how artificial intelligence is reshaping protein structure prediction and therapeutic design

Jens Meiler

Humboldt Professor and Director, Institute for Drug Discovery, University Leipzig, Leipzig, Saxony, Germany

Distinguished Professor, Department of Chemistry and Pharmacology, Vanderbilt University, Nashville, TN, USA

UNIVERSITÄT LEIPZIG

and

VANDERBILT UNIVERSITY

are partner universities: www.leipzig.vanderbilt.edu

Nobelpreis Chemie 2024

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker “for computational protein design”, the other half jointly to Demis Hassabis and John M. Jumper “for protein structure prediction”.

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

Prize share: 1/4

Ill. Niklas Elmehed © Nobel Prize Outreach

John M. Jumper

Prize share: 1/4

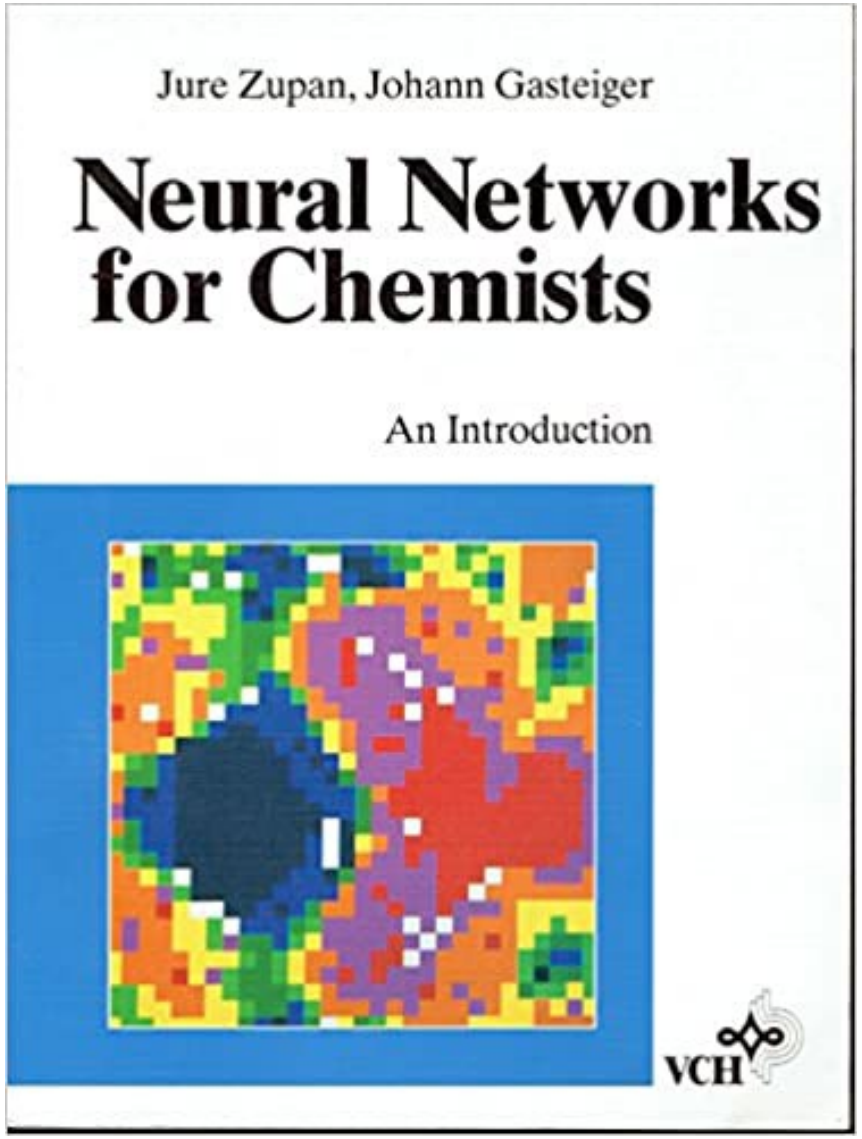
Nobelpreis Chemie 2024

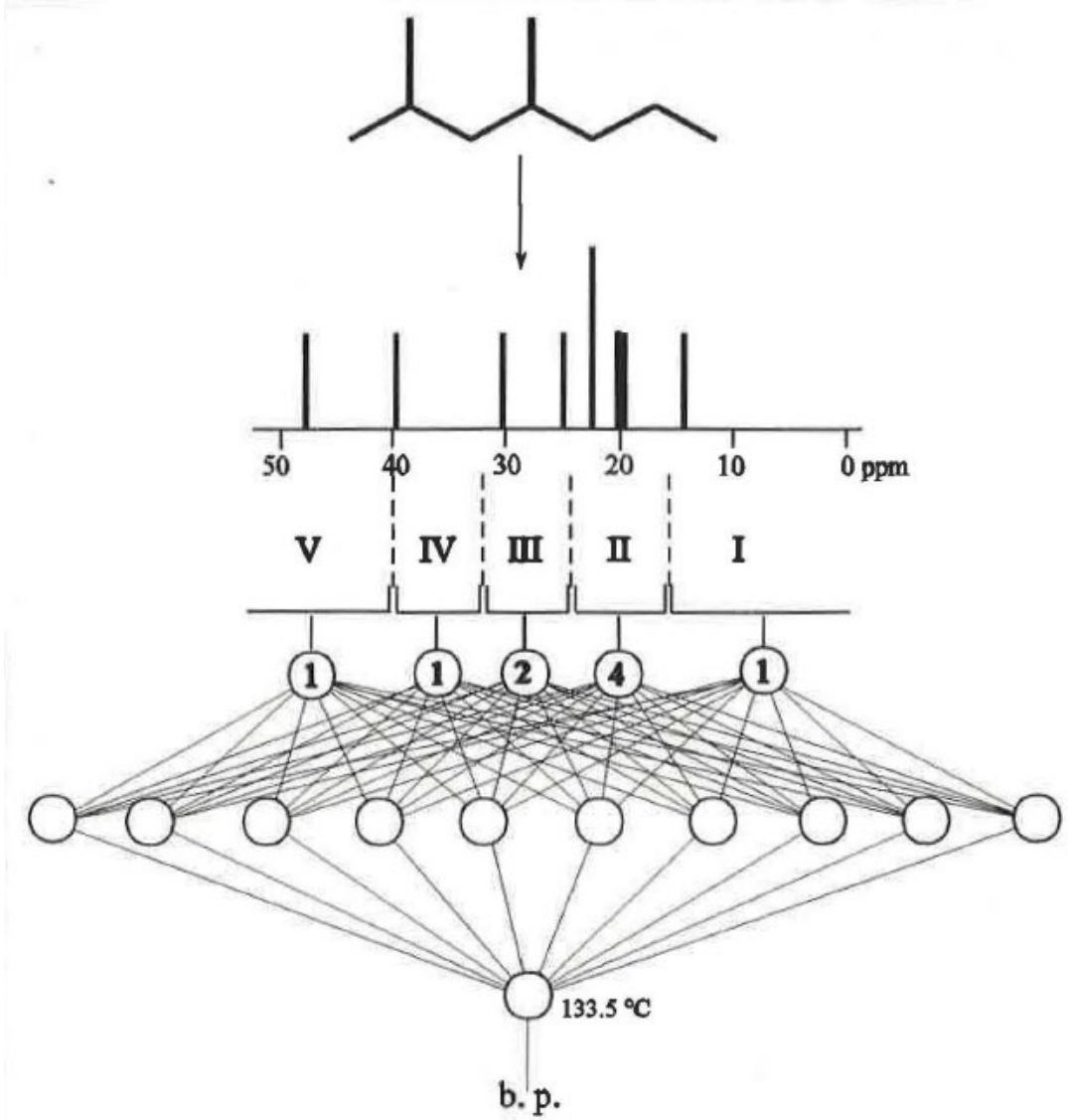
The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker “for computational protein design”, the other half jointly to Demis Hassabis and John M. Jumper “for protein structure prediction”.

Nobelpreis Chemie 2024

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker “for computational protein design”, the other half jointly to Demis Hassabis and John M. Jumper “for protein structure prediction”.

Since 1993(!) – Neural Networks for Chemists

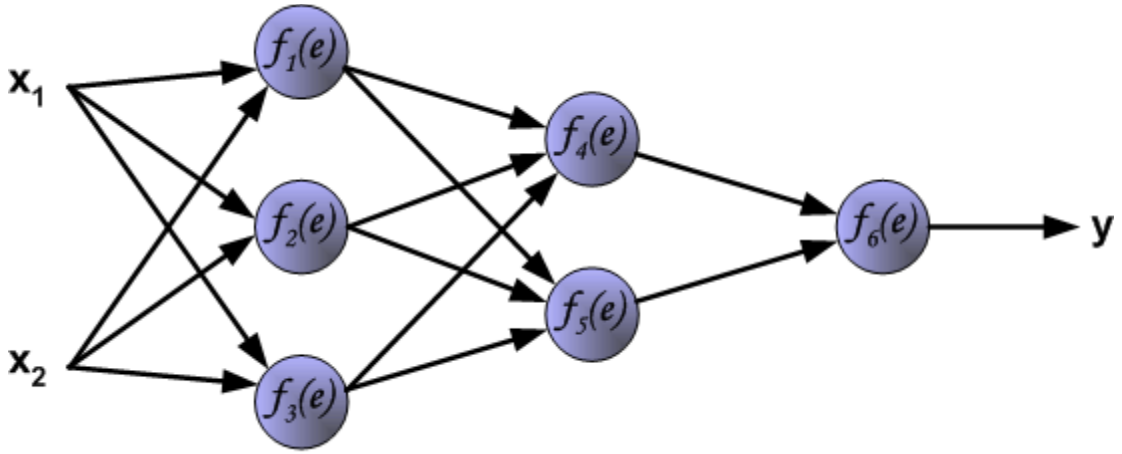




J. Meiler and R. Meusinger "Use of Neural Networks to Determine Properties of Alkanes from their 13C-NMR Spectra" in *Software - Entwicklung in der Chemie*; Gasteiger, J., Ed. Gesellschaft Deutscher Chemiker: Frankfurt am Main; **1995**; Vol. 10: p. 259-263.

Back Propagation of Errors is the most popular Training Algorithm

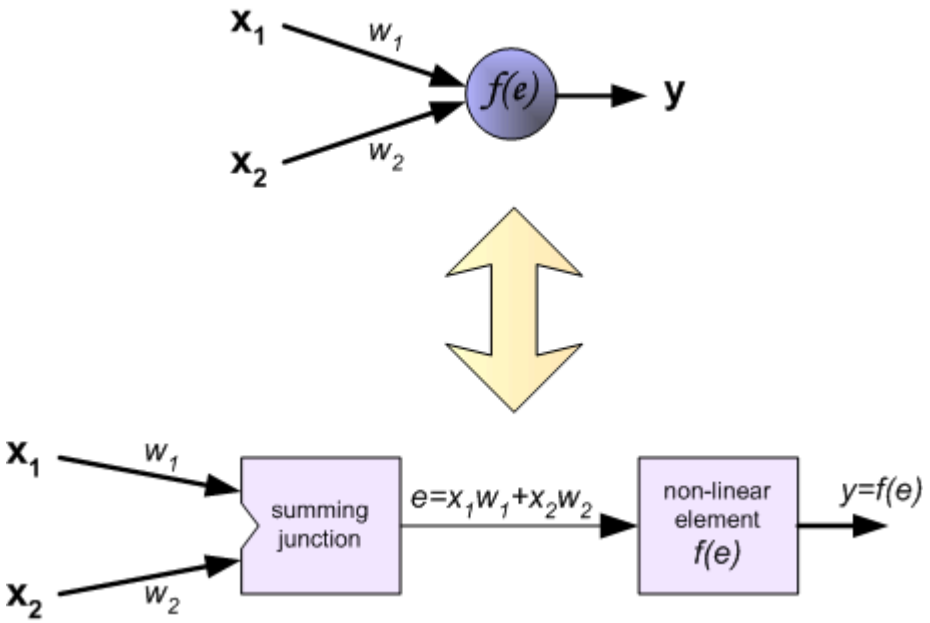
- Teaching process of multi-layer neural network employing *backpropagation* algorithm. To illustrate this process, consider the three layer neural network with two inputs and one output:



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

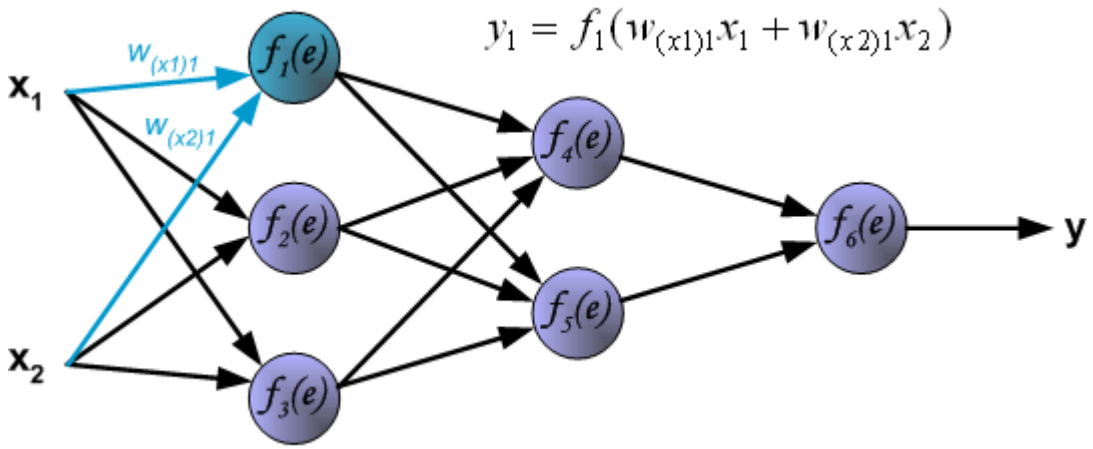
- Each neuron is composed of two units. First unit adds products of weights coefficients and input signals. The second unit realizes nonlinear function, called neuron activation function. Signal *e* is summed weighted input signal, and *y = f(e)* is output signal of nonlinear element. Signal *y* is also output signal of neuron:



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

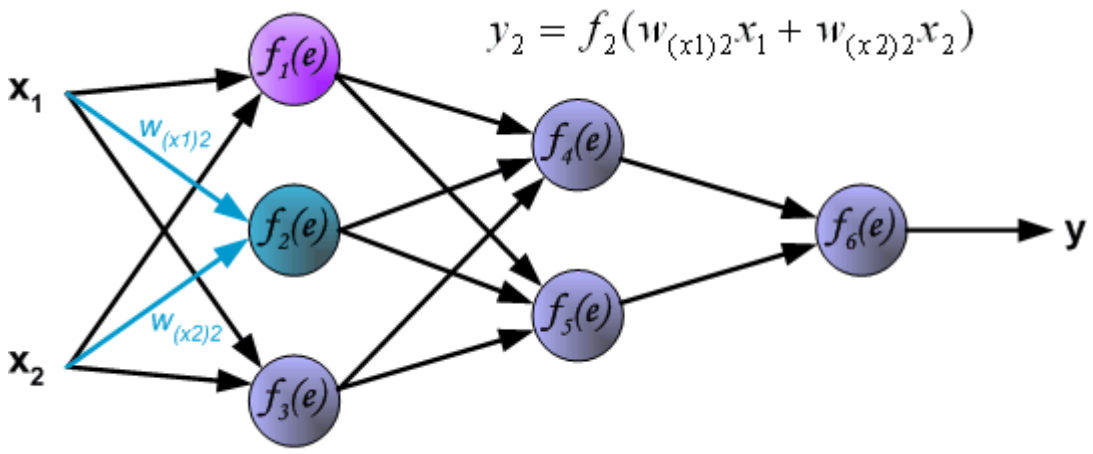
- To teach the neural network we need training data set. The training data set consists of input signals (x_1 and x_2) assigned with corresponding target (desired output) z. The network training is an iterative process. In each iteration weights coefficients of nodes are modified using new data from training data set. Modification is calculated using algorithm described below: Each teaching step starts with forcing both input signals from training set. After this stage we can determine output signals values for each neuron in each network layer. Pictures below illustrate how signal is propagating through the network, Symbols $w_{(xm)n}$ represent weights of connections between network input x_m and neuron n in input layer. Symbols y_n represents output signal of neuron n.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

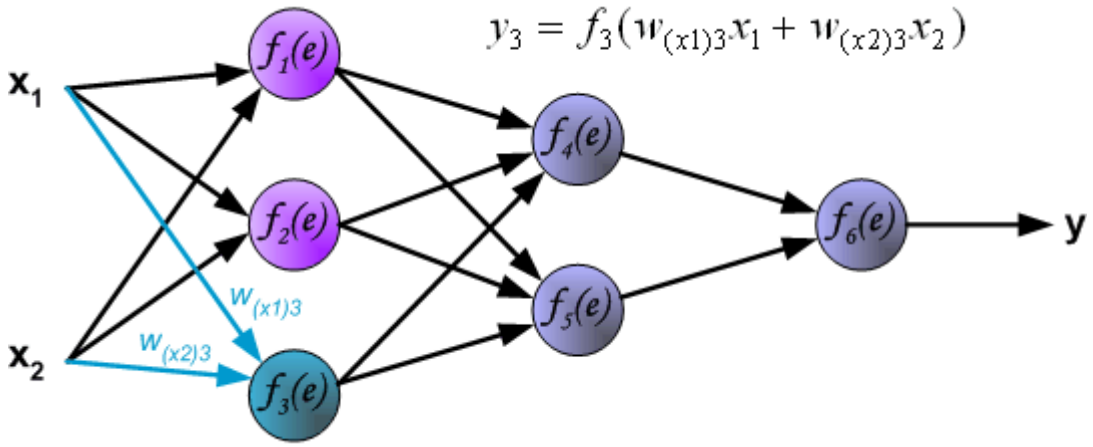
- To teach the neural network we need training data set. The training data set consists of input signals (x_1 and x_2) assigned with corresponding target (desired output) z. The network training is an iterative process. In each iteration weights coefficients of nodes are modified using new data from training data set. Modification is calculated using algorithm described below: Each teaching step starts with forcing both input signals from training set. After this stage we can determine output signals values for each neuron in each network layer. Pictures below illustrate how signal is propagating through the network, Symbols $w_{(xm)n}$ represent weights of connections between network input x_m and neuron n in input layer. Symbols y_n represents output signal of neuron n.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

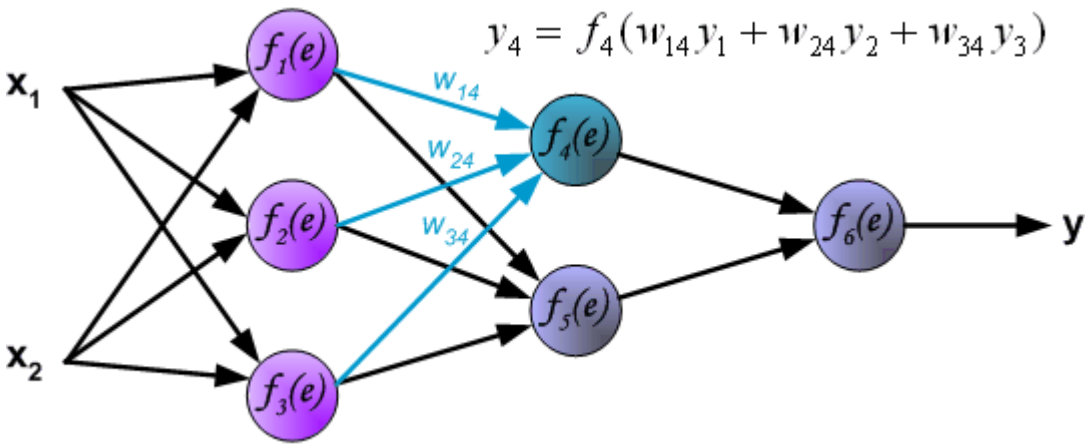
- To teach the neural network we need training data set. The training data set consists of input signals (x_1 and x_2) assigned with corresponding target (desired output) z. The network training is an iterative process. In each iteration weights coefficients of nodes are modified using new data from training data set. Modification is calculated using algorithm described below: Each teaching step starts with forcing both input signals from training set. After this stage we can determine output signals values for each neuron in each network layer. Pictures below illustrate how signal is propagating through the network, Symbols $w_{(xm)n}$ represent weights of connections between network input x_m and neuron n in input layer. Symbols y_n represents output signal of neuron n.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

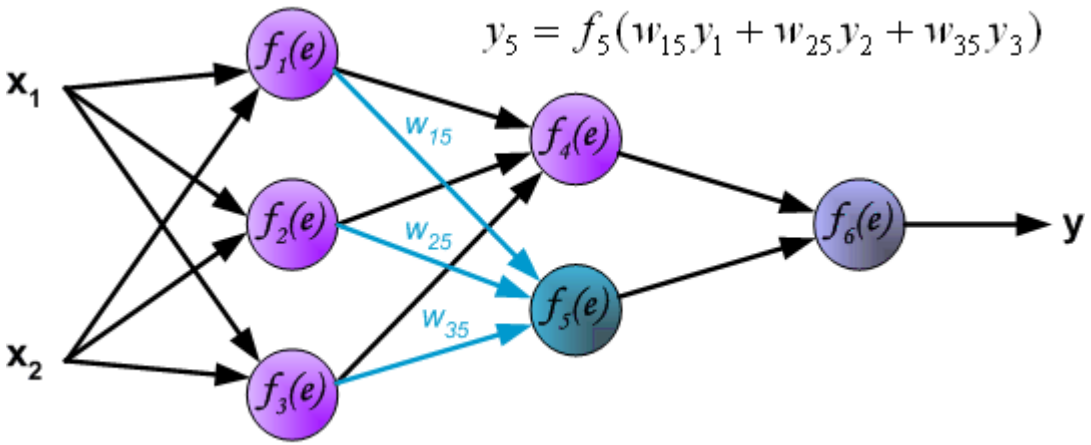
- Propagation of signals through the hidden layer. Symbols w_{mn} represent weights of connections between output of neuron *m* and input of neuron *n* in the next layer.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

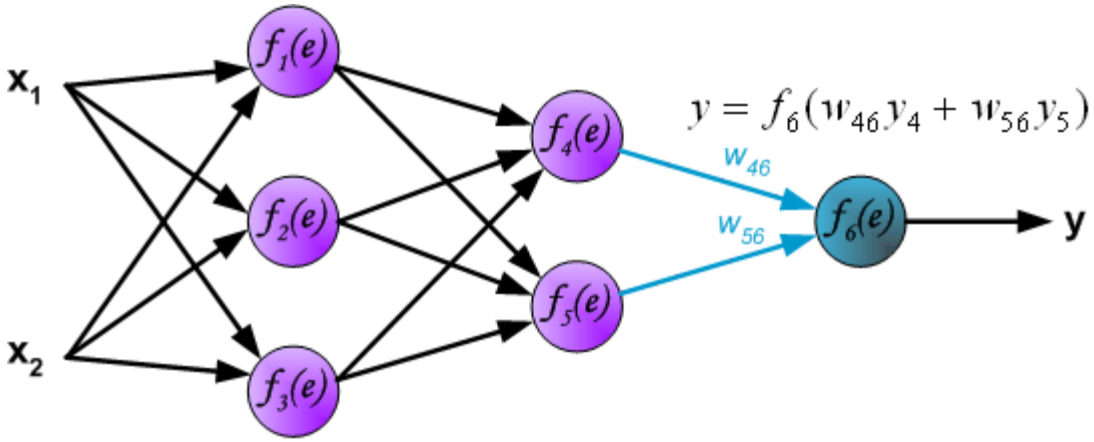
- Propagation of signals through the hidden layer. Symbols w_{mn} represent weights of connections between output of neuron *m* and input of neuron *n* in the next layer.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

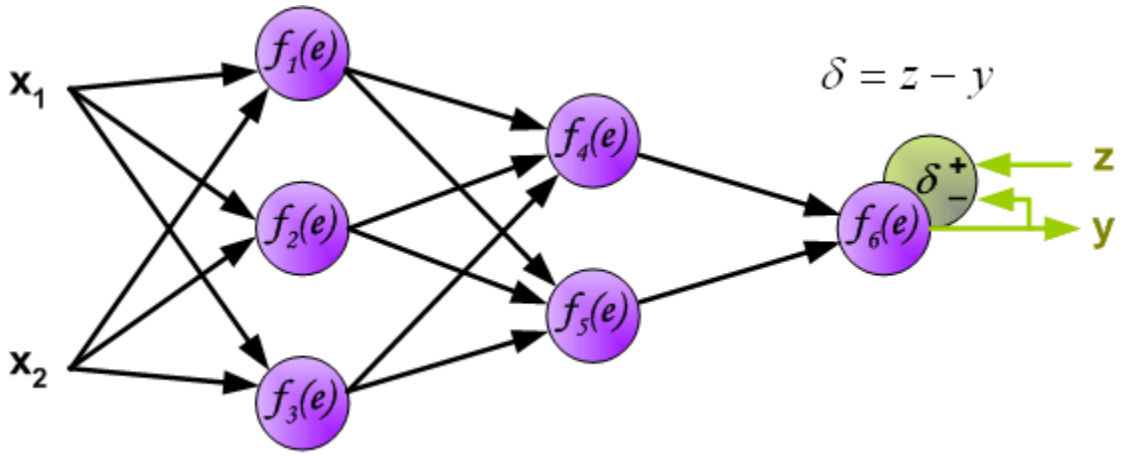
- Propagation of signals through the output layer.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

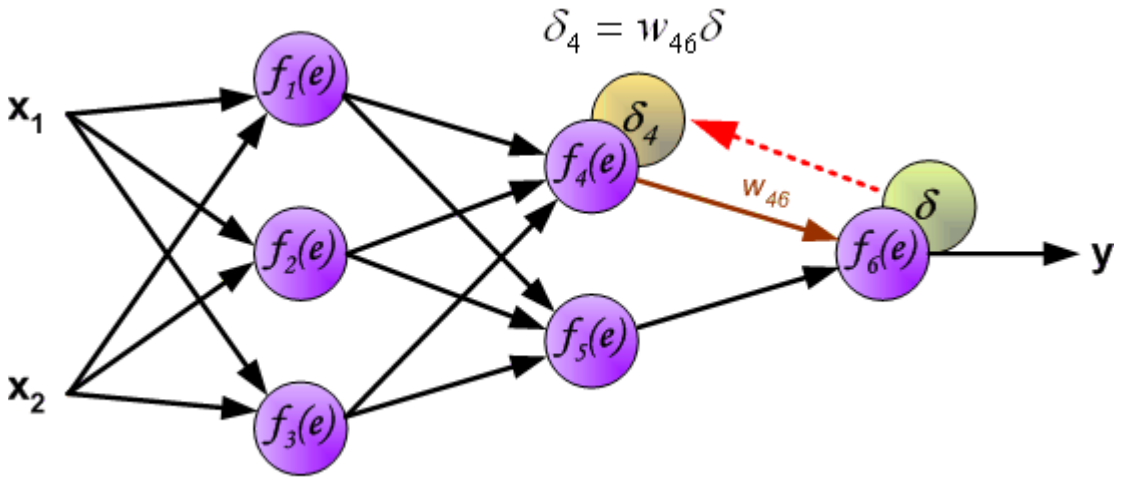
- In the next algorithm step the output signal of the network *y* is compared with the desired output value (the target), which is found in training data set. The difference is called error signal *d* of output layer neuron.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

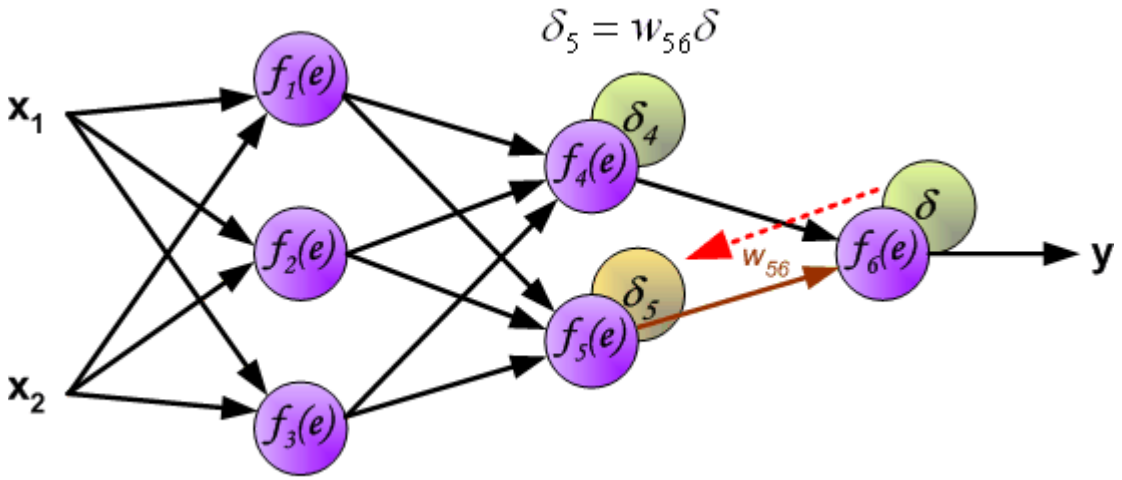
- It is impossible to compute error signal for internal neurons directly, because output values of these neurons are unknown. The idea is to propagate error signal *d* (computed in single teaching step) back to all neurons, which output signals were input for discussed neuron.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

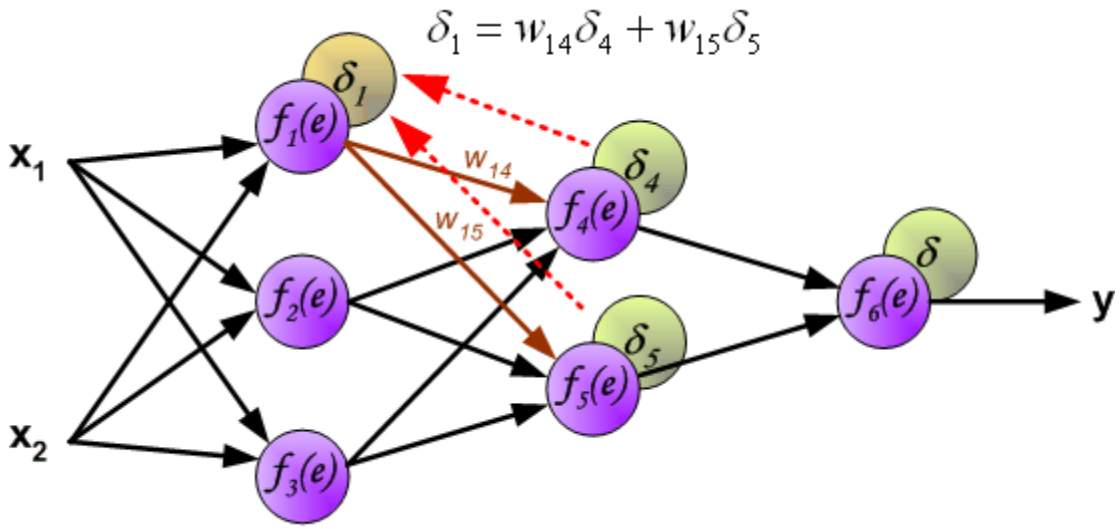
- It is impossible to compute error signal for internal neurons directly, because output values of these neurons are unknown. The idea is to propagate error signal *d* (computed in single teaching step) back to all neurons, which output signals were input for discussed neuron.



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

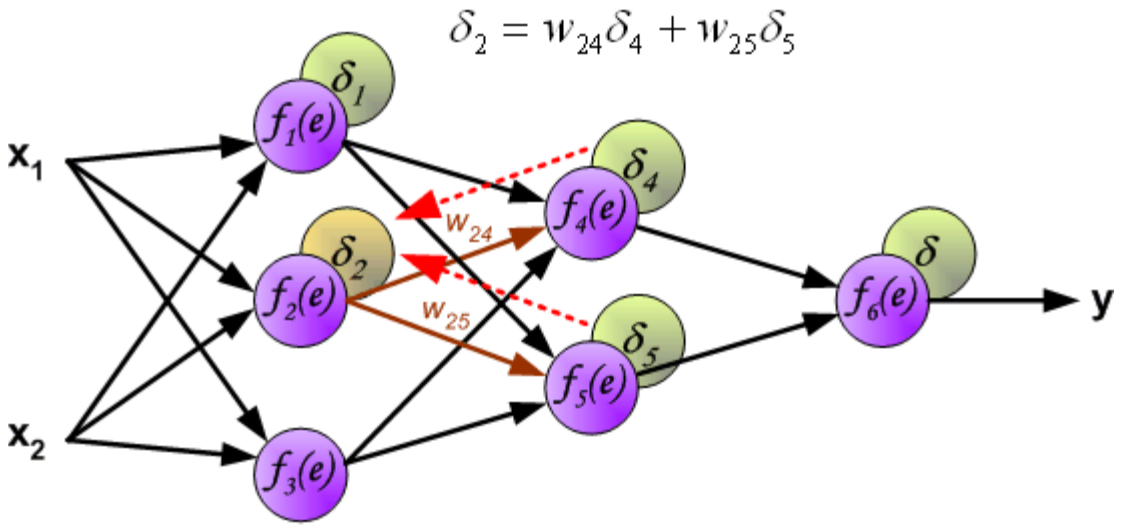
- The weights' coefficients w_{mn} used to propagate errors back are equal to this used during computing output value. Only the direction of data flow is changed - signals are propagated from output to inputs one after the other. This technique is used for all network layers. If propagated errors came from few neurons they are added. The illustration is below:



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

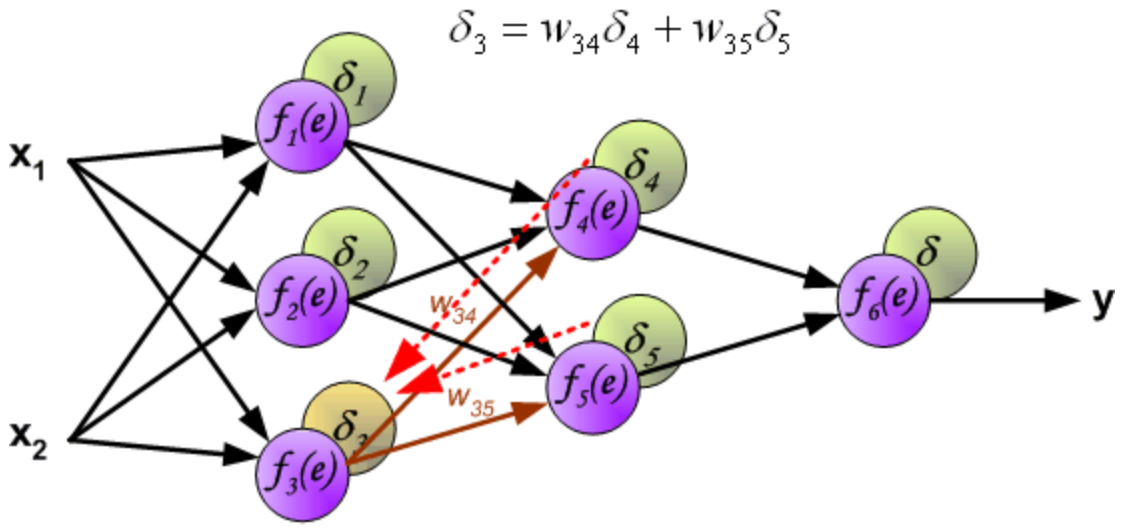
- The weights' coefficients w_{mn} used to propagate errors back are equal to this used during computing output value. Only the direction of data flow is changed - signals are propagated from output to inputs one after the other. This technique is used for all network layers. If propagated errors came from few neurons they are added. The illustration is below:



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

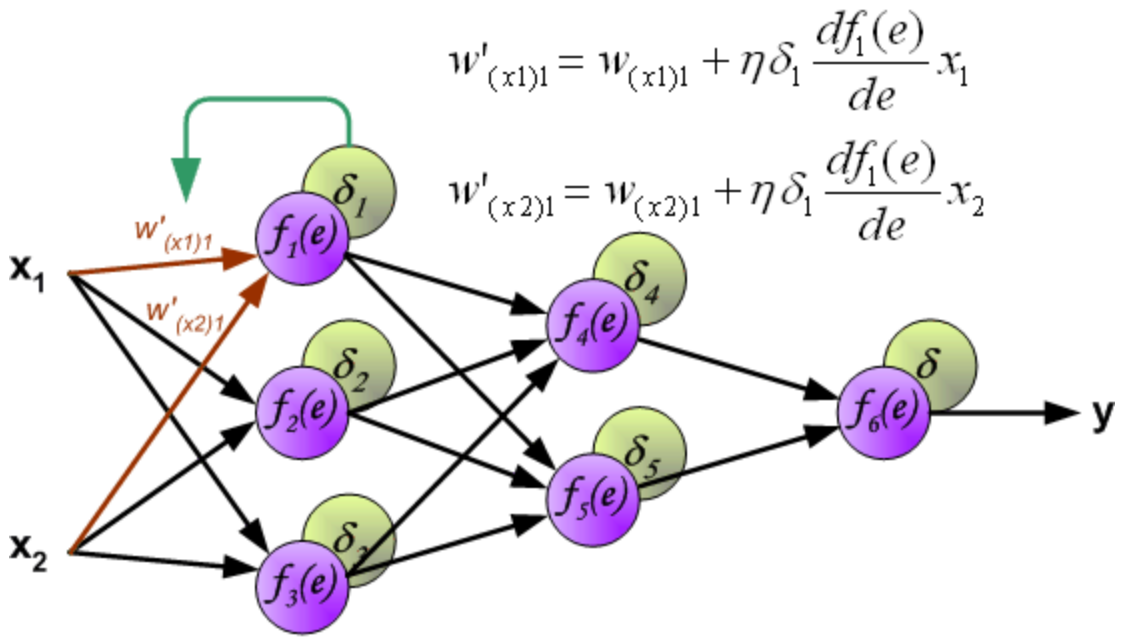
- The weights' coefficients w_{mn} used to propagate errors back are equal to this used during computing output value. Only the direction of data flow is changed - signals are propagated from output to inputs one after the other. This technique is used for all network layers. If propagated errors came from few neurons they are added. The illustration is below:



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified. In formulas below *df(e)/de* represents derivative of neuron activation function (which weights are modified).



