Molecular Dynamics and Docking Approaches to Characterize Drug Resistance and Assist Lead Discovery

Robert C. Rizzo

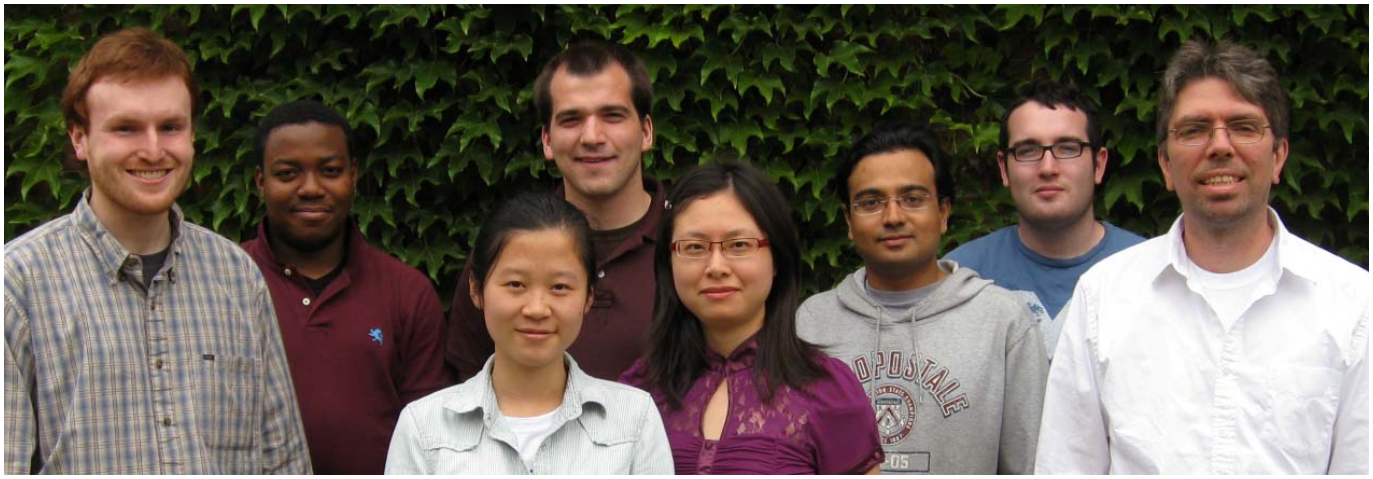
Department of Applied Mathematics and Statistics, and
Institute of Chemical Biology & Drug Discovery

Stony Brook University, Stony Brook, NY 11794

2011 IPAM Meeting

Acknowledgments

Current Students:
Trent Balius
Patrick Holden
Yulin Huang
Lingling Jiang
Sudipto Mukherjee
Jibril Ashiru-Balogun
Brian Shea



Collaborators:
HIV: Miriam Gochin (Touro/UCSF)
HIV: Mark Wilson (U of Nebraska)
HIV: Iwao Ojima ICB&DD (SBU)
Flu: Janet Hearing (SBU)
Kinases: Todd Miller (SBU)
DOCK: Tack Kuntz (UCSF), Brian Shoichet (UCSF), Dave Case (Scripps), Carlos Sosa (IBM)

Support:
Applied Math & Stat (SBU)
Computational Science Center at Brookhaven National Laboratory (NewYork Blue)
NYSTAR Investigator Award
School of Medicine (SBU) Influenza TRO, Carol M. Baldwin Breast Cancer Research Award
National Institutes of Health R01GM083669 (NIGMS), F31CA134201 (NCI, to TEB)

Lab Mission:

Develop, validate, and apply computational tools and protocols for drug discovery and design targeted towards human disease.

Driving Hypotheses:

Molecular recognition (drug binding, activity, potency, resistance) can be understood and predicted through interpretation of accurate computational results (structural and energetic).

Primary computational tools: Molecular Dynamics, docking

Goal: Quantify $\rightarrow$ Understand $\rightarrow$ Predict

Projects:

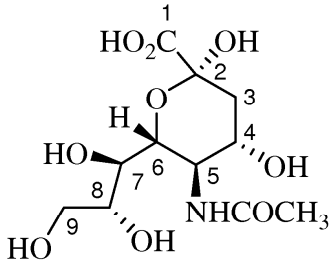
- Drug binding (activity, selectivity, resistance)
  - neuraminidase project: flu
  - kinase project: cancer
  - gp41 project: HIV

- All projects have lead discovery component
  - virtual screening
  - method / protocol development: DOCK

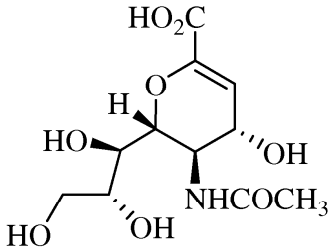
Flu Inhibitor Resistance

Chachra, R.; Rizzo, R. C. *J. Chem. Theory Comput.*, **2008**,*4*, 1526-1540

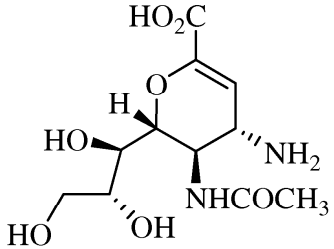
Exptl activities: neuraminidase (NA) subtype N9



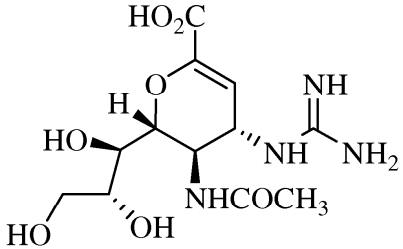
S00 (sialic acid)



S01



S02

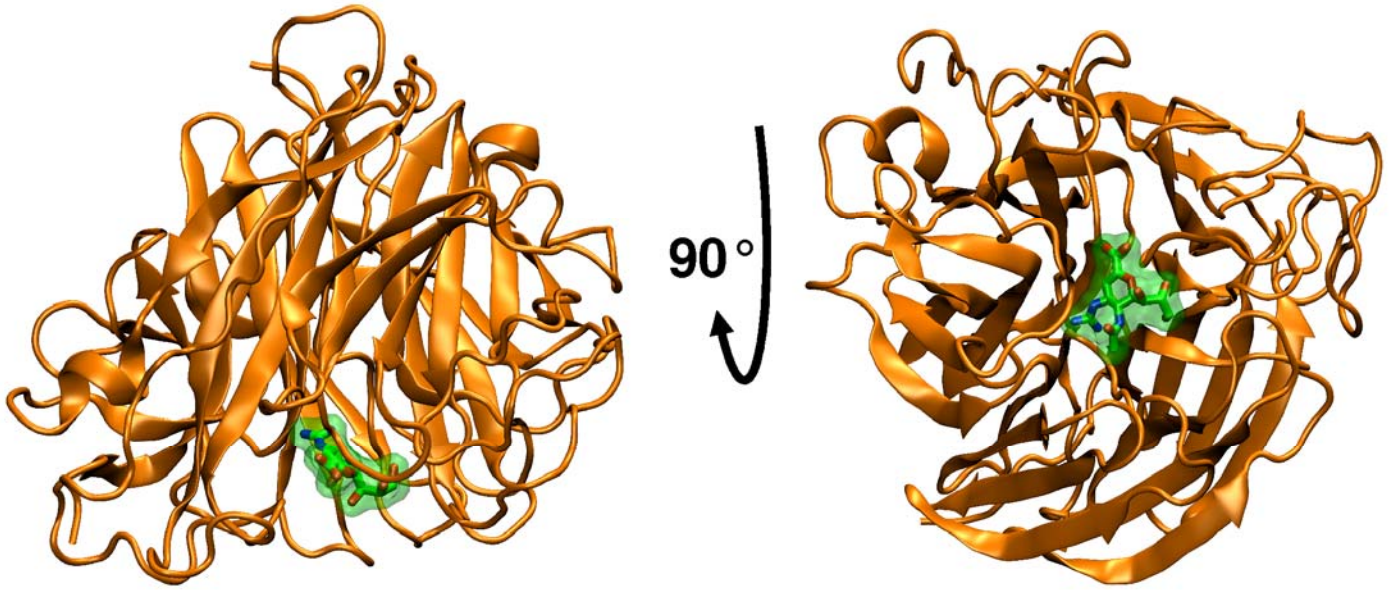


S03 (zanamivir)

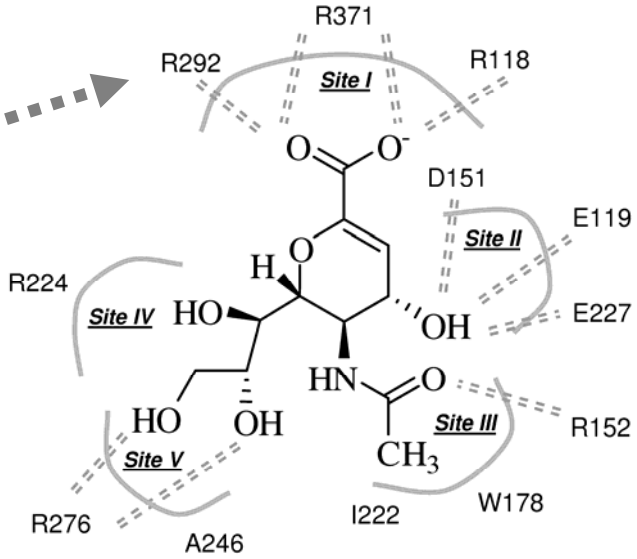
Code	WT K_i (uM)[b]	R292K K_i (uM)[b]	E119G K_i (uM)[c]
S00	55, 84.1	1820	73.7
S01	2.64	280	24.0
S02	0.148	14	0.0663
S03	0.002	0.033	0.0451

[a]**S00,** sialic acid, *N*-acetylneuraminic acid (Neu5Ac); **S01,** DANA, 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid, (Neu5Ac2en); **S02,** 4-amino Neu5Ac2en; **S03,** zanamivir, 4-guanidinio Neu5Ac2en. [b]McKimm-Breschkin, J.L.; et al. *J Virology* **1998**, *72* 2456-2462 for A/NWS/Tern/Australia/G70C (subtype N9). [c]Blick, T.J.; et al. *Virology* **1995**, *214* 475-484 for A/NWS/Tern/Australia/G70C (subtype N9). Free energies of binding estimated as $\Delta G_{exptl} \approx RTln(K_i$ in molar) at 37 °C.

Known inhibitor binding geometry (pose) with NA



N9 mutation site



Coordinates from
Smith, B.J.; et al.,
Protein Sci. **2001**, *10* 689-96

Schematic adapted from Stoll et al.,
Biochemistry **2003**, *42*, 718-727

All-atom computational modeling: Molecular Mechanics

- Each atom is represented as a sphere in 3-D space
  - atom type (C, N, H, O, S, etc)
  - radius (size)
  - partial charge

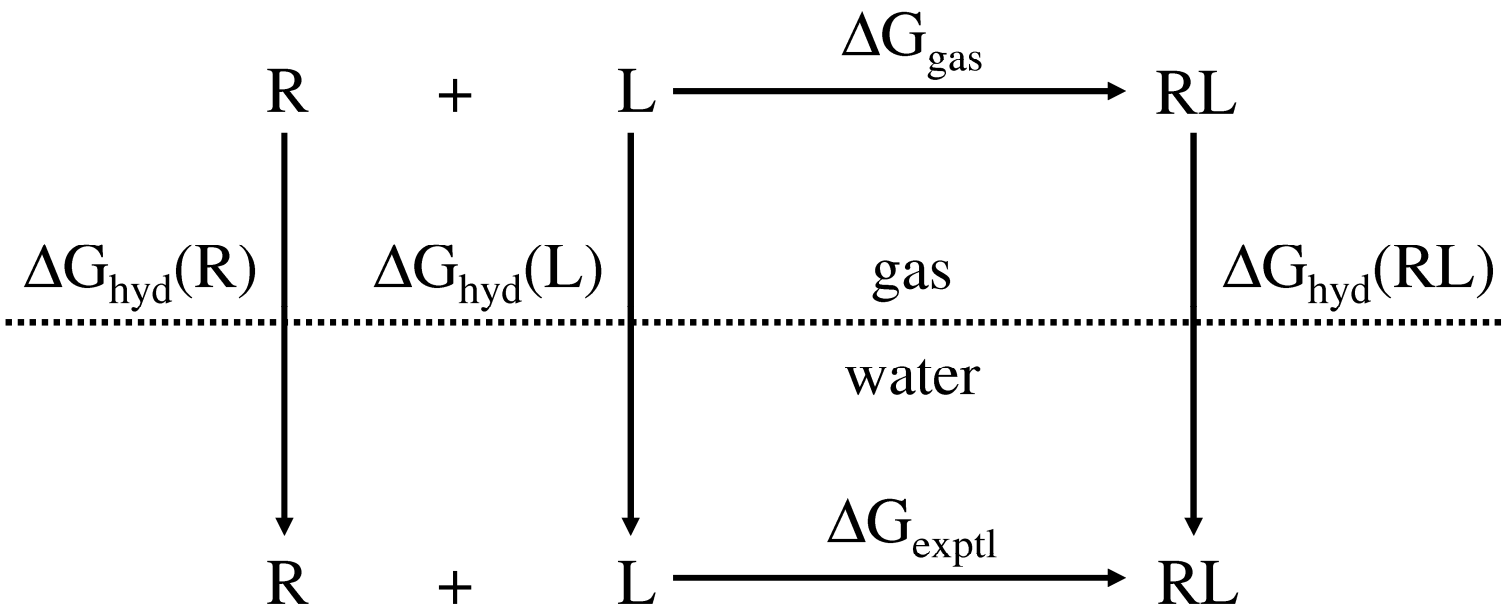
- Energy = bonds + angles + torsions + charge-charge + steric fit

- Simulation setups constructed using crystallographic coordinates and model building

- Explicit solvent molecular dynamics simulations and analysis

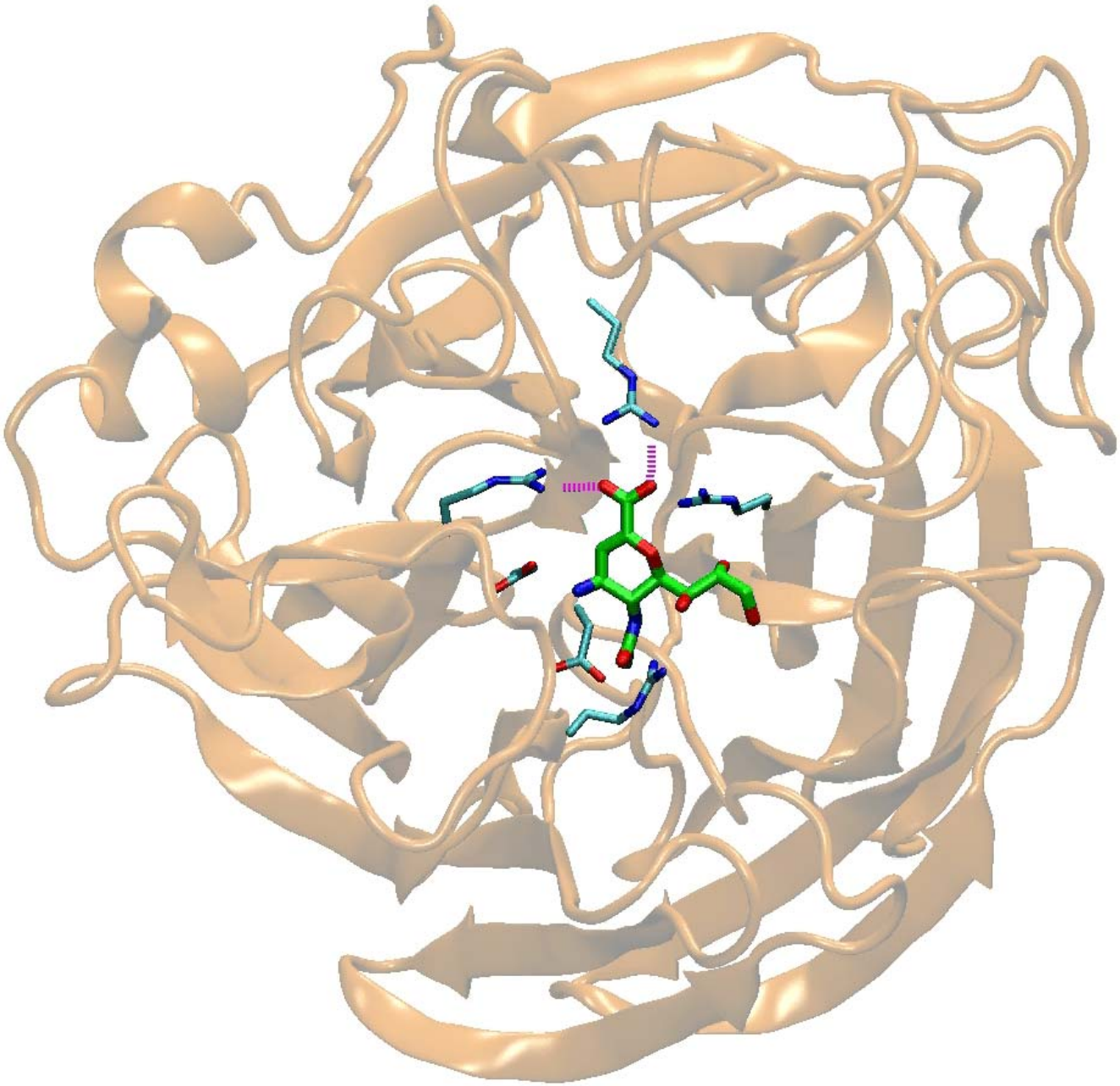
- Simulations yield detailed atomic level structures and energies

Computational protocol: Free energy of binding (ΔG_{bind}) estimates: MM-GBSA

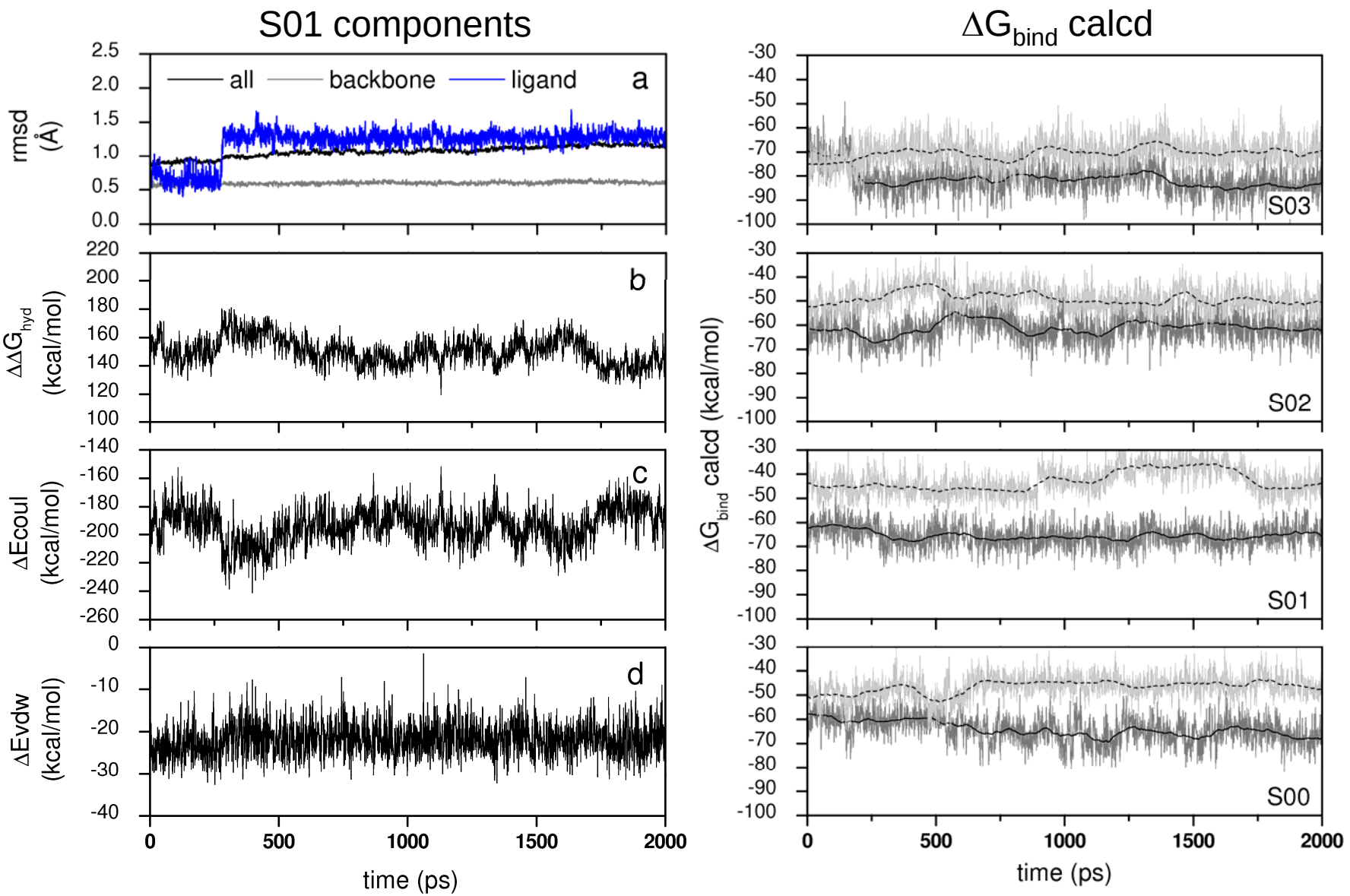


$$\Delta G_{bind} = \Delta G_{gas} + \Delta G_{hyd}(RL) - [\Delta G_{hyd}(R) + \Delta G_{hyd}(L)] \approx \Delta G_{exptl}$$

