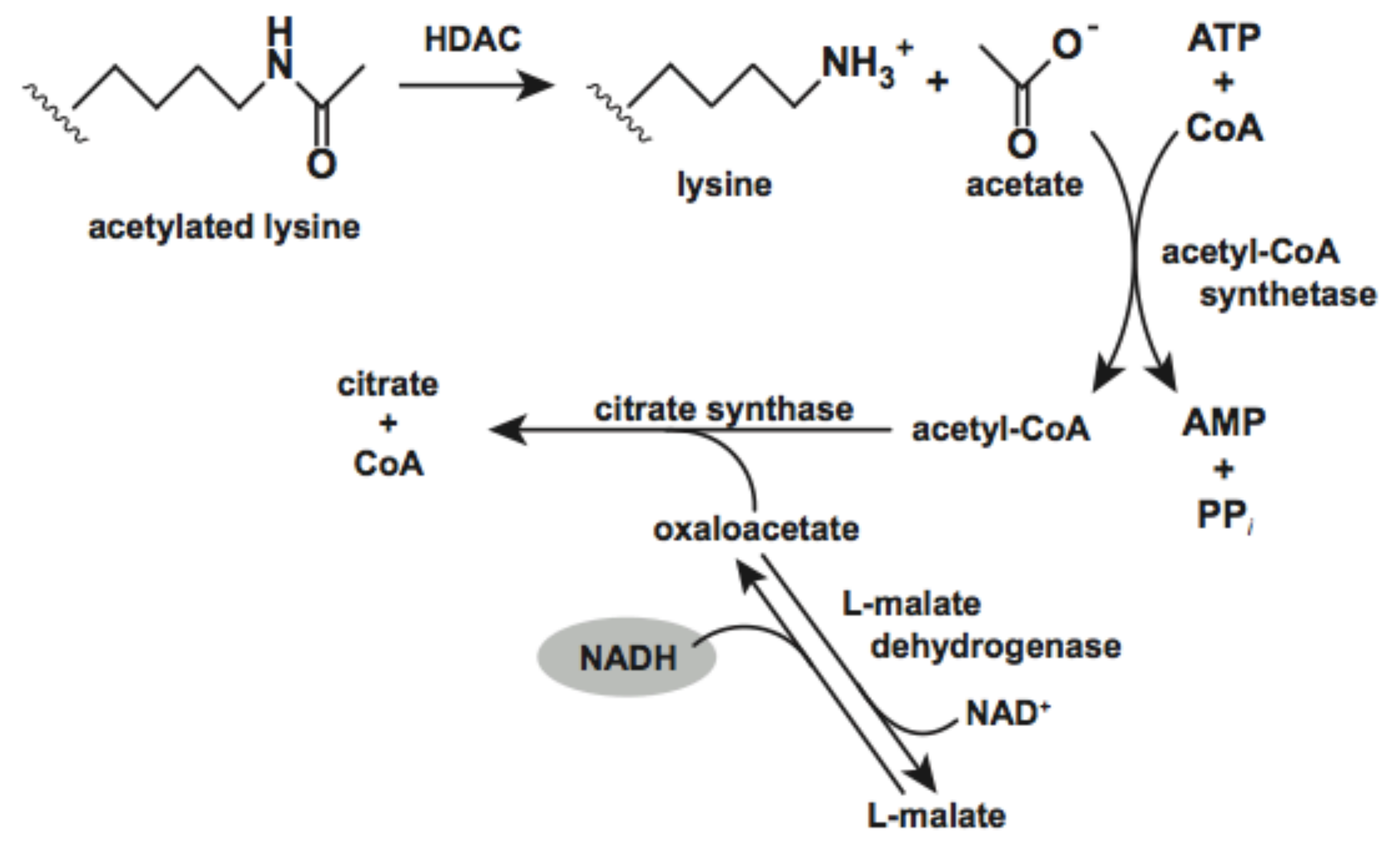
A PROTEIN MODIFICATION RESOURCE

Hornbeck *et al.* Proteomics. (2004)

PhosphoSite scanning: ~20'000 reported acetylated sites

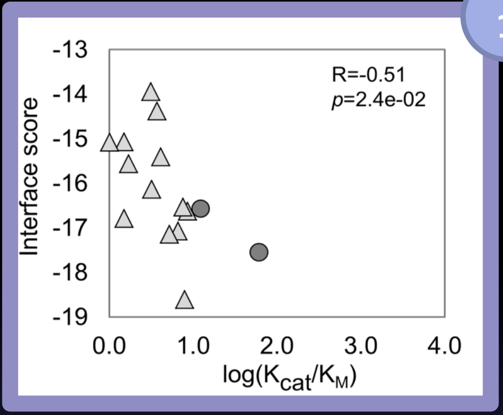


In vitro HDAC8 activity assay



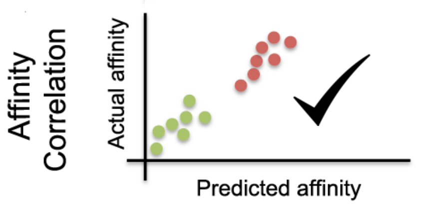
Wolfson et al., Analytical Biochemistry (2014)

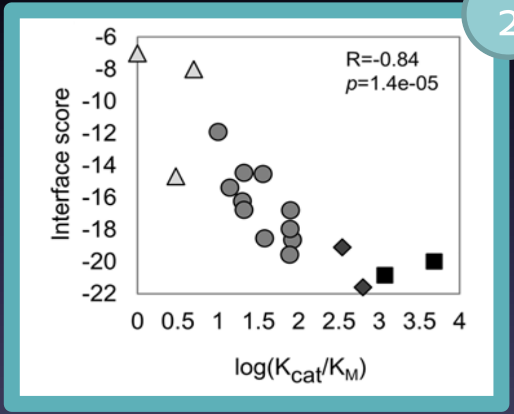
Prediction & experiment correlate



1

validation set 1
1st PhosphoSite screen
(n=20)