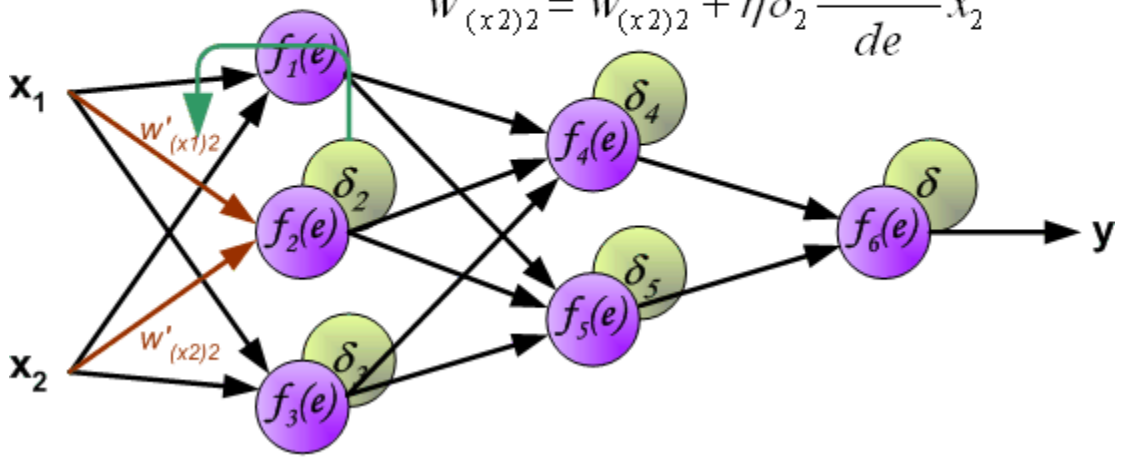
http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified. In formulas below *df(e)/de* represents derivative of neuron activation function (which weights are modified).

$$w'_{(x1)2} = w_{(x1)2} + \eta \delta_2 \frac{df_2(e)}{de} x_1$$

$$w'_{(x2)2} = w_{(x2)2} + \eta \delta_2 \frac{df_2(e)}{de} x_2$$



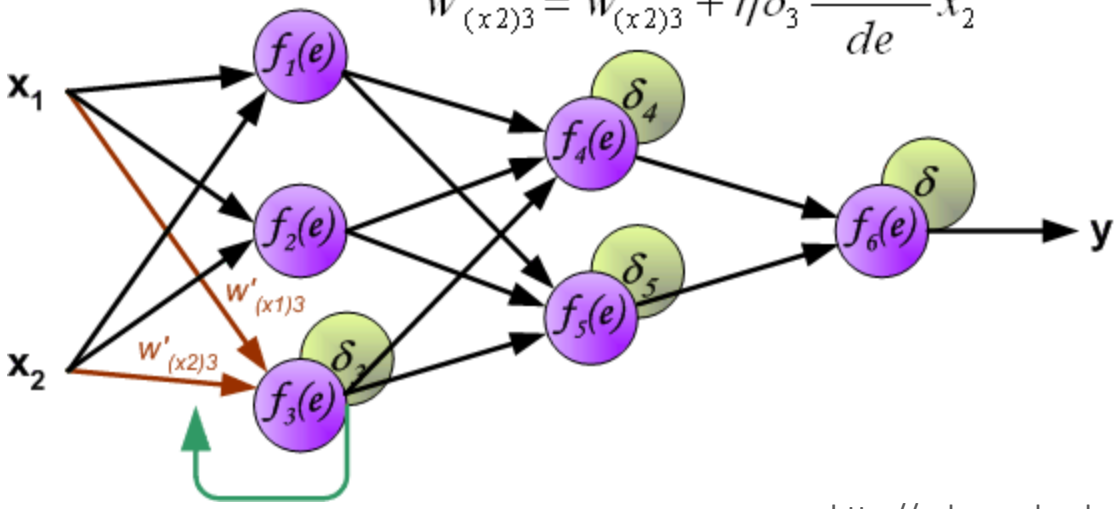
http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified. In formulas below *df(e)/de* represents derivative of neuron activation function (which weights are modified).

$$w'_{(x1)3} = w_{(x1)3} + \eta \delta_3 \frac{df_3(e)}{de} x_1$$

$$w'_{(x2)3} = w_{(x2)3} + \eta \delta_3 \frac{df_3(e)}{de} x_2$$



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

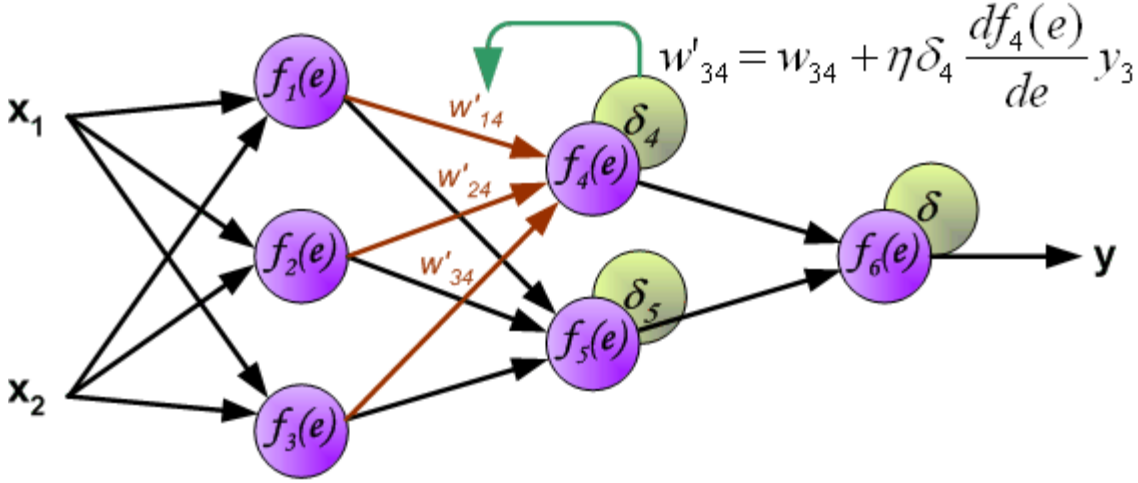
Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified.

$$w'_{14} = w_{14} + \eta \delta_4 \frac{df_4(e)}{de} y_1$$

$$w'_{24} = w_{24} + \eta \delta_4 \frac{df_4(e)}{de} y_2$$

$$w'_{34} = w_{34} + \eta \delta_4 \frac{df_4(e)}{de} y_3$$



http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

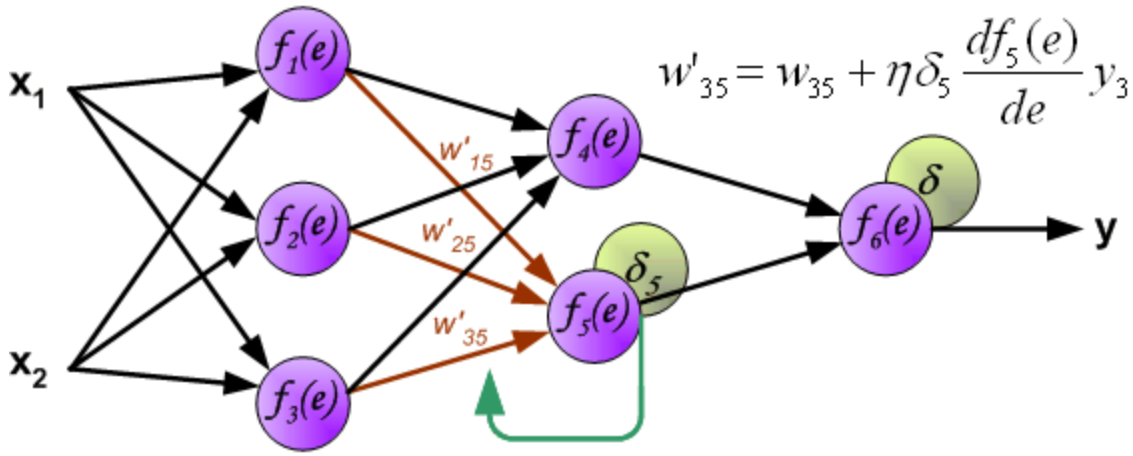
Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified.

$$w'_{15} = w_{15} + \eta \delta_5 \frac{df_5(e)}{de} y_1$$

$$w'_{25} = w_{25} + \eta \delta_5 \frac{df_5(e)}{de} y_2$$

$$w'_{35} = w_{35} + \eta \delta_5 \frac{df_5(e)}{de} y_3$$



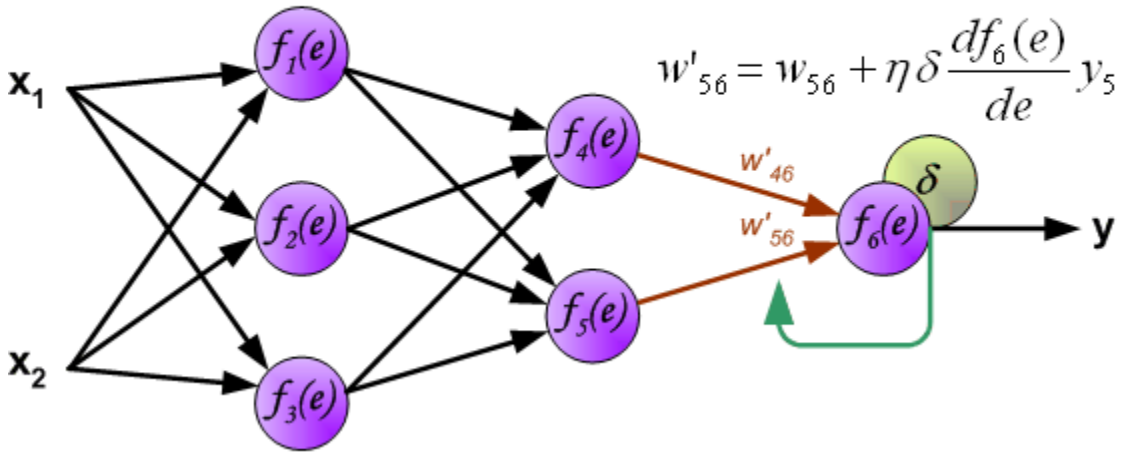
http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

Back Propagation of Errors is the most popular Training Algorithm

- When the error signal for each neuron is computed, the weights coefficients of each neuron input node may be modified. In formulas below *df(e)/de* represents derivative of neuron activation function (which weights are modified).

$$w'_{46} = w_{46} + \eta \delta \frac{df_6(e)}{de} y_4$$

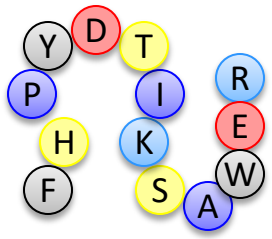
$$w'_{56} = w_{56} + \eta \delta \frac{df_6(e)}{de} y_5$$

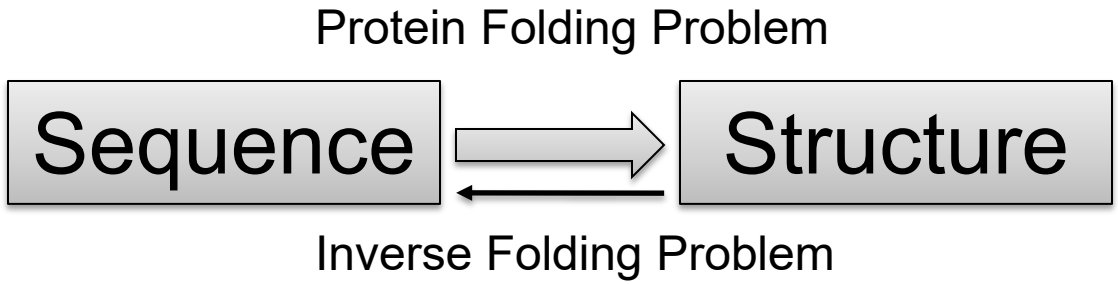


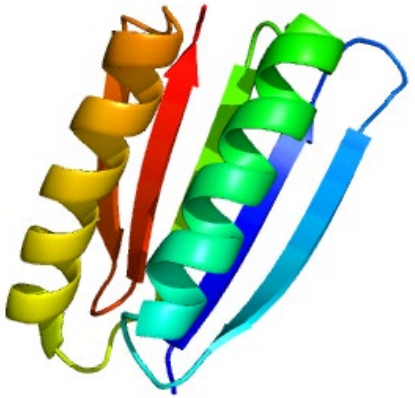
http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html

The (Inverse) Protein Folding Problem
Holy Grail of Comp. Structural Biology

meilerlab



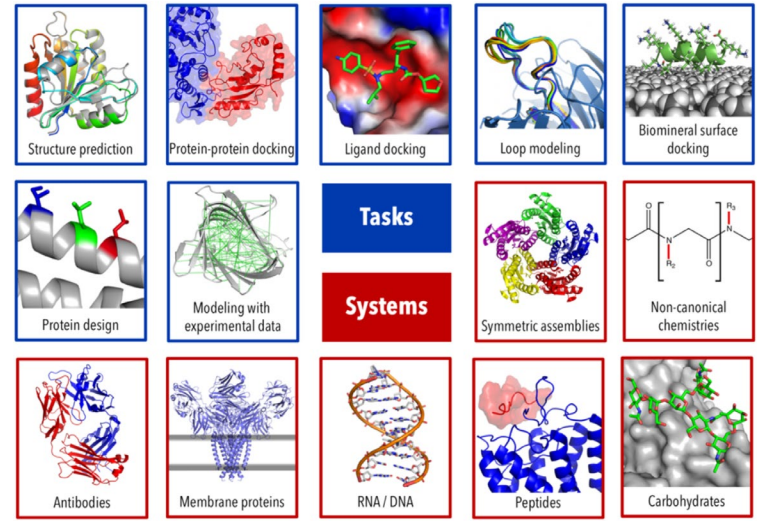


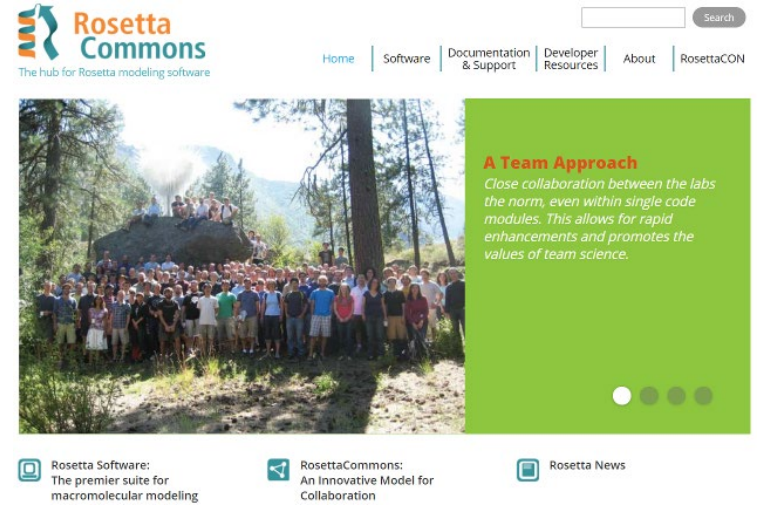


- Given a protein's AA sequence, what is its 3-dimensional fold , and how does it get there?
- Assume 100 conformations for each amino acid in a 100 amino acid protein ⇨ 10^{200} possible conformations!
- Cyrus Levinthal's paradox of protein folding,1968.

Rosetta: A Unified Framework for Protein Structure

meilerlab

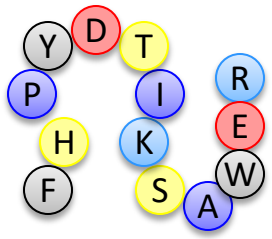




A
J. p. 665-680.

The (Inverse) Protein Folding Problem
Holy Grail of Comp. Structural Biology

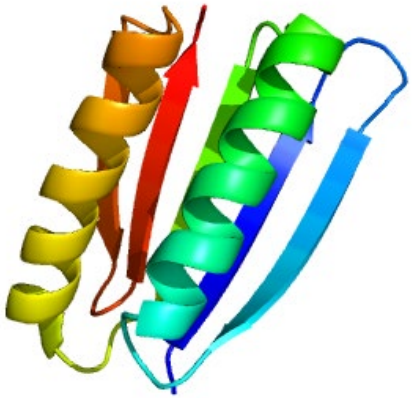
meilerlab



Protein Folding Problem

Sequence ⇌ Structure

Inverse Folding Problem

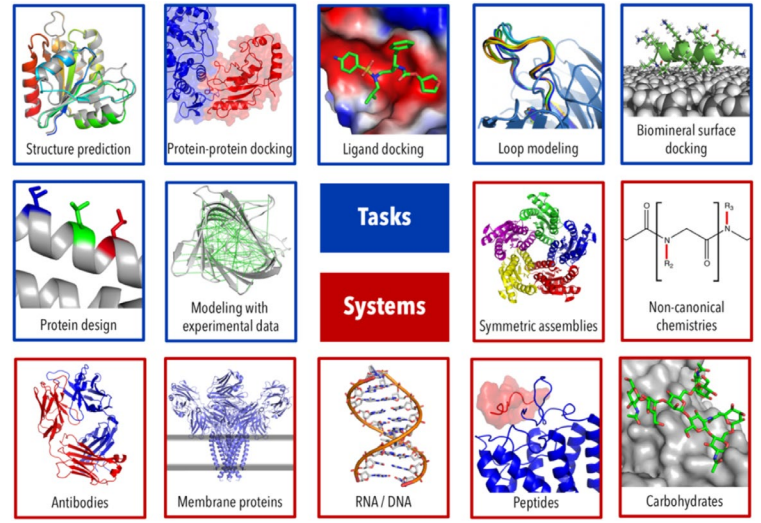


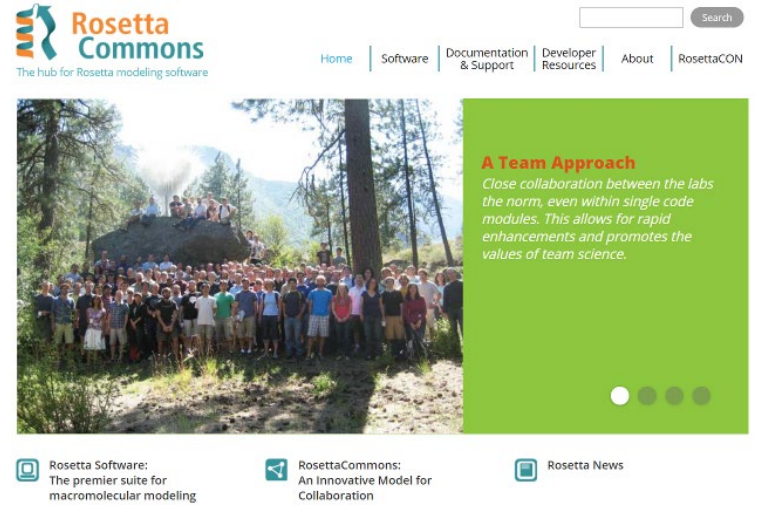
- Given a protein's AA sequence, what is its 3-dimensional fold , and how does it get there?
- Assume 100 conformations for each amino acid in a 100 amino acid protein ⇨ 10^{200} possible conformations!
- Cyrus Levinthal's paradox of protein folding,1968.

- Given a protein fold, which primary sequence(s) fold into it?
- Assume a total of 100 conformations for all 20 natural occurring amino acids side chains in a 100 amino acid protein ⇨ 10^{200} possible conformations!
- Earth is less than 10^{10} years old.

Rosetta: A Unified Framework for Protein Structure

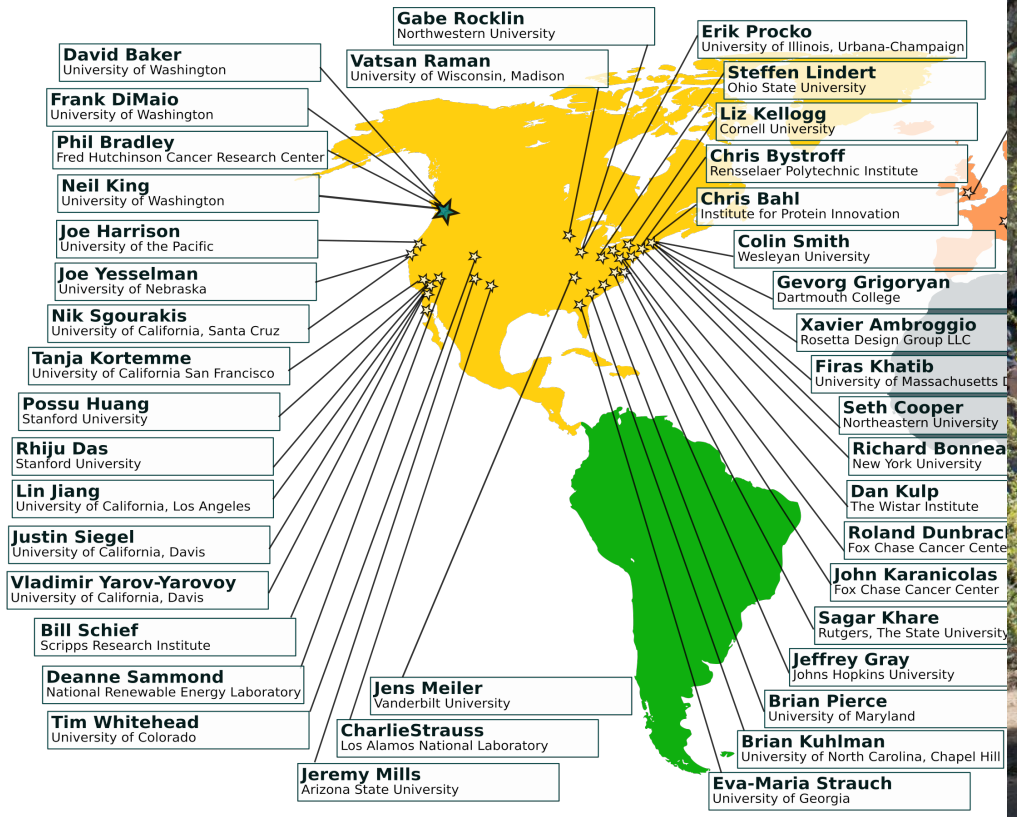
meilerlab





A.
J. K
. 665-680.

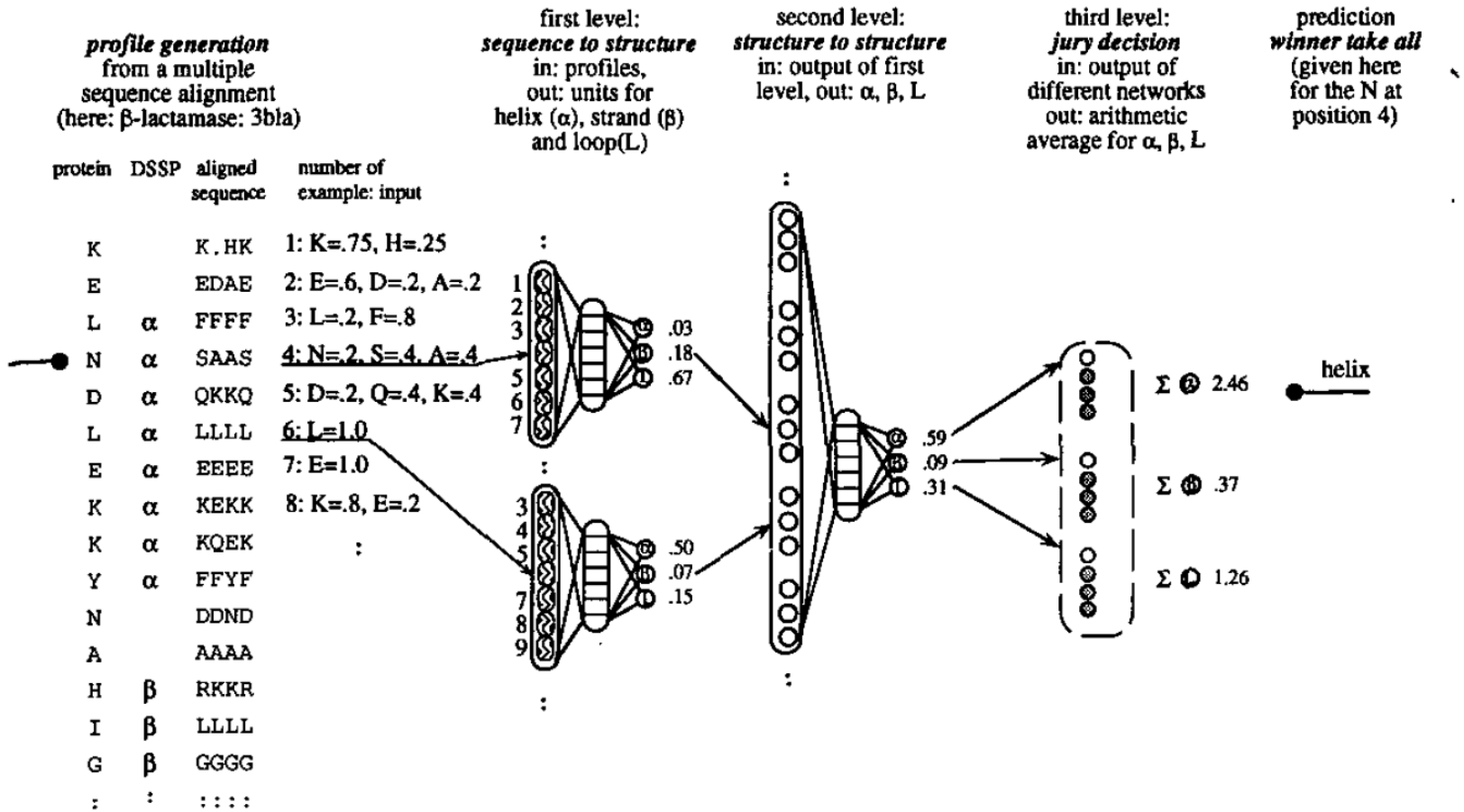
Rosetta: A Unified Framework for Protein Structure Perdiction and Design



A. Leaver-Fay, et al.; "ROSETTA3: an object-oriented software suite …"; Methods Enzymol; 2011; Vol. 487 p. 545-74.

J. K. Leman, et al.; "Macromolecular modeling and design in Rosetta: recent methods and frameworks"; Nat Methods; 2020; Vol. 17 (7): p. 665-680.

PhD - Prediction of protein secondary structure at better than 70% accuracy



B. Rost and C. Sander; "Prediction of protein secondary structure at better than 70% accuracy"; *J. Mol. Biol.*; **1993**; Vol. 232 (2): p. 584-99; J. Meiler, A. Zeidler, F. Schmaschke and M. Muller; "Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks"; *Journal of Molecular Modeling*; **2001**; Vol. 7 (9): p. 360-369.

BCL::Jufo9D >70% correct 9-state prediction, >80% SS, >90% TM

meilerlab

==REAL STATE==	PREDICTION tm_C	tr_C	sol_C	tm_S	tr_S	sol_S	tm_H	tr_H	sol_H
tm_C	63.7	13.3	1.9	4.8	2.0	1.1	6.0	6.9	0.5
tr_C	4.8	64.7	3.4	1.9	7.8	0.5	0.9	15.3	0.8
sol_C	3.8	23.0	43.9	2.4	2.5	8.7	2.1	2.4	11.2
tm_S	6.2	0.9	0.9	78.6	7.5	5.1	0.0	0.3	0.4
tr_S	3.0	12.0	1.7	14.2	66.5	1.8	0.6	0.1	0.0
sol_S	3.7	6.2	11.1	7.0	4.9	55.9	4.0	4.2	3.1
tm_H	2.0	3.8	0.1	0.4	0.1	0.1	80.1	13.2	0.4
tr_H	2.0	16.9	0.2	0.2	1.1	0.1	12.0	66.0	1.6
sol_H	1.4	6.0	9.0	0.8	0.7	2.9	2.0	8.8	68.5

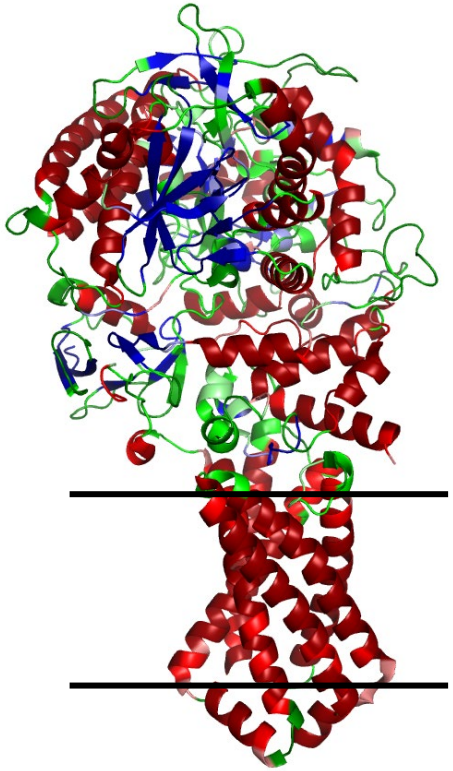
TM trans-membrane
TR transition
SOL solution

H helix
S strand
C coil

J. K. Leman, R. Mueller, M. Karakas, N. Woetzel and J. Meiler; "Simultaneous prediction of protein secondary structure and transmembrane spans"; *Proteins*; **2013**; Vol. 81 (7): p. 1127-40.

Example 1: Succinate dehydrogenase (1NEK)

- Secondary structure prediction

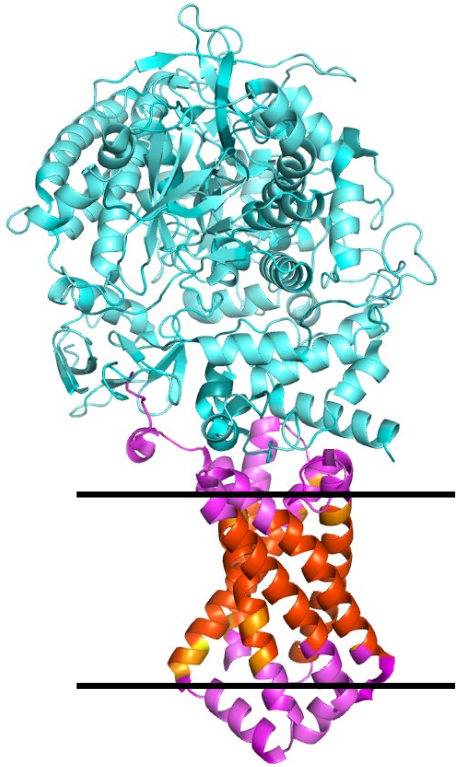


helix
strand
coil

membrane
transition
solution

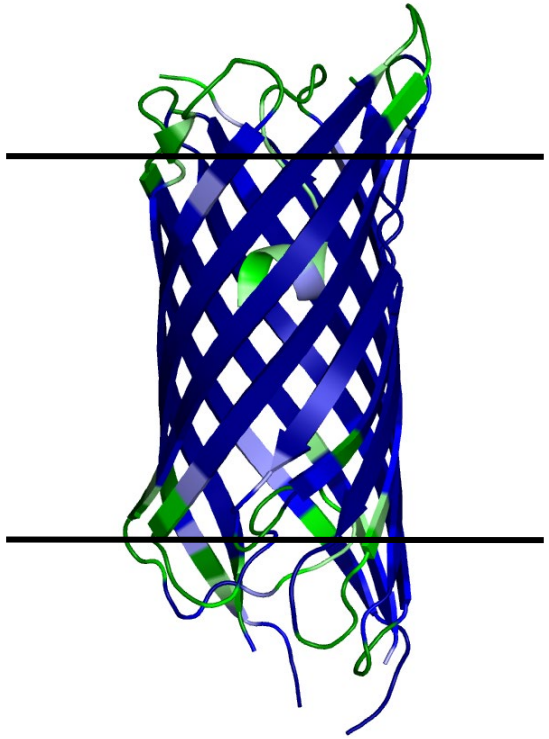
92% **81%**
residues correct

- Trans-membrane span identification



Example 2: EspP autotransporter beta-domain (2QOM)

- Secondary structure prediction

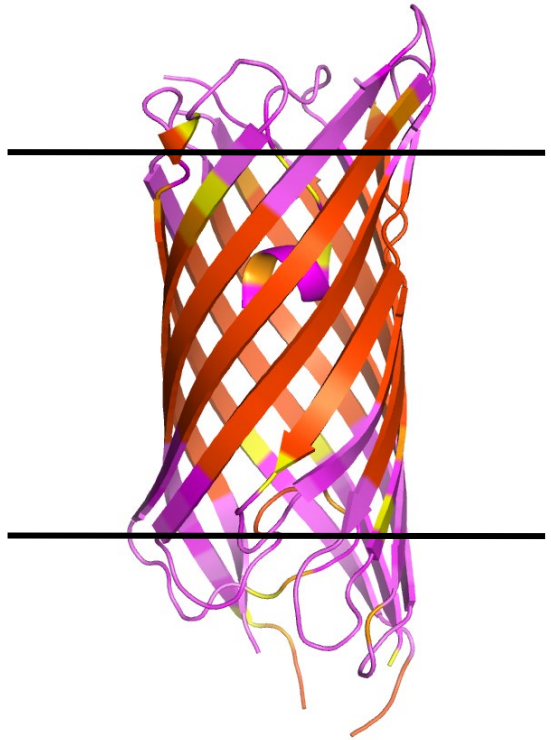


helix
strand
coil

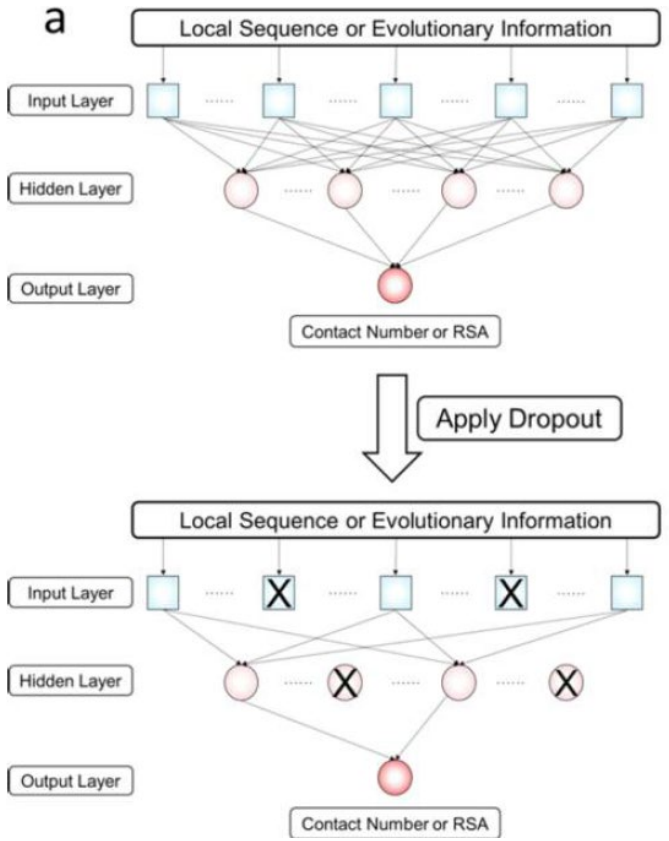
membrane
transition
solution

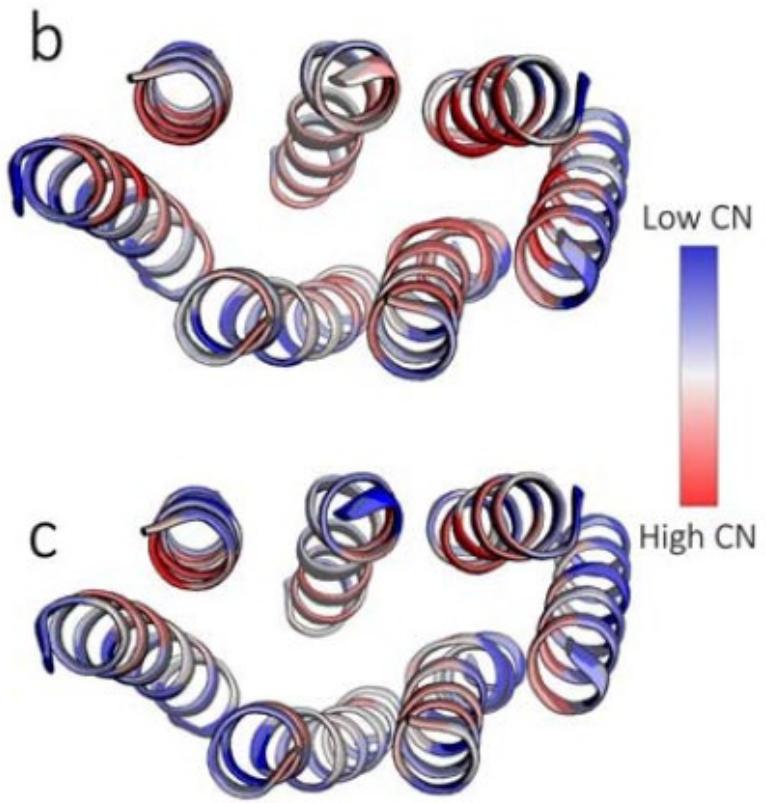
80% 74%
residues correct

- Trans-membrane span identification



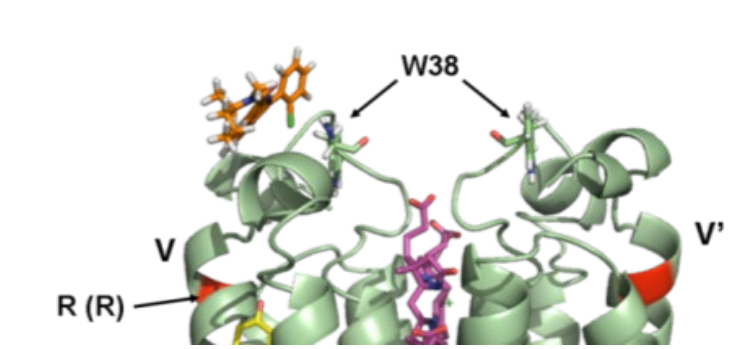
ANN – Derived Contact Numbers Improve Membrane Protein Structure Prediction





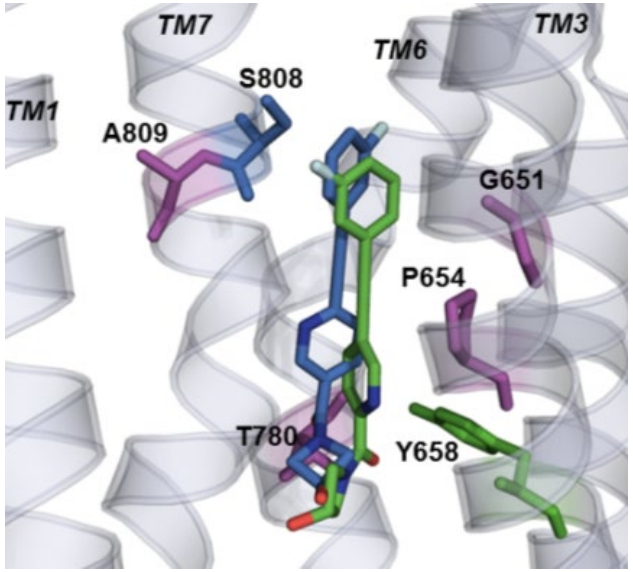
B. Li, J. Mendenhall, E. D. Nguyen, B. E. Weiner, A. W. Fischer and J. Meiler; "Accurate Prediction of Contact Numbers for Multi-Spanning Helical Membrane Proteins"; *J Chem Inf Model*; **2016**; Vol. 56 (2): p. 423-34.
B. Li, J. Mendenhall, E. D. Nguyen, B. E. Weiner, A. W. Fischer and J. Meiler; "Improving prediction of helix-helix packing in membrane proteins using predicted contact numbers as restraints"; *Proteins*; **2017**; Vol. 85 (7): p. 1212-1221.

Computational Structural and Chemical Biology



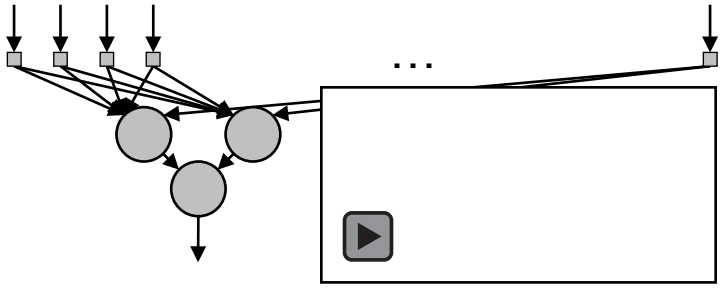
Weiner, et al., **Structure**, 2013, 21(7)

Protein structure prediction de novo and from limited experimental data

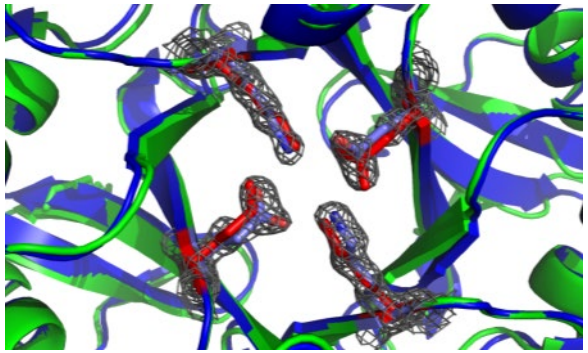


Dong, et al., **PLoS ONE**, 2013, 8(7)

Merging ligand- and structure-based computer-aided drug discovery



Muller, et al., **ChemMedChem**, 2012, 7(3)

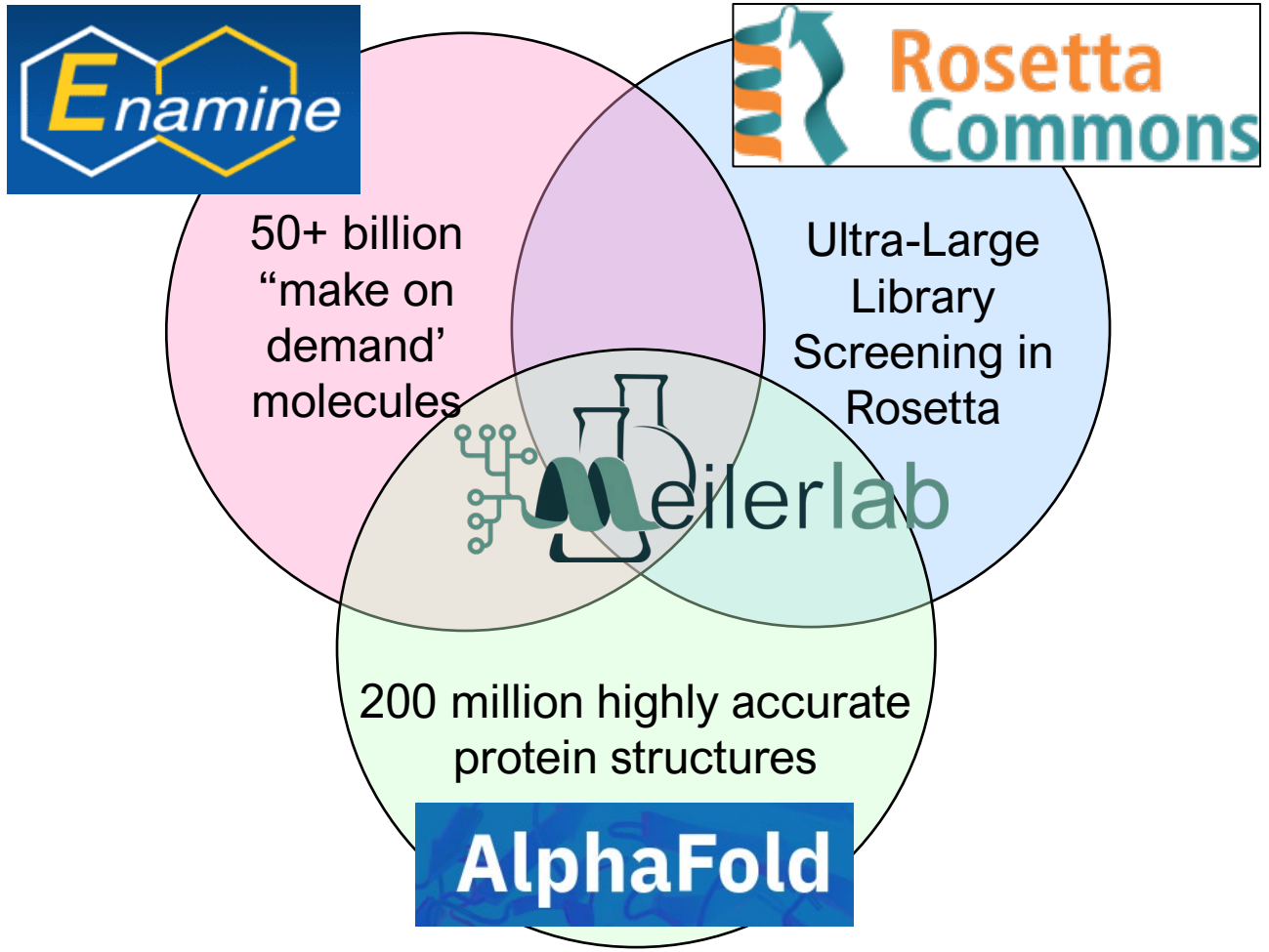


Fortenberry, et al., **JACS**, 2011, 133(45)

