CASP15 – Ensembles and Multiple Conformational States

6:45 Gaetano Montelione - Multiple Conformational States of Proteins
7:05 Dan Rigden - Multi-Conformation Protein Targets in CASP15
7:20 Rhiju Das - Multiple Conformational States of RNA in CASP15

CASP15 Ensemble Targets

T1109-T1110 pair (contact: Mark Wilson): These were released as two separate targets. Predictors were informed that: A pair of targets T1109 and T1110 represent a mutant and a wild-type structure of the isocyanide hydratase. Difference in the two structures is caused by the structurally disruptive D183A missense mutation.

T1158 (v0-4) (contact: Sergei Pourmal): A series of 5 targets representing the type IV ABC transporter, apo and mutant with different ligands. The conformational changes are rigid body movements of domains.

T1160-T1161 (contact Shunsuke Tagami): Ancient protein reconstruction; crystallization condition induced different folds for T1160 and T1161 (43/48 residues identical) . Released as two targets.

T1195-T1197 (contact Babis Kalodimos): Three kinase targets that were solved with an advanced NMR technique. For each of them, two to three structural conformations are present.

(RNA) R1136 and R1138 (contact: Ebbe Andersen, Denmark): Released as one target each, but we warned people in the weekly digest that the structures are available in two conformations:

(RNA) R1149 and R1156 (contact: Rachael Kretsch): Alternative conformations.

T1170, H1171 (v1,2), H1172 (v1-4), (contact Jiri Wald): A series of 3 targets: T1170 (A6:B0), H1171 (A6:B1); and H1172 (A6:B2); representing different intermediate conformational states of the multimolecular complex. A 15 bp dsDNA is present in all of them.