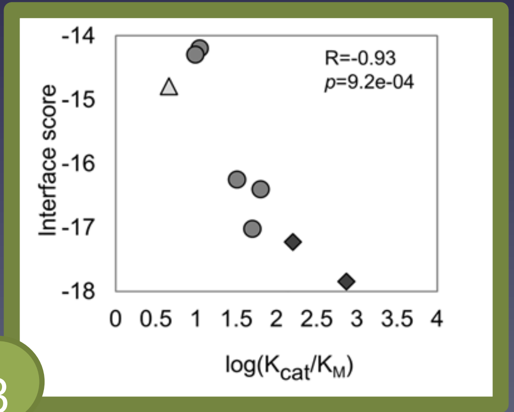
2

validation set 2
2nd PhosphoSite screen
(n=18)

prediction

reactive (k_{cat}/K_M >1000 $M^{-1}s^{-1}$) ■
reasonable (> 100 $M^{-1}s^{-1}$) ◆
mild (> 10 $M^{-1}s^{-1}$) ●
poor (> 1 $M^{-1}s^{-1}$) △

validation set 3
in vivo proteomic
assay* (n=8)

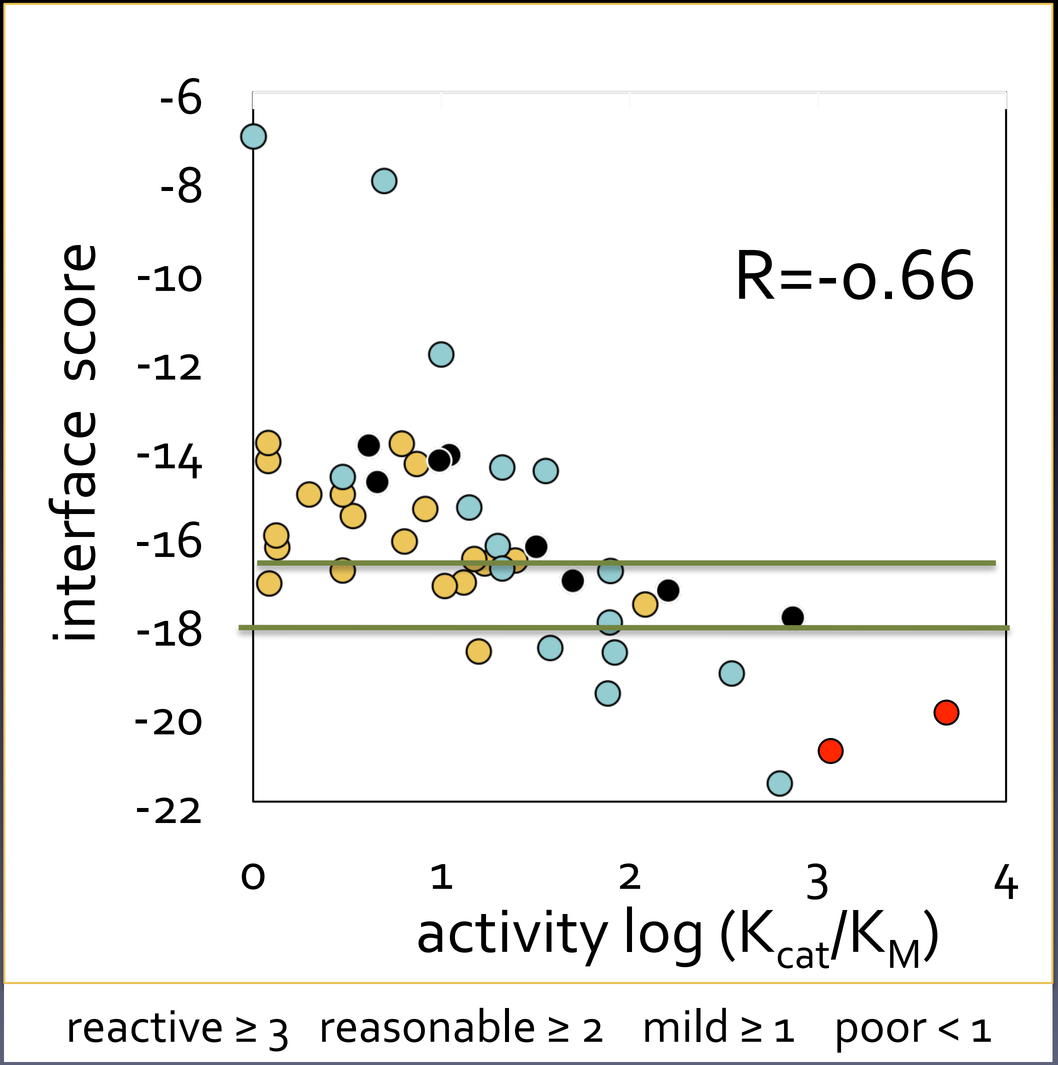


3

experiment

**Olson et al. ACS Chemical Biology 2014*

Summary of validation



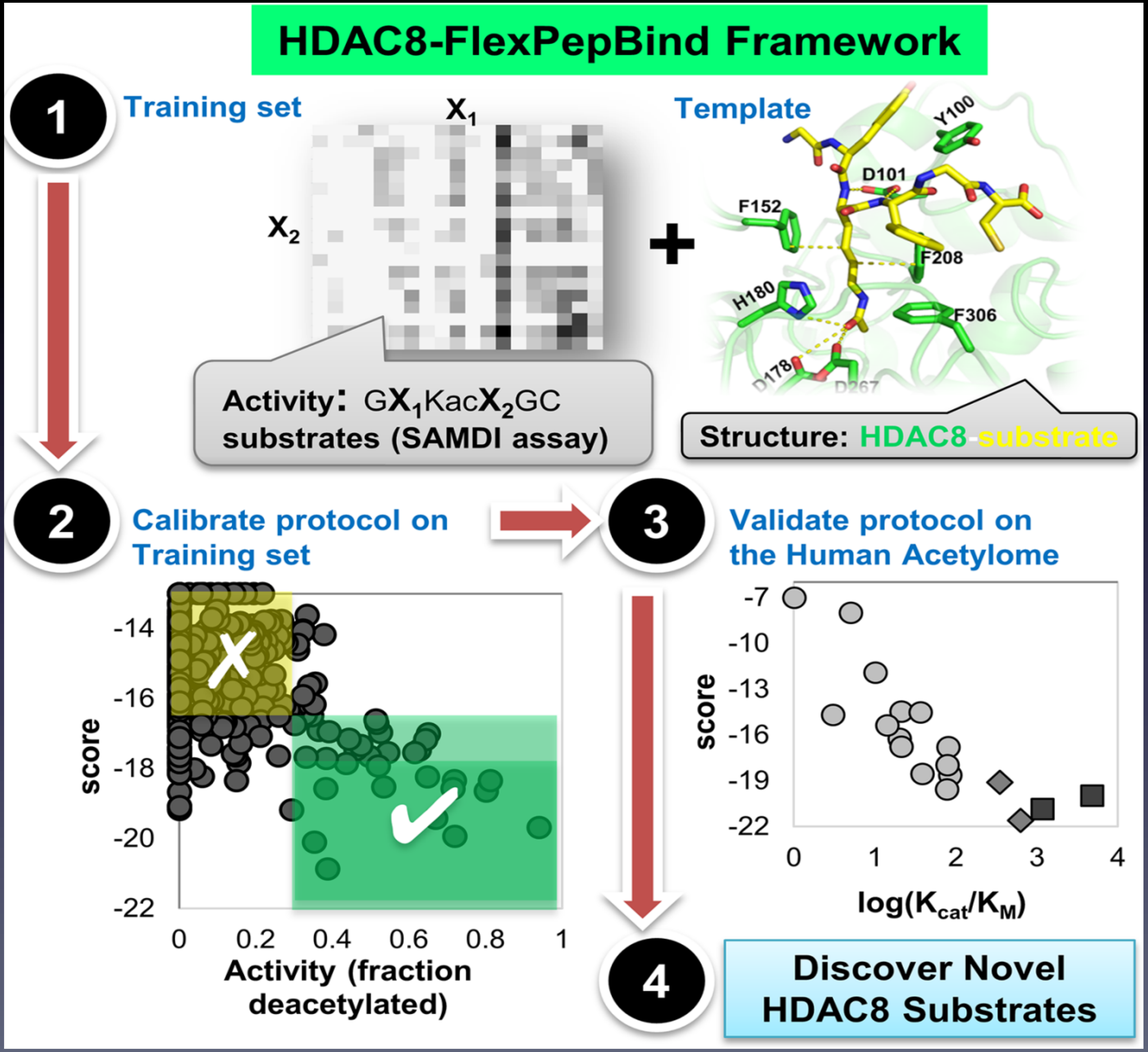
✓ good correlation *across* experiments

✓ stable thresholds

✓strongest measured HDAC8 substrates

- peptide reactive (>1000 K_{cat}/K_M)
- expected protein reactivity: x1000

Summary



Who are these substrates?

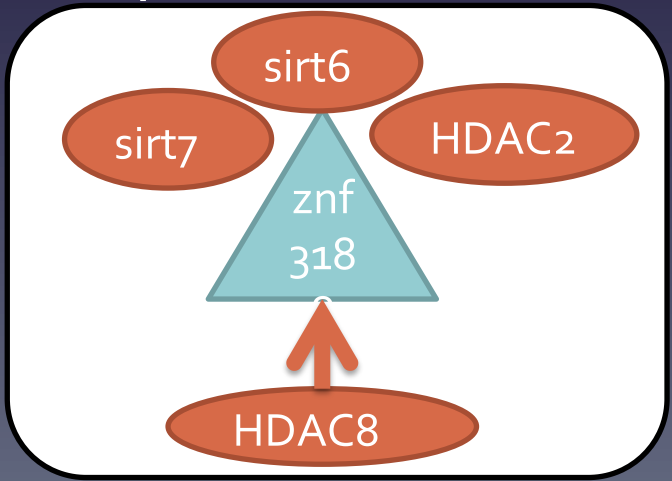
1

Strongest HDAC8 target: ZNF318