Srinivasan, J. ; et al. *J. Am. Chem. Soc*. **1998**, *120*, 9401-9409



Simulation stability / convergence

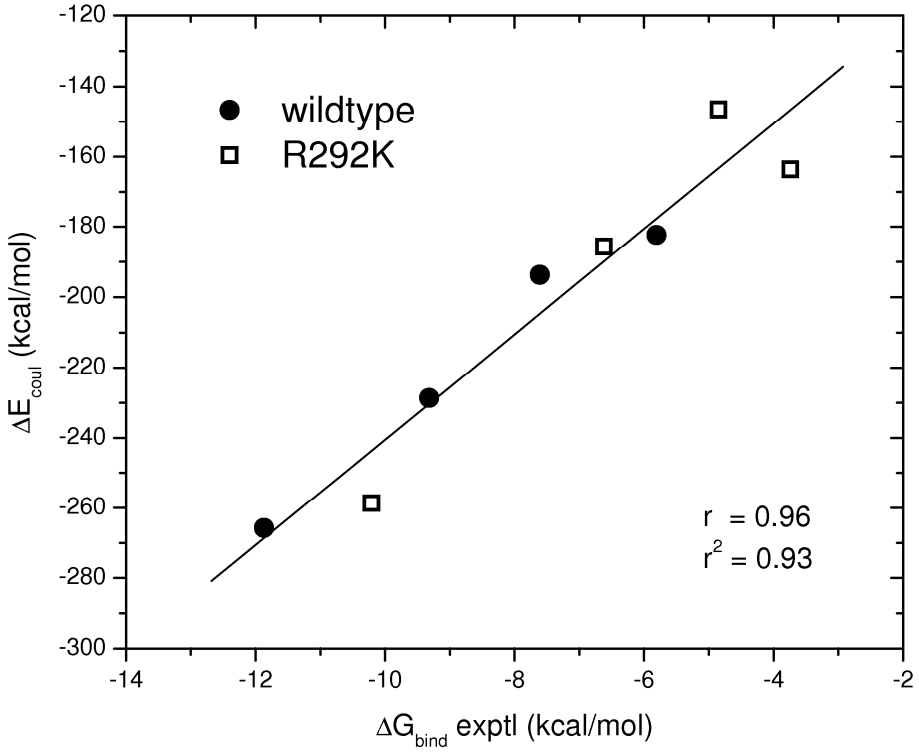


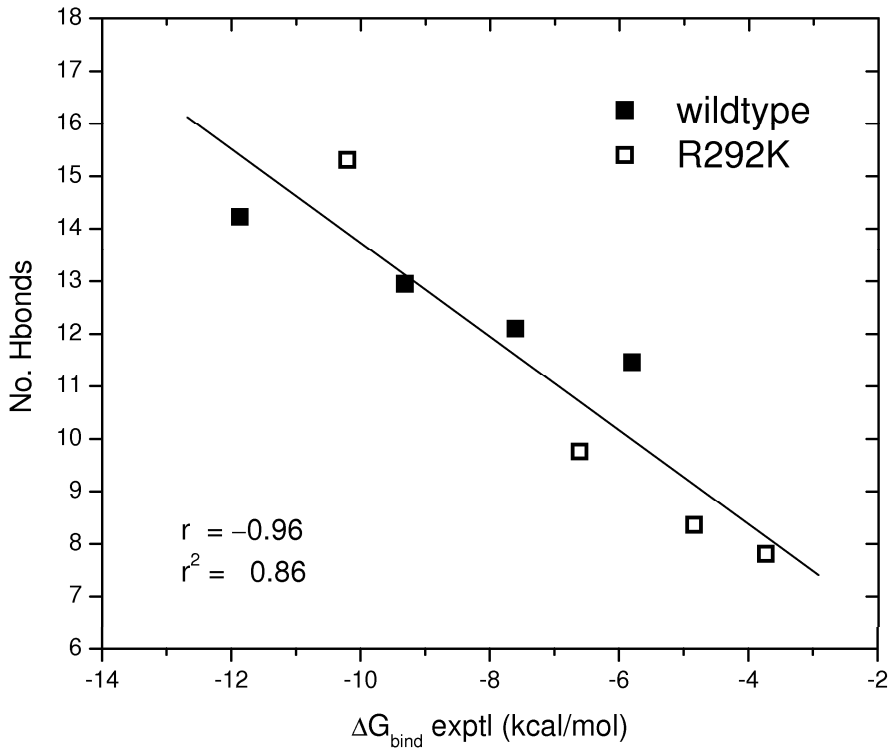
Energy component breakdown with exptl binding

system	ΔE_{vdw} A	ΔE_{coul} B	$\Delta\Delta G_{polar}$ C	$\Delta\Delta G_{nonpolar}$ D	ΔG_{bind} calcd = A+B+C+D	ΔG_{bind} expt $\approx RT \ln(K_i)$
$S00_{wildtype}$	-23.62 ± 0.09	-182.32 ± 0.31	146.36 ± 0.23	-4.35 ± 0.002	-64.07 ± 0.13	-5.81
$S01_{wildtype}$	-21.93 ± 0.09	-193.71 ± 0.29	154.70 ± 0.22	-4.28 ± 0.002	-65.58 ± 0.11	-7.61
$S02_{wildtype}$	-22.84 ± 0.11	-228.66 ± 0.33	194.81 ± 0.18	-4.41 ± 0.002	-61.18 ± 0.14	-9.32
$S03_{wildtype}$	-29.26 ± 0.10	-265.78 ± 0.29	218.79 ± 0.18	-4.70 ± 0.001	-81.26 ± 0.15	-11.87
$S00_{R292K}$	-23.98 ± 0.10	-163.63 ± 0.33	145.77 ± 0.27	-4.36 ± 0.003	-46.56 ± 0.11	-3.74
$S01_{R292K}$	-23.31 ± 0.08	-146.52 ± 0.25	131.67 ± 0.19	-4.37 ± 0.001	-42.85 ± 0.13	-4.85
$S02_{R292K}$	-21.40 ± 0.10	-185.63 ± 0.29	162.95 ± 0.21	-4.36 ± 0.002	-49.02 ± 0.12	-6.62
$S03_{R292K}$	-22.92 ± 0.10	-258.67 ± 0.30	216.12 ± 0.20	-4.63 ± 0.001	-70.49 ± 0.13	-10.21
	$r^2 = 0.23$	$r^2 = 0.93$	$r^2 = 0.88$	$r^2 = 0.62$	$r^2 = 0.76$	

ΔGb correlation

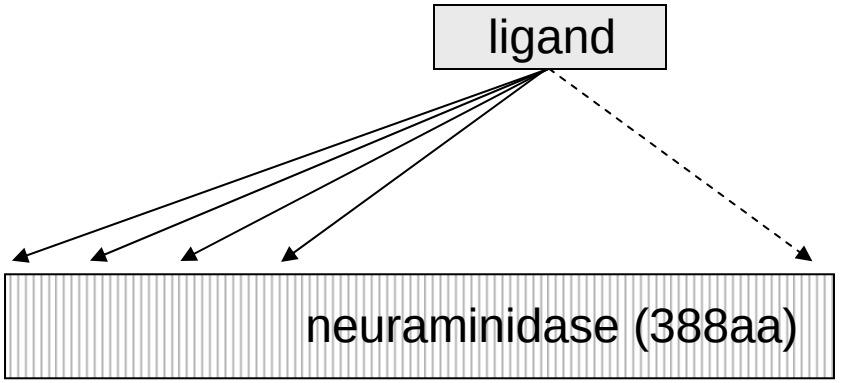
ΔE_{coul} and H-bond terms show strong correlation with ΔG_{bind} exptl



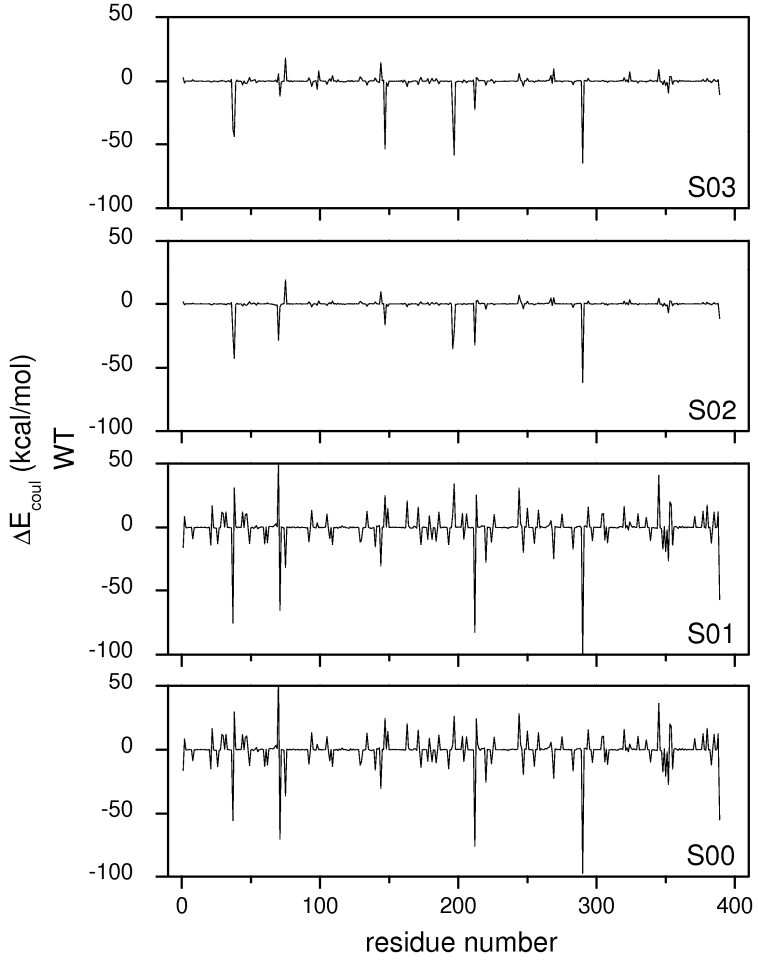


Which residues interact significantly with inhibitors? Per-residue energetic footprints (interaction signature)

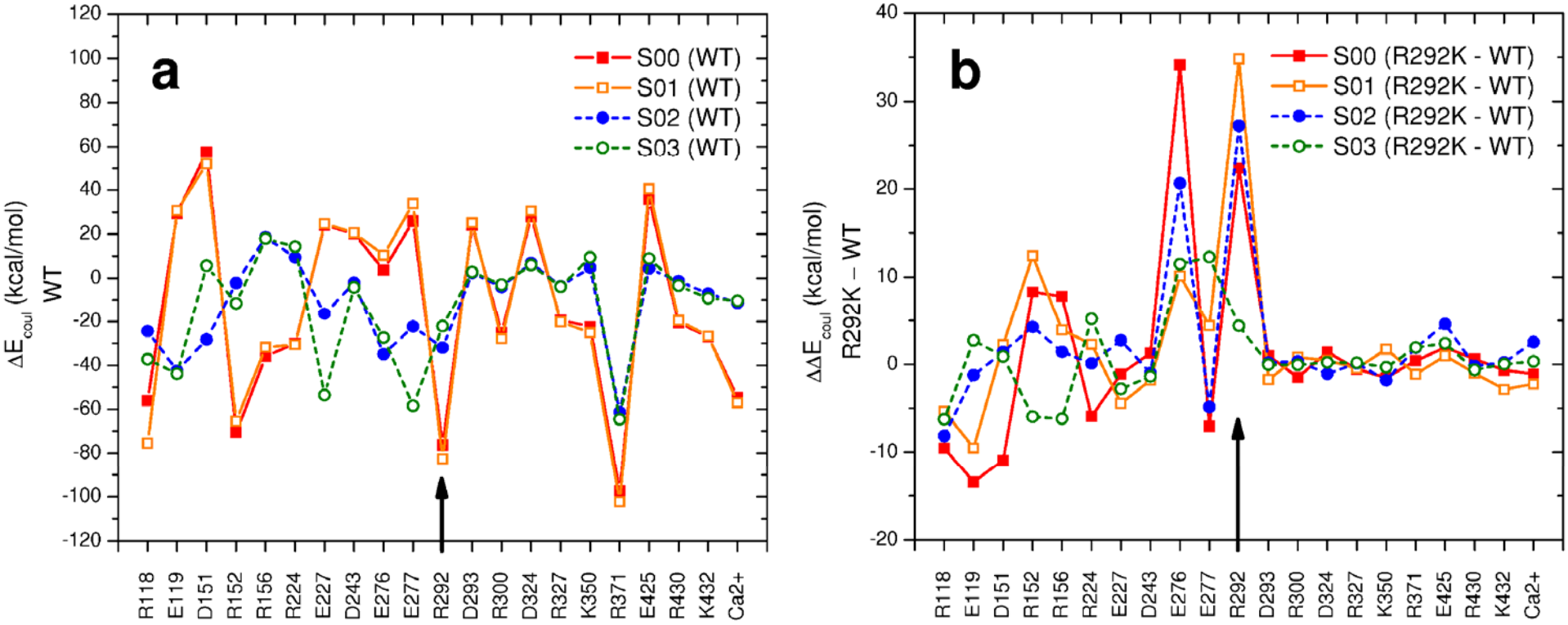
important receptor residues



WT Coulombic (ΔE_{coul}) footprints



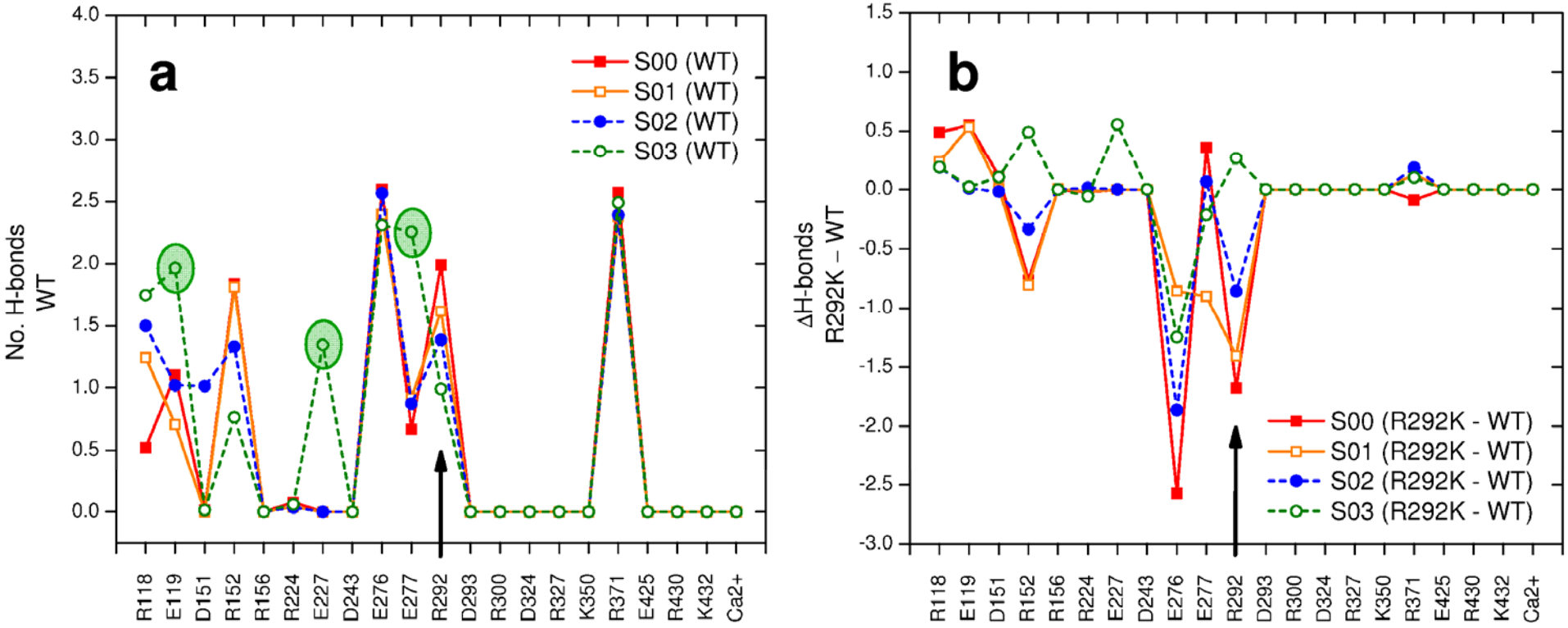
Coulombic footprints for: (a) wildtype and (b) mutant - wildtype



Ligands with formal charge of -1 (S00 and S01) have different WT footprint than do ligands with formal charge of 0 (S02 and S03).

profile suggests S03 relies less on R292K for intrinsic binding affinity

H-bond footprints for: (a) wildtype and (b) mutant - wildtype



S03 makes less H-bonds with R292

However, increased H-bonding observed at E119, E227, E277

S03 does not lose H-bonds as a result of the mutation

Conclusions (neuraminidase study)

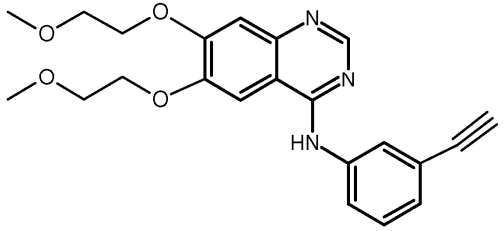
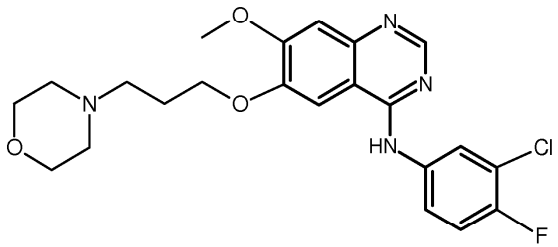
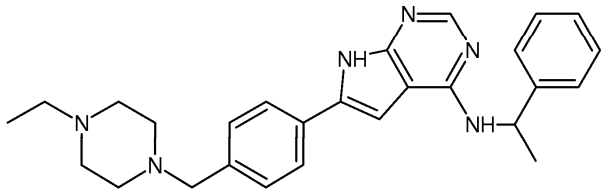
- WT/R292K simulations with N9 yield quantitative agreement with expt
- coulombic energy (ΔEcoul) correlates most strongly with exptl activities
- Hbond populations parallel exptl ordering
- S03: less reliance on R292K for intrinsic affinity, flatter ΔEcoul and ΔH-bond profiles
- S03: minor loss of $\Delta\Delta$Ecoul (no loss of H-bonds) when localized to 292
- S03: enhanced binding via E199, E227, and E277
- S03: less-ordered glycerol --- R292 interaction
- S03: greater avg distances to R292, shorter avg distances to E227 and E277

EGFR Inhibitor Resistance

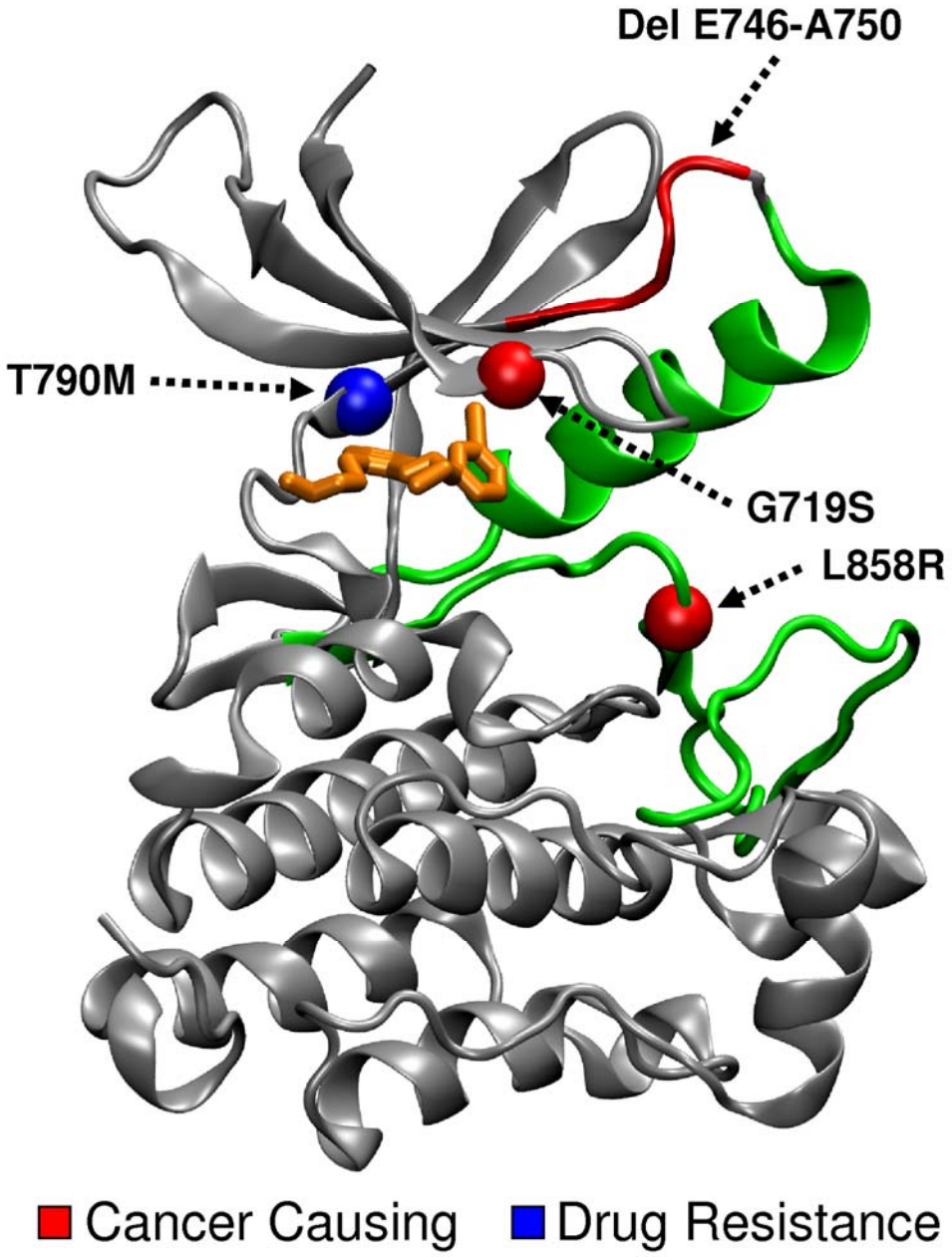
Balius, T. E.; Rizzo, R. C. *Biochemistry*, **2009,** *48,* 8435-8448

FR = ratio of activities (exptl or calcd)

Table 1. Experimental Fold Resistance (FR) values for ATP-competitive inhibitors with EGFR.

