From protein structure prediction to mechanistic insights into signaling and disease diagnostics

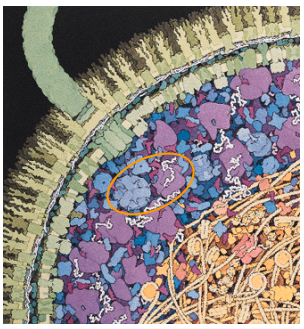
Holger Gohlke

C | P | C
Computational Pharmaceutical Chemistry
Heinrich-Heine-University Duesseldorf

C | B | C
Computational Biophysical Chemistry
Forschungszentrum Jülich

Mission

Understand, predict,
and
modulate
molecular and
cellular behavior
from an
atomistic perspective



Taken from David S. Goodsell, **The Machinery of Life**, 1997.

Application studies

$\alpha_5\beta_1$ integrin
Gohlke, ..., Häussinger, Hepatology 2013
Reinehr, Gohlke, ..., Häussinger, JBC 2010

$\alpha_{IIb}\beta_3$ integrin
Donner, ..., Gohlke, ..., Elvers, Cell Signal 2018
Pagani, Gohlke, Sci Rep 2018
Pagani, ..., Scharf, Gohlke, JBC 2018
Donner, ..., Gohlke, Elvers, Sci Signal 2016

TGR5
Greife, ..., Gohlke, Keitel, Seidel, Sci Rep 2016
Gertzen, ..., Keitel, Gohlke, Eur J Med Chem 2015
Spomer, ..., Gohlke, Keitel, JBC 2014

Glutamine synthetase
Frieg, ..., Häussinger, Gohlke, PLOS Comp Biol 2016

MDR3/Pdr5
Tanabe, ..., Gohlke, ..., Lamping, Antim. A. Chemo. 2019
Sachs, ..., Gohlke, ..., Teusch, Front Pharmacol 2019
Dröge, ..., Gohlke, ..., Keitel, J Hepatol 2017

ETR1
Schott-Verdugo, ..., Gohlke, Groth, Sci Rep 2019
Milic, ..., Gohlke, Groth, Sci Rep 2018

PEPC
Dick, ..., Groth, Gohlke, FEBS Lett 2017
Nguyen, ..., Gohlke, Groth, Sci Rep 2016

PPDK
Ciupka, Gohlke, Sci Rep 2017
Minges, ..., Gohlke, Groth, Sci Rep 2017

HDAC
Raudszus, ..., Gohlke, ..., Hansen, BMC 2019
Reßling, ..., Gohlke, ..., Hansen, Med Chem Comm 2019
Krieger, ..., Gohlke, ..., Hansen, J Med Chem 2017
Stenzel, ..., Gohlke, ..., Kurz, J Med Chem 2017
Diederich, ..., Gohlke, ..., Hansen, Chem Comm 2016
Marek, ..., Gohlke, ... Kurz, J Med Chem 2013

NHR2
Schanda, ..., Gohlke, ..., Wichmann, Haematol 2017
Metz, ..., Wichmann, Gohlke, JCIM 2013
Wichmann, ..., Gohlke, Grez, Blood 2010

Hsp90
Bhatia, ..., Gohlke, ..., Hauer, Blood 2018
Krieger, ..., Gohlke, Hansen, Chem Eur J 2017
Diedrich, Gohlke, ..., Hansen, Chem Eur J 2016
Spanier, ..., Gohlke, Kurz, J Org Chem 2014
Bopp, ..., Gohlke, Jose, BBA – Gen Subj 2016
Ciglia, ..., Groth, Gohlke, PLOS ONE 2014

Lipases
Verma, ..., Jaeger, Gohlke, J Comp Chem 2019
Skoczinski, ..., Gohlke, ..., Jaeger, Microb Cell Fact 2017
Rathi, ..., Jaeger, Gohlke, PLOS Comp Biol 2016
Rathi, ..., Jaeger, Gohlke, PLOS ONE 2015

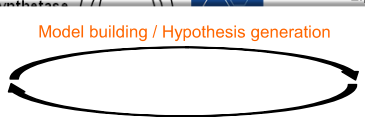
Aldolases
Bisterfeld, ..., Gohlke, Pietruszka, PLOS ONE 2016
Dick, ..., Gohlke, Pietruszka, Sci Rep 2016

NSR
Porta, ..., Smits, Gohlke, BMC 2019
Khosha, ..., Gohlke, ..., Smits, PLOS ONE 2016
Khoas, ..., Gohlke, Smits, Sci Rep 2016

Vifs
Gu, ..., Gohlke, Münk, J Virol 2018
Zhang, ..., Gohlke, Münk, Retrovirology 2016

FRET
Sanabria, ..., Gohlke, Seidel, Nature Comm 2020
Dimura, ..., Gohlke, Seidel, Curr Opin Struct Biol 2018
Kalinin, ..., Gohlke, Seidel, Nature Methods 2012

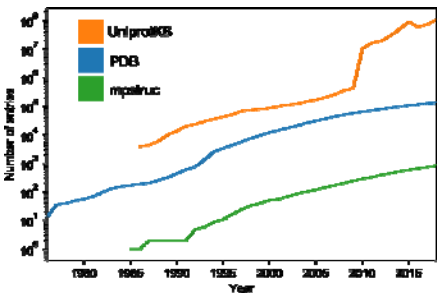
Model building / Hypothesis generation



Validation / Restraints for integrative modeling

From sequence to structure

Next-generation sequencing rates
by far exceed
experimental structure determination rates



From sequence to structure

Template-based structure prediction

Variability

Sequence
↓ determines
Structure
↓ determines
Function

⇒ **„Sufficiently similar" sequences adopt identical fold**

From sequence to structure

Template-based structure prediction

Variability

Sequence (Target)