peptide	log (Kcat/K_M)	min. I_sc	protein
FG K1275 FSW	3.68	-20.0	Zinc finger protein 318

K1275: unstructured & accessible
- G1274R in breast cancer
- S1277 phosphorylated

ZNF318 represses Androgen Receptor
- related partners:
  - sirt6, sirt7 & HDAC2



reactive ≥ 3 **reasonable ≥ 2** mild ≥ 1 poor < 1

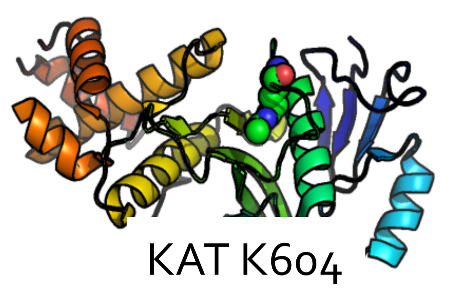
②

Targets: good old neighbors!

peptide	log (K_{cat}/K_M)	min. I_sc	protein
STPV K292 FISR	2.20	-17.23	CSRP2BP (HAT)
VS K350 GPF	2.09	-17.5	KAT6A (MYST3)
DH K604 TLY	1.18	-16.5	KAT6A
DS K1583 NAK	1.02	-17.1	CREB Binding Protein

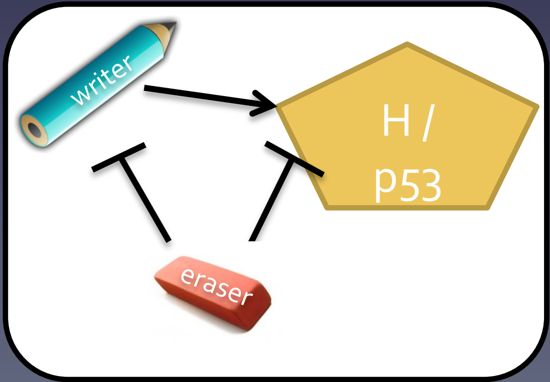
KAT acetylated for activation (@604)

- HDAC8 -> KAT6A
- sirt1 -> hMOF



KAT350 : in unstructured regions

- fine tuning?