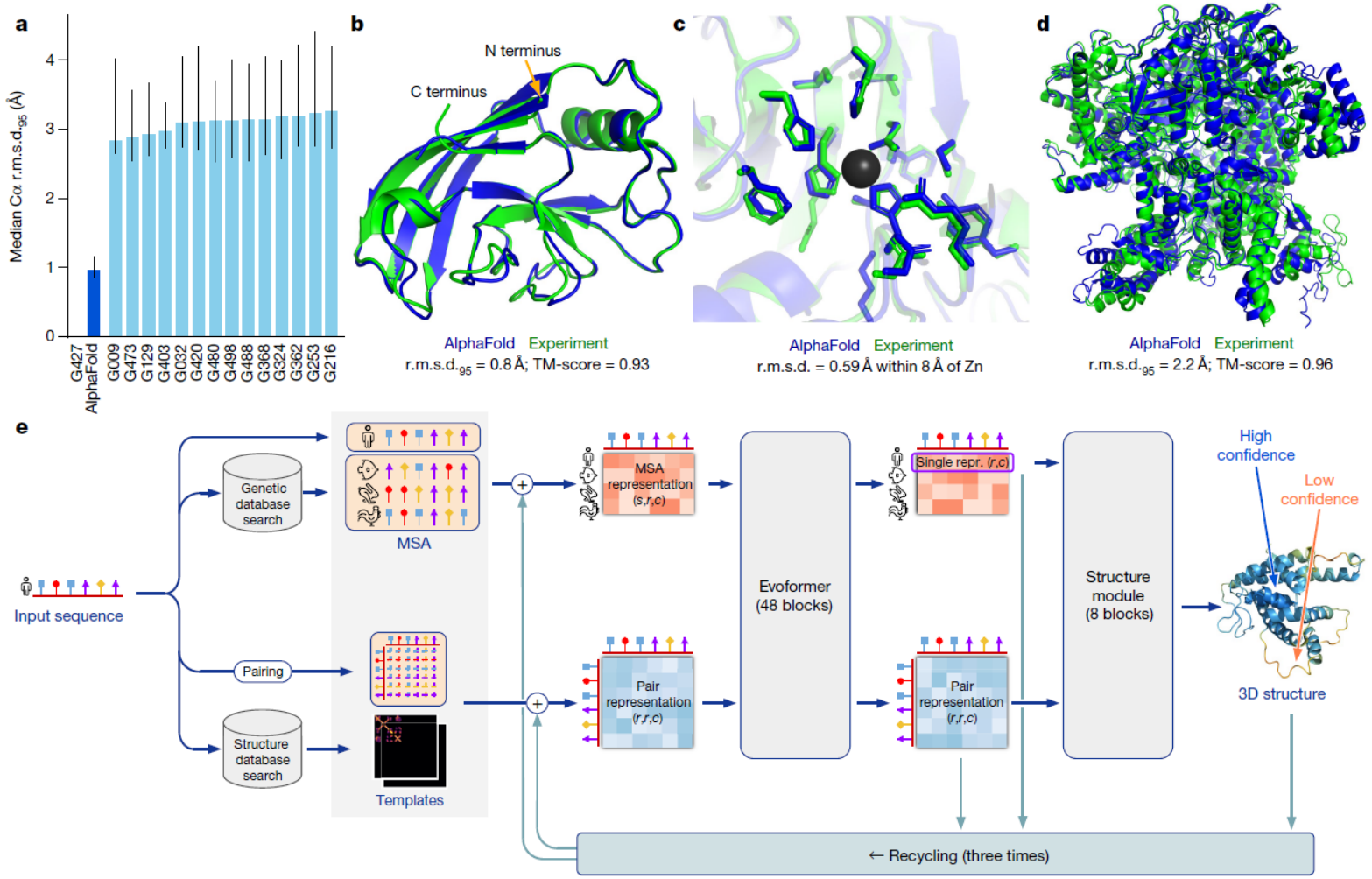
Design of large protein therapeutics, antibodies, and protein interfaces

Willis, et al., **PLoS Comp. Biol.**, 2013, 9(4)

A Paradigm Shift in Therapeutic Development in Academia: ***in silico* First**

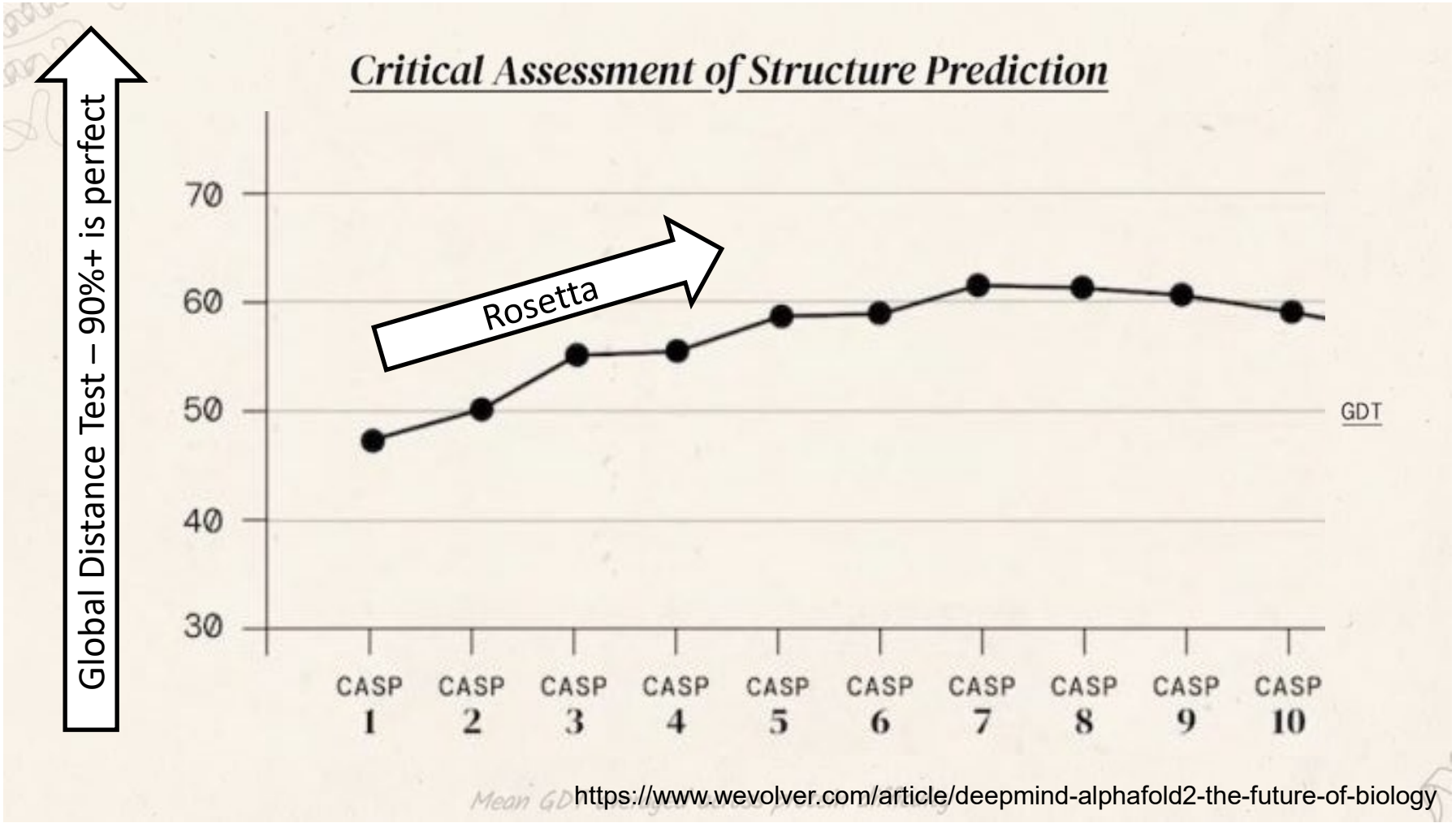


Highly accurate protein structure prediction with AlphaFold2



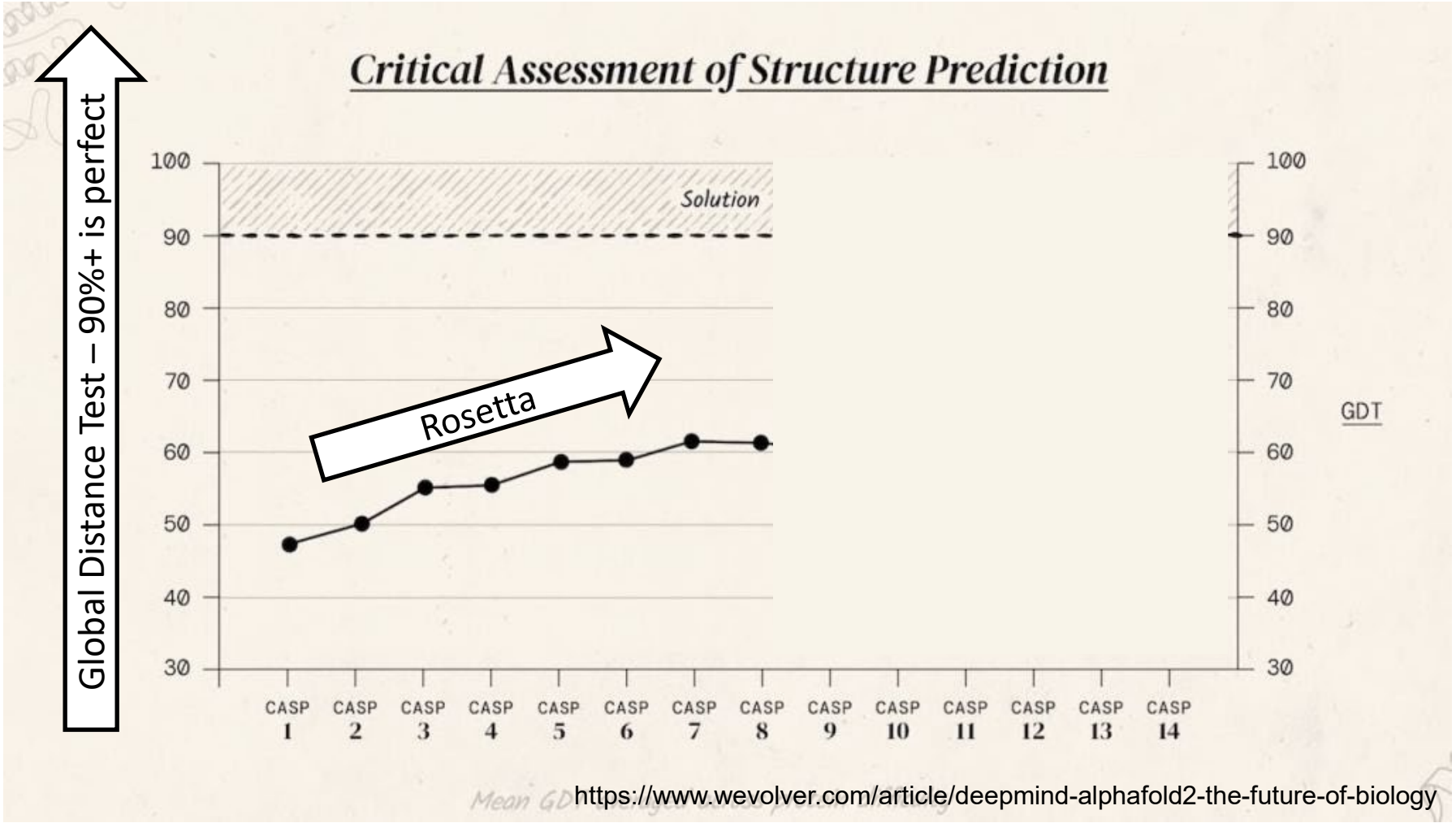
J. Jumper, et al.; "Highly accurate protein structure prediction with AlphaFold"; *Nature*; **2021**; Vol. 596 (7873): p. 583-589.

Critical Assessment of protein Structure Prediction (CASP) – established 1994



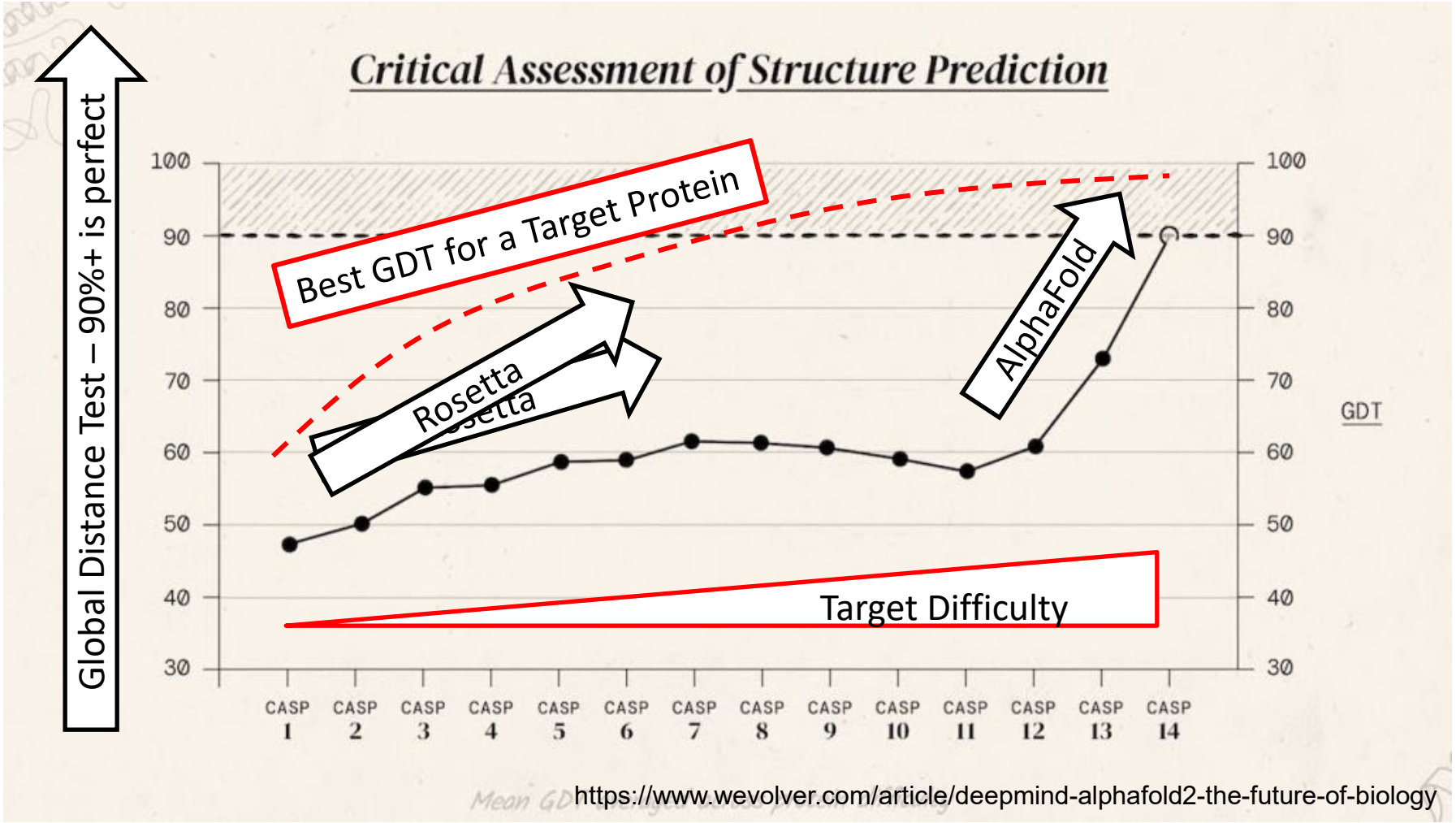
https://www.wevolver.com/article/deepmind-alphafold2-the-future-of-biology

Critical Assessment of protein Structure Prediction (CASP) – established 1994



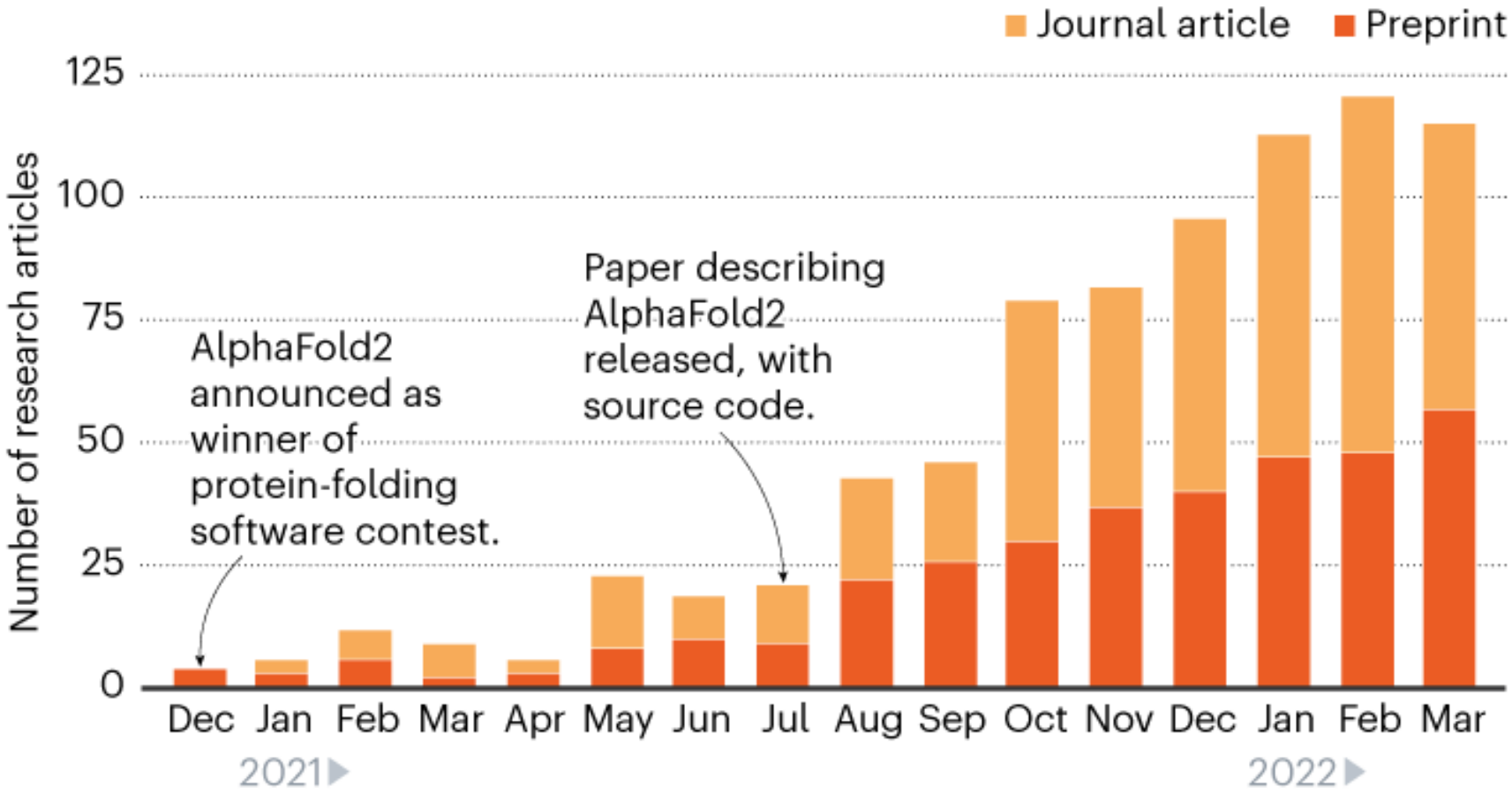
https://www.wevolver.com/article/deepmind-alphafold2-the-future-of-biology

Critical Assessment of protein Structure Prediction (CASP) – established 1994



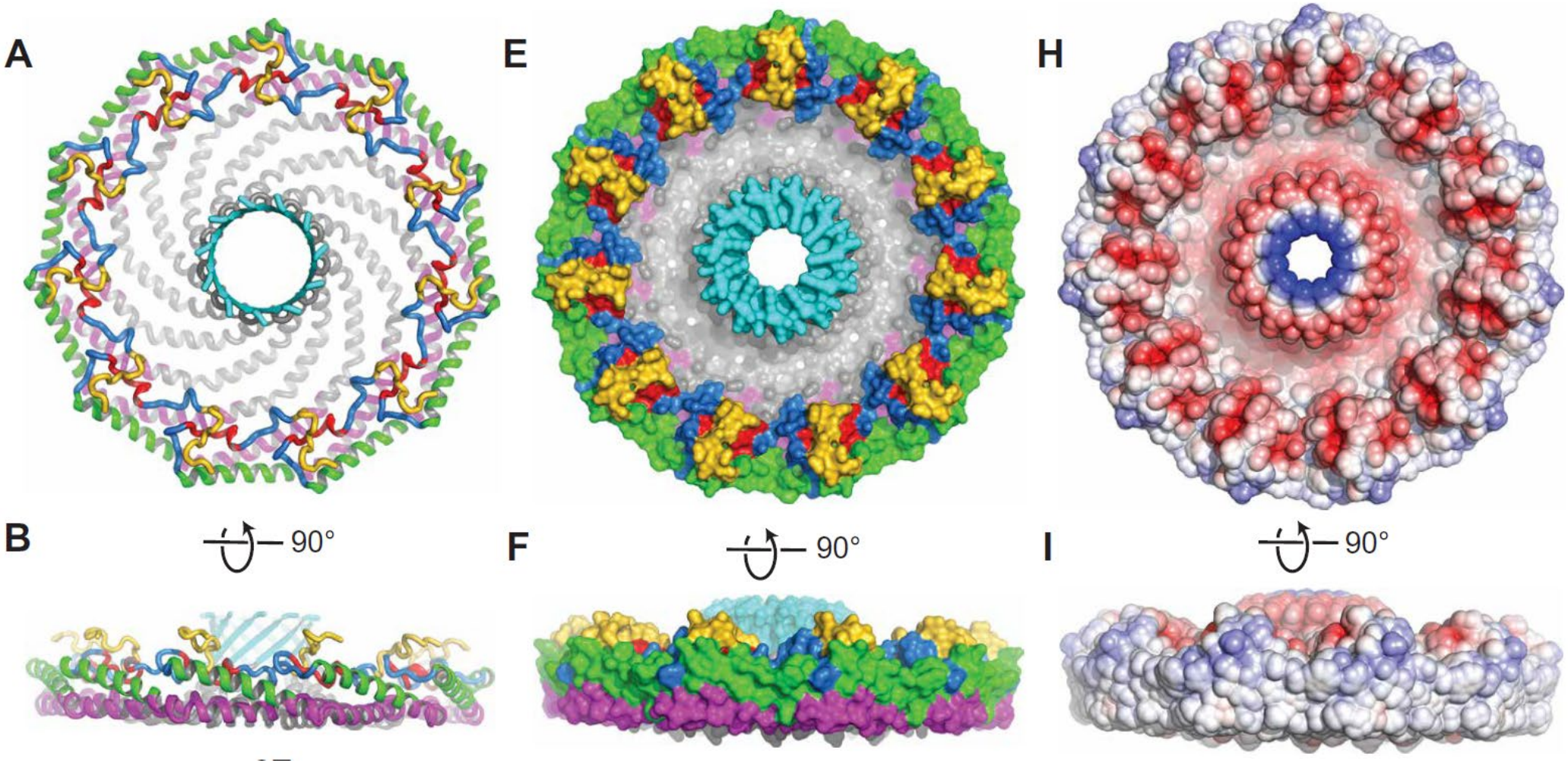
https://www.wevolver.com/article/deepmind-alphafold2-the-future-of-biology

AlphaFoldMania – The number of research papers and preprints



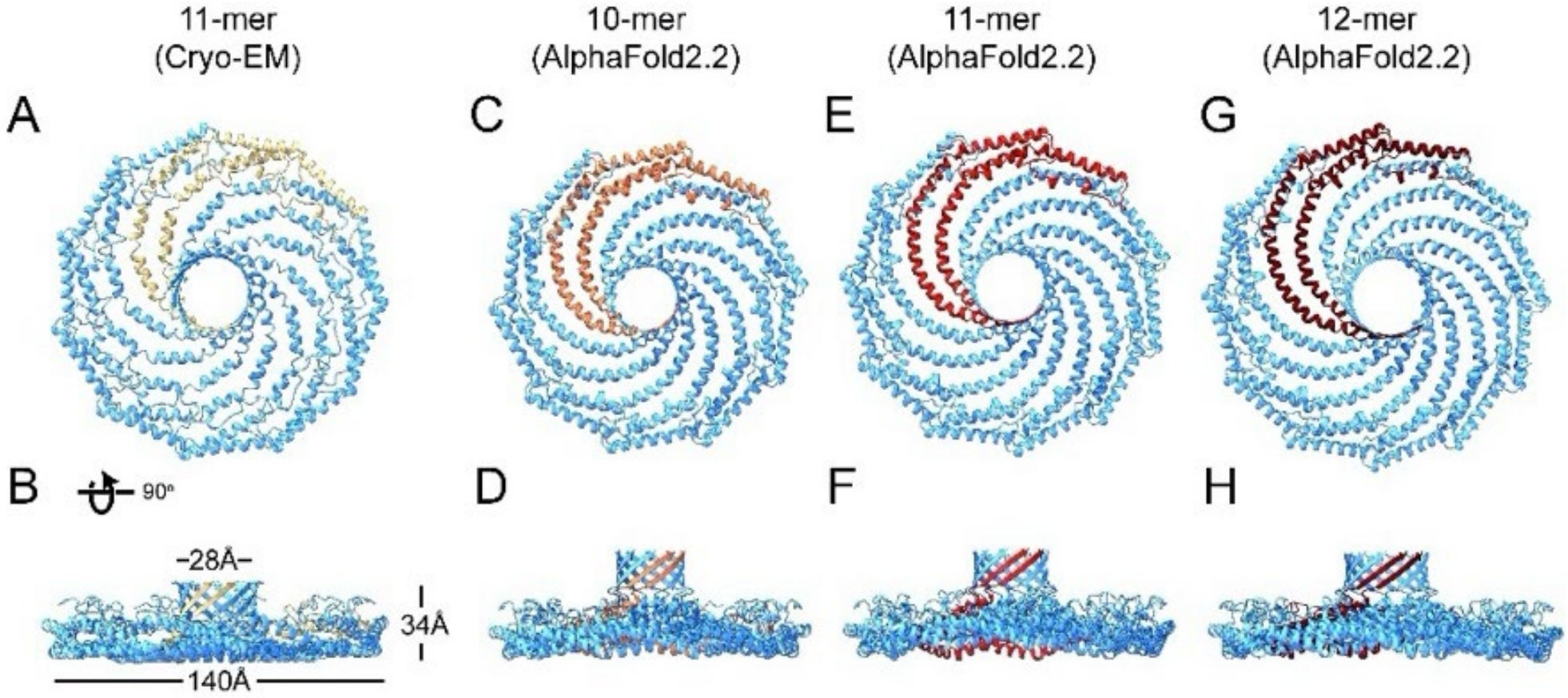
Nature, News Feature, 13 April 2022

Molecular Architecture of the Human Caveolin-1 Complex



J. C. Porta, B. Han, A. Gulsevin, J. Chung, Y. Peskova, S. Connolly, H. S. Mchaourab, J. Meiler, E. Karakas, A. K. Kenworthy and M. D. Ohi; "Molecular architecture of the human caveolin-1 complex"; *Science Advances*; **2022**; Vol. p.

Molecular Architecture of the Human Caveolin-1 Complex with AlphaFold2



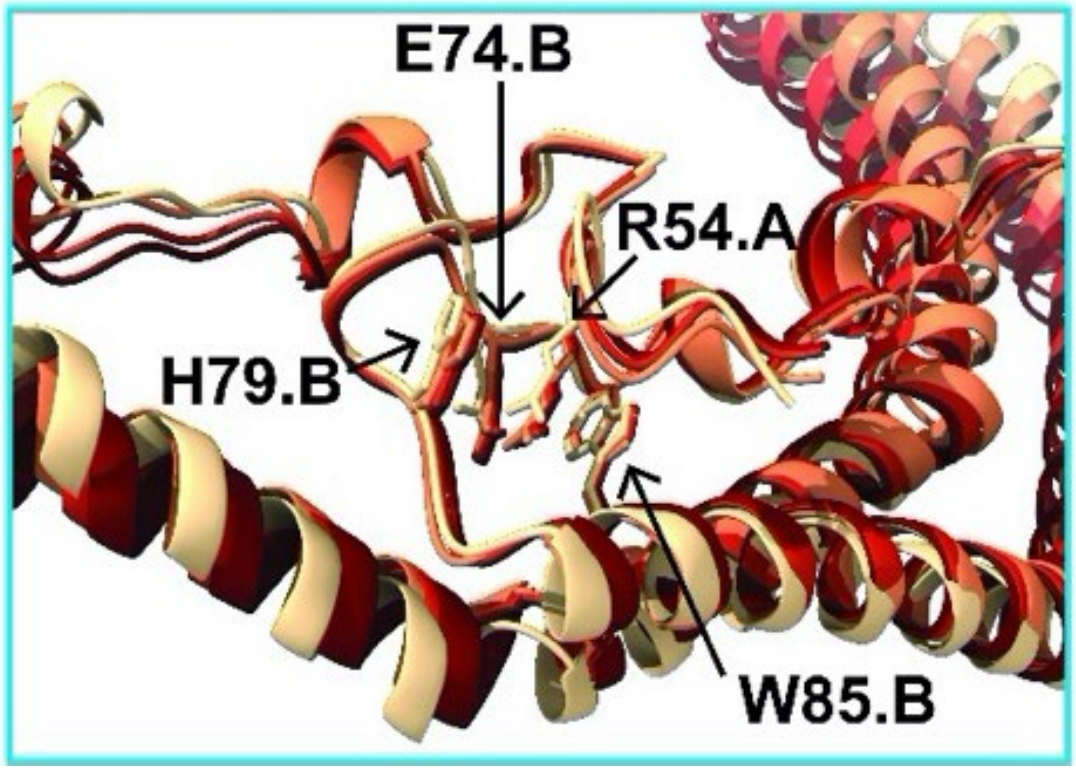
A. Gulsevin, B. Han, J. C. Porta, H. S. McHaourab, J. Meiler and A. K. Kenworthy; "Template-free prediction of a new monotopic membrane protein fold and assembly by AlphaFold2"; *Biophys J;* ***2022; Vol. p.***

J. C. Porta, B. et al. "Molecular architecture of the human caveolin-1 complex"; *Science Advances*; **2022**; Vol. p.

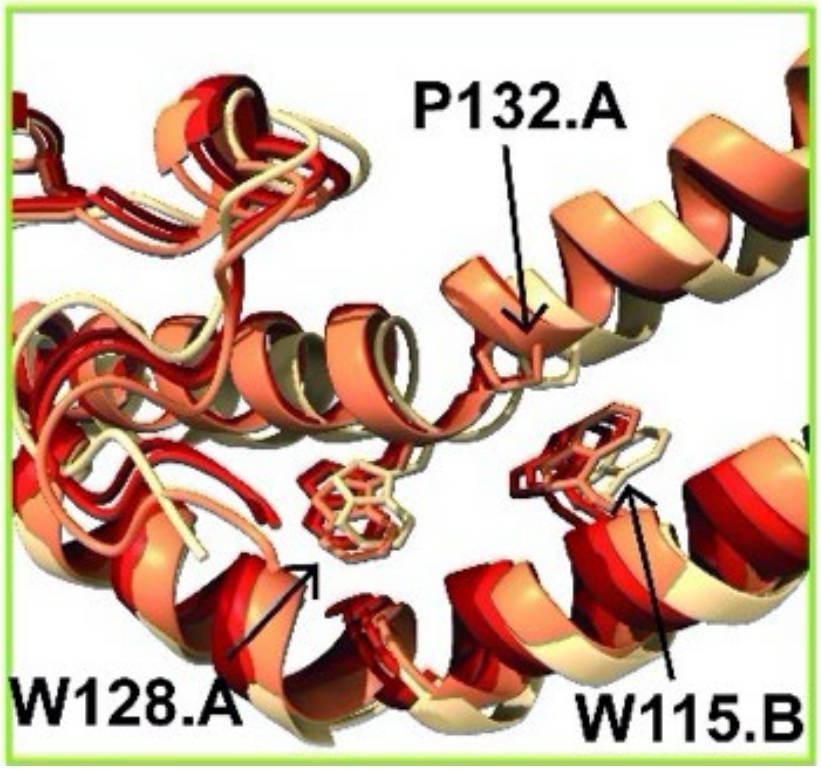
Molecular Architecture of the Human Caveolin-1 Complex with AlphaFold2

meilerlab

K



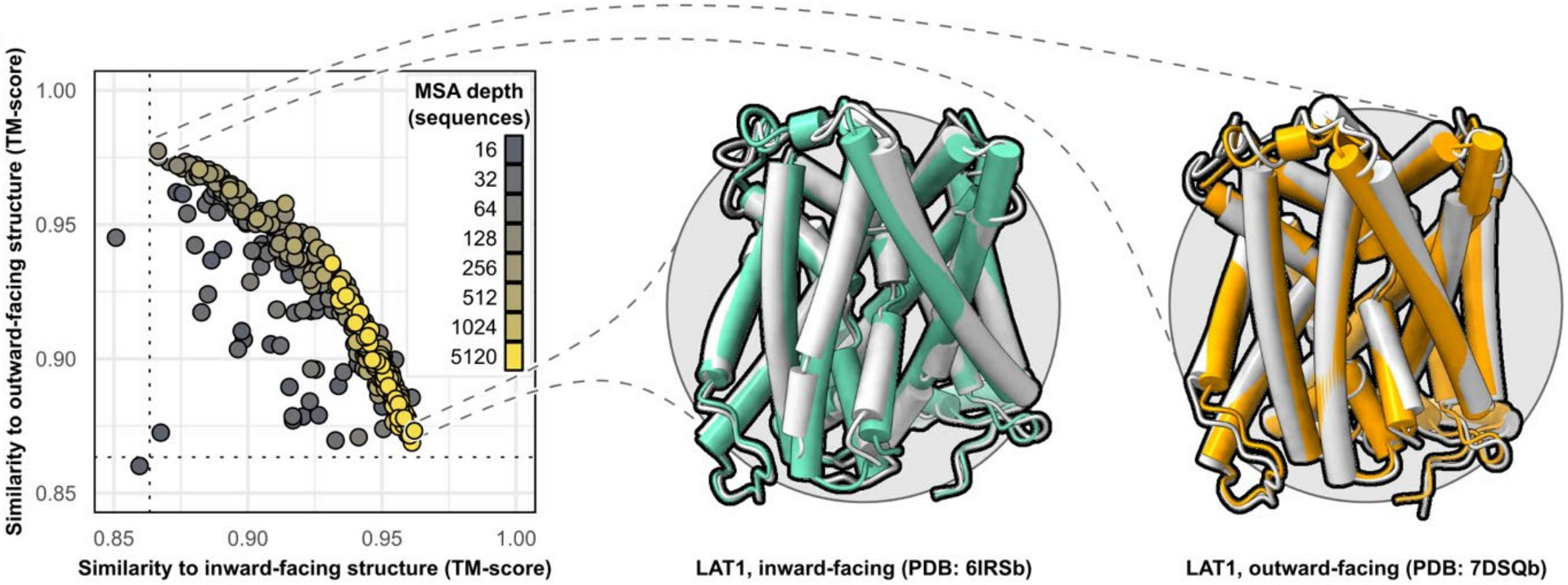
L



A. Gulsevin, B. Han, J. C. Porta, H. S. McHaourab, J. Meiler and A. K. Kenworthy; "Template-free prediction of a new monotopic membrane protein fold and assembly by AlphaFold2"; *Biophys J;* ***2022; Vol. p.***

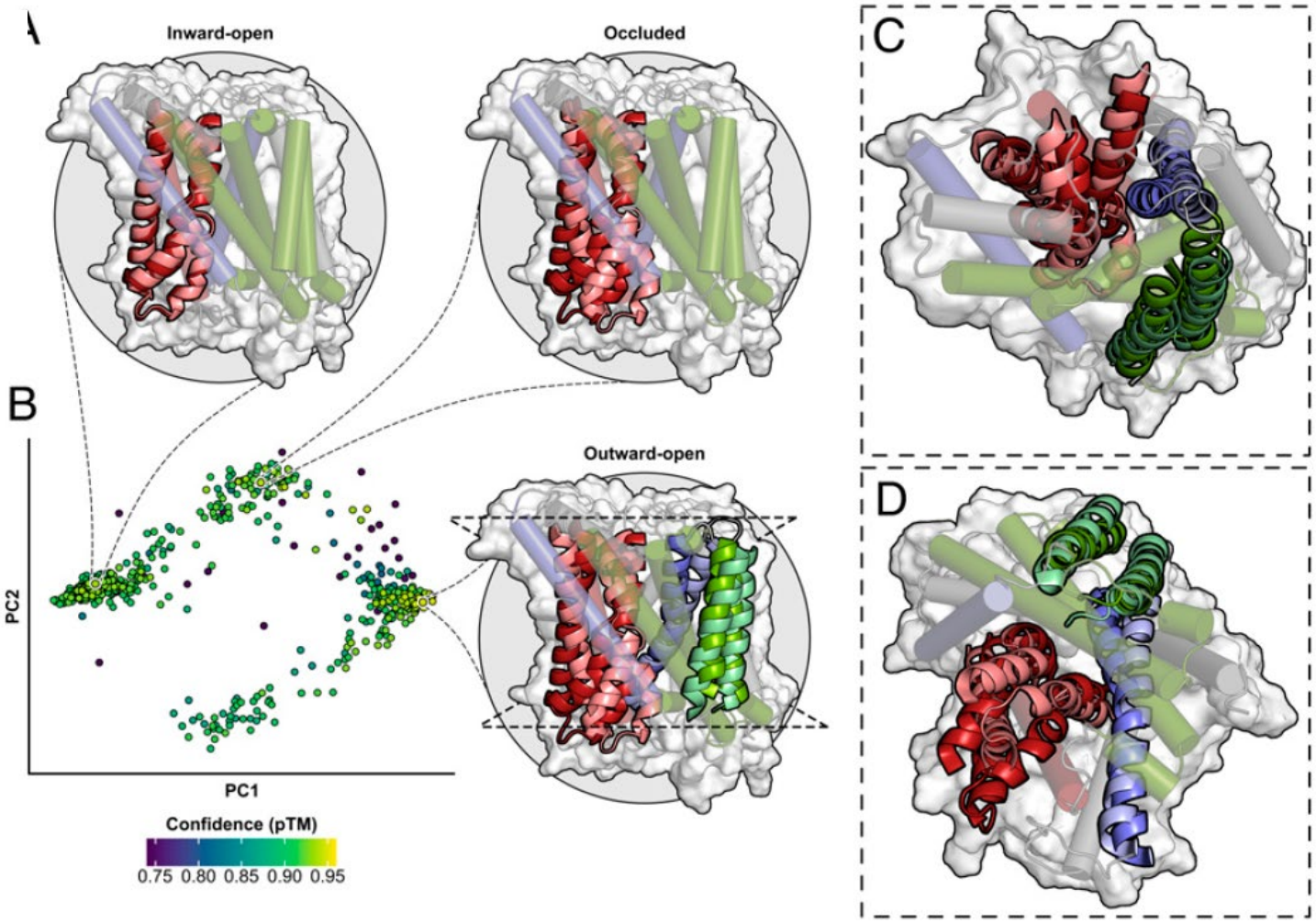
J. C. Porta, B. et al. "Molecular architecture of the human caveolin-1 complex"; *Science Advances*; **2022**; Vol. p.

Sampling Alternative Conformational States with AlphaFold2



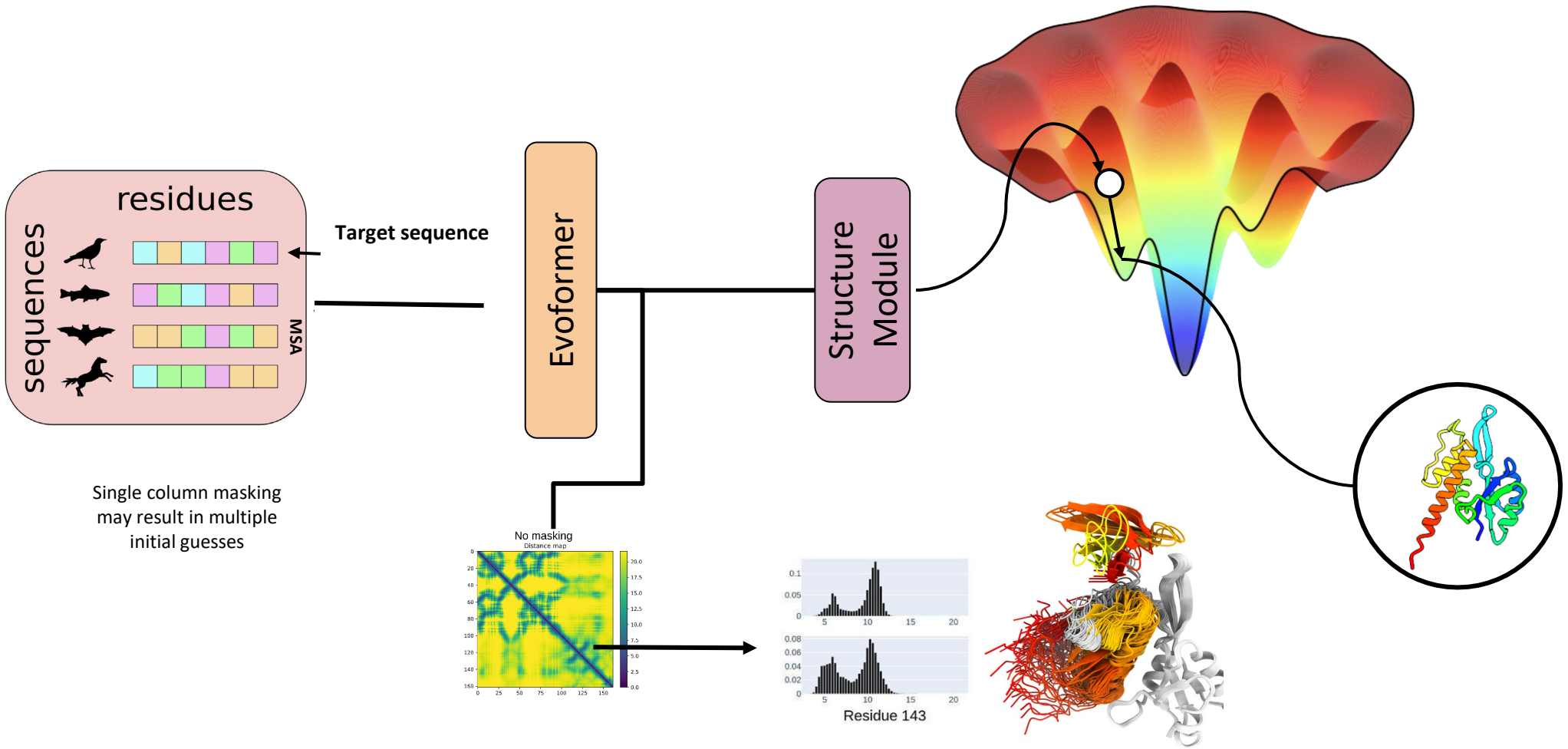
D. Del Alamo, D. Sala, H. S. Mchaourab and J. Meiler; "Sampling alternative conformational states of transporters and receptors with AlphaFold2"; *Elife*; **2022**; Vol. 11 p.