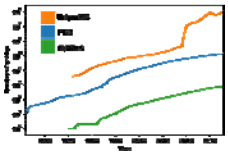
predict

Structure

determines

Function

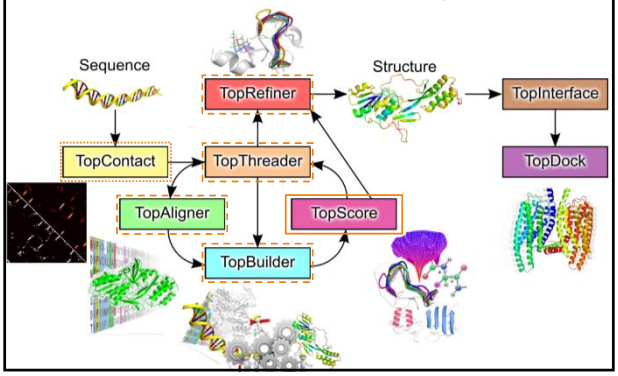
(Template(s))



⇒ **„Sufficiently similar" sequences adopt identical fold**

From sequence to structure

TopSuite: Consensus is key

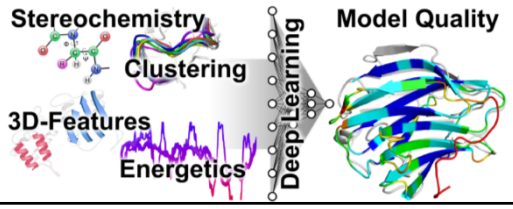


Protein structure prediction

Protein model quality assessment

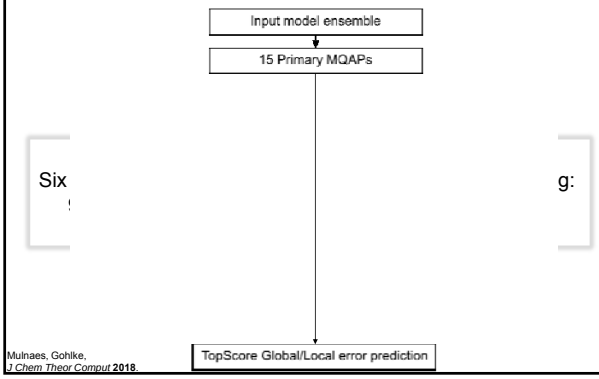
Is my structure prediction good? → Compare to known structure

IDDT score → Intra-molecular distance conservation
No superposition



TopScore

Implementation



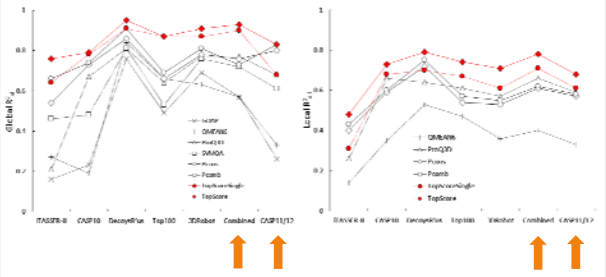
Mulnaes, Gohlke, *J Chem Theor Comput* **2018**.

TopScore

Comparative assessment

Global quality

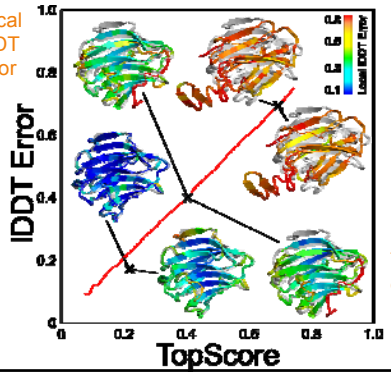
Local quality



TopScore

Scoring model quality („Combined")

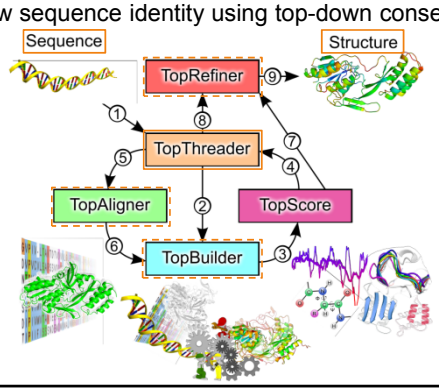
Local IDDT error



Local TopScore error prediction

TopModel

Template-based protein structure prediction
at low sequence identity using top-down consensus



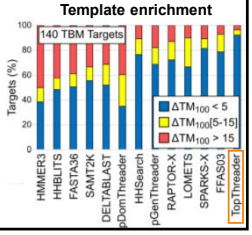
TopModel

Finding templates

The consensus problem:
Predictor performance is case-dependent → No one is always right

Which are the best template(s) for my protein?
1. Let DNNs find the *best* hit
2. Top-down consensus → Discard hits that do not fit the *best* hit

Template enrichment



TopModel

Finding templates

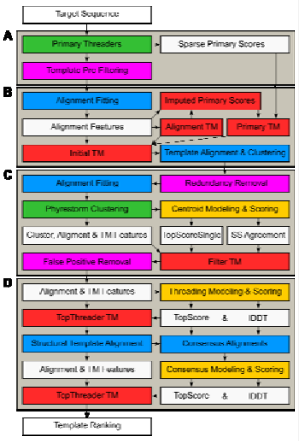
- Primary threading ⇒ Template(s) + score(s)

TopModel

Finding templates

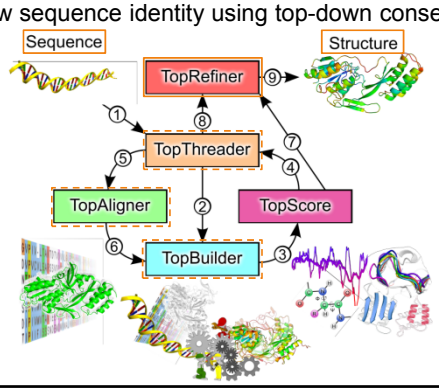
- Primary threading

- Pre-filtering
- Alignment fitting
- Score templates (DNN)
- Redundancy clustering
- False positive removal (DNN)
- Consensus generation
- Ranking (DNN)