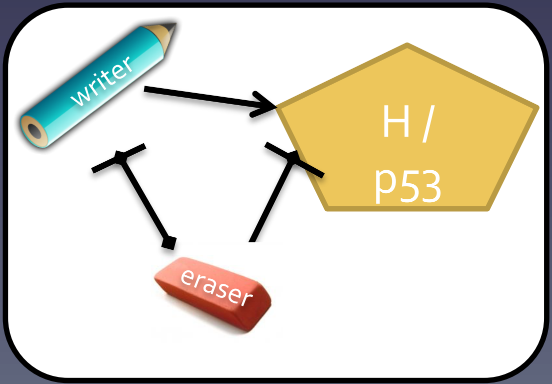


reactive ≥ 3 **reasonable ≥ 2** **mild ≥ 1** poor < 1

3 More chromatin remodelers …

peptide	log(Kcat/ Km)	min. I_sc	protein
KLIS **K1808 FDKL**	2.87	-17.85	ARID1A (chromatin remodeling)
SK**IQ K3579 QL**DQ	1.51	-16.25	MLL2 (methyltransferase)

- cross-talk between different modifications ?

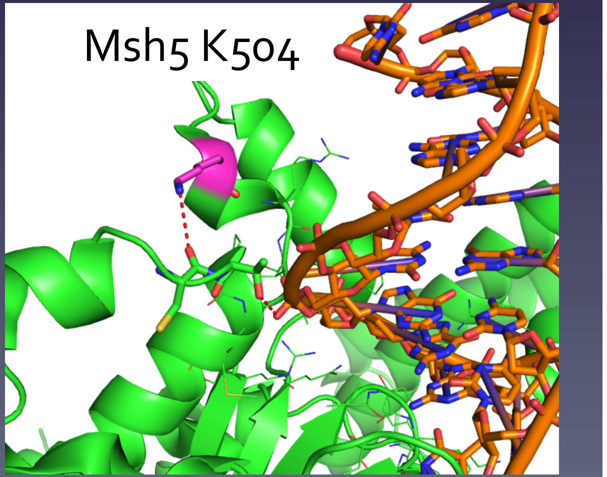


reactive ≥ 3 **reasonable ≥ 2** mild ≥ 1 poor < 1

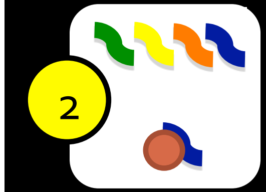
More cross-talk

peptide	log (K_{cat}/K_M)	min. I_sc	protein
SF K55 YAW	3.07	-20.8	EF1 α 1
LG K1017 FRR	2.54	-19.1	La-related protein 1
IS K504 YDR	1.92	-18.6	Msh6

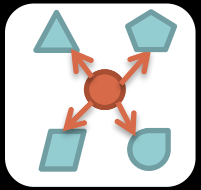
- EF1α1 - transcription/translation
- Larp1 - binds top5′ RNA
  - K1017 *adjacent* to RNA BD
- Msh6 - DNA mismatch repair
  - K504 *in periphery* of DNA



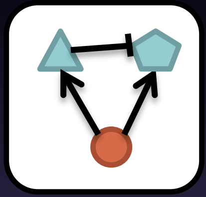
reactive ≥ 3 **reasonable ≥ 2** **mild ≥ 1** poor < 1



New targets - summary

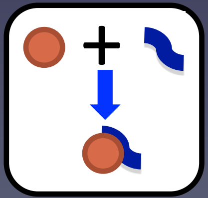


✓ found strongest known (peptide) HDAC8 substrates



✓ role of HDAC8 modifications
  - ✧*peripheral* fine-tuning
  - ✧extensive potential for crossregulation

✓ structure-based identification of substrates
  - ✧binding is the bottleneck for catalysis
  - ✧protocol works for **flexible & internal** sites



Why a structure-based approach ?

peptide	log (Kcat/Km)	min. I_sc
KLIS K1808 FDKL	2.87	-17.85
STPV K292 FISR	2.20	-17.23
KRILH K687 LLQN	1.70	-17.02
RVIGAK K106 DQY	1.80	-16.4
SKIQ K3579 QLDQ	1.51	-16.25
LGDG K387 MKS	0.66	-14.79
KLSG K167 EING	0.99	-14.29
KLGG K1087 QRAA	1.04	-14.19
TEIG K54 TLAEK	0.61	-13.96

structure and proteomic information complement each other

Proteomic identification of HDAC8 substrates

✓ *in vivo* data

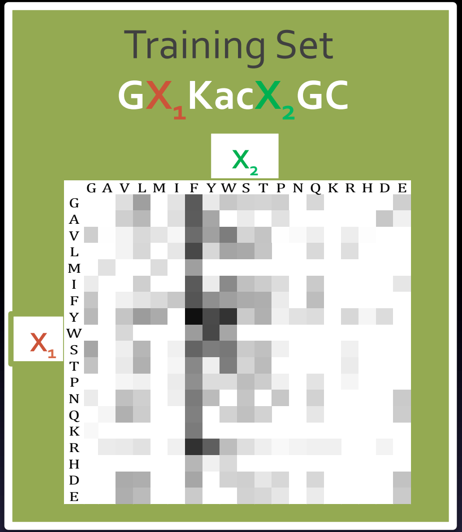
✧ many known targets missed

✧ weak targets (GKac preferred) -> bias?