Inhibitor	Structure	Experimental Fold Resistance[a]		
		L858R / WT	**L858R&T790M / L858R**	**G719S / WT**
erlotinib		6.25 / 17.5 nM[b] 0.36 FR −0.61 $\Delta\Delta G_{FR}$	>10000 / 12.5 nM[c] >800 FR >3.96 $\Delta\Delta G_{FR}$	–
gefitinib		2.4 / 35.3 nM[d] 0.068 FR −1.59 $\Delta\Delta G_{FR}$	10.9 / 2.4 nM[d] 4.54 FR 0.90 $\Delta\Delta G_{FR}$	123.6 / 53.5 nM[e] 2.31 FR 0.50 $\Delta\Delta G_{FR}$
AEE788		1.1 / 5.3 nM[d] 0.21 FR −0.92 $\Delta\Delta G_{FR}$	18.6 / 1.1 nM[d] 16.9 FR 1.68 $\Delta\Delta G_{FR}$	11.3 / 10.9 nM[e] 1.04 FR 0.02 $\Delta\Delta G_{FR}$

[a]Fold Resistance (FR) = ratio of experimental activities. $\Delta\Delta G_{FR}$ exptl ≈ RTln(FR) at 298.15 K in kcal/mol. [b]Ki values (nM) from Carey, K. D., et al. , Cancer Res 66, 8163-8171. (2006). [c] IC_{50} values (nM) from Ji, H., et al., Proc Natl Acad Sci U S A 103, 7817-7822. (2006). [d] Kd values (nM) from Yun, C. H., et al., Proc Natl Acad Sci U S A 105, 2070-2075. (2008). [e]Kd values (nM) from Yun, C. H., et al., Cancer Cell 11, 217-227. (2007).



EGFR inhibitor pose and key mutations

- Cancer Causing
  - L858R

    increase binding to erlotinib & gefitinib
  - Del E746-A750

    increase binding to erlotinib & gefitinib
  - G719S

    decrease binding to gefitinib

- Drug Resistance
  - T790M

    decrease binding to all

Coordinates from Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. *J. Biol. Chem*., **2002**, 277, 46265-46272

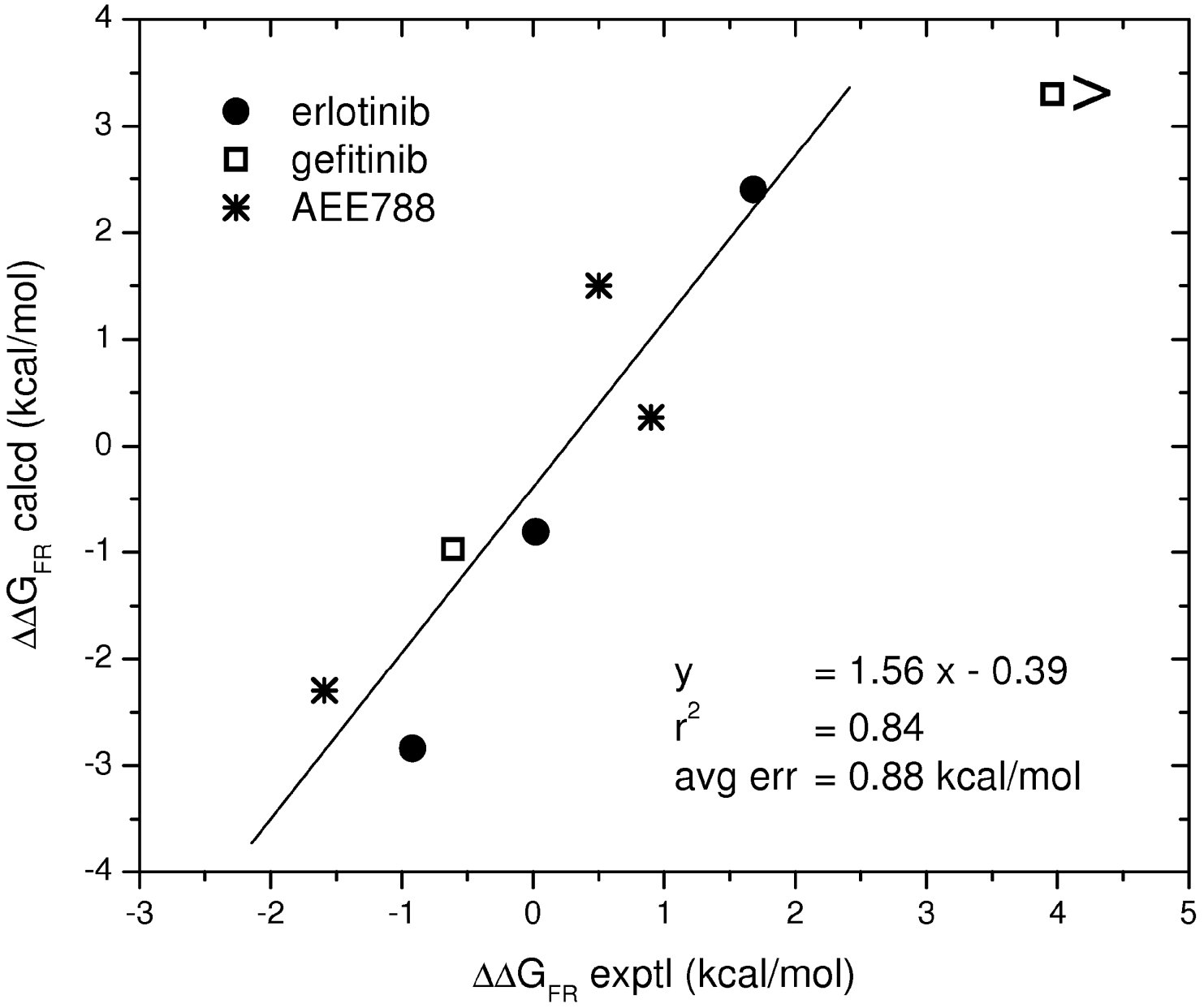
Relative Free Energies and Components

Table 2. Experimental versus calculated Fold Resistance (FR) energies ($\Delta\Delta G_{FR}$) and energy components for ligands with EGFR.

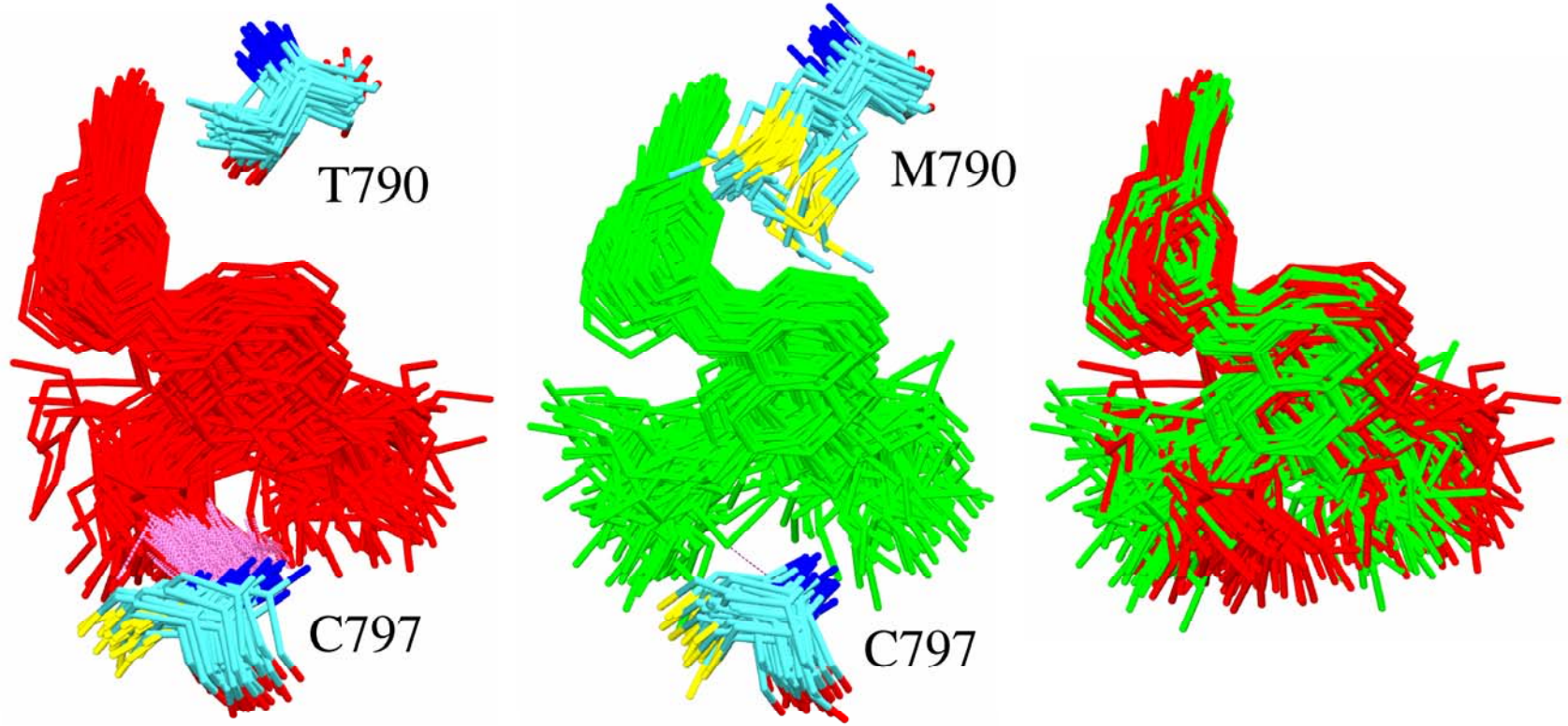
inhibitor	$\Delta\Delta E_{vdw}$ A	$\Delta\Delta E_{coul}$ B	$\Delta\Delta G_{polar}$ C	$\Delta\Delta G_{nonpolar}$ D	$\Delta\Delta G_{FR}$ calcd E =(A+B+C+D)	$\Delta\Delta G_{FR}$ exptl F
			L858R – WT			
erlotinib	–0.86 ± 0.06	–0.34 ± 0.13	0.21 ± 0.11	0.06 ± 0.003	–0.97 ± 0.07	–0.61
gefitinib	–0.99 ± 0.06	–0.72 ± 0.07	–0.58 ± 0.06	–0.01 ± 0.004	–2.30 ± 0.07	–1.59
AEE788	–2.41 ± 0.06	–0.48 ± 0.07	0.36 ± 0.06	–0.30 ± 0.005	–2.84 ± 0.06	–0.92
			L858R&T790M – L858R			
erlotinib	2.30 ± 0.06	7.42 ± 0.11	–6.56 ± 0.10	0.09 ± 0.003	3.30 ± 0.06	>3.96
gefitinib	–0.10 ± 0.05	–0.06 ± 0.07	0.49 ± 0.06	–0.06 ± 0.004	0.27 ± 0.06	0.90
AEE788	3.39 ± 0.07	3.15 ± 0.09	–4.33 ± 0.07	0.20 ± 0.004	2.40 ± 0.08	1.68
			G719S – WT			
erlotinib	–2.08 ± 0.06	–0.05 ± 0.12	–0.24 ± 0.11	0.04 ± 0.003	–2.38 ± 0.07	not reported
gefitinib	0.74 ± 0.07	–0.85 ± 0.07	1.59 ± 0.07	0.04 ± 0.004	1.50 ± 0.08	0.50
AEE788	–0.65 ± 0.06	–0.78 ± 0.06	0.55 ± 0.05	0.08 ± 0.005	–0.81 ± 0.07	0.02
r^2 =	0.70	0.47	0.19	0.30	0.84	7 data points[c]

[a]$\Delta\Delta G_{FR}$ calcd derived from the difference of two independent simulations (eg L858R – WT) computed using eqs 1-3. [b]$\Delta\Delta G_{FR}$ exptl values from Table 1. Correlations coefficients (r^2 values) obtained from fitting the change in each energy component to $\Delta\Delta G_{FR}$ exptl. All energies in kcal/mol ± standard errors of the mean from 5000 MD snapshots. [c]Data point for erlotinib with double mutant (>3.96) excluded from r^2 calculations given ambiguity in the experimental $\Delta\Delta G_{FR}$ measurement.

Correlation With Exptl Fold Resistance (ΔΔG)



Structure Comparison: Erlotinib

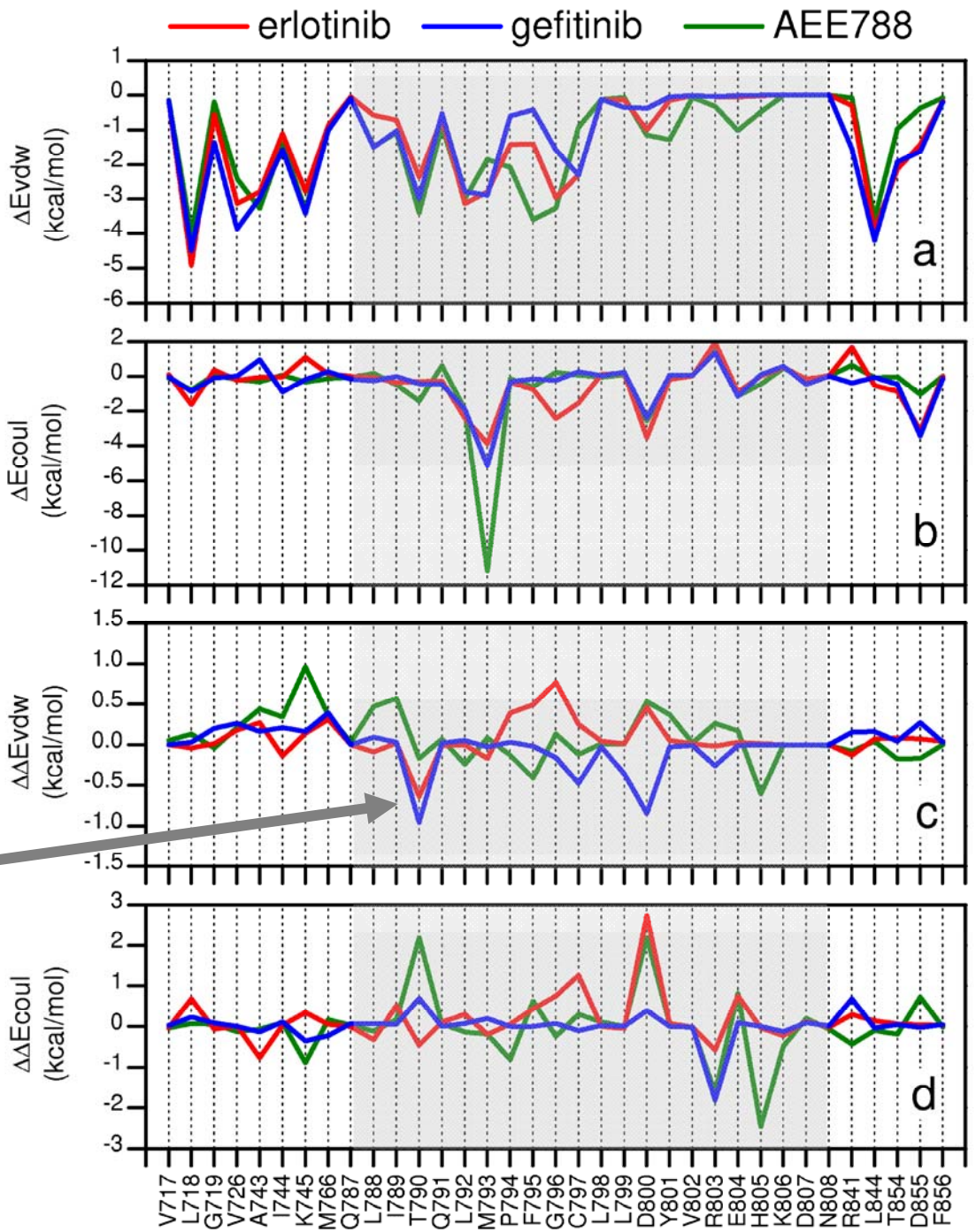


L858R simulations | **L858R&T790M simulations** | **overlay**

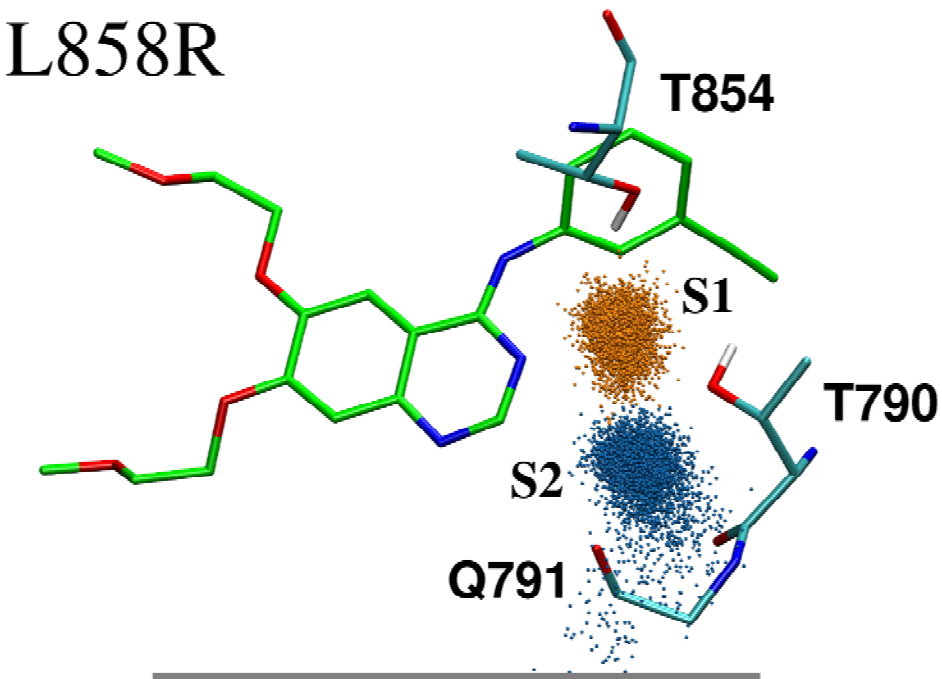
The T790M mutation does not lead to a steric clash with erlotinib however there is change in H-bonding at position C797