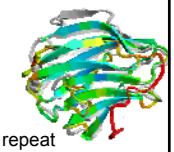
TopModel

Template-based protein structure prediction
at low sequence identity using top-down consensus



TopModel

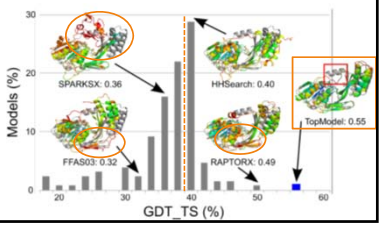
Refining solutions



1. TopScore tells us where we are wrong
2. Delete bad regions from initial predictions
3. Put good pieces together into new models and repeat

Advantages: Converge on the best - not on the average
Disadvantages: Speed

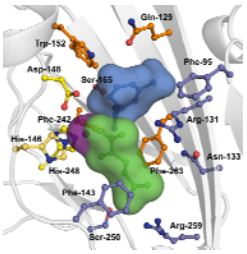
Prospective modeling of *Sa*NSR



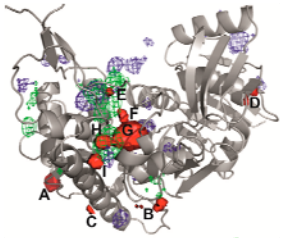
From sequence to structure

Prospective analysis

Ectoin hydroxylase



Alanopine dehydrogenase



- **Prediction of binding mode and catalytic residues**
- **Experimentally validated**

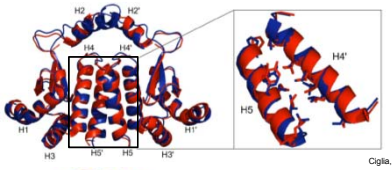
Widderich, N. ..., Mulnaes, D., Gohlke H. et. al. J. Mol. Biol. (2014)

Gohlke H. ..., Mulnaes D. et. al. J. Chem. Inf. Model. (2013)

From sequence to structure

Prospective analysis

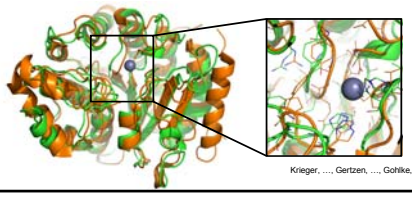
CTD of Hsp90



RMSD
Overall: 0.8 A
Interface: 0.7 A

Ciglia, ..., Gohlke, PLOS ONE 2014.

HDAC6



RMSD
Overall: 2.9 A
Active site: 1.6 A

Krieger, ..., Gertzen, ..., Gohlke, ... Hansen, J Med Chem 2017.

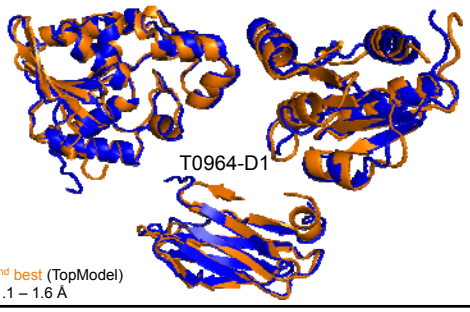
From sequence to structure

Prospective prediction for CASP13

T1016-D1

T1014-D2

T0964-D1

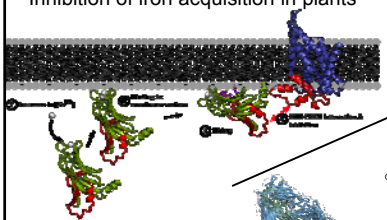


Best vs. 2nd best (TopModel)
RMSD = 1.1 – 1.6 Å

From sequence to structure

Prospective predictions on membrane proteins

Inhibition of iron acquisition in plants

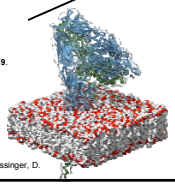


Khan, ..., Gohlke, Bauer, Ivanov, *Plant Physiol.* **2019.**

Dimer models of TGR5 in live cells

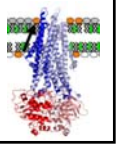
Greife, A., ..., Gertzen C.G.W., Spomer L., Häussinger D., Keitel, V., Gohlke H., Seidel C.A.M., *Sci Rep* **2016.**

Integrin activation by bile acids



Gohlke, H., ..., Häussinger, D. *Hepatology* **2013.**

Mechanism of ABCB4

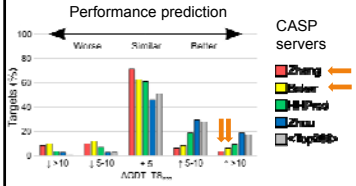


Dröge, C., ..., Gohlke, H., Schmitt, L., Kubitz, R., Häussinger, D., Keitel, V. **J. Hepatol.** 2017,

Further improvements

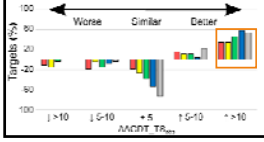
Domain identification

Performance prediction

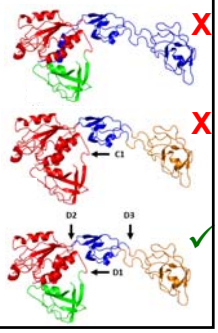


CASP servers

With domain parsing / modeling / recomb.

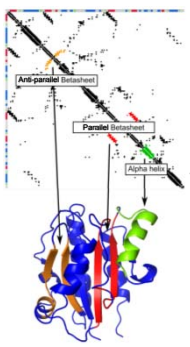


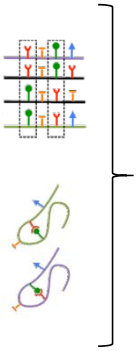
Dataset to develop TopDomain

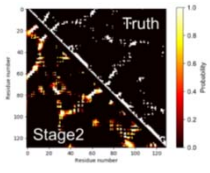


Further improvements

Coevolutionary information







Conclusion

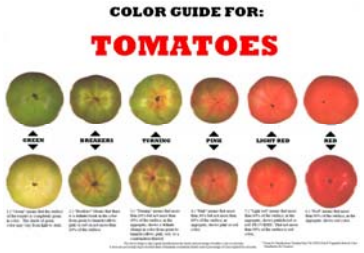
TopScore
JCTC **2018.**

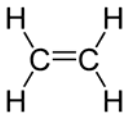
TopModel
JCTC **2020.**

Much can (still) be gained in protein structure prediction by

- optimally combining different methods,
- using ‚big data' training sets and
- deep neural networks.