Why a structure-based approach ?

peptide	log (Kcat/ Km)	min. I_sc
DI K42 YPL	**1.20**	**-18.6**
VS K350 GPF	**2.09**	-17.5
DS K1583 NAK	1.02	-17.1
SG K256 YDL	1.12	-17.1
VS K93 GTL	1.23	-16.6
SF K769 SDQ	1.39	-16.6
DH K604 TLY	1.18	-16.5
LS K747 FLR	2.80	-21.6
SF K55 YAW	3.07	-20.8
FG K1275 FSW	**3.68**	**-20.0**
TW K78 ANF	1.89	-19.6
LG K1017 FRR	2.54	-19.1
IS K504 YDR	1.92	-18.6
KI K31 RLR	1.58	-18.5
SL K269 EFY	1.89	-18.0
RL K713 YSQ	1.90	-16.8
SG K517 YYY	1.32	-16.8
SG K8 YFA	1.30	-16.2
TF K292 GVD	1.15	-15.4
KG K878 DAE	1.56	-14.5
GG K149 AFG	1.32	-14.5
RK K1970 GEP	1.00	-11.9

KLIS K1808 FDKL	2.87	-17.85
STPV K292 FISR	2.20	-17.23
KRILH K687 LLQN	1.70	-17.02
RVIGAK K106 DQY	1.80	-16.4
SKIQ K3579 QLDQ	1.51	-16.25
KLGG K1087 QRAA	1.04	-14.19


Training Set

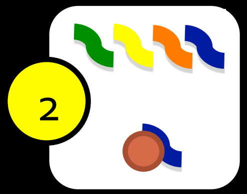
GX_1KacX_2GC

sequence based motif Kac[FY]

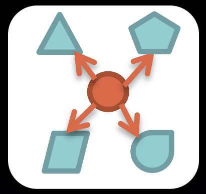
✓ identifies 6/7 strong substrates

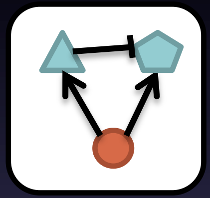
✧ many false positives

✧ weak substrates missed



What next ?





- ✓ found strongest known (peptide) HDAC8 substrates

- ➢ biology of new targets & functional impact
- ➢ differences and similarities among HDAC substrate specificities
    - ❖ targeted drugs ?

- ➢ extend to new systems (lipids, sugars, nucleotides…)

Back to the context

Jamal Fahoum

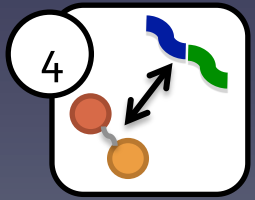
Emma-Joy Dodson

Shahar Rotem

Jordan Chill
Keren Adamsky
NMR Bar Ilan

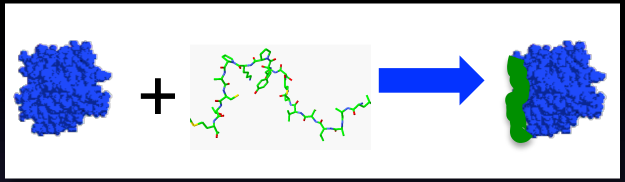
Rami Aqeilan
Saleh Khawaled

Reuven Wiener



Back to the challenge

Nature’s challenge



$\Delta G_{BIND} = \Delta H - T\Delta S$

✧ **small** interface

✧ **flexible** peptide

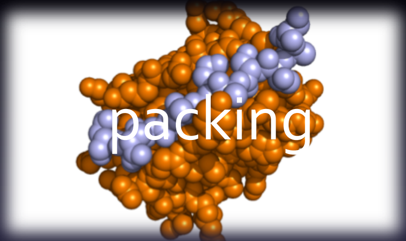
➢ how do peptides do it?

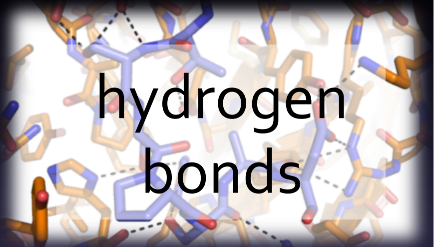
peptiDB

London, Movshovitz-Attias, Schueler-Furman (2010) Structure

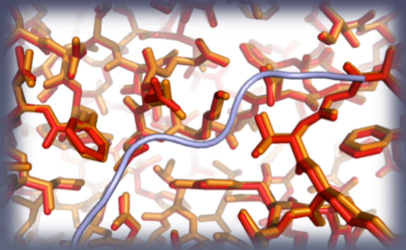
Nature’s solutions

① increase ΔH





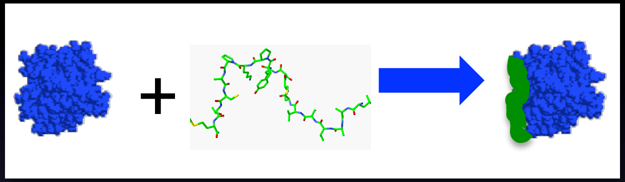
② decrease ΔS



prearranged
receptor
& peptide

Back to the challenge

Nature's challenge



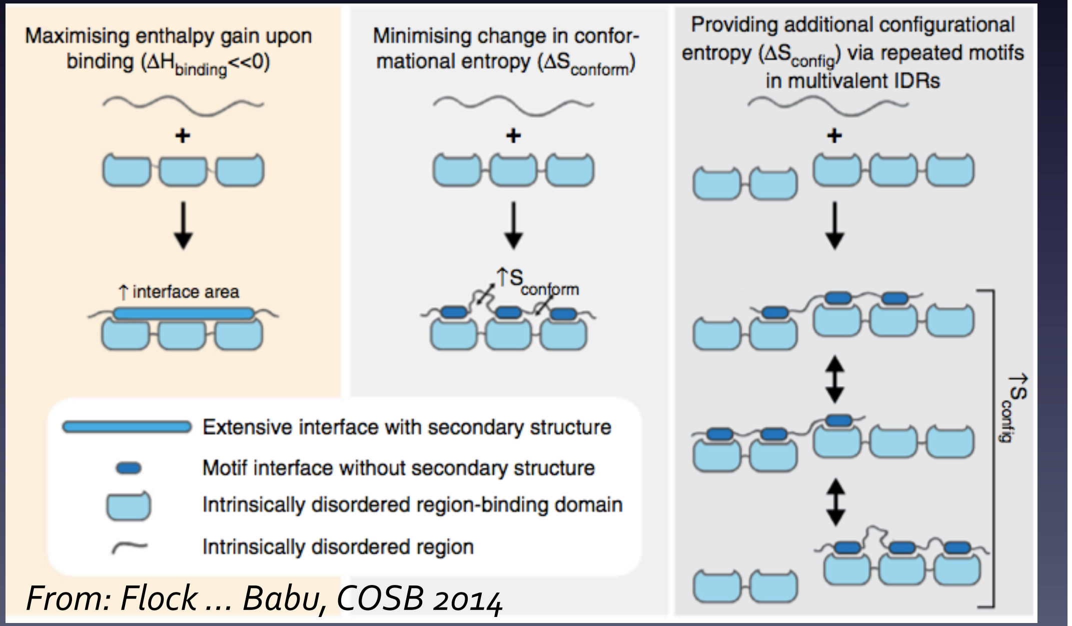
$$\Delta G_{BIND} = \Delta H - T\Delta S$$