Integrating Limited Experimental Data: NMR, EPR, MassSpec, cryo-EM, …

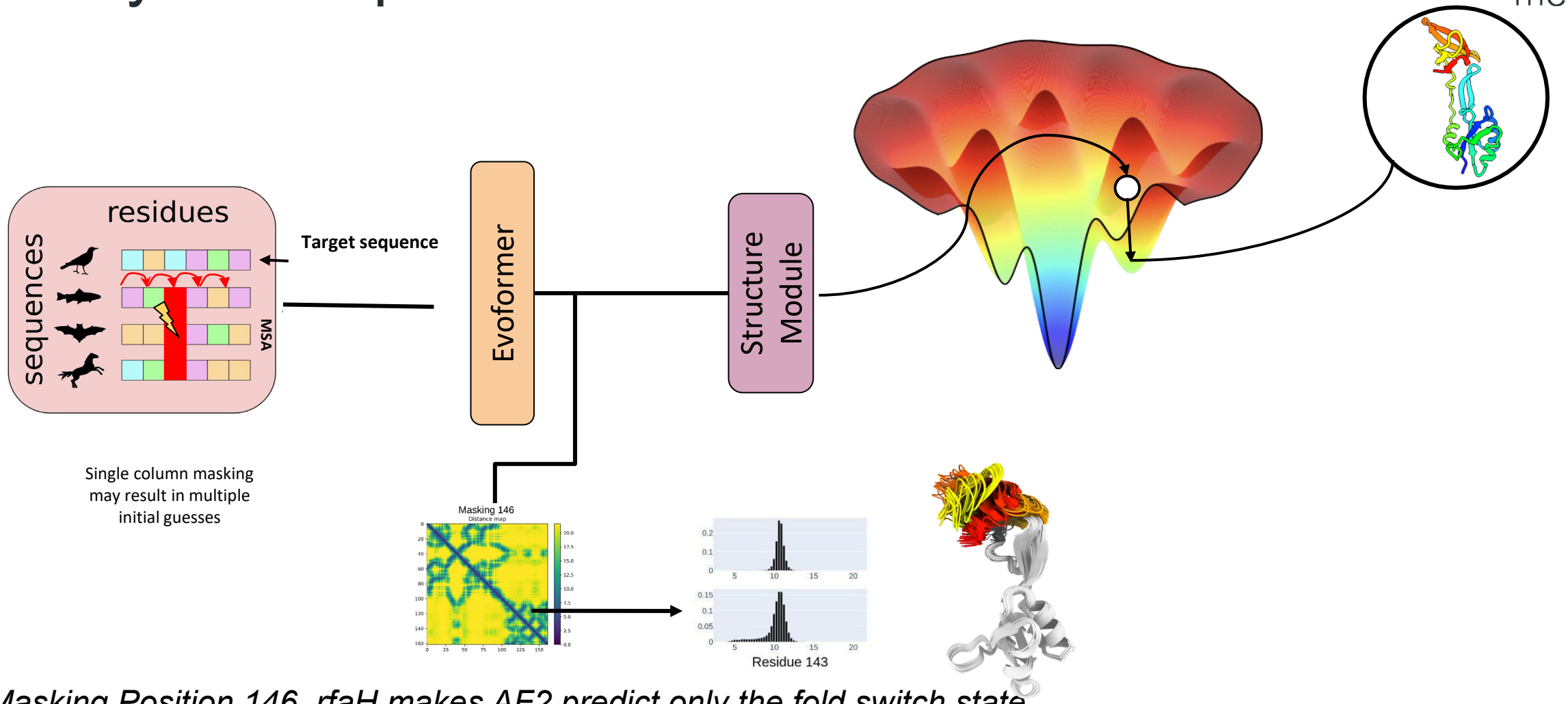


D. Del Alamo, L. DeSousa, R. M. Nair, S. Rahman, J. Meiler and H. S. Mchaourab; "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter"; *Proc Natl Acad Sci U S A;* ***2022; Vol. 119 (34): p. e2206129119***

Explaining by removing: Demystifying AF2 ability to sample conformations – rfaH



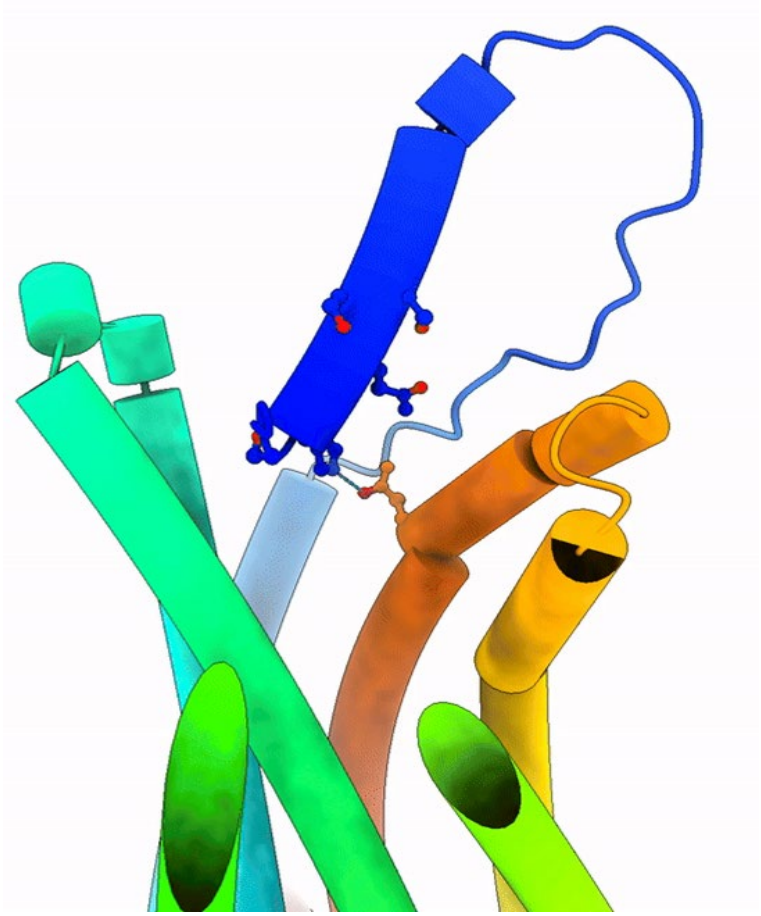
Explaining by removing: Demystifying AF2 ability to sample conformations – rfaH

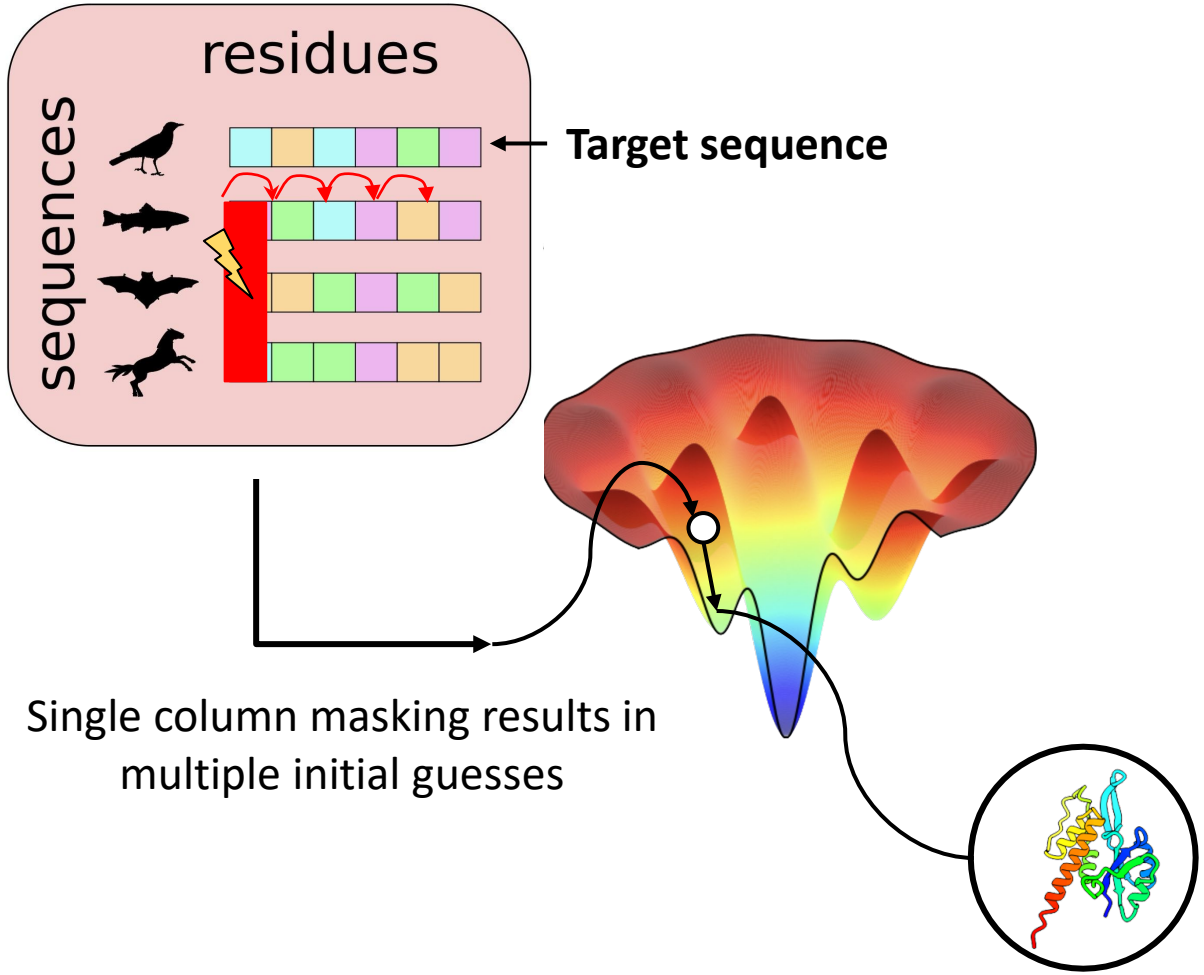


Masking Position 146 rfaH makes AF2 predict only the fold switch state

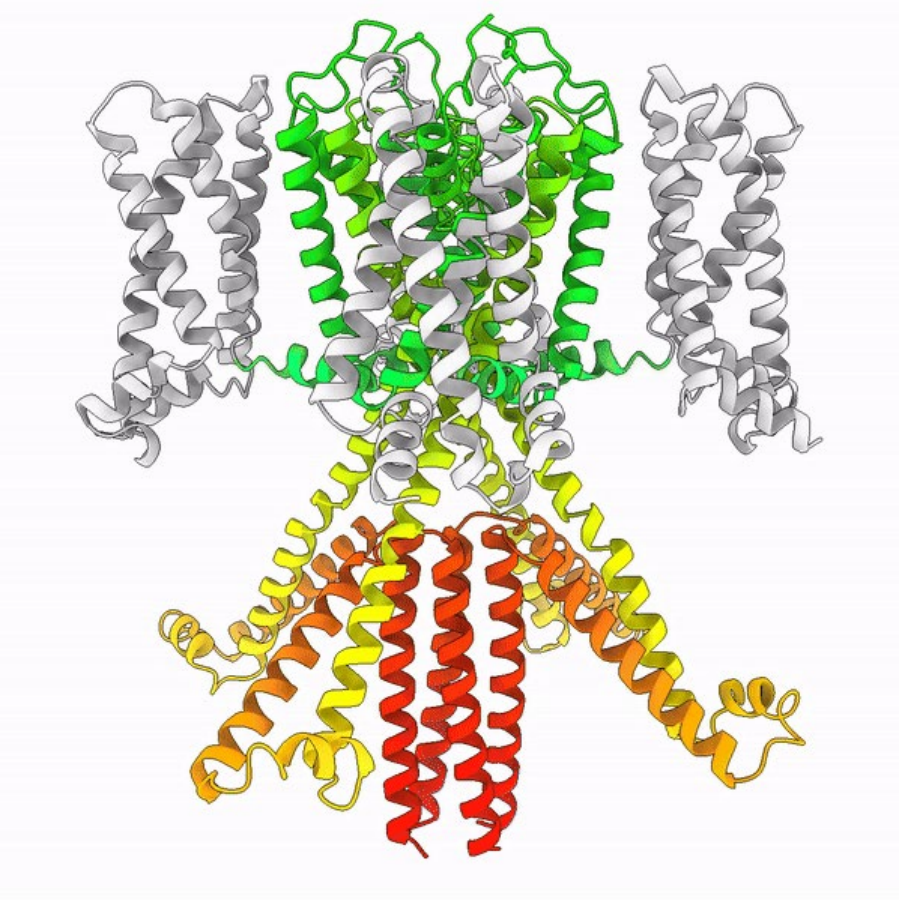
Demystifying AF2 ability to sample conformations

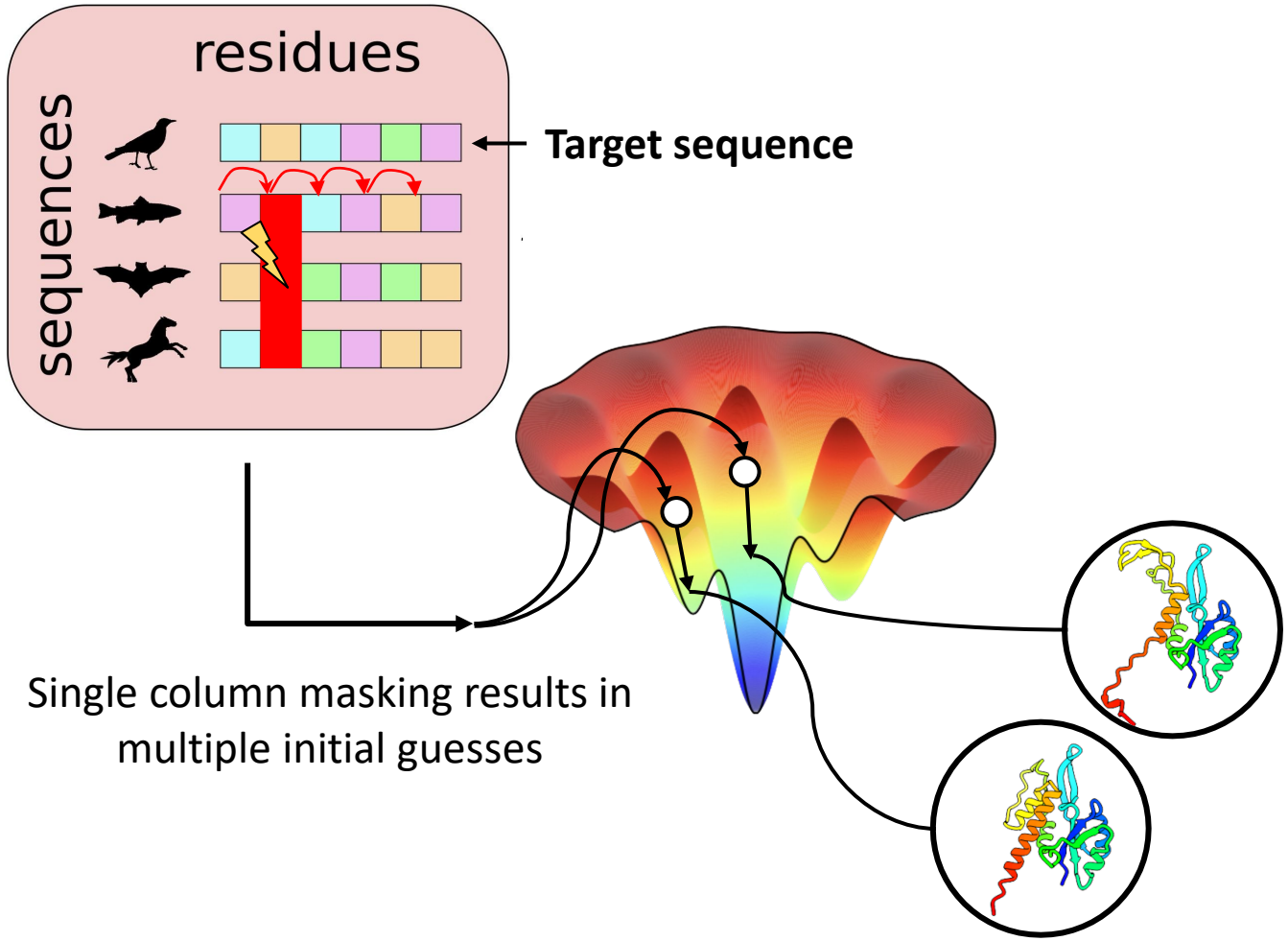
MC4R auto inhibiting peptide





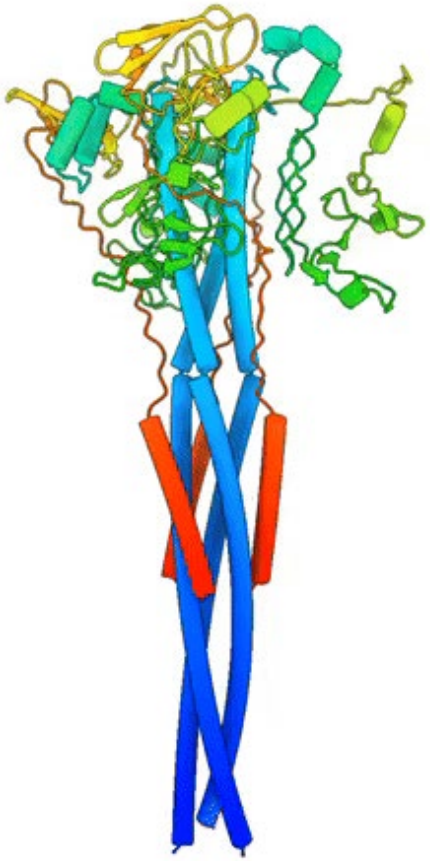
Demystifying AF2 ability to sample conformations
KCNQ1 potassium channel

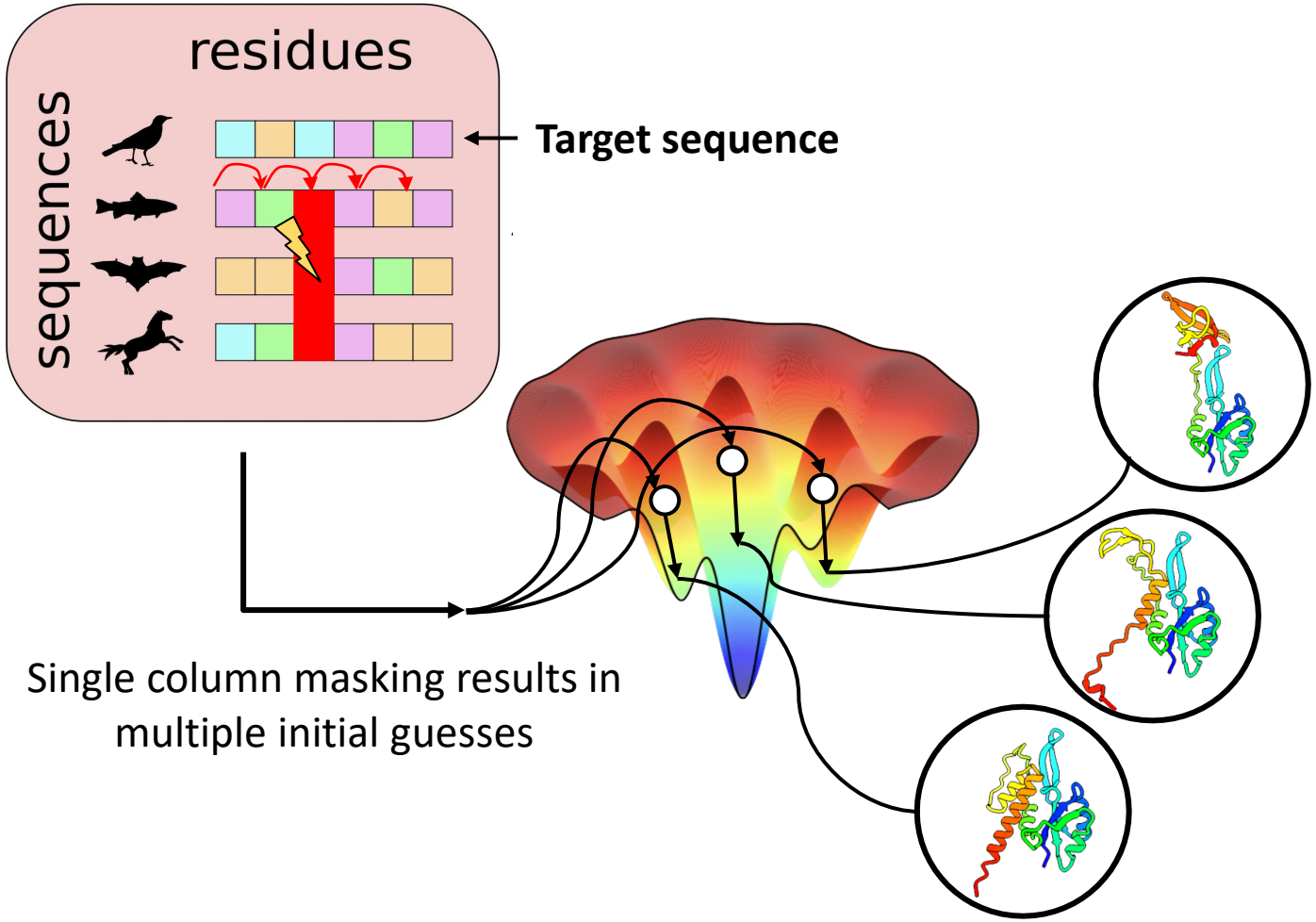




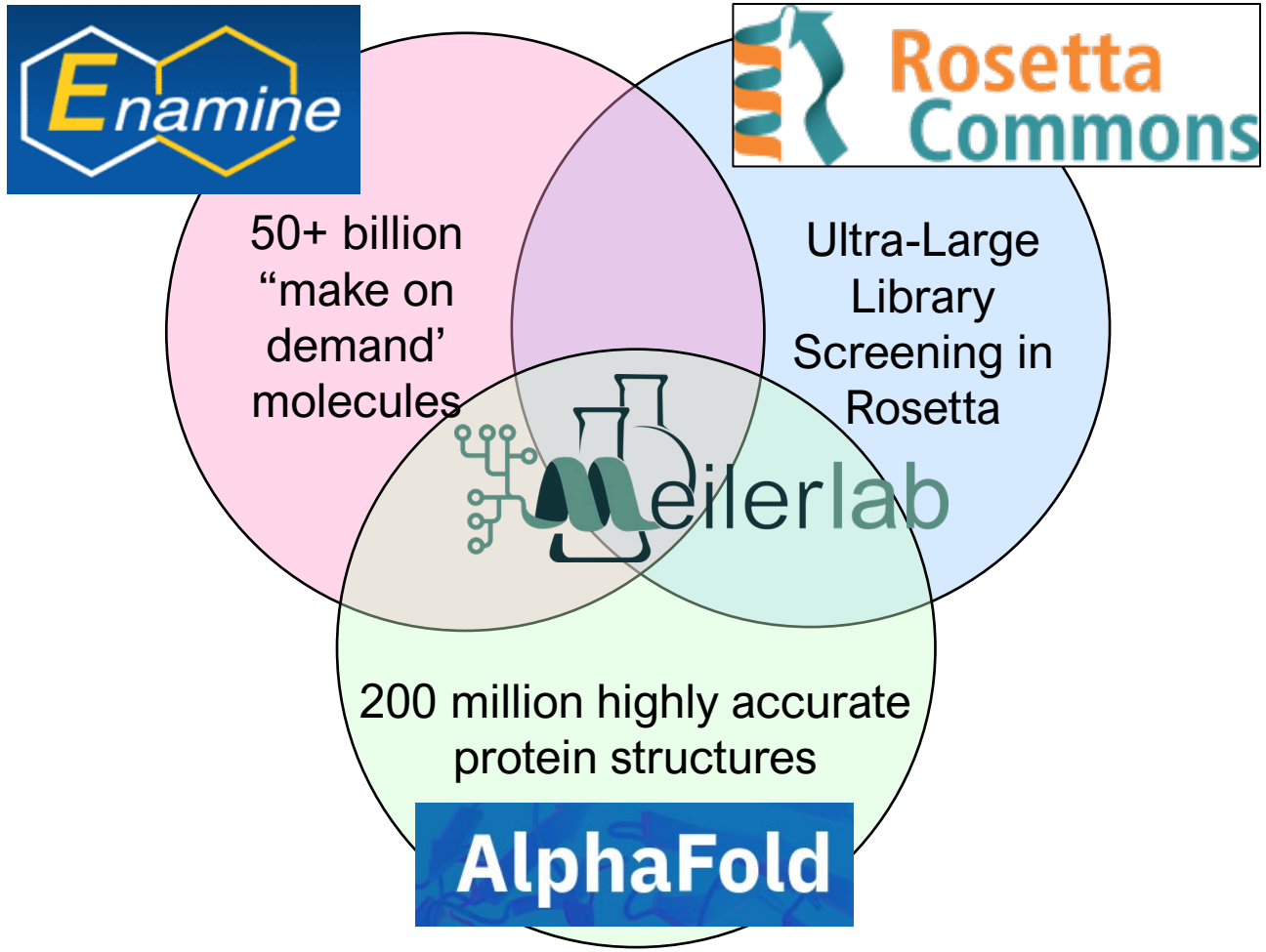
Demystifying AF2 ability to sample conformations
Respiratory Syncytial Virus Glycoprotein

meilerlab

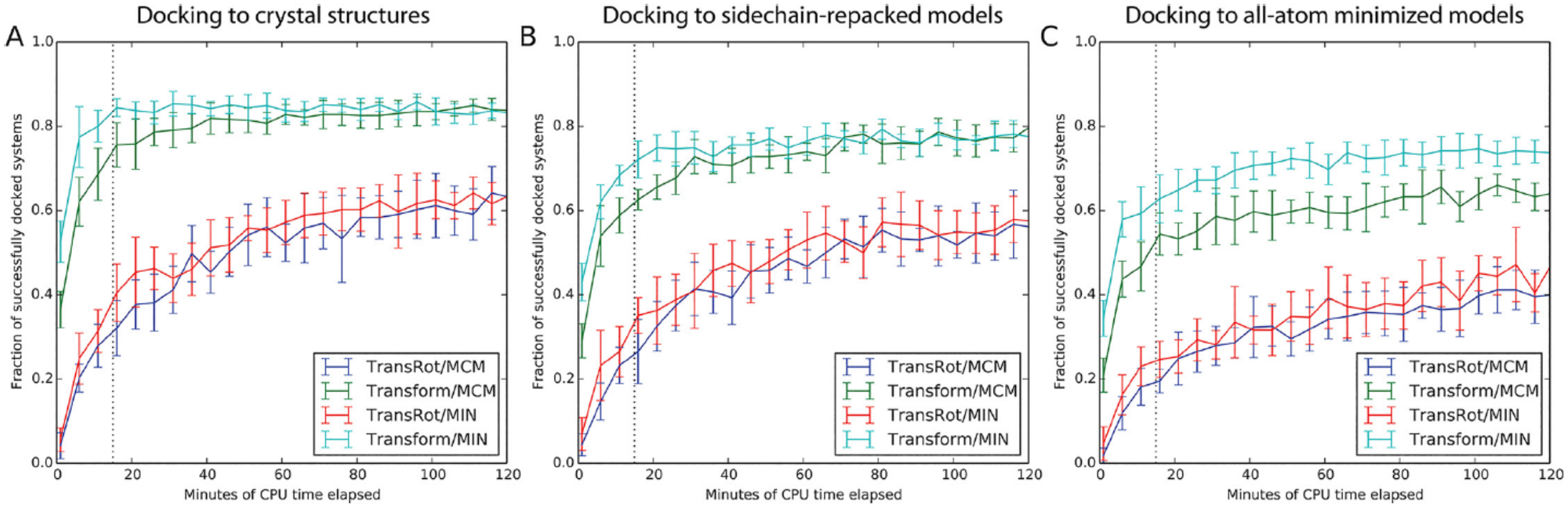




A Paradigm Shift in Therapeutic Development in Academia: ***in silico* First**

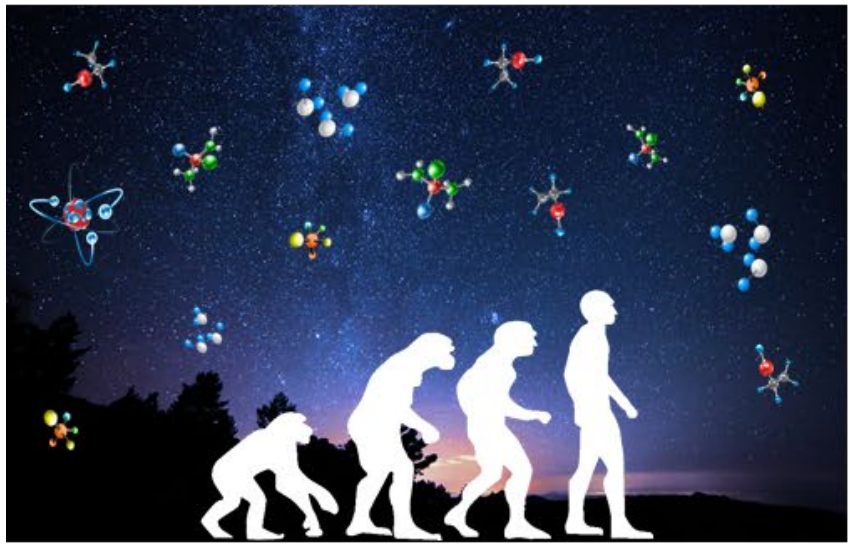


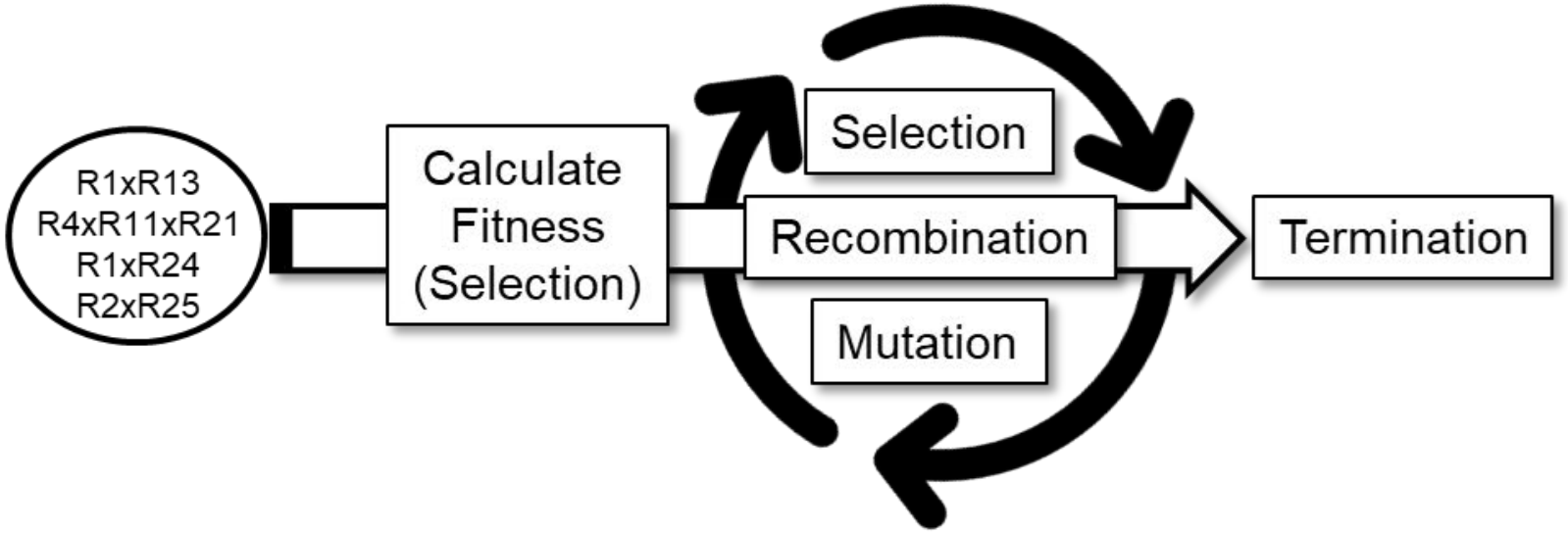
Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand

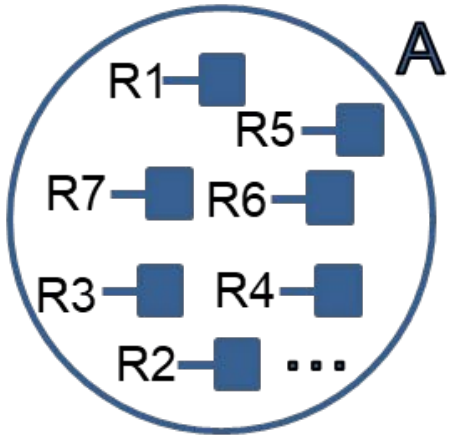


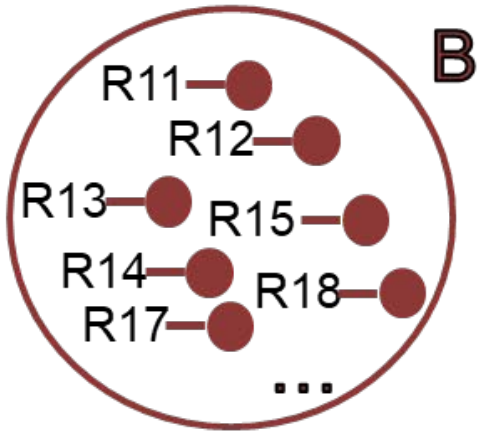
DeLuca, S., Khar, K., & Meiler, J. (2015). Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One, 10(7), e0132508.* || Meiler, J., & Baker, D. (2006). RosettaLigand: protein-small molecule docking with full side-chain flexibility. *Proteins, 65(3), 538-548.*

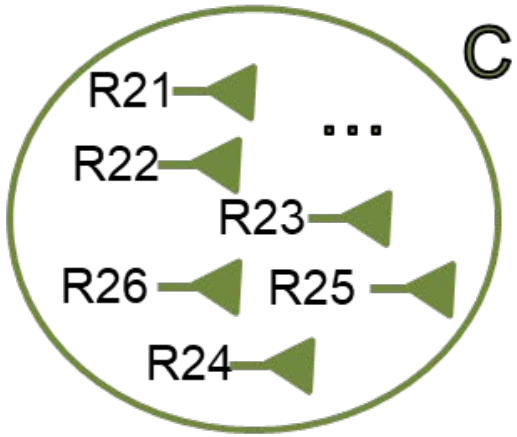
Rosetta Evolutionary Ligand – REvoLd: Ultra-Large Library Screening with an Evolutionary Algorithm











$A + B \Longrightarrow P_{AB}$

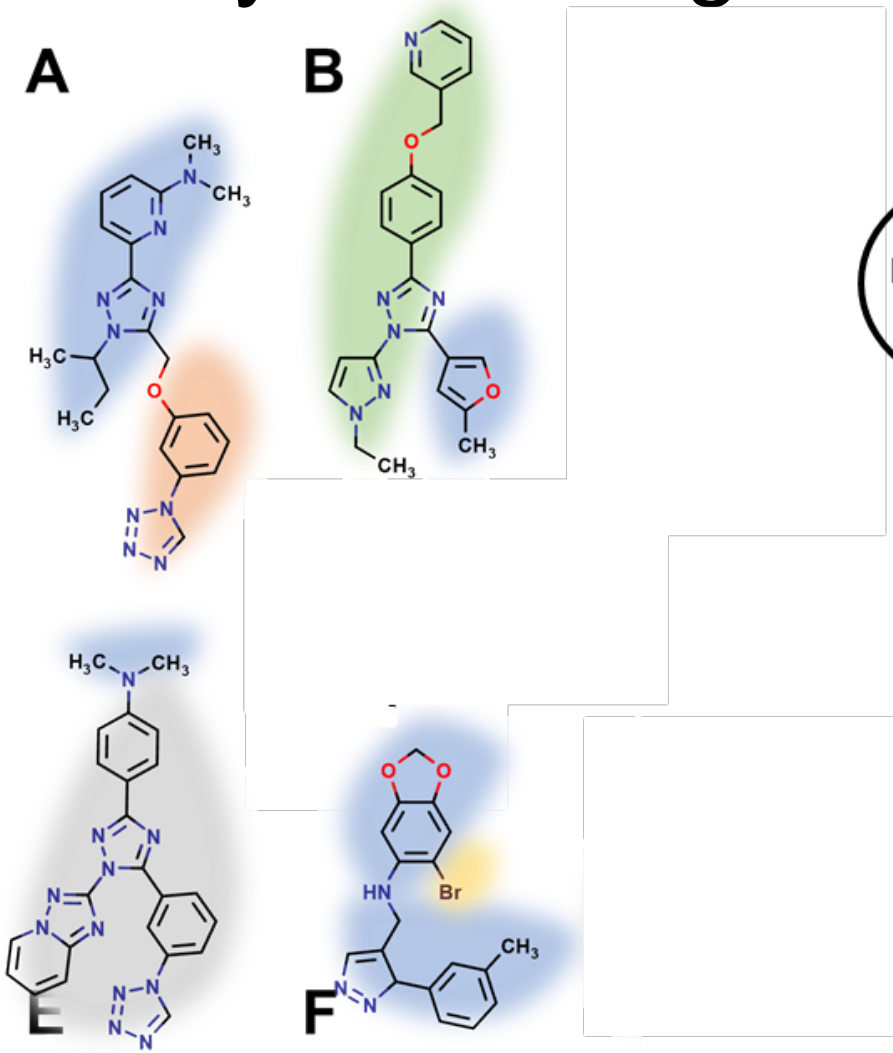
$A + B + C \Longrightarrow P_{ABC}$

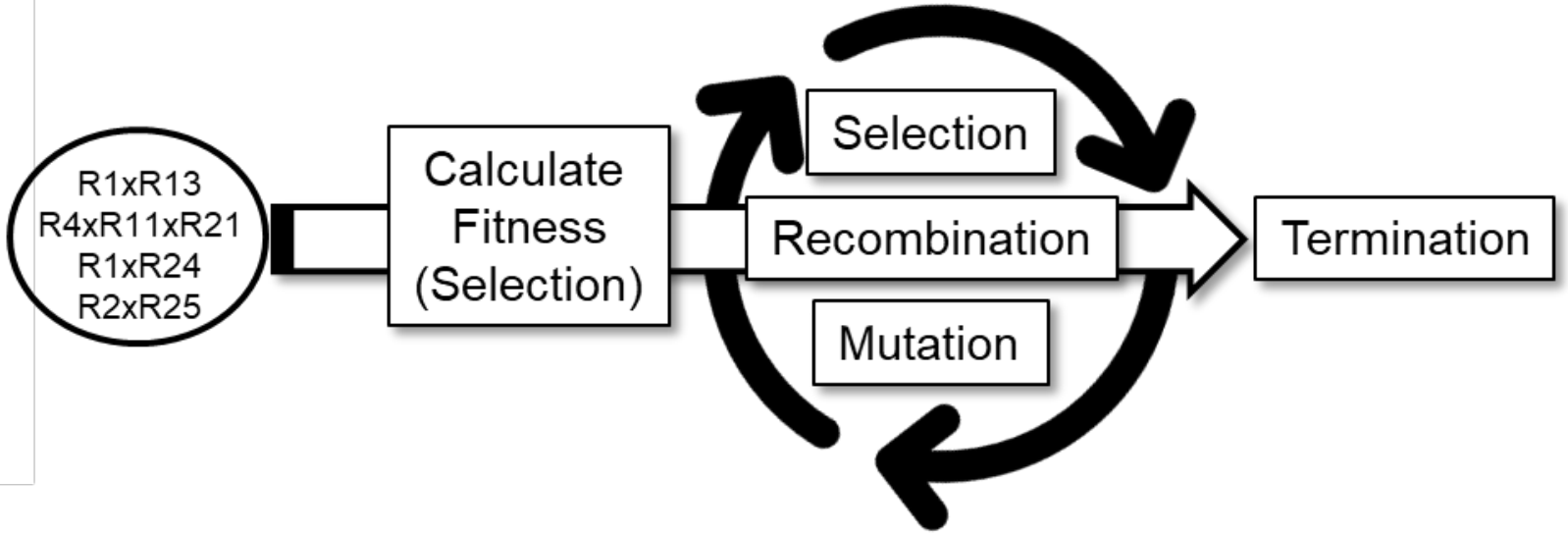
10k x 5k → 50mio

10k x 5k x 600 → 3bn

Eisenhuth P. et al. (arXiv:2404.17329v1)

Rosetta Evolutionary Ligand – REvoLd: Ultra-Large Library Screening with an Evolutionary Algorithm