Energetic Footprints

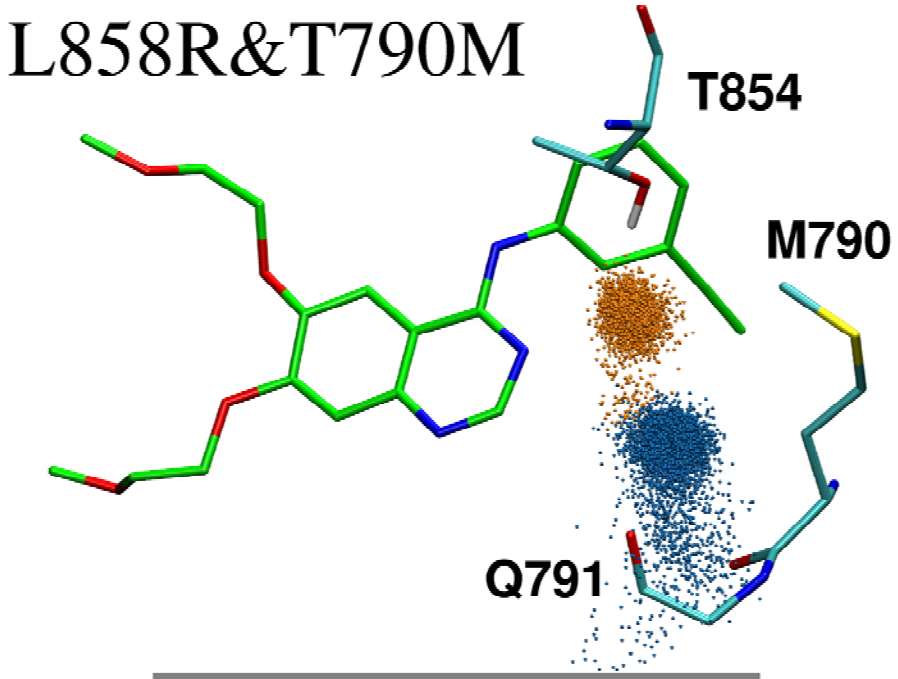
- Highly conserved and variant regions (suggests convergence)
- H-bonds visible at M793 in Coulombic graphs
- Delta FP appear flatter for gefitinib (less affected by mutations)
- No steric clash at T790 as previously proposed
- Computed losses suggests affinity differences for ATP are not the sole cause of drug resistance (recently proposed)



Water-mediated ligand binding and changes that result from T790M appear important



	S1 count	S2 count
erlotinib	4628	4168
gefitinib	4544	4385
AEE788	2489	5076



	S1 count	S2 count
erlotinib	3279	4111
gefitinib	3126	2536
AEE788	172	3563

Conclusions (EGFR Study)

- Good agreement with activity data ($\Delta\Delta G_{FR}$ calcd vs. exptl), $r^2 = 0.84$, N=7
- VDW is the most correlated term ($\Delta\Delta G_{FR}$)
- FP regions with similar and dissimilar energies suggest convergence/reproducibility
- Coulombic energies mirror H-bond trends
  - AEE788 shows largest interactions at M793
- Flatter ΔFP profiles for gefitinib shows agreement with exptl FR trend
- T790M resistance does not appear to be caused by a steric clash
  - increased favorable VDW interactions occur at M790
- Results suggest resistance not solely a function of increased affinity for ATP
  - strong correlations without ATP considerations, water-mediated changes

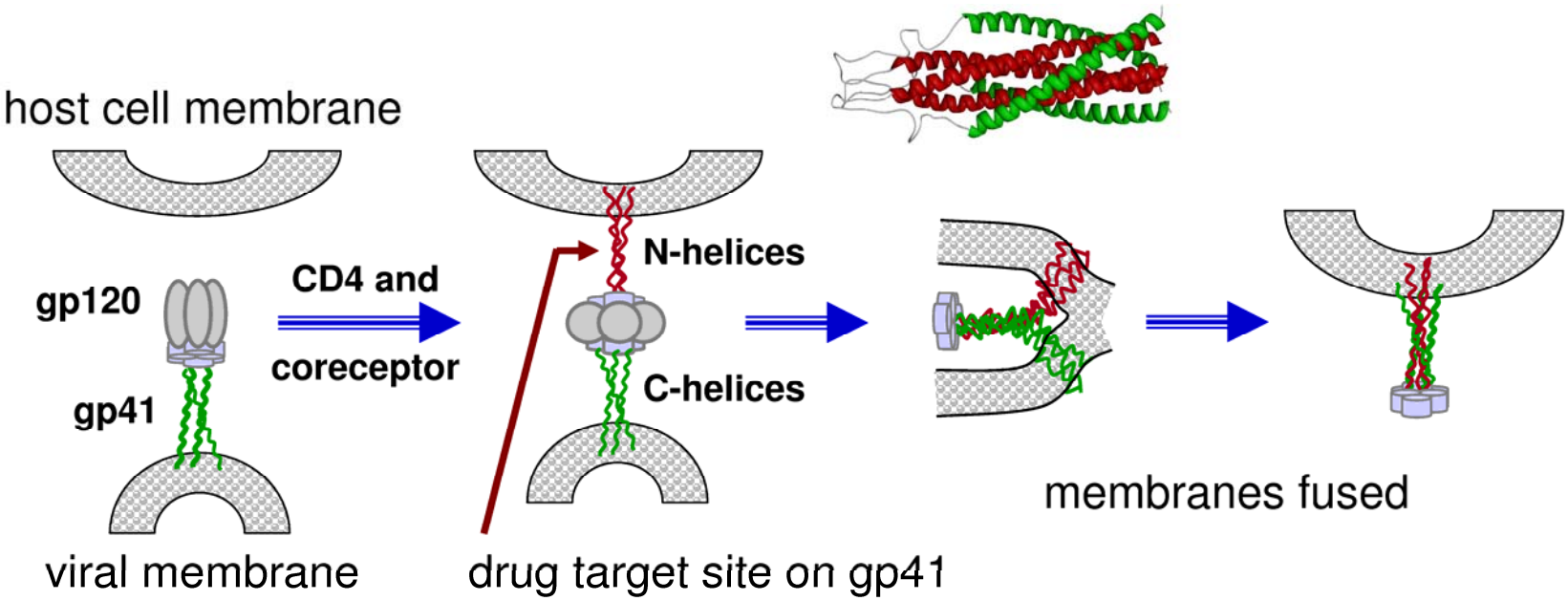
HIV Inhibitor Resistance

McGillick, B. E.; Balius, T.E.; Mukherjee, S.; Rizzo, R. C. *Biochemistry*, **2010**, *49,* 3575-3592

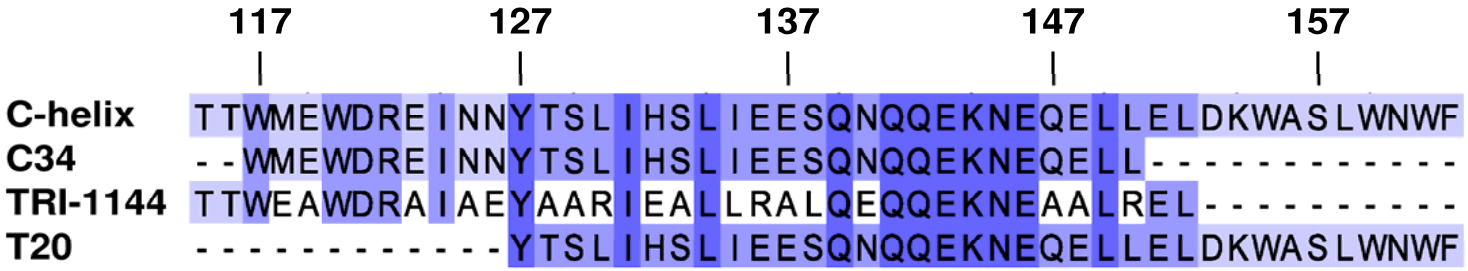
Viral Entry (membrane fusion)

Pre-fusion **Intermediate** **Fusogenic** **Post-fusion**

(exptl structures)

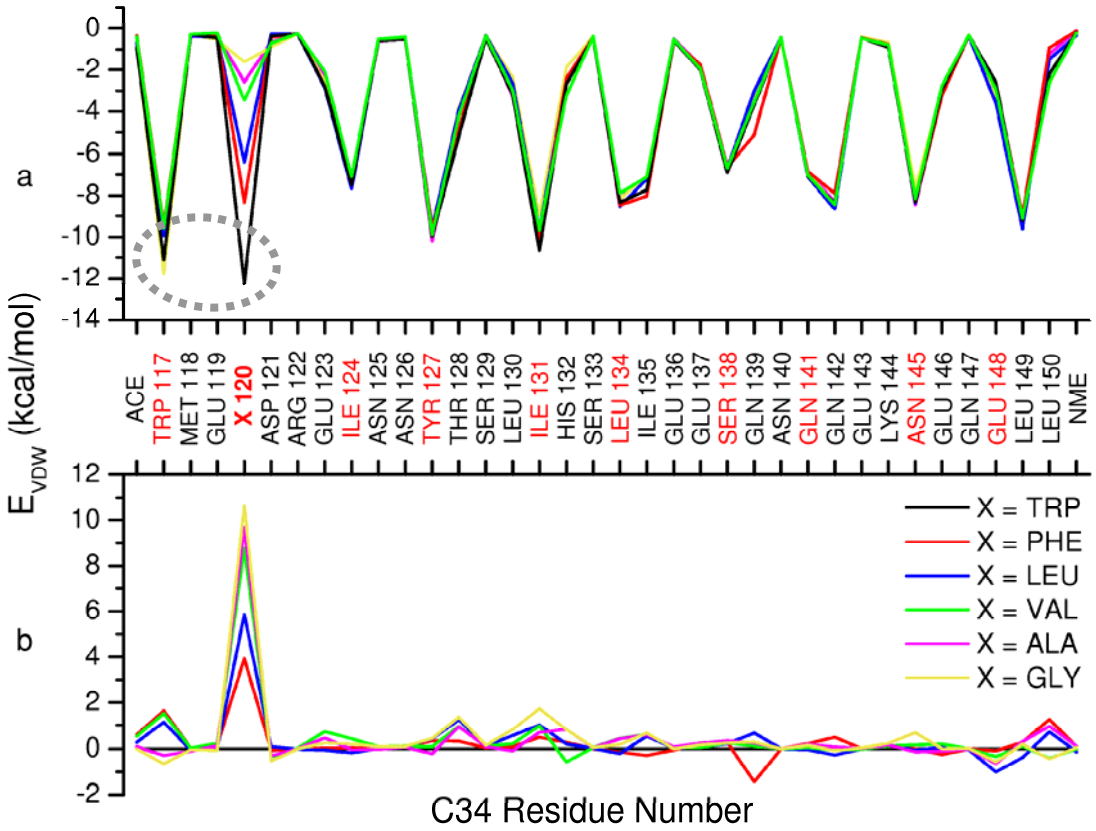


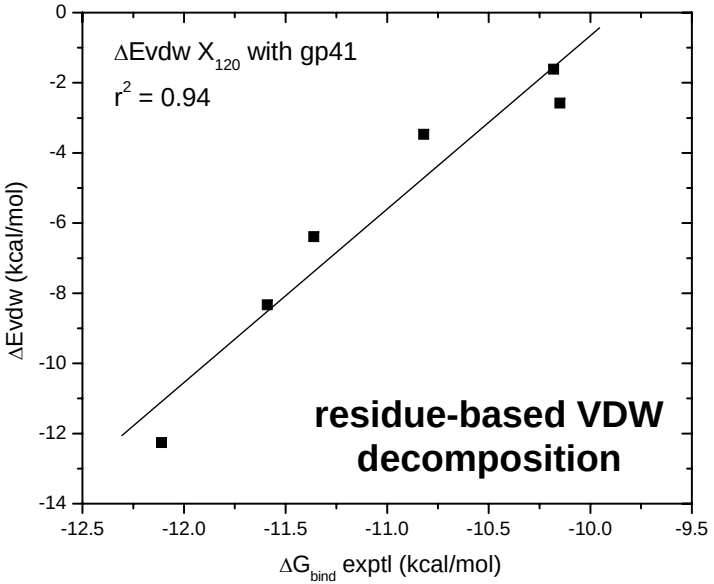
Adapted from Kim, P.; Eckert, D. M.; *PNAS* **2001,** *98*, 11187-11192



C34 and TRI-1144 are experimental inhibitors, T20 approved as Fuzeon 2003

Energy/footprint decomposition: C34

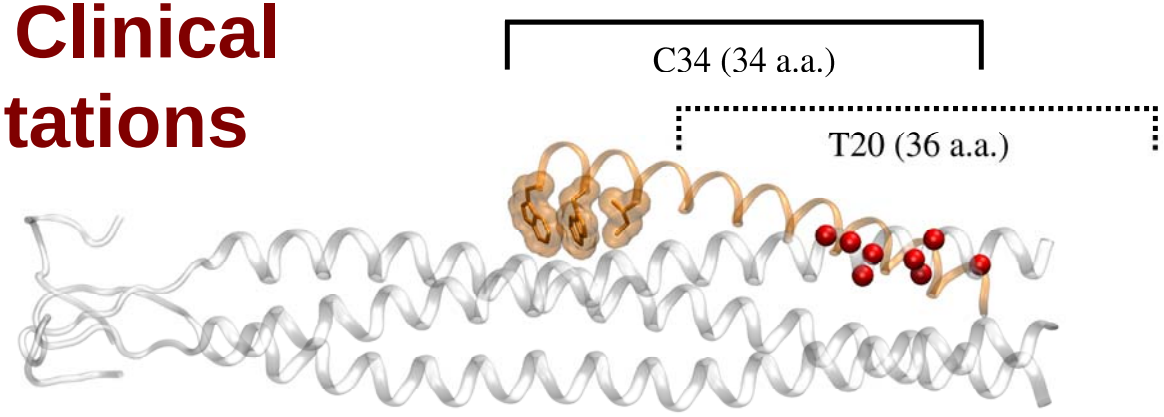




Peptide Inhibitor	ΔE_{coul} A	ΔE_{vdw} B	ΔG_{polar} C	$\Delta G_{nonpolar}$ D	$\Delta G_{electro}$ A+C	$\Delta G_{MM\text{-}GBSA}$ A+B+C+D	ΔG_{bind} exptl[b]
Trp120	-1090.66 ± 0.81	-144.58 ± 0.11	1142.36 ± 0.76	-16.55 ± 0.01	51.70	-109.42 ± 0.12	-12.11
Phe120	-1121.06 ± 0.81	-136.01 ± 0.10	1173.36 ± 0.77	-16.24 ± 0.01	52.30	-99.94 ± 0.11	-11.59
Leu120	-1145.73 ± 0.92	-133.66 ± 0.11	1192.61 ± 0.86	-16.23 ± 0.01	46.88	-103.00 ± 0.12	-11.36
Val120	-1070.50 ± 0.79	-129.34 ± 0.10	1119.90 ± 0.74	-15.51 ± 0.01	49.40	-95.45 ± 0.12	-10.82
Ala120	-1086.44 ± 0.90	-129.84 ± 0.10	1136.34 ± 0.88	-15.62 ± 0.01	49.90	-95.55 ± 0.11	-10.15
Gly120	-1078.06 ± 0.81	-127.48 ± 0.10	1124.30 ± 0.75	-15.62 ± 0.01	46.24	-96.85 ± 0.13	-10.18
	$r^2 = 0.22$	$r^2 = 0.84$	$r^2 = 0.27$	$r^2 = 0.84$	$r^2 = 0.32$	$r^2 = 0.75$	

Strockbine, B.; Rizzo, R. C. *Proteins: Struct. Func. Bioinformatics*, **2007**, *67*, 630-642

T20 Clinical Mutations



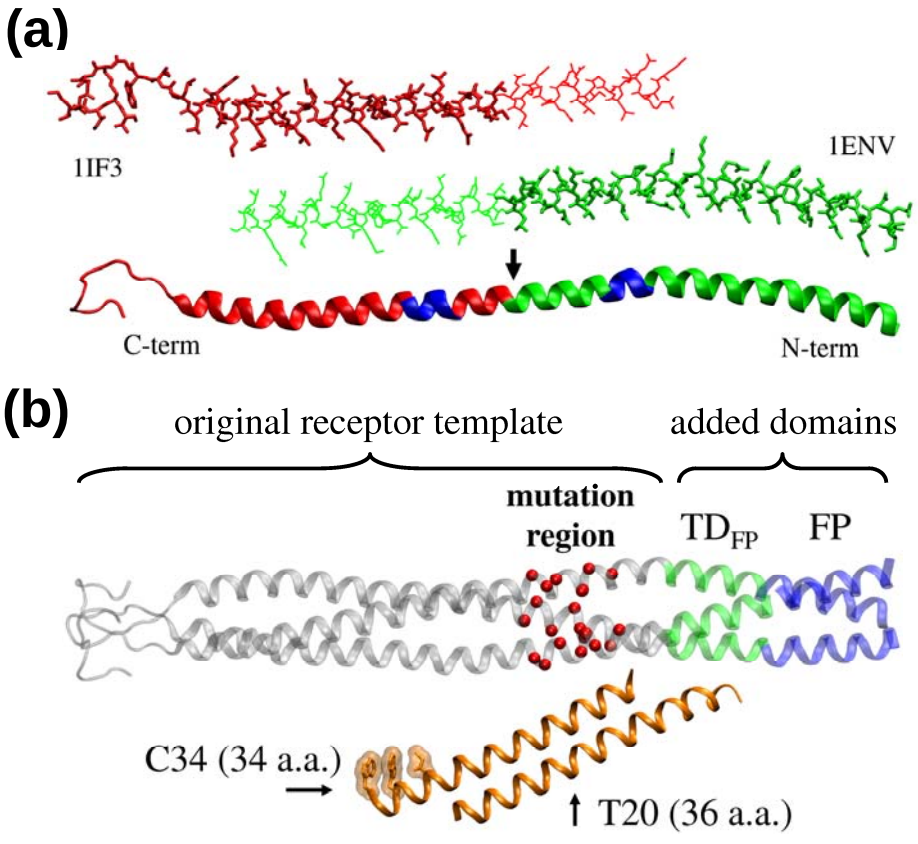
	mutation	FR = mutant/wildtype	$\Delta\Delta G_{FR} = RT \ln FR$[f]	
*	L33Q	924[a]	4.05	
*	L33S	64,[a] 108[a]	2.62[g]	2.46, 2.77
	G36D	8,[b] 10,[c] 46[d]	1.62[g]	1.23, 1.36, 2.27
	G36E	39[c]	2.17	
	G36S	7,[b] 7[c]	1.15	
*	G36V	45[d]	2.26	
*	I37K	212[e]	3.17	
	I37T	13[e]	1.15	
	V38A	16,[b] 18,[c] 54[d]	1.90[g]	1.64, 1.71, 2.36
*	V38E	494,[d] 1100[c]	3.92[g]	3.68, 4.15
	V38M	17,[d] 26[e]	1.81[g]	1.68, 1.93
*	Q40H	21,[b] 28,[c] 34[d]	1.95[g]	1.80, 1.97, 2.09
*	Q40K	1268[a]	4.23	
	Q41R	983,[a] 1137,[a] 1433,[a]	4.19[g]	4.08, 4.17, 4.31
	N42T	4,[b] 26[d]	1.38[g]	0.82, 1.93
	N43D	18,[b] 23,[c] 37[d]	1.90[g]	1.71, 1.86, 2.14

T20-gp41 structures not reported

→ model construction

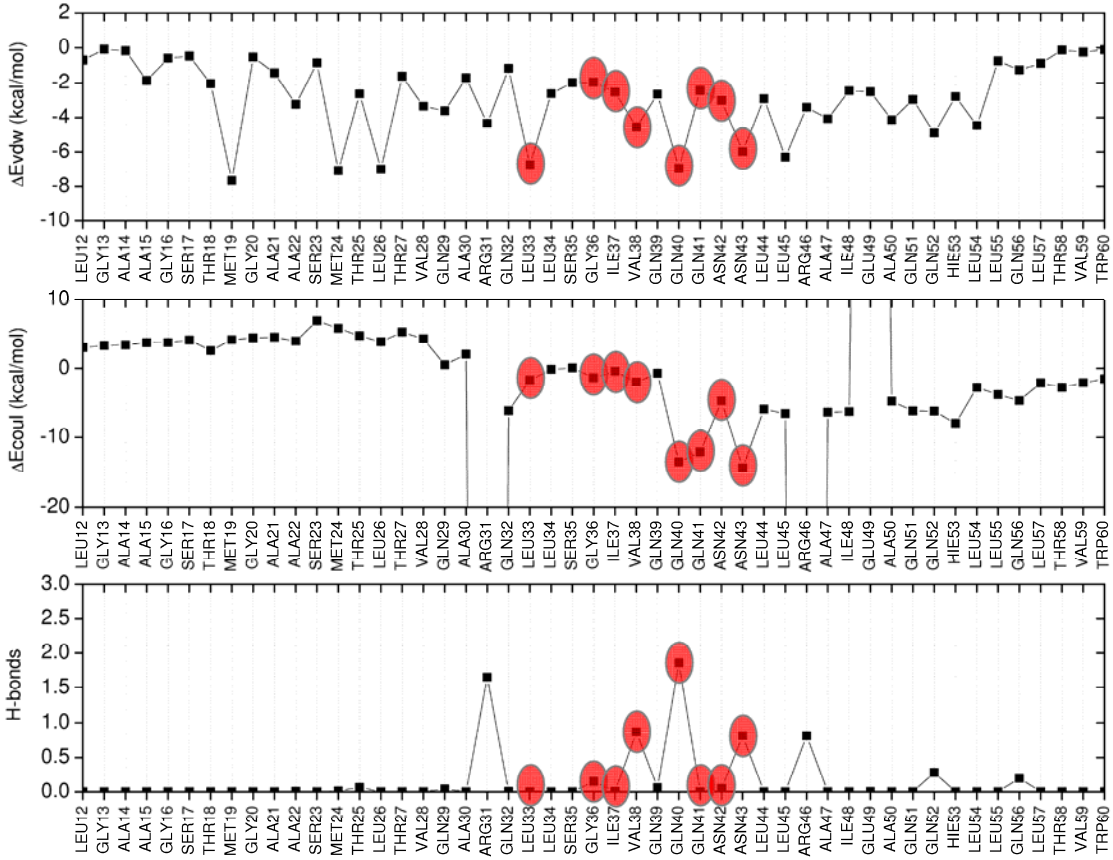
See McGillick, B. E.; et al *Biochemistry*, **2010**, 49, 3575–3592 for experimental references

