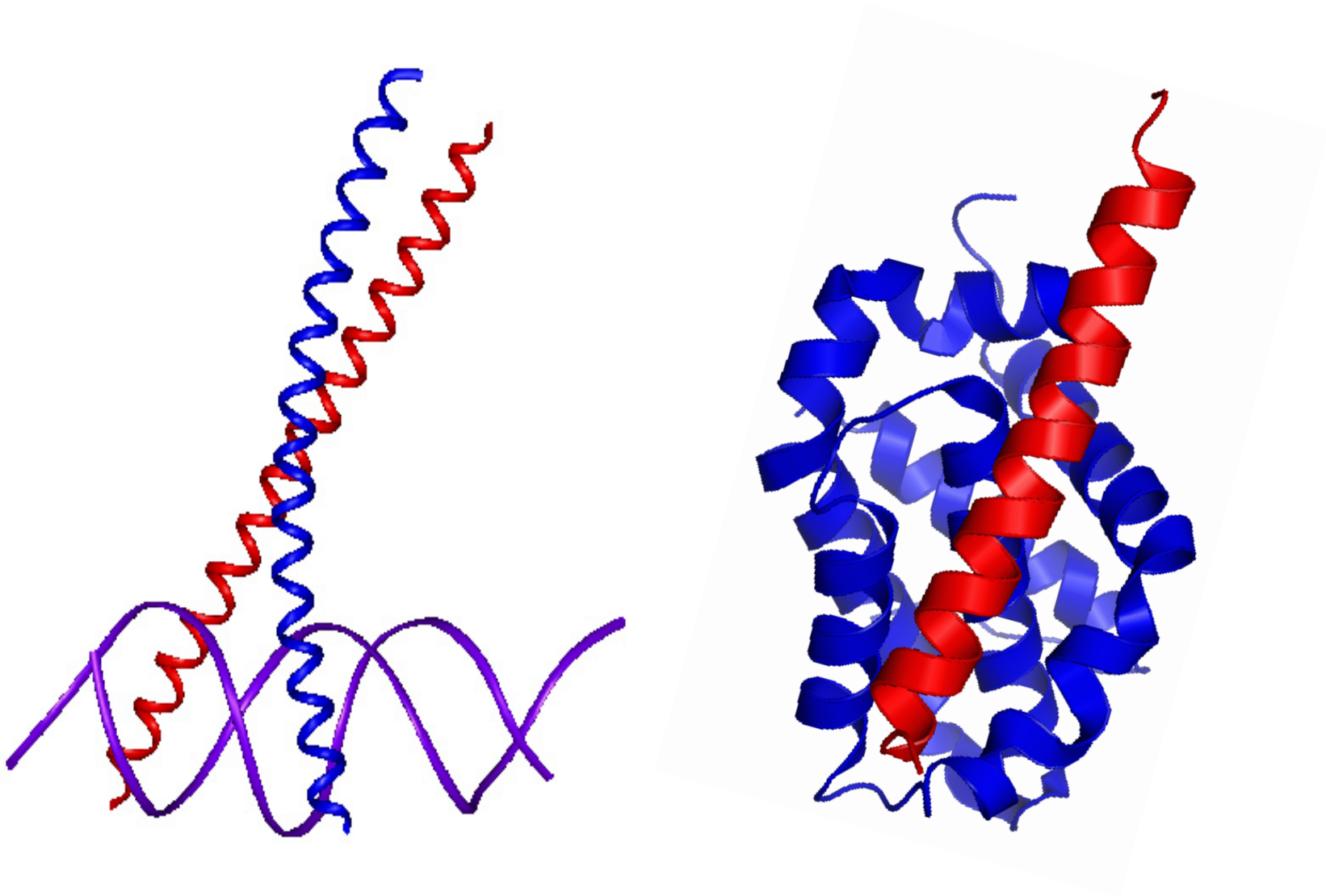
Measuring and modeling protein-peptide interaction specificity

Amy E. Keating

RosettaCon

July 30, 2012

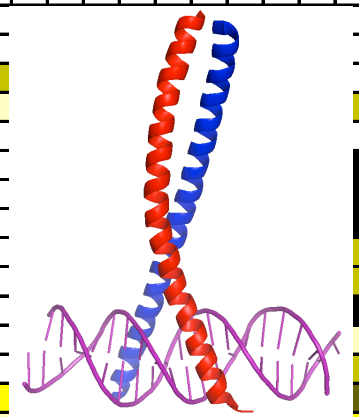
KEATING LAB



GOALS: understand specificity, predict interactions, design specific binders

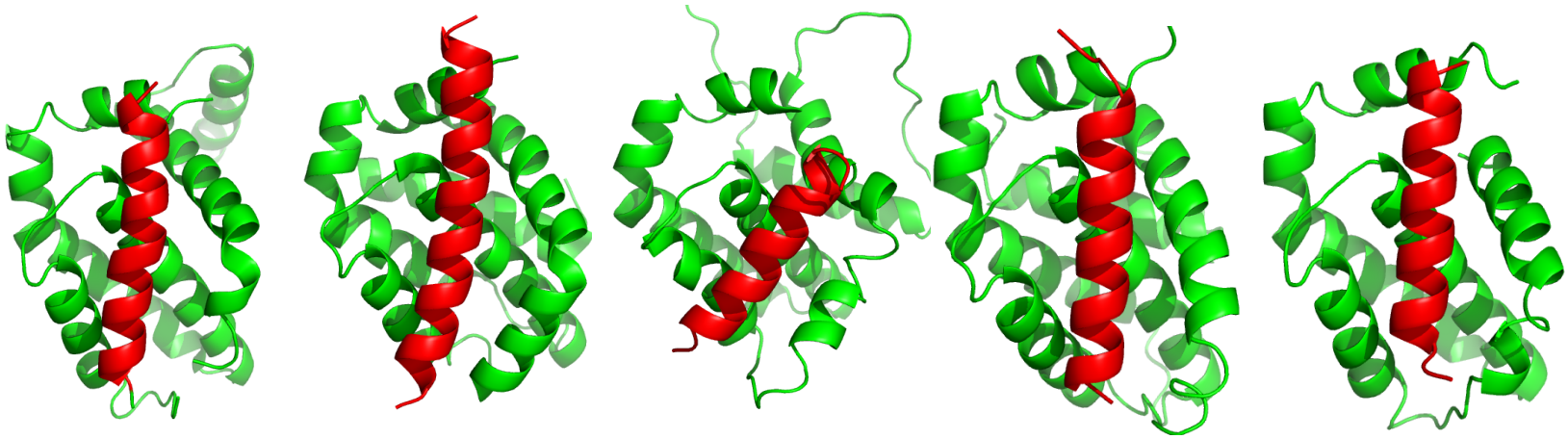
Human bZIP interaction network (36 x 36)

Family	Protein	DDIT3	CEBPG	CEBPA	CEBPE	CREB1	CREB3	CREB3L3	CREB3L1	CREBZF	XBP1	ATF6	ATF6B	NFIL3	DBP	HLF	ATF2	JUN	JUNB	FOS	FOSL1	ATF4	ATF5	ATF3	BATF	BATF2	BATF3	MAFF	MAFG	MAF	MAFB	NFE2	NFE2L1	NFE2L2	NFE2L3	BACH1	BACH2
DDIT3	DDIT3																																			
CEBPG	CEBPG																																			
CEBP	CEBPA																																			
CEBP	CEBPE																																			
CREB	CREB1																																			
OASISA	CREB3																																			
OASISA	CREB3L3																																			
OASISB	CREB3L1																																			
CREBZF	CREBZF																																			
XBP1	XBP1																																			
ATF6	ATF6																																			
ATF6	ATF6B																																			
NFIL3	NFIL3																																			
PAR	DBP																																			
PAR	HLF																																			
ATF2	ATF2																																			
JUN	JUN																																			
JUN	JUNB																																			
FOS	FOS																																			
FOS	FOSL1																																			
ATF4	ATF4																																			
ATF4	ATF5																																			
ATF3	ATF3																																			
BATF	BATF																																			
BATF	BATF2																																			
BATF	BATF3																																			
SMAF	MAFF																																			
SMAF	MAFG																																			
LMAF	MAF																																			
LMAF	MAFB																																			
NFE2	NFE2																																			
NFE2	NFE2L1																																			
NFE2	NFE2L2																																			
NFE2	NFE2L3																																			
BACH	BACH1																																			
BACH	BACH2																																			



Kd(nM)	
<10	
10-50	
50-250	
250-1000	
>1000	
Non	

~15 pro-death BH3 domains bind selectively to 5 prosurvival Bcl-2 receptors



	Bcl-x_L	Mcl-1	Bcl-w	Bfl-1	Bcl-2
Bim	✓	✓	✓	✓	✓
Bad	✓	✗	✓	✗	✓
Noxa	✗	✓	✗	✓	✗
Mule	✗	✓	✗	✗	✗
Bok	✗	✓	✗	✗	✗
Puma	✓	✓	✓	✓	✓

To better understand/predict/design specific interactions:

1. Collect lots of data

2. Attempt to model it

To better understand/predict/design specific interactions:

1. Collect lots of data

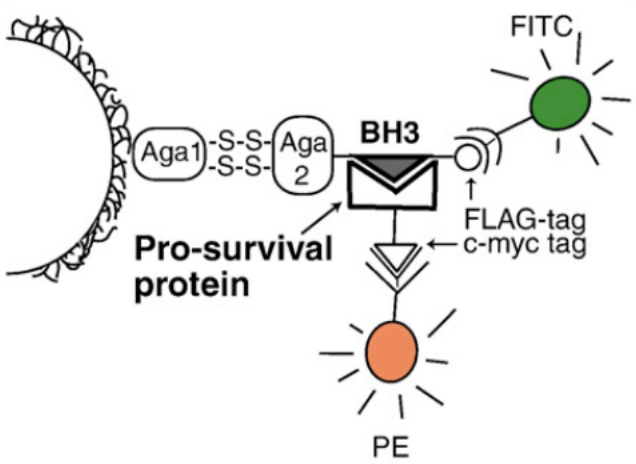
2. Attempt to model it

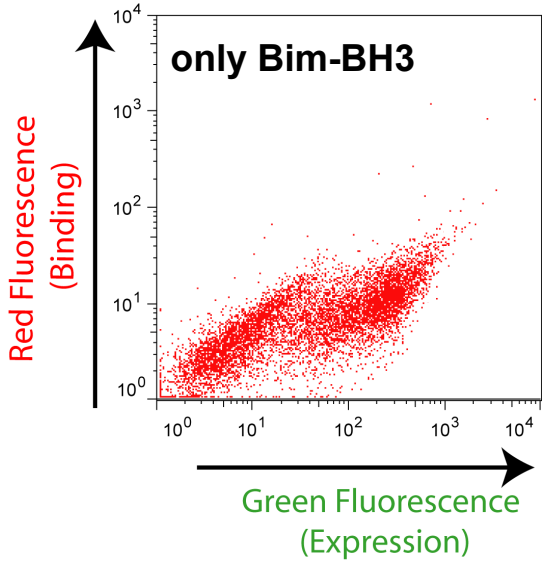
Coiled-coil work on posters by

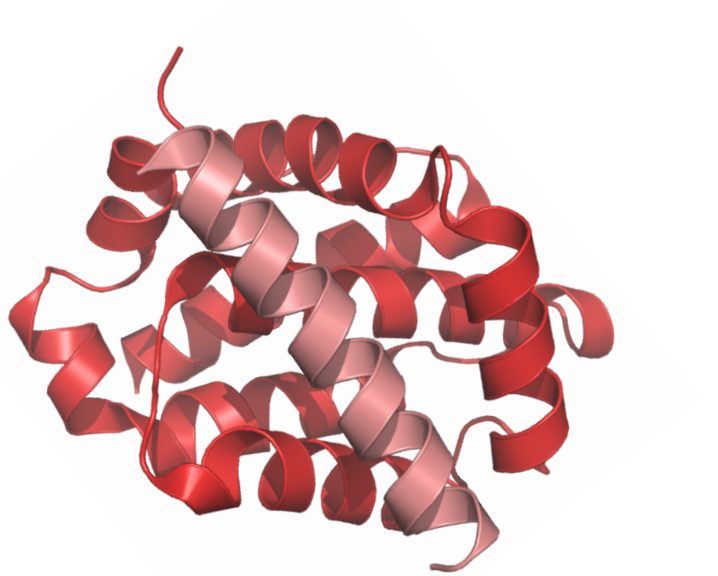
Chris Negron
incorporating anti-parallel states into design

Vladimir Potapov
specificity models assessed using large experimental bZIP benchmarks

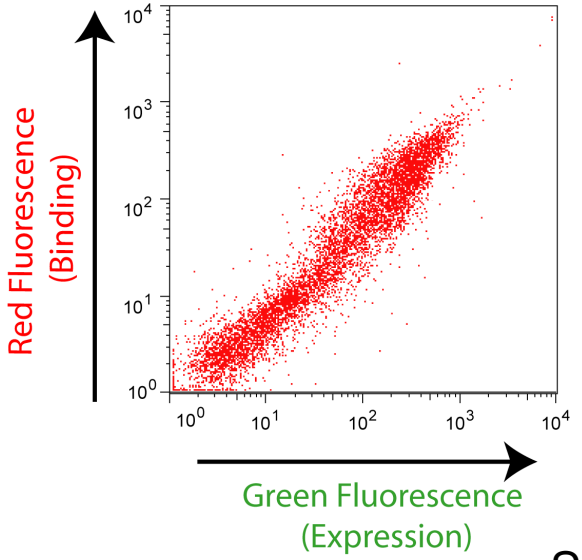
Data from yeast-surface display library screening





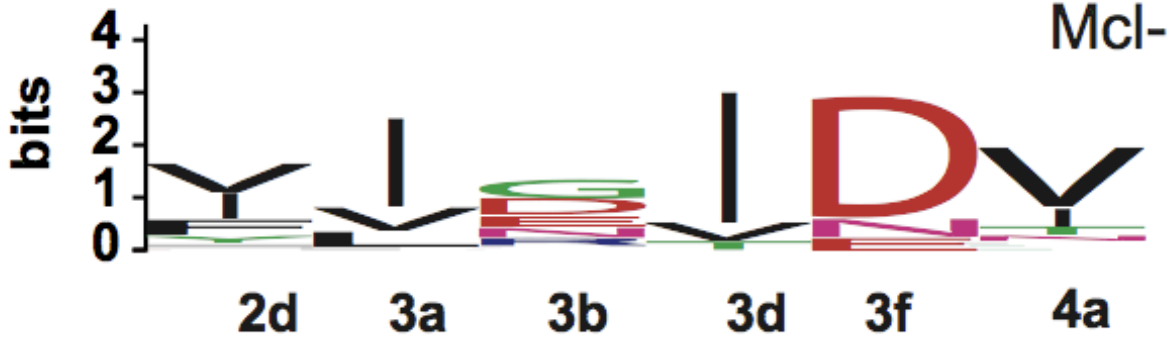


Bim-BH3 + 100 nM Bcl-x_L

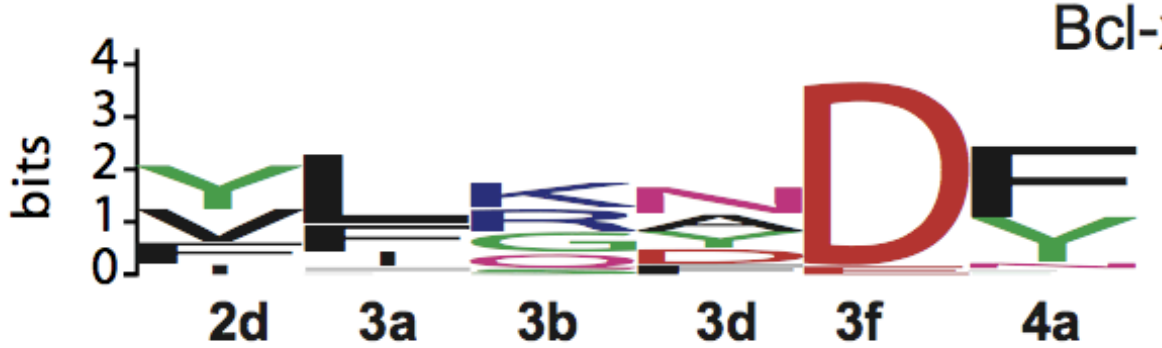


Sanjib Dutta

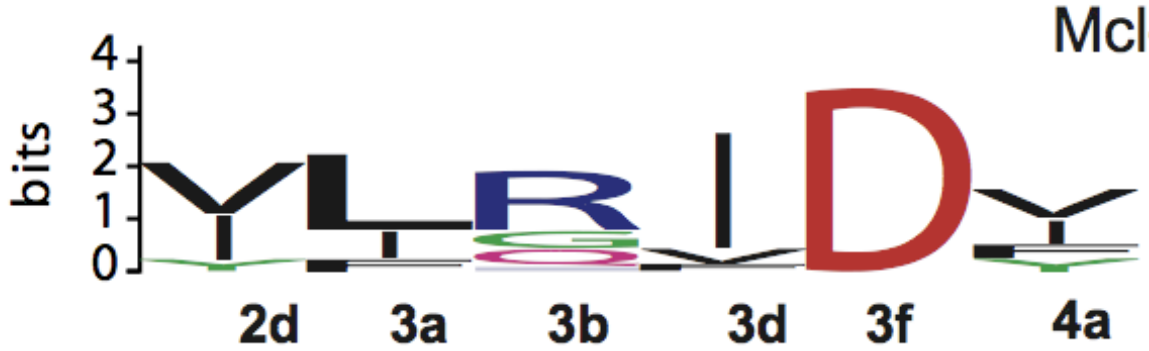
33 Mcl-1 specific unique clones



40 Bcl-x_L specific unique clones



17 unique clones bind both



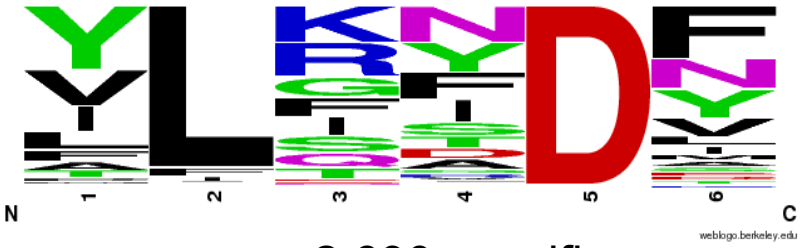
Dutta et al. JMB 2010

Data-based model I

Use yeast-display sorting data to define PSSM weights

high-affinity pool (Mcl-1)

high-specificity pool (Mcl-1)



e.g. 2,690 specific sequences

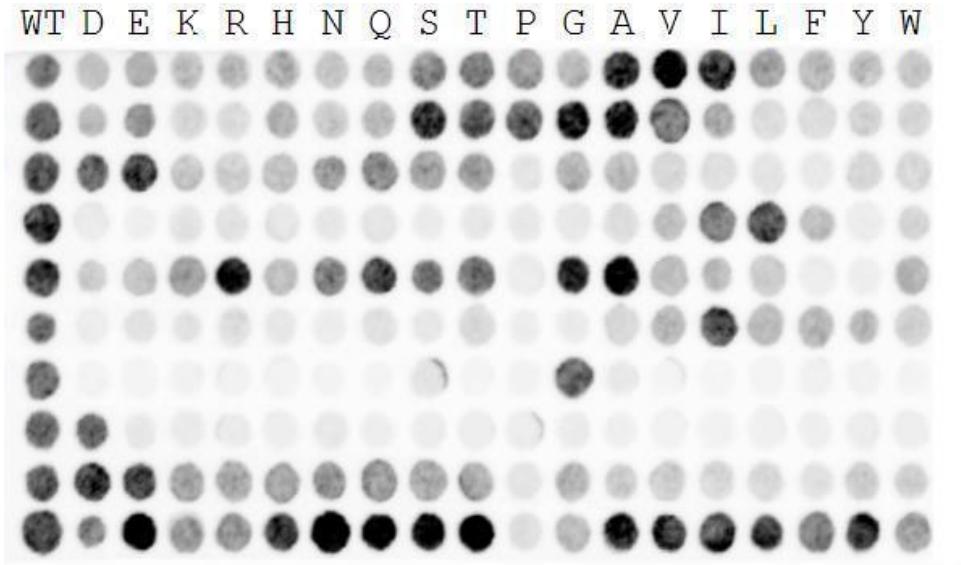
$$\textbf{Score PSSM}_{\textbf{DEEP}} = -\Sigma_i \ln(P_{i,j})$$