$A + B \rightarrow P_{AB}$

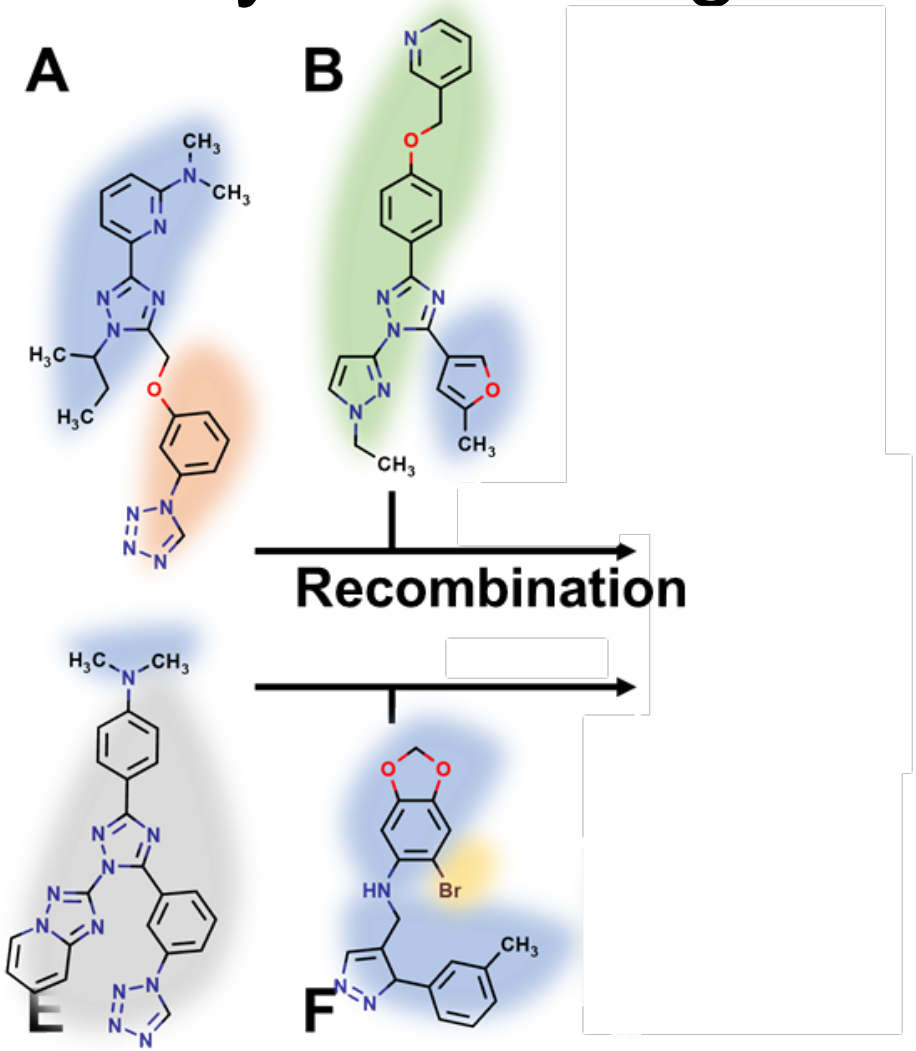
$A + B + C \rightarrow P_{ABC}$

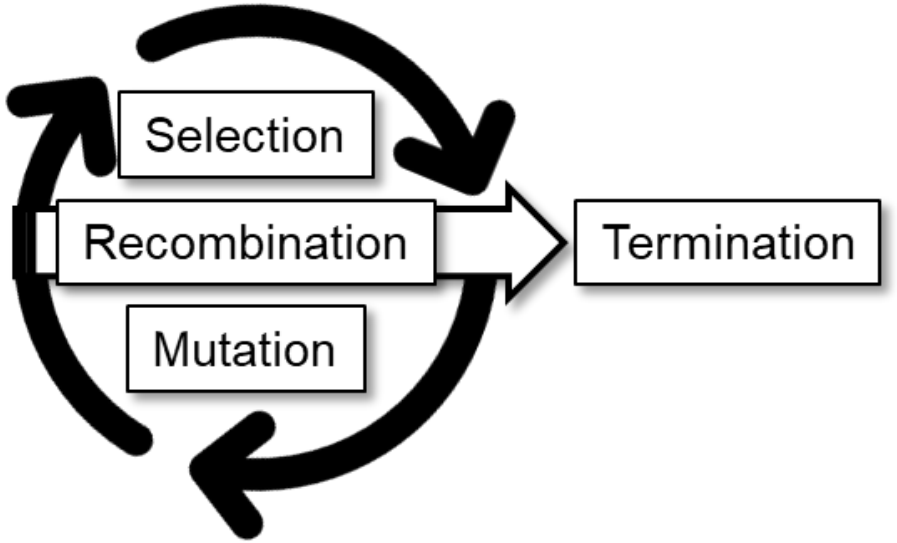
10k x 5k → 50mio

10k x 5k x 600 → 3bn

Eisenhuth P. et al. (arXiv:2404.17329v1)

Rosetta Evolutionary Ligand – REvoLd: Ultra-Large Library Screening with an Evolutionary Algorithm





$A + B \rightarrow P_{AB}$

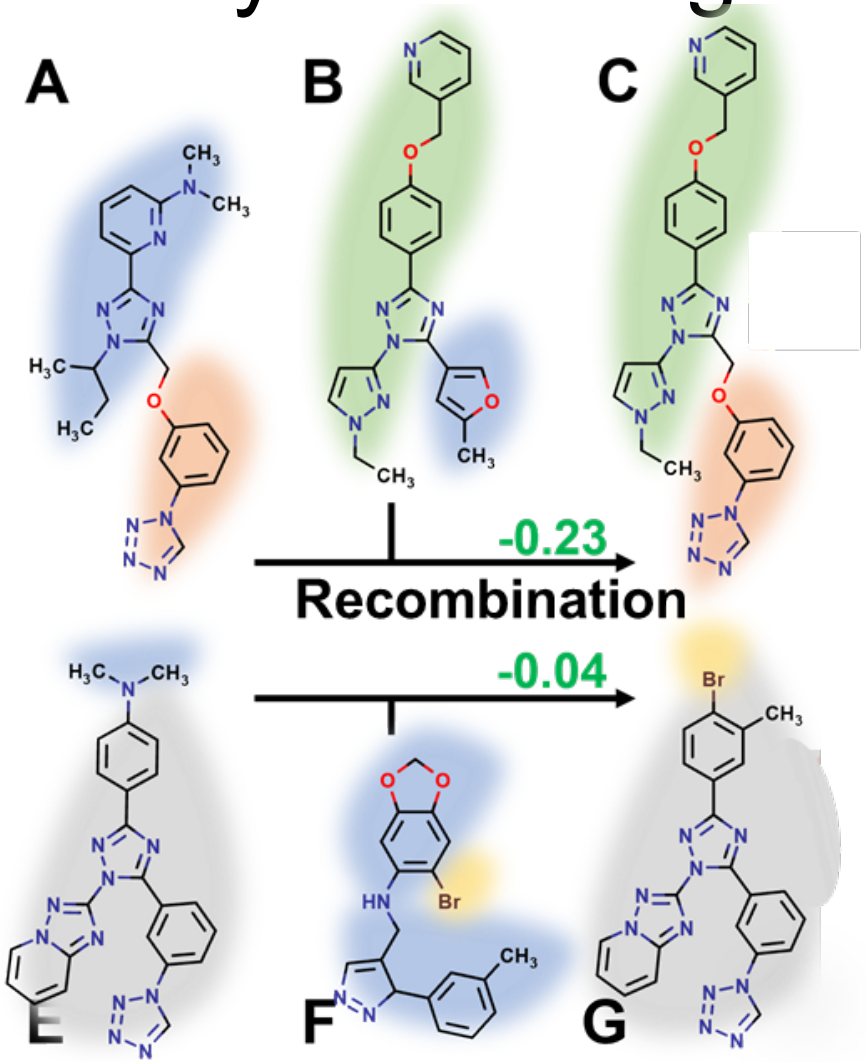
$A + B + C \rightarrow P_{ABC}$

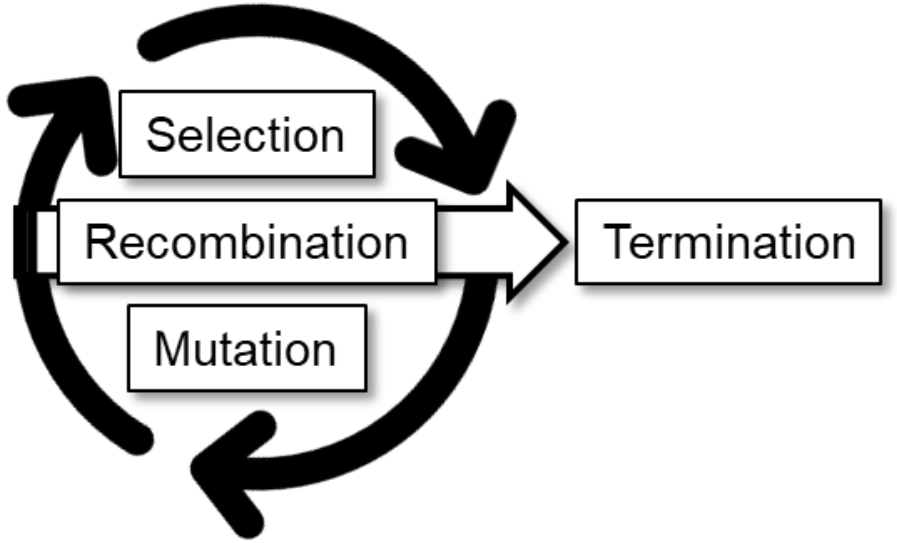
10k x 5k → 50mio

10k x 5k x 600 → 3bn

Eisenhuth P. et al. (arXiv:2404.17329v1)

Rosetta Evolutionary Ligand – REvoLd: Ultra-Large Library Screening with an Evolutionary Algorithm





$A + B \rightarrow P_{AB}$

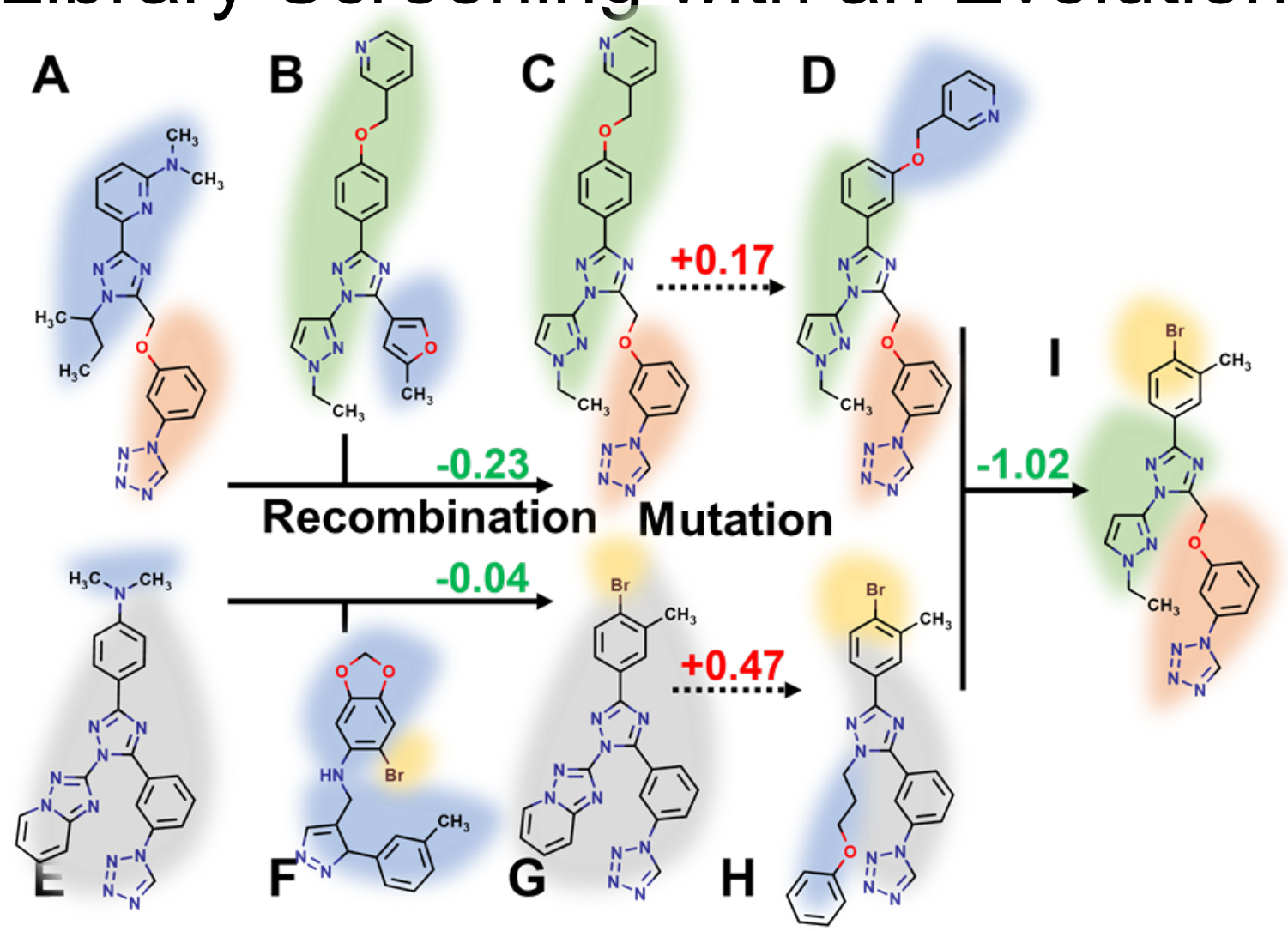
$A + B + C \rightarrow P_{ABC}$

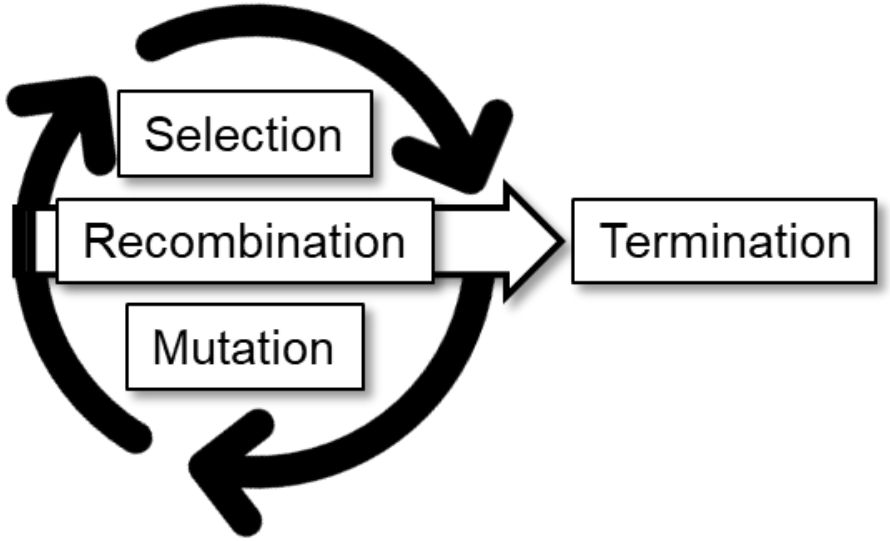
10k x 5k → 50mio

10k x 5k x 600 → 3bn

Eisenhuth P. et al. (arXiv:2404.17329v1)

Rosetta Evolutionary Ligand – REvoLd: Ultra-Large Library Screening with an Evolutionary Algorithm





$A + B \Rightarrow P_{AB}$

$A + B + C \Rightarrow P_{ABC}$

10k x 5k → 50mio

10k x 5k x 600 → 3bn

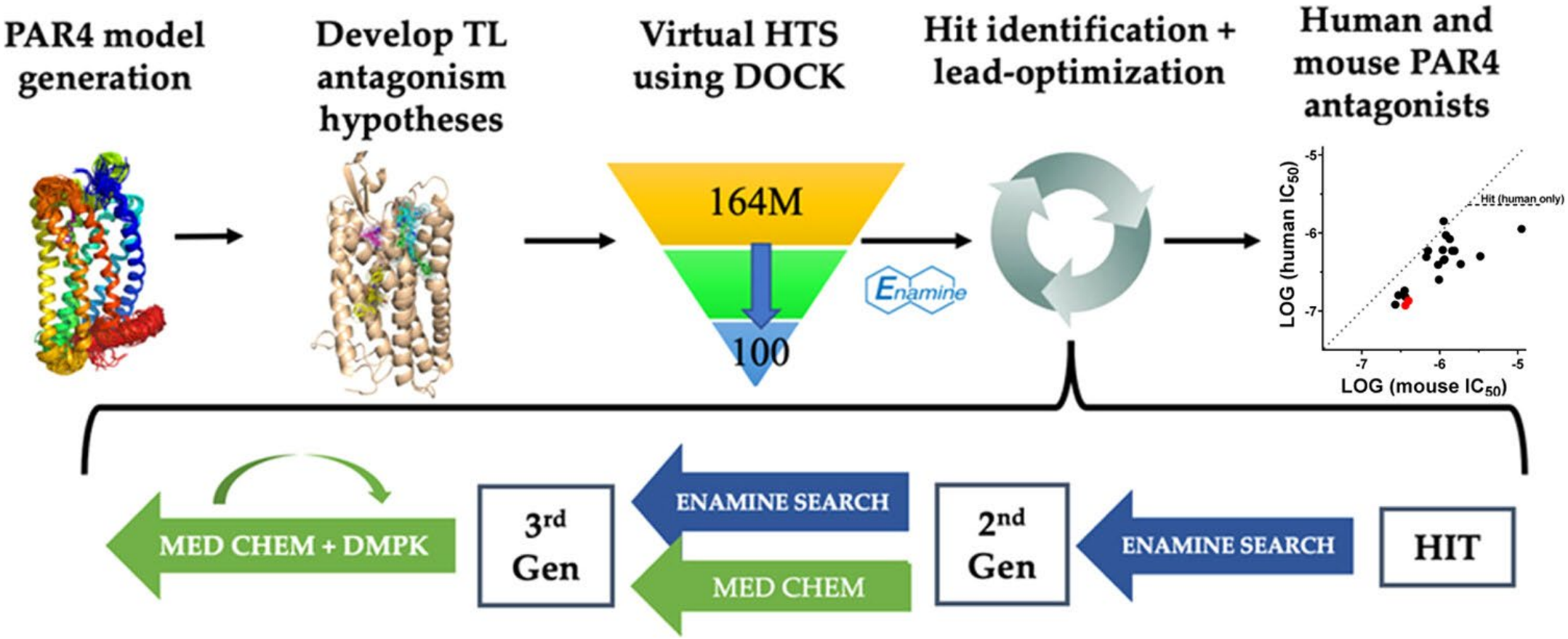
Eisenhuth P. et al. (arXiv:2404.17329v1)

Ultra-Large Library Screening Projects Using Computational Models with Increasing Library Sizes

meilerlab

Target	Structure	Database size	Hits/tested	Comment	Med. Chem
GPCR PAR4 with DOCK	Homology model	164 m	1/92	1 prelim hit (5 µM), optimized to 100 nm *Smith, S. T., …, M., Shoichet, B. K., Lindsley, C. W., Meiler, J., & Hamm, H. E. (2024). ACS Pharmacol Transl Sci, 7(4), 1086*	Completed
aGPCR GPR114	Homology model	303 m	5/88	5 prelim hits (7 nM – 10 µM)	Ongoing
WDR40 LLRK2 (CACHE)	Crystal structure	20 bn	1/102 4/37 (follow-up)	1 initial validated hit, 4 follow-up hits (KD 34 µM – 290 µM)	Ongoing
GPCR MC4R	CryoEM structure	20 bn	13/82	QSAR model as filter, activity currently investigated	Planned
GPCR Y4R	CryoEM structure	20 bn	1/14	1 inital validated hit (EC_{50} ~10 µM, first small molecule agonist)	Ongoing
GPCR GPR3	AlphaFold2 model	31 bn	9/91	9 hits (not reported EC_{50} values)	Planned
GPCR CMKLR1	Homology model +CryoEM structure	31 bn	(5)/97	5 prelim hit, currently investigation	Planned
aGPCR GPR110	CryoEM structure	20 bn	(3)/82	3 prelim hits, currently investigated	Planned

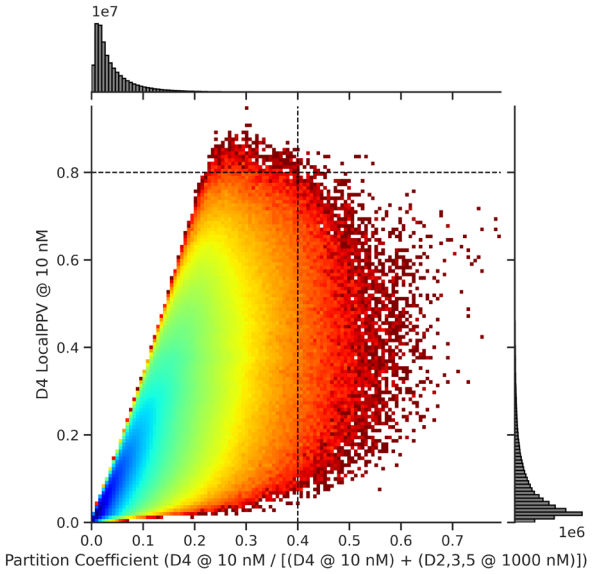
Ultra-Large Library Screen Identifies new Class of Antagonists Targeting PAR4

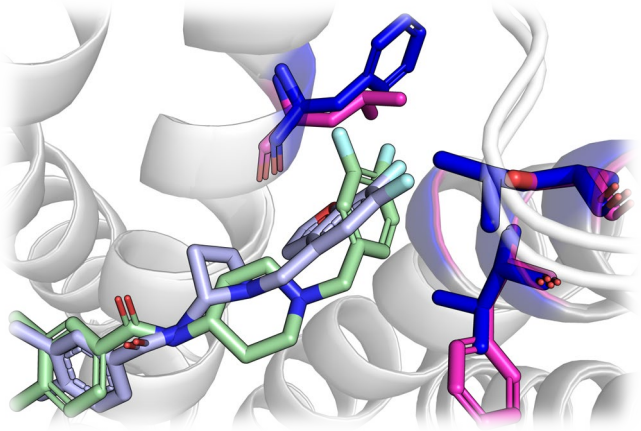


Smith, S. T., Cassada, J. B., Von Bredow, L., Erreger, K., Webb, E. M., Trombley, T. A., Kalbfleisch, J. J., Bender, B. J., Zagol-Ikapitte, I., Kramlinger, V. M., Bouchard, J. L., Mitchell, S. G., Tretbar, M., Shoichet, B. K., Lindsley, C. W., Meiler, J.*, & Hamm, H. E.* (2024). Discovery of Protease-Activated Receptor 4 (PAR4)-Tethered Ligand Antagonists Using Ultralarge Virtual Screening. *ACS Pharmacol Transl Sci, 7(4), 1086-1100.*

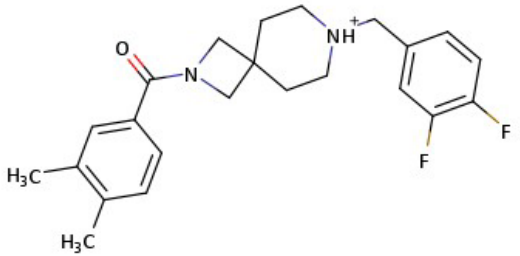
Computer-Aided Design and Biological Evaluation of Diazaspirocyclic and Selective D4R Antagonists

meilerlab

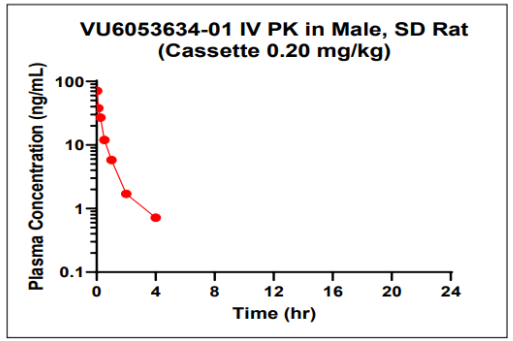




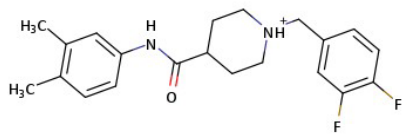
DRD2 DRD4



Derivatives by
Caleb A. Jones, Ph.D.



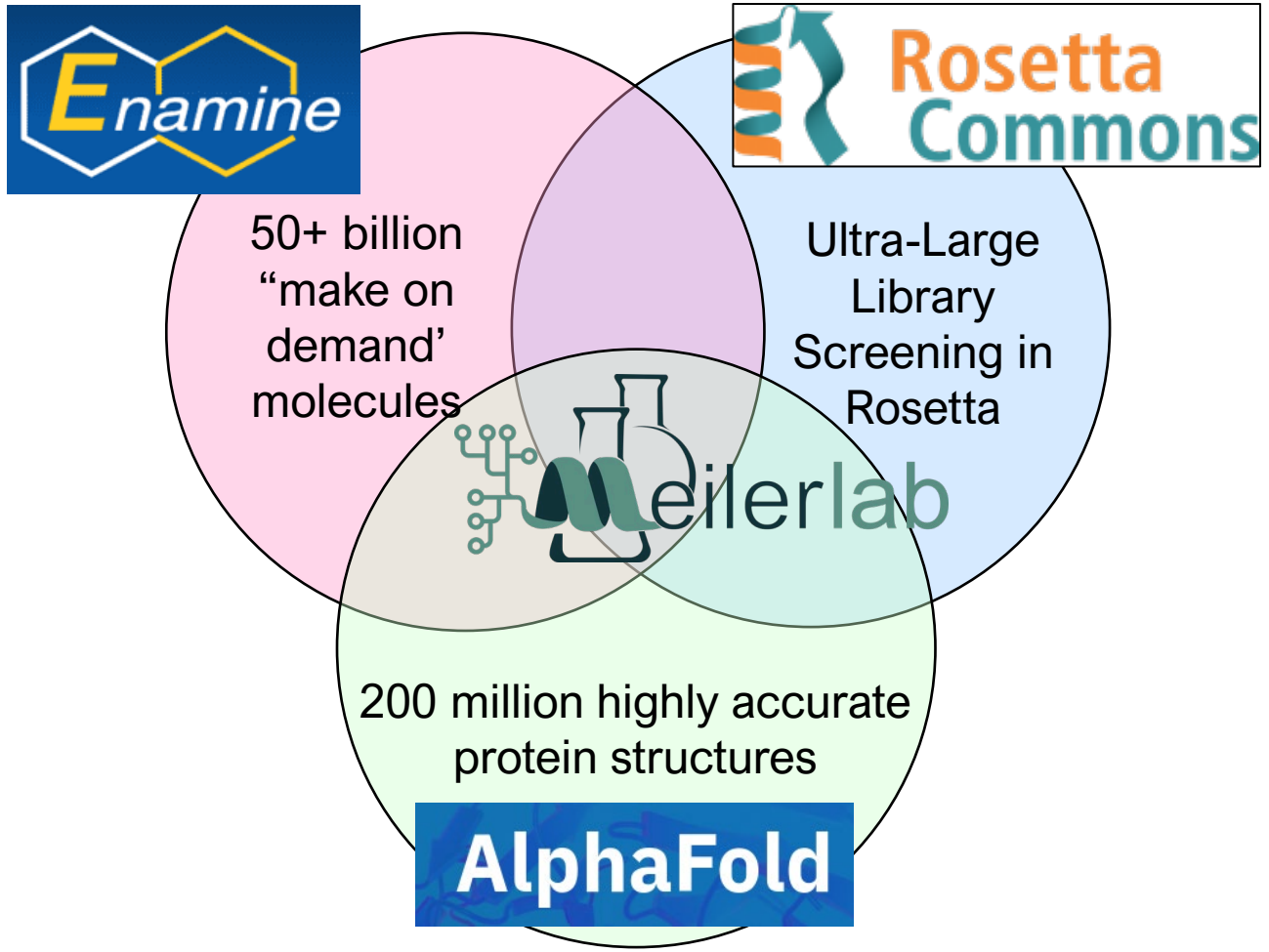
VU6053634-01 (MW: 384.2; LLOQ: 1.0 ng/ml)			
Time	Plasma Concentration (ng/mL)		
(hr)	1	2	Mean
0.033	71.5	-	-
0.117	37.6	-	-
0.266	27.4	-	-
0.533	12.3	-	-
1	5.83	-	-
2	1.73	-	-
4	0.72	-	-
7	BLQ	-	-
24	BLQ	-	-
Elim. $t_{1/2}$ (hr)	1.05	-	-
MRT (hr)	0.79	-	-
CL_p (mL/min/kg)	116	-	-
V_{ss} (L/kg)	5.52	-	-
AUC (hr*ng/ml)	28.7		



Jones, C. A. H., Brown, B. P., Schultz, D. C., Engers, J., Kramlinger, V. M., Meiler, J.*, & Lindsley, C. W.* (2024). Computer-Aided Design and Biological Evaluation of Diazaspirocyclic D(4)R Antagonists. *ACS Chem Neurosci, 15(12), 2396-2407.*

Virtual high-throughput screen → Structure-based computational SAR exploration → Experimental focused library design → Lead Optimization

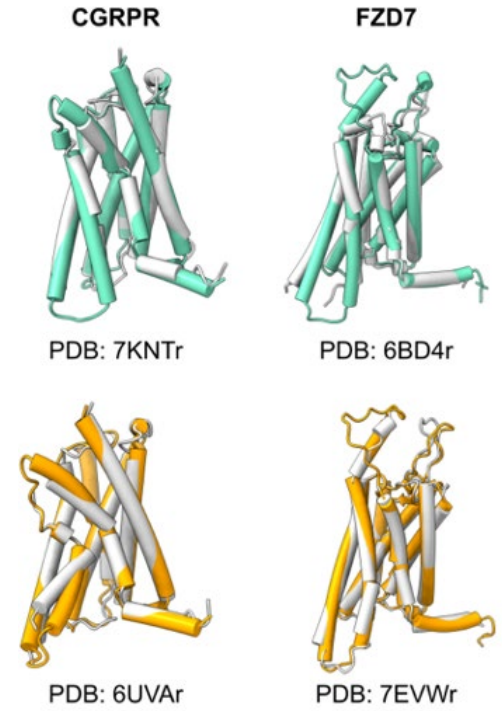
A Paradigm Shift in Therapeutic Development in Academia: ***in silico* First**



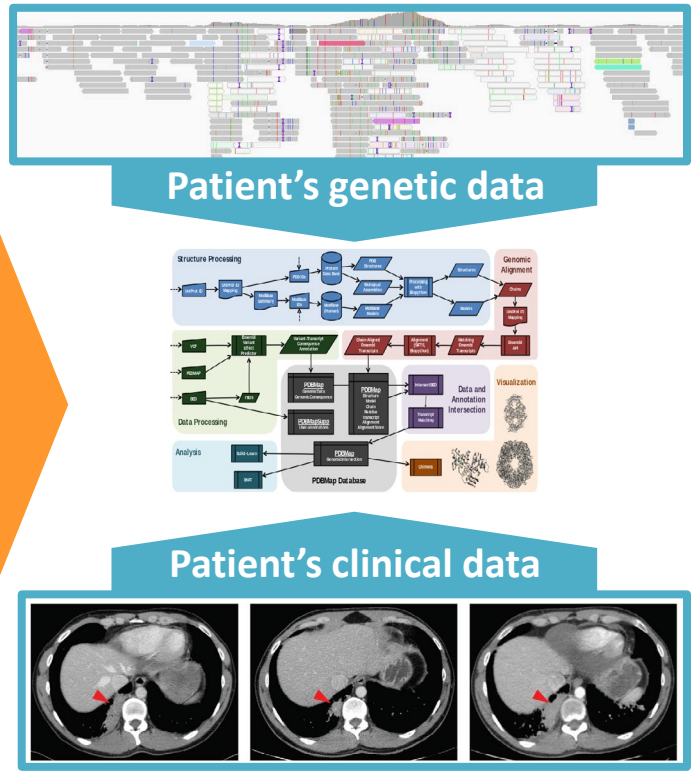
Precision Molecular Medicine with Artificial Intelligence Algorithms

meilerlab

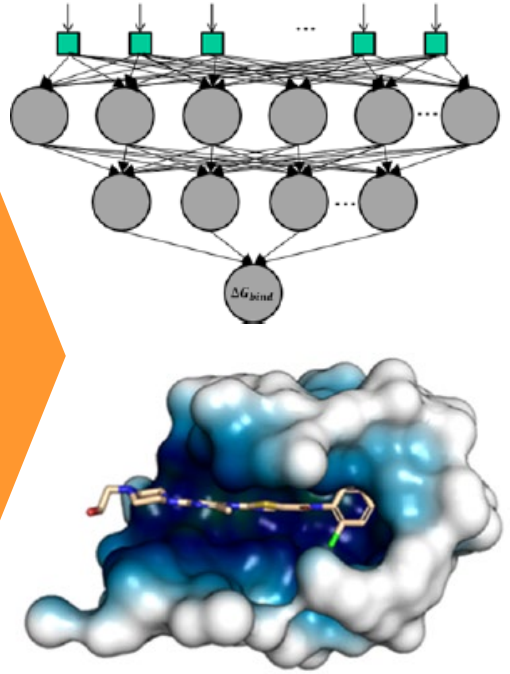
3D structure of protein receptors in different functional **conformations predicted by AI** methods (Alphafold, Rosetta)



Disease-causing mutations detected in the genome of an individual patient are used to **personalize the model**



Computational **screening** of 50 billion candidate drug molecules against the **personalized protein model** using AI methods.



Del Alamo, Sala, McHaourab, **Meiler**: "Sampling alternative conformational states of transporters and receptors with AlphaFold2"; *Elife*; **2022** || Brown, Mendenhall, Geanes, **Meiler**; "General Purpose Structure-Based Drug Discovery Neural Network Score Functions with Human-Interpretable Pharmacophore Maps"; *J Chem Inf Model*; **2021** || Schuss, Vu, Schubert, Du, Mishra, Tough, Stichel, Weaver, Emmitte, Cox, **Meiler**, Beck-Sickinger: "Highly Selective Y4 Receptor Antagonist Binds in an Allosteric Binding Pocket"; J Med Chem; **2021** || Gallant, Sheehan, Shaver, Bailey, Lipson, Chandramohan, Red Brewer, York, Kris, Pietenpol, Ladanyi, Miller, Ali, **Meiler**, and Lovly; "EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib"; *Cancer Discov*; **2015**

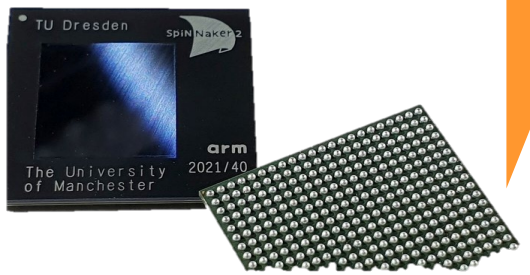
Personalized Molecular Medicine with AI Algorithms on SpiNNaker2

meilerlab

SpiNNaker2 Chip

Designed in the Human Brain Project

- SpiNNaker2 is the only machine that can support DNN models with 10^{14} parameters (i.e. brain-size)
- Unprecedented low energy bill and very fast compared to existing GPU hardware
- Built on most advanced European semiconductor technology (22FDX)
- Outperforming Nvidia, Intel Loihi, Google TPU on **real-time AI**

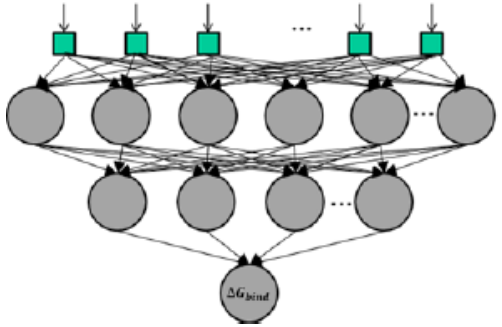


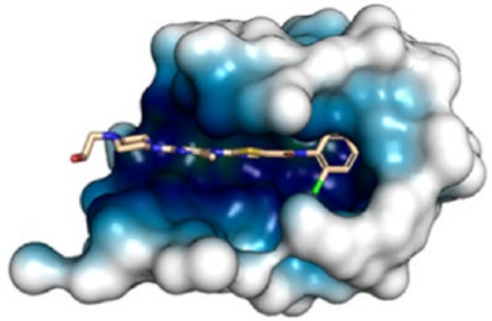
SpiNNcloud Supercomputer

World-largest real-time AI platform

- 10,000,000 processors on 70,000 chips in 16 racks
- 3 PFLOPS CPU / 0.4 ExaOPS in AI accelerator
- Deployment: 01/2023

Computational **screening** of 21 billion candidate drug molecules against the **personalized protein model** using AI methods.