Ethylene and ripening





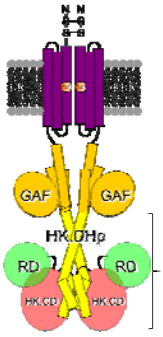
- Reducing concentration of ethylene extends 60% postharvest life
- Efforts:
  - -Molecular competitors
  - -Ethylene absorbers
  - -Temperature

https://youtu.be/k3sFpxyUwY0

Saltveit, M. (1998).

37

Ethylene ETR1



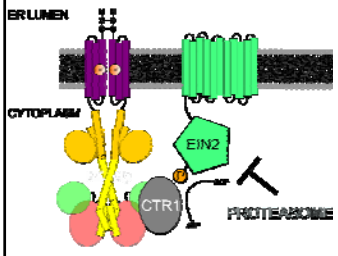
Two component system

- Transmembrane sensor domain
- Signal transmission Multimerization
- Dimerization/Phosphotransferase
- Phosphotransferase
- Catalysis

Mayerhofer, H., Panneerselvam, S., Kaljunen, H., Tuukkanen, A., Mertens, H. D. T., & Mueller-Dieckmann, J. (2015).
Milić, D., Dick, M., Mulnaes, D., Pfleger, C., Kinnen, A., Gohlke, H., & Groth, G. (2018).

38

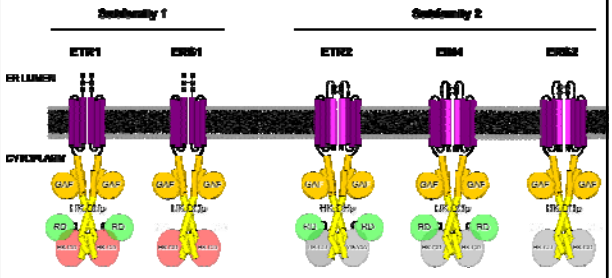
ETR1 - Mechanism of action



Light, K. M., Wisniewski, J. A., Vinyard, W. A., & Kieber-Emmons, M. T. (2016).

39

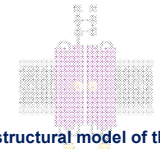
Ethylene receptor families



Shakeel, S. N., Wang, X., Binder, B. M., & Schaller, G. E. (2013).

40

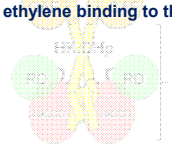
ETR1 domains



No structure/homolog available!

Alanine scanning of function related residues*

1) Make a structural model of the TM domain (N-ter 117 residues)
2) Assemble a dimeric model
3) Measure copper stoichiometry & tryptophan scanning (AG Groth)
4) Evaluate ethylene binding to the model



ERS1 crystal

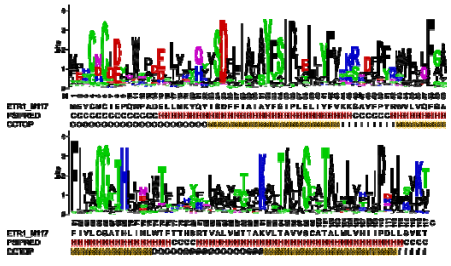
Crystallized

Crystallized

Müller-Dieckmann, H. J., Grantz, A. A., & Kim, S. H. (1999).
*Wang, W., Esch, J. J., Shiu, S.-H., Agula, H., Binder, B. M., Chang, C., … Bleecker, A. B. (2006).
Mayerhofer, H., Panneerselvam, S., Kaljunen, H., Tuukkanen, A., Mertens, H. D. T., & Mueller-Dieckmann, J. (2015).
Milić, D., Dick, M., Mulnaes, D., Pfleger, C., Kinnen, A., Gohlke, H., & Groth, G. (2018).

41

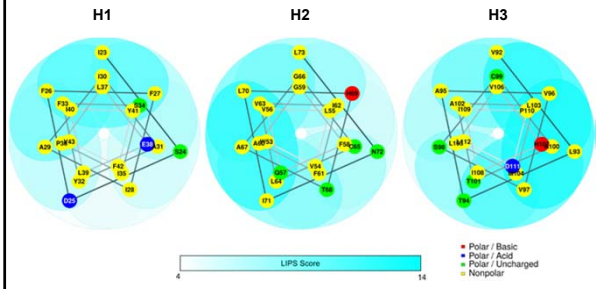
ETR1 linear predictors



Fits with topology of Cm-ERS1

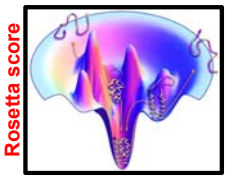
McGuffin, L. J., Bryson, K., & Jones, D. T. (2000).
Ma, B., Cui, M.-L., Sun, H.-J., Takada, K., Mori, H., Kamada, H., & Ezura, H. (2006).
Dobson, L., Reményi, I., & Tusnády, G. E. (2015).

ETR1 linear predictors



Adamian, L., & Liang, J. (2006).

Rosetta Membrane ab initio



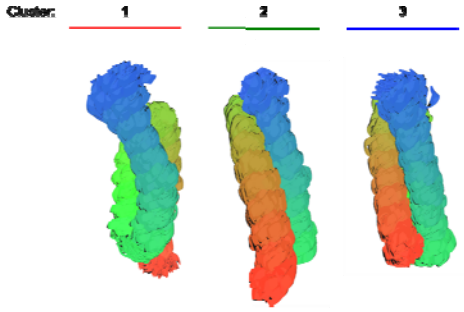
membrane_abinitio2

Sequence (1-117)
Secondary structure
Transmembrane topology
LIPS membrane accessibility

100.000 “centroid” monomer models

Yarov-Yarovoy, V., Schonbrun, J., & Baker, D. (2006).
Dill KA, MacCallum JL. (2012).