$P_{i,j}$ = frequency of residue j at site i in sequences from deep sequencing of yeast display pools

DeBartolo JMB 2012

Data from SPOT arrays

I2d	MRPEIW**X**AQELRRIGDEFNAYYARRV
A2e	MRPEIWI**X**QELRRIGDEFNAYYARRV
E2g	MRPEIWIAQ**X**LRRIGDEFNAYYARRV
L3a	MRPEIWIAQE**X**RRIGDEFNAYYARRV
R3b	MRPEIWIAQEL**X**RIGDEFNAYYARRV
I3d	MRPEIWIAQELRR**X**GDEFNAYYARRV
G3e	MRPEIWIAQELRRI**X**DEFNAYYARRV
D3f	MRPEIWIAQELRRIG**X**EFNAYYARRV
E3g	MRPEIWIAQELRRIGD**X**FNAYYARRV
F4a	MRPEIWIAQELRRIGDE**X**NAYYARRV

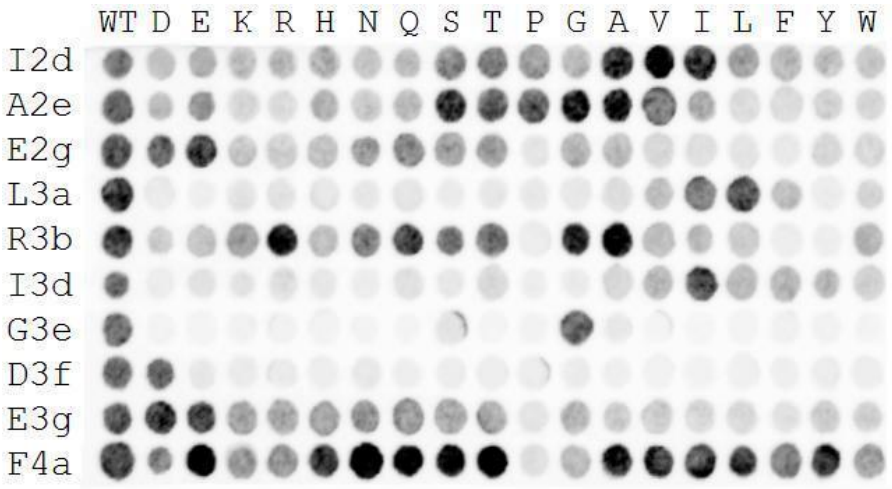


Substitution array probed with 100 nM Mcl-1

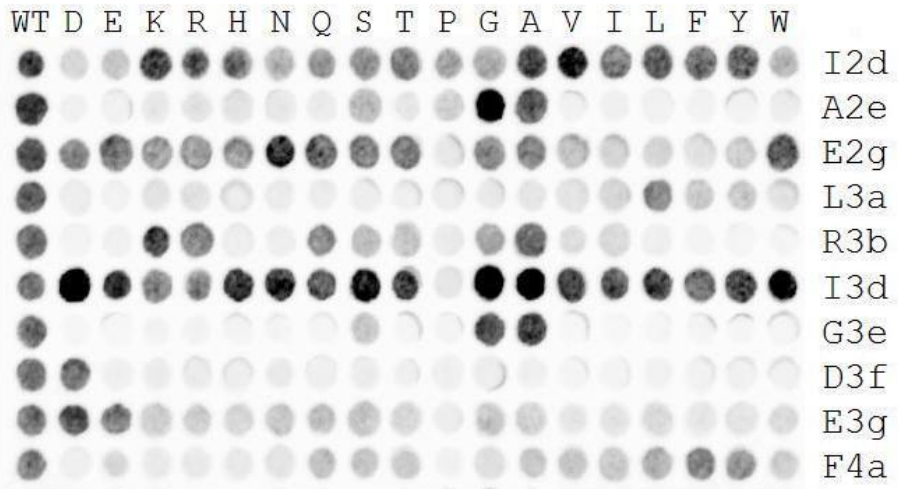
Emiko Fire, Stefano Gullá, Dick Cook

SPOT arrays for Bim-BH3 substitution analysis

100 nM Mcl-1



100 nM Bcl-x_L



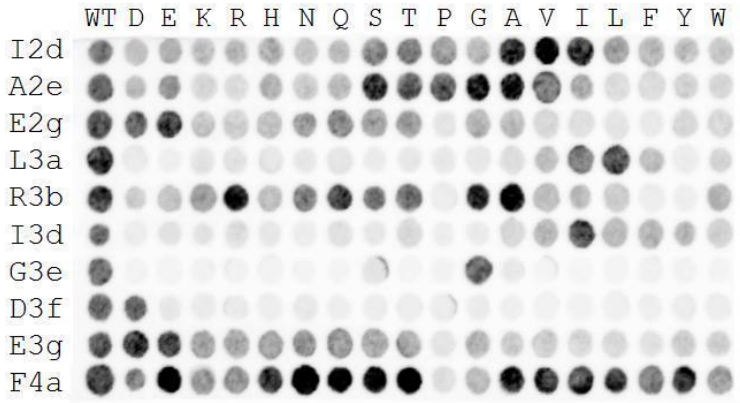
Emiko Fire, Stefano Gullá

Data-based model II

Use SPOT substitution array data to define PSSM weights

$$\textbf{Score PSSM}_{SPOT} = \Sigma_i \log (I_{i,j}/I_{i,WT\_res})$$

$$I_{i,j} = \textbf{SPOT intensity for residue j at site i}$$



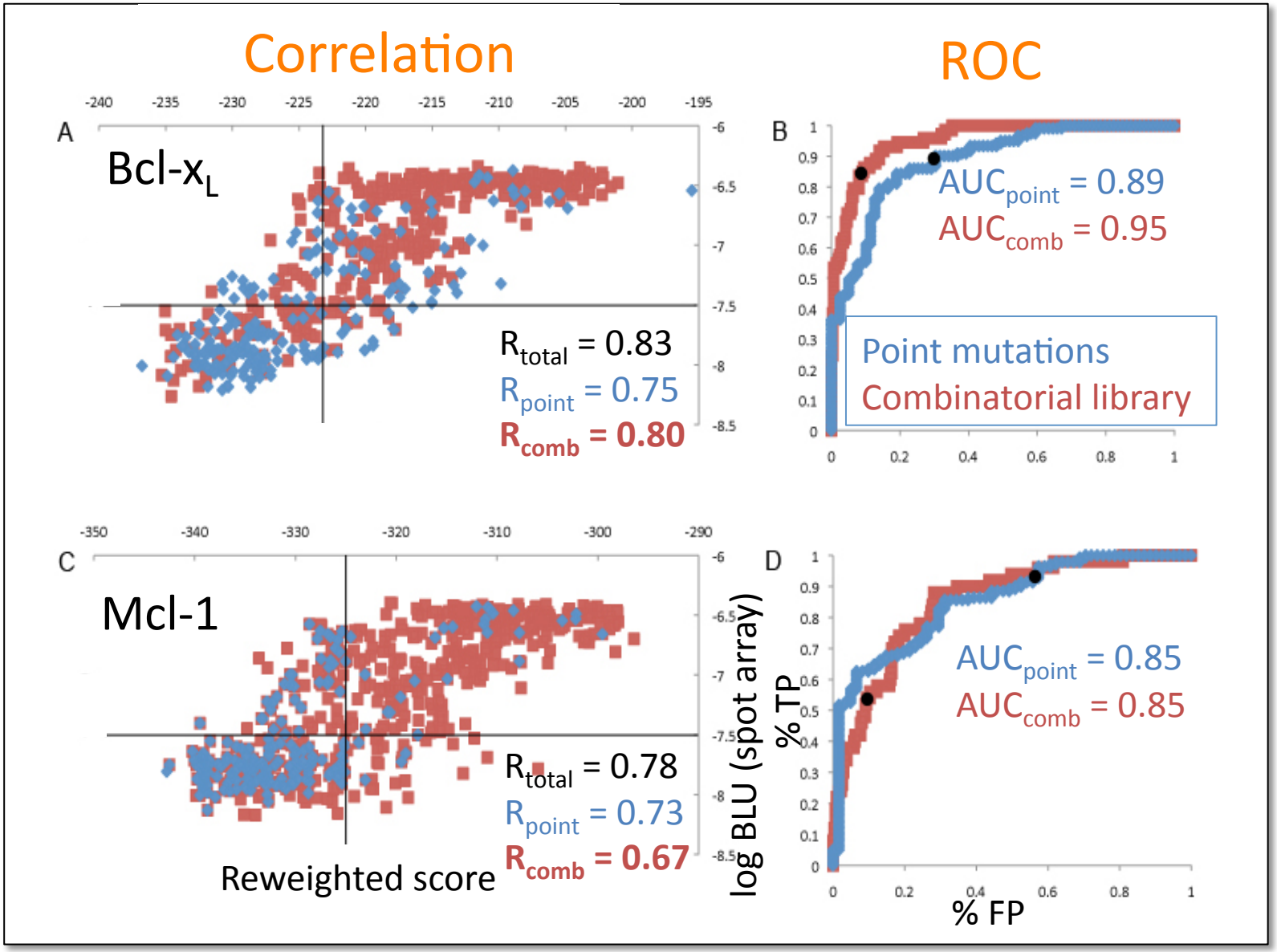
T. Scott Chen; Dutta et al. JMB 2010; DeBartolo et al. JMB 2012

Structure-based computational models

Rosetta!