T1189/R1189 (contact Su Zhaoming): T1190/R1190 RNA-protein complexes

Multiple Conformational Modeling in CASP

Yuanpeng Janet Huang
Roberto Tejero
Theresa Ramelot
GVT Swapna
Gaohua Liu

Kelly Brock
Chris Sander
Debbie Marks

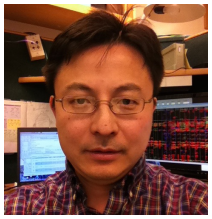

Yuanpeng Janet Huang


Roberto Tejero


Andriy Kryshtafovych

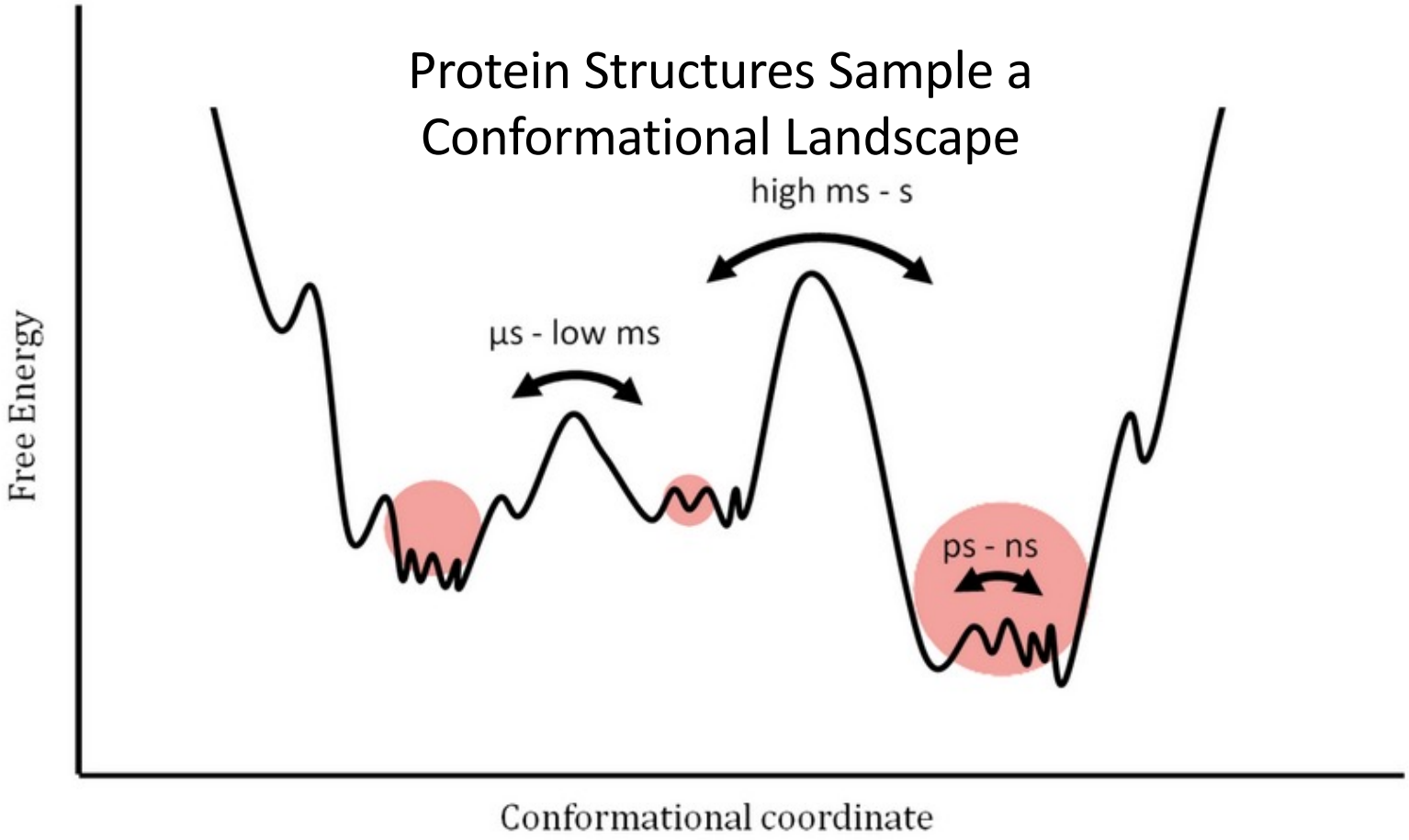

Theresa Ramelot


GVT Swapna


Gaohua Liu

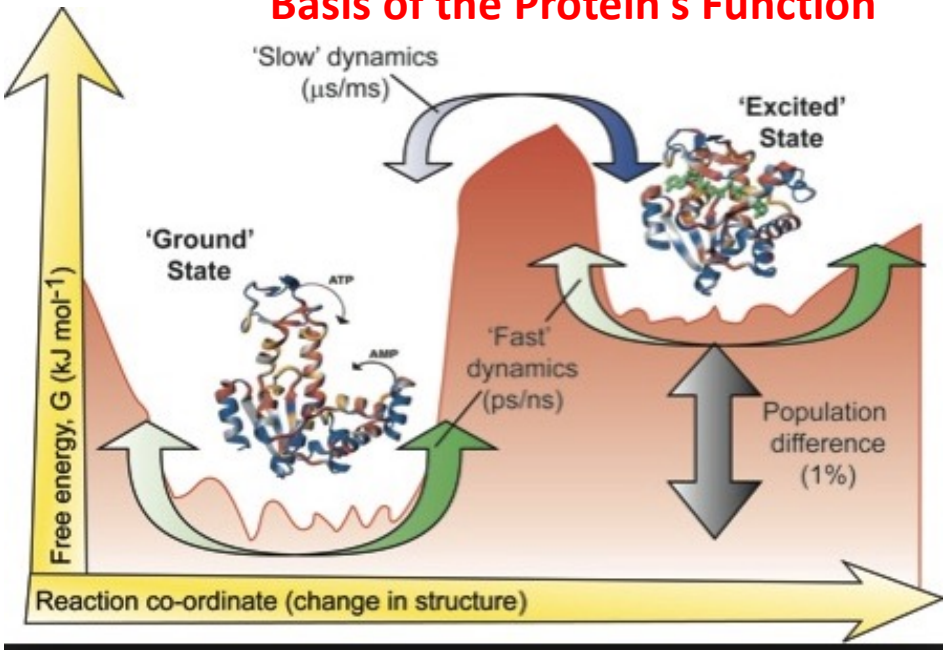
G.T. Montelione
CASP15
Antalya, Turkey
Dec 12, 2022

