First-principle-based and data-driven design platform for small molecule drugs and antimicrobial peptides

Acknowledgement: Profs Yunching Becky Chen, Vincent Shu, Chin Yu, Qing Zhang, Praexisio Inc. and students/postdoc researchers in my lab

Presenter:
Lee-Wei Yang
（楊立威）

Professor and Director,
Inst. of Bioinformatics and Structural Biology, **NTHU, Taiwan** ;
Director of **PhD program in Biomedical Artificial Intelligence**, NTHU
Group coordinator, Complex Systems and Mathematical Biology,
National Center for Theoretical Sciences (Physics Division);
TIGP Bioinformatics & CBMB PhD program, Academia Sinica;
Honorary Professor, University of Liverpool, UK
Visiting Professor, University of Osaka (2018),
Co-founder of Praexisio Inc. (Drug discovery with harnessing the dancing targets)

NCTS Winter School
02/2202

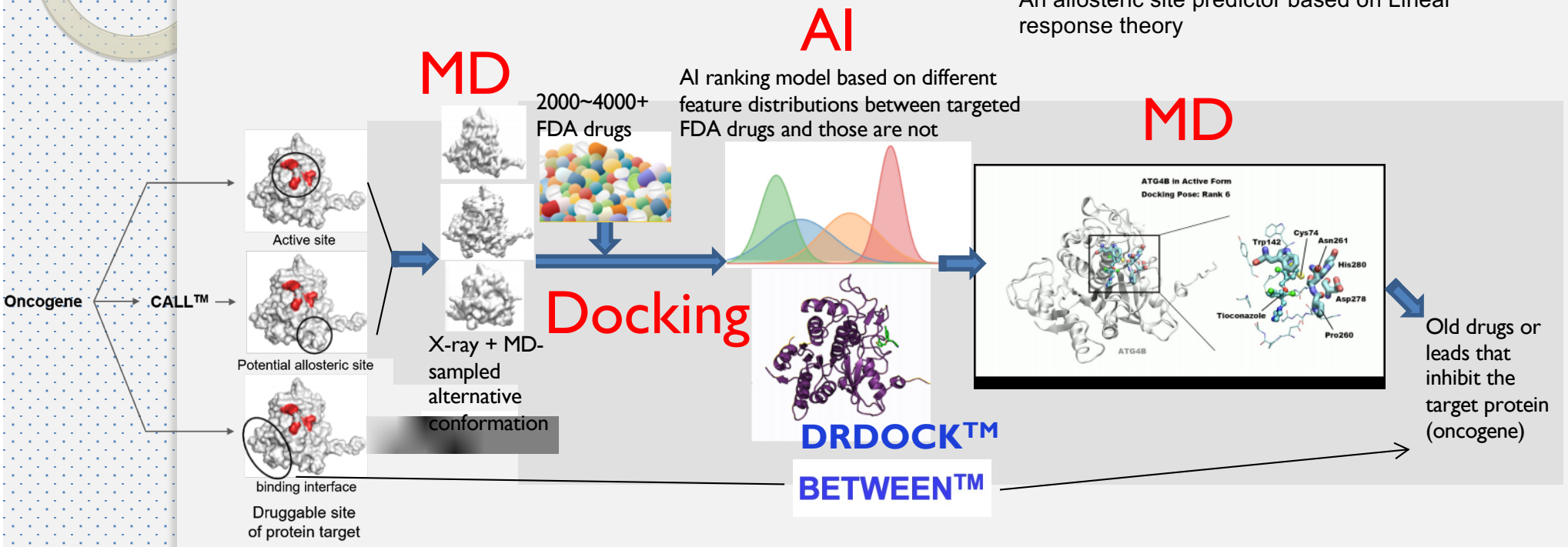
Talk overview

- Small drug design - state of the art for MD-aided drug screening and new drug discovery
- Find PPI blockers, re-define specificity
- Therapeutic peptides
- Other methods to explore protein dynamics – Normal Mode Analysis and Linear Response Theories

Current pipeline in Yang's lab to screen drugs with protein dynamics accentuated

CALL™ (LWY's Yang et al. 2014; Li et al., 2017; Huang et al.,2019):

- An allosteric site predictor based on Linear response theory



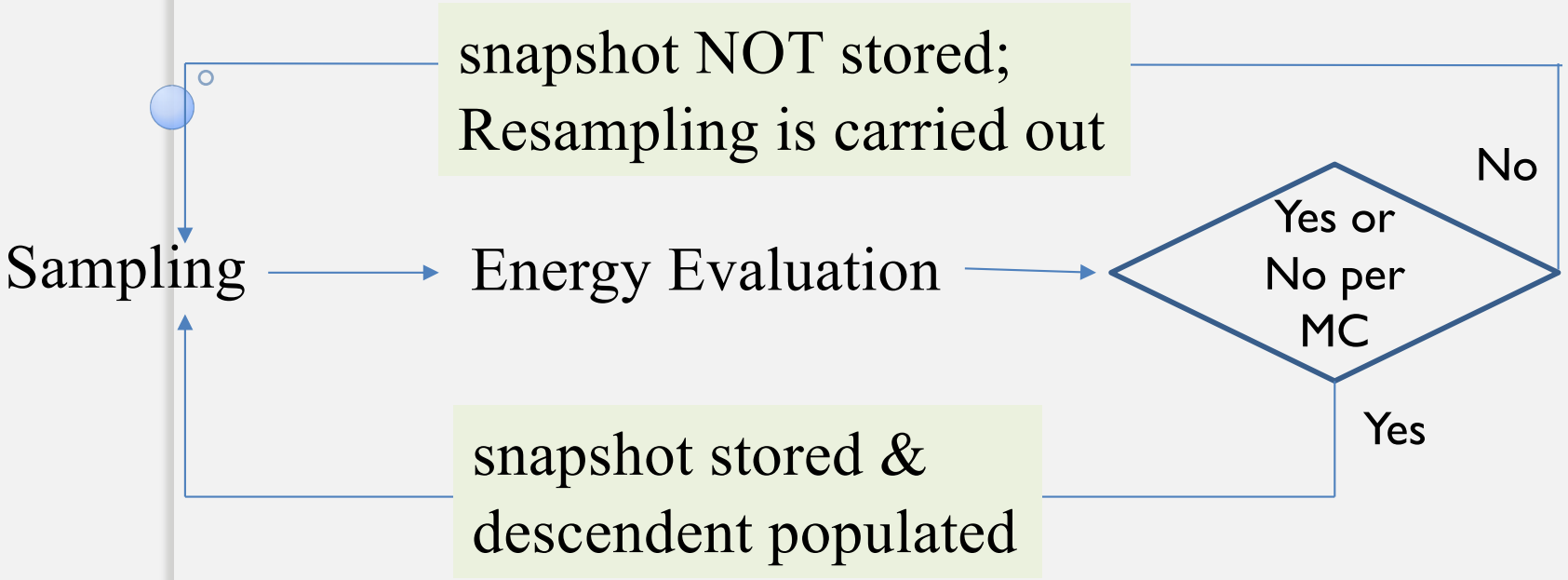
BETWEEN™ (LWY's Lin.,2020):

- A drug screening that integrates docking, entropy-weighted contact, MD-refined bind free energy
- For screening the binding interface

DRDOCK™ (LWY's Tsai et al.2021):

- A drug screening that integrates docking, AI-based ranking score and MD-refined bind free energy
- For the druggable site that has a cavity

SMALL MOLECULE DOCKING



Journal of COMPUTATIONAL CHEMISTRY

Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function

GARRETT M. MORRIS,[1] DAVID S. GOODSELL,[1]
ROBERT S. HALLIDAY,[2] RUTH HUEY,[1] WILLIAM E. HART,[3]
RICHARD K. BELEW,[4] ARTHUR J. OLSON[1]

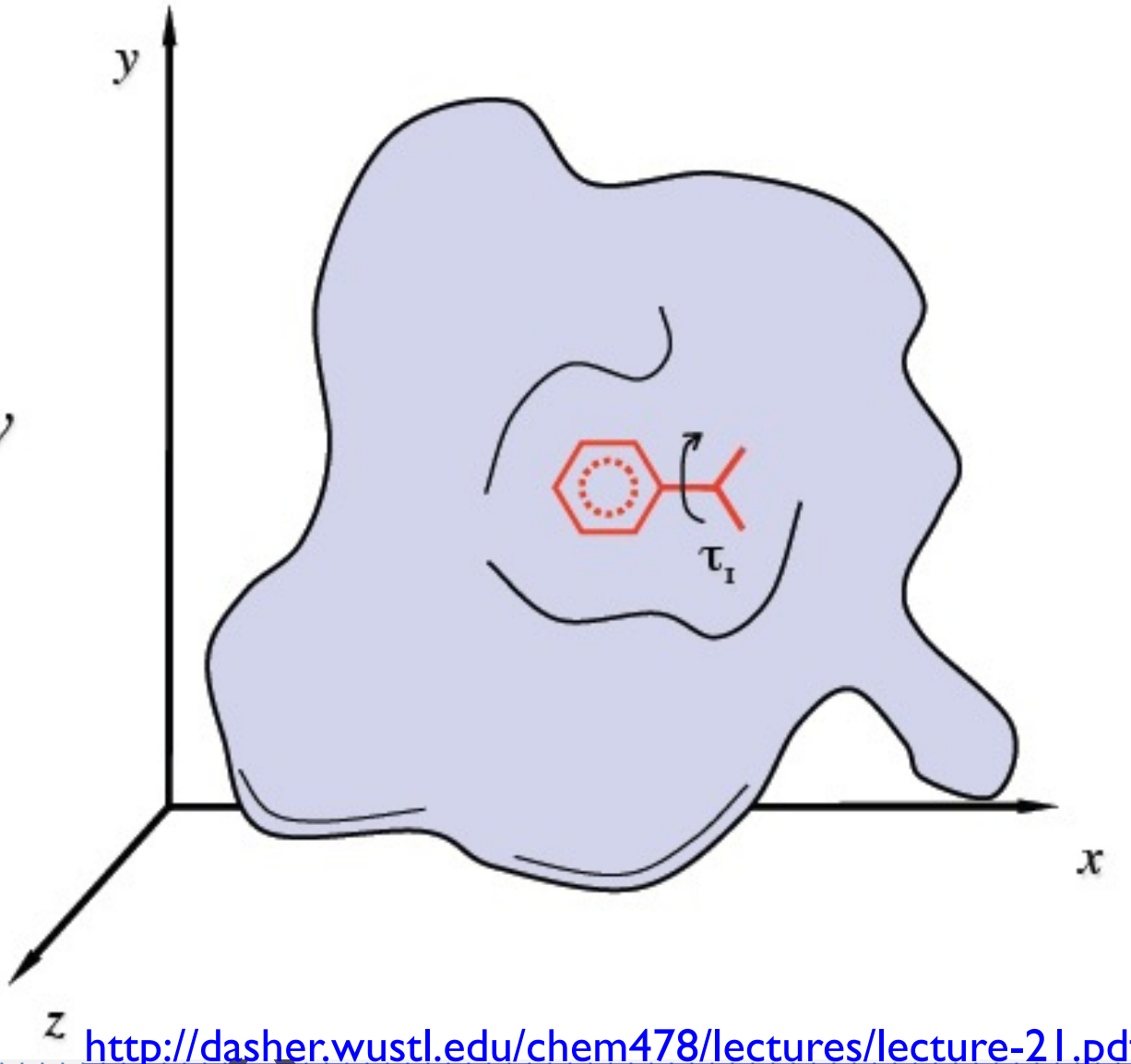
[1] *Department of Molecular Biology, MB-5, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000*
[2] *Hewlett-Packard, San Diego, California*
[3] *Applied Mathematics Department, Sandia National Laboratories, Albuqurque, NM*
[4] *Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA*

Simulated Annealing (Goodsell et al. 1990)

https://en.wikipedia.org/wiki/Simulated_annealing

Defining a Docking

- Position
  - x, y, z
- Orientation
  - qx, qy, qz, qw
- Torsions
  - $\tau_1, \tau_2, \ldots \tau_n$



http://dasher.wustl.edu/chem478/lectures/lecture-21.pdf

Genetic algorithm

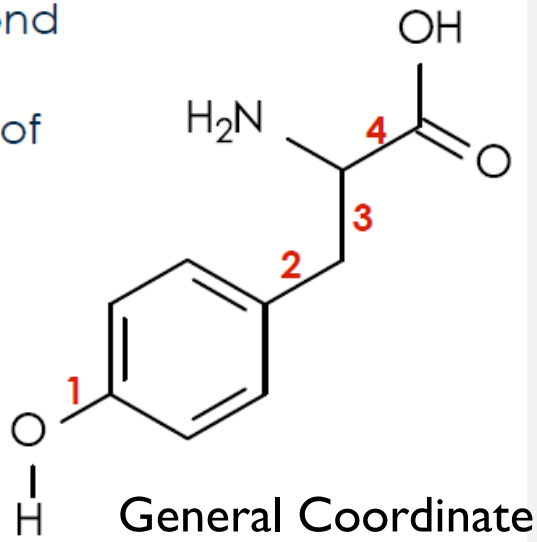
(Morris et al. 1998)

Use of a Genetic Algorithm as a sampling method

- Each conformation is described as a set of rotational angles.
- 64 possible angles are allowed to each of the bond in the ligand.
- Each plausible dihedral angle is codified in a set of binary bits (2^6=64)
- Each conformation is codified by a so called chromosome with 4 x 6 bits (0 or 1)

111010.010110.001011.010010

Φ_1 Φ_2 ...

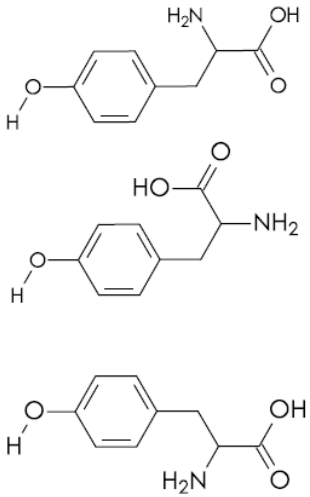


General Coordinate System

$\Phi_1 = 1\times2^5 + 1\times2^4 + 1\times2^3 + 0\times2^2 + 1\times2^1 + 0\times2^0 = 58°$

Genetic algorithm

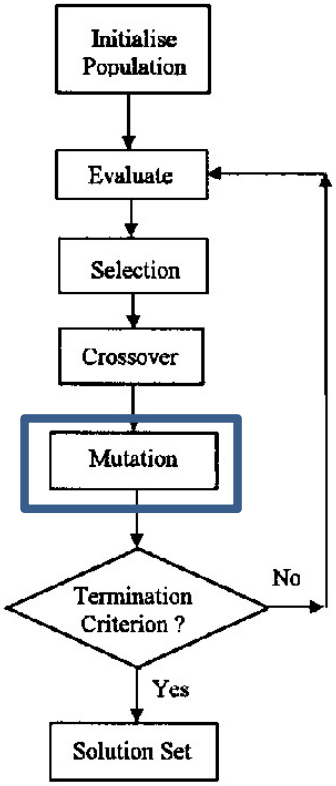
Population (ie, set of chromosomes or configurations)



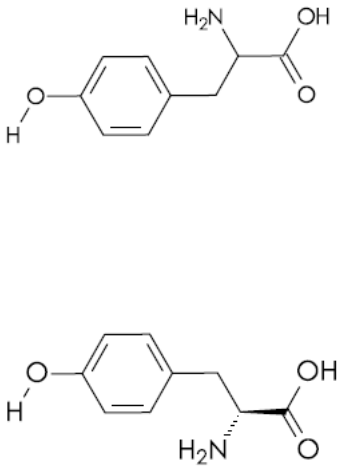
011010.010110.011010.010111 ← Chromosome
111010.010110.001011.010010
001010.010101.000101.010001
101001.101110.101010.001000
001010.101000.011101.001011

Gene

Genetic algorithm



3.Genetic operator : Mutation

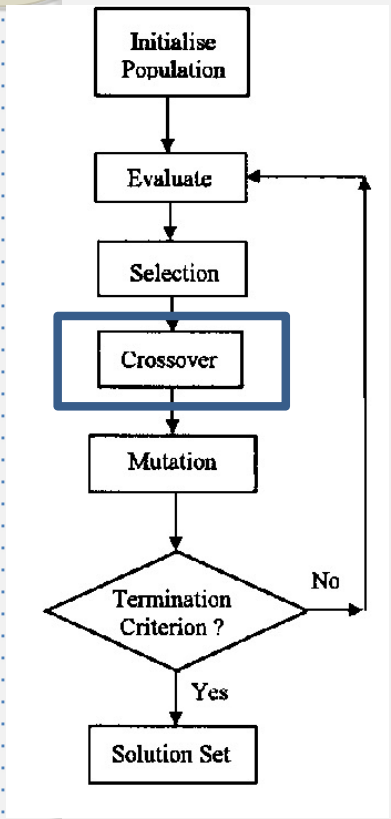


011010.010110.011010.010111

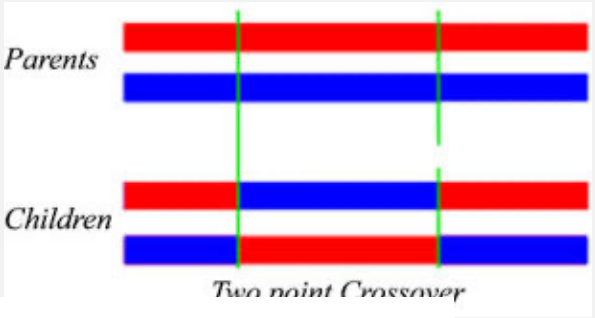
Single mutation

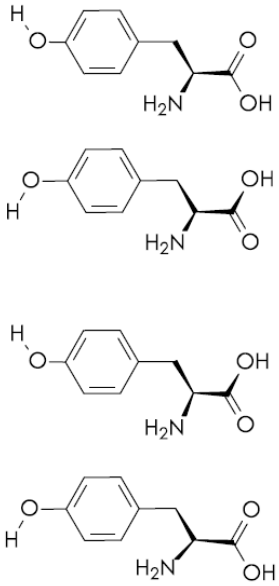
011010.011110.011110.010111

Genetic algorithm



3.Genetic operator : Crossover





001010.010101.000101.010001

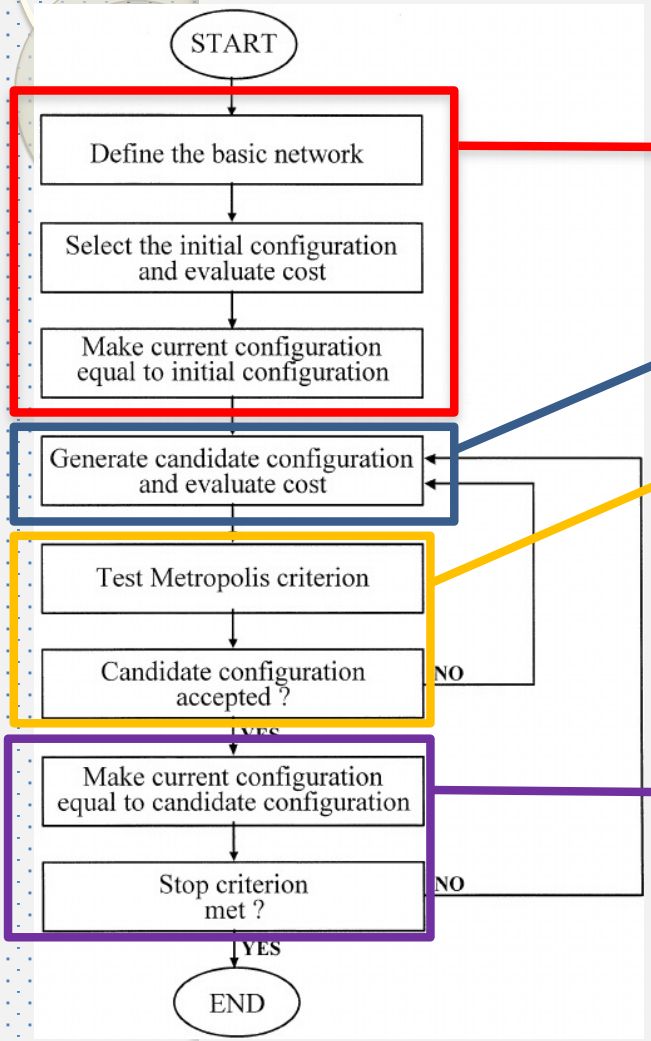
011010.010110.011010.010111

Recombination

001010.010101.011010.010111

011010.010110. 000101.010001

Monte Carlo Sampling

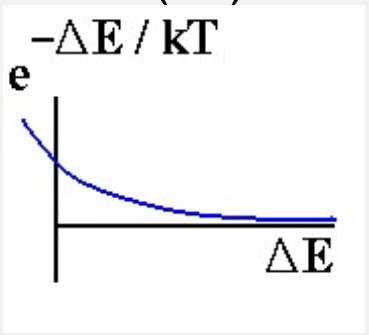


Step1. set initial temperature and use random number generator generate random molecular conformation