Bcl-x_L & Mcl-1: Slide from Ora's talk last year

London et al. Biochemistry 2012



STATIUM: A very fast structure-based model

P_{ij} : Probability of amino acid pair ***ij*** at structure of pair **res_1-res_2**
P_i : Probability of amino acid ***i*** in the PDB
P_j : Probability of amino acid ***j*** in the PDB

$$-\sum_{N}\log\left(\frac{P_{ij}}{P_i P_j}\right)$$

Critical features:
- Only experimental input is the template structure
- Fixed backbone and no side chain atoms beyond Cβ
- Can score sequences at a rate of 10^6 s^{-1}

identify interacting residue pairs

AA pair probs. from PDB

res_1	res_2
BH3-4a	Receptor-45
BH3-4a	Receptor-86
BH3-3a	Receptor-91
BH3-3c	BH3-3g

etc…

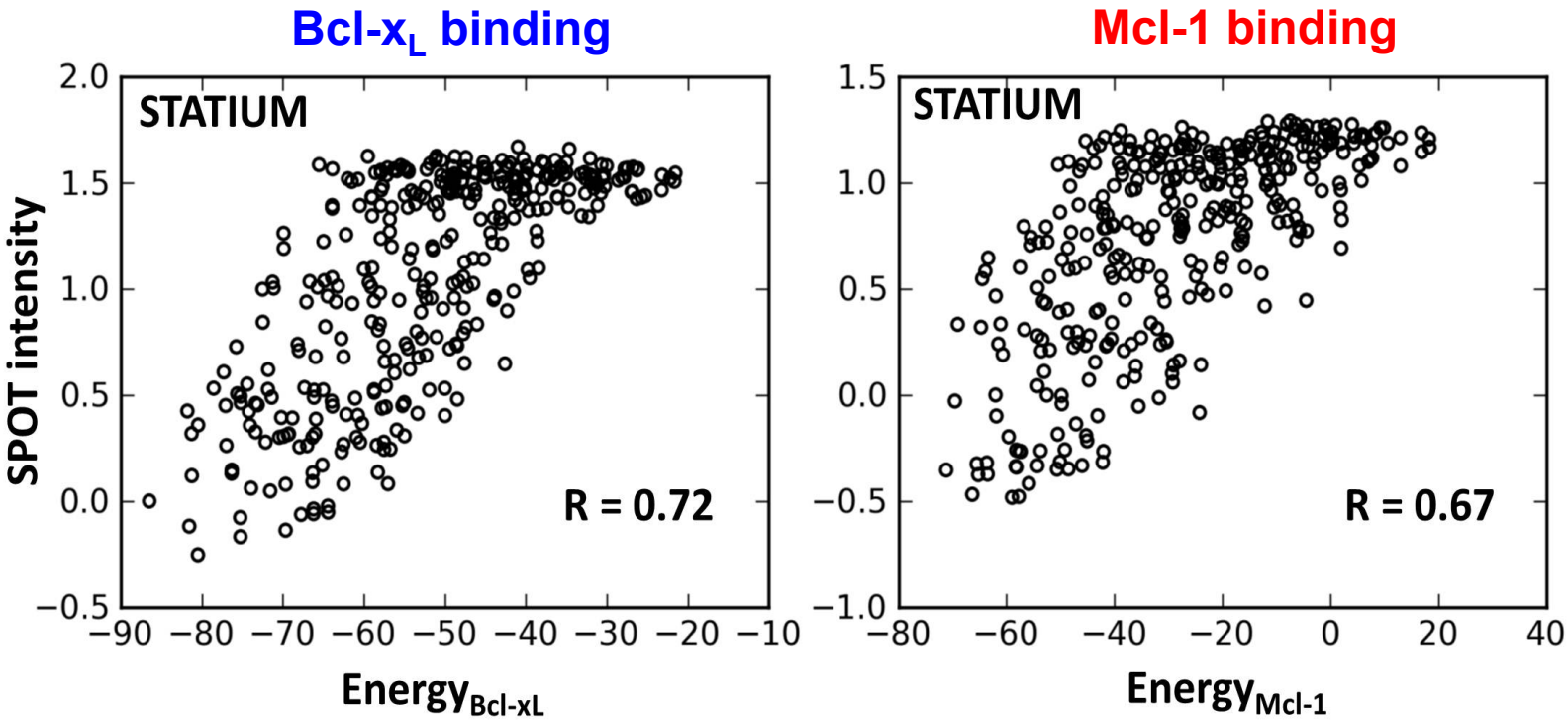
calculate distances

res_1-res_2	$Cβ_{distance}$	$Cα_{distance}$
	6.1	7.5
	6.2	8.8
	6.1	8.3
	6.5	6.8

How much accuracy is lost to speed?

Test 1: Experimental SPOT intensities vs. STATIUM

360 peptides including 1-5 mutations in Bim BH3



Performance of FlexPepBind on the same dataset

unoptimized: **R = 0.62, 0.62**

optimized: **R = 0.8, 0.67**

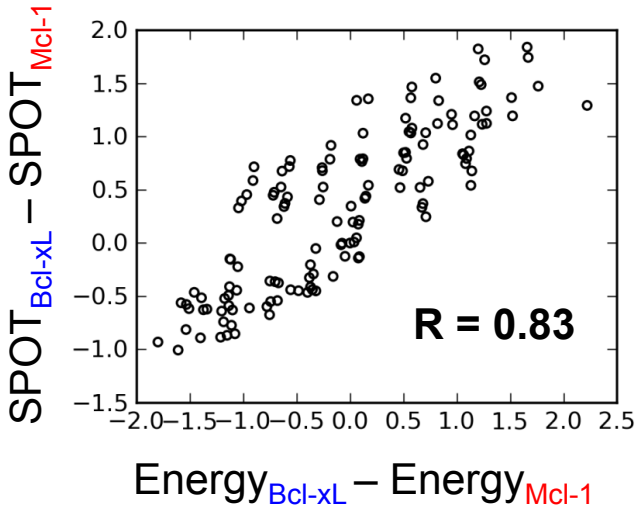
(London et al. Biochemistry)

Joe DeBartolo JMB 2012

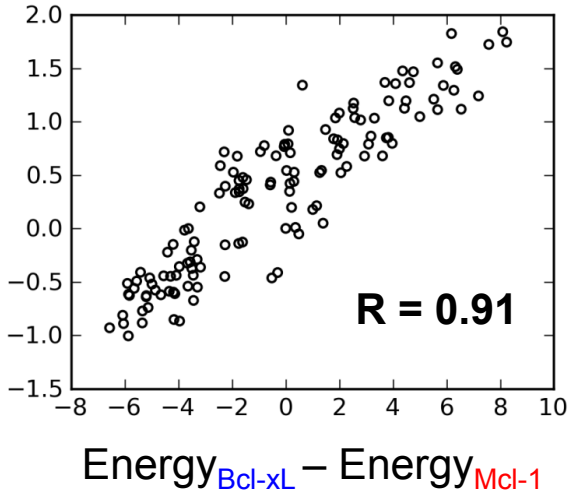
Test 2: Distinguish Mcl-1 vs. Bcl-x_L SPOT signals