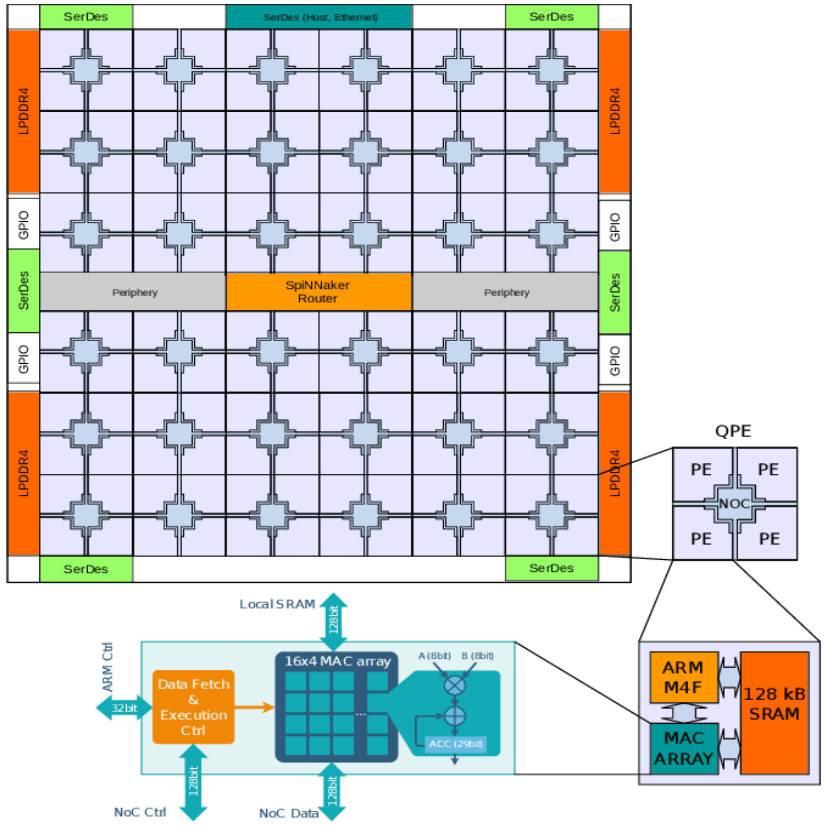




Yan, Stewart, Choo, Vogginger, Partzsch, Höppner, Kelber, Eliasmith, Furber, and **Mayr**: "Comparing Loihi with a SpiNNaker 2 prototype on low-latency keyword spotting and adaptive robotic control"; Neuromorph. Comput. Eng. **2021** || Höppner, Yan, Dixius, Scholze, Partzsch, Stolba, Kelber, Vogginger, Neumärker, Ellguth, Hartmann, Schiefer, Hocker, Walter, Liu, Garside, Furber, **Mayr**: "The SpiNNaker 2 Processing Element Architecture for Hybrid Digital Neuromorphic Computing"; preprint arXiv:2103.08392 [cs.AR]; **2021**

The SpiNNaker2 Platform: Between GPUs and QC 10^{10} Cores and 10^{14} Connections

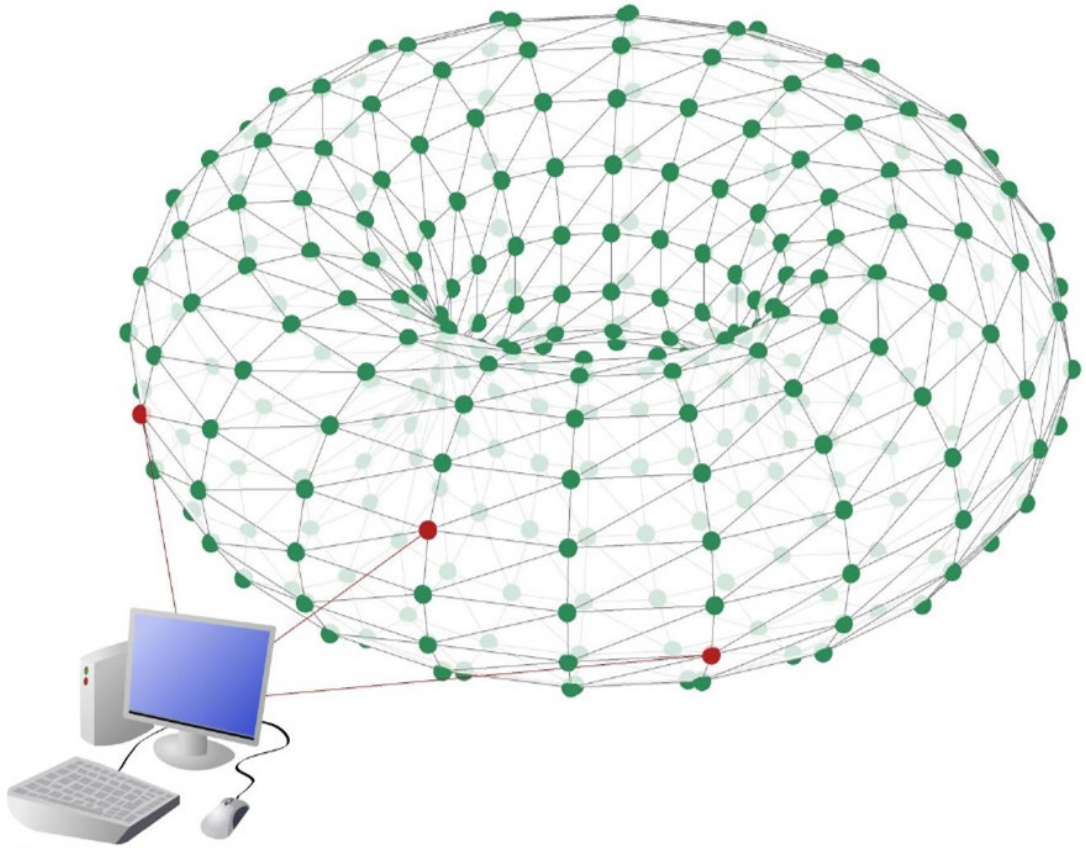
meilerlab



Yan, Stewart, Choo, Vogginger, Partzsch, Höppner, Kelber, Eliasmith, Furber, and **Mayr**: "Comparing Loihi with a SpiNNaker 2 prototype on low-latency keyword spotting and adaptive robotic control"; Neuromorph. Comput. Eng. **2021** || Höppner, Yan, Dixius, Scholze, Partzsch, Stolba, Kelber, Vogginger, Neumärker, Ellguth, Hartmann, Schiefer, Hocker, Walter, Liu, Garside, Furber, **Mayr**: "The SpiNNaker 2 Processing Element Architecture for Hybrid Digital Neuromorphic Computing"; preprint arXiv:2103.08392 [cs.AR]; **2021**

The SpiNNaker2 Platform: Between GPUs and QC 10^{10} Cores and 10^{14} Connections

meilerlab



Yan, Stewart, Choo, Vogginger, Partzsch, Höppner, Kelber, Eliasmith, Furber, and **Mayr**: “Comparing Loihi with a SpiNNaker 2 prototype on low-latency keyword spotting and adaptive robotic control”; Neuromorph. Comput. Eng. **2021** || Höppner, Yan, Dixius, Scholze, Partzsch, Stolba, Kelber, Vogginger, Neumärker, Ellguth, Hartmann, Schiefer, Hocker, Walter, Liu, Garside, Furber, **Mayr**: “The SpiNNaker 2 Processing Element Architecture for Hybrid Digital Neuromorphic Computing”; preprint arXiv:2103.08392 [cs.AR]; **2021**

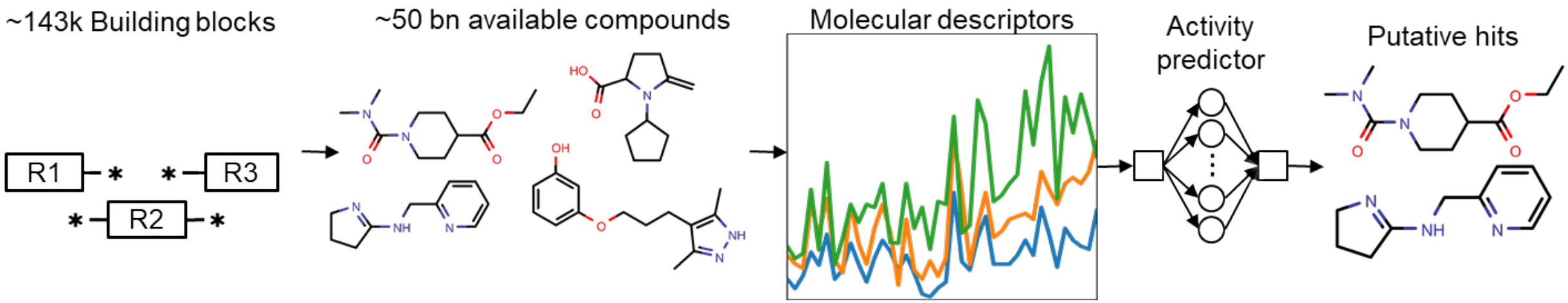
Advantages of SpiNNaker2 for Non-cognitive/non-AI Numerical Problems

meilerlab

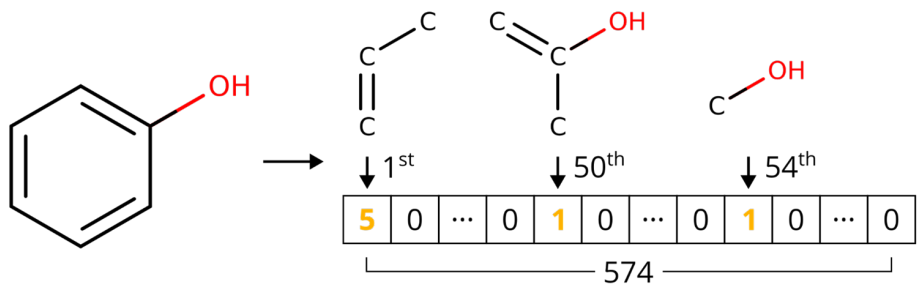
- Extreme parallelism of simple operations (think neurons...)
- (Search for) Sparse solutions in high-dimensional numerical rooms
- Stochastic computation/stochastic state representations
- Solving systems of locally coupled differential equations in a mesh/network topology (e.g. Neuron models, but also FEM and similar)

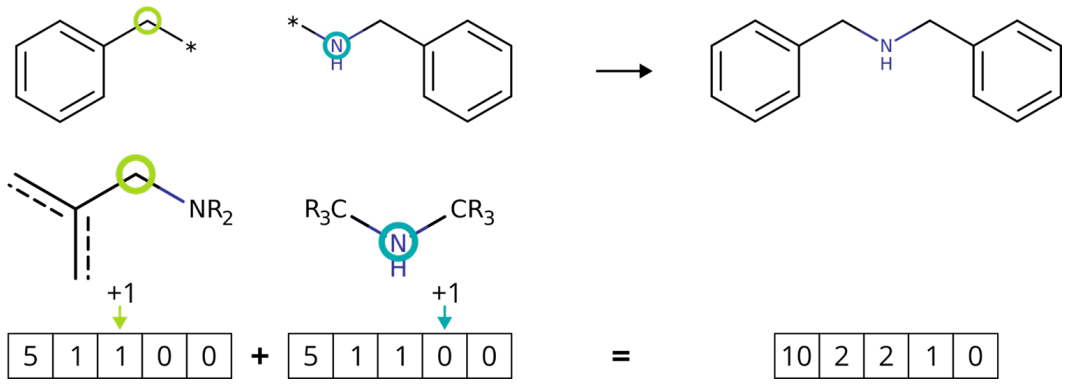
	HPC	SpiNNaker2	Quantum Computing
Parallelism	10^5 cores	10^{14} synaptic updates/msec	$>10^{25}$ quantum entanglements
Stochastic Computation	Only in software, 10^{10} stochastic decisions/sec	Hardware accelerators, 10^{17} stochastic decisions/sec	Inherent in Qubits, $>10^{30}$ stochastic decisions/sec
Sparsity in high-dimensional spaces	Not supported	Fully supported	Fully supported
FEM-type tesselations	10^5 elements, boundary condition updates µs to ms	10^7 elements in torus, boundary condition updates <10µs	Potentially very fast convergence, but tessalation limited to #Qubits: 10^2-10^3

Ultra-Large Library Screening on the SpiNNaker2 Platform



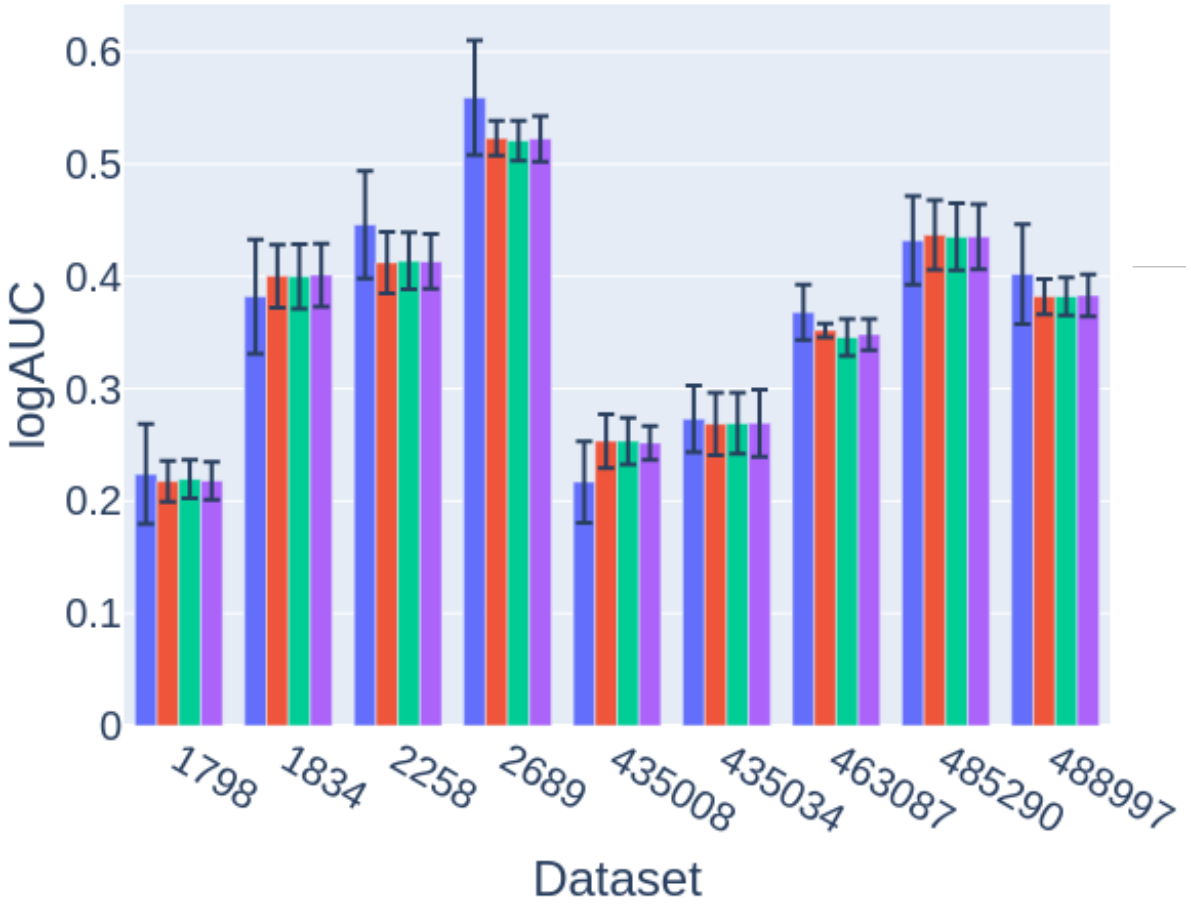
BCL::Mol2D descriptors from each atom's immediate neighborhood





Vu, O., Mendenhall, J., Altarawy, D., & Meiler, J. (2019). BCL::Mol2D-a robust atom environment descriptor for QSAR modeling and lead optimization. *J Comput Aided Mol Des, 33(5), 477-486. https://doi.org/10.1007/s10822-019-00199-8*

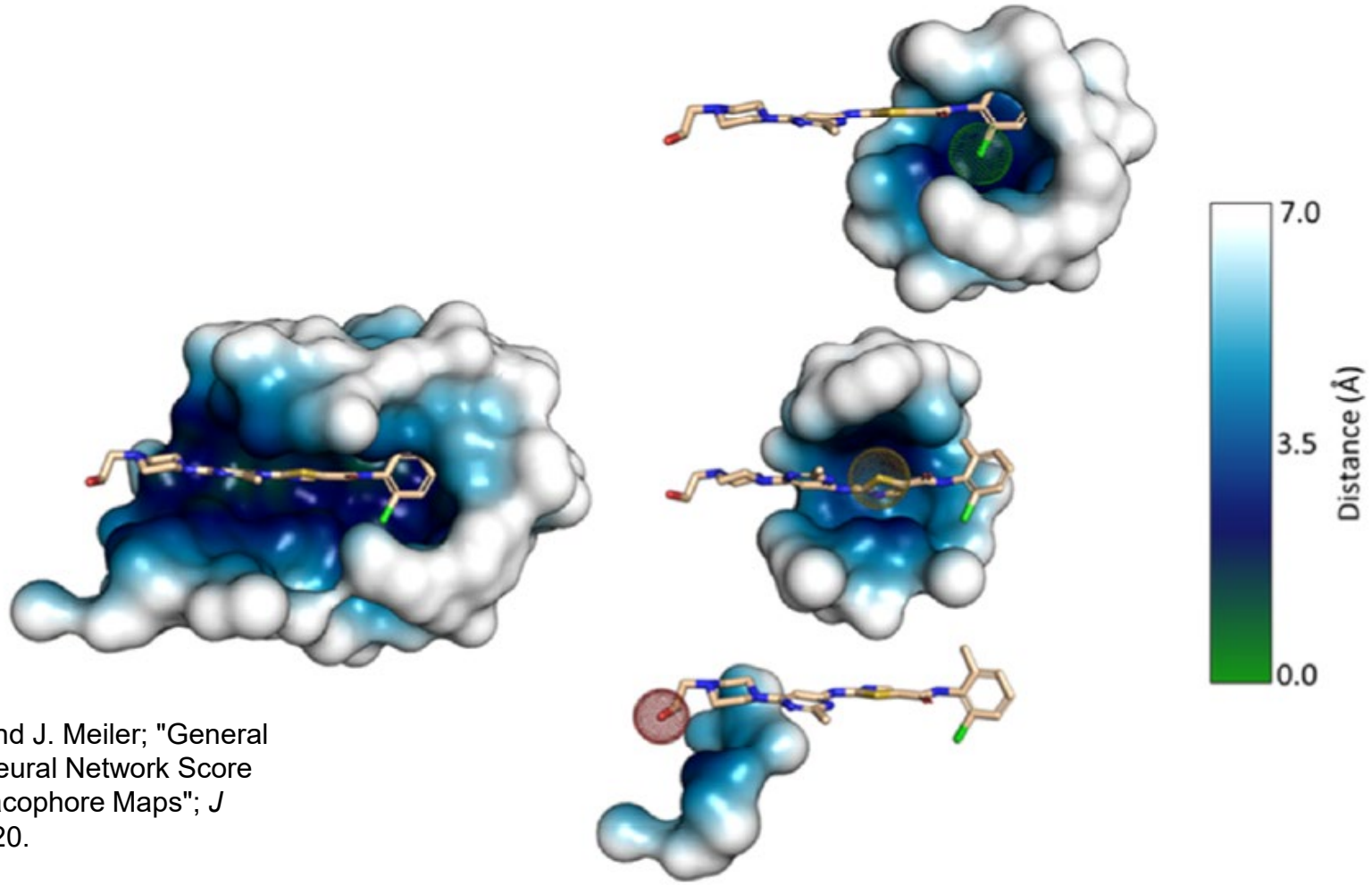
Benchmarking Computer-Aided Drug Discovery with a Well-Curated Dataset is Critical for AI



- Published performance of the Benchmark
- Float32 CPU/GPU implementation of the ANN
- Integer8 SpiNNaker2 implementation of the ANN
- Integer8 + BitShift SpiNNaker2 implementation

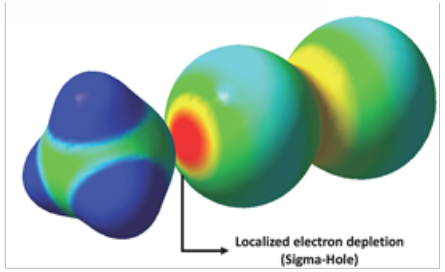
Butkiewicz, M., Lowe, E. W., Jr., Mueller, R., Mendenhall, J. L., Teixeira, P. L., Weaver, C. D., & Meiler, J. (2013). Benchmarking ligand-based virtual High-Throughput Screening with the PubChem database. *Molecules, 18(1), 735-756.* || Butkiewicz, M., Wang, Y., Bryant, S. H., Lowe, E. W., Jr., Weaver, D. C., & Meiler, J. (2017). High-Throughput Screening Assay Datasets from the PubChem Database. *Chem Inform, 3(1).*

General Purpose Structure-Based Drug Discovery Neural Network



B. P. Brown, J. Mendenhall, A. R. Geanes and J. Meiler; "General Purpose Structure-Based Drug Discovery Neural Network Score Functions with Human-Interpretable Pharmacophore Maps"; *J Chem Inf Model*; **2021**; Vol. 61 (2): p. 603-620.

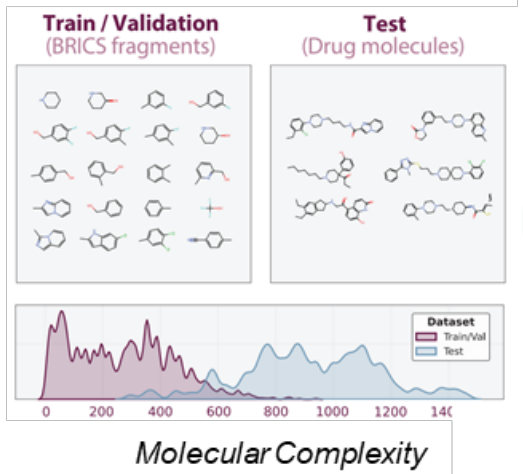
ElektroNN: Predicting Electron Densities of Drug-like Molecules from Small Fragments

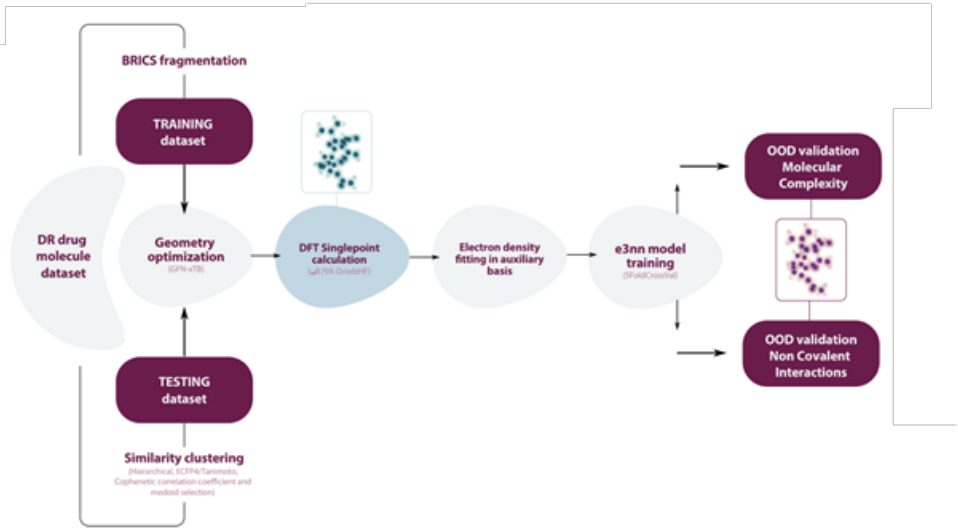


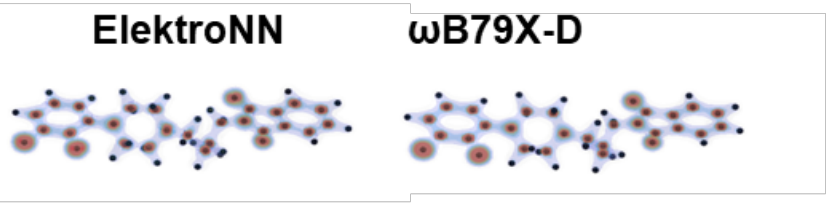
- Total Energy $E = f(\rho(x,y,z))$
- Electrostatic Potential $V = f(\rho(x,y,z))$
- (Hellmann-Feynman) Force $\vec{F} = f(\rho(x,y,z))$
- Interaction energy $E_{int} = f(\rho(x,y,z))$

  $?$
- Biological activity? $\approx f(\rho(x,y,z))$

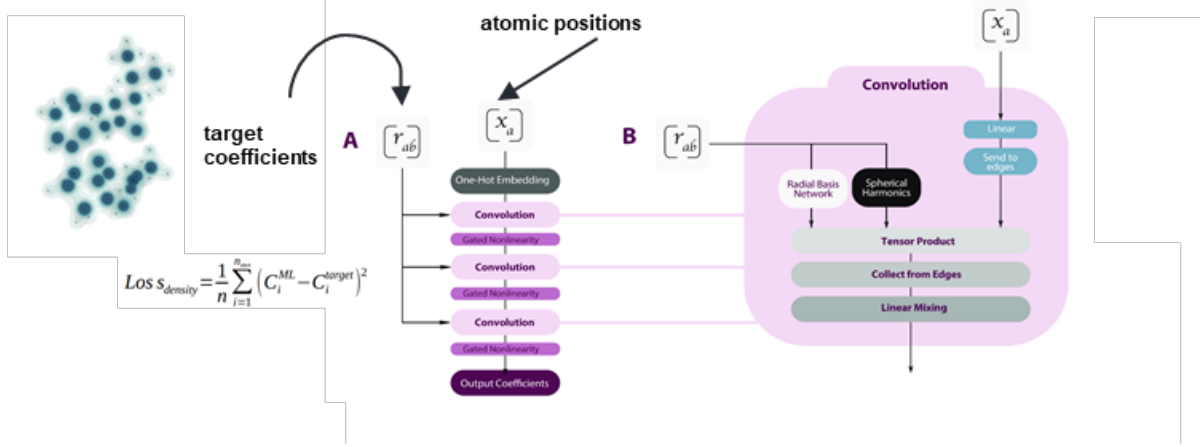
→ Descriptor design





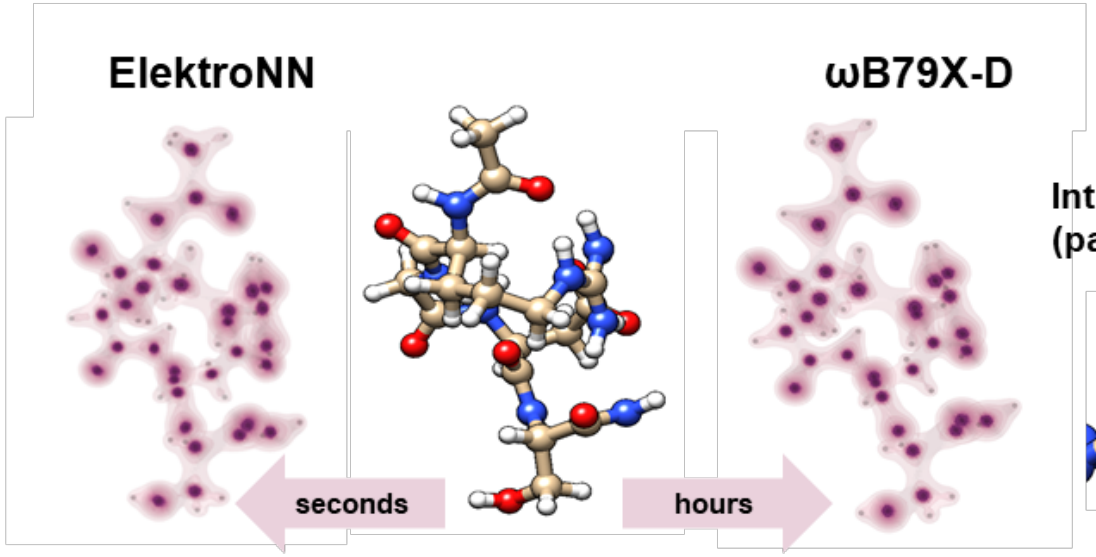


	Target Ele. Count (Mean)	Predicted Ele. Count (Mean)	Error per Atom (%)
Overall (all folds)	209.41	209.17	0.30

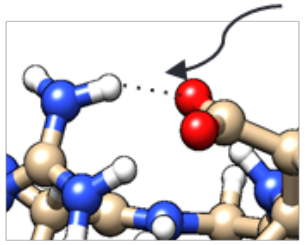


$$Loss_{density} = \frac{1}{n}\sum_{i=1}^{n_{ex}} \left(C_i^{ML} - C_i^{target}\right)^2$$

ElektroNN: Generalization to non-covalent Interactions in Ligands, Peptides, and Proteins

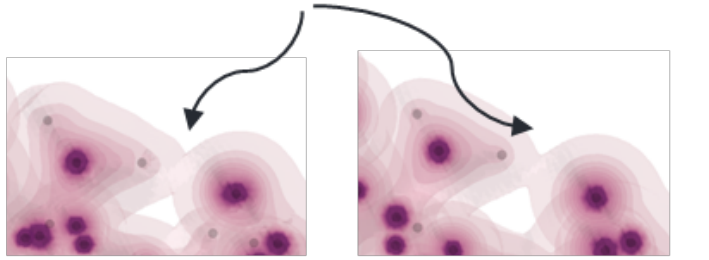


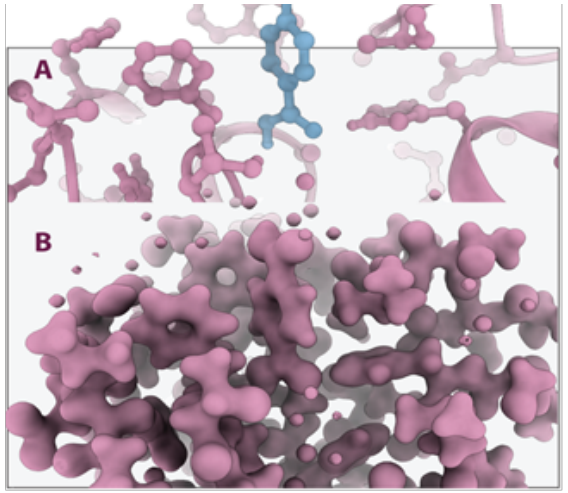
Intramolecular hydrogen bond
(partially covalent; "sharing of electrons")



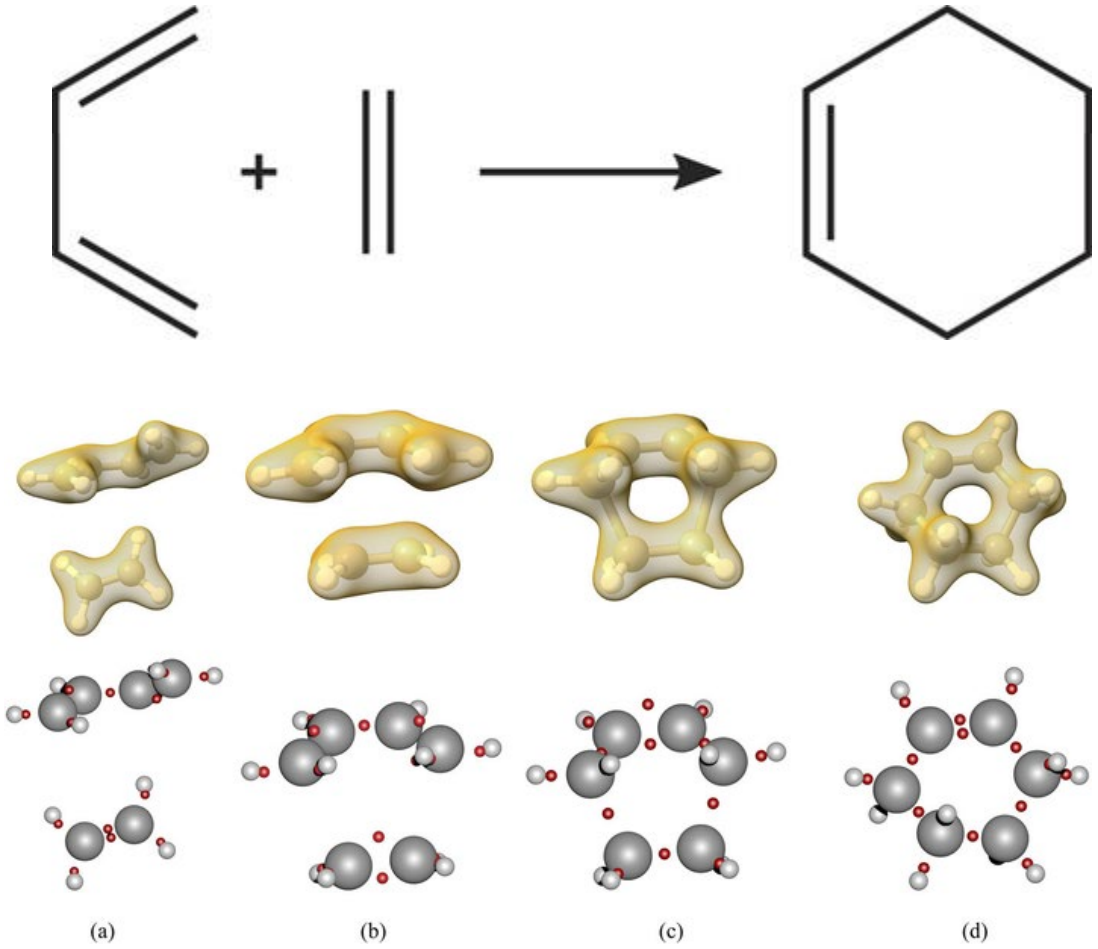
Peptide (5 conformations each)	Charge	Error per Atom (%)
VRN	1	0.30
KGD	0	0.35
DGEA	-2	0.29
APGL	0	0.25
RGD	0	0.30

Accumulation of electron density



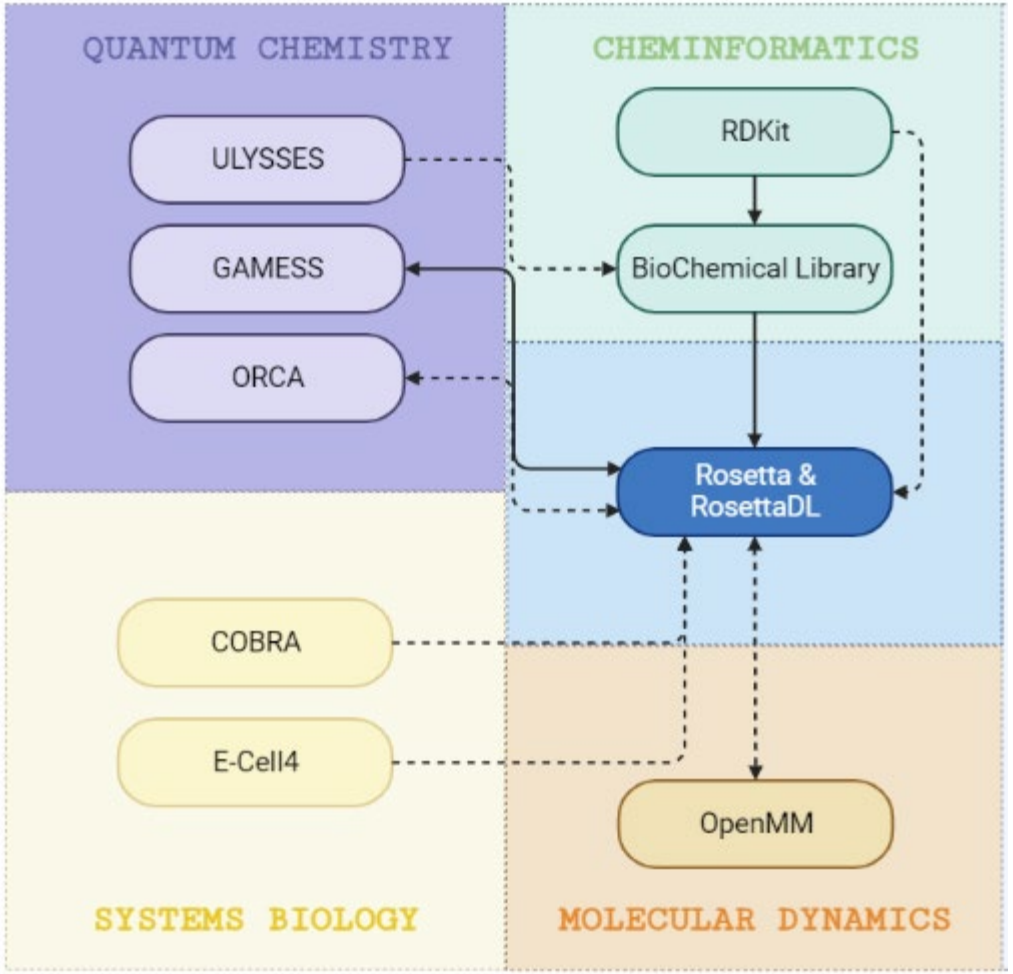


ElektroNN: Predicting Electron Density changes for the Diels-Alder Reaction



RosettaQM is coming soon …

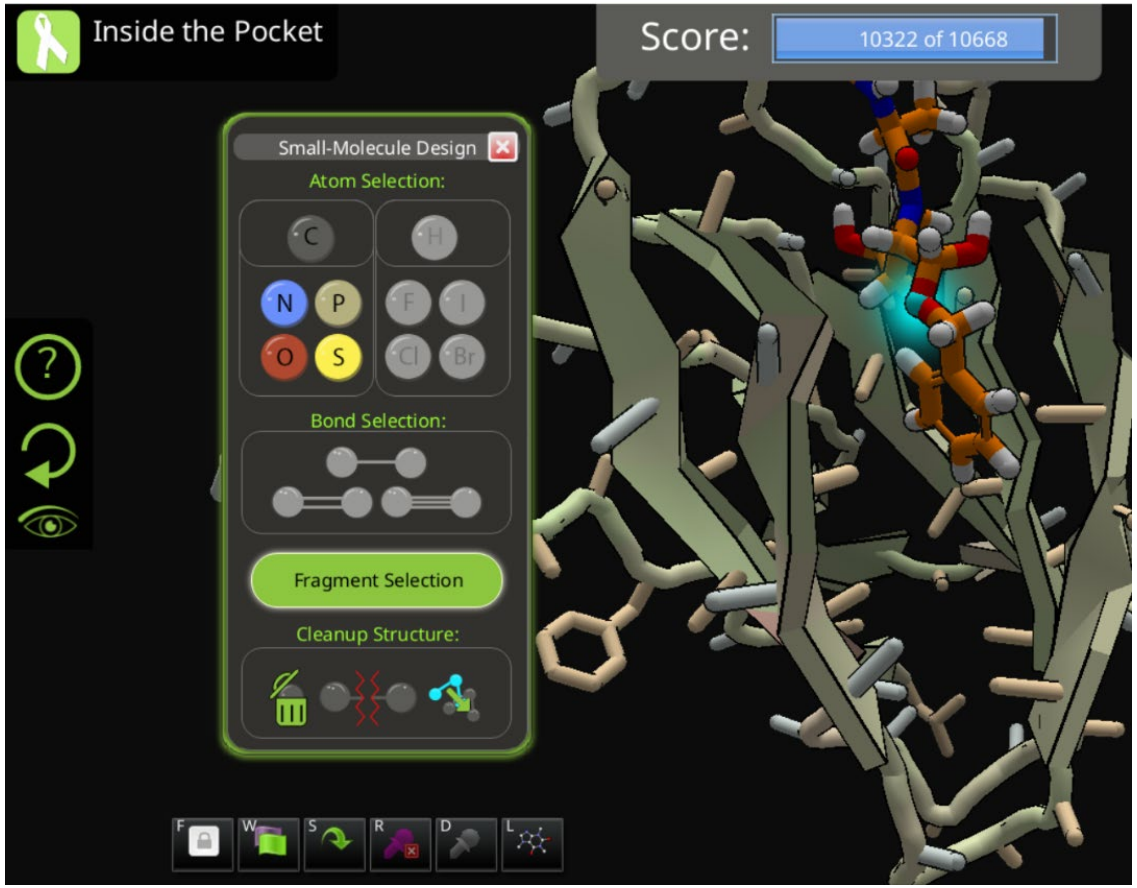
- Enzymes
- Non-canonical Amino Acids
- Covalent Inhibitors

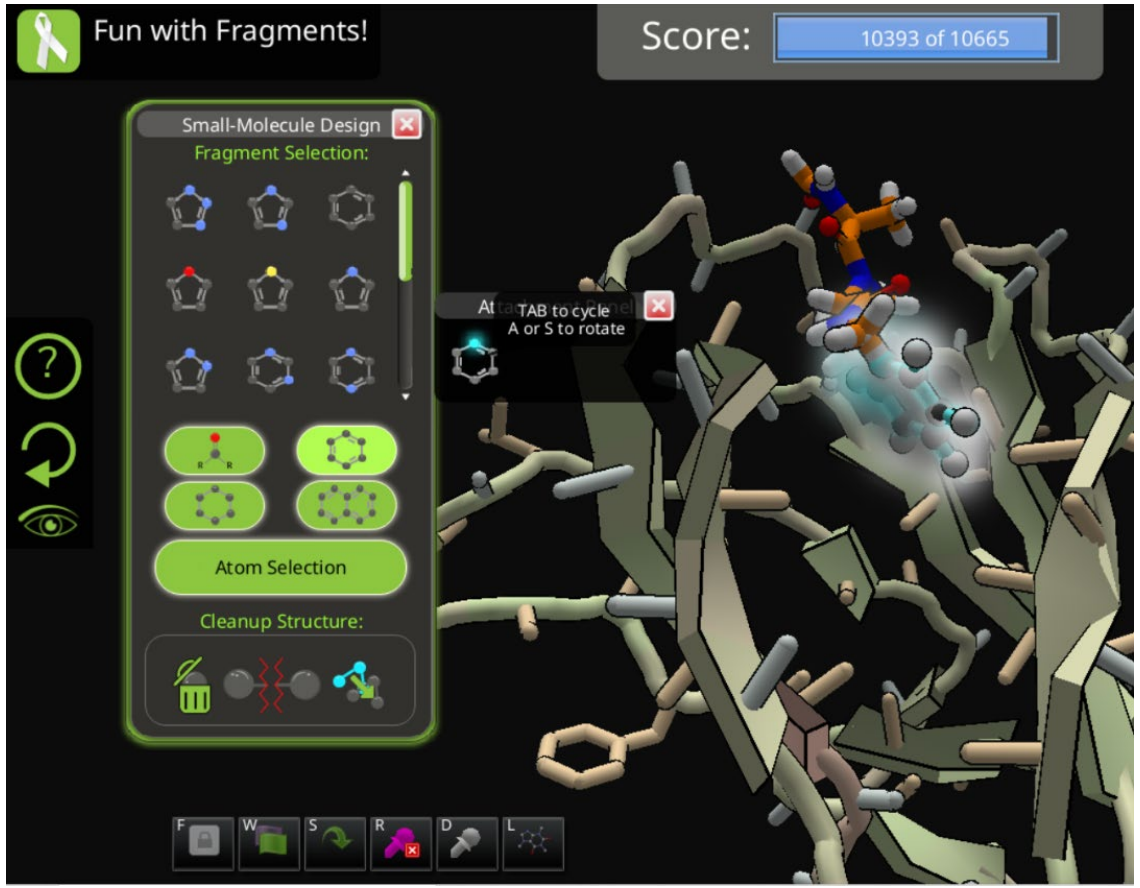


DrugIt: Crowd-Sourcing Drug Design via a Computer Game to Engage Citizen Scientists

meilerlab

- Players can place atoms and bonds

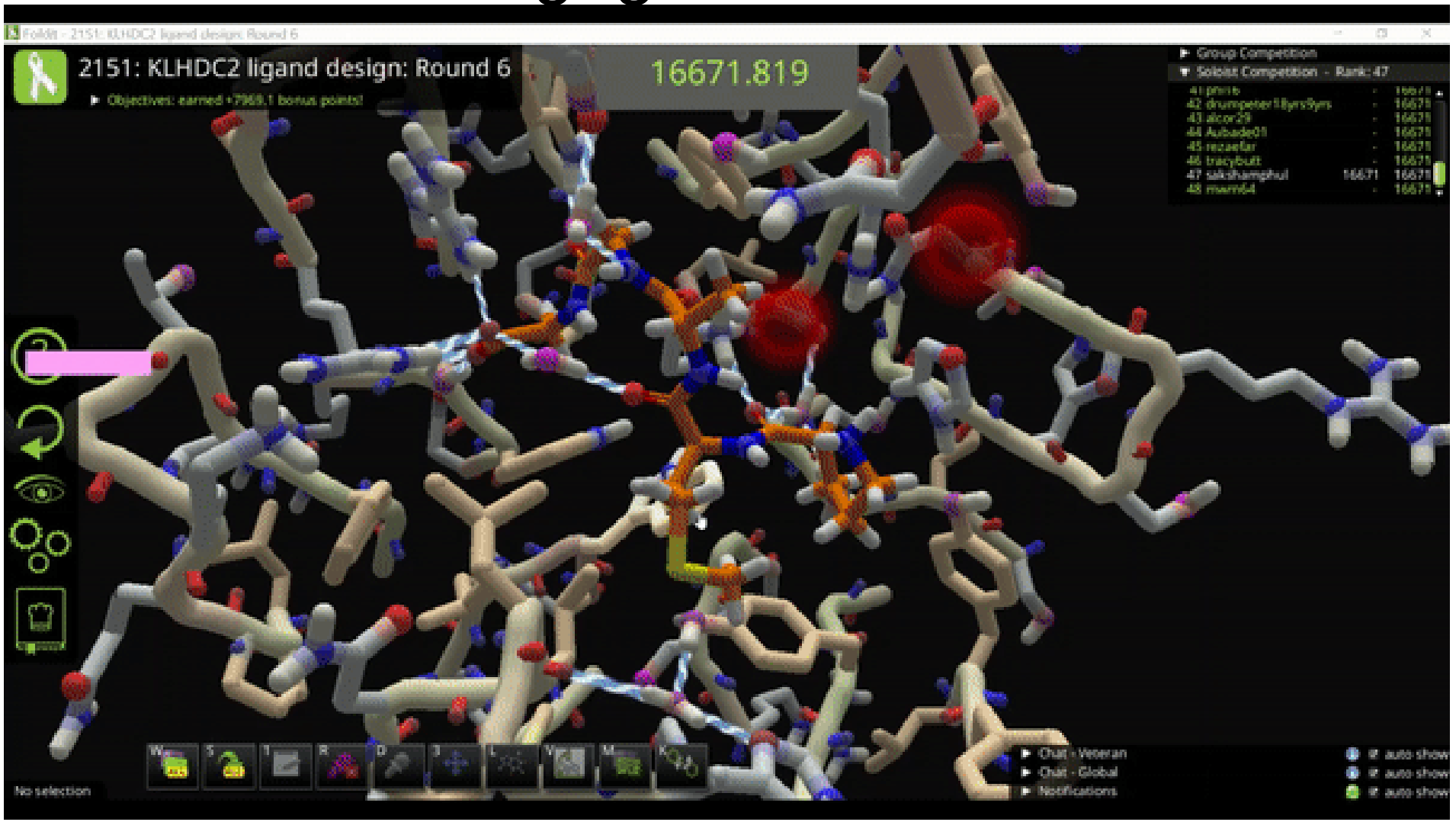




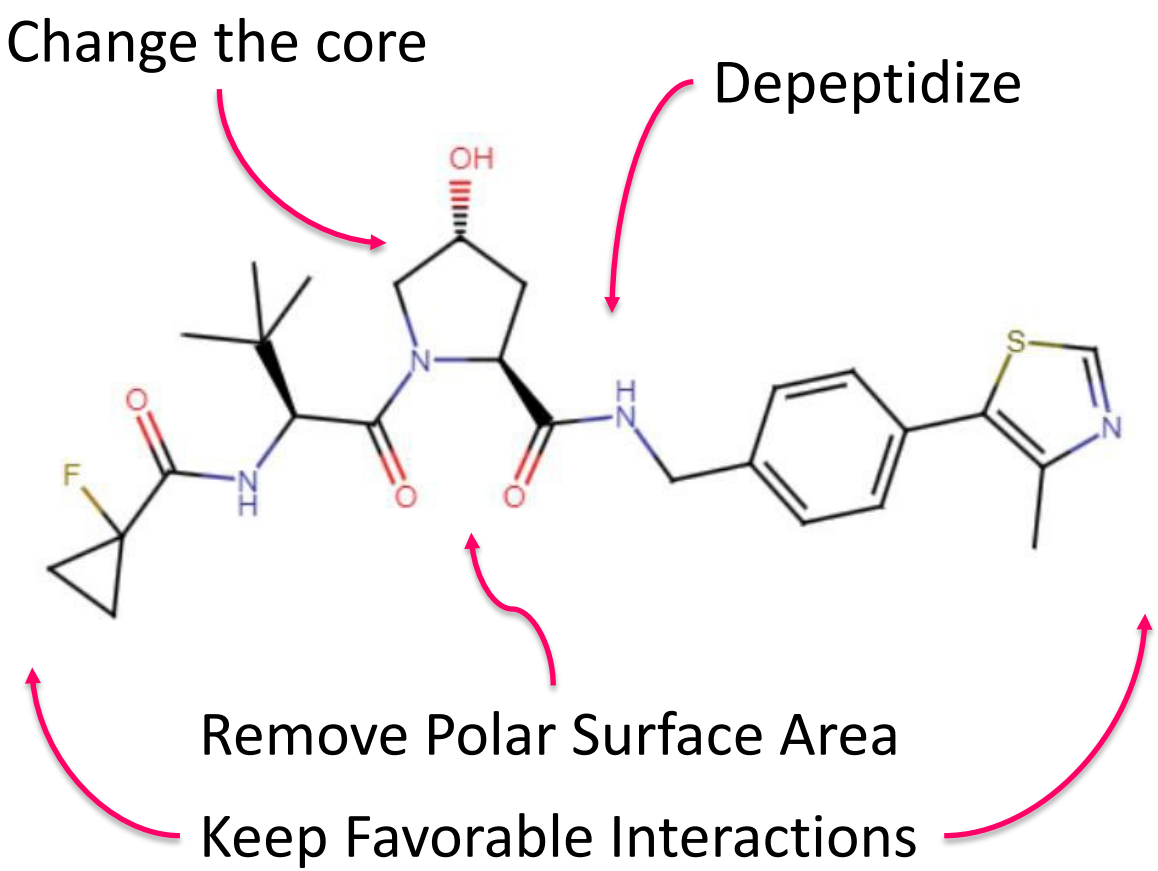
- Players can also select from list of fragments