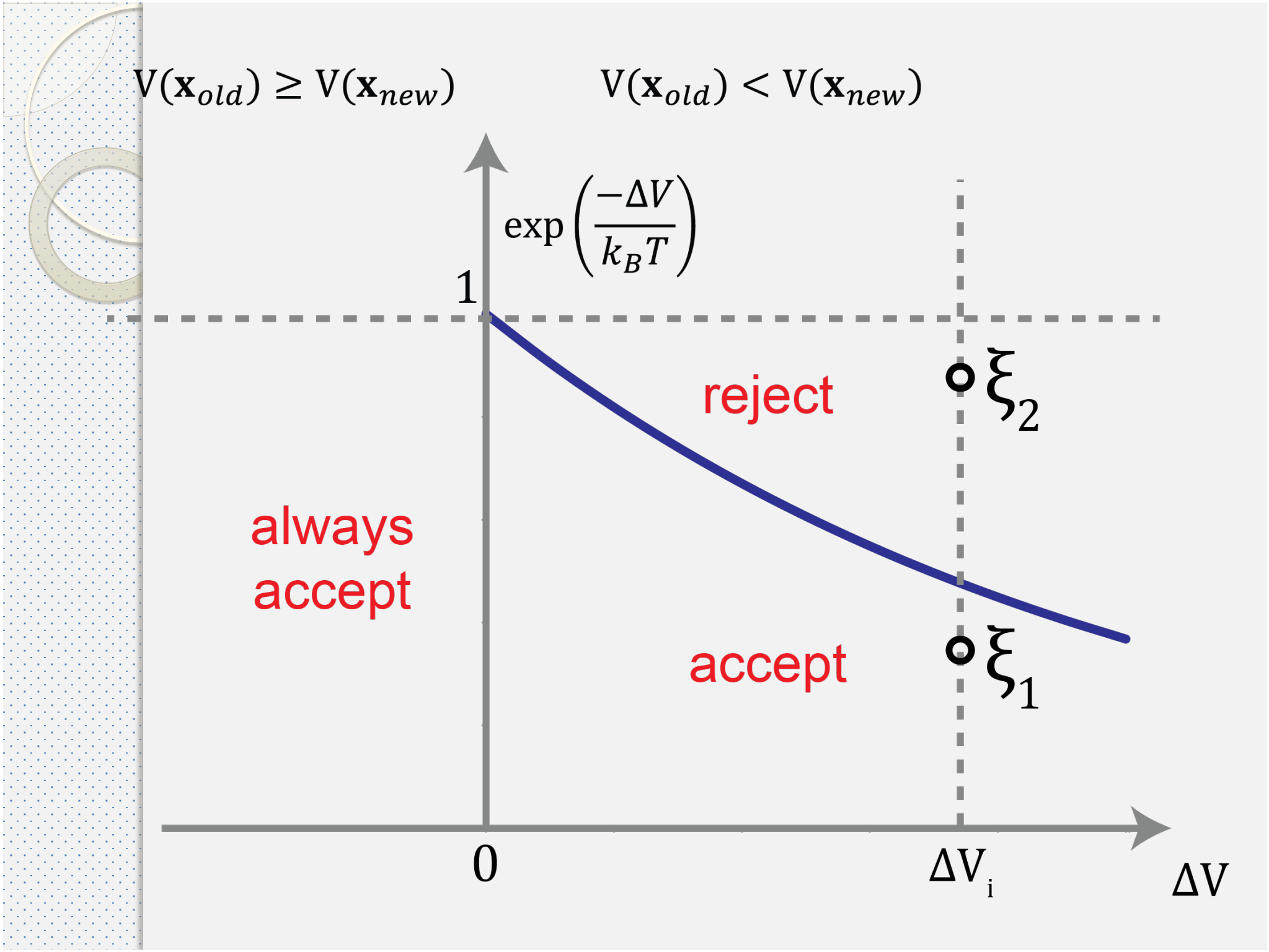
Step2. generate a new molecular conformation and calculate molecular conformation energy

Step3. Compare energy between previous(EA) and current (EB) conformations.

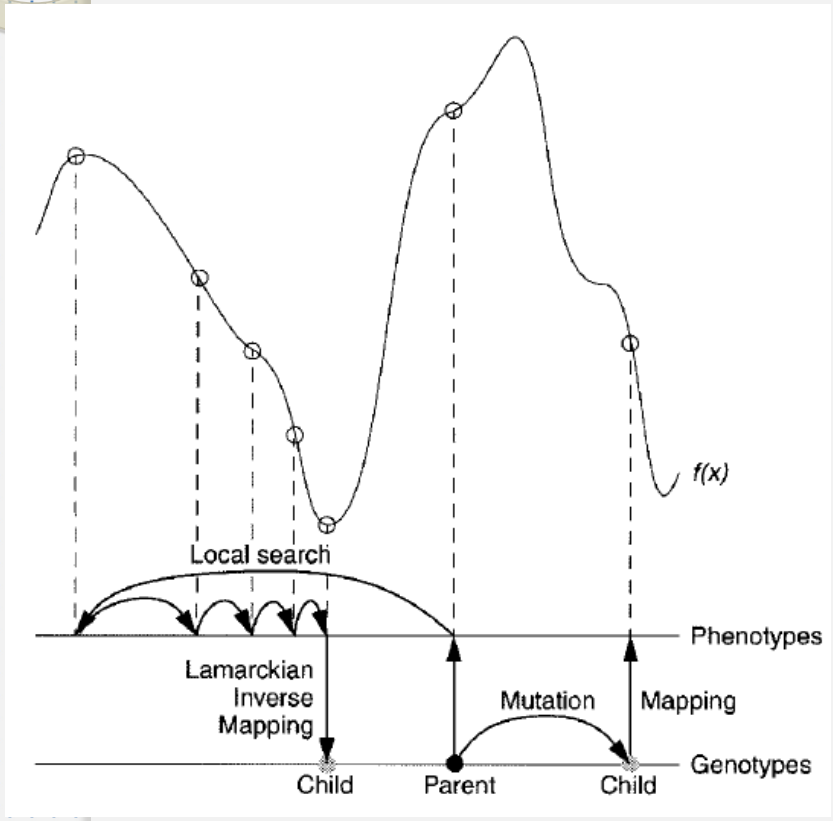
$$P = \begin{cases} 1 & ,if\ \Delta E \le 0 \\ e^{\frac{-\Delta E}{T}} & ,if\ \Delta E > 0 \end{cases}$$



Step4. simulated annealing iterates over time until one of the termination criteria is met. termination criteria: predetermined number of iterations, or termination temperature is reach.


$V(\mathbf{x}_{old}) \geq V(\mathbf{x}_{new})$
$V(\mathbf{x}_{old}) < V(\mathbf{x}_{new})$
$\exp\left(\frac{-\Delta V}{k_B T}\right)$
1
reject
ξ_2
always
accept
accept
ξ_1
0
ΔV_i
ΔV

Lamarckian genetic algorithm



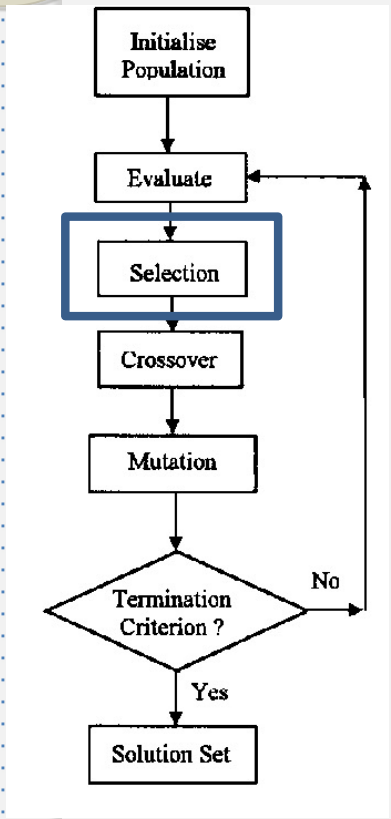
The local search method is *adaptive*, in that it adjusts the step size depending upon the recent history of energies:

1.a user-defined number of consecutive failures, or increases in energy, cause the step size to be doubled

2.a user-defined number of consecutive successes, or decreases in energy, cause the step size to be halved

translational step size :0.2 A°
orientational and torsional step sizes :58°

Genetic algorithm



2. selection:

The integer number of offspring

$$n_o = \frac{f_w - f_i}{f_w - \langle f \rangle} \qquad f_w \neq \langle f \rangle$$

f_i: the fitness of the individual

$< f >$: the mean fitness of the population

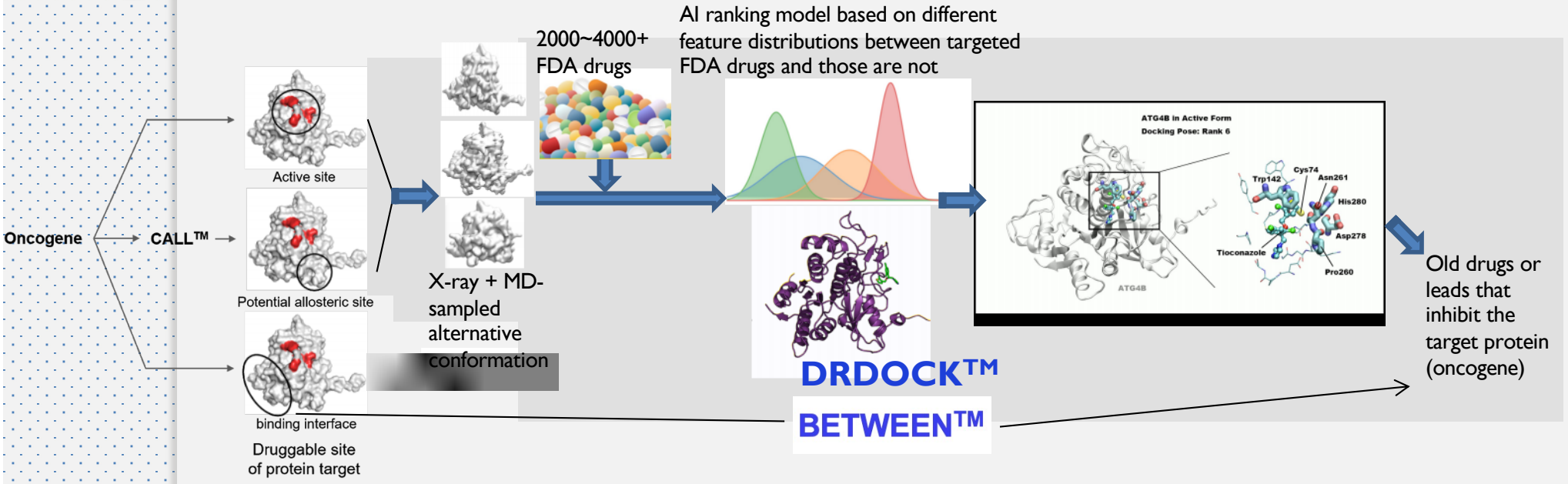
The fitness of the worst individual, or highest energy

Problems - affinity and specificity/selectivity

Current pipeline in Yang's lab to screen drugs with protein dynamics accentuated

CALL™ (LWY's Yang et al. 2014; Li et al., 2017; Huang et al.,2019):

- An allosteric site predictor based on Linear response theory

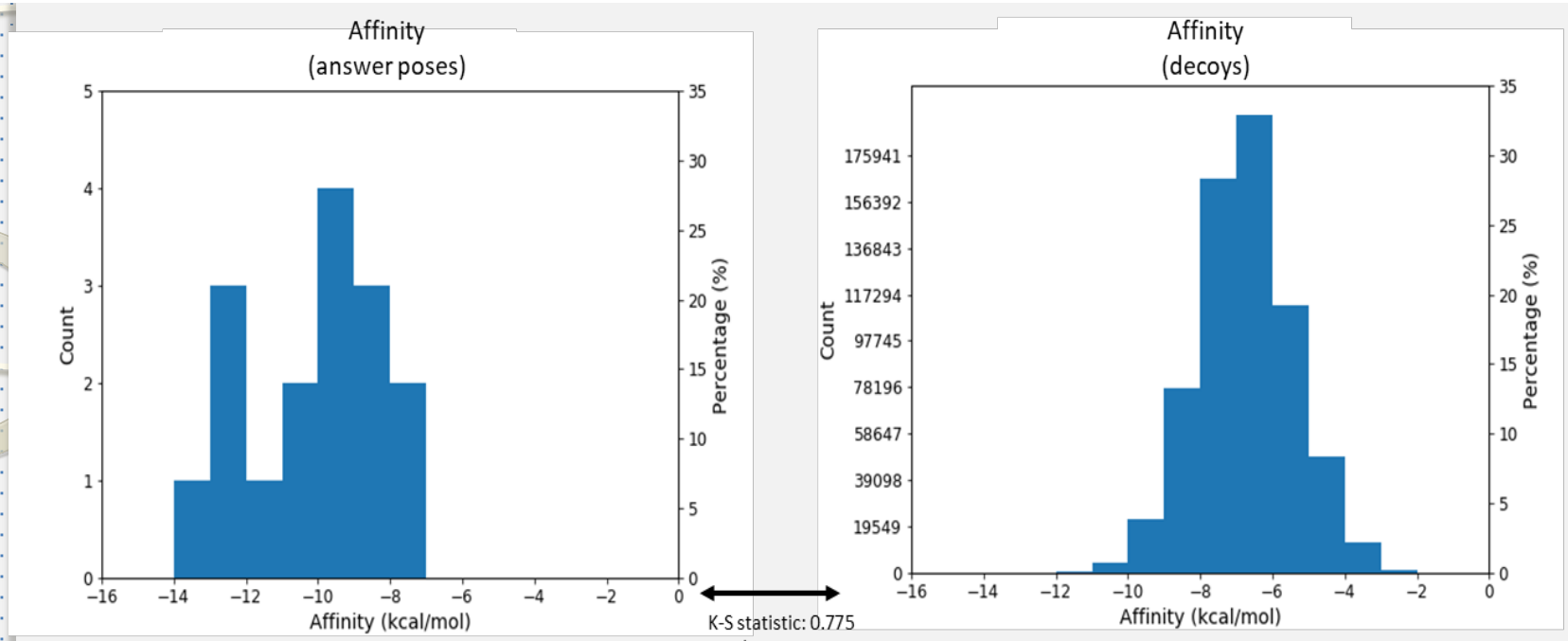


BETWEEN™ (LWY's Lin.,2020):

- A drug screening that integrates docking, entropy-weighted contact, MD-refined bind free energy
- For screening the binding interface

DRDOCK™ (LWY's Tsai et al.2021):

- A drug screening that integrates docking, AI-based ranking score and MD-refined bind free energy
- For the druggable site that has a cavity

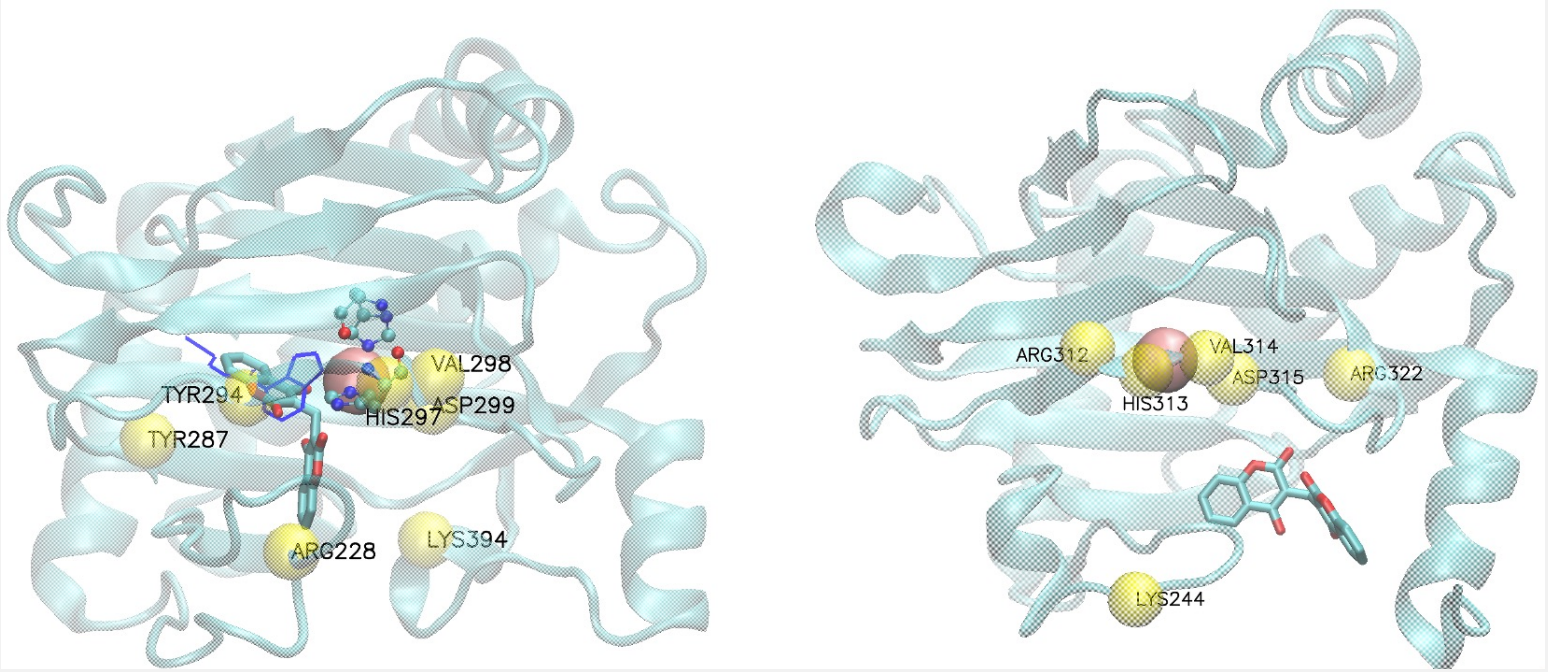


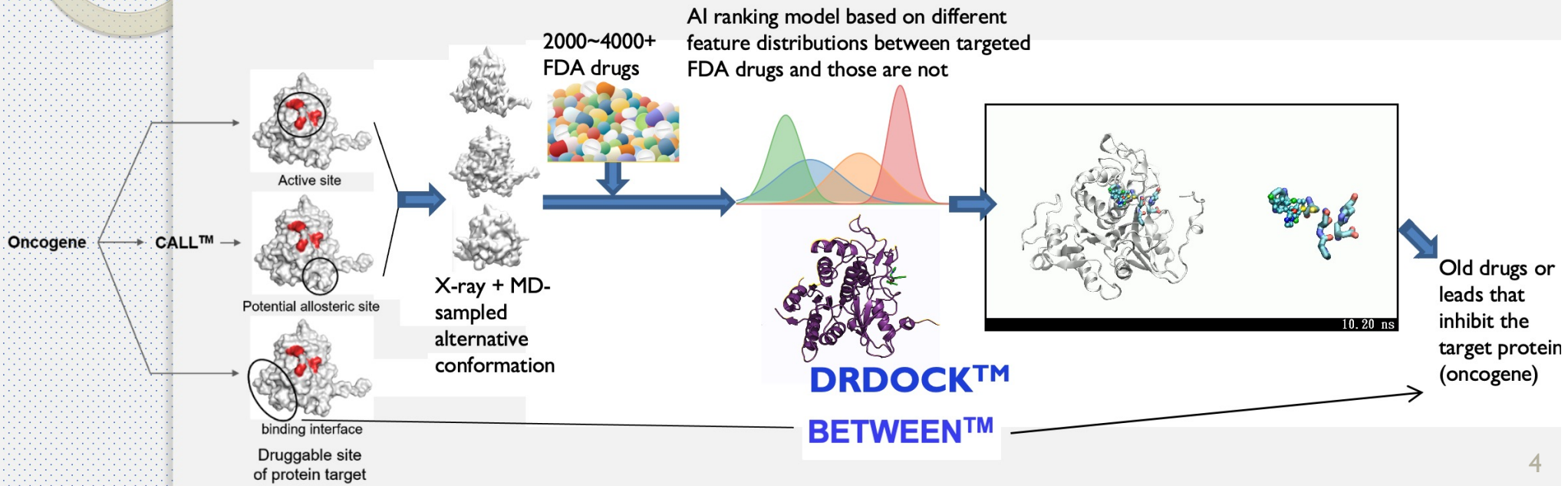
https://www.biorxiv.org/content/10.1101/2021.01.31.429052v4

$$LOD\ Score\ of\ pose\ x = \sum_{f} log \frac{P_f^T(f(x))}{P_f^F(f(x))}$$

, where x represented a sampled docked pose. f represented the feature generation function that took an input pose x and returned the corresponding feature value. $f \in$ { the docking affinity, the distance to the drug target site, the size of poses cluster }. P_f^T was the value distribution of feature f for a docked pose sampled from a true binder. P_f^F was the value distribution of feature f for a docked pose sampled from a decoy. Thus, the LOD score represented how more likely a sampled pose was resulted from a true binder or decoy. The distribution of P_f^T and P_f^F were derived from the docking results of the 2016 drugs on 16 selected crystallized complex structures composing of proteins as drug targets and FDA-approved drugs that was also included in the 2016 drugs according to Westbrook et al., 2019 (Table SX). The 2016 drugs including the one originally in the complex were

Find New Leads for EglN2 but not EglN1





New drug derived from combining fragments from on-target drugs and systemic drugs

(100 times better than the best of the 80 compounds screened and provided by a leading AI-based drug discovery company)

On-target drug efficiency

(FDA 2) (FDA 1)

Identify systemic drug

(FDA 1) (FDA 2)

The new drug integrate from on-target fragment and systemic fragments

IC50 ≤ 20 nM

4 FDA drugs, A-X, A'''-C, A*-D, A'''-B, were screen out from **DRDOCK™** to inhibit EglN2's **active site**.

Immunoblotting assays showed A"-B is the best drug in **molecular level**.