✧ **small** interface

✧ **flexible** peptide

Nature's solutions

③ multiplicity



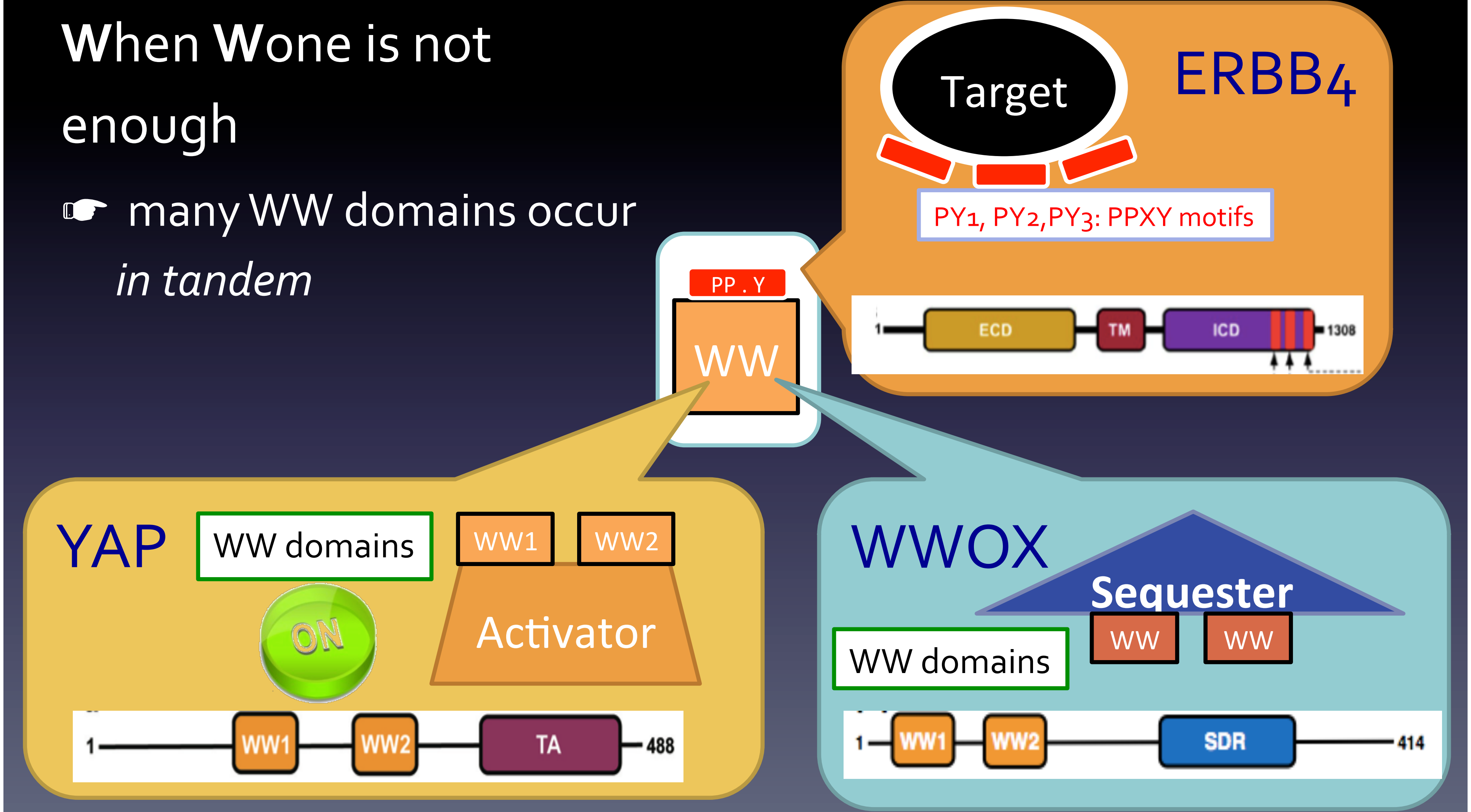
From: Flock … Babu, COSB 2014

➢ opportunities for regulation by context!