Protein Dynamics



www.mdpi.com

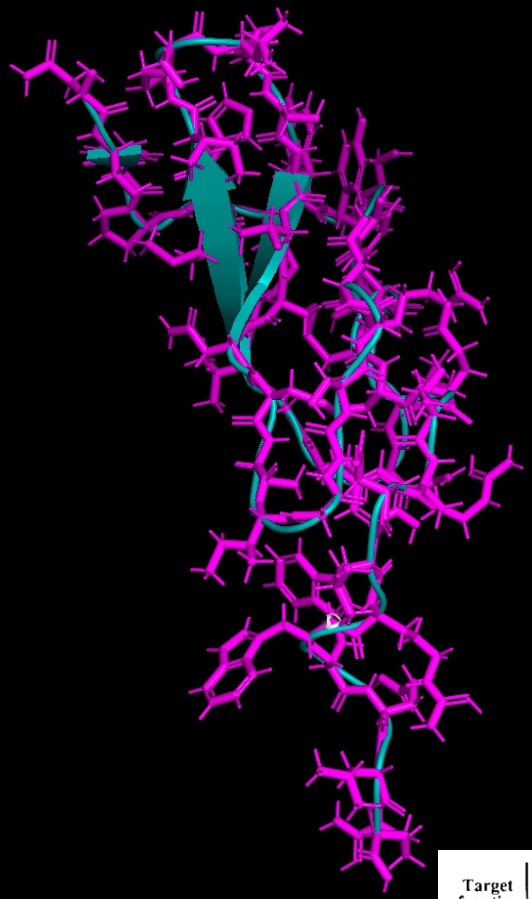
The Conformational Landscape Provides the Basis of the Protein's Function