At the cellular level, A-X shows a strong tumor suppressor effect on breast ductal carcinoma (MCF7, T47D), but less on non-tumorigenic breast epithelial cell line (MCF10A).

New drug (A"-X) that integrates **fragment A"** from **on-target** drug (A''-B) and **fragment X** from **systemic** drug (A-X) together also showed specific inhibition to EglN2 but not to EglN1 which share 70% sequence identity.

Only ~**400 nM** of A''-X can inhibit 50% of the breast ductal carcinoma.

(Unpublished data, confidential)

Data are by courtesy of collaborator Prof. Q Zhang

Our new drug:
IC50 = 50nM; EC50 = 200nM

Using pharmacophors analysis to add one functional group to boost the efficacy of *Atomwise*'s top compound (IC50=2500nM; EC50>11,000nM)

Triple negative breast cancer cell line MDA-MB-231

By the end of next year, we can make similar design with a bigger database in 10 days. 6 out of 7 drugs/leads we screened/designed were effective, while Atomwise gave 1/80.

(Unpublished data, confidential)

Table 2. Rank of the docking results for effective FDA drugs repurposed to inhibit EglN2

	drugid	name	docking_rank_in_egln2_X_ray	docking_rank_in_egln2_c0	docking_rank_in_egln2_c1
1	01025	A'''- C	74	224	94
2	01426	A - X	296	170	12
3	01631	A'' - B	191	119	208
4	01450	A* - D	197	222	245

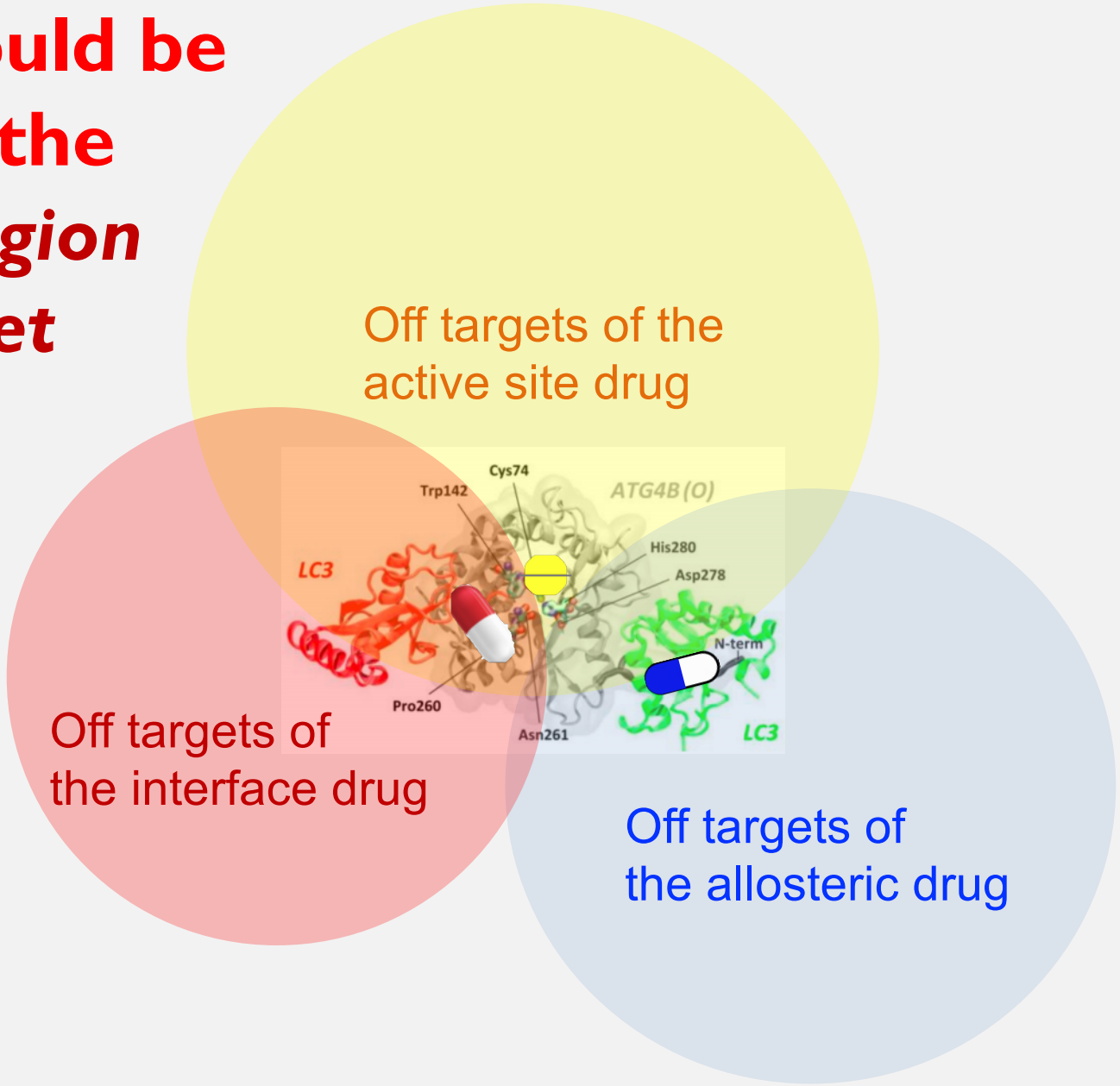
EglN2 and the best FDA drugs were more prioritized in MD conformations than that with the x-ray one

As previously discussed with Figure 3, A''-B is the most inhibitory at the molecular level (with cell-free translation and Western Blot) and A-X is the most breast-cancer-suppressive. The two drugs were prioritized by alternative conformations C0 and C1, where A-X is much favored in C1 conformation and A''-B is prioritized by C0 among a total of 2000+ FDA drugs. The ranking of the drug in EglN1 is only 129 higher than that using the EglN2 x-ray structure, when at least 400 higher is needed as the selection criterium. The second best EglN2 inhibitor, A*-D cannot be selected for a similar reason.

Conventional thinking of the pharma: high affinity = specificity

Yet, evolution-imposed similarity for functional lives at molecular level makes off-target effects inevitable

Specificity could be conferred at the *overlapped region of on/off-target effects.*



FDA/Old drugs (and/or their fragments) are mostly safer, synthesizable, cell-penetrating

Theranostics

Impact factor 8.766

ISSN: 1838-7640

Theranostics 2018; 8(3):830-845. doi:10.7150/thno.22012

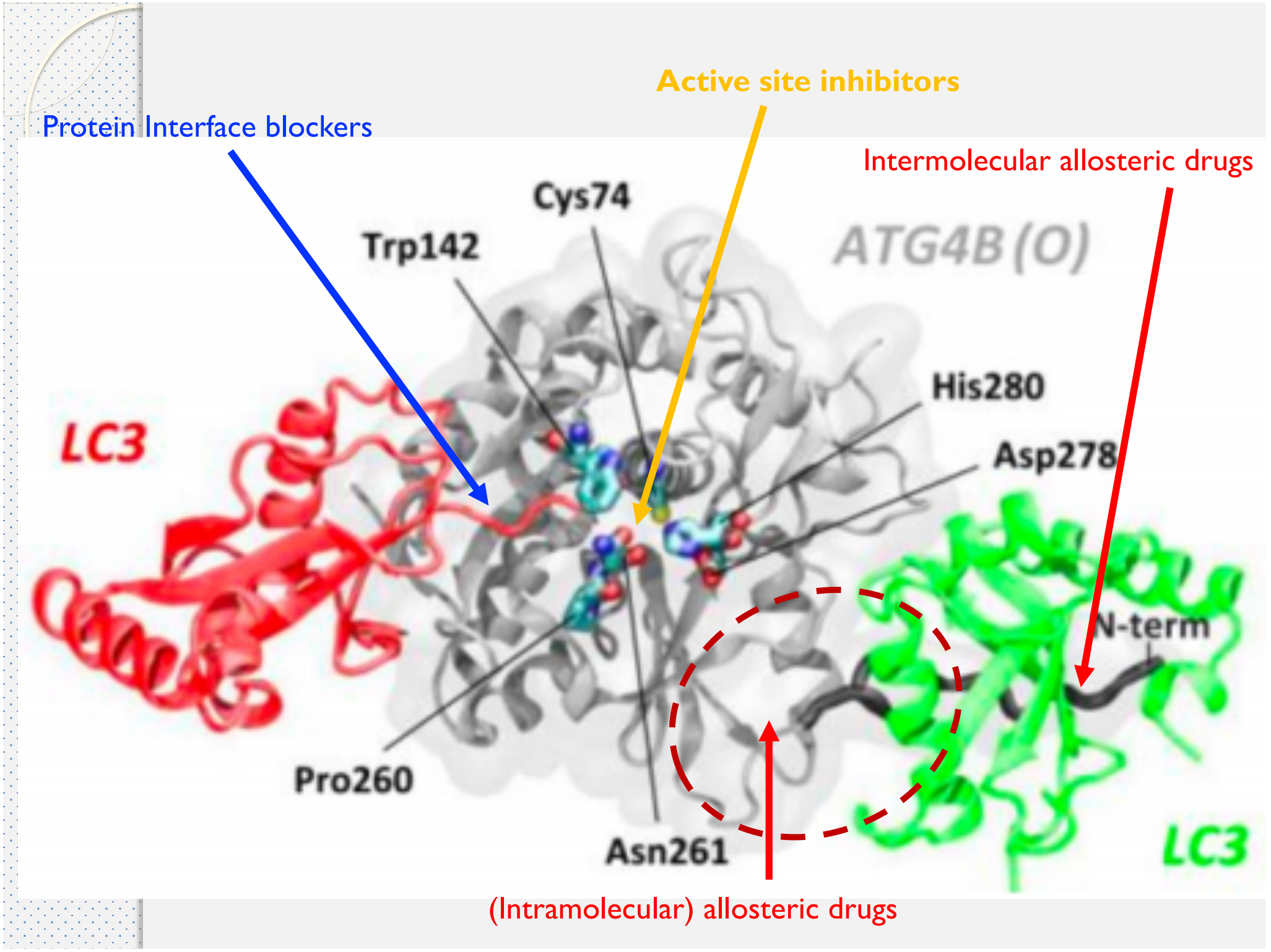
Research Paper

PDF

Drug Repurposing Screening Identifies Tioconazole as an ATG4 Inhibitor that Suppresses Autophagy and Sensitizes Cancer Cells to Chemotherapy

Pei-Feng Liu[1*], Kun-Lin Tsai[2*], Chien-Jen Hsu[1*], Wei-Lun Tsai[3,4], Jin-Shiung Cheng[3], Hsueh-Wei Chang[5,6], Chung-Wai Shiau[7], Yih-Gang Goan[8], Ho-Hsing Tseng[1], Chih-Hsuan Wu[1], John C. Reed[9#], Lee-Wei Yang[2,10], Chih-Wen Shu[1]

1. Department of Medical Education and Research, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan;
2. Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsinchu, Taiwan;
3. Division of Gastroenterology and Hepatology, Department of Internal Medicine, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan;
4. School of Medicine, National Yang-Ming University, Taipei, Taiwan;
5. Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung, Taiwan;
6. Institute of Medical Science and Technology, National Sun Yat-Sen University, Kaohsiung, Taiwan;
7. Institute of Biopharmaceutical Sciences, National Yang-Ming University, Taipei, Taiwan.
8. Division of Thoracic Surgery, Department of Surgery, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan;
9. Sanford-Burnham Medical Research Institute, Program on Apoptosis and Cell Death Research, and Conrad Prebys Center for Chemical Genomics, La Jolla, CA, USA.
10. Physics Division, National Center for Theoretical Sciences, Hsinchu 30013, Taiwan.
* Co-first authors
# Present address: Roche. Basel, Switzerland


Protein Interface blockers
Active site inhibitors
Intermolecular allosteric drugs
Cys74
Trp142
ATG4B (O)
His280
LC3
Asp278
N-term
Pro260
Asn261
LC3
(Intramolecular) allosteric drugs

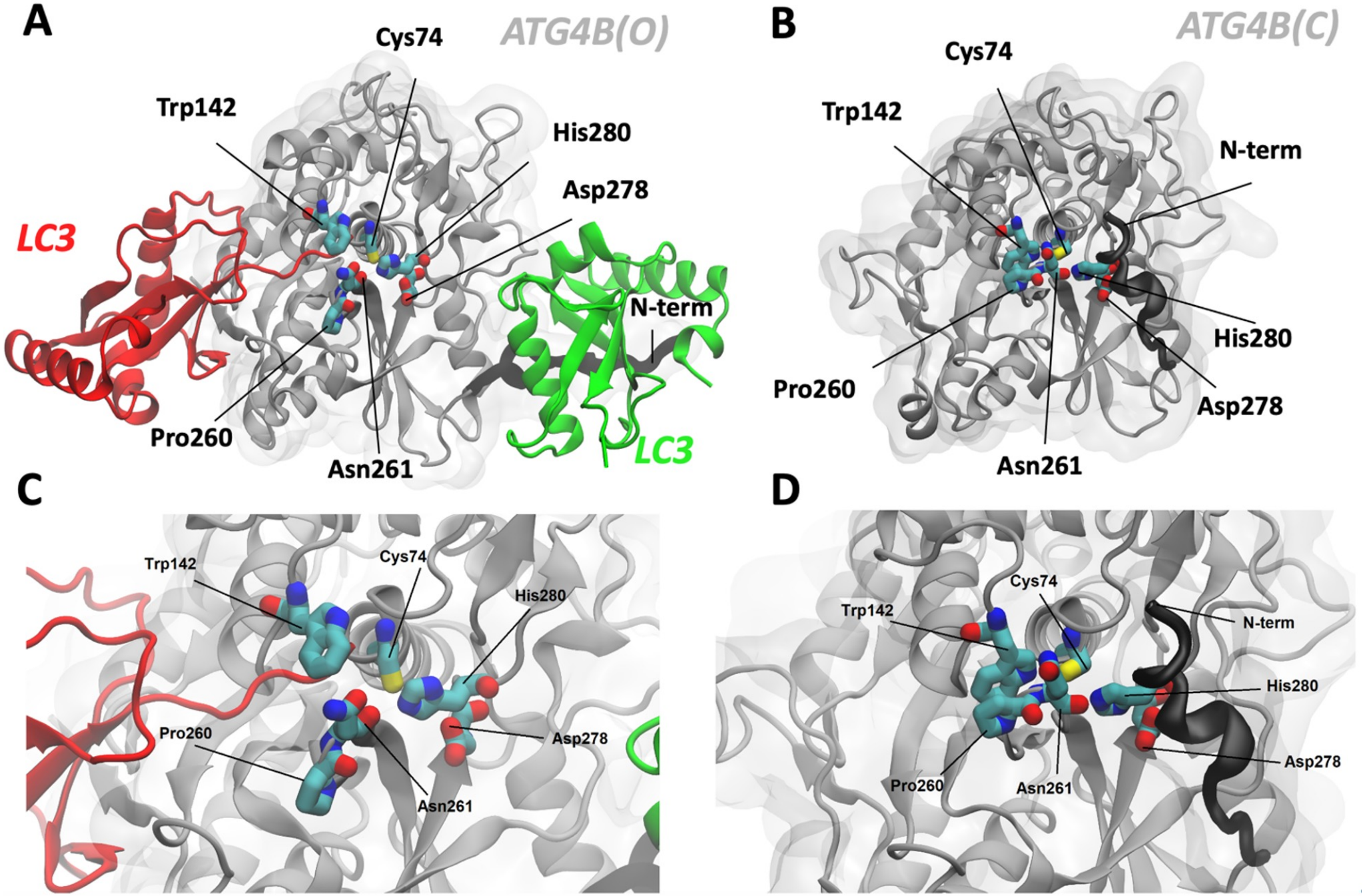
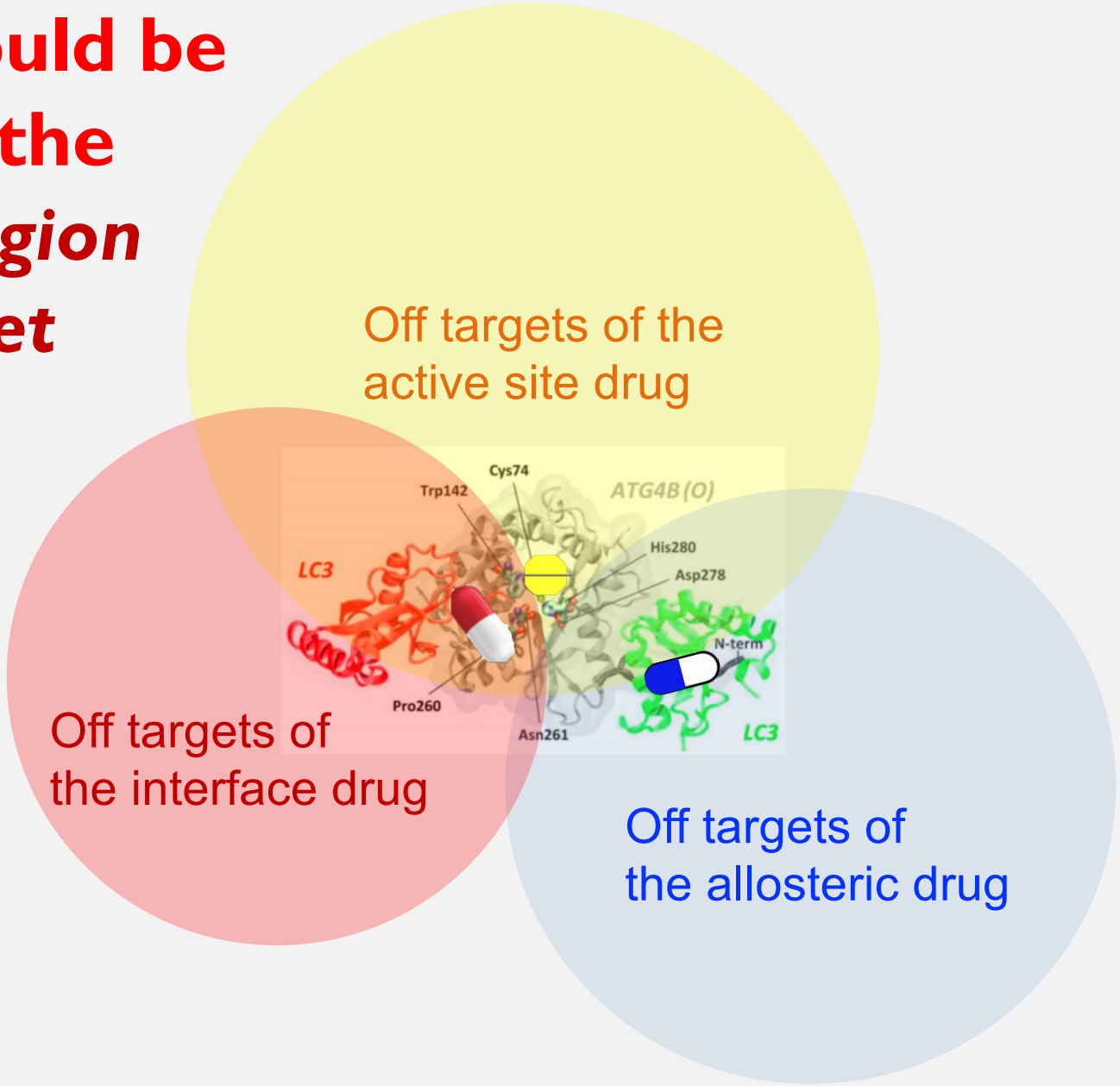