Communication by SMS

When Wone is not enough

☛ many WW domains occur *in tandem*



Target

ERBB4

PY1, PY2,PY3: PPXY motifs

PP . Y

WW

YAP

WW domains

ON

WW1

WW2

Activator

WWOX

Sequester

WW domains

WW

WW

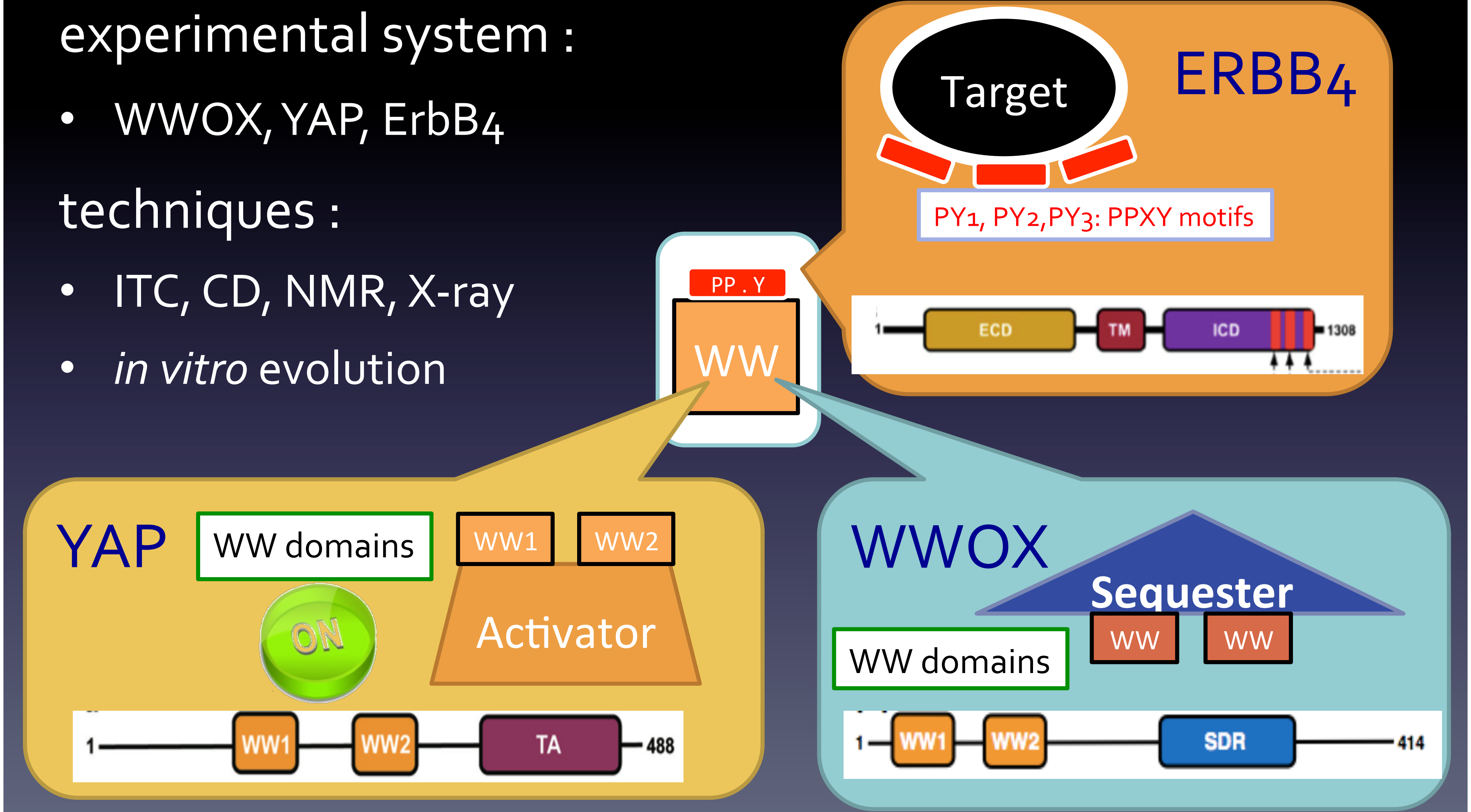
Communication by SMS

experimental system :

- WWOX, YAP, ErbB4

techniques :

- ITC, CD, NMR, X-ray
- *in vitro* evolution



ERBB4

Target

PY1, PY2,PY3: PPXY motifs

PP . Y

WW

YAP

WW domains

WW1

WW2

ON

Activator

WWOX

Sequester

WW domains

WW

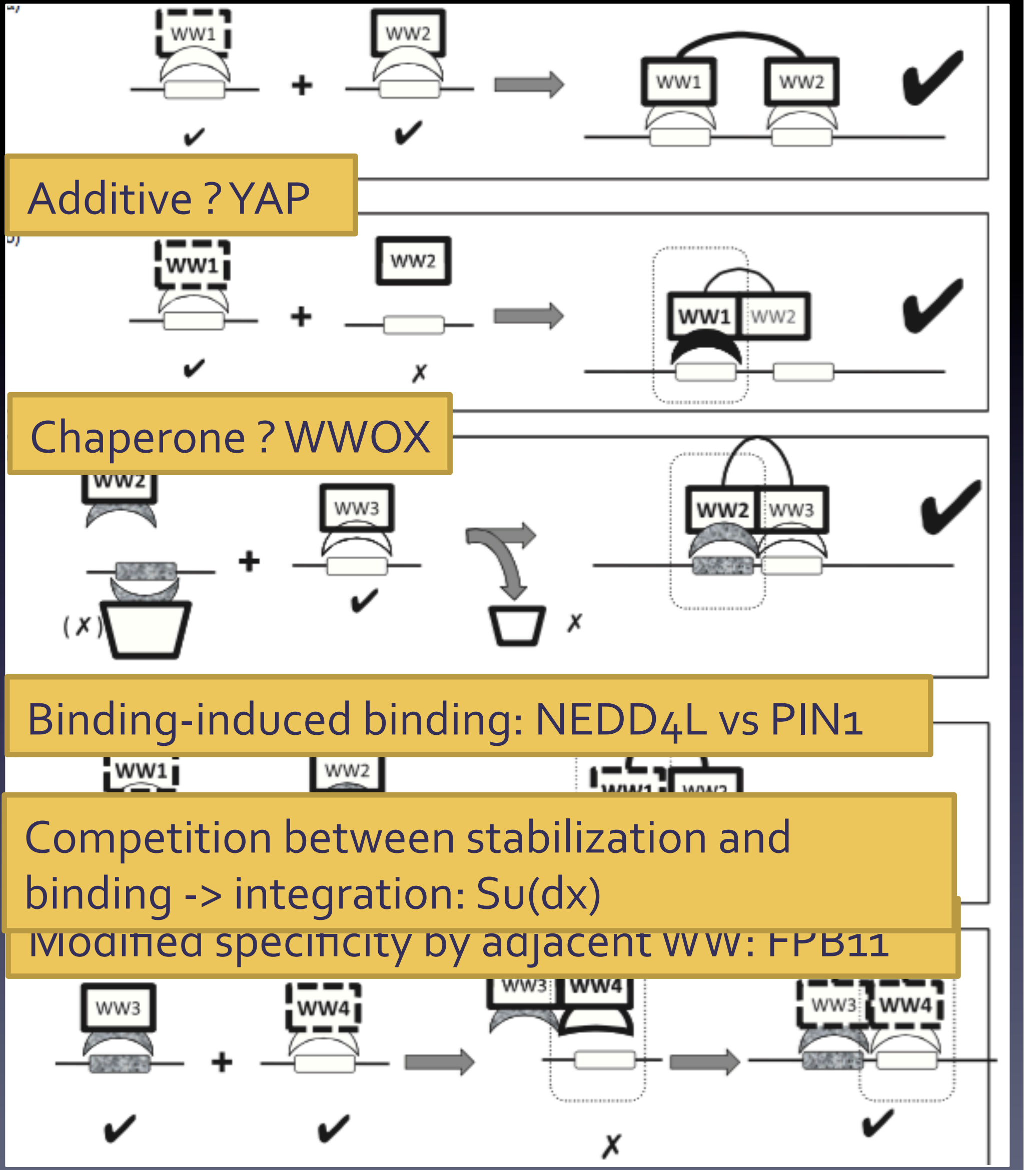
WW

When Wone is not enough

50% of WW domains occur *in tandem*

- ☛ *context* crucial for complex regulation
- ☛ variable integration strategies

Dodson,...., Schueler-Furman (2015). Versatile communication strategies among tandem WW domain repeats. Experimental Biology and Medicine, 240:351