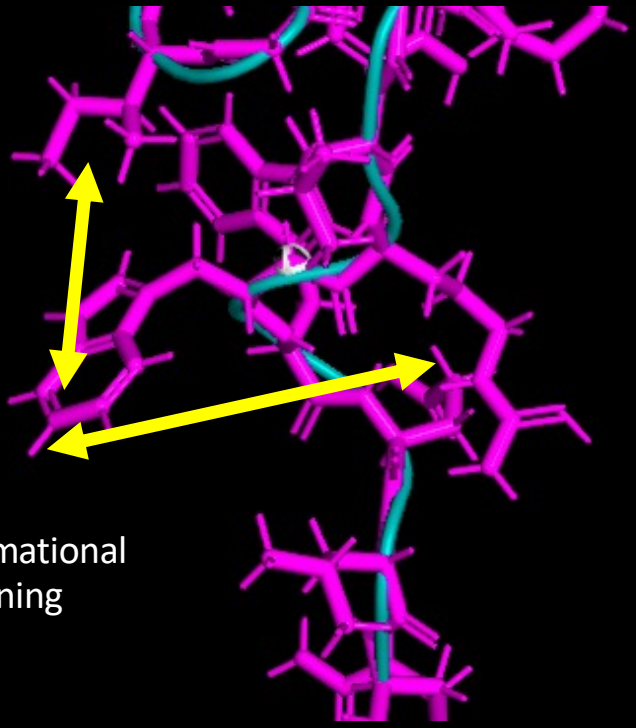
Andrew Baldwin
research.chem.ox.ac.uk

Ground Truth Data

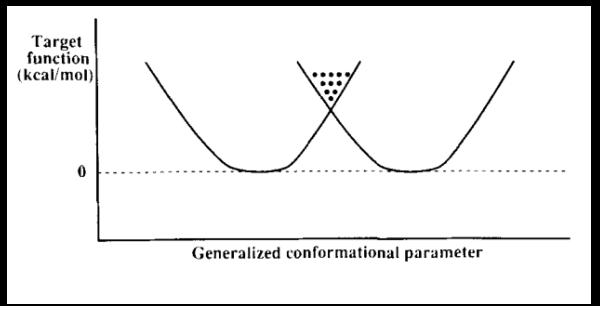
- X-ray Crystallography
  - Can see alternative conformations in different crystal forms
  - Multiple conformations in asymmetric unit
  - Can sometimes fit density to multiple conformations
  - Do crystal lattice interactions shift conformational distributions; e.g. stabilize low populated states
- CryoEM
  - Can generate multiple models from cryoEM data
  - Need a lot of data
  - Effects of freezing?
- NMR
  - Chemical shift, NOEs, RDCs, Paramagnetic effects
  - Particularly sensitive to motions
  - Limited in size (< 50 kDa)
  - Exchange broadening can make peaks “disappear”
- Fluorescence Energy Transfer, Chemical Cross Linking; Small Angle X-ray scattering



3EGF Montelione, G.T.; … Wüthrich, K.; Scheraga, H.A. Proc Natl Acad Sci 1987

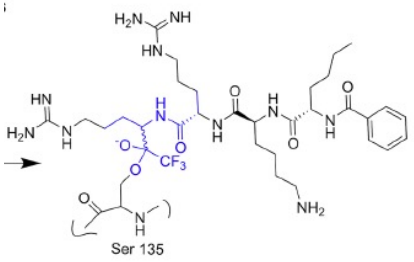


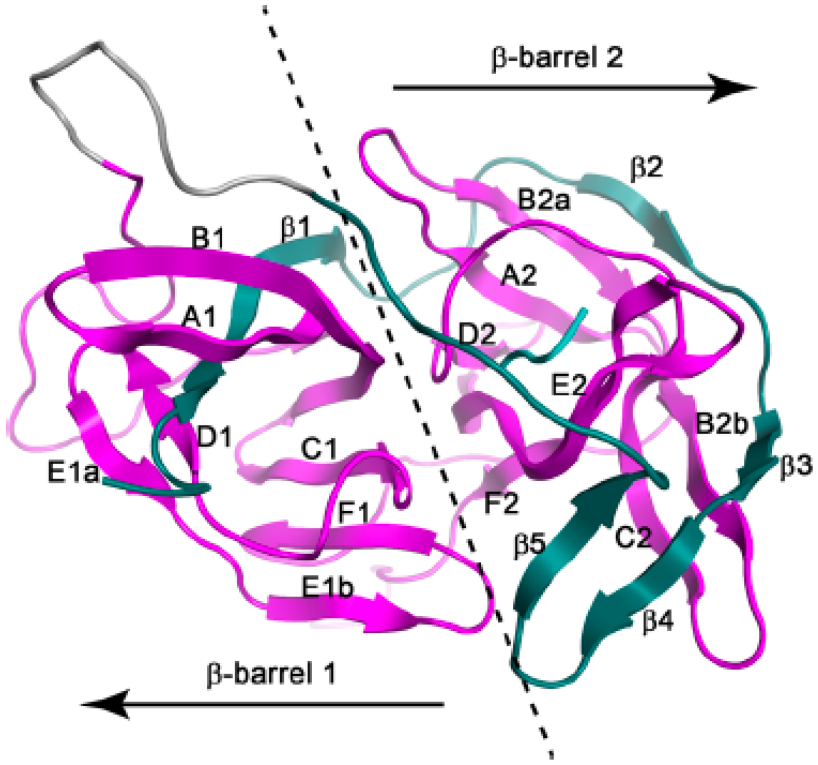
Conformational Pinning



Tejero, R.; Bassolino-Klimas, D.; Bruccoleri, R.E.; Montelione, G.T. **Protein Science** 1996