A
ATG4B(O)
Cys74
Trp142
His280
Asp278
LC3
N-term
Pro260
Asn261
LC3
B
ATG4B(C)
Cys74
Trp142
N-term
His280
Pro260
Asp278
Asn261
C
Trp142
Cys74
His280
Pro260
Asp278
Asn261
D
Cys74
Trp142
N-term
His280
Asp278
Pro260
Asn261

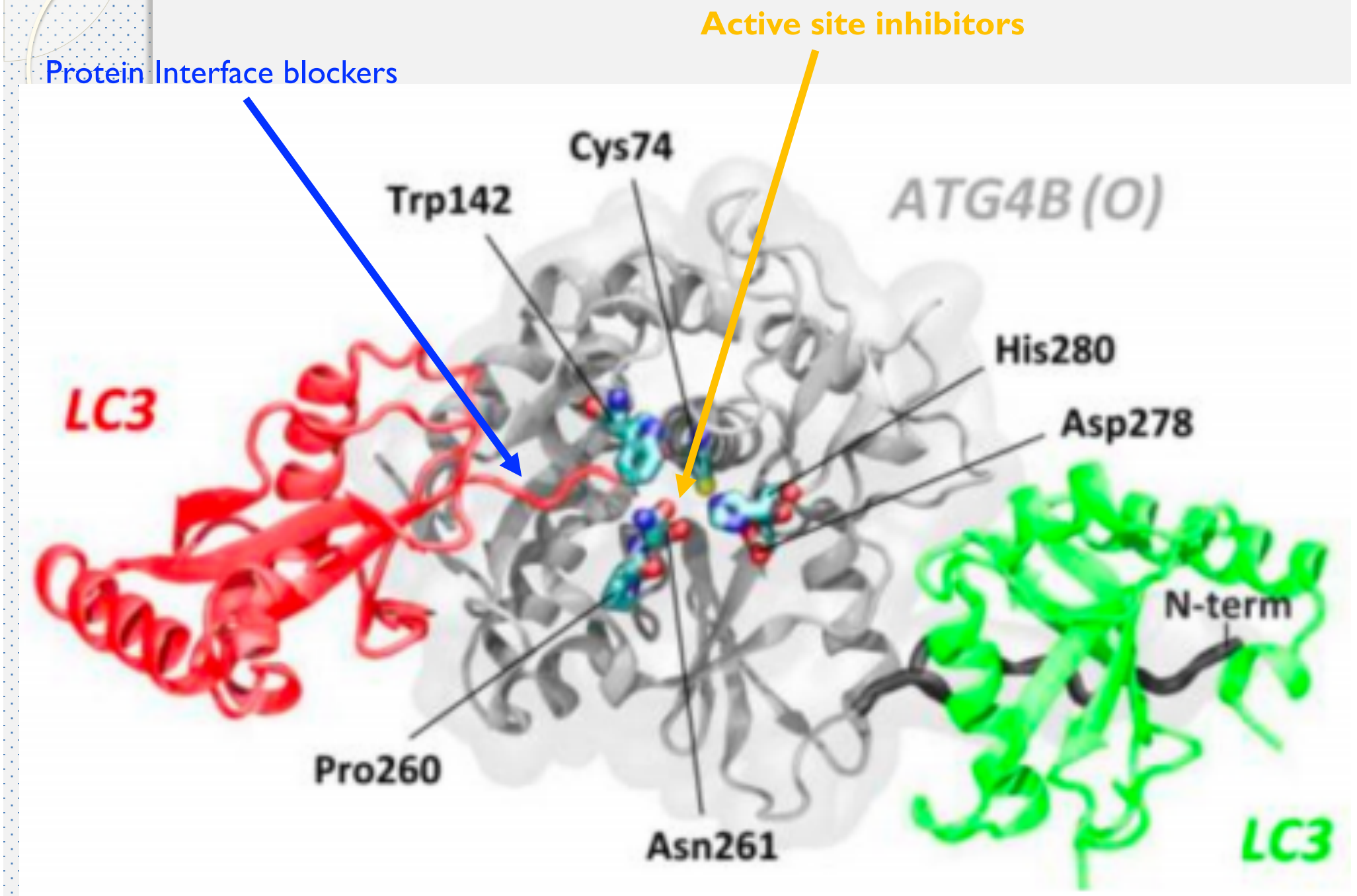
Table 1. List of ATG4B inhibitors

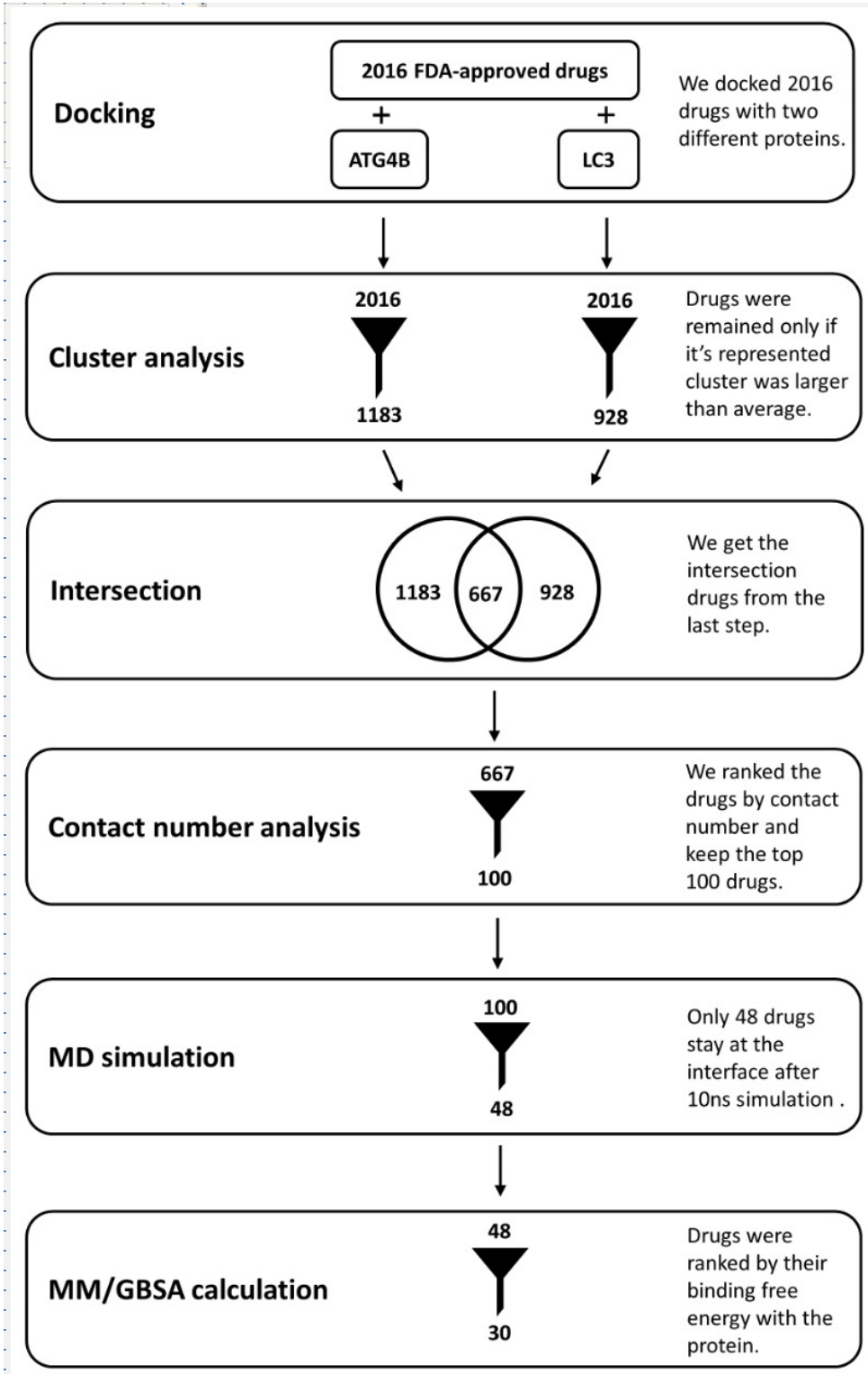
Name/compound ID	ATG4B conformation	Ligand binding sites	IC50	References
NSC185058	Closed-form	The exit of the active site nearby the ATG4B N-terminus	51 μM	Akin et al., 2014
LV-320	Closed-form	The exit of the active site nearby the ATG4B N-terminus	24.5 μM	Bosc et al., 2018
S130	Closed-form	The groove formed by the folded N-terminal tail	3.24 μM	Fu et al., 2018
Tioconazole	**Open-form**	The entry of the active site	1.79 μM	Liu et al., 2018
Compound 1 and 17	**Open-form**	The entry of the active site	10 and 12 μM	Endo et al., 2019
VB	**Open-form**	The interface of LC3 binding site	~5 μM	Yang et al
Allo1 and Allo2	**Open-form**	Potential Allosteric site	~10 μM	Yang et al
N7 and N9	Open-form	Intermolecular allosteric site on LC3	N.A.	Yang et al

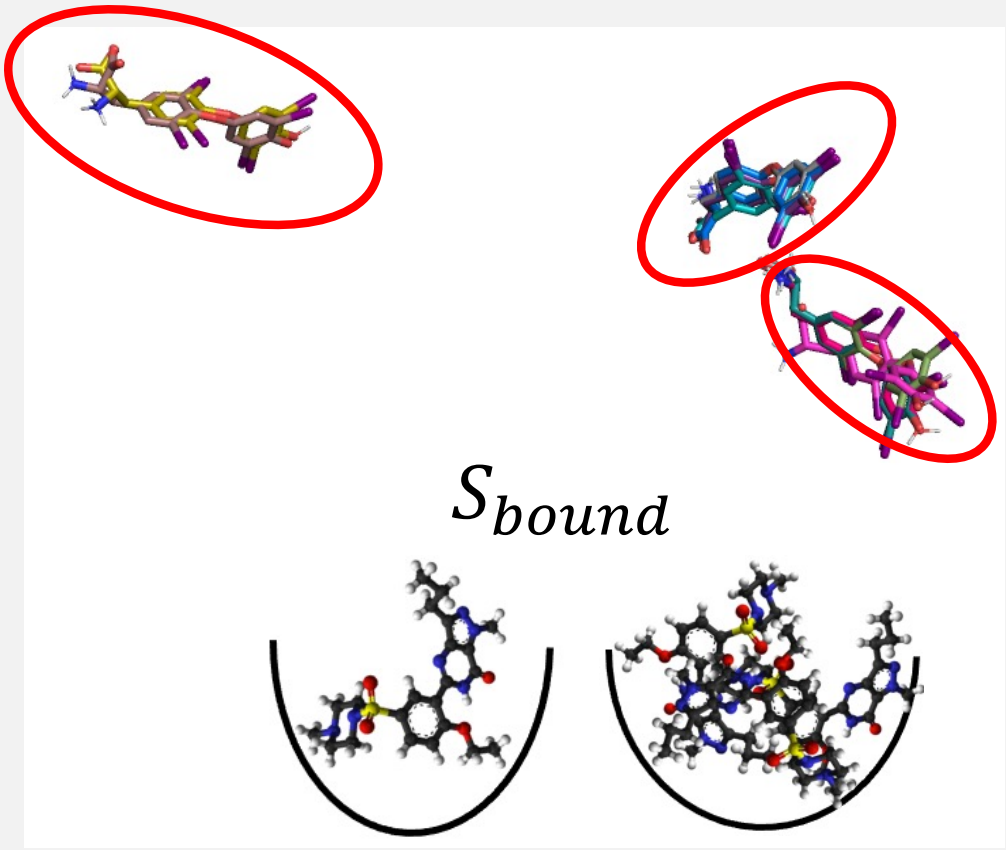
Specificity could be conferred at the *overlapped region of on/off-target effects.*



FDA/Old drugs (and/or their fragments) are mostly safer, synthesizable, cell-penetrating


Active site inhibitors
Protein Interface blockers
Cys74
Trp142
ATG4B (O)
His280
LC3
Asp278
N-term
Pro260
Asn261
LC3





S_{bound}

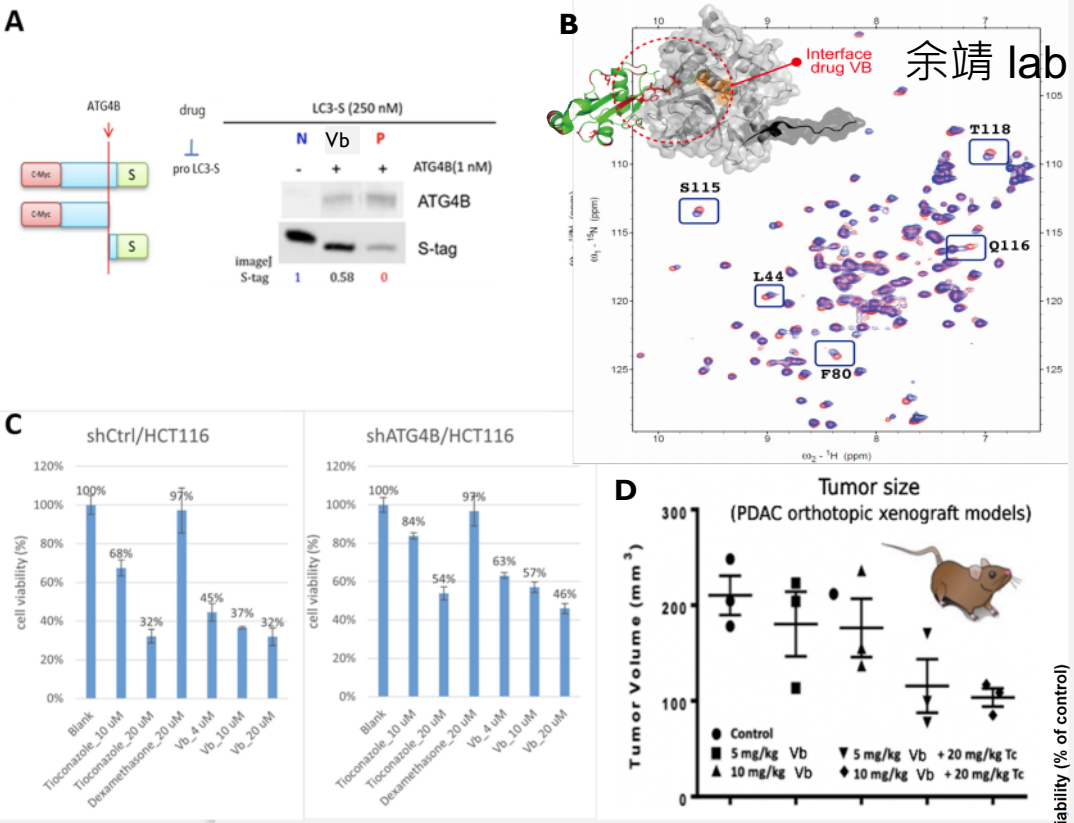
(Small) (Big)

- $\Delta S = S_{bound} - S_{free}$
- $\Delta S < 0$

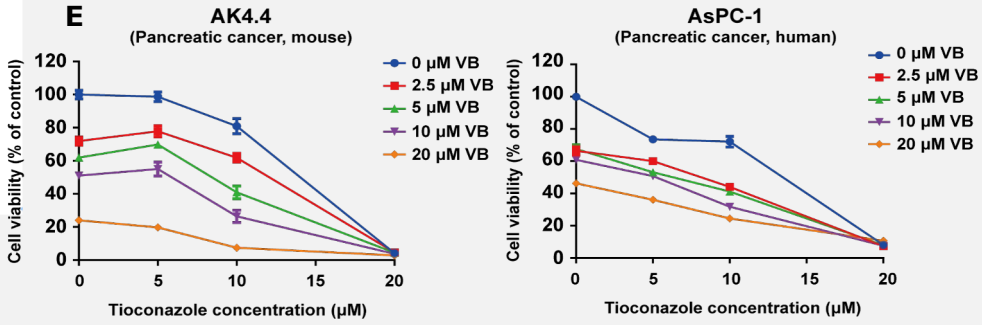
- $\Delta G = \Delta H - T\Delta S$

Synergistic effects (協同效應) by co-drugging active site and protein-protein interface for Cancer cure and SARS-CoV2 suppression