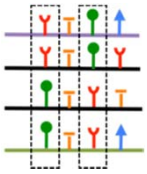
Model clusters



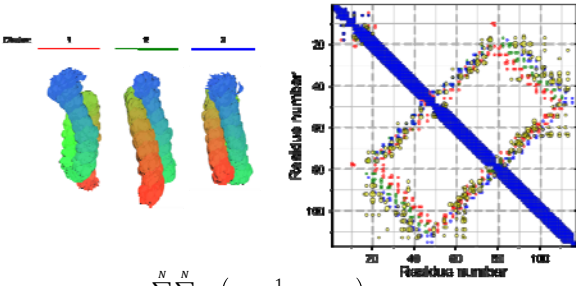
RMSD$_{15\text{-}117}$ 1-22 Å

Protein co-evolution

What is the probability of two residues being in contact given that they coevolve?

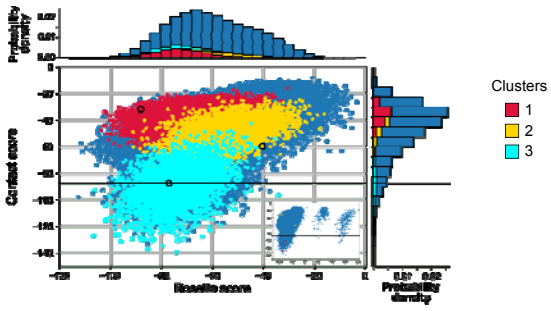


ETR1 contact prediction



$$Contact\ score\ (c,r) = \sum_{i}^{N}\sum_{j\neq i}^{N} c_{i,j}\left(\frac{1}{1+e^{-3(r_{i,j}-8\,\text{Å})}}-1\right)$$

Model distributions

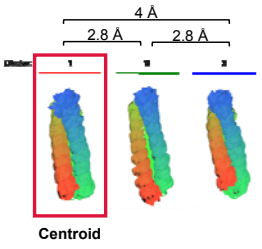


From 100.000 to 5.217 models

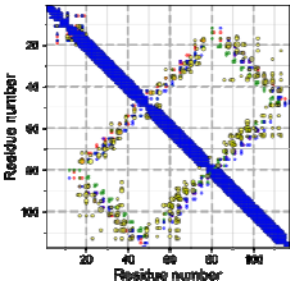
Antala, S., Ovchinnikov, S., Kamisetty, H., Baker, D., & Dempski, R. E. (2015).
Zhang, T., Liu, J., Fellner, M., Zhang, C., Sui, D., & Hu, J. (2017).

48

Filtered model clusters



$\langle RMSD^2 \rangle^{1/2}$ = 3.1-3.2 Å



Only left handed models remain!

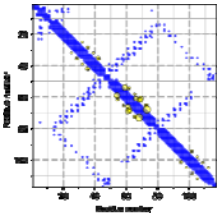
49

Dimer model

-Metal binding sites are buried
-Homodimers cause contact signal in neighboring residues
-Helix lipophilicity

HADDOCK

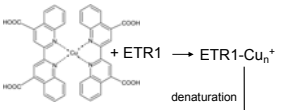
REMD / GBSW
Implicit membrane refinement

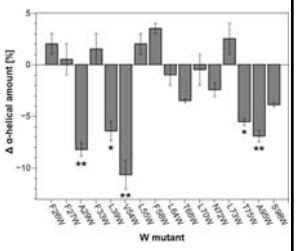


Dutta, A., & Bahar, I. (2010).
Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003).
Im, W., Lee, M. S., & Brooks, C. L. (2003).

50

Copper stoichiometry & Tryptophan mutants



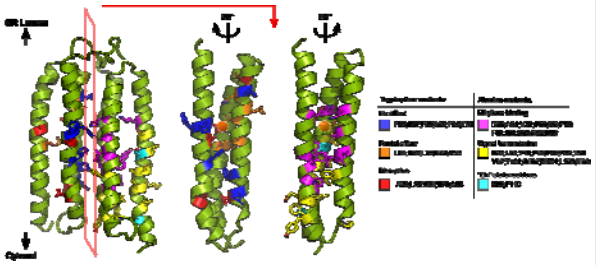


1 Cu(I) per ETR1 monomer

Tryptophan mutants show secondary structure disruption

51

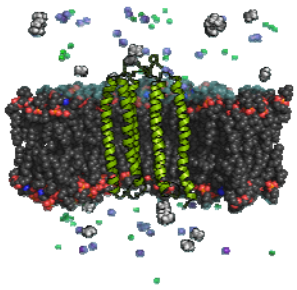
ETR1 mutants in the dimer model



Wang, W., Esch, J. J., Shiu, S.-H., Agula, H., Binder, B. M., Chang, C., ... Bleecker, A. B. (2006).

52

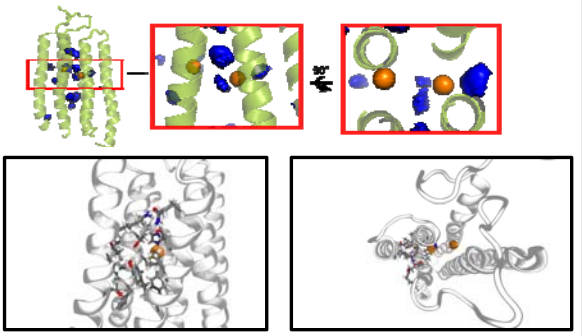
ETR1 / ethylene molecular dynamics



10 replicas of 1 µs
24 C_2H_4 molecules
DOPC:DOPE 3:1
2 Cu^+ ions placed proximal to C65/H69

53

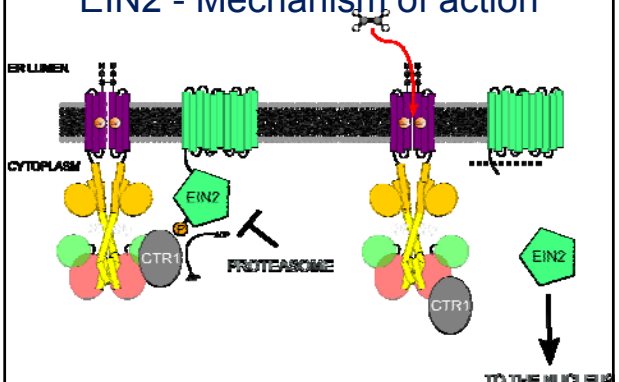
ETR1 / ethylene binding



Without incorporating ethylene-copper electron exchange!