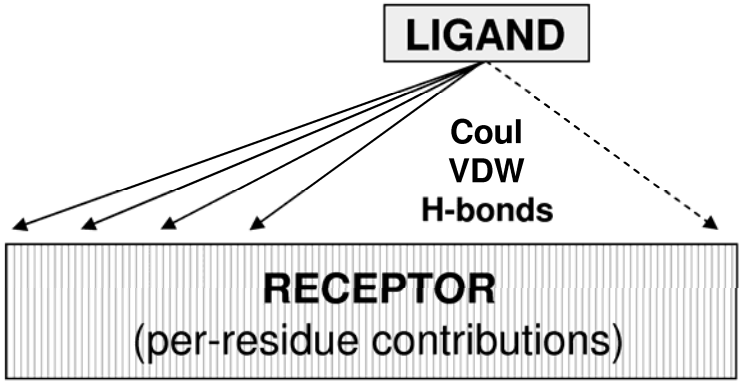
T20 Model & MD Protocol



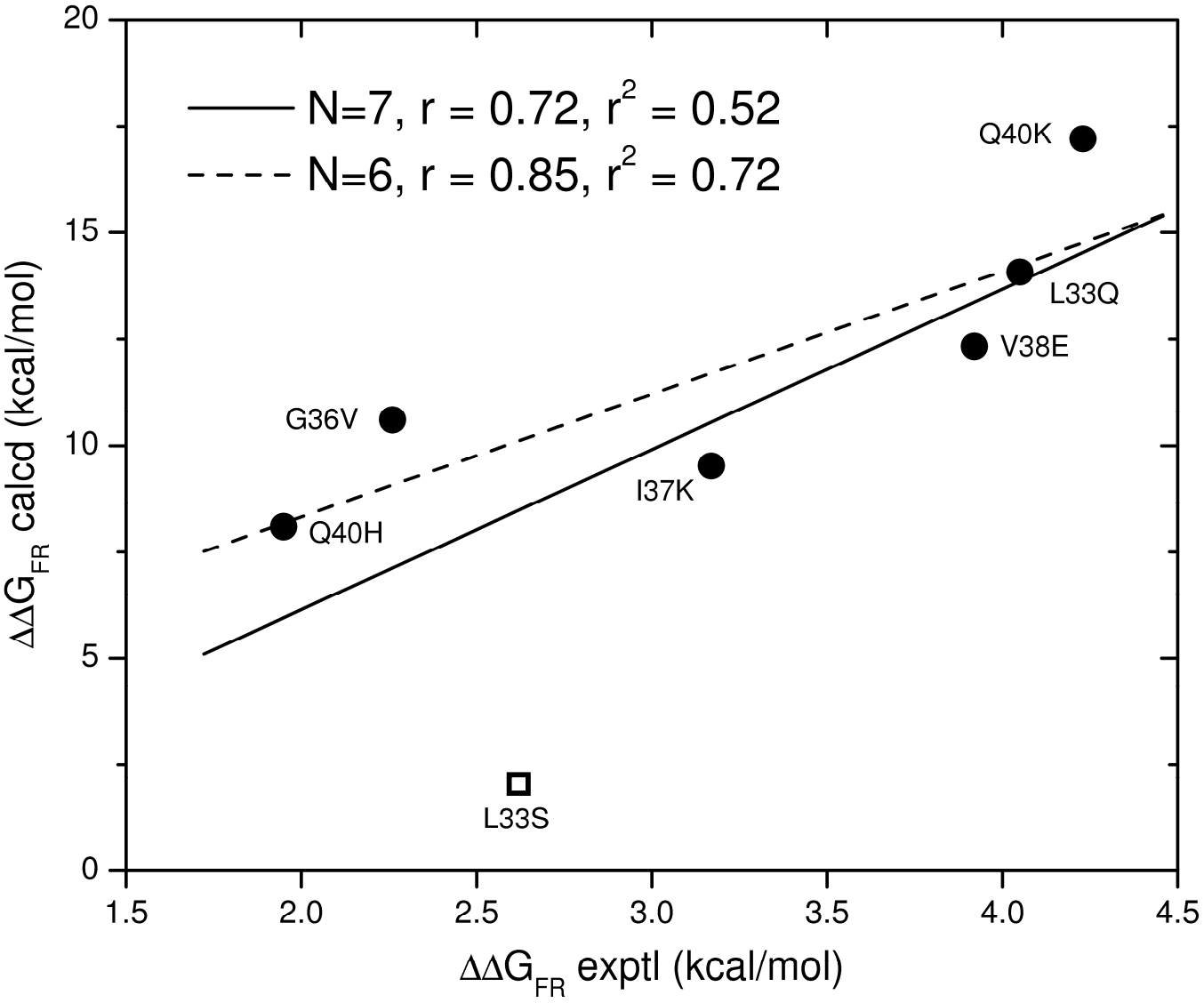
- Equilibration: Energy Minimization and MD
- Force Field: CHARMM27 (protein and lipids), TIP3P (water)
- Production Runs: NAMD 10,000ps (saved every 1 ps)
- NPT: T=310K, P=1.01325bar
- Complex
  - Periodic Boundary $60x60x200Å^3$
  - Analysis Details
    - Energy decomposition, footprinting, H-bond
    - MM-GBSA: binding energy method

Molecular footprints reveal energetically important gp41 residues which map to clinical mutation sites

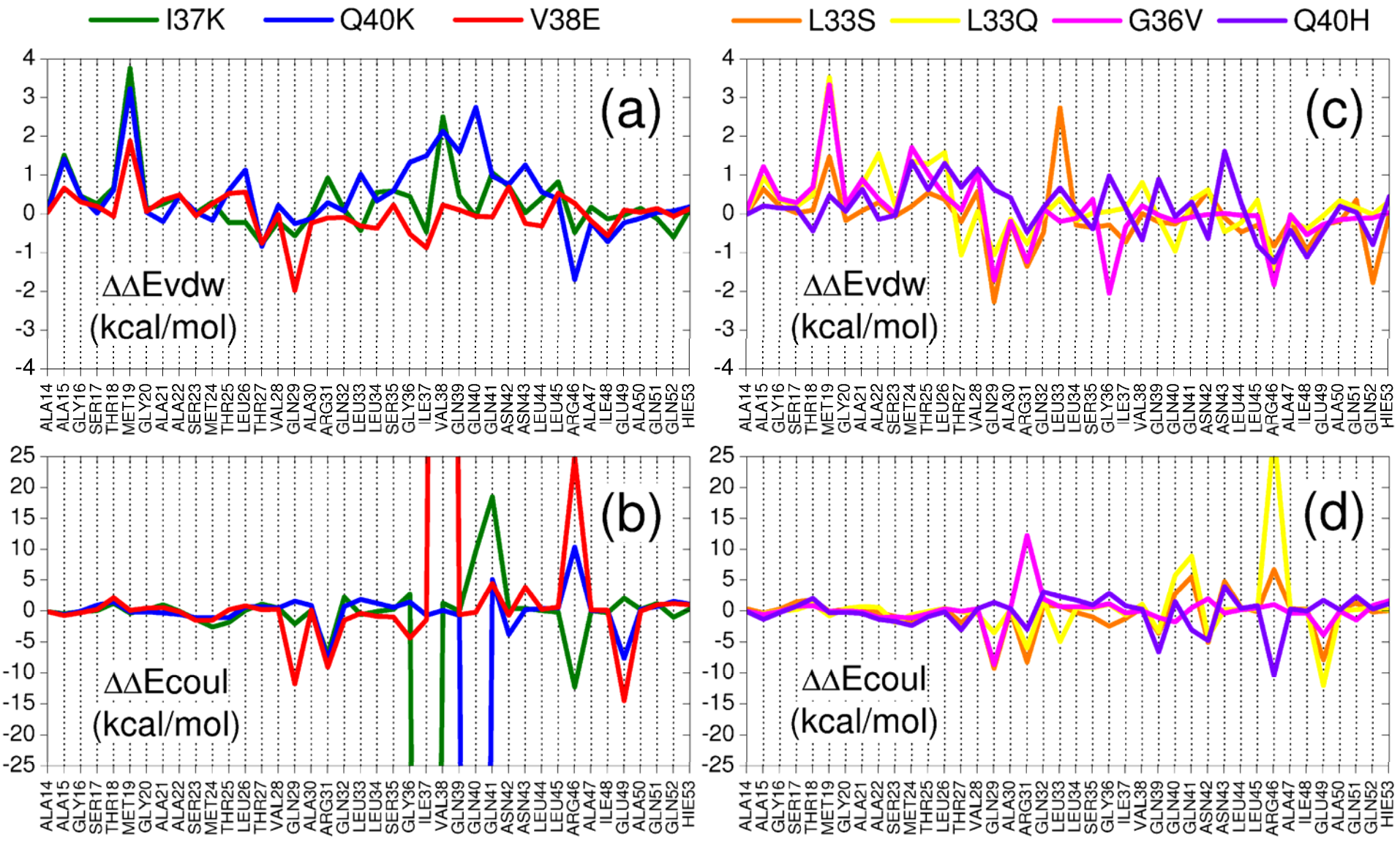




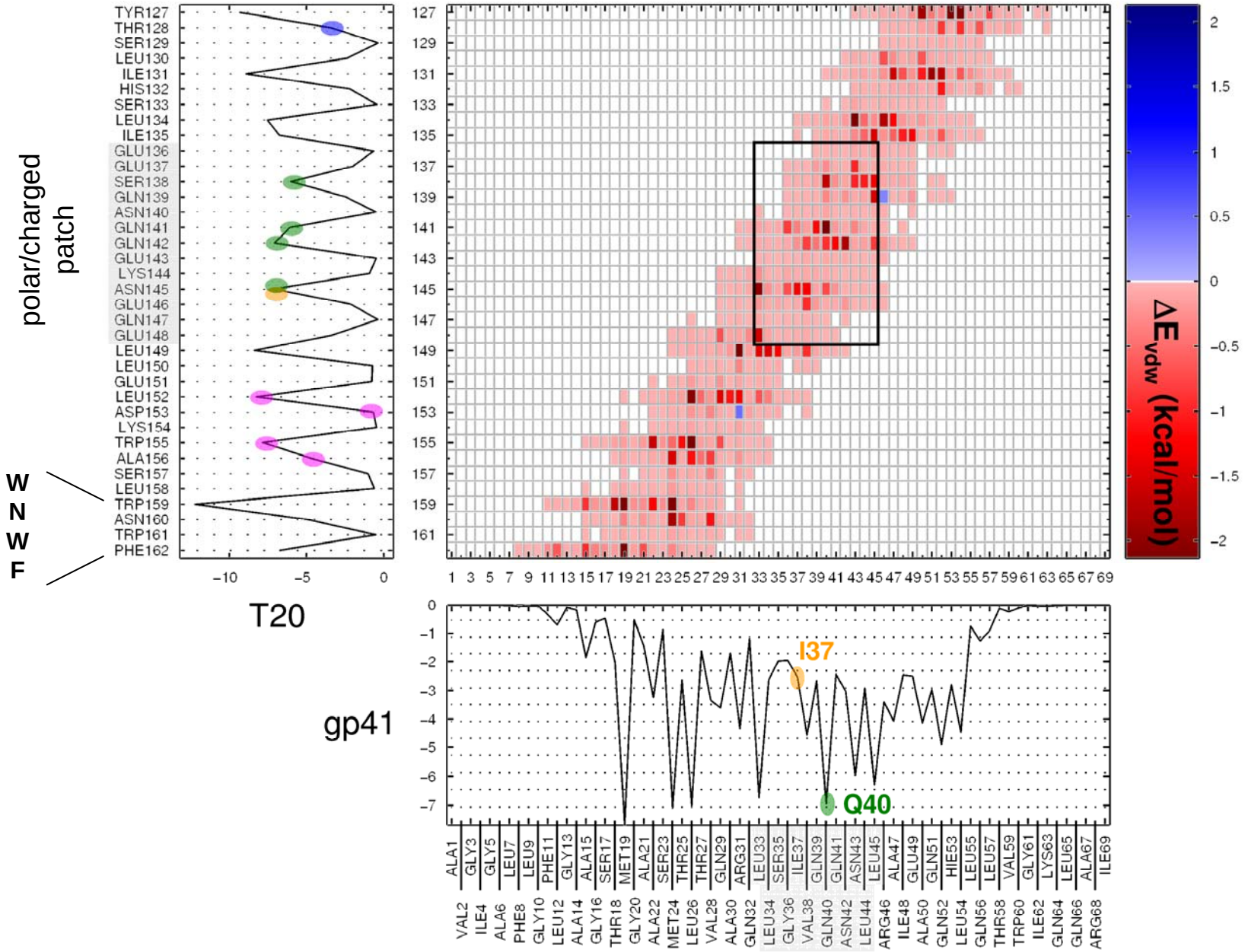
Correlation with Experiment Fold Resistance



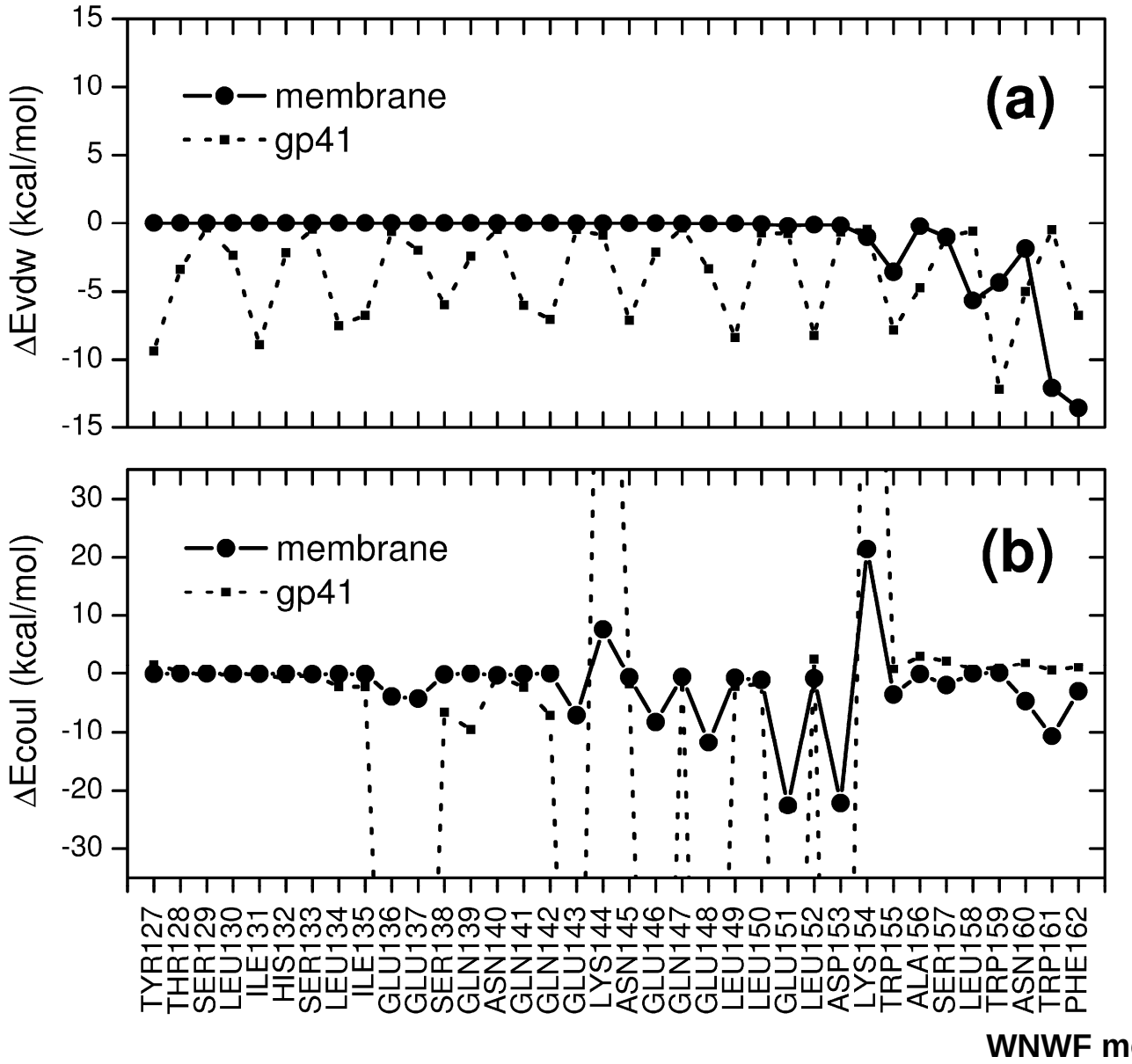
gp41-based footprint interactions (ΔΔE)



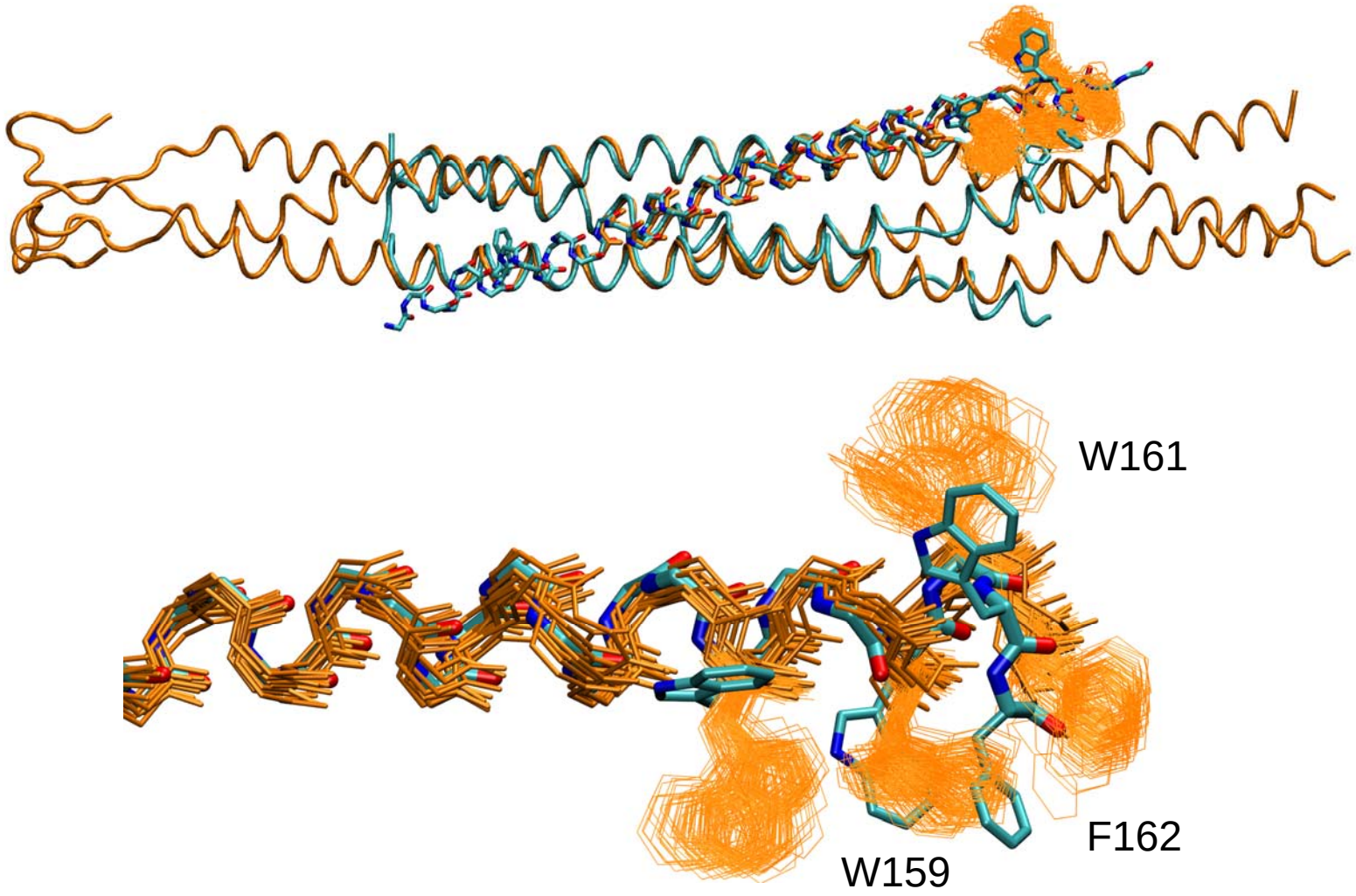
Steric packing footprint interaction matrix



T20-based footprints with membrane



Recent Crystal Structure Similar to T20-gp41 Model



orange: McGillick, B. E.; Balius, T.E.; Mukherjee, S.; Rizzo, R. C. *Biochemistry*, **2010**, 49, 3575–3592
blue: Buzon, V; et al. *PLoS Pathogens*, **2010**, 6, e1000880

Conclusions (T20-gp41 study)

General conclusions about the complex:

- Stable simulations (models include fusion peptide region embedded in membrane)
- Footprints yield qualitative agreement with clinically observed resistance sites
- Point mutation simulations yield good quantitative agreement (r2 = 0.7, N=6)

Analysis suggests:

- Mutations disrupt H-bonding and reduce favorable contact with gp41 at M19
- Charged mutations yield significant Coulombic changes that reduce favorable VDW
- Q40K > I37K resistance involves interaction differences in the initial (wildtype) state
- L33S vs L33Q resistance involves packing differences in the final (mutated) state.