~140 peptides including 1-5 mutations in Bim BH3

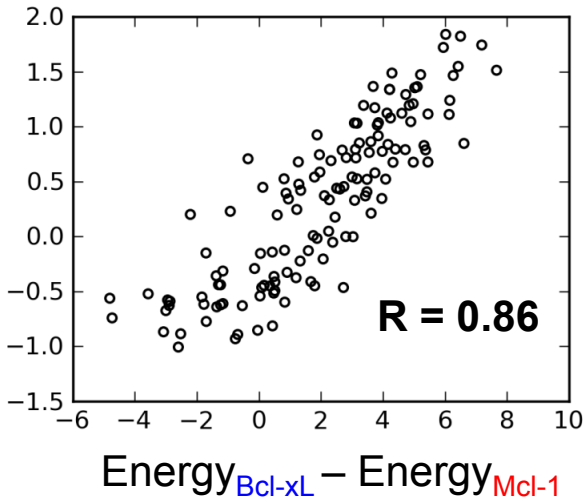
$PSSM_{SPOT}$

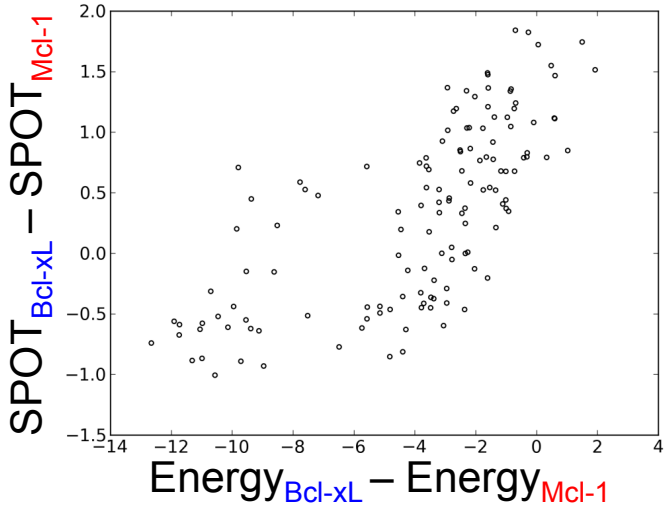


$PSSM_{DEEP}$



STATIUM





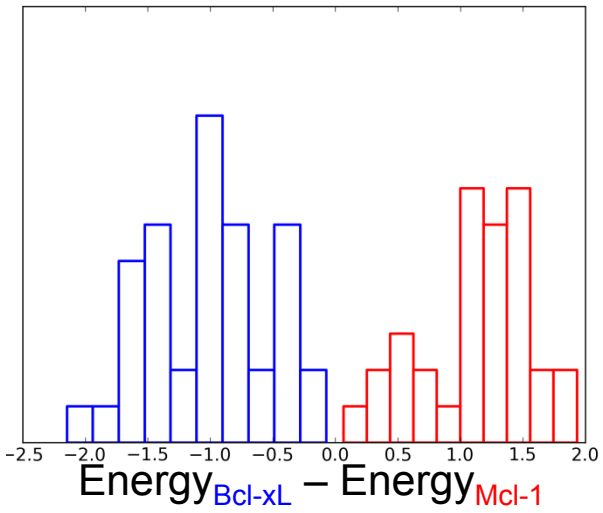
FlexPep Bind

R = 0.75

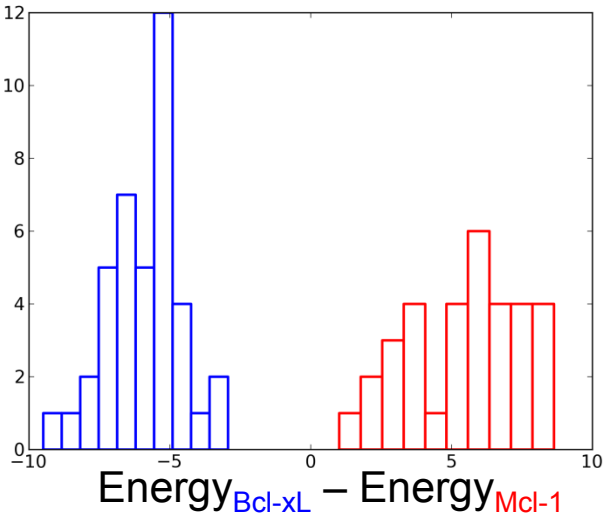
Joe DeBartolo JMB 2012

Test 3: Distinguish Mcl-1 vs. Bcl-x_L specific sequences identified by yeast screening

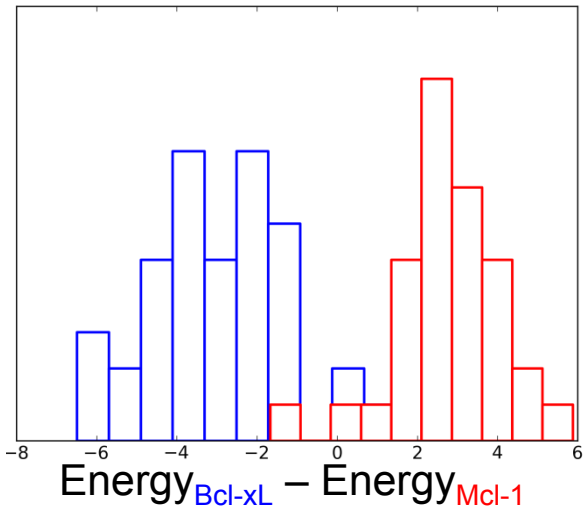
$PSSM_{SPOT}$



$PSSM_{DEEP}$

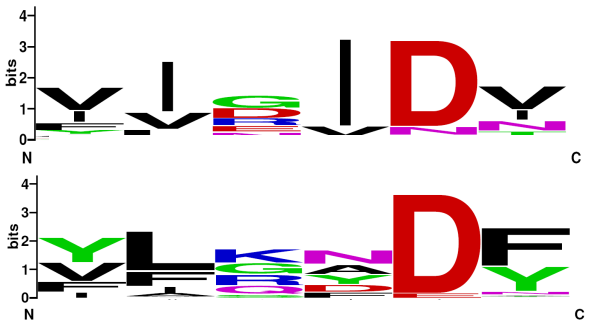


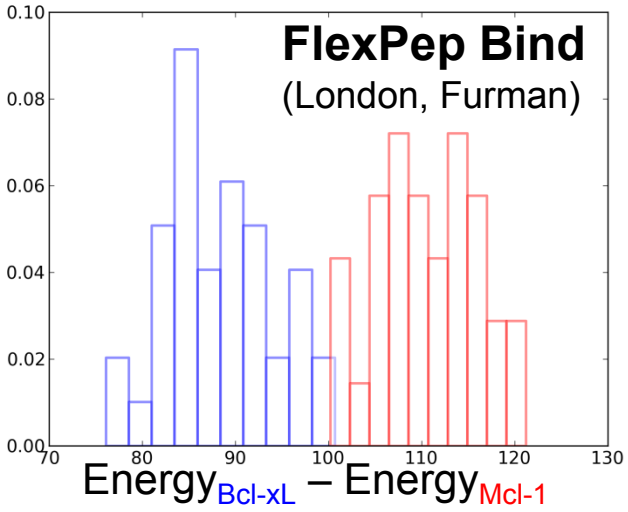
STATIUM



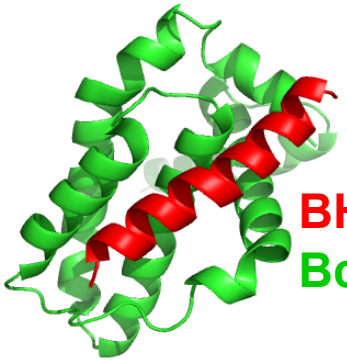
33 Mcl-1 specific sequences

40 Bcl-x_L specific sequences





Joe DeBartolo
JMB 2012



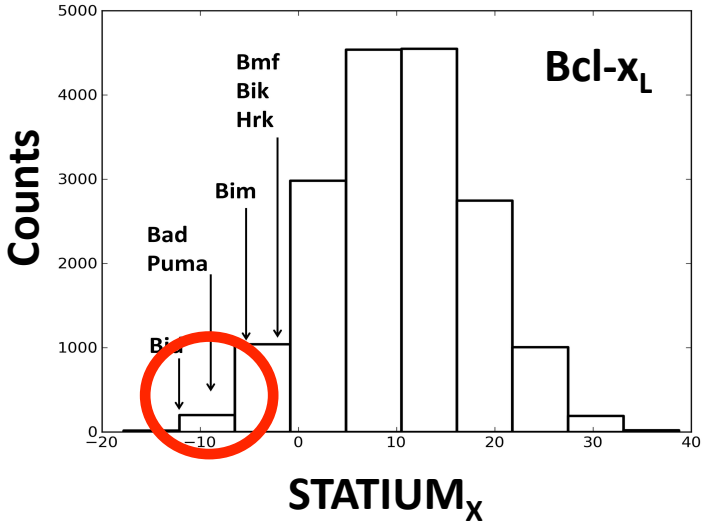
BH3 domain
Bcl-2 receptor

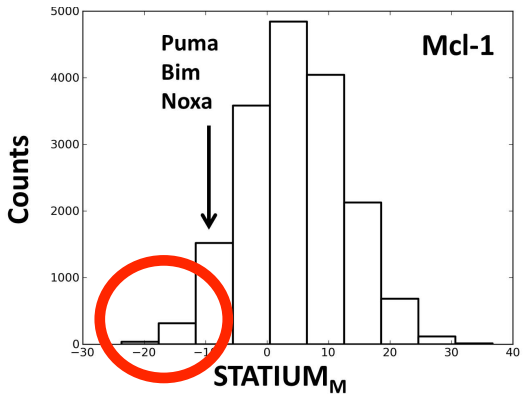
Methods used to discover the current Bcl-2 interactome

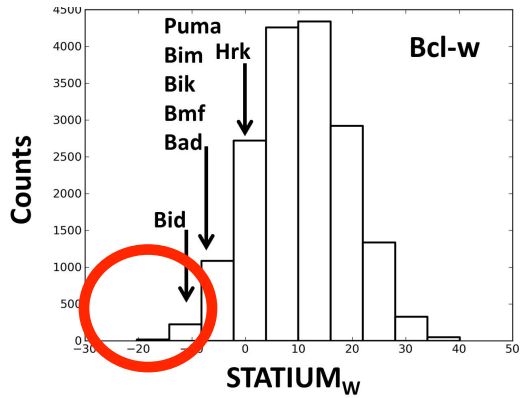
BH3 Domains	First method used to discover Bcl-2 interaction	
Bim:	Phage λ screening of cDNA libraries	O'Connor et al. EmboJ 1998
Bid:	Phage λ screening of cDNA libraries	Wang et al. Genes Dev. 1996
Bad:	Yeast two-hybrid	Yang et al. Cell 1995
Bok:	Yeast two-hybrid	Hsu et al. PNAS 1997
Hrk:	Yeast two-hybrid	Inohara et al. EmboJ 1997
Bmf:	Yeast two-hybrid	Puthalakath et al. Science 2001
Beclin:	Yeast two-hybrid	Liang et al. J Virol. 1998
Nip1/2/3:	Yeast two-hybrid	Boyd et al. Cell 1994
Noxa:	Coimmunoprecipitation	Oda et al. Science 2000
Puma:	Coimmunoprecipitation	Nakano et al. Mol. Cell 2001
Mule:	Coimmunoprecipitation	Zhong et al. Science 2005
Bik:	Coimmunoprecipitation	Gillissen et al. EmboJ 2003

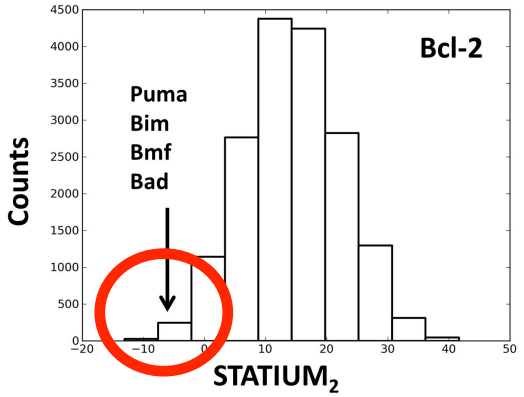
There may be many more BH3 domains.
Can we find them?