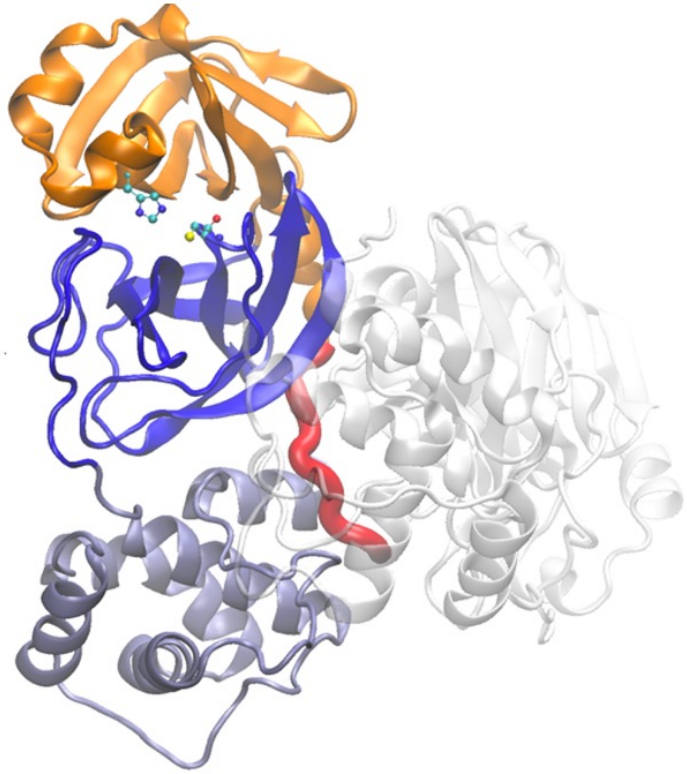
Inhibit ATG4B-LC3 interaction

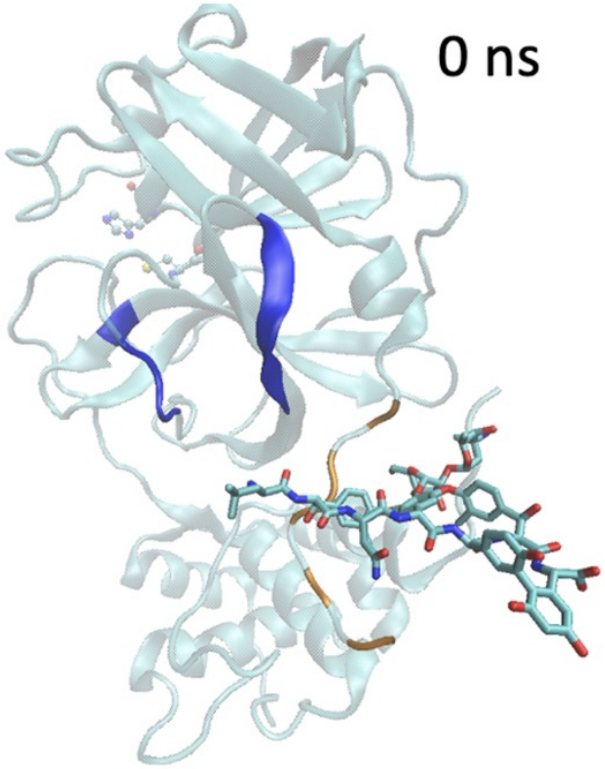


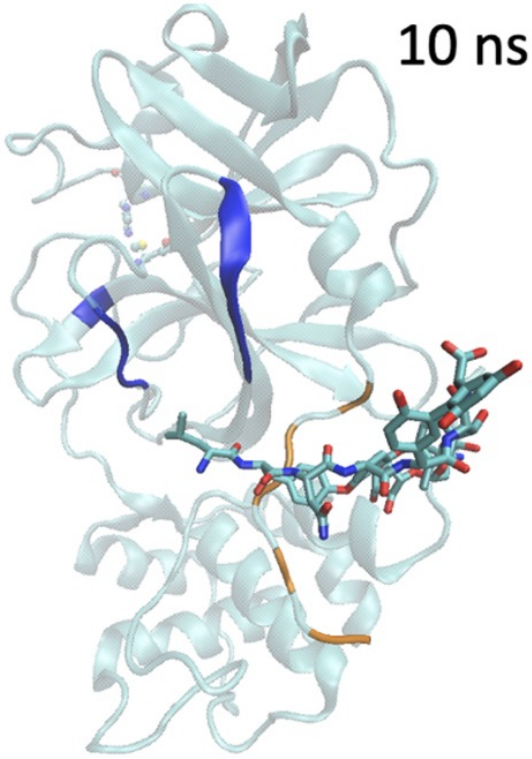
from 陳韻晶 lab



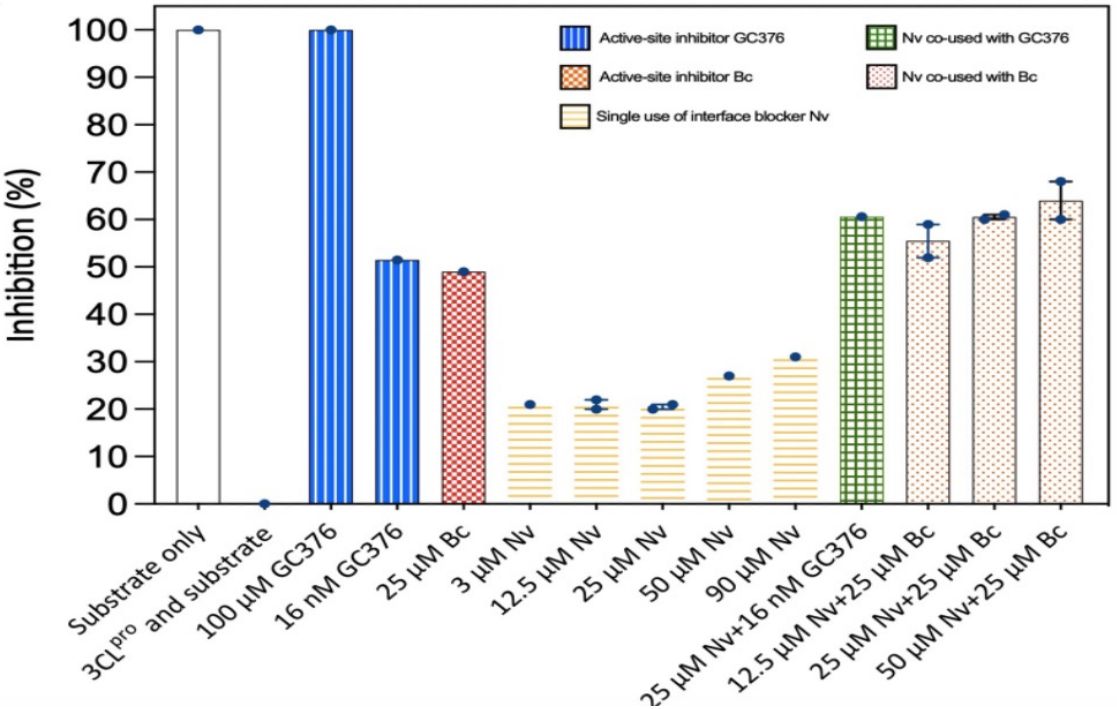
from 陳韻晶 lab





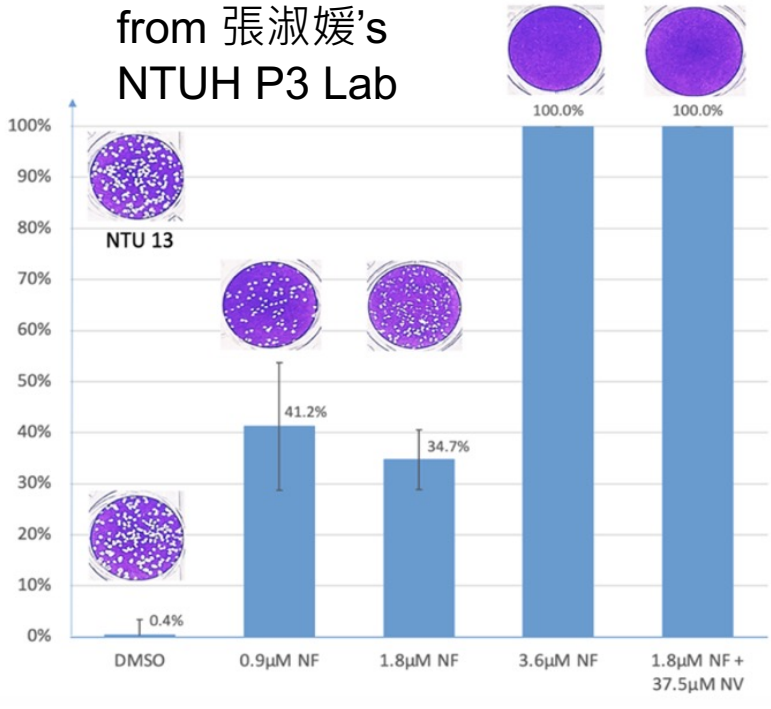


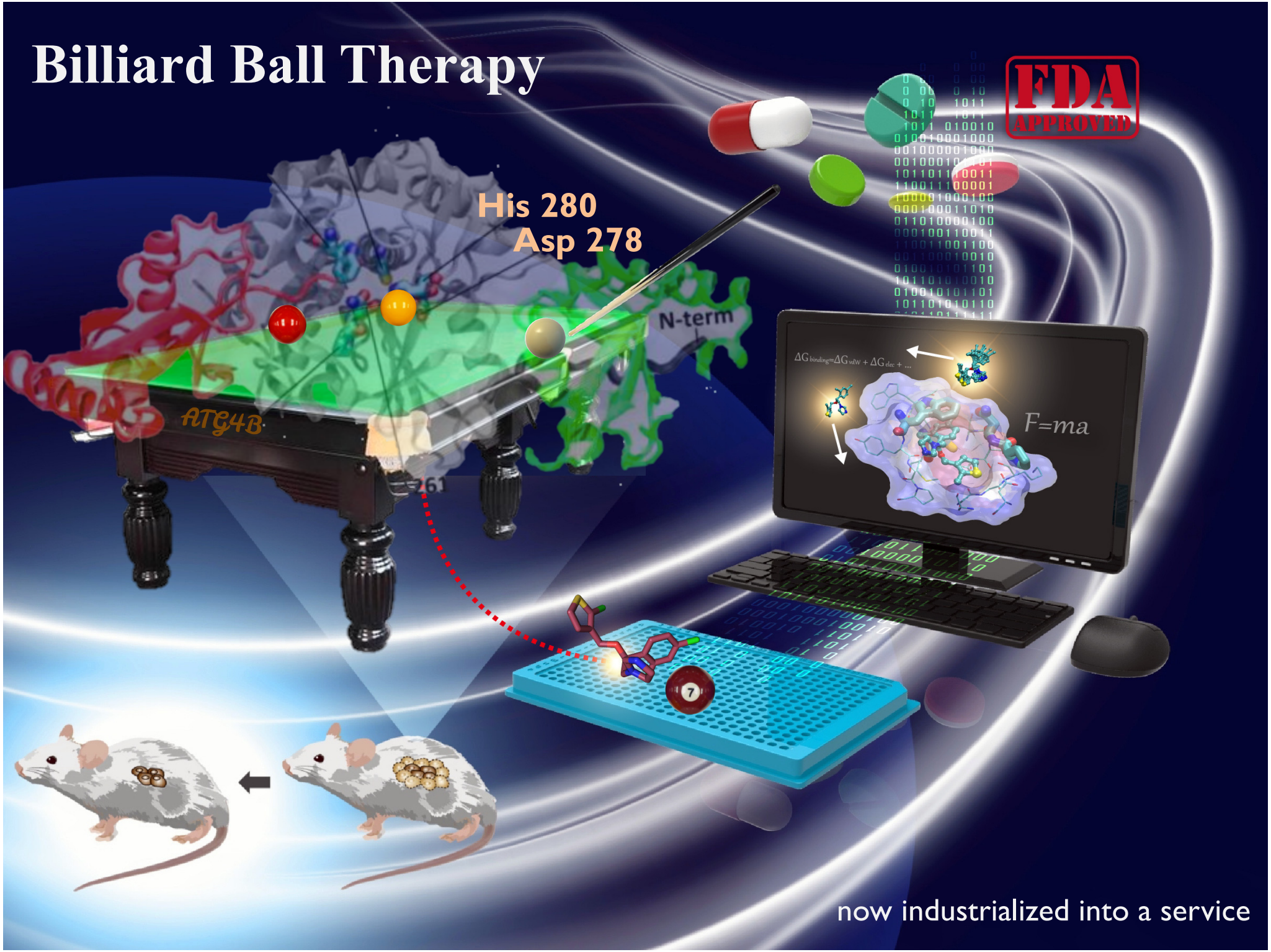
interface drug and active site drug co-used to inhibit 3CL^{pro} catalysis



Virus Inhibition %

from 張淑媛's NTUH P3 Lab




Billiard Ball Therapy
FDA
APPROVED
His 280
Asp 278
N-term
ATG4B
261
F=ma
now industrialized into a service

DRDOCK

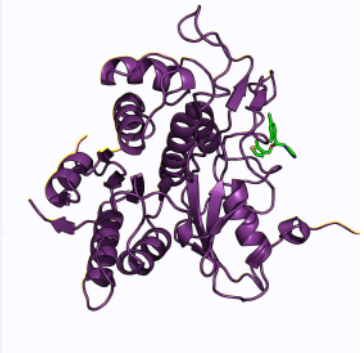
Drug Repurposing – Where will the next success lie?

Home | Tutorials | DynOmics 1.0 | Yang's Lab | Server Status | Example

MD (sampling) → Docking → MD (binding affinity)

What is *DRDOCK*?

DRDOCK is a web server that performs online automatic virtual drug screening of 2016 FDA–approved drugs on a user–submitted target protein. Its goal is to facilitate the drug repurposing to extend the possible treatments toward life–threaten and epidemic diseases. This is achieved by combining the molecular docking and molecular dynamic (MD) simulations to prioritize drugs of strong binders from non–binder. The only things the user required to provide are a well–patched protein structure made of a single peptide chain in PDB format and the target residues IDs to which the drugs expected to bind. *DRDOCK* is freely accessed without login requirements. Find more in Tutorials.



PDB file: 選擇檔案 未選擇任何檔案

Chain ID: ex: A Model No: ex: 1

Residues range: ex: 1-100

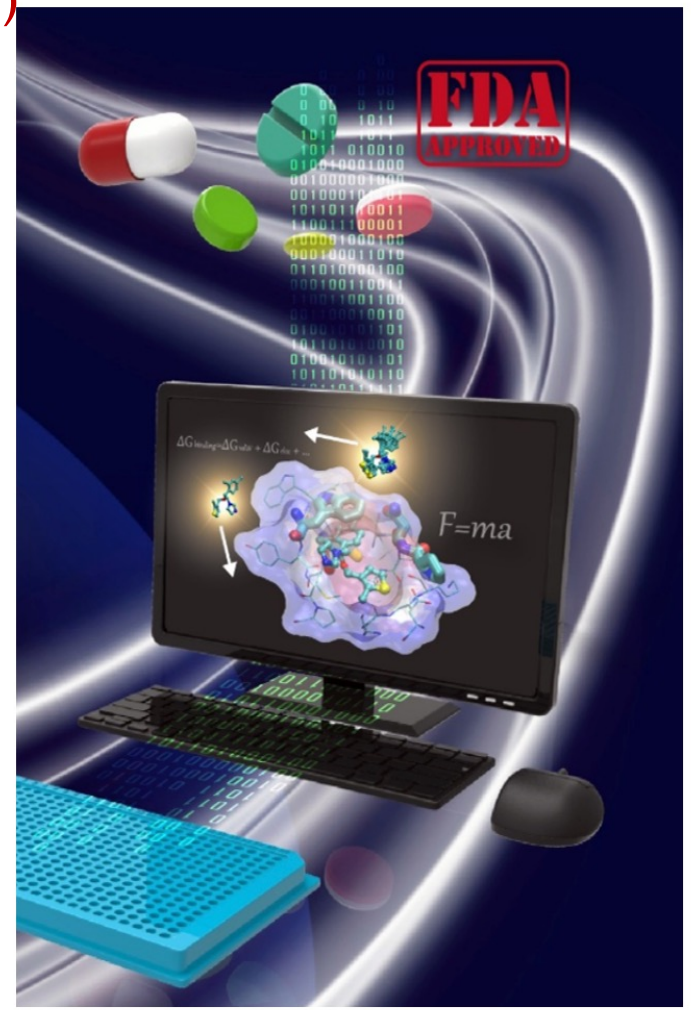
Target sites: ex: 101-102,201

Email:

☐ Relax the structure by MD before docking

Submit

or AlphaFold2 modeling



NTHU

Reference: Tsai,KL. and and Yang,LW. (2021) DRDOCK: A DrugRepurposing platform integrating automated docking, simulation and a log–odds–based drug ranking scheme. **doi:** 10.1101/2021.01.31.429052 *(To be peer–reviewed)*

Contact: The server is maintained by the Yang Lab at the Institute of Bioinformatics and Structural Biology at National Tsing Hua University, Taiwan.

For questions and comments please contact Kun–Lin Tsai.

DRDOCK

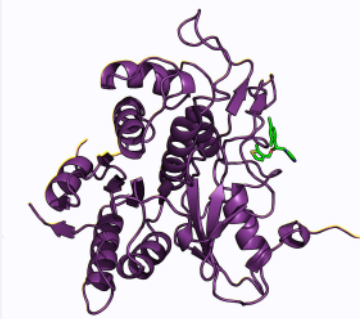
Drug Repurposing – Where will the next success lie?

Home | Tutorials | DynOmics 1.0 | Yang's Lab | Server Status | Example

MD (sampling) → Docking → MD (binding affinity)

What is *DRDOCK*?

DRDOCK is a web server that performs online automatic virtual drug screening of 2016 FDA–approved drugs on a user–submitted target protein. Its goal is to facilitate the drug repurposing to extend the possible treatments toward life–threaten and epidemic diseases. This is achieved by combining the molecular docking and molecular dynamic (MD) simulations to prioritize drugs of strong binders from non–binder. The only things the user required to provide are a well–patched protein structure made of a single peptide chain in PDB format and the target residues IDs to which the drugs expected to bind. Find more in Tutorials.



Protein sequence:

AlphaFold2 (AI-based structure prediction)

```
>6WKS_1|Chain A[auth AAA]|2'-O-
methyltransferase|Severe acute>
SSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATLPKGIMMNVA
KYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLP
TGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPK
TKNVTKENDSKEGFFTYICGFIQQKLALGGSVAIKITEHSWNADL
YKLMGHFAWWTAFVTNVNASSSEAFLIGCNYLGKPREQIDGYVMH
ANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKEGQINDM
```

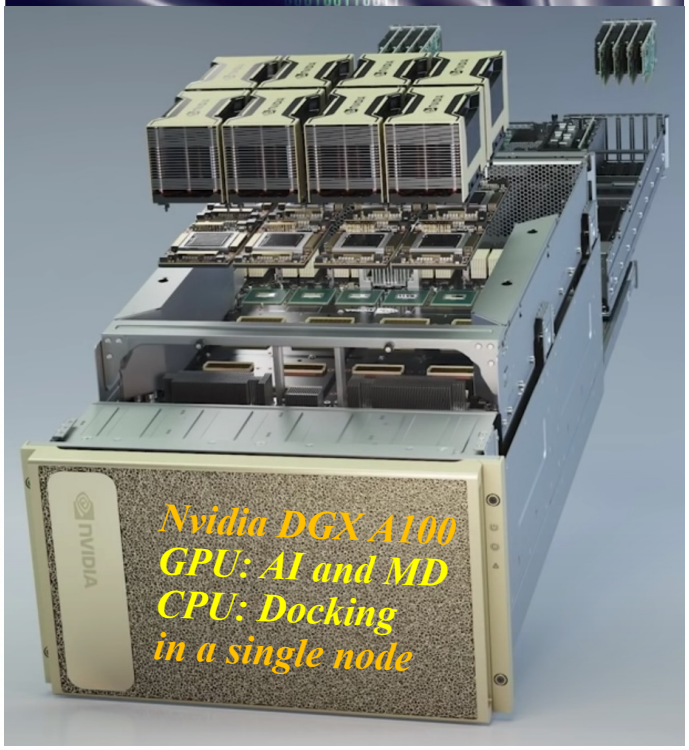
Target sites: ex: 101-102,201

Email:

☐ Relax the structure by MD before docking

Submit

or Upload protein structure



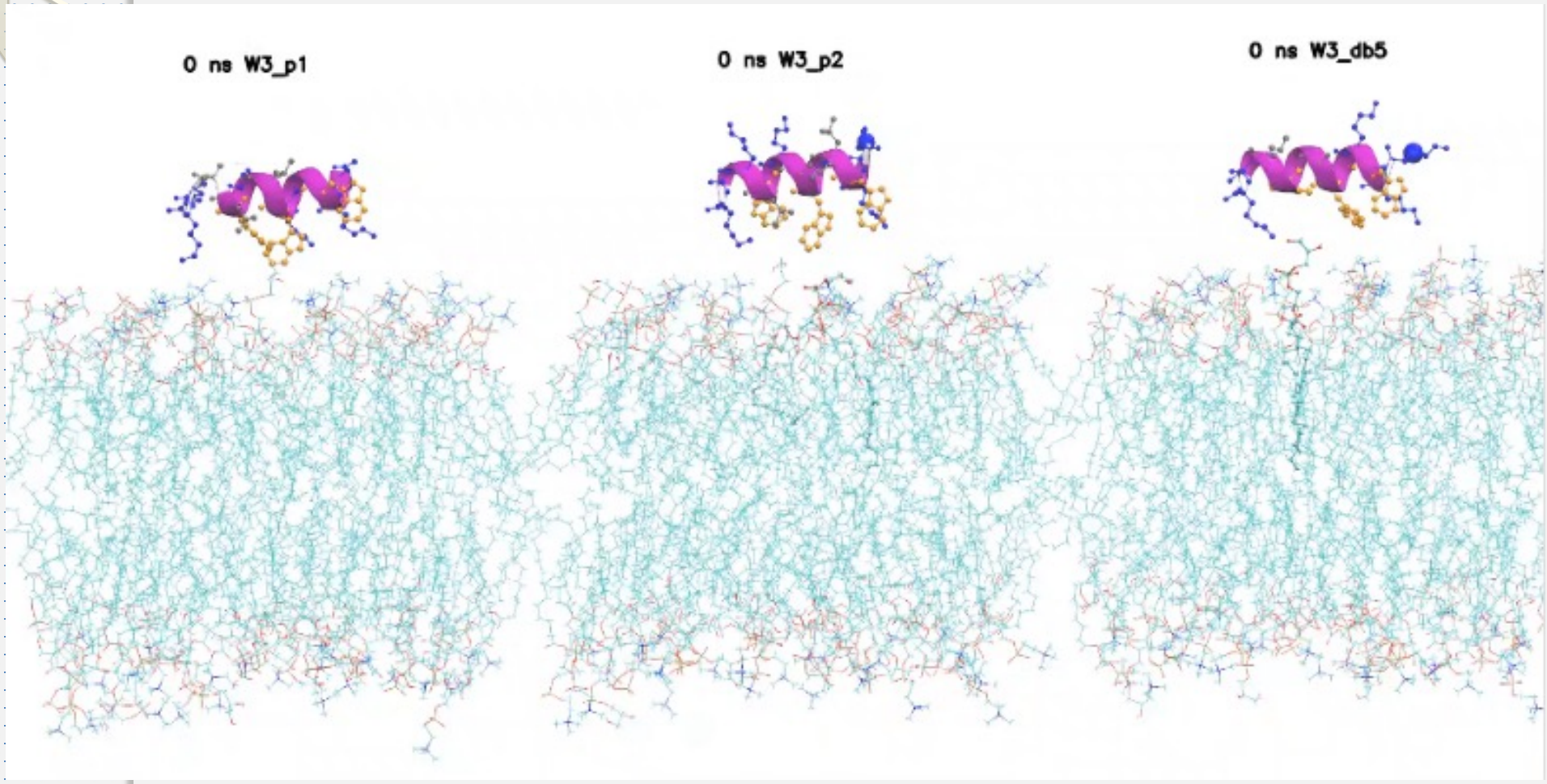
Reference: Tsai,KL. and and Yang,LW. (2021) DRDOCK: A DrugRepurposing platform integrating automated docking, simulation and a log–odds–based drug ranking scheme. *(To be peer–reviewed)*

Contact: The server is maintained by the Yang Lab at the Institute of Bioinformatics and Structural Biology at National Tsing Hua University, Taiwan.