- identified favorable interactions between the C-terminal end of T20 (WNWF motif), residues on gp41 (including the fusion peptide), and head groups in the adjacent membrane

Results suggest a complete T20 binding site would contribute to a stable complex, which could help to explain why prior studies, that employed truncated gp41 constructs, reported that C-terminal T20 residues may not interact with gp41.

Results suggest that modified T20 peptides designed to increase favorable contact with membrane could lead to enhanced activity.

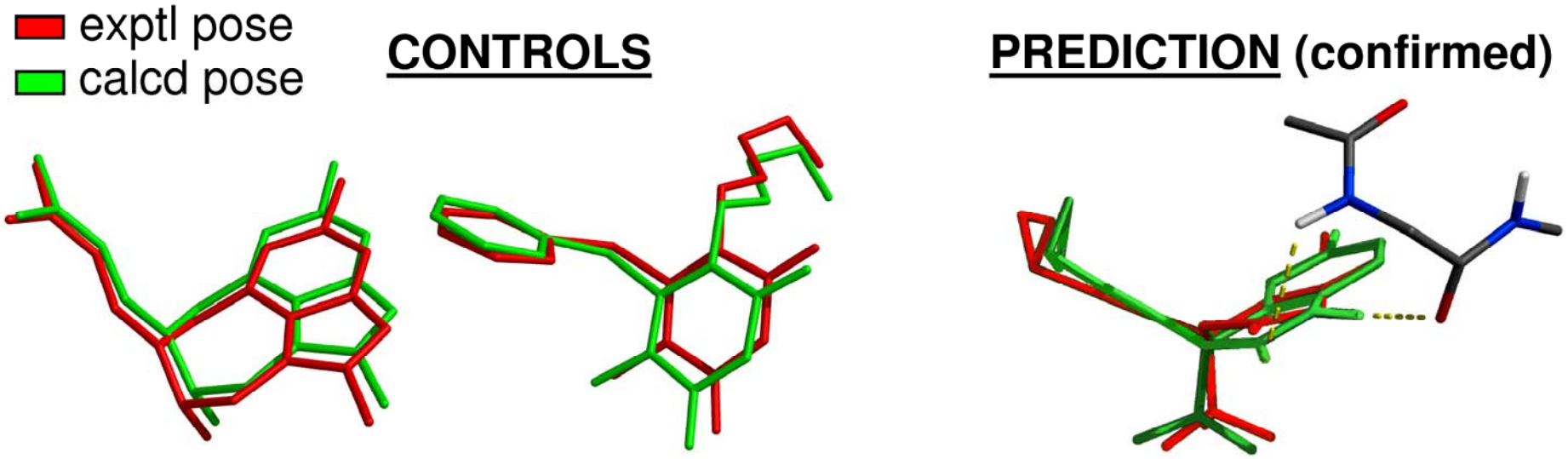
DOCK Development Projects

Developers (partial list)

- Tack Kuntz Group (UCSF)
- Brian Shoichet Group (UCSF), John Irwin (UCSF)
- David Case Group (Rutgers), Scott Brozell (Half Moon Bay)
- Demetri Moustakas (AstraZeneca), Terry Lang (Berkeley)
- Rizzo Group (Stony Brook)
- Latest version: DOCK 6.4 (2010)

Mukherjee, S.; Balius, T.E.; Rizzo, R. C. *Journal of Chemical Information and Modeling*, **2010**, *50***,** 1986-2000.
Balius, T.E.; Mukherjee, S.; Rizzo, R. C. *Journal of Computational Chemistry,* **2011,** *in press*

Importance of controls (protocol validation)



Rizzo, R. C.; Wang, D.-P.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.*, **2000**, *122*, 12898-12900
Ren J.; Milton J.; Weaver K. L.; Short S. A.; Stuart D. I.; Stammers D. K. *Structure,* **2000,** *8,* 1089-1094

Specific goals: Stony Brook project

- Testset construction (termed SB2010 testset)

- Three core DOCK experiments:
  RGD = Rigid Ligand Docking (tests orienting code)
  FAD = Fixed Anchor Docking (tests flexible ligand growth)
  FLX = Flexible Ligand Docking (tests both)

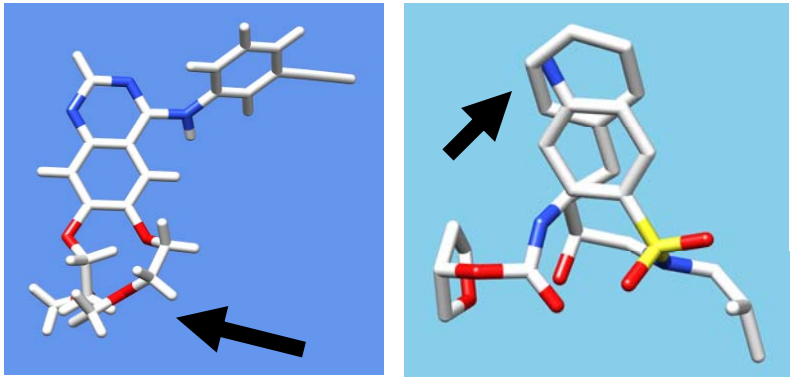
- Primary analysis :
  success rate (pose reproduction to within 2.0 Å of x-ray ref)
  sampling failures (pose not generated)
  scoring failures (pose generated but not selected)

  Note: success + sampling failures + scoring failures = 100%

FAD, and FLX protocols: ligand internal clashes

Most problematic for flexible molecules

lack of torsional energy term



Repulsive vdw term (intra-ligand)

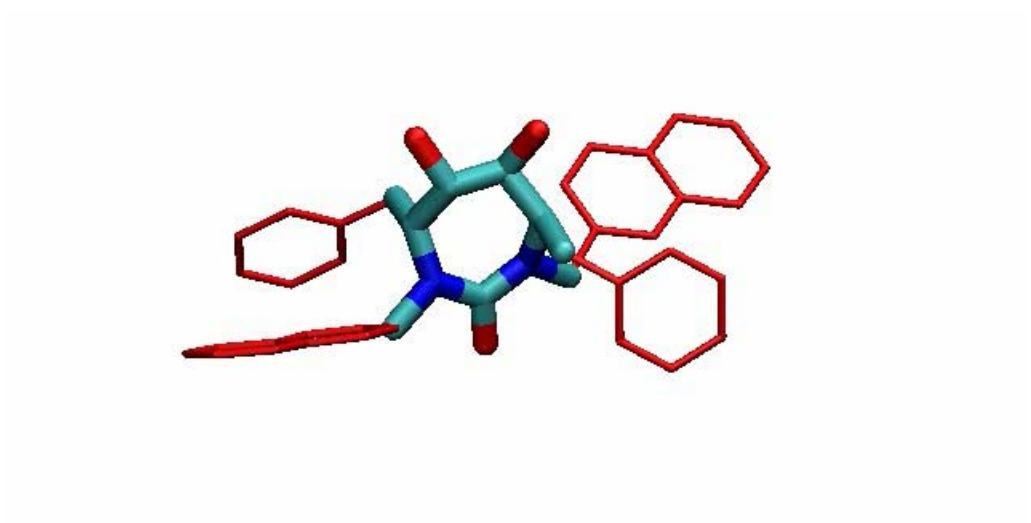
interaction pair list for speedup

Input parameter "use_internal_energy" [yes no] anchor & grow, rigid docking, (minimizations)

Optimized (repaired) active atom flag logic segment-based energy evaluations and multiple anchor orienting array

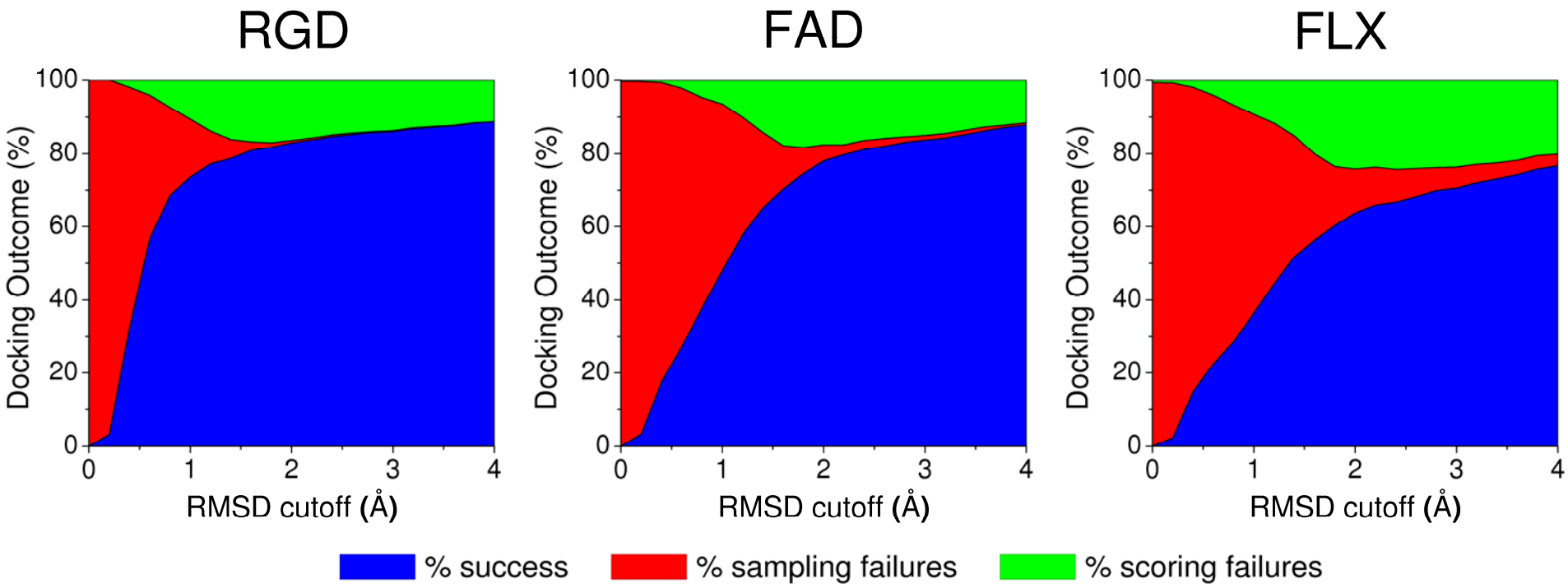
Implementation of growth trees (FAD and FLX protocols)

Growth tree examples: HIVPR



Multiple anchor option enabled

Pose reproduction statistics: global rmsd spectrum



At any RMSD: total success + sampling failures + scoring failures = 100%

N=780 systems

Pose reproduction statistics: by flexibility

Table 2. Success Rates for Rigid (RGD), Fixed-Anchor (FAD, and Flexible (FLX) Ligand Docking.

Docking Method	No. of RB	Subset Size	Total success[a]		Sampling failures[a]		Scoring failures[a]		Avg time
			(No.)	(%)	(No.)	(%)	(No.)	(%)	(min/mol)
A	B	C	D	E	F	G	H	I	J
RGD	7 or less	423	352	83.2	3	0.7	68	16.1	2.0
RGD	8 to 15	268	225	83.9	1	0.4	42	15.7	5.8
RGD	15 plus	89	70	78.7	1	1.1	18	20.2	13.6
RGD	all	780	647	82.9	5	0.6	128	16.4	4.6
FAD	7 or less	423	370	87.5	4	0.9	49	11.6	0.9
FAD	8 to 15	268	192	71.6	14	5.2	62	23.1	11.2
FAD	15 plus	89	47	52.8	16	18.0	26	29.2	67.1
FAD	all	780	609	78.1	34	4.4	137	17.6	11.9
FLX	7 or less	423	315	74.5	18	4.3	90	21.3	3.5
FLX	8 to 15	268	148	55.2	41	15.3	79	29.5	17.3
FLX	15 plus	89	35	39.3	35	39.3	19	21.3	74.4
FLX	all	780	498	63.8	94	12.1	188	24.1	16.6

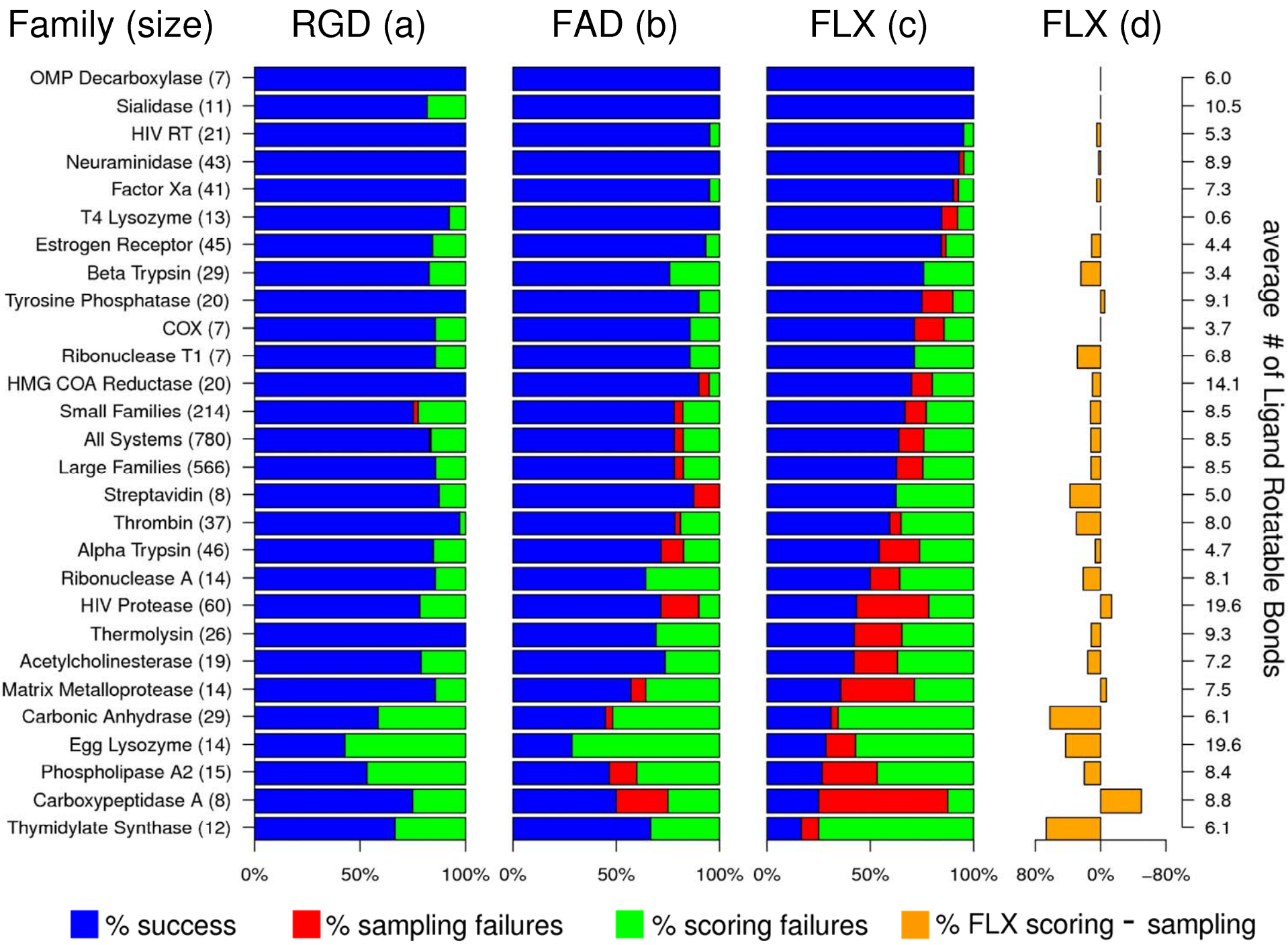
[a]Computed using a 2.0 Å rmsd definition of success using experimentally observed ligand poses as reference.

Total success + sampling failures + scoring failures = 100% (N=780 systems)

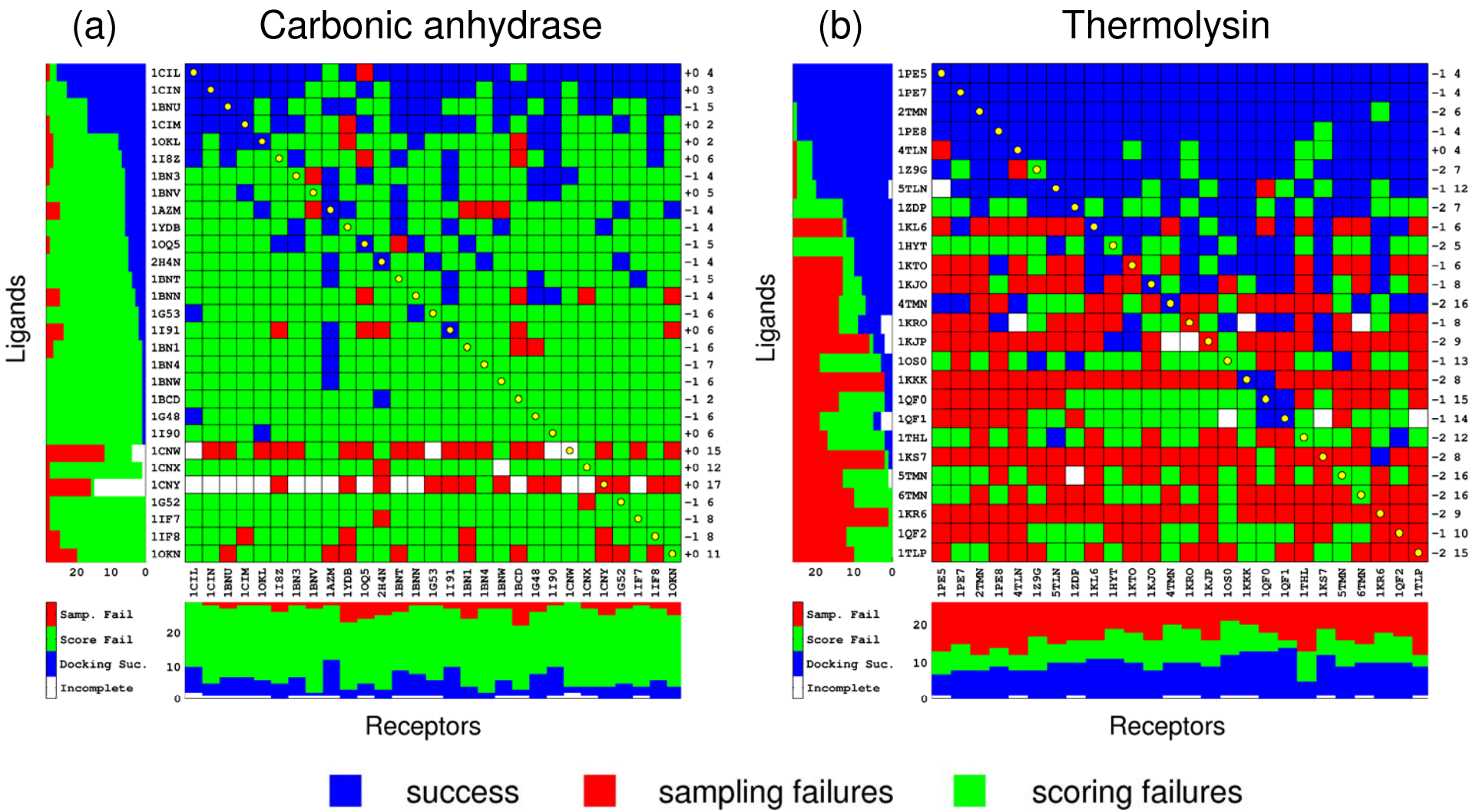
On-the-fly flexible growth (FLX) shows improvment