DrugIt: Crowd-Sourcing Drug Design via a Computer Game to Engage Citizen Scientists

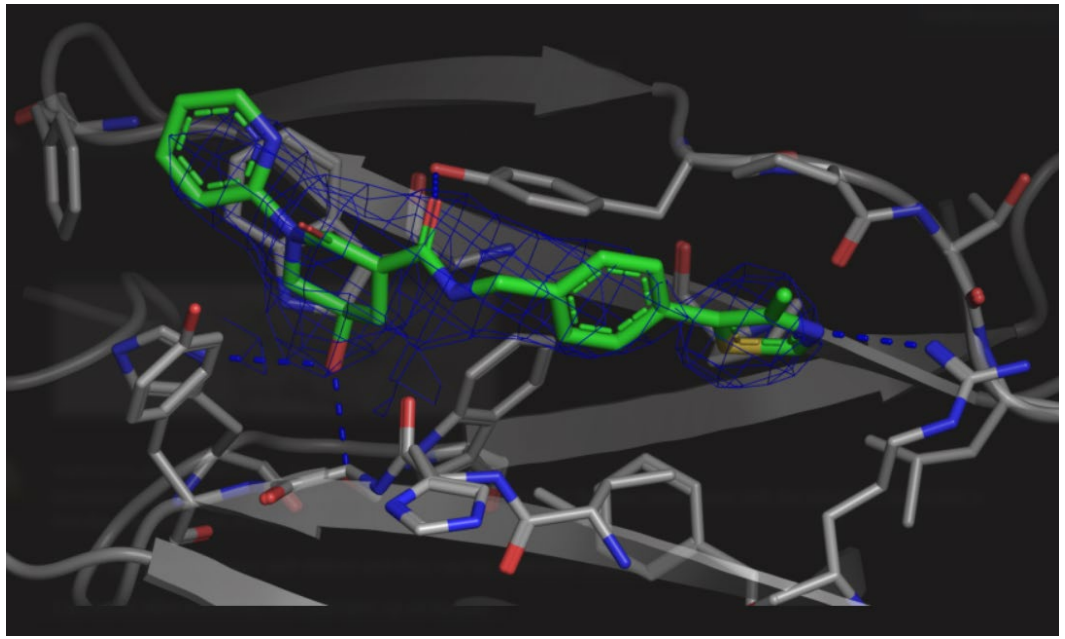
meilerlab



Drugit: Crowd-Sourced Development of E3 Ligase Binders with Boehringer Ingelheim



Cmpd	IC50	NMR Confirmation	Efflux Model
Parent	0.15 µM	✓	>10
1	258 µM	✓	<3



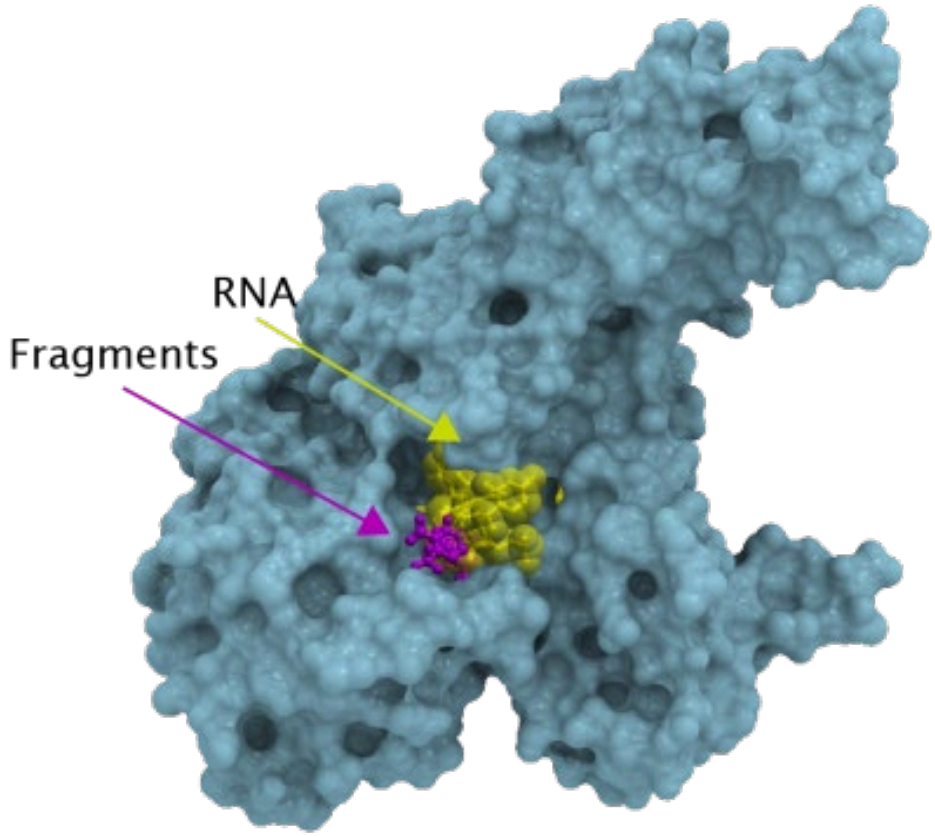
https://chemrxiv.org/engage/chemrxiv/article-details/64596c281ca6101a45015aa7

CACHE: Critical Assessment of Computational Hit-Finding Experiments

CACHE CHALLENGE

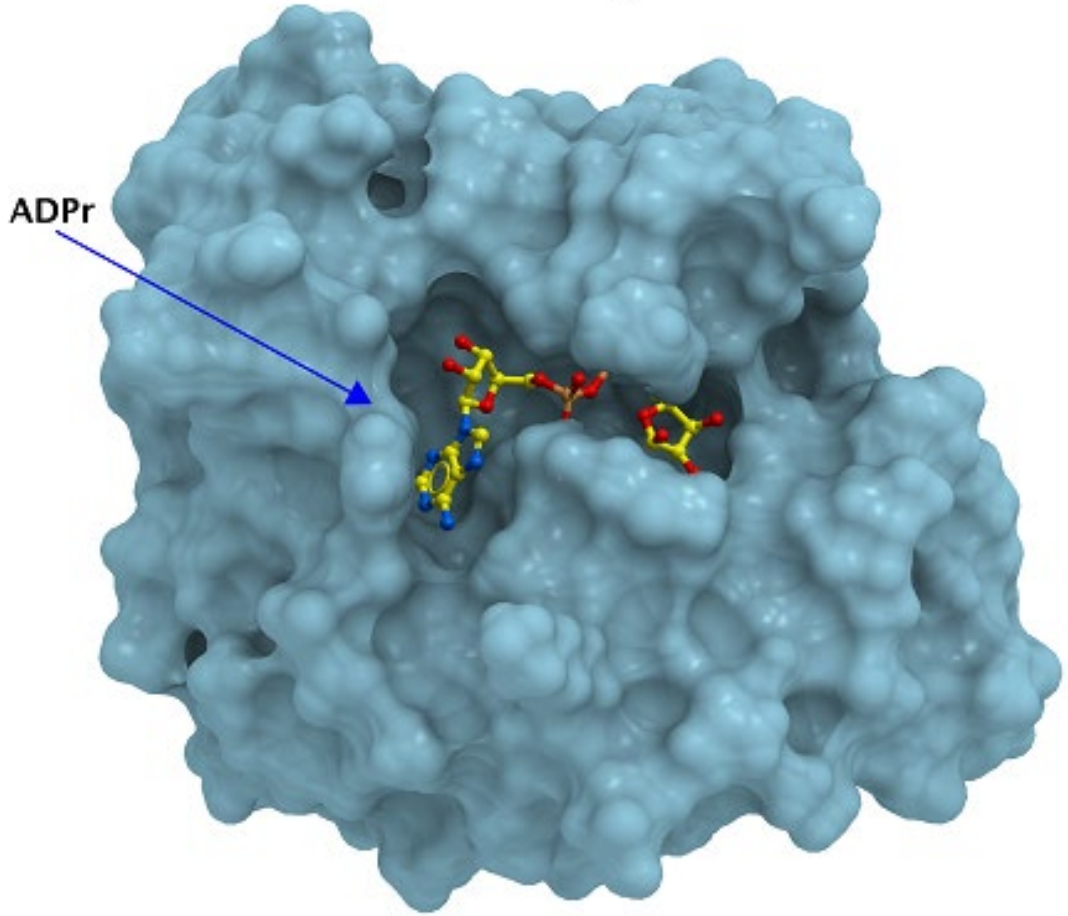
meilerlab

CACHE Challenge #2



Crystal structures of SARS-CoV-2 NSP13/helicase bound to fragments (5RLH, 5RLZ, 5RML, 5RMM) and RNA (7CXM).

CACHE Challenge #3

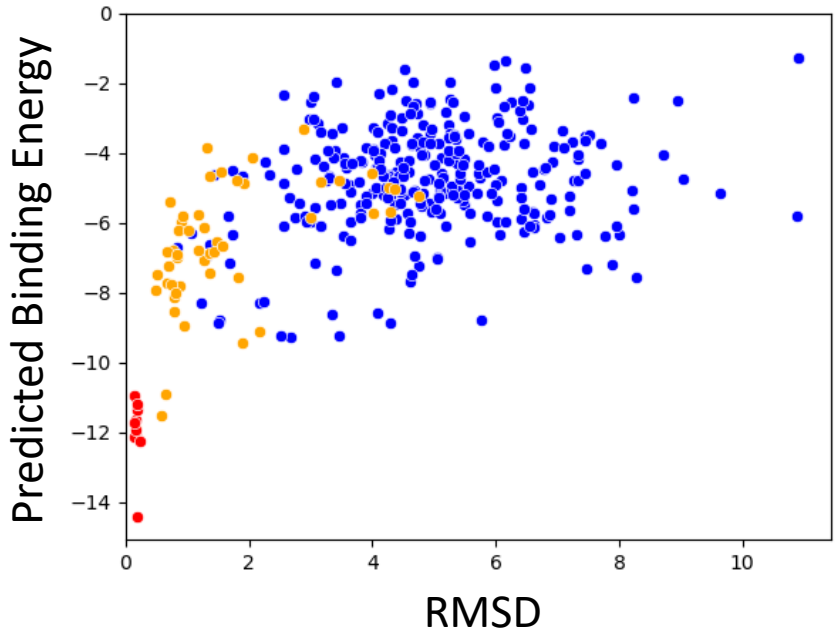


Crystal structure of SARS-CoV-2 NSP3 macrodomain bound to ADPr (7KQP).

CACHE #2: Hit Identification by Players and Post-Processing

- 7598 total compounds submitted
- 1411 from Enamine REAL set

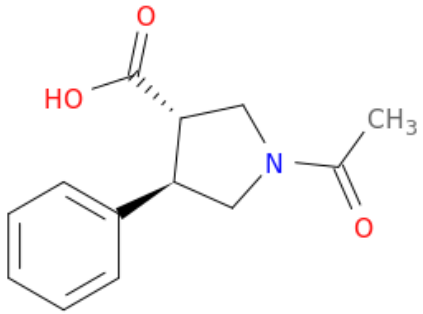
Primary Evaluation Measure: Rosetta Redocking



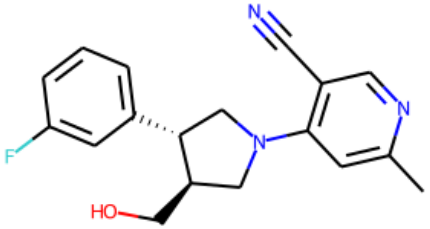
Predicted Binding Energy
10.1371/journal.pone.0240450
"Funnel Quality":
Pnear metric: 10.1038/nature19791

CACHE #2: SARS-CoV2 Helicase

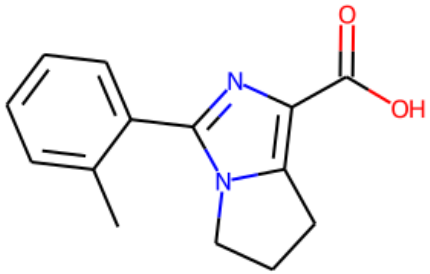
- Round 1: Hit Identification. 76 compounds tested



Starting Molecule

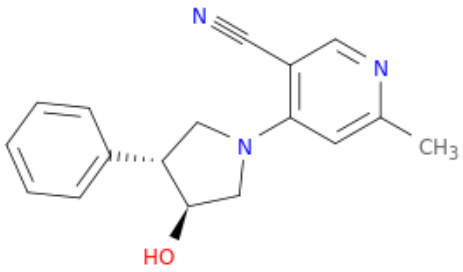


CACHE_1414_40

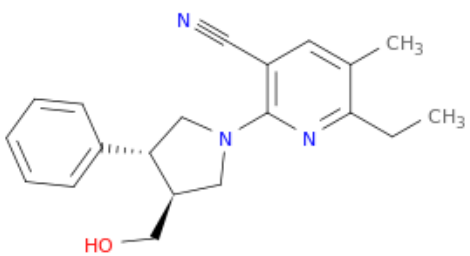


CACHE_1414_34

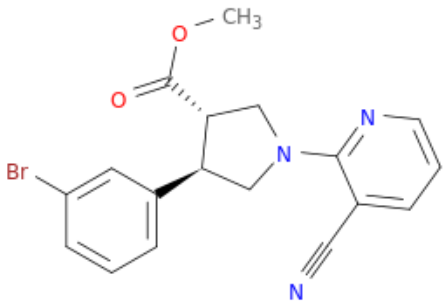
- Round 2: Hit Optimization. 34 compounds tested



CACHE-HO_1414_13



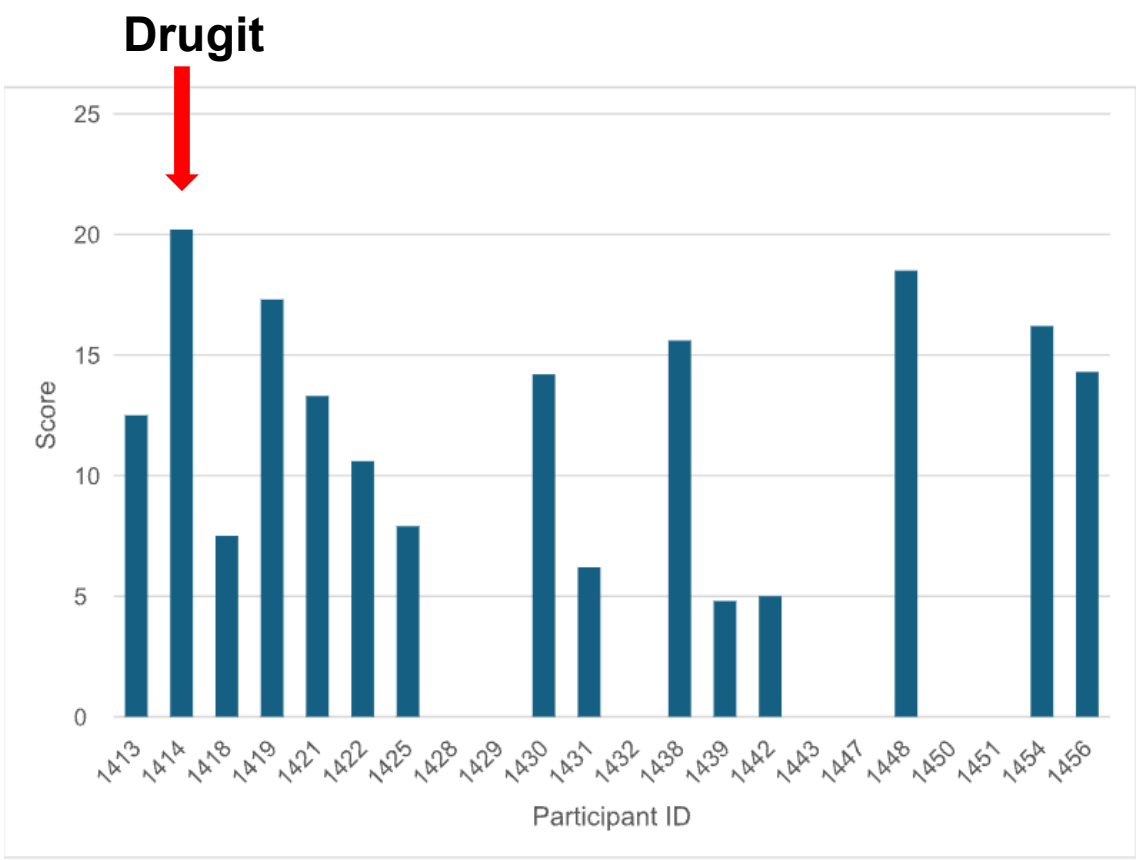
CACHE-HO_1414_14



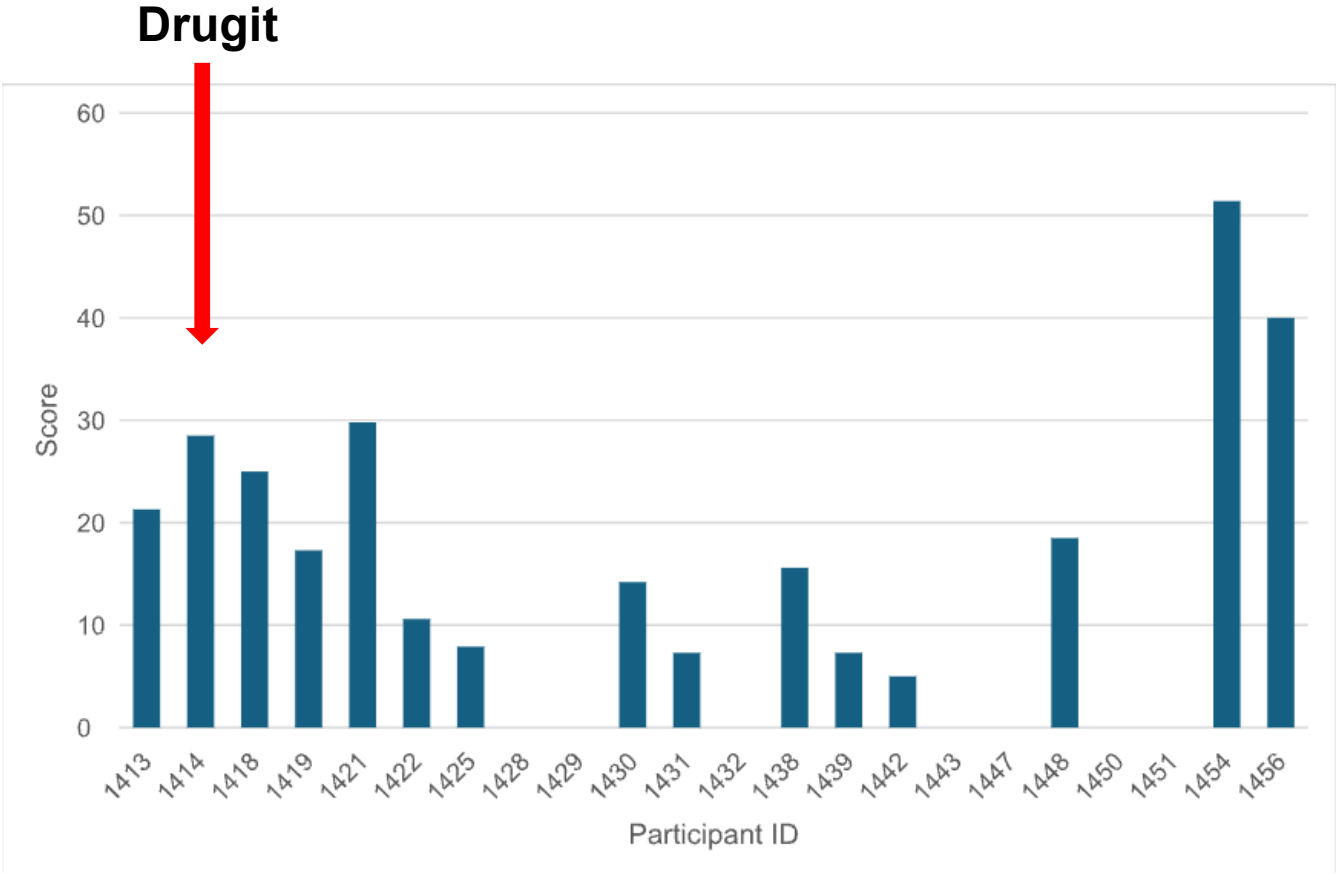
CACHE-HO_1414_20

CACHE #2: SARS-CoV2 Helicase

- Best Compound

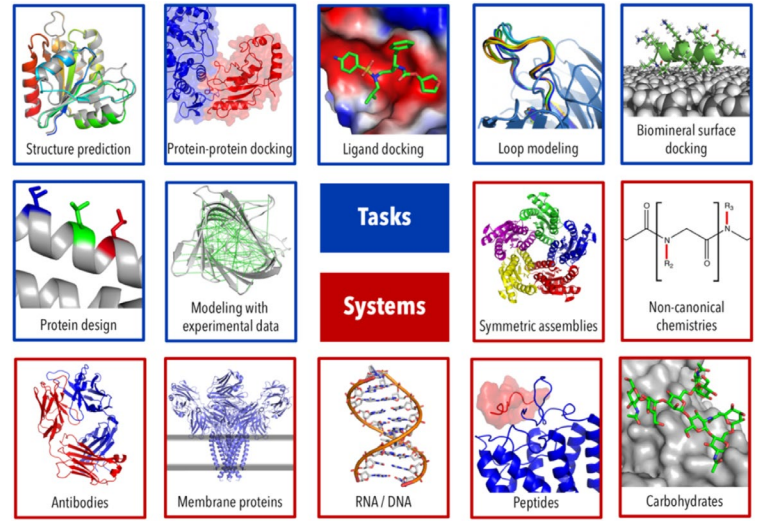


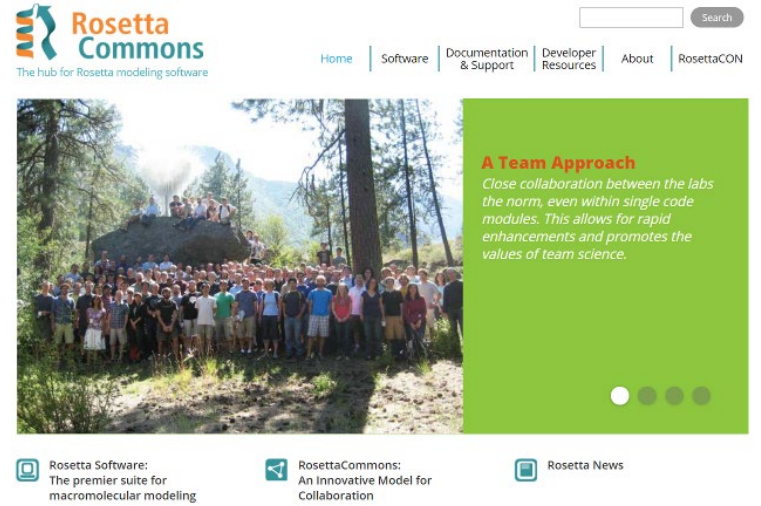
- Combined Compound Scores



Rosetta: A Unified Framework for Protein Structure

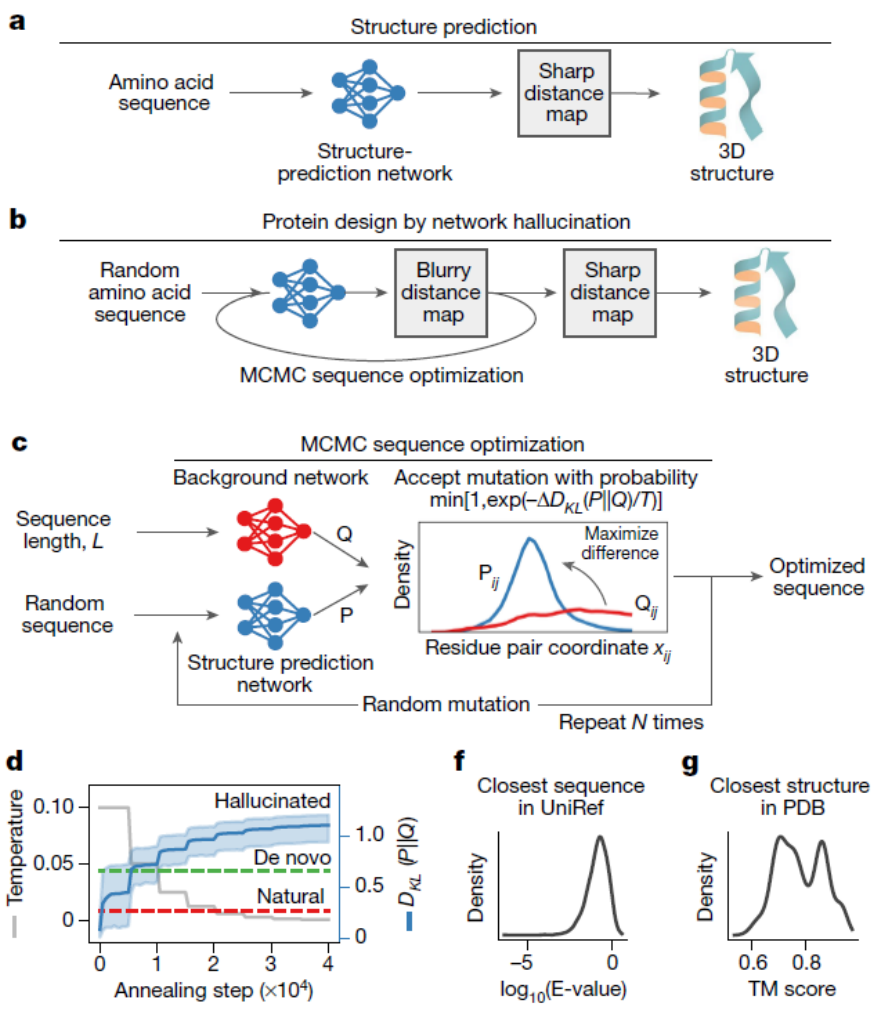
meilerlab

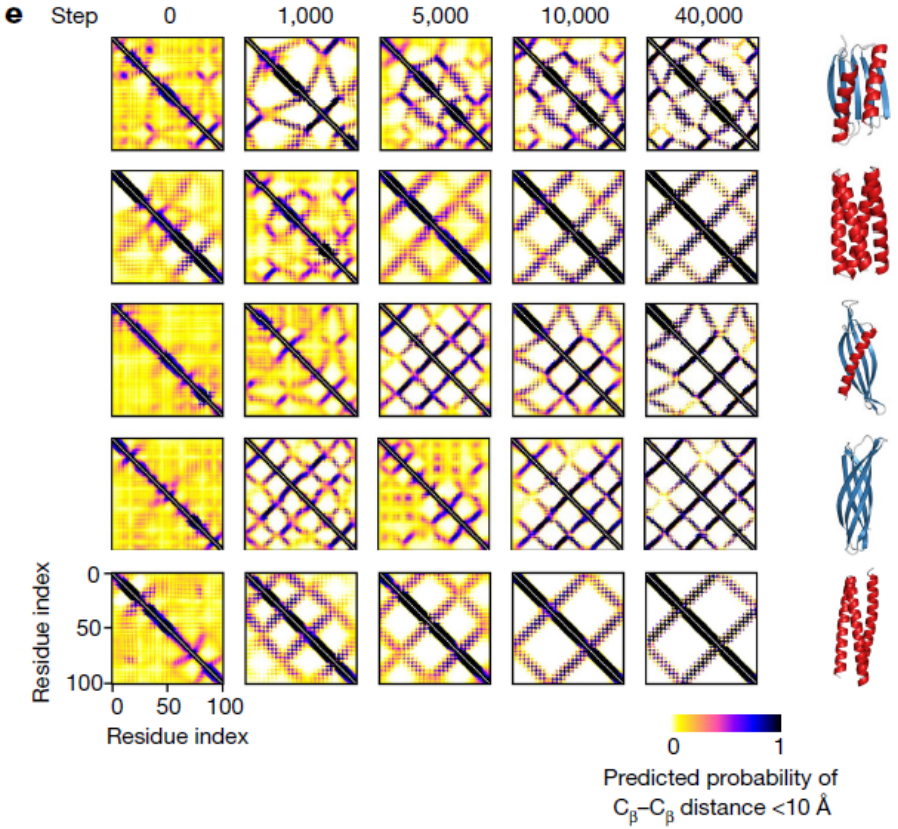




A.
J. K
. 665-680.

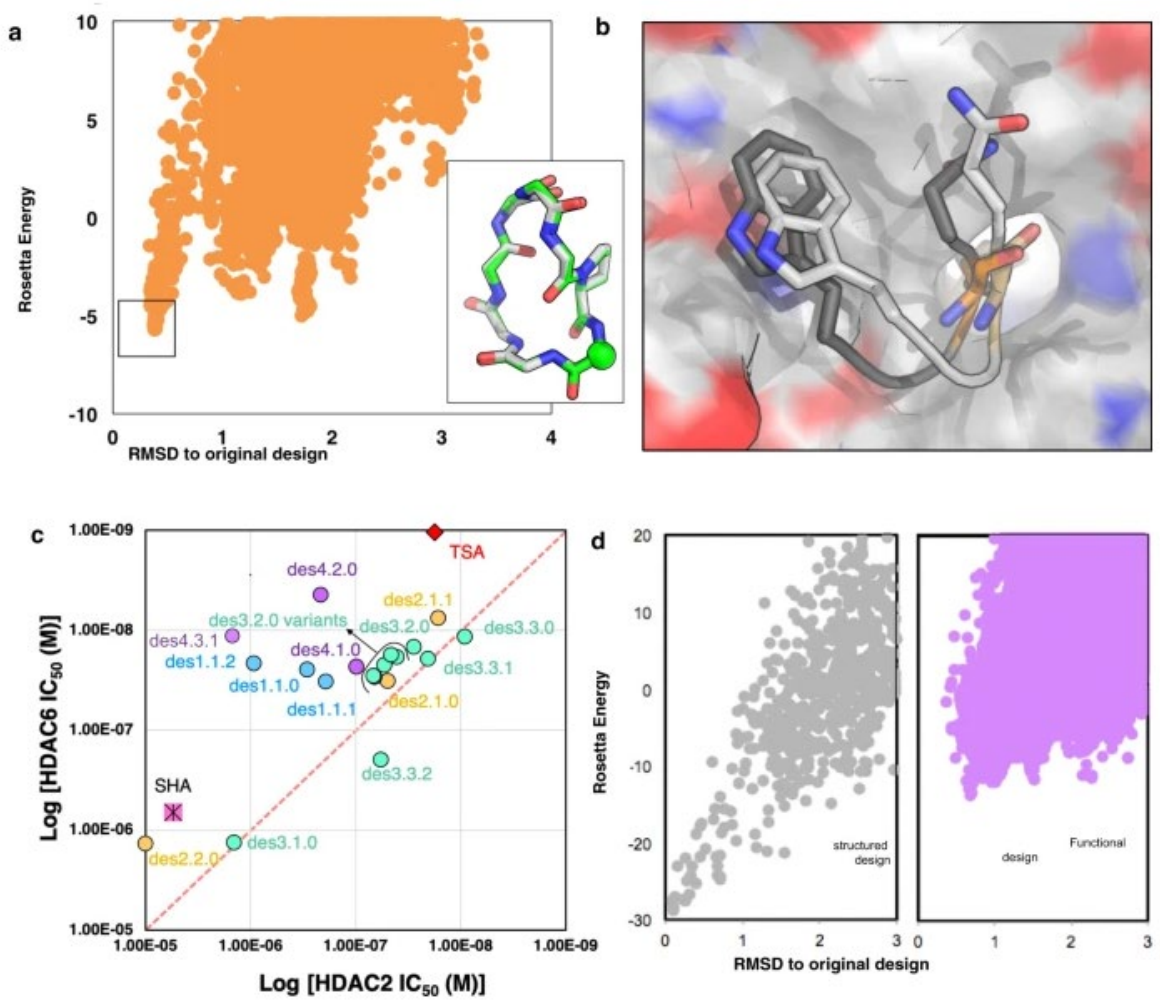
De novo Protein Design by Deep Network Hallucination





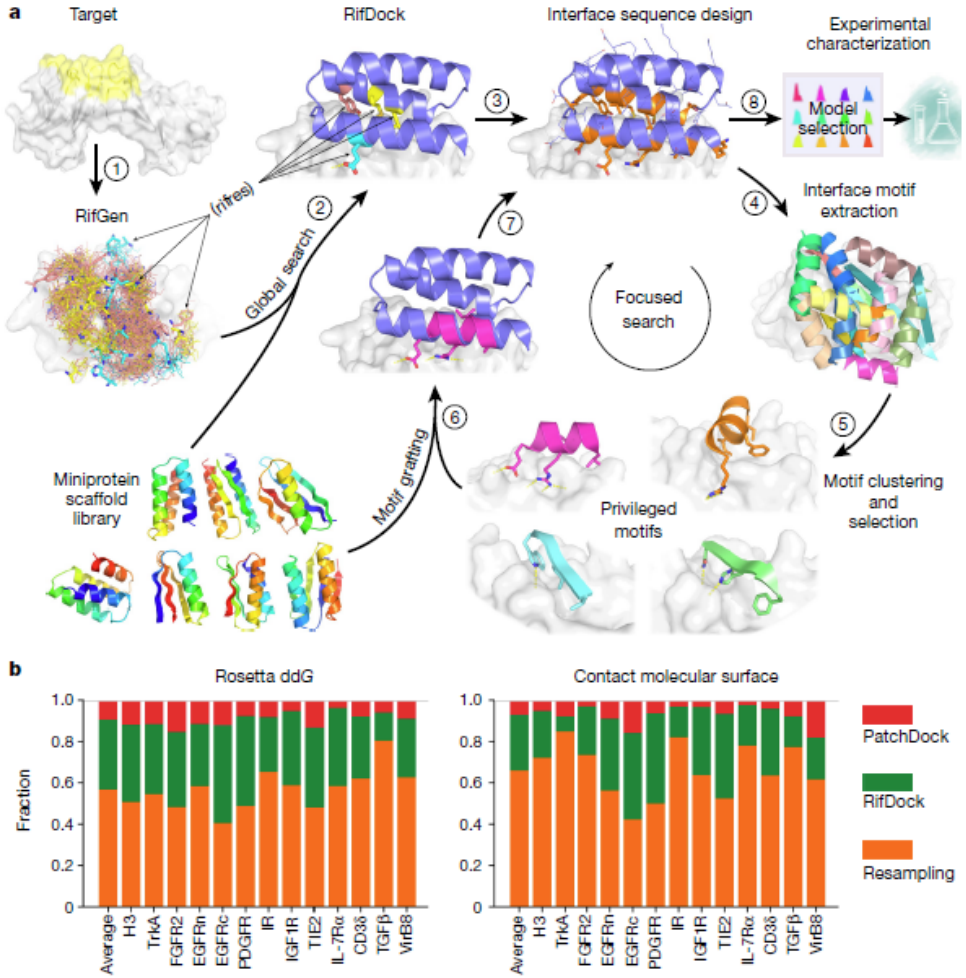
I. Anishchenko, S. J. Pellock, T. M. Chidyausiku, T. A. Ramelot, S. Ovchinnikov, J. Hao, K. Bafna, C. Norn, A. Kang, A. K. Bera, F. DiMaio, L. Carter, C. M. Chow, G. T. Montelione and D. Baker; "De novo protein design by deep network hallucination"; *Nature;* ***2021; Vol. 600 (7889): p. 547-552.***

Anchor extension: a structure-guided approach to design cyclic peptides



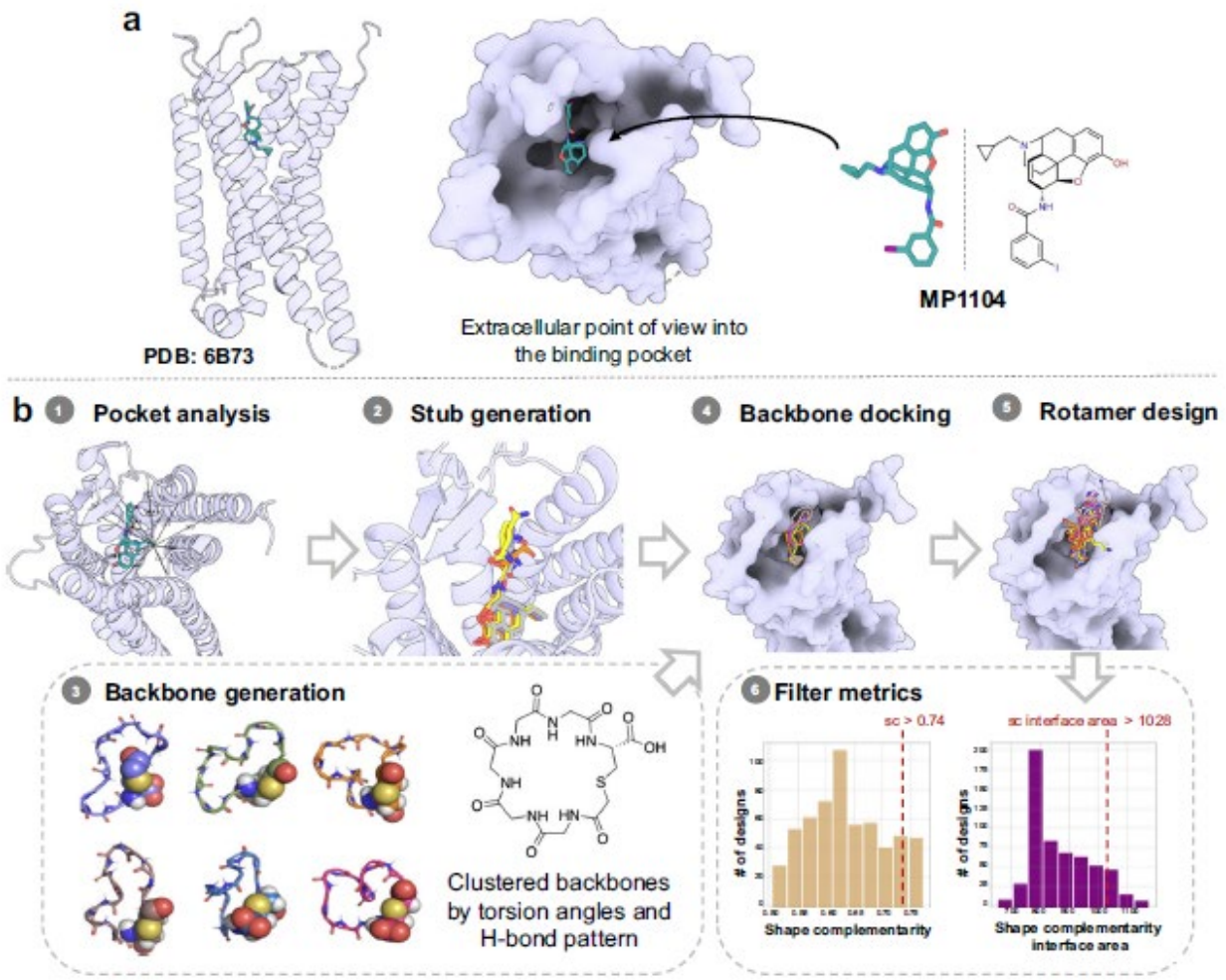
P. Hosseinzadeh, P. R. Watson, T. W. Craven, X. Li, S. Rettie, F. Pardo-Avila, A. K. Bera, V. K. Mulligan, P. Lu, A. S. Ford, B. D. Weitzner, L. J. Stewart, A. P. Moyer, M. Di Piazza, J. G. Whalen, P. J. Greisen, D. W. Christianson and D. Baker; "Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites"; *Nat Commun;* ***2021; Vol. 12 (1): p. 3384.***