NTHU

For questions and comments please contact Kun–Lin Tsai.

Antimicrobial peptides (AMPs)



Therapeutic Peptide Design

nature communications

Article | Open Access | Published: 10 January 2022

Helical structure motifs made searchable for functional peptide design

Cheng-Yu Tsai, Emmanuel Oluwatobi Salawu, Hongchun Li, Guan-Yu Lin, Ting-Yu Kuo, Liyin Voon, Adarsh Sharma, Kai-Di Hu, Yi-Yun Cheng, Sobha Sahoo, Lutimba Stuart, Chih-Wei Chen, Yuan-Yu Chang, Yu-Lin Lu, Simai Ke, Christopher Llynard D. Ortiz, Bai-Shan Fang, Chen-Chi Wu, Chung-Yu Lan ✉, Hua-Wen Fu ✉ & Lee-Wei Yang ✉

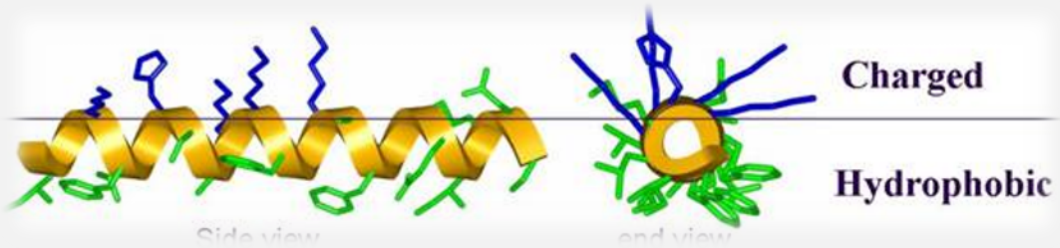
Nature Communications **13**, Article number: 102 (2022) | Cite this article

Antimicrobial peptides (AMPs)





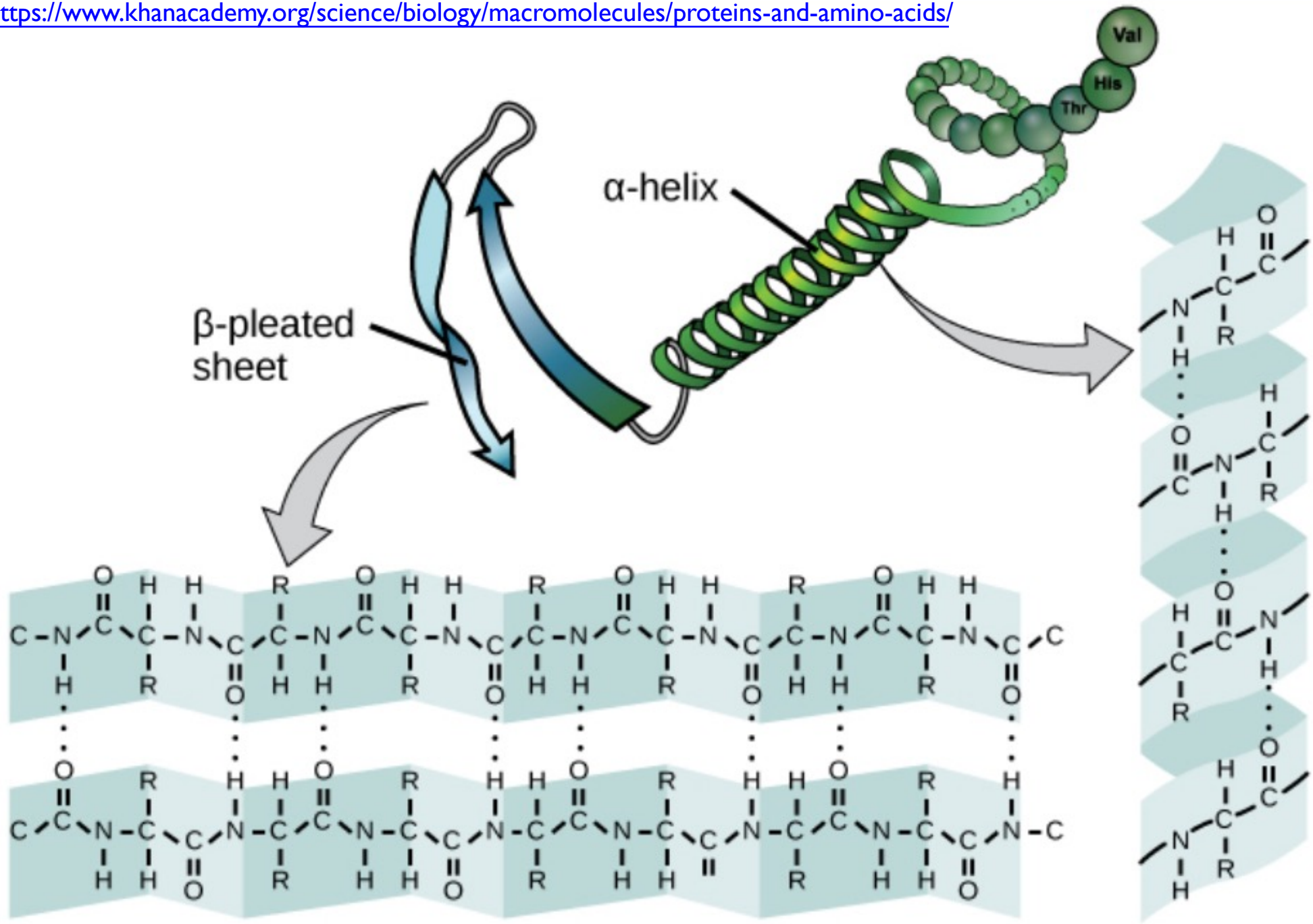
- It is a innate immune defensive system, which are wide spread in bacteria to amphibians, insects, fish, plants, mammals, and even viruses.
- Almost the antimicrobial peptides (generally 12~50 residues) have the common properties : **cationic** and **amphipathic**

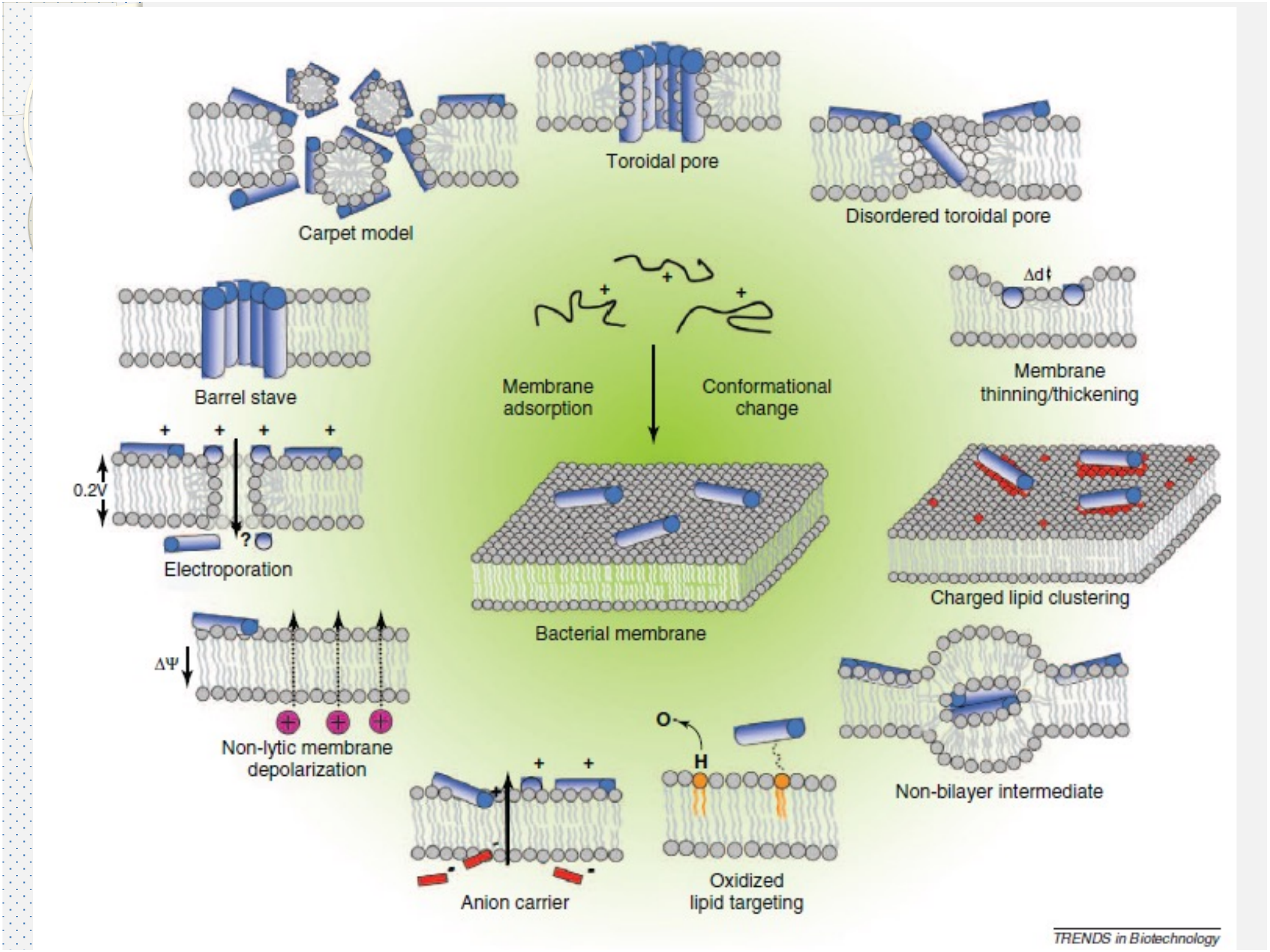


Nature Immunology ***2001***, *2, 1133*

Secondary Protein Structure

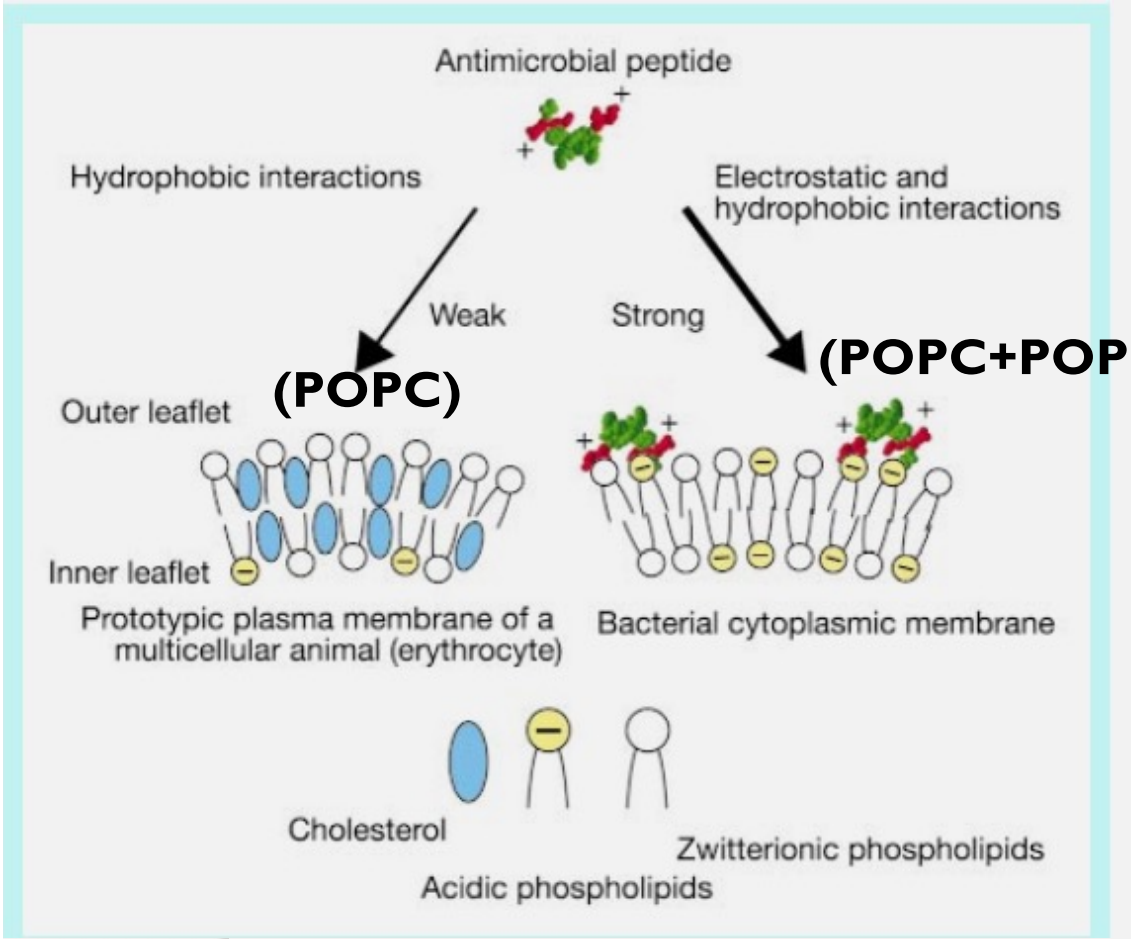
https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/



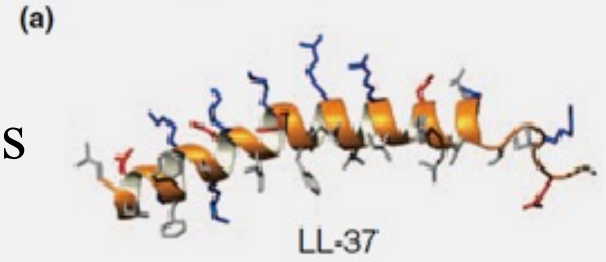

Carpet model
Toroidal pore
Disordered toroidal pore
Barrel stave
Membrane adsorption
Conformational change
Δd
Membrane thinning/thickening
0.2V
?
Electroporation
Bacterial membrane
Charged lipid clustering
ΔΨ
Non-lytic membrane depolarization
Non-bilayer intermediate
O
H
Anion carrier
Oxidized lipid targeting
TRENDS in Biotechnology

Characteristics of AMPs

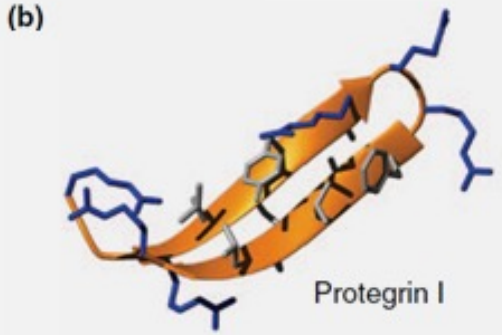
- Generally, 12~50 residues
- cationic and amphipathic
- diversity of secondary structures



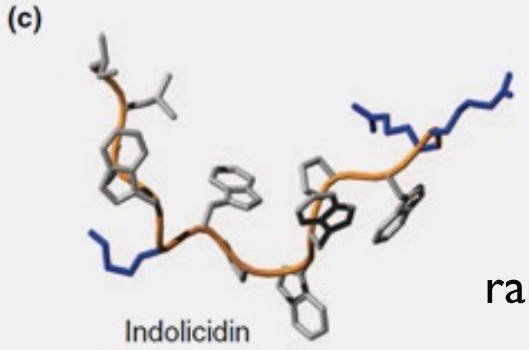
➢ *Nature*, Vol. 415, 2002



α-helical



β-sheet



random coiled

➢ *Trends in Biotechnology*, Vol. 29, No. 9, 2011

https://dyn.life.nthu.edu.tw/design

Pattern Search | BLAST | User Guide

~1.7 million alpha-helices

Pattern Query:

Enter your template to search helix:

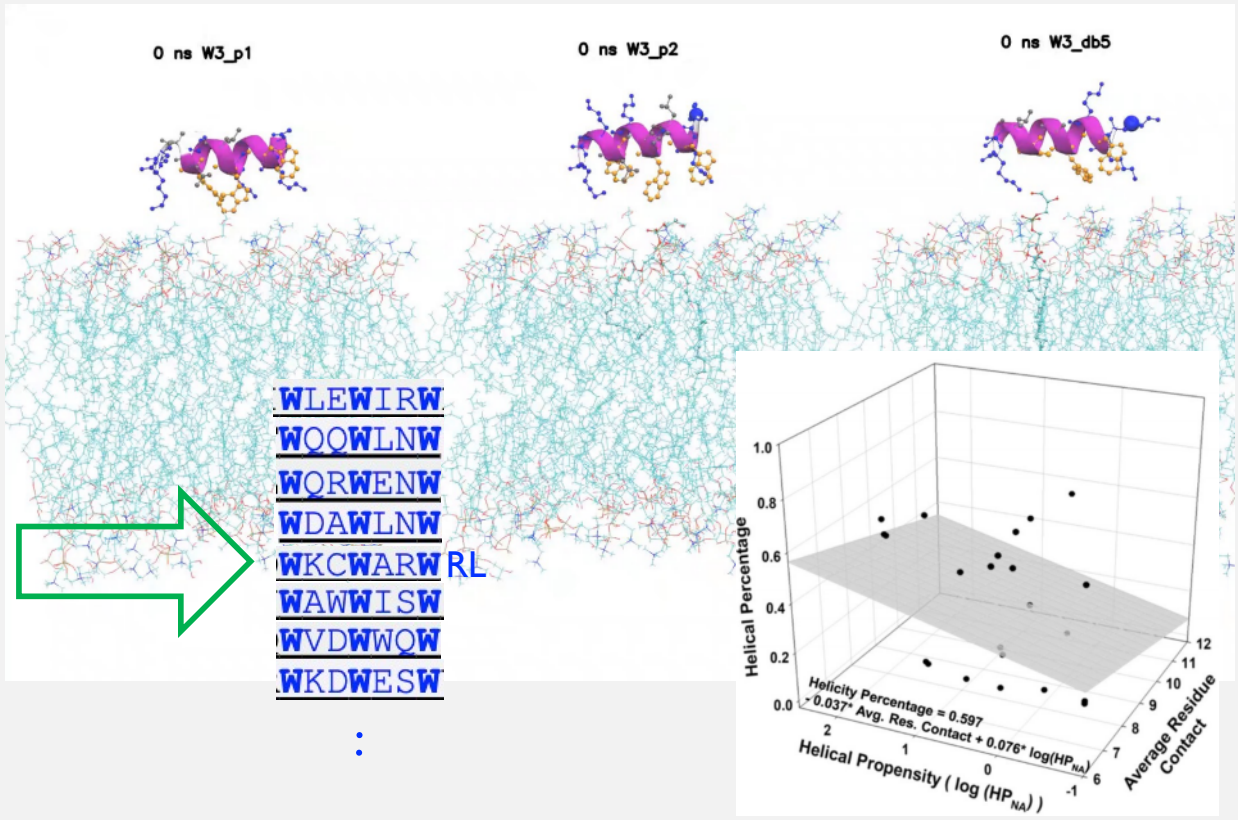
W 2 W 2 W

Examples:

Email Address (Optional)

Email Address

Submit



#	U#	Matched Sequence (MS)	Matched Pattern	Full Helix (FH)	PDB ID: Chain	Positions in PDB (MS), (FH) & File Download	Helical Propensity↕	Contact↕	Interacting Partners	Helicity%↕
1	1	WLEWIRW	W2W2W	DVQQLLIWLEWIRWESD	4eba:F	(287,293) (280,296)	1.189	8.857	4eba:F (324, 338)	0.360
10	2	WQQWLNW	W2W2W	DLATFWQQWLNW	3tsu:A	(651,657) (646,657)	1.131	7.857		0.392
11	3	WQRWENW	W2W2W	VNSYLWQRWENWFWNVTLR	2lzq:A	(10,16) (5,23)	1.105	8.000		0.385
12	4	WDAWLNW	W2W2W	WDAWLNWFR	1z8h:D	(198,204) (198,206)	1.089	7.286		0.410
13	5	WKCWARW	W2W2W	NVEDWKCWARWRLIRARA	3zuk:B	(283,289) (279,296)	1.055	9.143	3zuk:B (415, 434)	0.339
15	6	WAWWISW	W2W2W	AGEWLGSWTIFYWAWWISWSPFVGMFLAR	4llh:A	(371,377) (359,387)	1.039	10.571	4llh:A (449, 479) 4llh:A (511, 525) 4llh:A (527, 545)	0.285
25	7	WVDWWQW	W2W2W	QWVDWWQWWVK	4n7k:L	(259,265) (258,268)	0.985	7.286	4n7k:M (81, 88) 4n7k:M (178, 192)	0.402
86	8	WKDWESW	W2W2W	SVLRKALHDSLHDCSHWFYTRWKDWESWYSQSF	3zrh:A	(481,487) (460,492)	0.799	8.857	3zrh:A (522, 533)	0.330
87	9	WTAWSTW	W2W2W	WTAWSTWKYC	3cb7:A	(106,112) (106,115)	0.638	7.714		0.360
88	10	WPEWWNW	W2W2W	GWPEWWNWWLE	3wmm:L	(268,274) (267,277)	0.345	8.857	3wmm:M (82, 90) 3wmm:M (180, 193)	0.296
89	11	WPEWWGW	W2W2W	WPEWWGWWL	7prc:L	(259,265) (259,267)	-0.020	8.429		0.284

*At a minimum, 7–8 amino acids are needed for isolated helices due to hydrogen bond formation

***Helicity%** for matched sequence in this Table ranges from 0 to 1.

*The 3D visualization is empowered by 3Dmol (Rego and Koes, 2015, Bioinformatics, 31, 1322–1324)

W3-P2

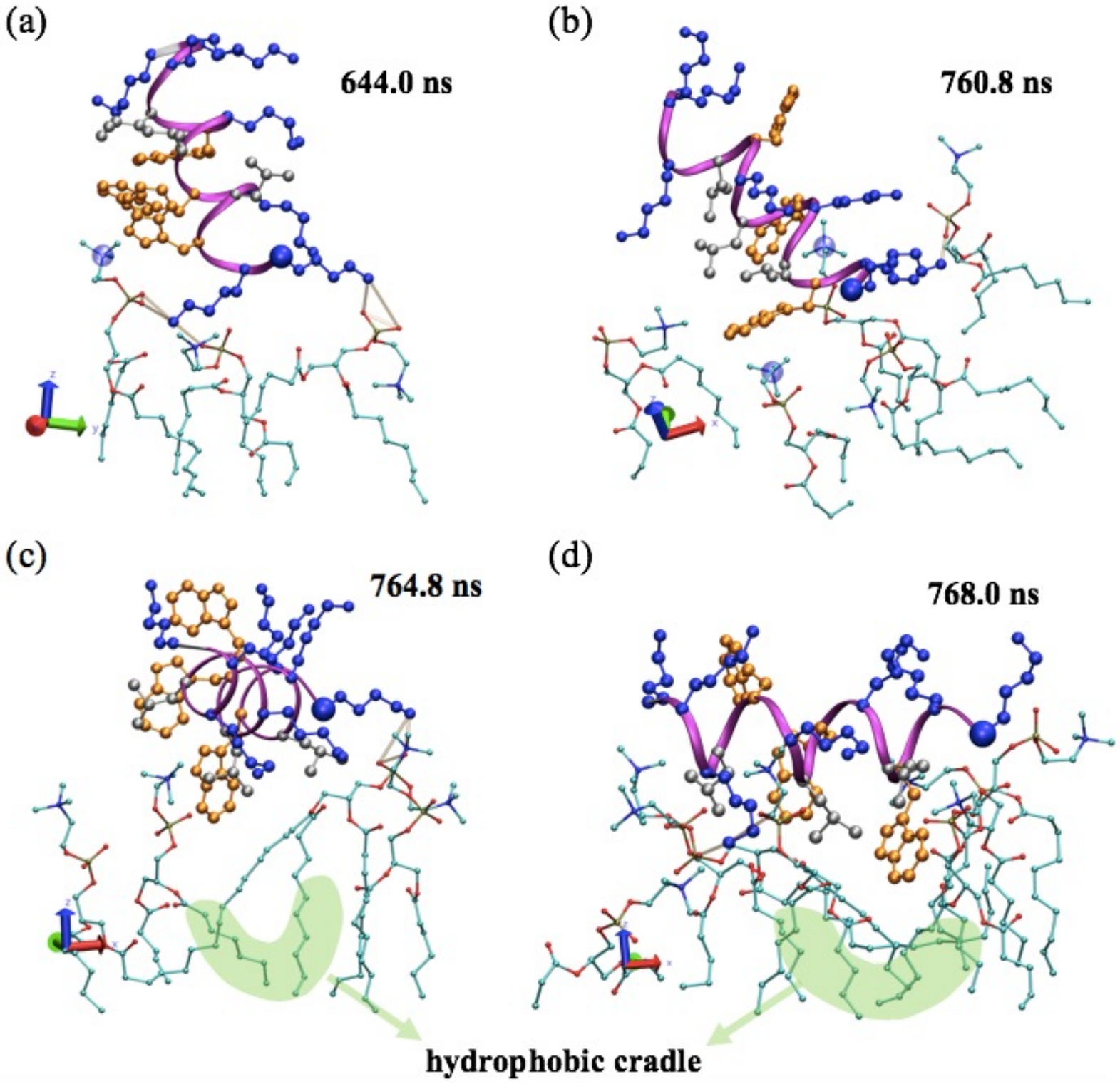
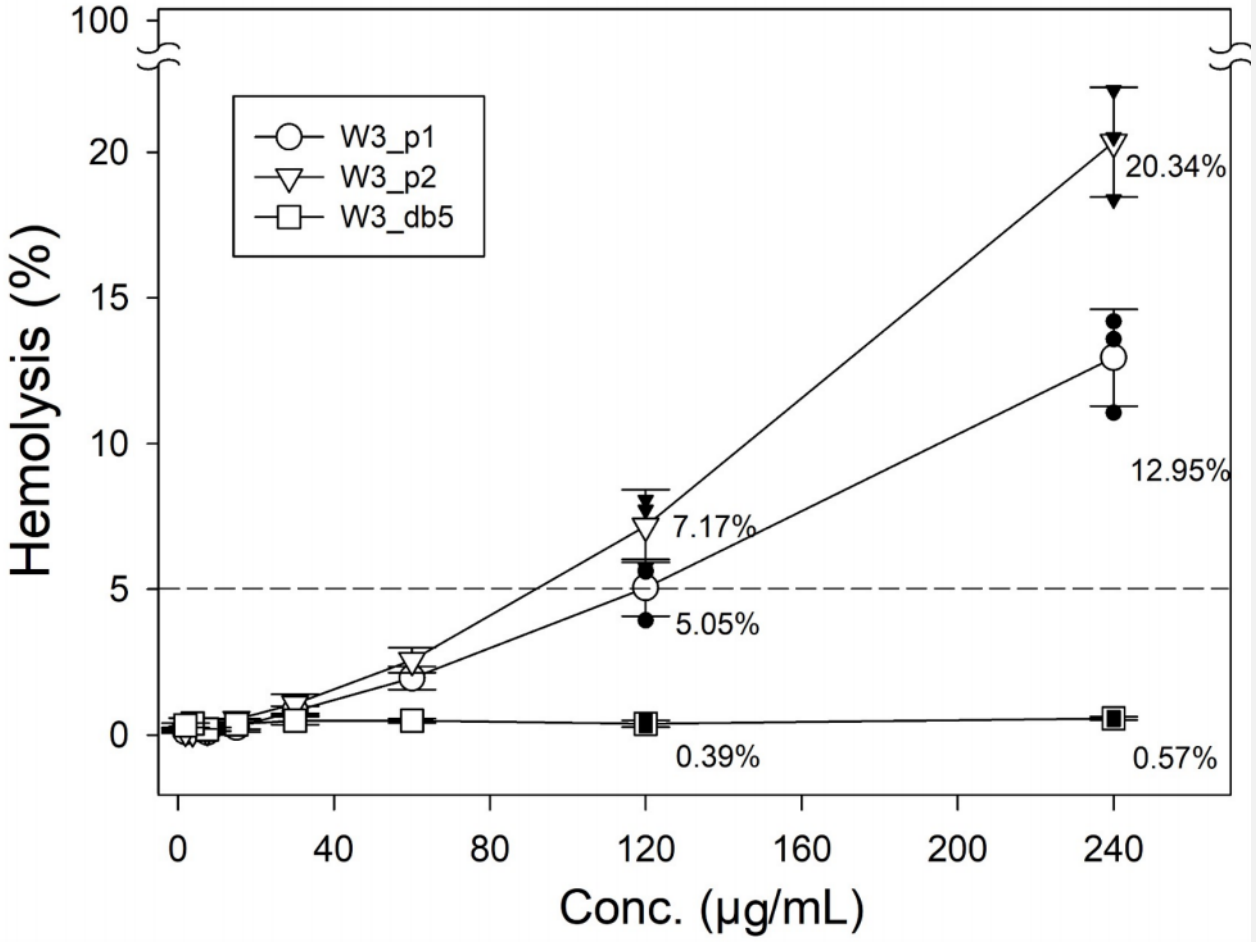
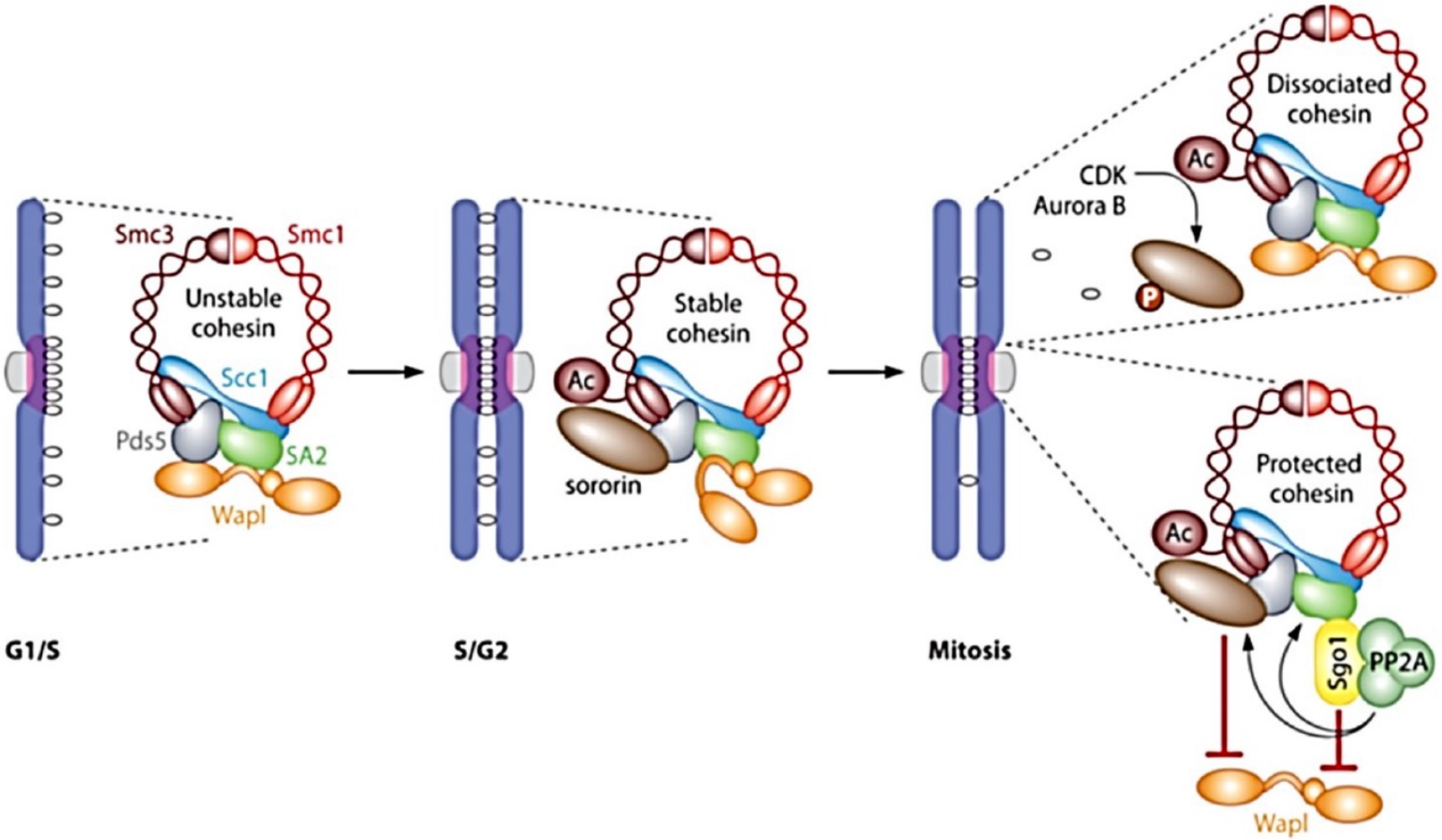


Table 3. The anti-fungal, antibacterial activity and hemolytic toxicity of TP-DB-designed AMP (W3_db5) and other AMPs containing the "W2W2W" motif.

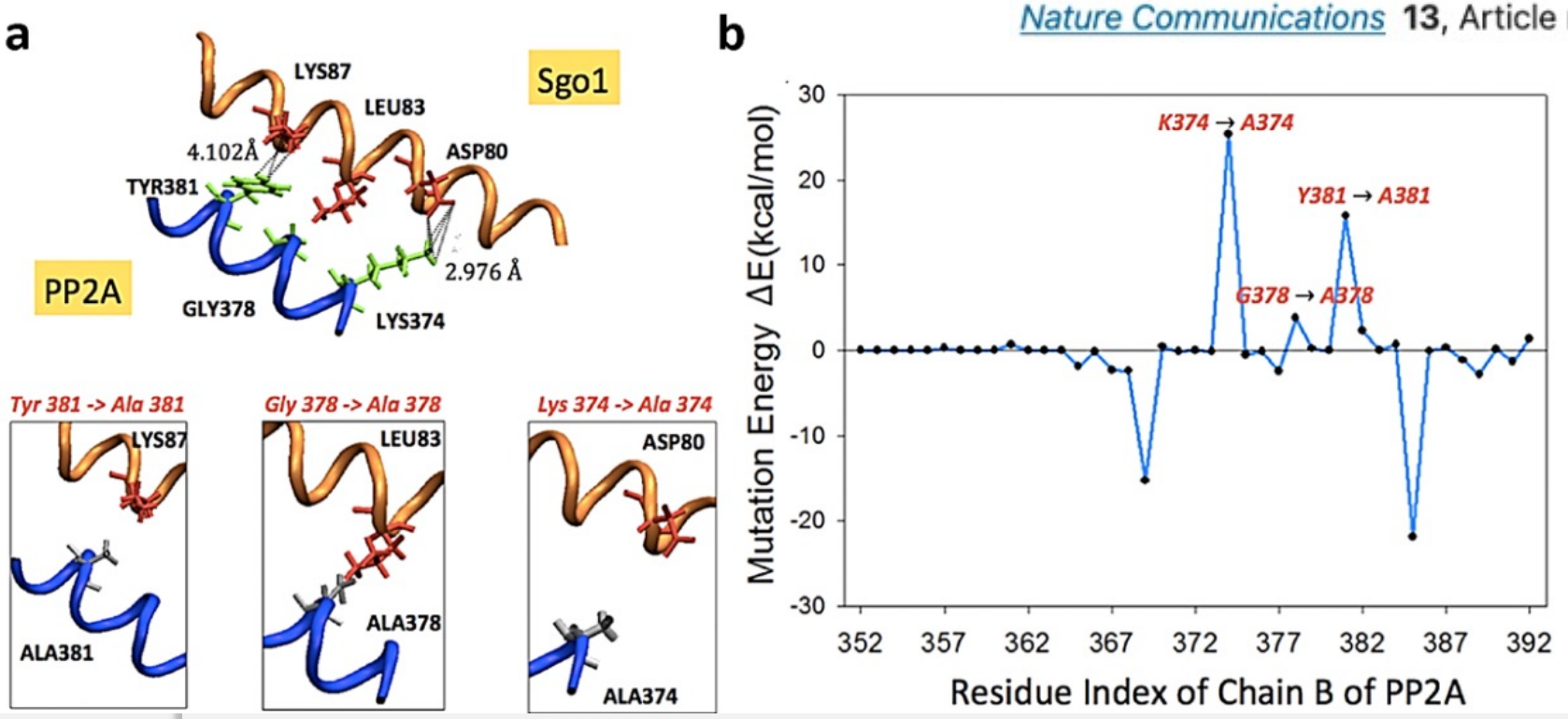
Name	Sequence	MIC$_{90}$ vs. *C. albicans* (μg/mL)	MIC$_{90}$ vs. *E. coli* (μg/mL)	MHC$_5$ (μg/mL)
W3_p1	KK **W**RK **W**LK **W**LA KK	7.5	15	120
W3_p2	KK **W**LK **W**LK **W**LK KK	7.5	15	90
W3_db5	KK **W**KC **W**AR **W**RL KK	3.75	30	>> 240
W3_n1	KK **W**GN **W**GG **W**RL KK	N/A	>> 120	N/A
W3_n2	KK **W**KD **W**ES **W**RL KK	N/A	>> 120	N/A

Therapeutic Index
= MHC/MIC
~= 80 times better than the original template AMPs.





Supplementary Figure 4 The overview of Sgo1-PP2A protecting centromeric cohesin. During mitosis, phosphorylation of SA2 and Sororin cause cohesin to dissociate from the chromosome arm. Cdk1-mediated phospho-Sgo1 recruits PP2A at centromere to counteract phosphorylation of SA2 and Sororin and maintains centromeric cohesin to tether sister chromatids together. Image is adopted from Fig. 3 in Marston's review paper[9] where the permission (Permission for the reuse of Supplementary Figure 4.pdf) is obtained through the Copyright Clearance Center of the publisher (https://marketplace.copyright.com/rs-ui-web/mp).