54

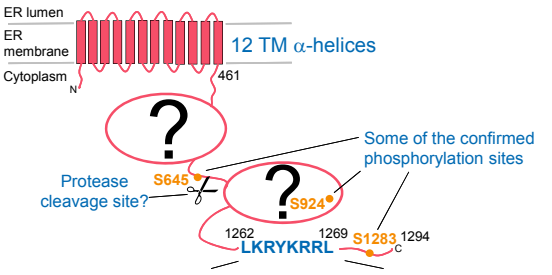
EIN2 - Mechanism of action



Light, K. M., Wisniewski, J. A., Vinyard, W. A., & Kieber-Emmons, M. T. (2016).

57

There are still many unknowns about structural organization of EIN2, but …

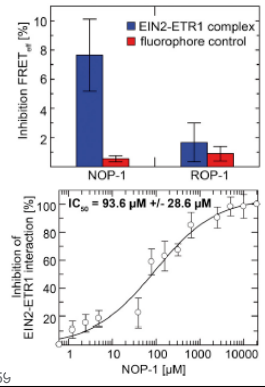


Nuclear localization signal (NLS)

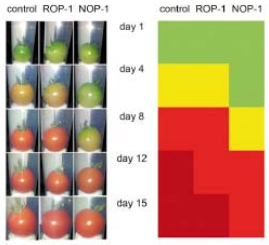
- necessary for EIN2 C-END nuclear localization
- crucial for translational repression of EBF1/2
- required for formation of EIN2–ETR1 complex

58

NLS-based peptides inhibit EIN2–ETR1 complex formation



NLS-based octapeptide:
NOP-1 = H_2N–LKRYKRRL–$CONH_2$
Random octapeptide:
ROP-1 = H_2N–EFLYMSVN–$CONH_2$

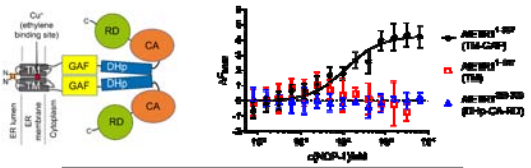


Bisson & Groth (2015). *Mol. Plant* **8**, 1165.
Bisson *et al.* (2016). *Sci. Rep* **6**, 30634.

59

GAF domain of AtETR1 is involved in binding of the inhibitory peptide NOP-1

Microscale thermophoresis - **AtETR1 constructs** labeled with thiol-reactive **Alexa Fluor 488**

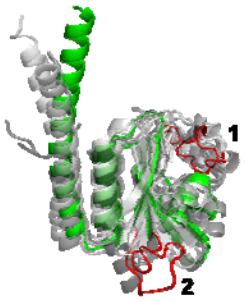


Truncated AtETR1 (amino acid range)	AtETR1 (sub)domains	Peptide	K_d/nM
full length: 1–738	TM–GAF–DHp–CA–RD	NOP-1	88 ± 41
1–589	TM–GAF–DHp–CA	NOP-1	101 ± 37
1–407	TM–GAF–DHp	NOP-1	83 ± 25
1–307	TM–GAF	NOP-1	104 ± 24
1–157	TM	NOP-1	No binding
306–738	DHp–CA–RD	NOP-1	No binding

62

Model of the GAF domain

Monomer generation



Generated by TopModel
Evaluated by TopScore

GAF (71% correct)
Templates
Flexible regions (~50% error)

66

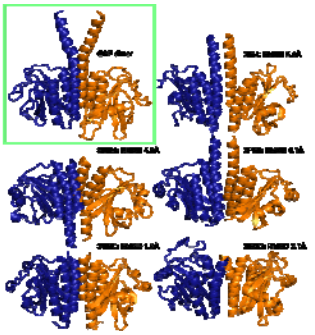
Model of the GAF domain

Dimer generation

ETR1 forms a dimer mediated by their GAF domains

TopDock (unpublished)

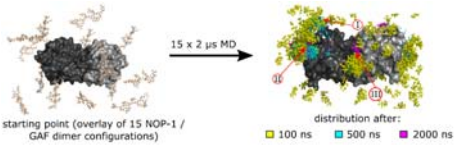
- Struct.-based hom.search
- Predict interface contacts
- Restrained docking



67

Binding site prediction

MD simulations of free ligand diffusion

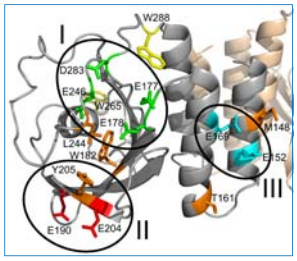


Three sites predicted

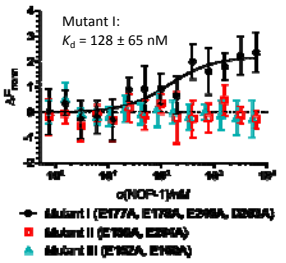
68

Site I is NOT involved in the binding of NOP-1

Predicted NOP-1 binding sites in the GAF domain



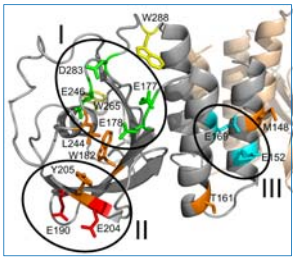
Microscale thermophoresis
AtETR1$^{1–307}$ Ala mutants + NOP-1