Design of protein-binding proteins from the target structure alone



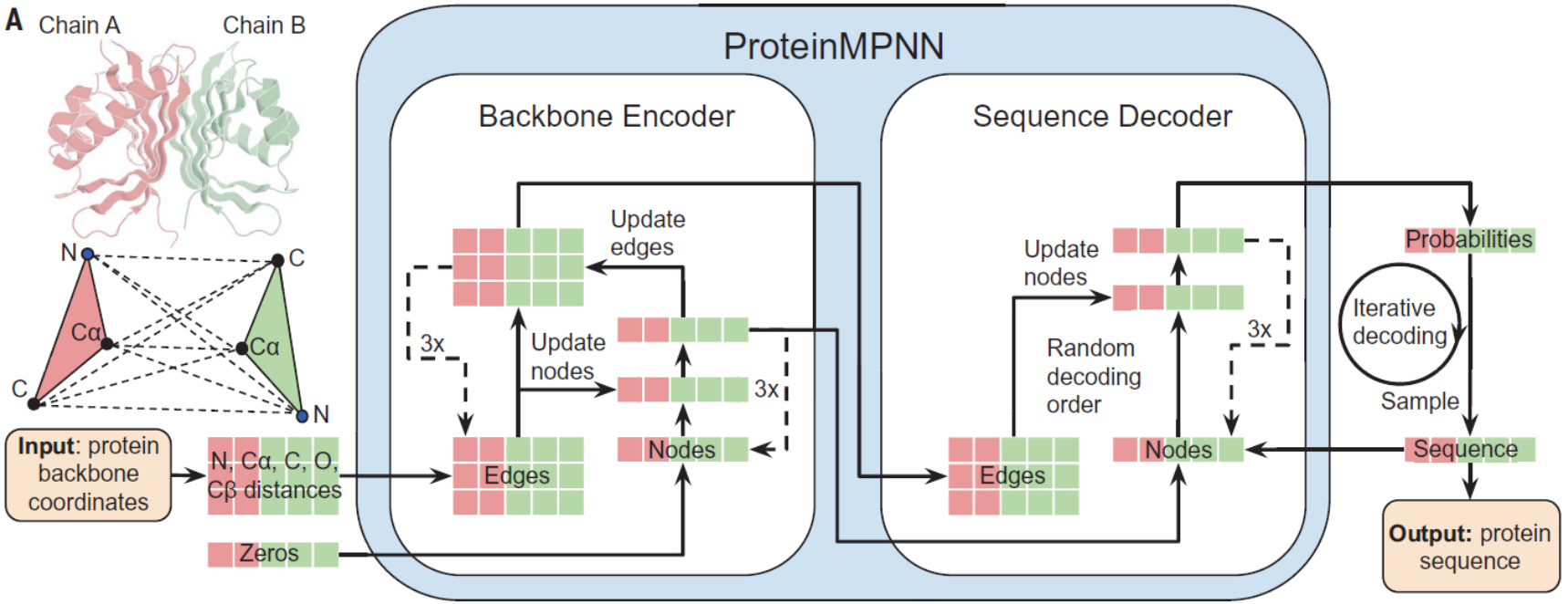
Cao, L. X., Coventry, B., Goreshnik, I., Huang, B. W., Sheffler, W., Park, J. S., Jude, K. M., Markovic, I., Kadam, R. U., Verschueren, K. H. G., Verstraete, K., Walsh, S. T. R., Bennett, N., Phal, A., Yang, A., Kozodoy, L., DeWitt, M., Picton, L., Miller, L., . . . Baker, D. (2022). Design of protein-binding proteins from the target structure alone. *Nature, 605(7910), 551-+.*

Design of peptide-drug conjugate ligands of the kappa-opioid receptor



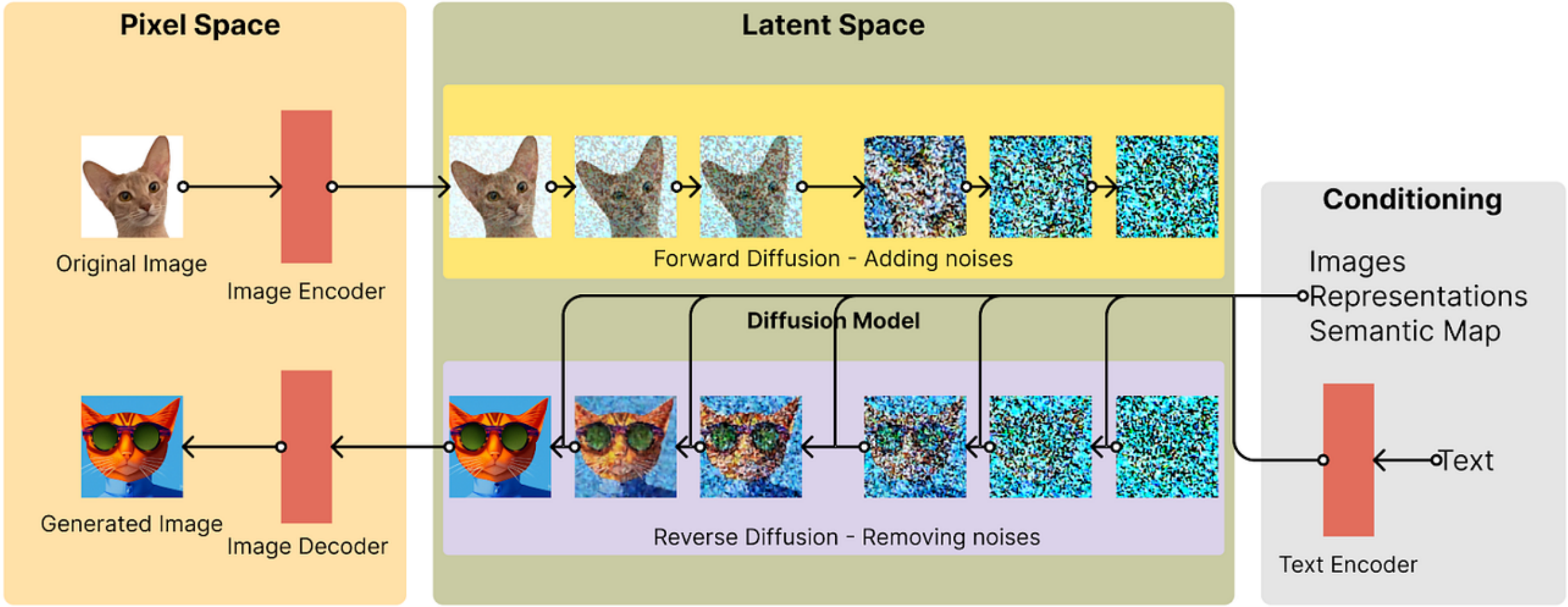
Muratspahic, E., Deibler, K., Han, J., Tomasevic, N., Jadhav, K. B., Olive-Marti, A. L., Hochrainer, N., Hellinger, R., Koehbach, J., Fay, J. F., Rahman, M. H., Hegazy, L., Craven, T. W., Varga, B. R., Bhardwaj, G., Appourchaux, K., Majumdar, S., Muttenthaler, M., Hosseinzadeh, P., . . . Gruber, C. W. (2023). Design and structural validation of peptide-drug conjugate ligands of the kappa-opioid receptor. *Nat Commun, 14(1), 8064.*

Robust deep learning-based protein sequence design using ProteinMPNN

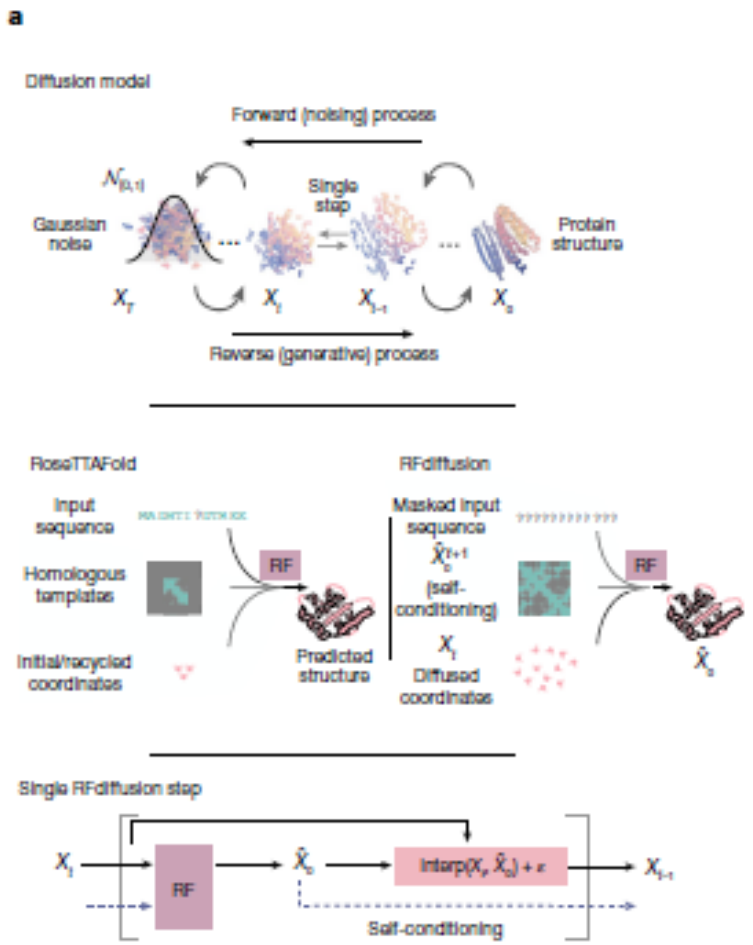


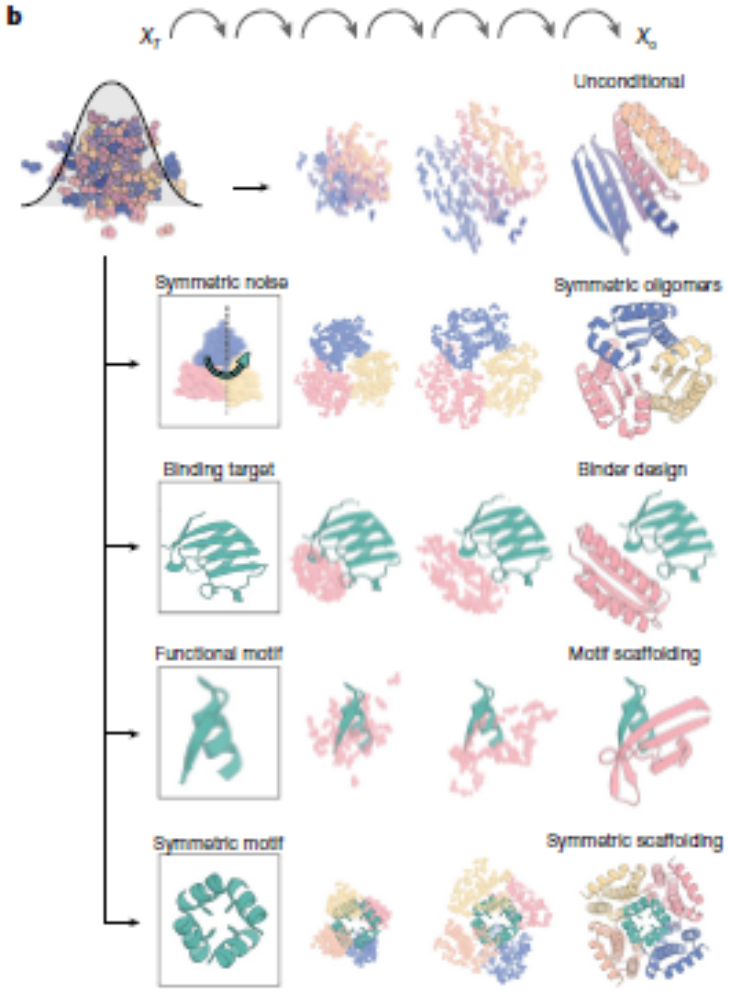
J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King and D. Baker; "Robust deep learning-based protein sequence design using ProteinMPNN"; *Science;* ***2022; Vol. 378 (6615): p. 49-56.***

How Stable Diffusion works in a Nutshell



De novo design of protein structure and function with RFDiffusion





Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., . . . Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. *Nature, 620(7976), 1089-1100.*

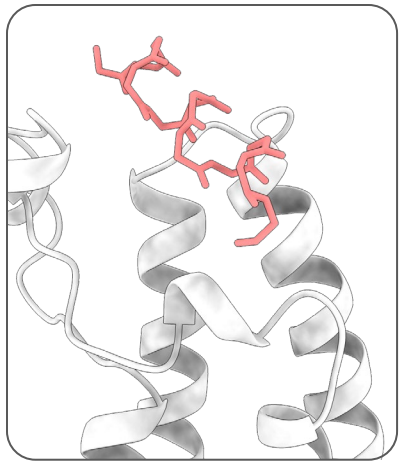
Modified Version of RFDiffusion for (Cyclic) Peptides

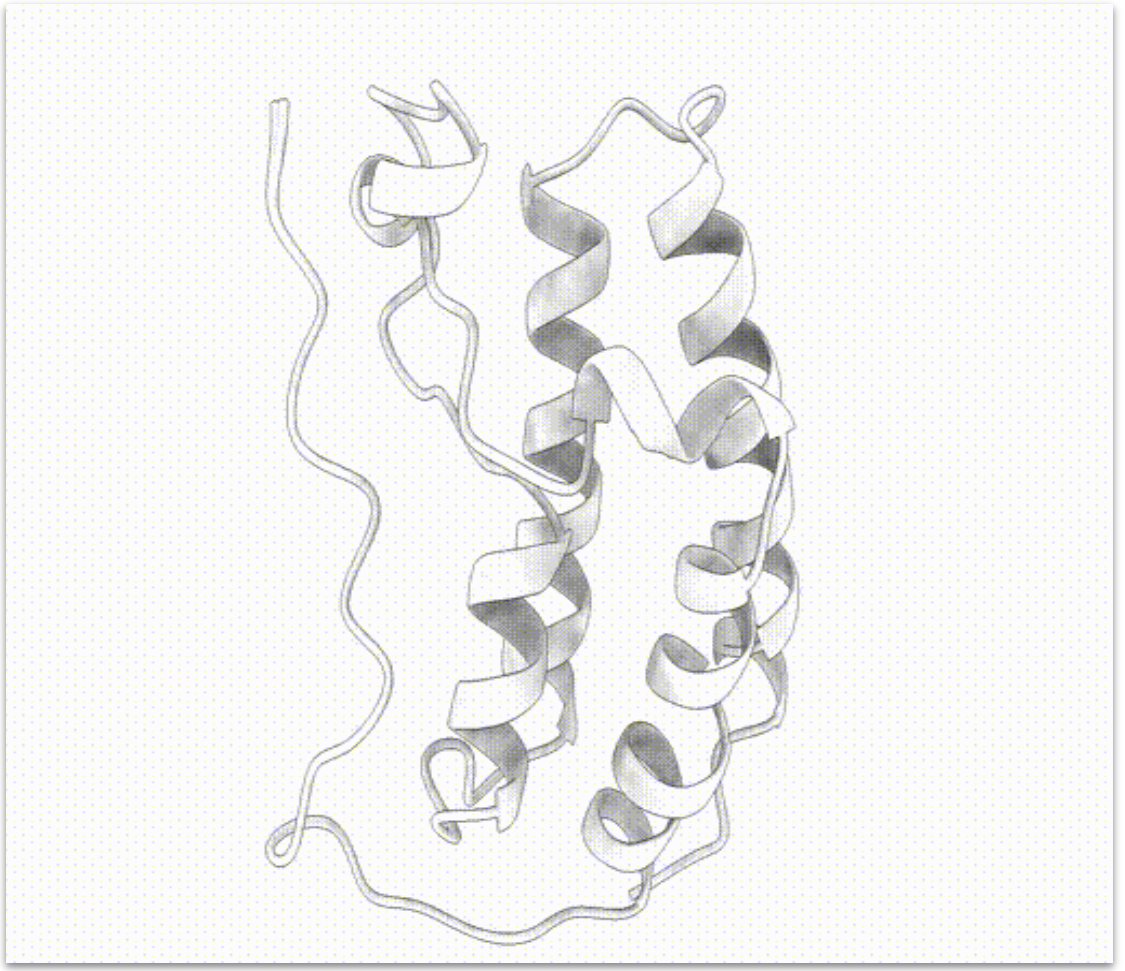
Normal positional embedding

0	1	2	3	4	5	6	7
-1	0	1	2	3	4	5	6
-2	-1	0	1	2	3	4	5
-3	-2	-1	0	1	2	3	4
-4	-3	-2	-1	0	1	2	3
-5	-4	-3	-2	-1	0	1	2
-6	-5	-4	-3	-2	-1	0	1
-7	-6	-5	-4	-3	-2	-1	0

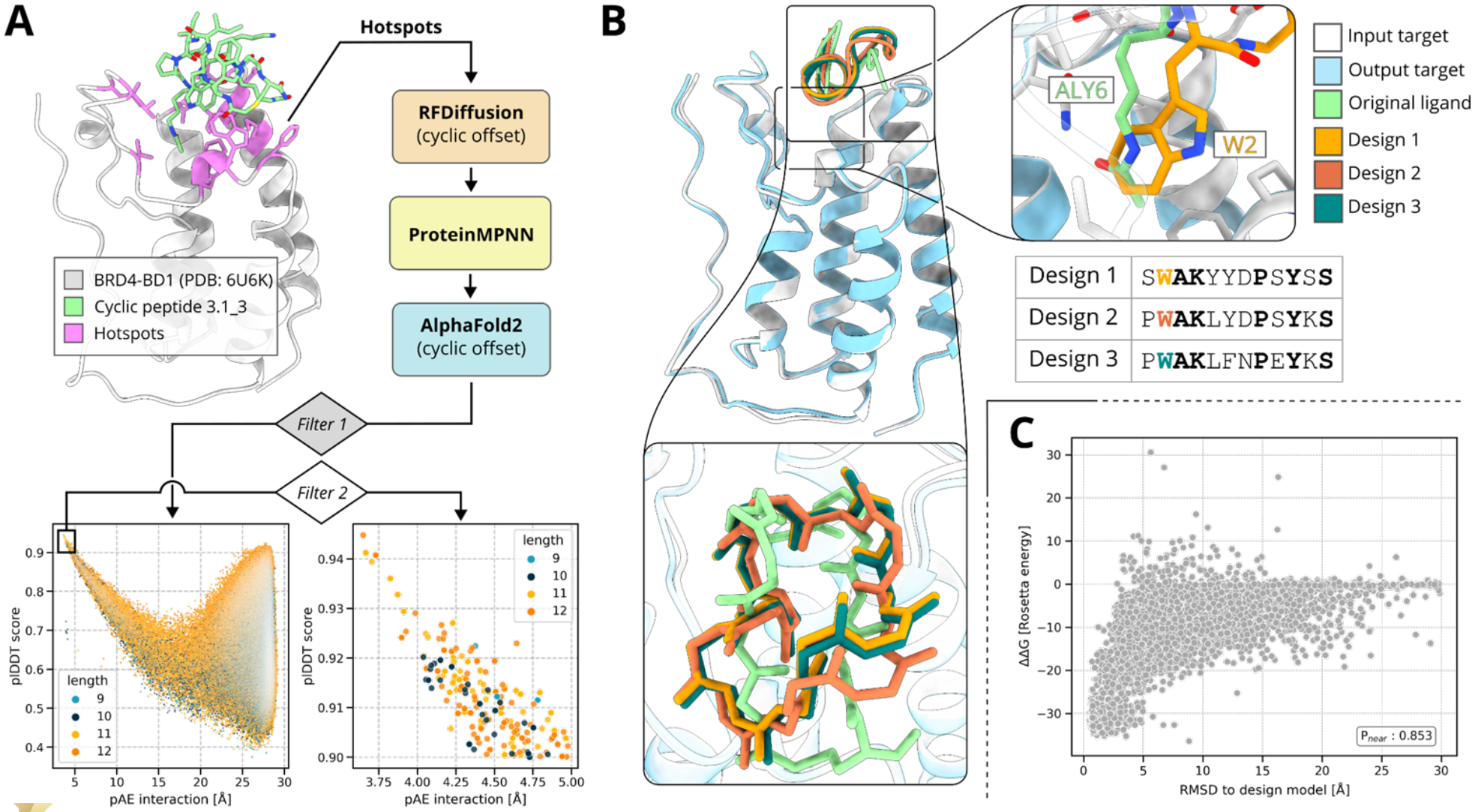
Cyclic positional embedding

0	1	2	3	-4	-3	-2	-1
-1	0	1	2	3	-4	-3	-2
-2	-1	0	1	2	3	-4	-3
-3	-2	-1	0	1	2	3	-4
-4	-3	-2	-1	0	1	2	3
3	-4	-3	-2	-1	0	1	2
2	3	-4	-3	-2	-1	0	1
1	2	3	-4	-3	-2	-1	0





Modified Version of RFDiffusion for (Cyclic) Peptides



Acknowledgements

Current Members:

Alan Perez-Ratzke
Alexander Fürll
Alexander Zlobin
Ali Kanso
Ana Chang-Gonzalez
Anja Landsmann
Amy (Minh) Tran
Carie Fortenberry
Chris Jurich
Chris Moth
Christian Schmidt
Christina Mercado
Claiborne Tydings
Cristina Martina
Dieter Hoffmann
Dominik Rieger
Drew Neufer
Emma Web
Eric Bell
Fabian Liessmann
Felipe Engelberger
Gustavo Araiza
Hannes Junker
Ivan Yushin
Jakob Riccabona
Jan Rossa
Jannis De Riz
Joe DeCorte
Johannes Klier
Johnny Siegert
Jordan Davis
Julia Beliaeva
Kaitlyn Ledwitch
Kelley Colopietro
Kuniko Hunter
Kyle Brown
Lance Liu
Lorenz Beckmann
Luisa Kaermer
Lukas v. Bredow
Marija Sorokina
Markus Fischer
Mateusz Sklodowski
Max Beining
Maximilian Hofmann
Paul Eisenhuth
Paul Kluge
Qianzhen Shao
Rajarshi SinhaRoy
Ratna Wolfram
Rocco Moretti
Sean Murray
Sevilay Gülesen
Shan Jiang
Sophie Maclure
Stefan Arnold
Te-Yu Kao
Thomas Scott
Tom Schlegel
Toni Oesterreich
Tracy (Yidan) Tang
Onur Elaldi
Victoria Most
Vivian Ehrlich
Zhaoqian Su

Selected Alumni:

Abdullah Al Mamun
Alican Gulsevin
Annalen Bleckmann
Alex Sevy
Axel Fischer
Benjamin Brown
Brett Kroncke
Brian Bender
Brian Weiner
Clara Schoeder
David Nannemann
Diego del Alamo
Eli McDonald
Georg Kuenze
Hope Woods
Jordon Willis
Julia Koehler Leman
Liz Dong Nguyen
Moritz Ertelt
Nils Woetzel
Steffen Lindert

Selected Collaborators:

Jim Crowe, Heidi Hamm, Craig Lindsley, Brian Shoichet, Hassane Mchaourab, Chuck Sanders, Lars Plate, Jonathan Schlebach, Annette Beck-Sickinger, Clara Schoeder, …

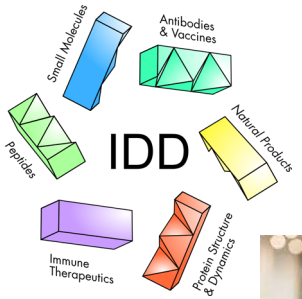
Funding: NIH NIAID U19 AI117905, NIH NILBI R01HL122010, NIH NIDA R01DA046138, NIH NLM LM013434, ARPA-H

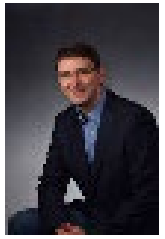
DFG CRC 1423, DFG CRC 1052, DFG TRR 386, DAAD 57616814, BMBF, CEPI, Alexander von Humboldt Foundation

BAYER, BIONTECH, BASF

Institute for Drug Discovery at University Leipzig in Germany

Institute for Drug Discovery





Maik Tretbar group leader Medicinal chemistry

2019

The Institute for Drug Discovery (IDD) is founded

Georg Kuenze group leader Computational Biology

2020

SFB1423 is funded

DFG Deutsche Forschungsgemeinschaft

Alexander von HUMBOLDT STIFTUNG

€5 Mio start-up funding

Alexander von HUMBOLDT STIFTUNG

Torben Schiffner receives the Sofja Kovalevskaja prize (€ 1,7 Mio start-up funding)

2021

Leo Kaysser becomes professor for pharm. Biology

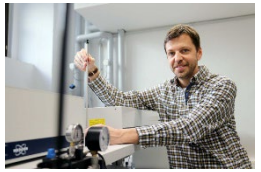
Christina Lamers is named junior professor for medicinal chemistry

Christian Woelk group leader lipd nano-particles

2022

ScaDS.AI DRESDEN LEIPZIG

National Competence Center Artificial Intelligence



The IDD installs an EPR and DEER facility

KIWI funding ~€1,5 Mio

BIONTECH

Bundesministerium für Bildung und Forschung

DFG Deutsche Forschungsgemeinschaft

TRR-386 is funded by the DFG

HYP*MOL

Vaccine design professorship

2023

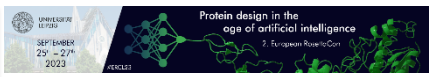
Partnership of the IDD with CEPI to design new vaccines for the future

CEPI

€1,8 Mio funding

State of Saxony will fund a new research building for the IDD (€50 Mio)

Clara Schoeder was appointed junior professor for the developement of immunotherapeutic drugs



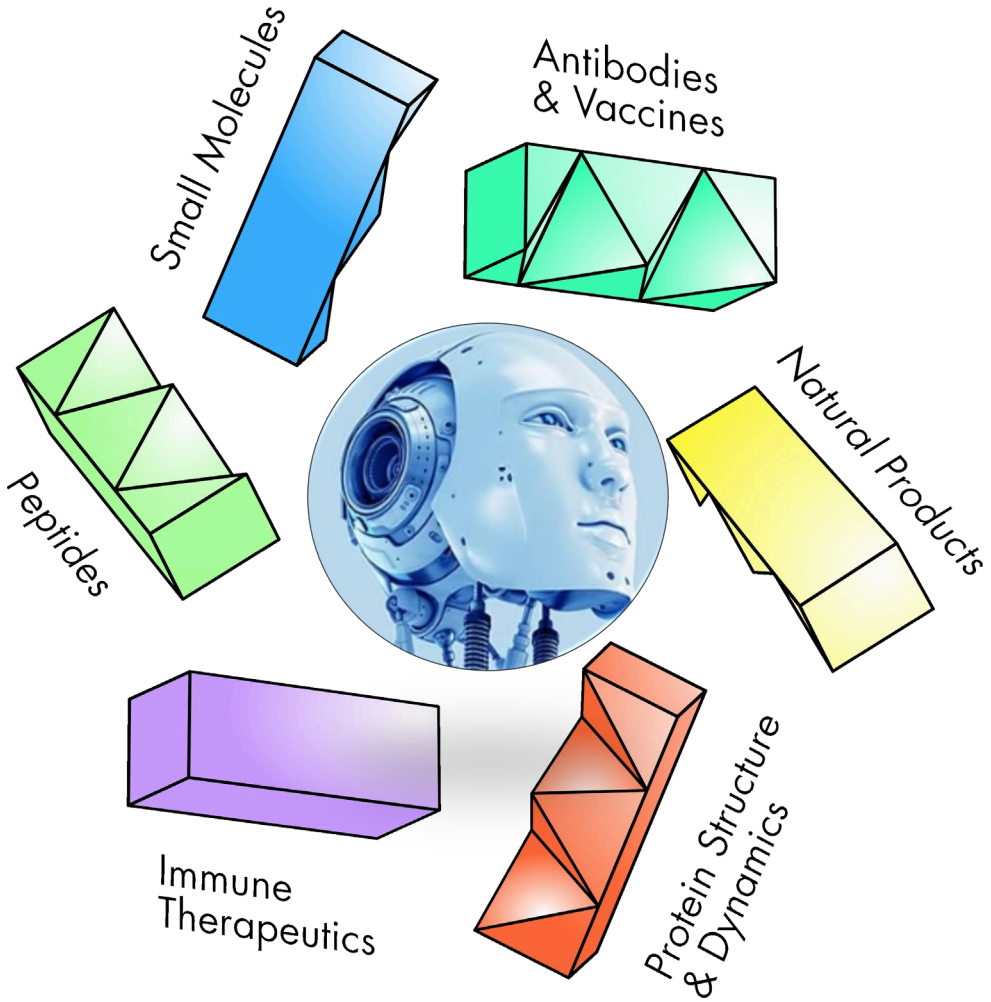
Rosetta Commons

The IDD and RosettaCommons host European RosettaCon in Leipzig

2024

***In silico* First:** Putting Computation and Artificial Intelligence in the Center

Institute for Drug Discovery



New Integrated Research Building (2026)

AI High-Performance Computing Center (2026)

ScaDS.AI DRESDEN LEIPZIG

National Competence Center Artificial Intelligence (2022)

Ultra-Large Library Screening Projects Using Computational Models with Increasing Library Sizes

meilerlab

Target	Structure	Database size	Hits/tested	Comment	Med. Chem
GPCR PAR4 with DOCK	Homology model	164 m	1/92	1 prelim hit (5 µM), optimized to 100 nm *Smith, S. T., …, M., Shoichet, B. K., Lindsley, C. W., Meiler, J., & Hamm, H. E. (2024). ACS Pharmacol Transl Sci, 7(4), 1086*	Completed
aGPCR GPR114	Homology model	303 m	5/88	5 prelim hits (7 nM – 10 µM)	Ongoing
aGPCR GPR133	Homology model / CryoEM structure	303 m	1/55	1 prelim hit (5 µM) not validated	Terminated
WDR40 LLRK2 (CACHE)	Crystal structure	20 bn	1/102 4/37 (follow-up)	1 initial validated hit, 4 follow-up hits (KD 34 µM – 290 µM)	Ongoing
GPCR MC4R	CryoEM structure	20 bn	13/82	QSAR model as filter, activity currently investigated	Planned
GPCR Y4R	CryoEM structure	20 bn	1/14	1 inital validated hit (EC_{50} ~10 µM, first small molecule agonist)	Ongoing
GPCR GPR3	AlphaFold2 model	31 bn	9/91	9 hits (not reported EC_{50} values)	Planned

Ultra-Large Library Screening Projects Using Computational Models with Increasing Library Sizes

meilerlab

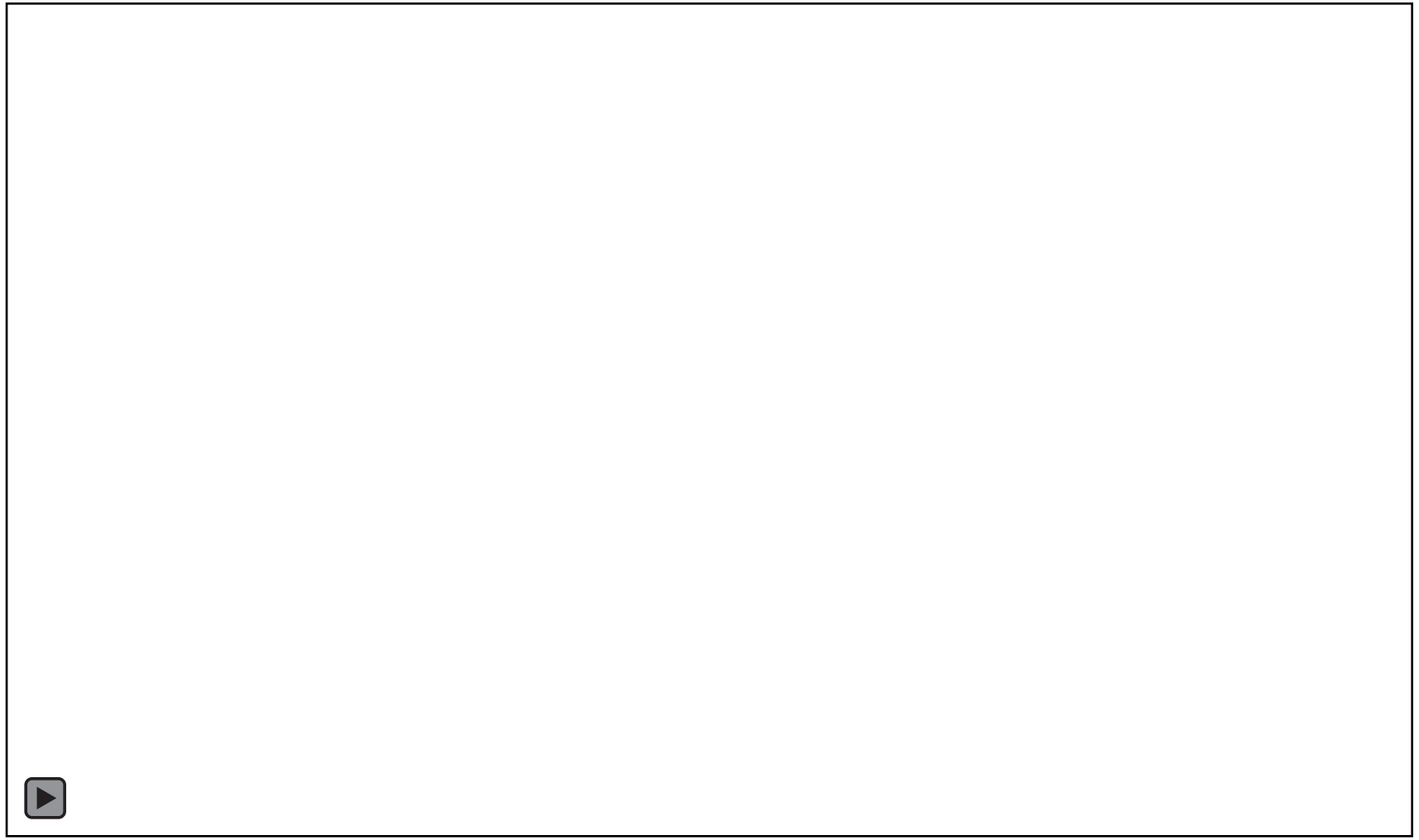
Target	Structure	Database size	Hits/tested	Comment	Med. Chem
GPCR CMKLR1	Homology model +CryoEM structure	31 bn	(5)/97	5 prelim hit, currently investigation	Planned
aGPCR GPR110	CryoEM structure	20 bn	(3)/82	3 prelim hits, currently investigated	Planned
HisF	Crystal structure	31 bn	Ongoing	Ongoing	Control
CFTR mutant	Homology model	31 bn	Ongoing	Ongoing	Possible
KCNQ1 mutant	Homology model	31 bn	Ongoing	Ongoing	Possible
GPCR PAR3	AlphaFold models	31 bn	Ongoing	Ongoing	Possible
GPCRs PAR4, mGluRs, mAChRs, DRs	AlphaFold models of states to be targeted	31 bn	Ongoing	Ongoing	Planned

The Future of Artificial Neural Networks in Biomedical Research – Some Thesis

1. All problems that have infinite/near infinite data available for training will be smashed (think language processing, sequence problems in biochemistry, protein structure)
2. New architectures and structures of ANNs will emerge that will be parallel in size to or larger then the human brain (10^{14} connections) with substructures matching in complexity
3. The biggest challenge for biomedical research will emerge with limited datasets that forbid training of super-large ANNs; Expert Knowledge will Design the Optimal ANN
4. For the next Decade (at least), you need to be an expert in machine learning and structural/chemical biology to contribute to progress in a meaningful way
5. We will start an honest discussion on ethics of artificial intelligence as these systems will start to act human-like on many levels all the way to having self-awareness


DrugIt: Crowd-Sourcing Drug Design via a Computer Game to Engage Citizen Scientists

meilerlab



Computational Algorithms for Enzyme Design in Rosetta

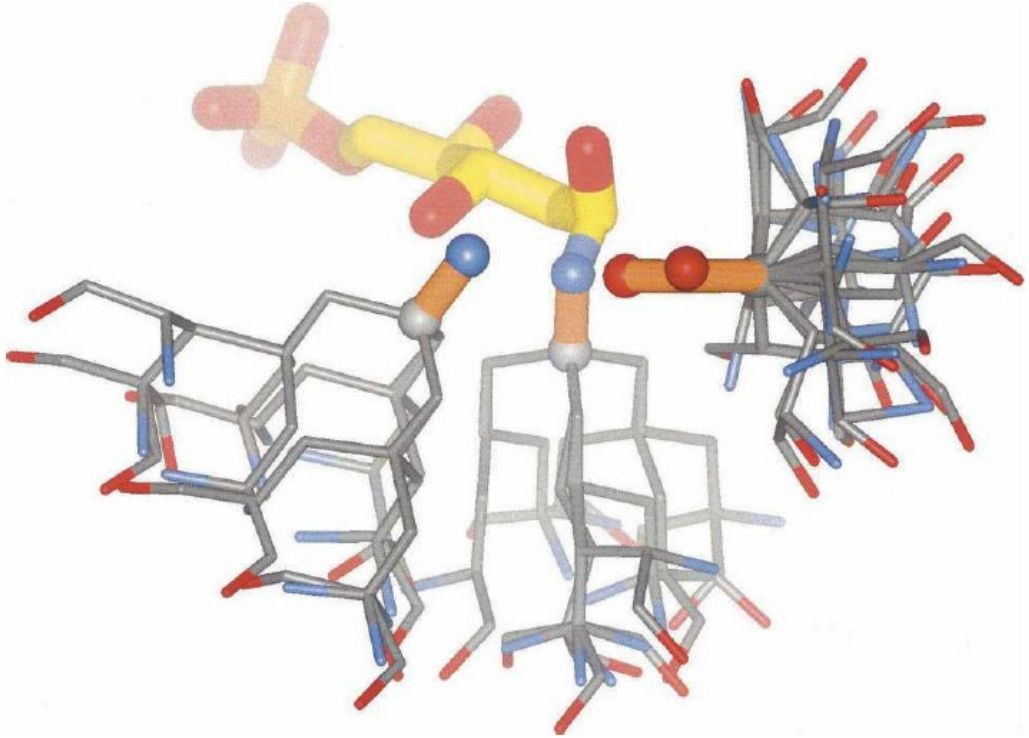


Figure 6. Inverse rotamer tree for deoxyribose-phosphate aldolase (DERA) active site. The transition state is colored in yellow, and the key functional groups of the catalytic residues are in gold. The remainder of the side chains in the rotamer trees are shown using thinner lines in CPK coloring.

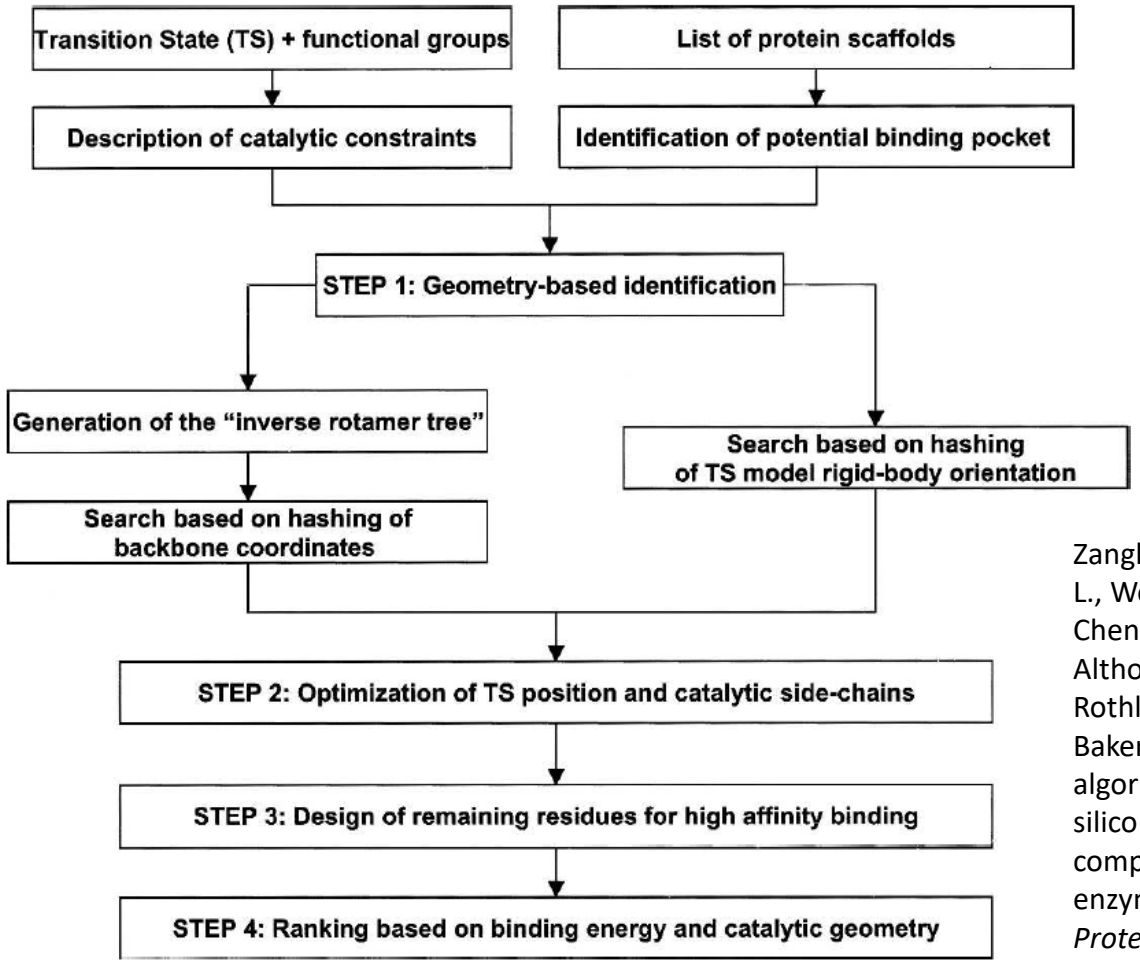


Figure 1. Diagram of the computational enzyme design procedure.

Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., Rothlisberger, D., & Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci, 15(12), 2785-2794. https://doi.org/10.1110/ps.062353106*

At Vanderbilt University we collaborate with the Warren Center for Neuroscience Drug Discovery

meilerlab

WARREN CENTER
FOR NEUROSCIENCE DRUG DISCOVERY

Our Mission

To translate advances in basic science to novel therapeutics by de-risking innovative approaches for the treatment of serious brain disorders.