Additive ? YAP

Chaperone ? WWOX

Binding-induced binding: NEDD4L vs PIN1

Competition between stabilization and binding -> integration: Su(dx)

Modified specificity by adjacent WW: FPB11

THANKS!

Fierke Lab U Michigan

Noah Wolfson
Caleb Joseph
Carol Fierke

Current lab members:

Nawsad Alam, Orly Marcu, Emma-Joy Dodson, Jamal Fahoum, Adily Mashash, Yuval Sedan, Oriel Goldstein, Aviou Zini, Alon Lalazry, Debbie Shalev, Shahar Rotem

Previous lab members

Nir London (Weizmann Institute), Barak Raveh (UCSF), Dan Reshef
Lavanya Pushpam, Michal Sperber, Vered Fishbain, Lior Zimmerman, Assaf Lavi, Dana Movshovitz-Attias (CMU)

NMR Jordan Chill, Keren Adamsky, Bar Ilan University

ITC & CD Assaf Friedler, Daniel Harries

Crystallography Reuven Wiener

***In vivo* assays** Rami Aqeilan, Saleh Khawaled

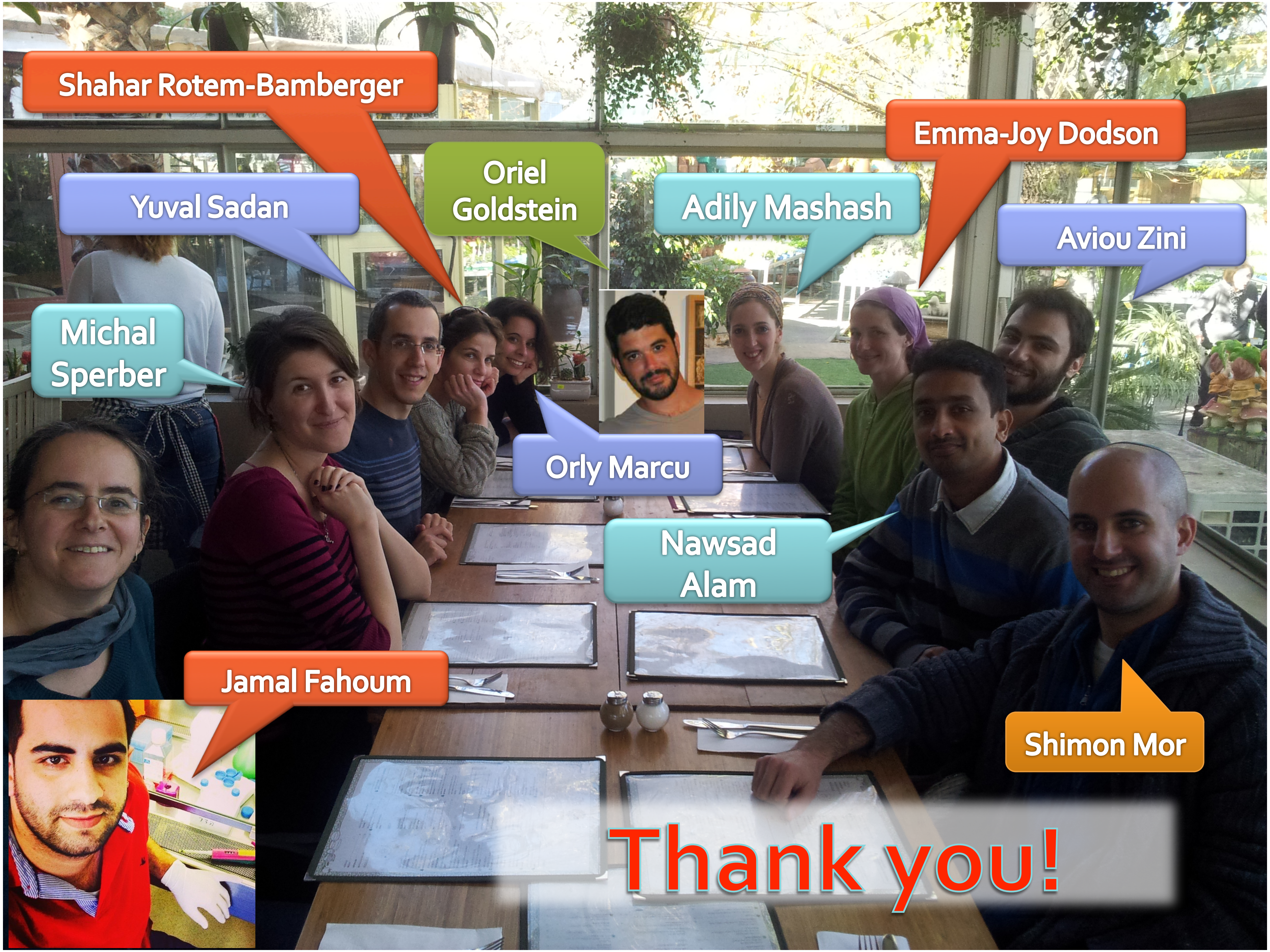
Dima Kozakov

Boston U & SUNY Stony Brook

Rosetta Community

BSF erc

Furman Lab @ HU


Shahar Rotem-Bamberger
Emma-Joy Dodson
Oriel Goldstein
Yuval Sadan
Adily Mashash
Aviou Zini
Michal Sperber
Orly Marcu
Nawsad Alam
Jamal Fahoum
Shimon Mor

Thank you!