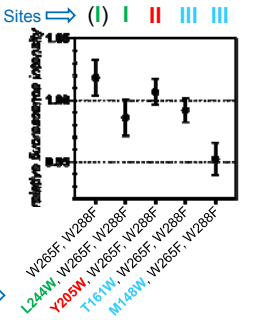
69

Site III is the binding site of NOP-1

Predicted NOP-1 binding sites in the GAF domain



Tryptophan fluorescence quenching
AtETR1$^{1–307}$ Trp mutants + NOP-1 (in 10-fold excess)

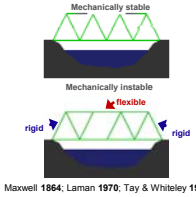


Tryptophan fluorescence reporters ⇨

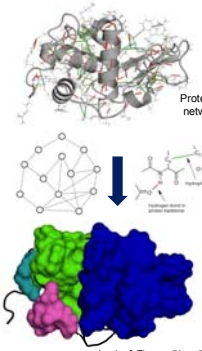
70

Potential influence of NOP-1 on ETR1 signal transmission

Structural frameworks



Constraint Network Analysis (CNA)

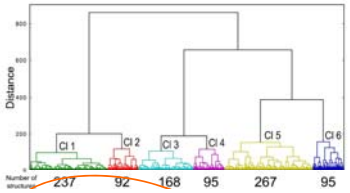


Maxwell **1864**; Laman **1970**; Tay & Whiteley **1984**; Katoh & Tanigawa **2009**.

Jacobs & Thorpe, *Phys. Rev. Lett.*, **1995**
Jacobs et al., *Proteins*, **2001**

71

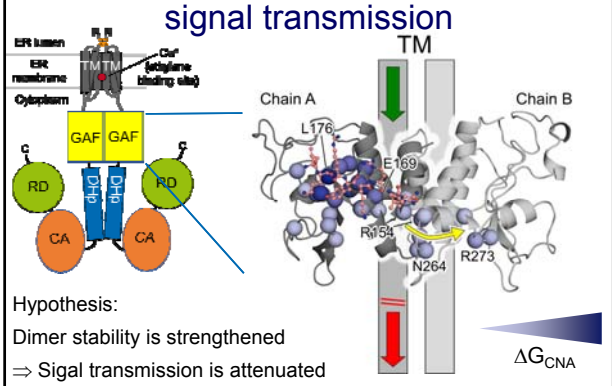
Potential influence of NOP-1 on ETR1 signal transmission



Clustering of NOP-1 binding modes at site III

73

Potential influence of NOP-1 on ETR1 signal transmission



Hypothesis:

Dimer stability is strengthened

⇒ Sigal transmission is attenuated

Milić, D., Dick, M., Mulnaes, D., Pfleger, C., Kinnen, A., Gohlke, H., Groth, G., *Sci. Rep.* **2018**

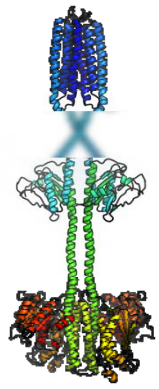
Conclusions

- Atomistic model for ETR1 TM sensor domain
- Model shows agreement with experimental mutations
- Predicted ethylene binding pathway

- Identified the interacting domain of NOP-1 – GAF
- Provided a model for how NOP-1 can interfere with signal transmission in ETR1

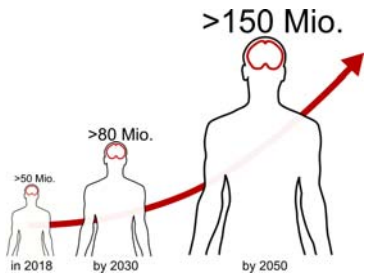
Outlook

- TM/cytoplasmic domains connection
- Receptor-copper-ethylene interaction
- Full receptor model of signal transmission



The global impact of dementia

Alzheimer's disease is the leading cause of dementia



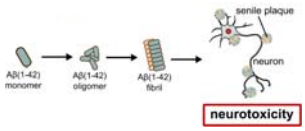
Alzheimer's Association®, *2019 Alzheimer's disease facts and figures*, 2019.
Alzheimer's Disease International, *World Alzheimer Report 2018*, 2018.

The global impact of dementia

Alzheimer's disease is the leading cause of dementia

- Currently, there is no cure for AD
- Changes in the brain begin 20 years or more before any AD-related symptoms are expected to occur
- Key to slowing or stopping the progression of AD is a therapeutic intervention in the very early stages

⇒ It is essential to conclusively recognize AD-related pathological hallmarks



Sanabria-Castro, A., *et al.*, Ann Neurosci, 2017
Selkoe, D.J., Hardy, J., EMBO Mol Med, 2016

Detection of amyloid fibrils

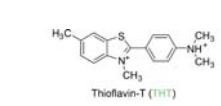
THT and CR are well-known as potent fluorescent dyes for amyloids

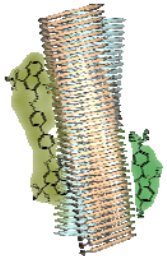
The exact nature of how both probes bind to the Aβ(1-42) fibril remained elusive.

Essential for the systematic search for novel molecular probes.

Detection of amyloid fibrils

THT and CR are well-known as potent fluorescent dyes for amyloids

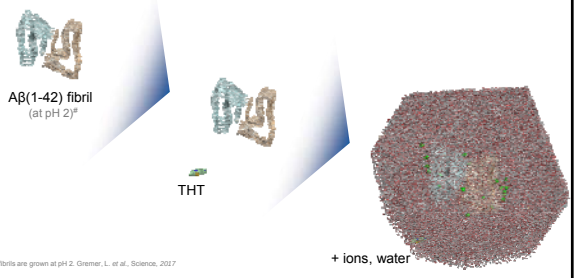




Gremer, L. *et al.*, Science, 2017

MD simulations of dye binding

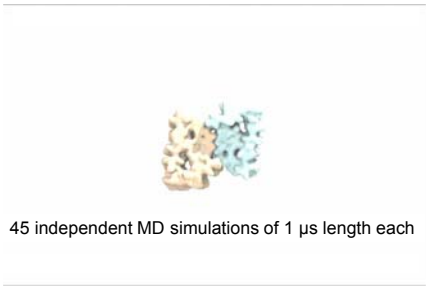
The dyes spontaneously recognize energetically preferred binding epitopes