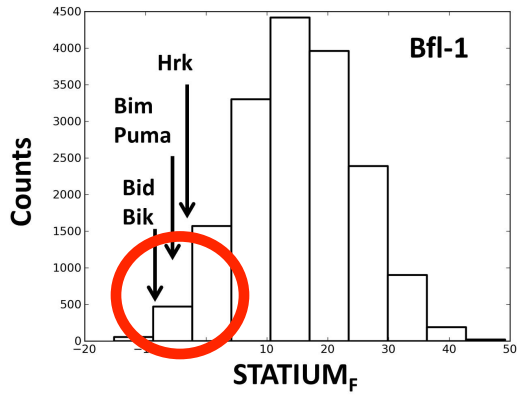
There are 5 human Bcl-2 receptors
Scan the human genome with 5 STATIUM models



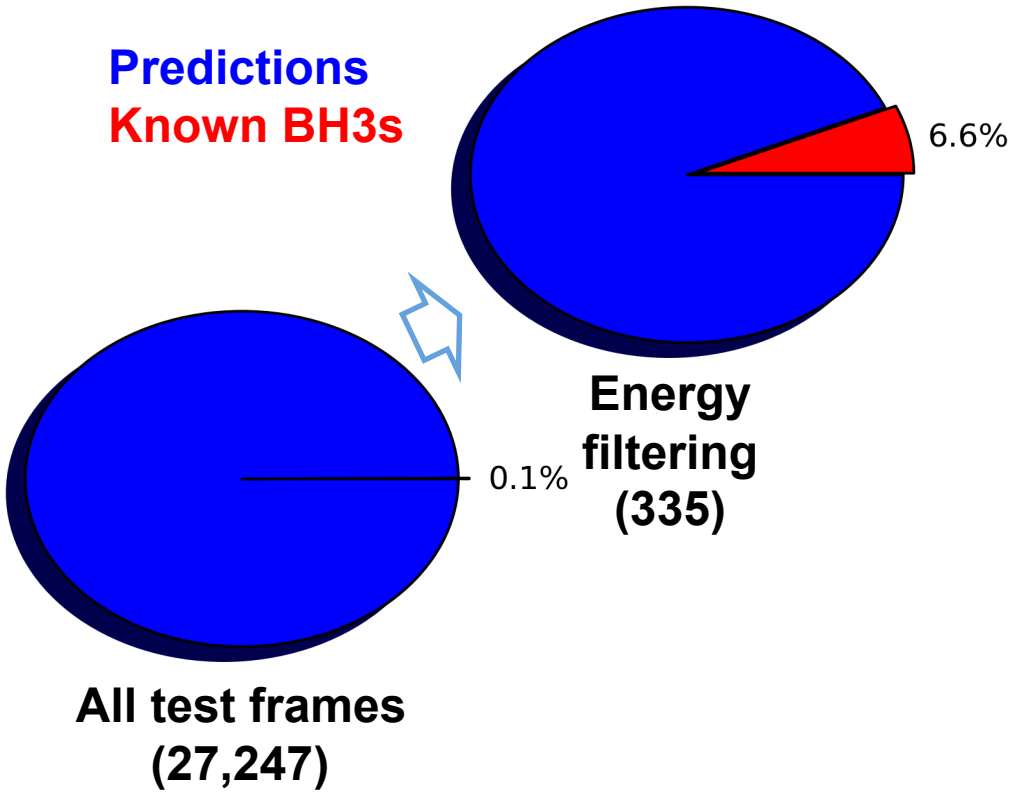




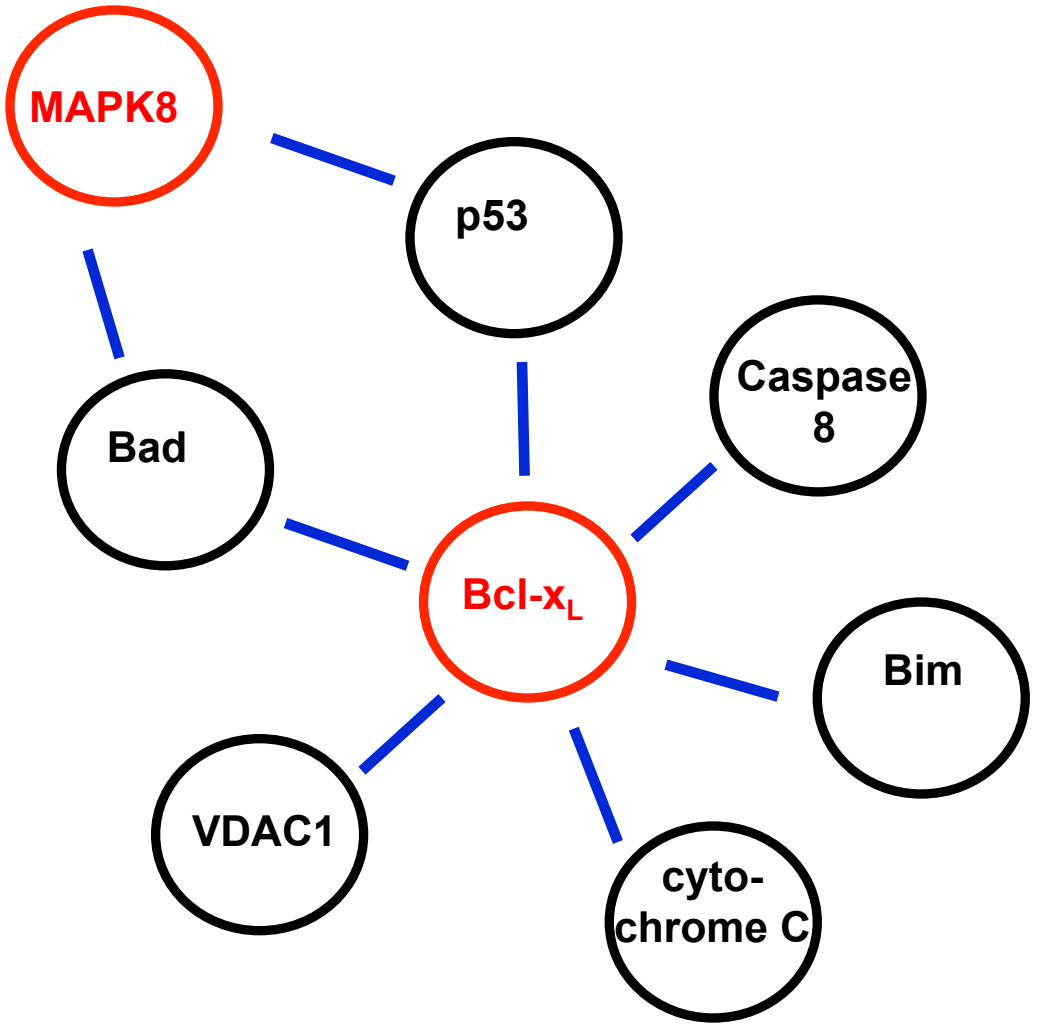




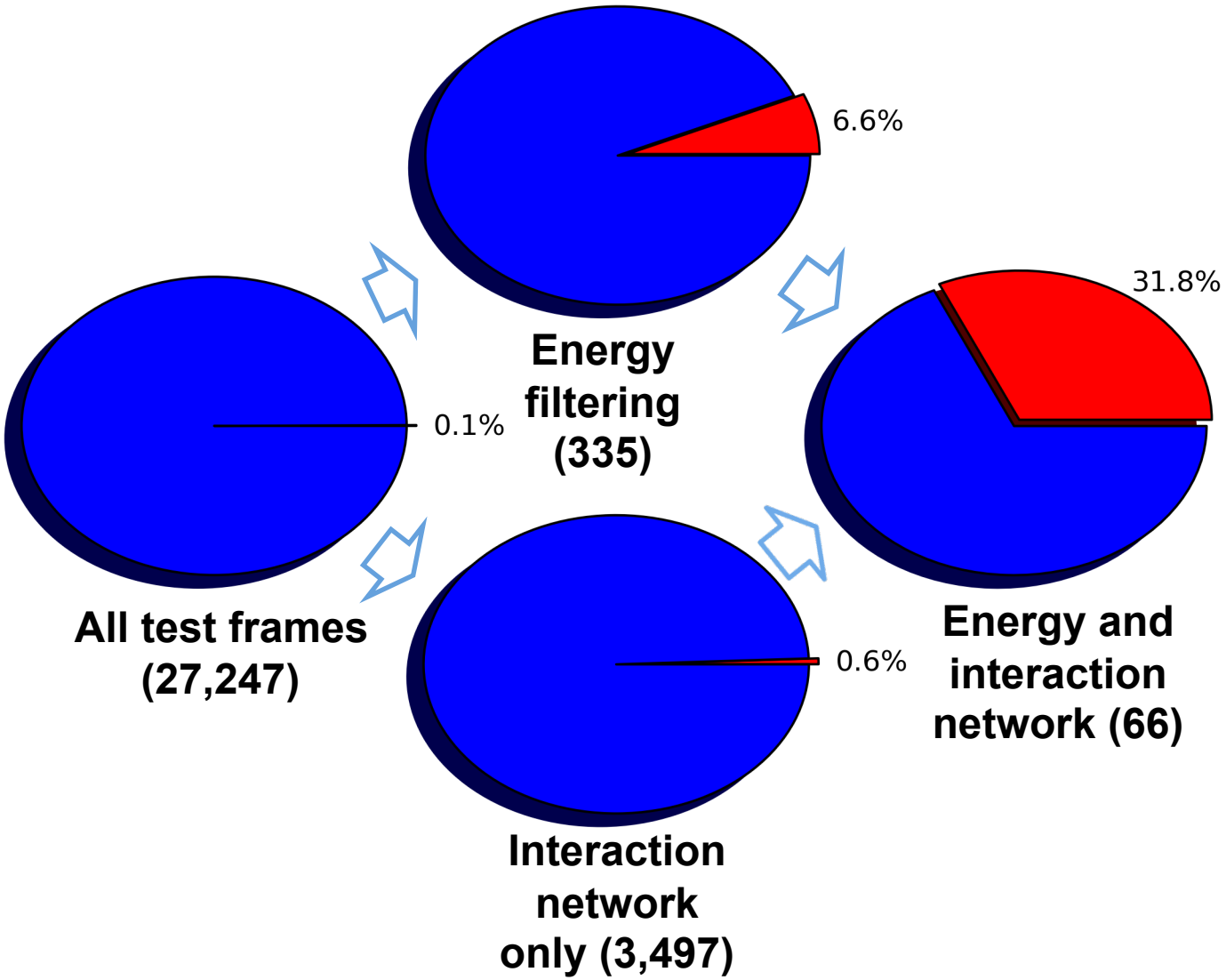
Increasing enrichment in known binders



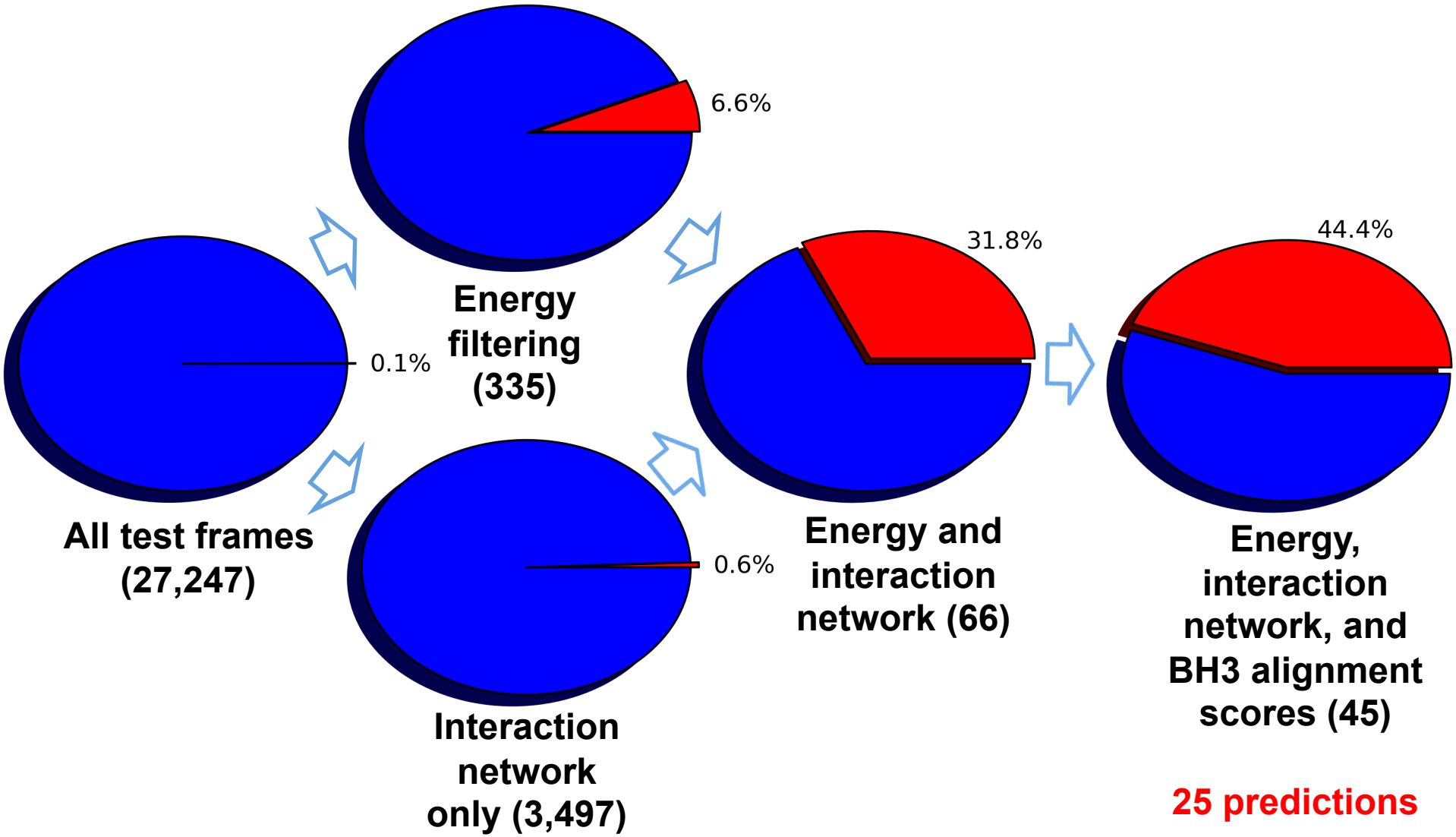
Interaction network screening (HPRED data)