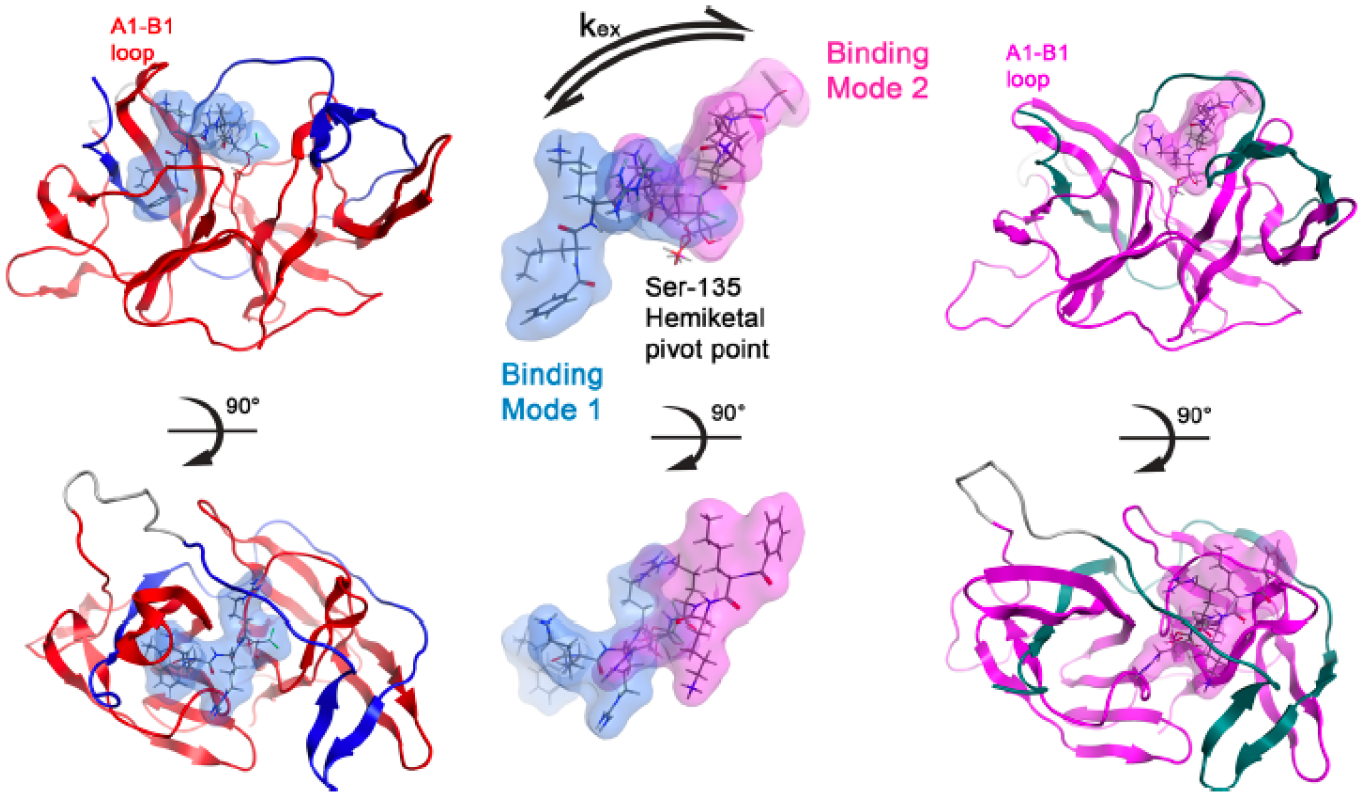
Solution NMR Structure of DENV2-NS2B-NS3pro Protease Complex
$^{2}H,^{15}N,^{13}C$, ILVA Me Labeled