Supplementary Table 1. PP2A-like patterns searched against the developed helical peptide database

Pattern	Query	# of Peptides Found	# of Unique Peptides Found
K *** G ** Y	K 3 G 2 Y	398	69
A *** K *** G	A 3 K 3 G	2981	391
G ** Y *** A	G 2 Y 3 A	2079	225
A *** K *** G ** Y	A 3 K 3 G 2 Y	15	3
K *** G ** Y *** A	K 3 G 2 Y 3 A	3	2
A *** K *** G ** Y *** A	A 3 K 3 G 2 Y 3 A	0	0
A**/***K**/***G**/***Y**/***A	A 2,3 K 2,3 G 2,3 Y 2,3 A	34	3

W3-P2: KKWLKWLKWLKKK

W3-201: KWLKWLKWLKKKK

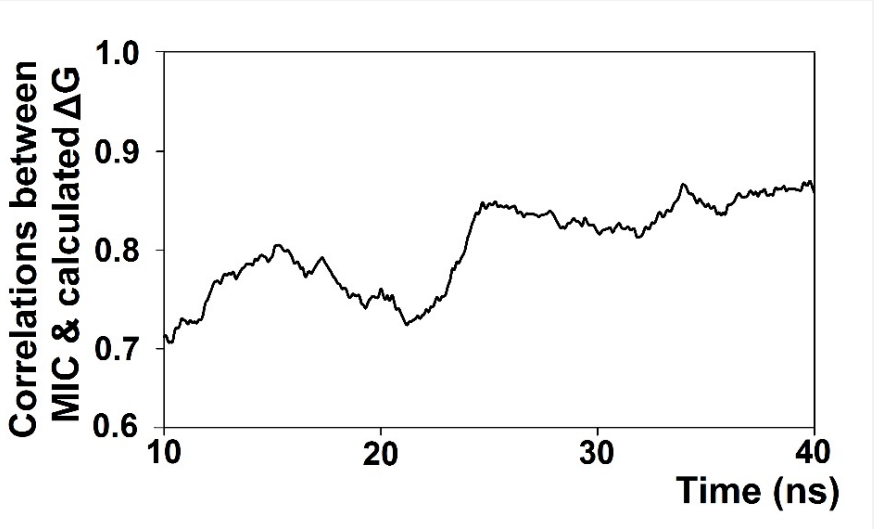
W3-202: WLKWLKWLKKKKK

W3-203: LKWLKWLKKKKKW

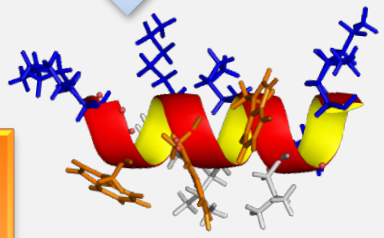
⋮

W3-212: KKKWLKWLKWLKK

Step-wise circular shuffling of antimicrobial peptides

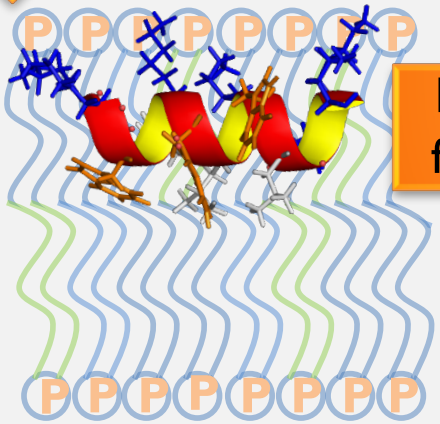


Peptides are made into helical shape and equilibrated in a water box



weak force

Restrain the peptide at -6 Å below P for 5 – 10 ns **(Sink)**

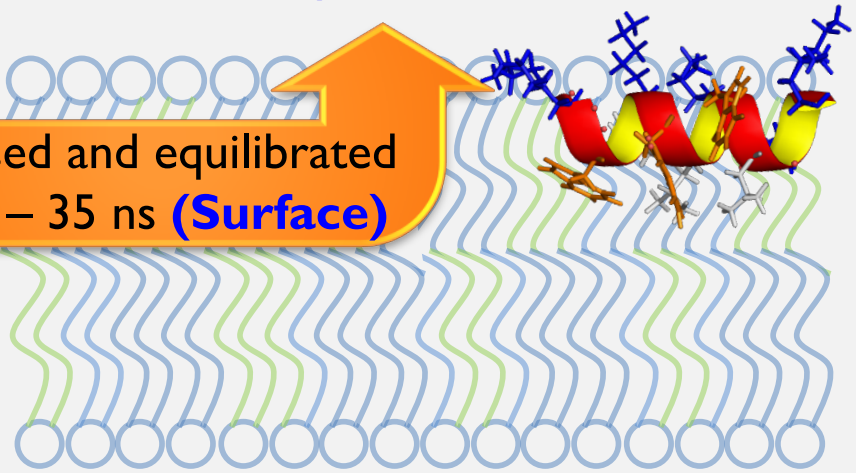


PC:PG = 3:1

(US Patent: US10810329B2)

Calculate ΔG_p = $-k_B T \ln(\langle N_i \rangle / \langle N_o \rangle)$

Released and equilibrated for 25 – 35 ns **(Surface)**



US, Taiwan & China Patent : https://patentscope.wipo.int/search/en/detail.jsf?docId=US192676438

W3-P2: KKWLKWLKWLKKK

W3-201: KWLKWLKWLKKKK

W3-202: WLKWLKWLKKKKK

W3-203: LKWLKWLKKKKKW

⋮

W3-212: KKKWLKWLKWLKK

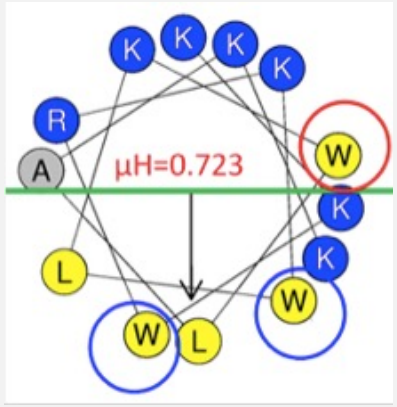
Step-wise circular shuffling of antimicrobial peptides

Calculation of the **mean amphipathic moment <μH>**

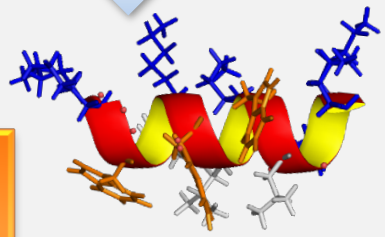
$$<\mu H>=\frac{1}{N}\left\langle\left[\sum_{n=1}^{N}H_n\sin(n\delta)\right]^2+\left[\sum_{n=1}^{N}H_n\cos(n\delta)\right]^2\right\rangle^{1/2}$$

where, δ=100° for an **alpha helix**

Nature, 1982, 299: 371-374.



Peptides are made into helical shape and equilibrated in a water box

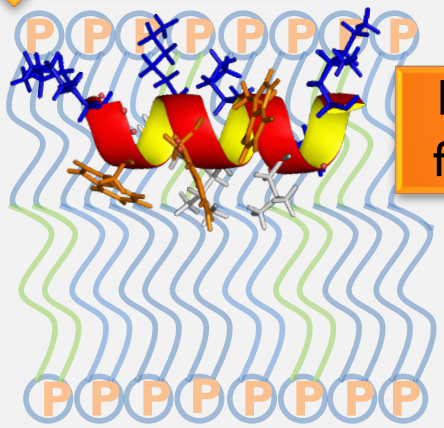


weak force

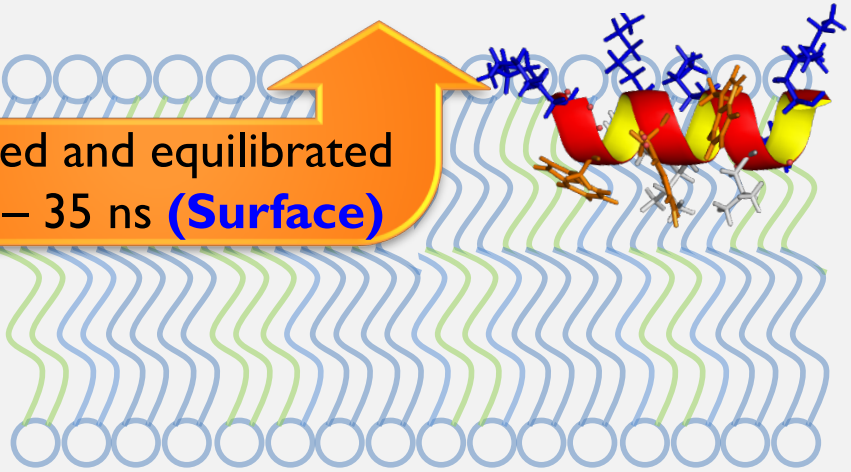
(Voon, 2021)

Calculate ensemble hydrophobic moment in PC space

Restrain the peptide at -6 Å below P for 5 – 10 ns **(Sink)**

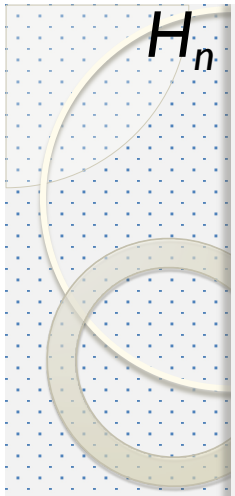


Released and equilibrated for 25 – 35 ns **(Surface)**

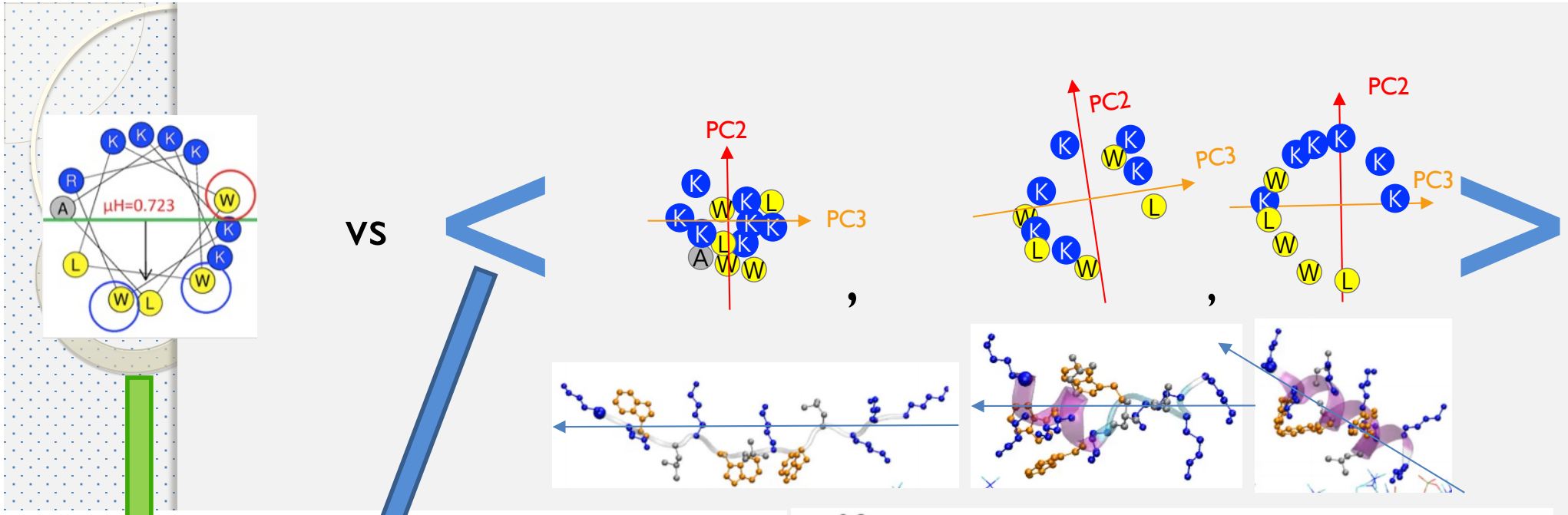


PC:PG = 3:1

US, Taiwan & China Patent : https://patentscope.wipo.int/search/en/detail.jsf?docId=US192676438



Amino Acid	Hydrophobicity
Ala	0.310
Arg	-1.010
Asn	-0.600
Asp	-0.770
Cys	1.540
Gln	-0.220
Glu	-0.640
Gly	0.000
His	0.130
Ile	1.800
Leu	1.700
Lys	-0.990
Met	1.230
Phe	1.790
Pro	0.720
Ser	-0.040
Thr	0.260
Trp	2.250
Tyr	0.960
Val	1.220



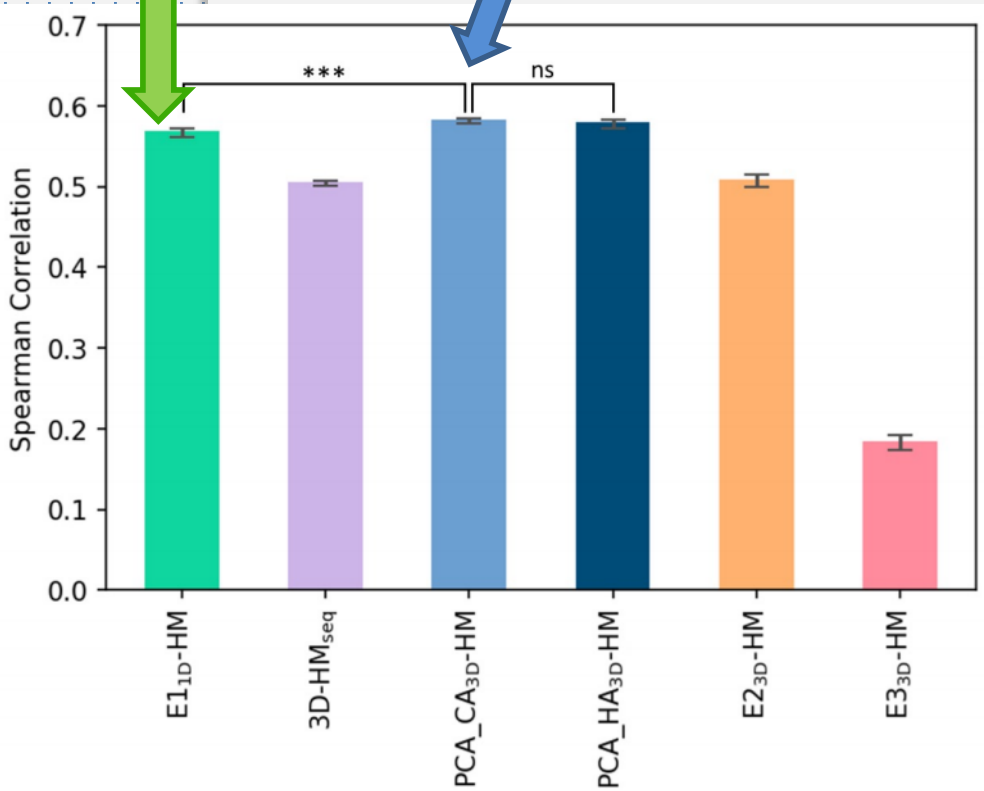


Figure 18. Spearman correlation for all 4 series of peptides at 20 to 40ns

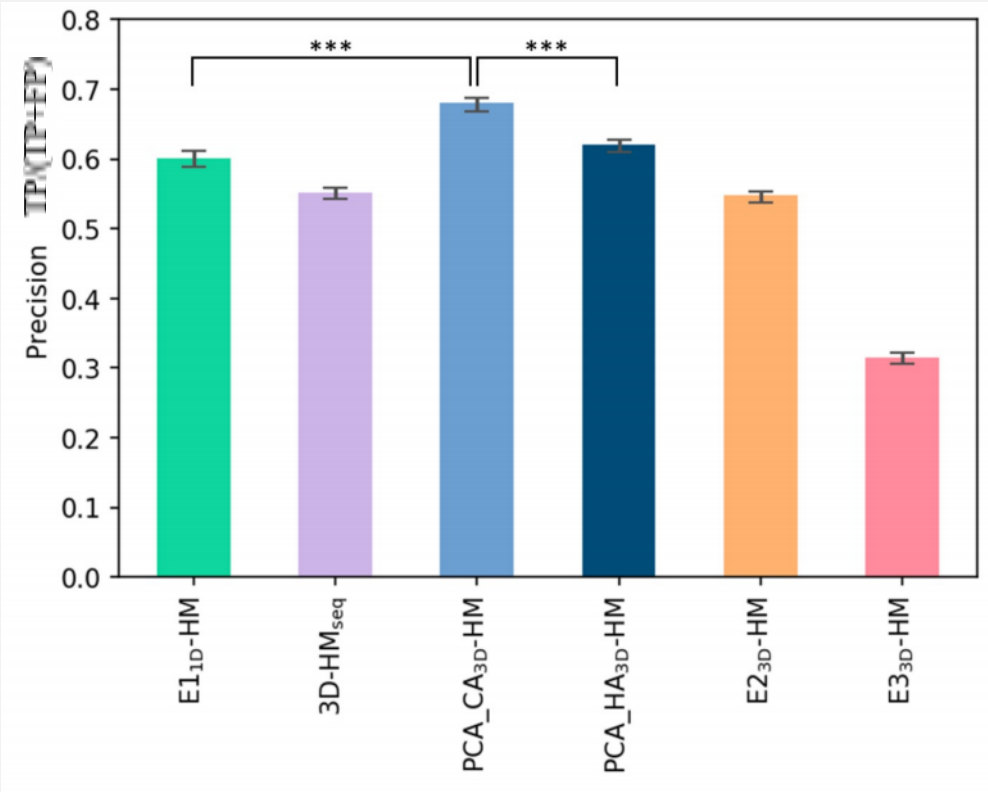


Figure 19. Precision for all 4 series of peptides at 20 to 40ns

Table 2. Circular permutants of peptide FK13

Mutants	Peptide Name	Sequence
Wild Type	FK13_W	FKRIVQRIKDFLR
Mutant 1	FK13_2	KRIVQRIKDFLRF
Mutant 2	FK13_3	RIVQRIKDFLRFK
Mutant 3	FK13_4	IVQRIKDFLRFKR
Mutant 4	FK13_5	VQRIKDFLRFKRI
Mutant 5	FK13_6	QRIKDFLRFKRIV
Mutant 6	FK13_7	RIKDFLRFKRIVQ
Mutant 7	FK13_8	IKDFLRFKRIVQR
Mutant 8	FK13_9	KDFLRFKRIVQRI
Mutant 9	FK13_10	DFLRFKRIVQRIK
Mutant 10	FK13_11	FLRFKRIVQRIKD
Mutant 11	FK13_12	LRFKRIVQRIKDF
Mutant 12	FK13_13	RFKRIVQRIKDFL

Table 3. Circular permutants of peptide R5KT8W

Mutants	Peptide Name	Sequence
Wild Type	R5KT8W_W	GLLKKIKWLL
Mutant 1	R5KT8W_2	LLKKIKWLLG
Mutant 2	R5KT8W_3	LKKIKWLLGL
Mutant 3	R5KT8W_4	KKIKWLLGLL
Mutant 4	R5KT8W_5	KIKWLLGLLK
Mutant 5	R5KT8W_6	IKWLLGLLKK
Mutant 6	R5KT8W_7	KWLLGLLKKI
Mutant 7	R5KT8W_8	WLLGLLKKIK
Mutant 8	R5KT8W_9	LLGLLKKIKW
Mutant 9	R5KT8W_10	LGLLKKIKWL

Mutants	Peptide Name	Sequence
Wild Type	CM15_W	KWKLFKKIGAVLKVL
Mutant 1	CM15_2	WKLFKKIGAVLKVLK
Mutant 2	CM15_3	KLFKKIGAVLKVLKW
Mutant 3	CM15_4	LFKKIGAVLKVLKWK
Mutant 4	CM15_5	FKKIGAVLKVLKWKL
Mutant 5	CM15_6	KKIGAVLKVLKWKLF
Mutant 6	CM15_7	KIGAVLKVLKWKLFK
Mutant 7	CM15_8	IGAVLKVLKWKLFKK
Mutant 8	CM15_9	GAVLKVLKWKLFKKI
Mutant 9	CM15_10	AVLKVLKWKLFKKIG
Mutant 10	CM15_11	VLKVLKWKLFKKIGA
Mutant 11	CM15_12	LKVLKWKLFKKIGAV
Mutant 12	CM15_13	KVLKWKLFKKIGAVL
Mutant 13	CM15_14	VLKWKLFKKIGAVLK
Mutant 14	CM15_15	LKWKLFKKIGAVLKV

Collaborators
Prof Yunching Becky Chen,
Prof Vincent Shu,
Prof Chin Yu,
Prof Qing Zhang,
Prof Bai-Shan Fang
Prof Hua-Wen Fu
Prof Chung-Yu Lan
Prof Zhong-Hong Lin
Praexisio Inc

Former and Current Lab members
Dr Hongchun Li
Dr Emmanuel O Salawu
Dr Kun-Lin Tsai
Dr Anny Cheng
Dr Bang-Chieh Huang
Dr Yuan-Yu Chang
Cheng-Yu Tsai
Christopher L Ortiz
Liyin Voon
Kai-Di Hu
Sobha Sahoo
Lutimba Stuart

Computing resources
National Center High Performance Computing (NCHC), Taiwan
High Performance Computing Infrastructure (HPCI), Japan

Funding ($$)
Mystery of Science and Technology (MOST)
NTHU SPROUT fund

Thank you.