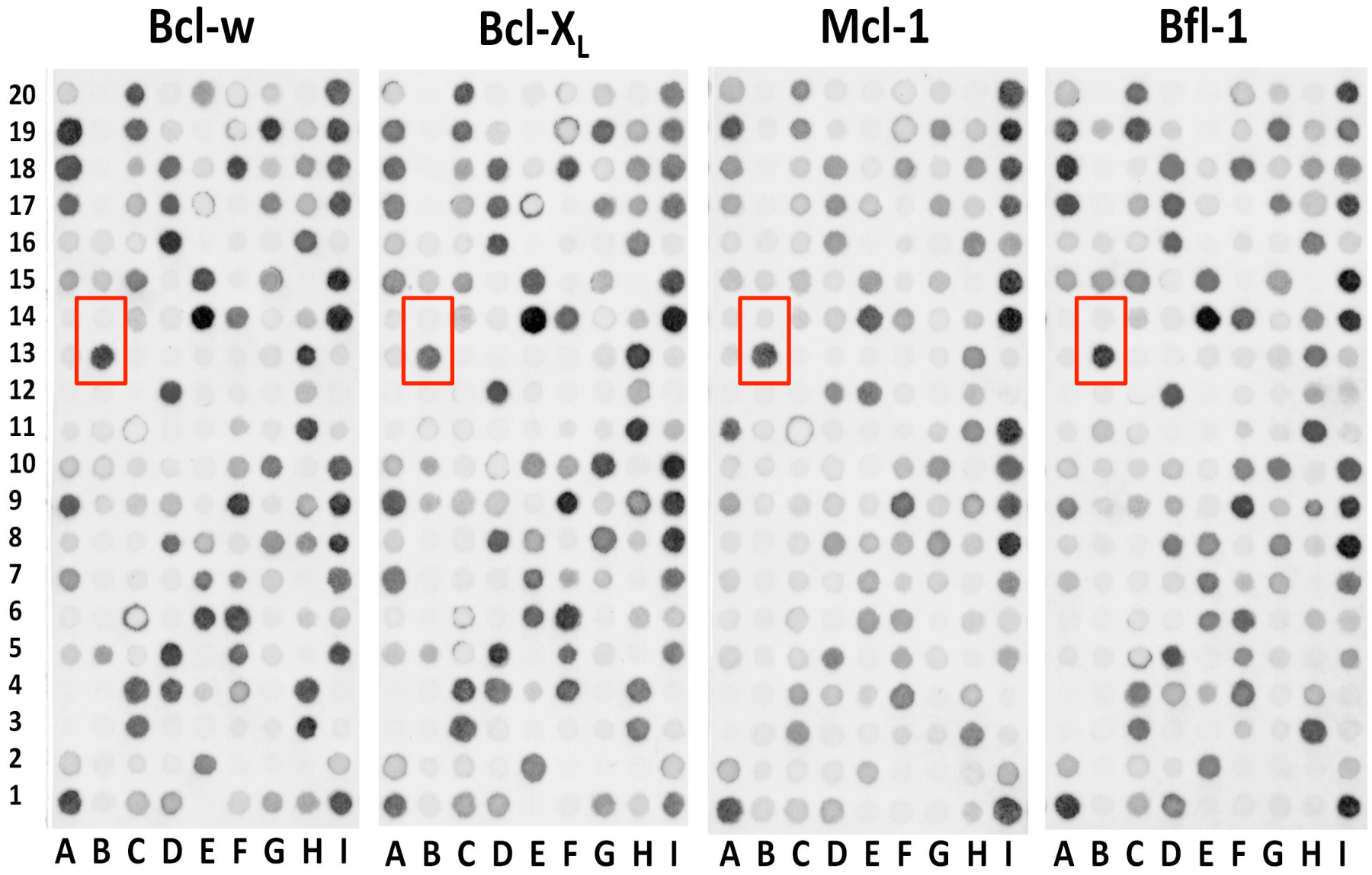
Increasing enrichment in known binders



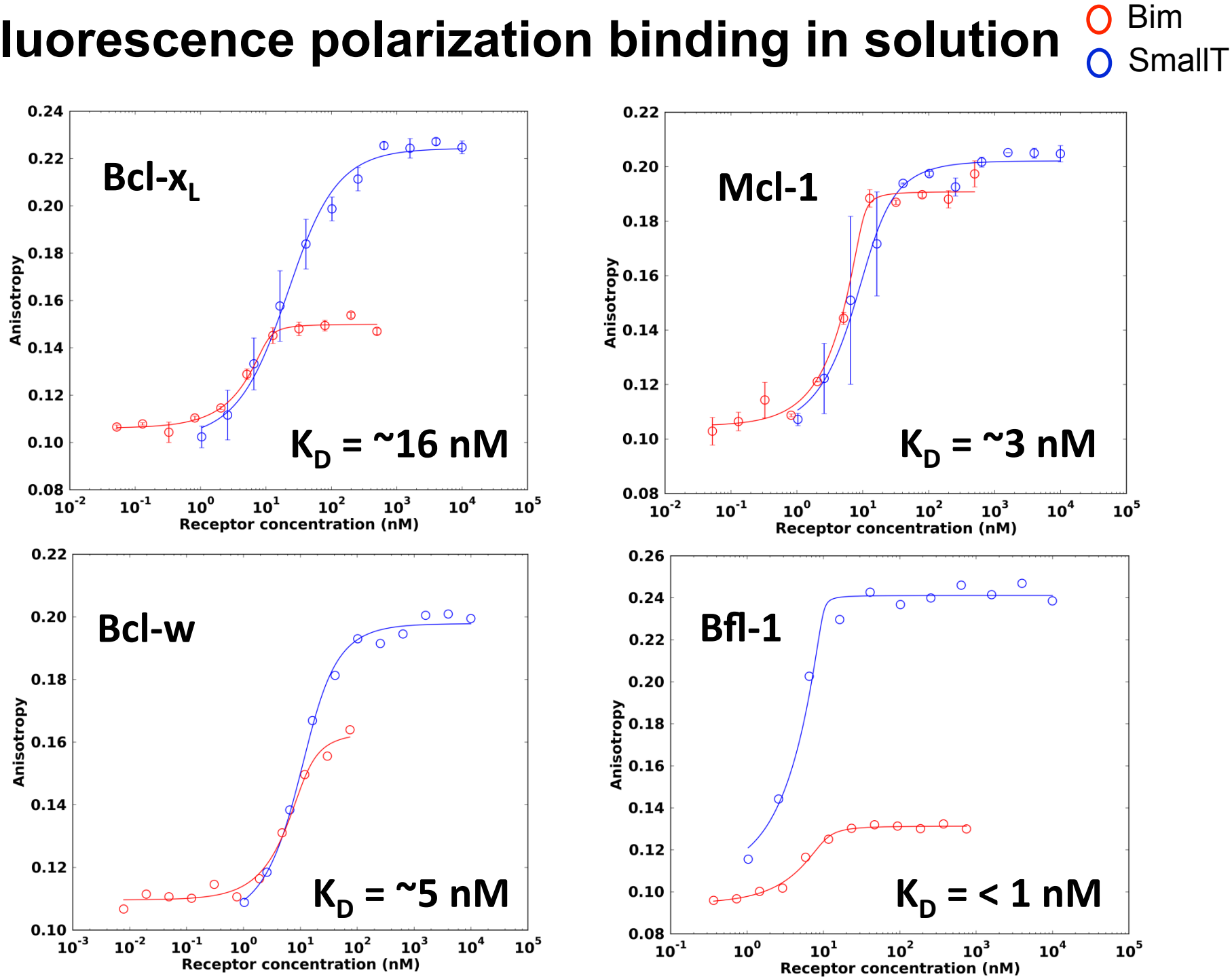
Increasing enrichment in known binders



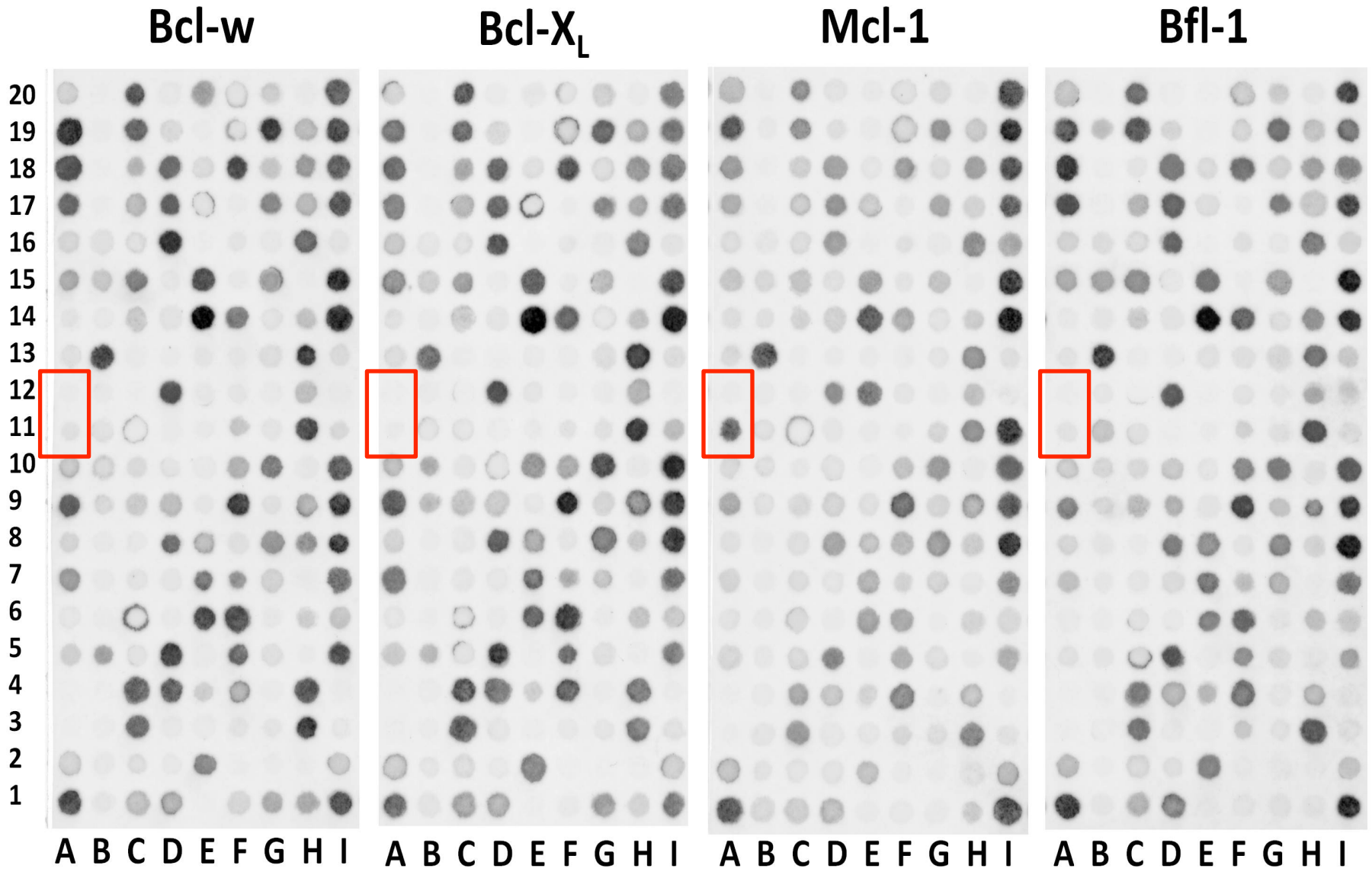
SPOT arrays reveal many potential genome hits



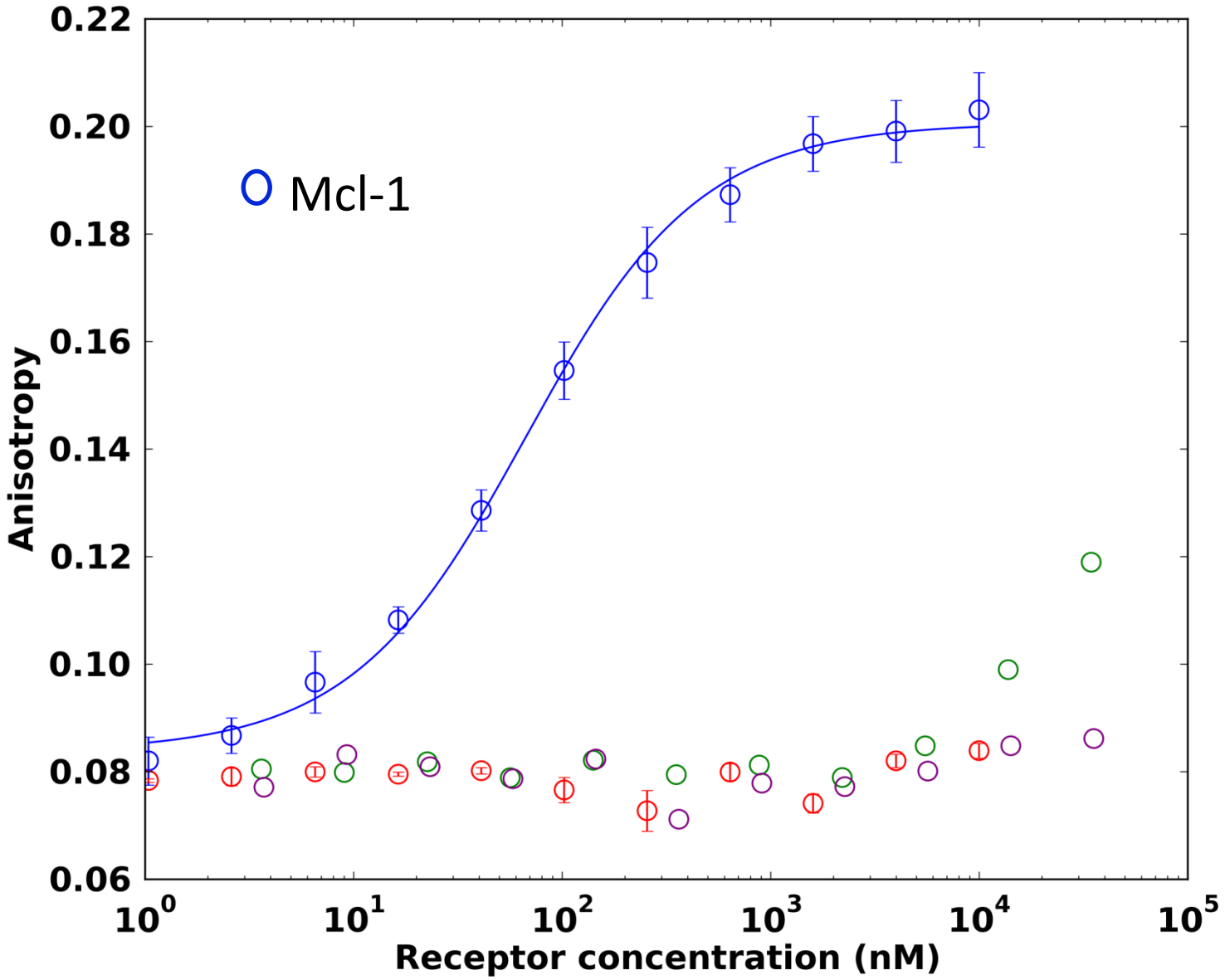
Fluorescence polarization binding in solution



Some hits are selective for specific receptors



Solution assays reveal binding to Mcl-1 at ~60 nM



Current/Future directions

STATIUM: When does it work? Why does it work? How can it be improved? Can it be used in conjunction with FlexPep Bind, which accounts much better for structural flexibility and detailed atomic interactions?

- Sanjib Dutta
- Emiko Fire
- Stefano Gullá

- Joe DeBartolo
- T. Scott Chen

NIGMS, NSF funded computing cluster, MIT BioMicro Center