#The fibrils are grown at pH 2. Gremer, L. *et al.*, Science, 2017

Simulations of THT binding

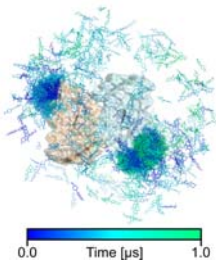
THT repeatedly binds to and unbinds from the Aβ(1-42) fibril



45 independent MD simulations of 1 µs length each

MD simulations of dye binding

Towards a binding mode model



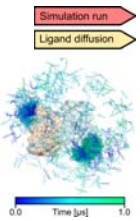
225,000 conformations from MD ensemble

?

binding mode model
structural and energetic characterization

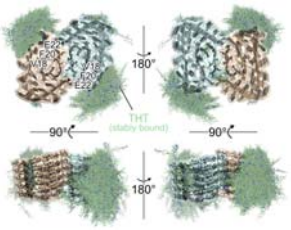
MD simulations of dye binding

Towards a binding mode model



MD simulations of dye binding

Towards a binding mode model



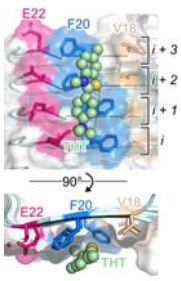
225,000 conformations from MD ensemble

stably bound conformations

binding mode model
structural and energetic characterization

THT predominantly binds to the ^{18}VFE^{22} motif

THT likely binds to the fibril surface



Free energy of binding → $\Delta G_{bind.} = RT \ln(K_D)$ ← Dissociation constant

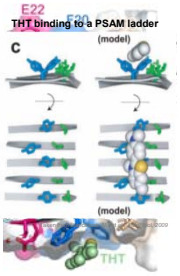
R, T = constants

$K_{D,comp.}$	$K_{D,exp}$
112 - 470 nM (ΔG^0_{bind} = -9.06 ± 0.42 kcal mol^{-1})	790 - 1740 nM (THT to Aβ(1-40) fibrils)

Rodriguez-Rodriguez, C., *et al.* Metallomics, 2015
Yona, R. L., *et al.*, ChemMedChem, 2008

THT predominantly binds to the ^{18}VFE^{22} motif

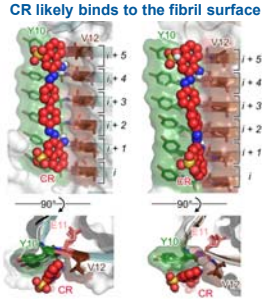
The proposed binding site agrees with experimental observables



- Relative fast binding kinetics of THT binding to fibrils suggests that THT has easy accessibility to its binding site. LeVine III, H., Arch Biochem Biophys, 1997
- THT does not bind to monomers; a minimal THT binding site on the fibril surface covers four consecutive β-strands. Biancalana, M., et al., J Mol Biol, 2009; Wu, C., et al., J Mol Biol, 2009
- Similar side chain orientation found for THT binding to peptide self-assembly mimics (PSAM). Biancalana, M., et al., J Mol Biol, 2009
- In amyloid fibrils, THT is suggested to bind to grooves formed by side chains on β–sheets on surface. Kreba, M.R.H. et al., J Struct Biol, 2005

CR predominantly binds to the ^{10}YEV^{12} motif

CR likely binds to the fibril surface

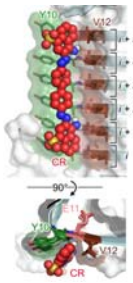


$K_{D,comp.}$	$K_{D,exp}$
0.8 - 70 nM (ΔG^0_{bind} = -11.1 ± 1.31 kcal mol^{-1})	48 - 1500 nM (CR to Aβ(1-40) fibrils)
3 - 1458 nM (ΔG^0_{bind} = -9.8 ± 1.8 kcal mol^{-1})	

Zhen, Z., et al., J Med Chem, 2009; Wang, Y., et al., J Labelled Compd Radiopharm, 2008

CR predominantly binds to the ^{10}YEV^{12} motif

The proposed binding site agrees with experimental observables



- Different binding sites for THT-like and CR-like ligands. Groenning, M., J Chem Biol, 2010
- Most studies suggest that the long axis of CR is perpendicular to the direction of the β-strands. Groenning, M., J Chem Biol, 2010
- The affinity of CR-like probes decreases significantly after replacing the amino by hydroxyl groups. Cai, L., Curr Med Chem, 2007
- Planarity around the diazo group is likely important. Cai, L., Curr Med Chem, 2007
- Speculation of multiple CR binding sites
  1. complexed by mainly hydrophobic contacts
  2. complexed by mainly ionic contacts

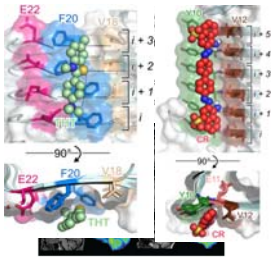
  Groenning, M., J Chem Biol, 2010

Conclusion

Towards improved diagnostic strategies of Alzheimer's disease

- Binding modes for THT and CR to the Aβ(1-42) fibril were identified by unbiased MD simulations and calculations of binding affinities
- Proposed binding modes agree with previous experimental observations
- The binding modes may provide a starting point for the systematic search and design of novel and improved molecules that bind to Aβ(1-42) fibrils

⇒ Essential for conclusively diagnosing amyloid fibrils-related diseases *in vitro*



Acknowledgment

- Daniel Mulnaes
- Stephan Schott-Verdugo
- Lena Müller
- Markus Dick
- Christopher Pfleger
- Dalibor Millic
- Georg Groth
- Benedikt Frieg
- Lothar Gremer
- Henrike Heise
- Dieter Willbold
- Computing time
- Funding