DOCK Version	#RB Subset	Succes (%)
DOCK 4	7orless	60.9%
DOCK 6.3	7orless	74.2%
Stony Brook (6.4)	7orless	74.5%
DOCK 4	8to15	29.8%
DOCK 6.3	8to15	41.0%
Stony Brook (6.4)	8to15	55.2%
DOCK 4	15+	4.4%
DOCK 6.3	15+	7.8%
Stony Brook (6.4)	15+	39.3%
DOCK 4	all	43.8%
DOCK 6.3	all	55.1%
Stony Brook (6.4)	all	63.8%

Pdb code	Drug/Inhibitor	#RB	Indication
8KME	SEL2770	29	anticoagulant
1AGM	acarbose	22	type 2 diabetes
1G7F	PNU177496	19	diabetes
1CNY	benzene sulfonamide	17	glaucoma
4TMN	peptidomimetic	16	antibacterial
1SDT	indinavir	16	antiviral HIV/AIDS
1HWK	atorvastatin	15	heart disease
1KPM	vitamin E	13	vitamin supplement
1H22	huperazine analog	13	alzheimers
1HFC	hydroxamate inhibitor	13	anticancer
1NJJ	geneticin	12	antibiotic
1A4G	zanamivir	10	antiviral influenza

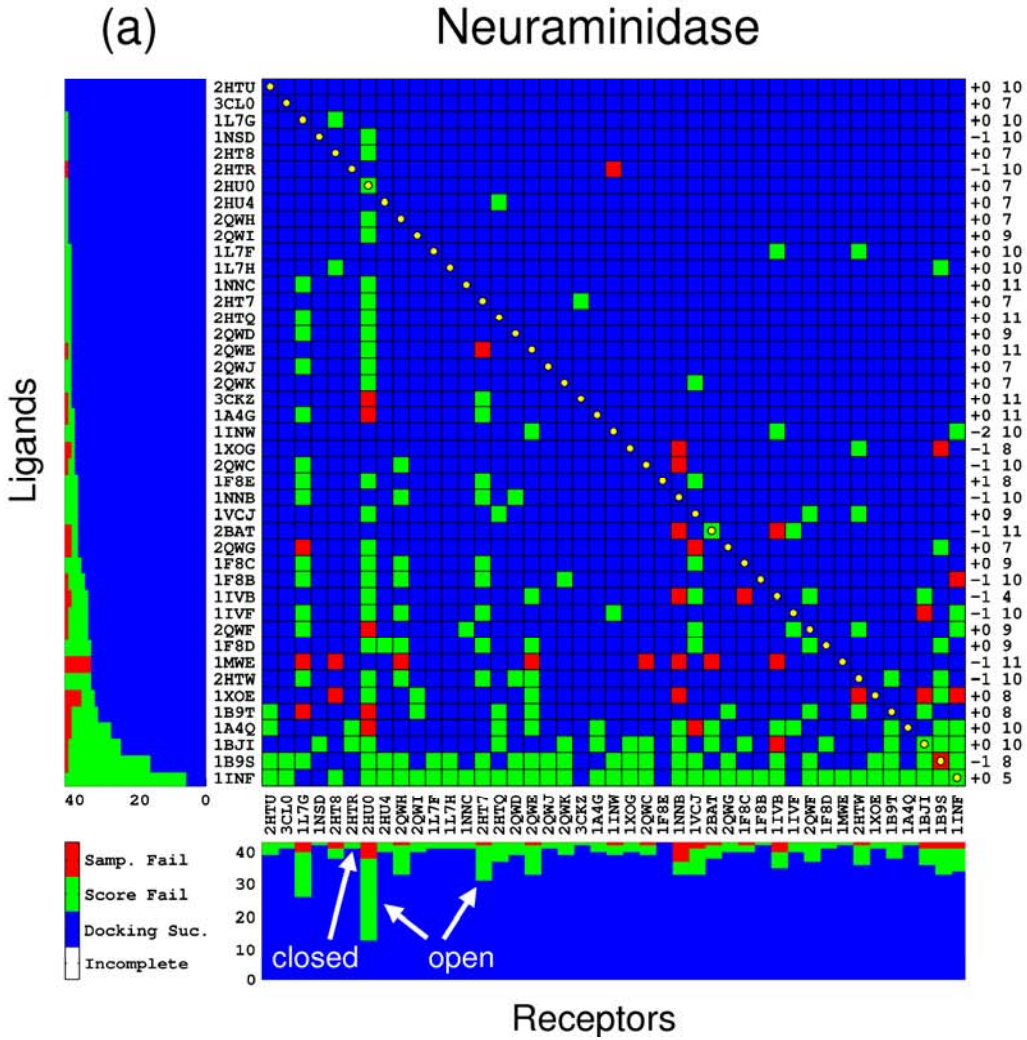

Family (size)
RGD (a)
FAD (b)
FLX (c)
FLX (d)
OMP Decarboxylase (7)
Sialidase (11)
HIV RT (21)
Neuraminidase (43)
Factor Xa (41)
T4 Lysozyme (13)
Estrogen Receptor (45)
Beta Trypsin (29)
Tyrosine Phosphatase (20)
COX (7)
Ribonuclease T1 (7)
HMG COA Reductase (20)
Small Families (214)
All Systems (780)
Large Families (566)
Streptavidin (8)
Thrombin (37)
Alpha Trypsin (46)
Ribonuclease A (14)
HIV Protease (60)
Thermolysin (26)
Acetylcholinesterase (19)
Matrix Metalloprotease (14)
Carbonic Anhydrase (29)
Egg Lysozyme (14)
Phospholipase A2 (15)
Carboxypeptidase A (8)
Thymidylate Synthase (12)
6.0
10.5
5.3
8.9
7.3
0.6
4.4
3.4
9.1
3.7
6.8
14.1
8.5
8.5
8.5
5.0
8.0
4.7
8.1
19.6
9.3
7.2
7.5
6.1
19.6
8.4
8.8
6.1
average # of Ligand Rotatable Bonds
0%
50%
100%
0%
50%
100%
0%
50%
100%
80%
0%
−80%
% success
% sampling failures
% scoring failures
% FLX scoring - sampling

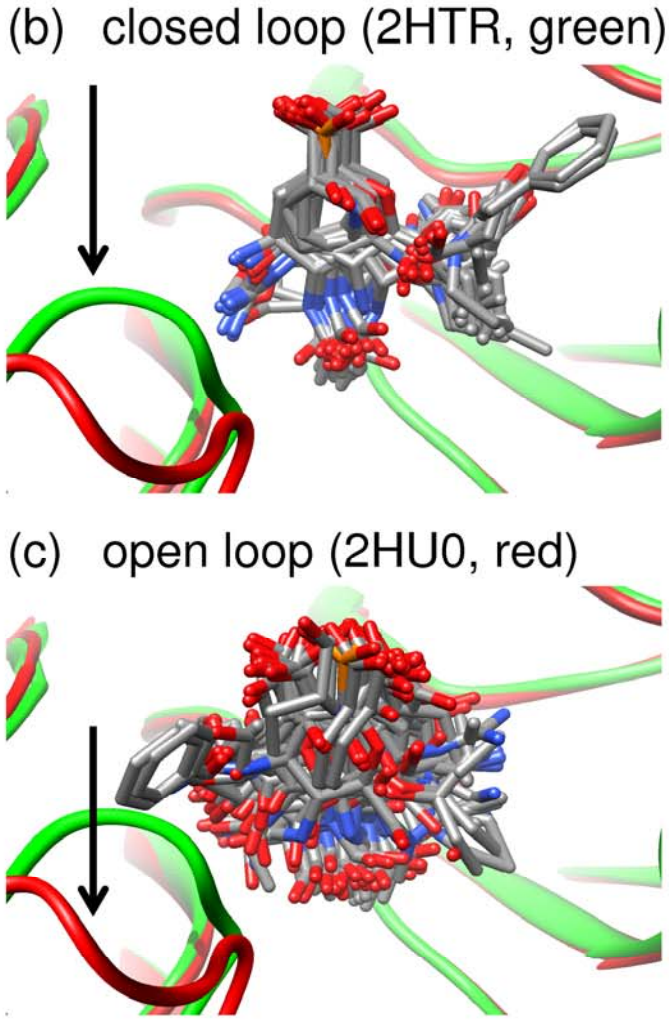
Family-based crossdocking heat maps



diagonal elements represent cognate protein-ligand pairs

Neuraminidase studies (43 x 43 matrix)





See Birch, L.; et al *J. Comput.-Aided Mol. Des.* **2002**, *16*, 855-869 for a related study

Recent ACS symposium (pose identification)

As provided (mods by Roe and Brozell)

ASTEX (N=85)						
RMSD	**Mean**	**Std Dev**	**Median**	**Min**	**Max**	**Success%**
Top Score	2.61	2.64	1.23	0.18	10.25	68.2
Least (32)	1.31	0.99	1.15	0.18	6.45	89.4

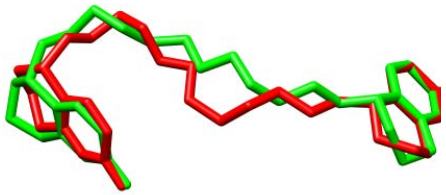
Stony Brook Prep

ASTEX84			
Bias	Finished	Success	SampFail
lig.10ps.trans	(84/84)=100.0%	(53/84)=63.0%	(6/84)=7.1%
lig.min.trans	(84/84)=100.0%	(59/84)=70.2%	(7/84)=8.3%
lig.min	(84/84)=100.0%	(54/84)=64.2%	(8/84)=9.5%

Progressive removal of bias in ligand starting conditions

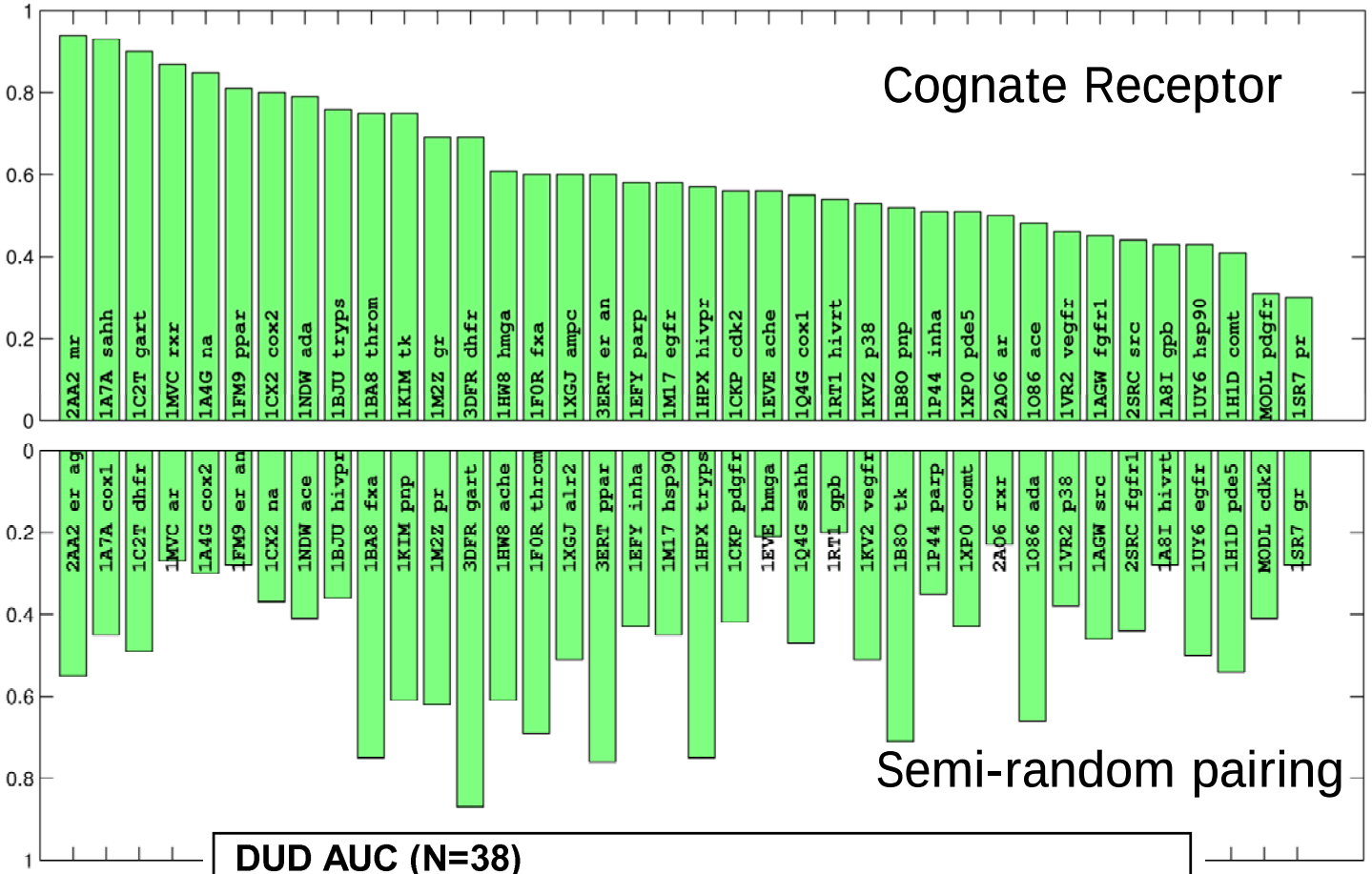
Initial studies: symmetry corrected rmsds

System	1HWR
Std RMSD	7.28
Hungarian	0.37
Min RMSD	0.09



System	1H22
Std RMSD	12.32
Hungarian	0.83
Min RMSD	0.13

DUD enrichment AUC



DUD AUC (N=38)			
Cognate		**Null Hypothesis**	
Median AUC	0.575	Median AUC	0.45
Mean AUC	0.61	Mean AUC	0.47
Std Dev	0.17	Std Dev	0.17
Max AUC	0.94	Max AUC	0.87
Min AUC	0.3	Min AUC	0.2

Footprint similarity (FPS) score examples

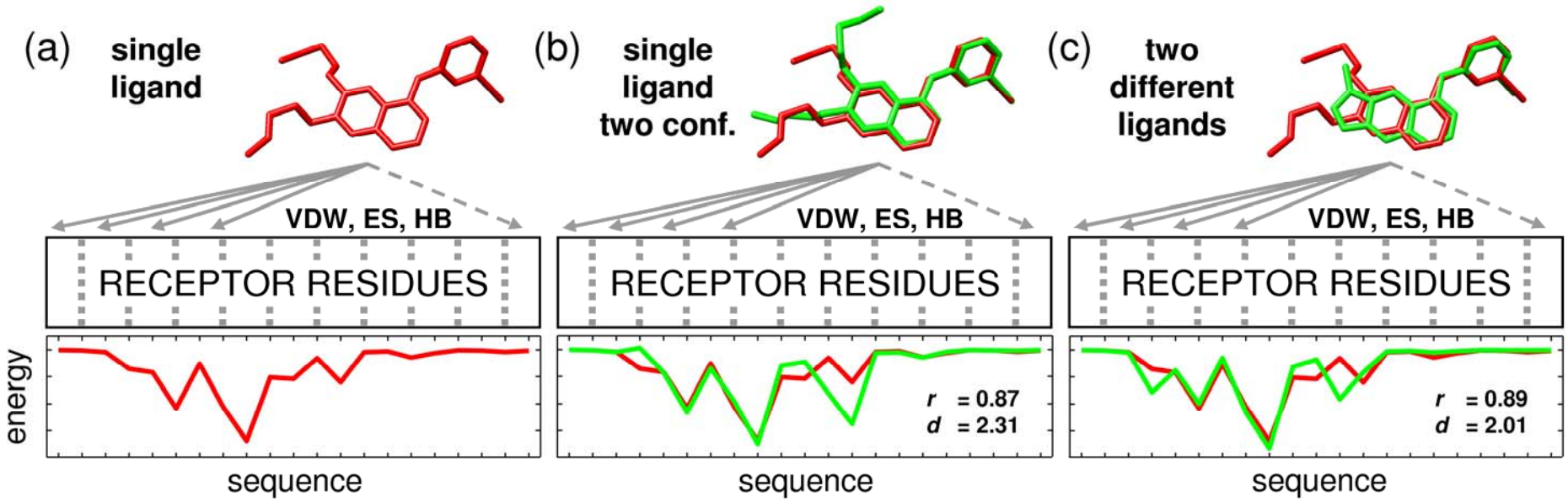


Table I. Examples of possible reference types to derive molecular footprints.

Reference Types	Description
Known inhibitor	FDA-approved drug or experimental inhibitor validated to bind
Natural substrate	Native peptide or cofactor
Transition state	Predicted transition state geometry for a chemical reaction
Modified structure	Key functionality/substructure (side-chain mediating protein-protein interactions)
Text file footprint	Modified entries to increase/decrease importance of select residues (resistance mutations)
Ensemble-weighted	Averaged footprints derived from MD/MC simulations

Similar work using binary footprints:

Deng, Z. *et al.*, *J Med Chem* **2004**, *47*, 337-44

Brewerton, S. C., *Curr Opin Drug Discov Devel* **2008**, *11*, 356-64 (review article)

FPS evaluations: pose identification SB2010

Table III. Pose identification success using Footprint similarity (FPS) vs DOCK Cartesian energy (DCE) methods to rescore rigid (RGD), fixed anchor (FAD) and flexible ligand (FLX) pose ensembles.

Row	Ligand Ensemble	FPS Standard Pearson A	FPS Standard Euclidean B	FPS Threshold Pearson C	FPS Normalized Euclidean D	DCE E
			VDW+ES			
1	RGD	691 (89.2%)[a]	718 (92.6%)	683 (88.1%)	707 (91.2%)	627 (80.9%)
2	FAD	642 (85.8%)	638 (85.3%)	644 (86.1%)	652 (87.2%)	606 (81.0%)
3	FLX	563 (82.8%)	565 (83.1%)	556 (81.8%)	574 (84.4%)	489 (71.9%)
			VDW			
4	RGD	687 (88.6%)	684 (88.3%)	662 (85.4%)	687 (88.6%)	445 (57.4%)
5	FAD	638 (85.3%)	630 (84.2%)	621 (83.0%)	638 (85.3%)	464 (62.0%)
6	FLX	545 (80.1%)	539 (79.3%)	525 (77.2%)	545 (80.1%)	309 (45.4%)
			ES			
7	RGD	579 (74.7%)	583 (75.2%)	576 (74.3%)	579 (74.7%)	398 (51.4%)
8	FAD	601 (80.3%)	573 (76.6%)	598 (79.9%)	603 (80.6%)	460 (61.5%)
9	FLX	521 (76.6%)	505 (74.3%)	513 (75.4%)	522 (76.8%)	314 (46.2%)
			VDW+ES+HB			
10	RGD	670 (86.5%)	726 (93.7%)	590 (76.1%)	685 (88.4%)	633 (81.7%)
11	FAD	621 (83.0%)	643 (86.0%)	590 (78.9%)	632 (84.5%)	606 (81.0%)
12	FLX	557 (81.9%)	564 (82.9%)	501 (73.7%)	561 (82.5%)	492 (72.4%)

[a] Number of molecules in which the pose identified was ≤ 2 Å from the x-tal structure pose followed by success rates in parenthesis. Pose ensembles (RGD = 775, FAD = 748, FLX= 680) derived from docking runs reported by Mukherjee et al.[7]

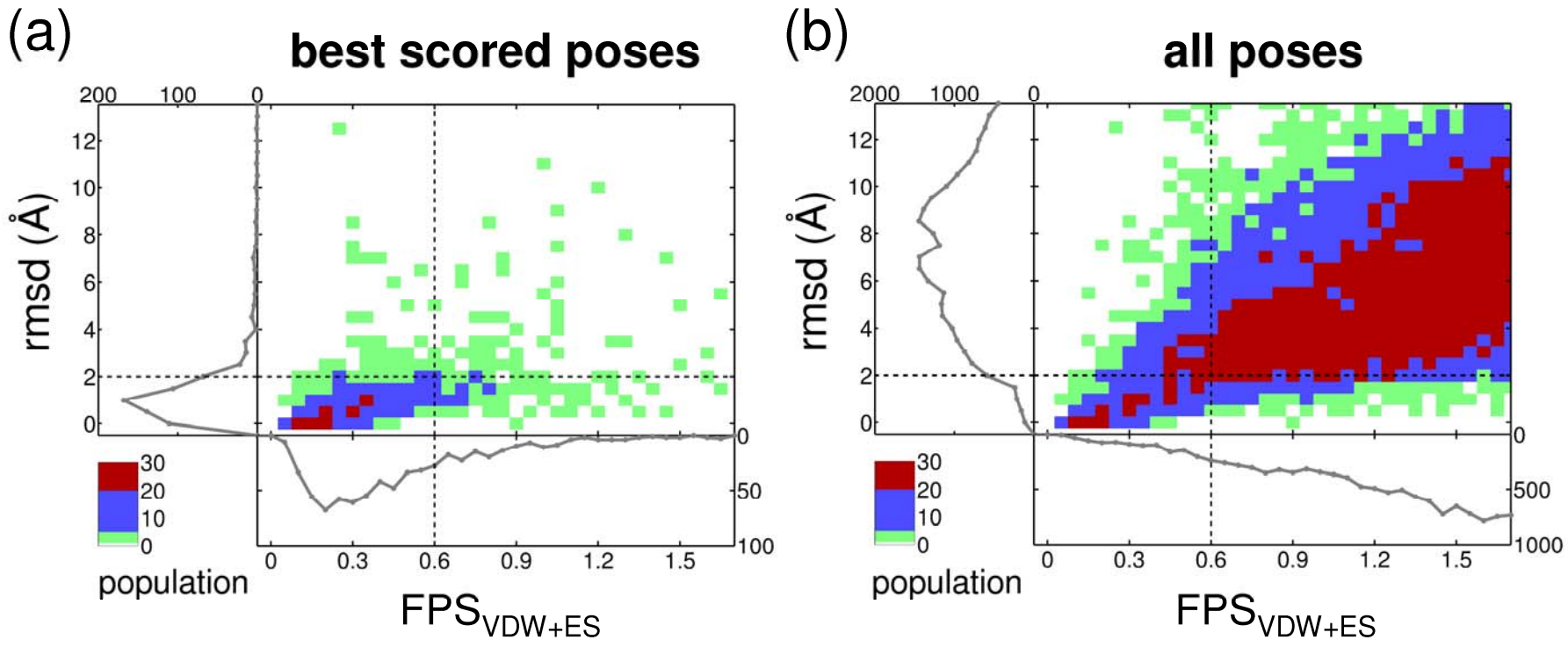


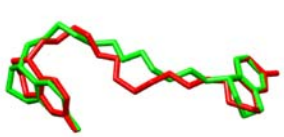
Table IV. FLX results scored with FPS_{VDW+ES} for three differing footprint threshold cutoffs using a 2 Å rmsd to separate positive from negative regions.

Set	Cutoff	Positive	Negative	Predicted Positive	Predicted Negative	True Positive	False Positive	True Negative	False Negative
best scored[a]	0.3	574	106	251	429	240	11	95	334
	0.6			507	173	458	49	57	116
	0.9			618	62	537	81	25	37
all poses[b]	0.3	965	25,865	295	26,535	261	34	25,831	704
	0.6			1,185	25,645	577	608	25,257	388
	0.9			3,026	23,804	759	2267	23,598	206

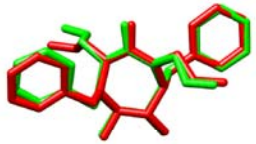
[a]N = 680. [b]N = 26,830.

False positives (rmsd > 5 Å; FPS < 0.6)

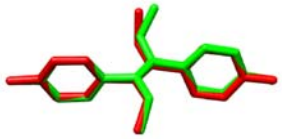
1H22
12.32, 0.27
symmetry



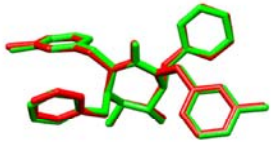
1HWR
7.28, 0.30
symmetry



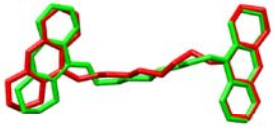
3ERD
7.01, 0.31
symmetry



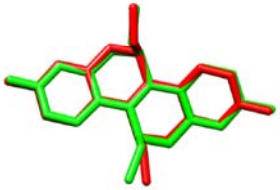
1MER
8.56, 0.32
symmetry



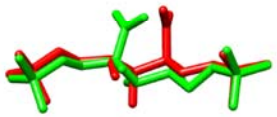
2CKM
13.08, 0.32
symmetry



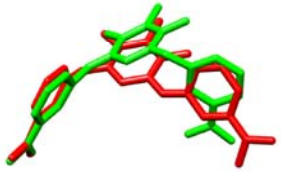
1L2J
6.85, 0.37
symmetry



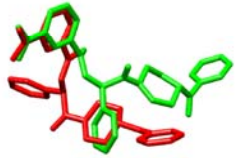
1RBO
6.79, 0.40
partial sym.



1QB6
7.48, 0.42
partial sym.



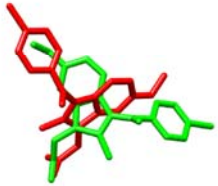
1EB2
6.01, 0.46
non-specific



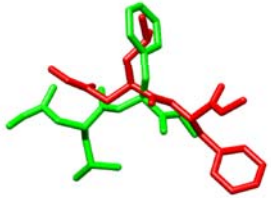
1J07
5.15, 0.46
partial sym.

1ROB
5.03, 0.49
non-specific

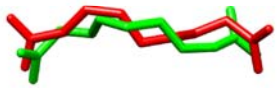
3H1X
6.66, 0.49
non-specific



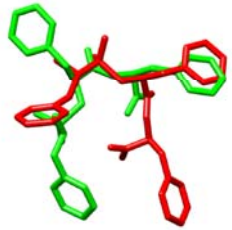
1HFC
6.35, 0.51
non-specific



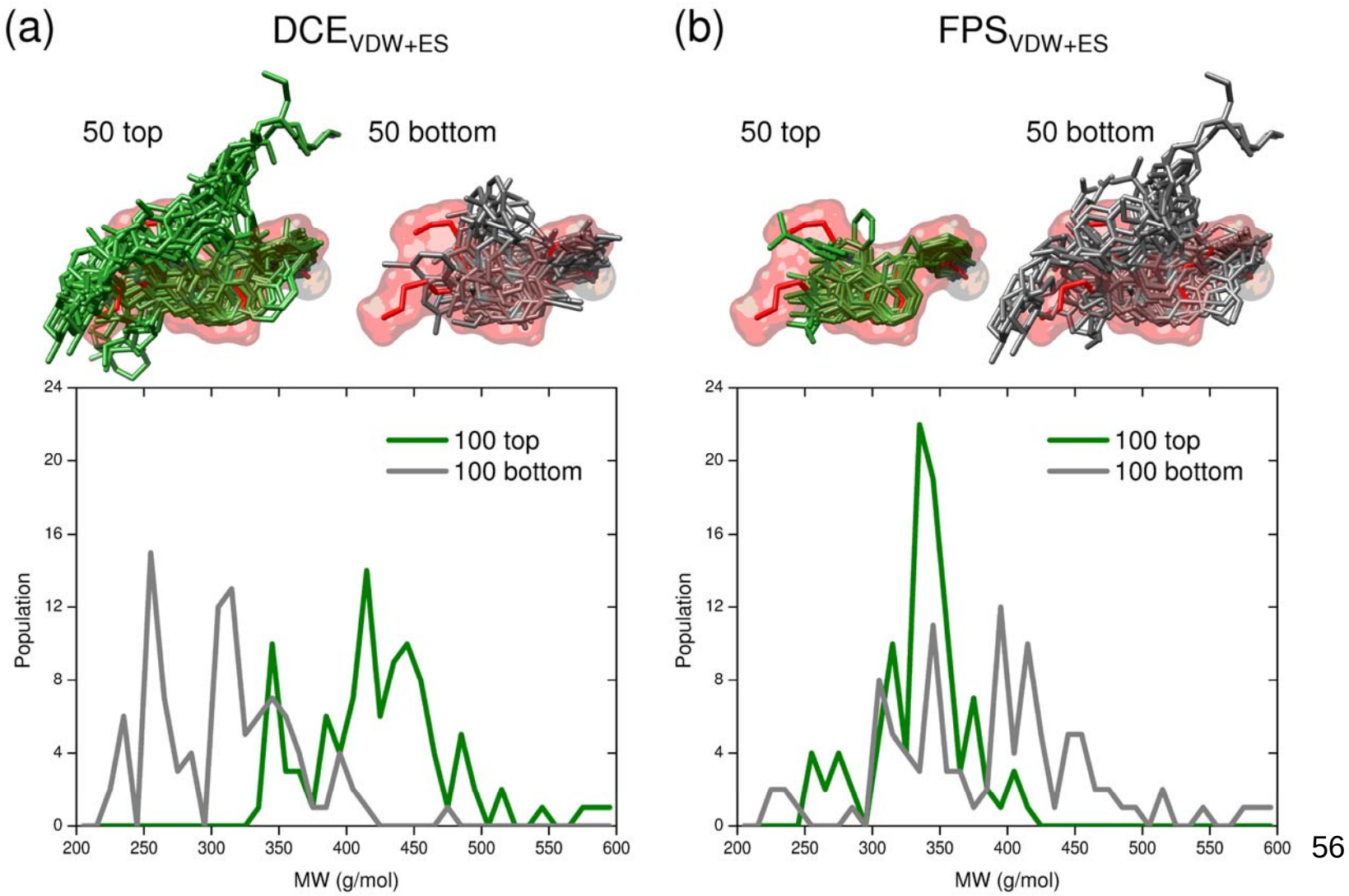
1TUF
7.43, 0.53
symmetry



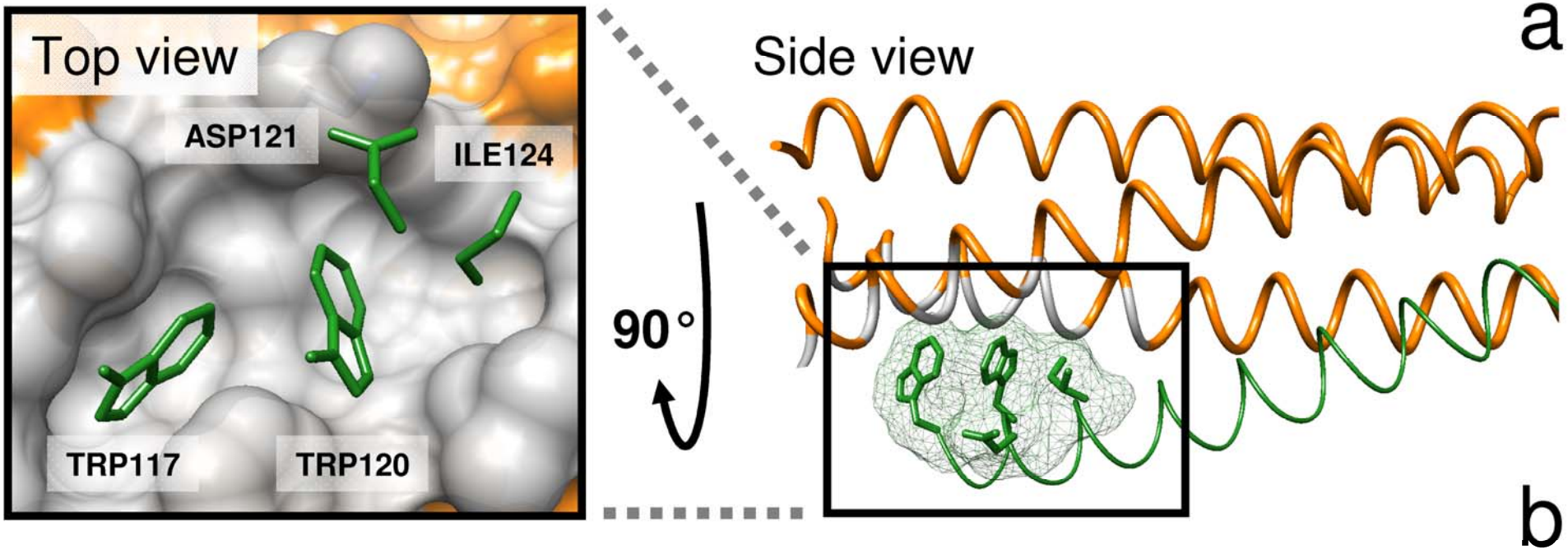
1OS0
6.69, 0.56
non-specific

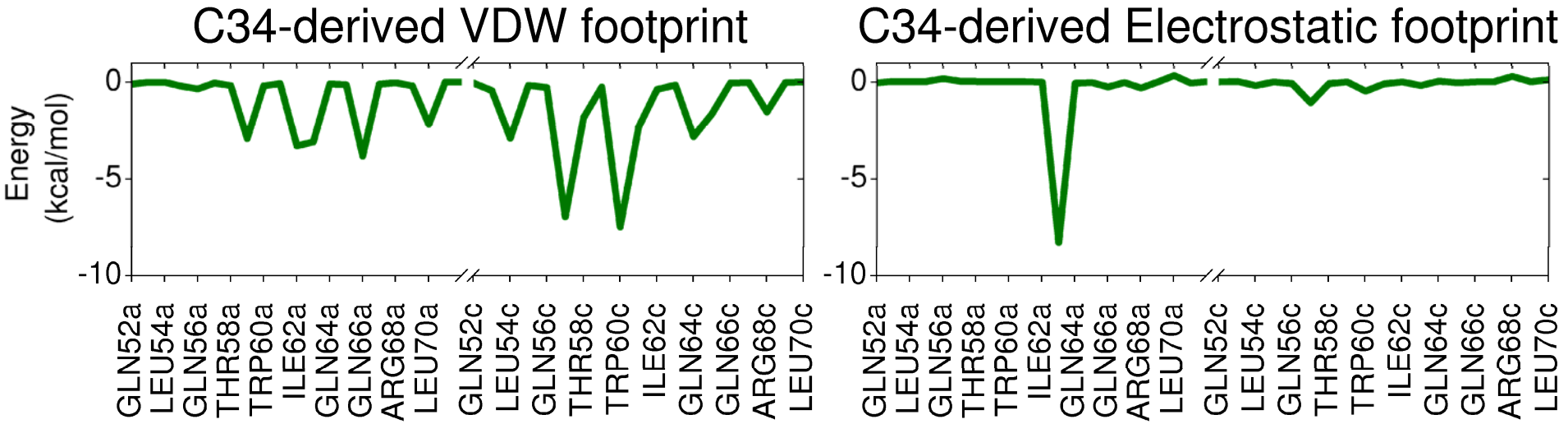


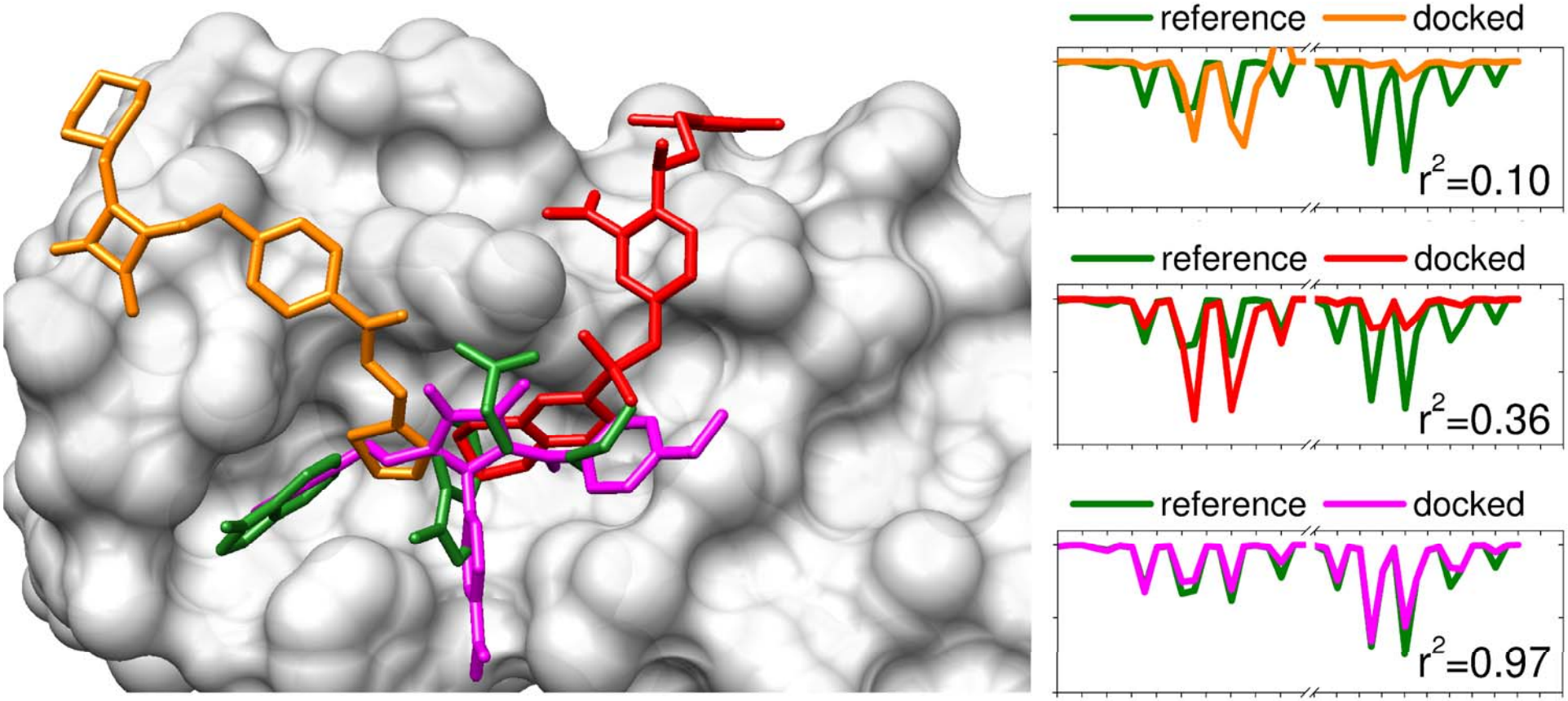
Enrichment studies: EGFR (different MW bias for top scoring compounds)



HIVgp41 studies






reference
docked
r²=0.10
reference
docked
r²=0.36
reference
docked
r²=0.97

Conclusions / summary

Test set development:

- characterized RGD, FAD, FLX protocols for pose reproduction
- family-based analysis
- cross docking experiments
- vs protocol developed for BlueGene

Code-base enhancements:

- internal ligand geometry issues repaired
- growth trees (pruning statistics)
- footprint-based continuous scoring
- available as DOCK6.4

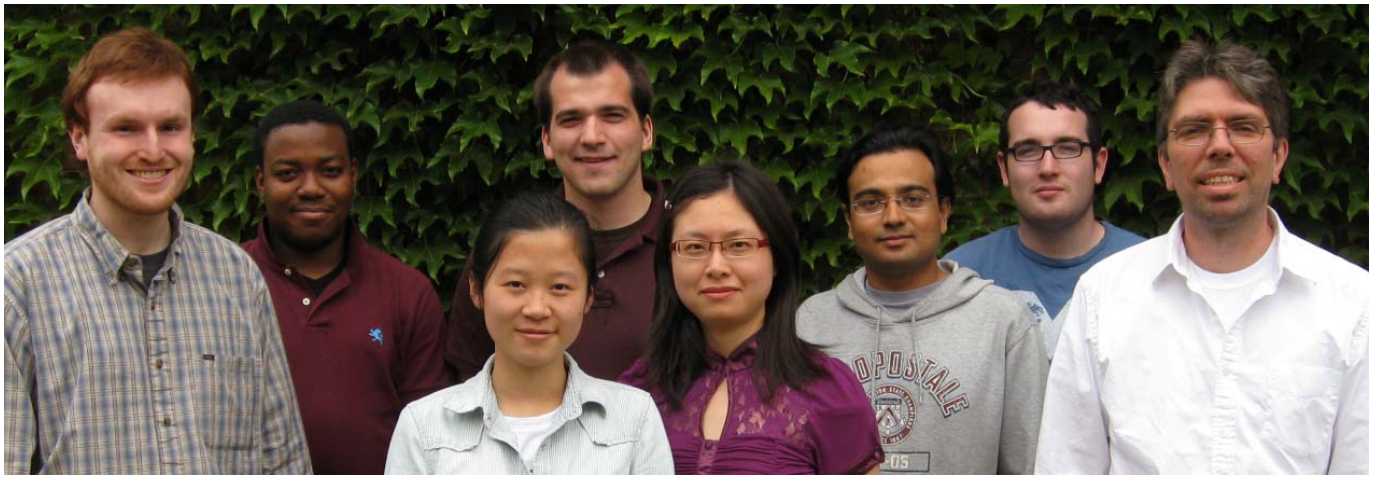
Footprints:

- identify ligands and/or poses that look like a reference query
- pose reproduction: improvement in success vs standard score
- initial enrichments studies promising (DUD)
- ongoing project example (gp41)

Acknowledgments

Current Students:

Trent Balius
Patrick Holden
Yulin Huang
Lingling Jiang
Sudipto Mukherjee
Jibril Ashiru-Balogun
Brian Shea



Collaborators:

HIV: Miriam Gochin (Touro/UCSF)
HIV: Mark Wilson (U of Nebraska)
HIV: Iwao Ojima ICB&DD (SBU)
Flu: Janet Hearing (SBU)
Kinases: Todd Miller (SBU)
DOCK: Tack Kuntz (UCSF), Brian Shoichet (UCSF), Dave Case (Scripps), Carlos Sosa (IBM)

Support:

Applied Math & Stat (SBU)
Computational Science Center at Brookhaven National Laboratory (NewYork Blue)
NYSTAR Investigator Award
School of Medicine (SBU) Influenza TRO, Carol M. Baldwin Breast Cancer Research Award
National Institutes of Health R01GM083669 (NIGMS), F31CA134201 (NCI, to TEB)