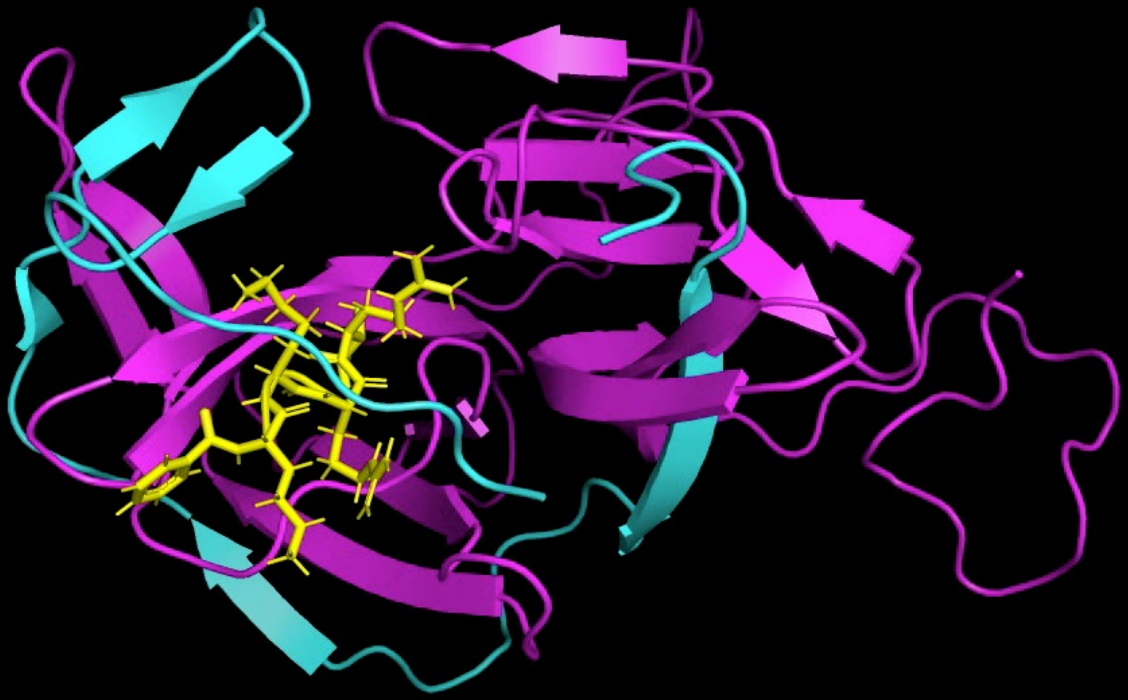




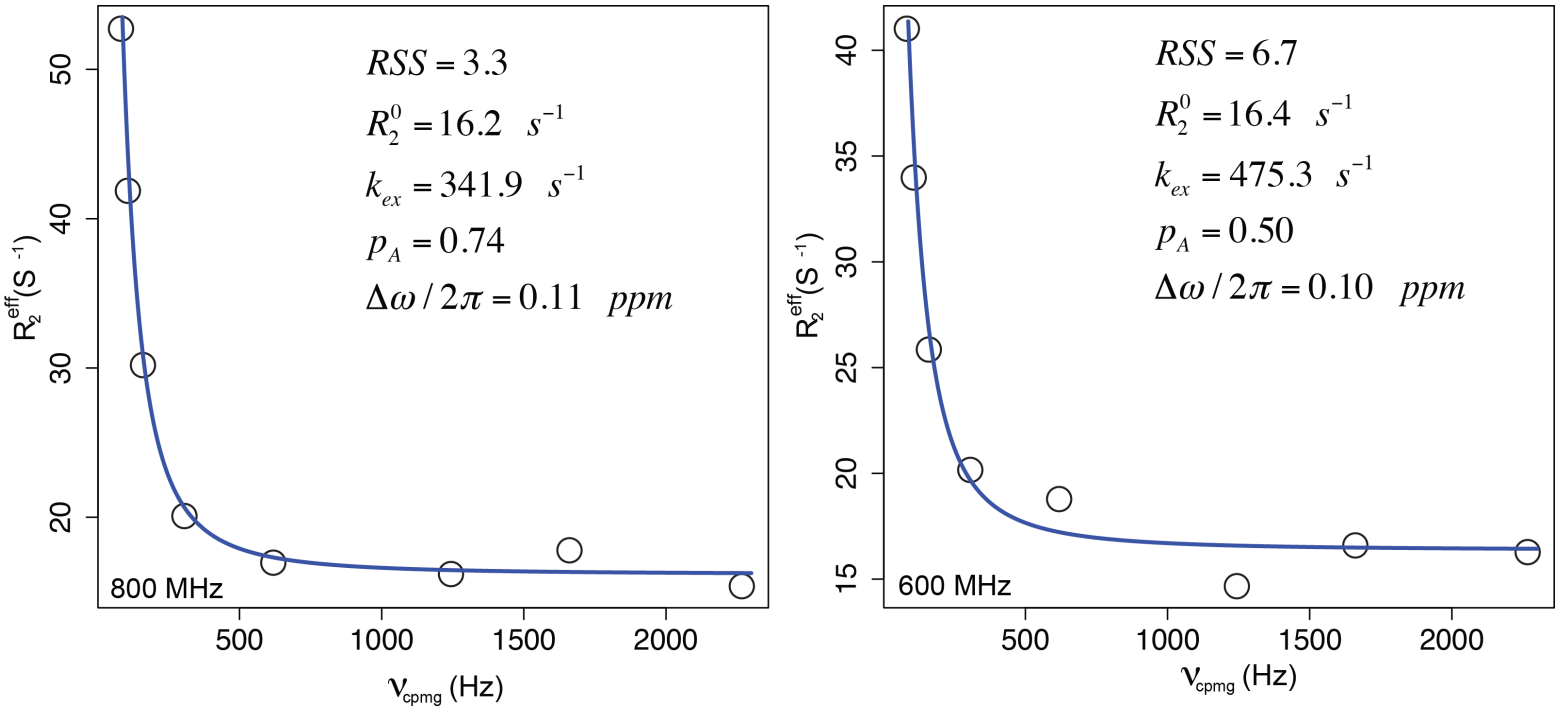
G. Liu

NMR Reveals Two Non-Overlapping Inhibitor Binding Sites in DENV2-NS2B-NS3pro Protease Complex





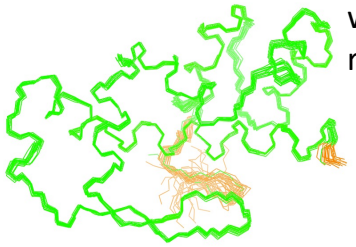
NMR Relaxation Dispersion (CPMG) Bound Inhibitor Resonances Reveal Dynamic Interconversion Between Two Sites on the Millisecond Timescale



For NMR ensemble: well-defined region based on convergence (Cyrange) and ^{15}N relaxation

CASP14 Target T1055

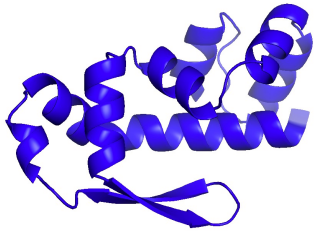
NMR Model



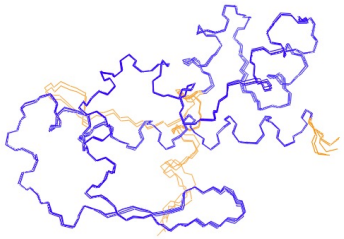
well-defined region:
residues 305 - 426

NMR Ensemble

Common Region for RMSD / GDT comparison: 310-426



AF2 Model



AF2 Ensemble

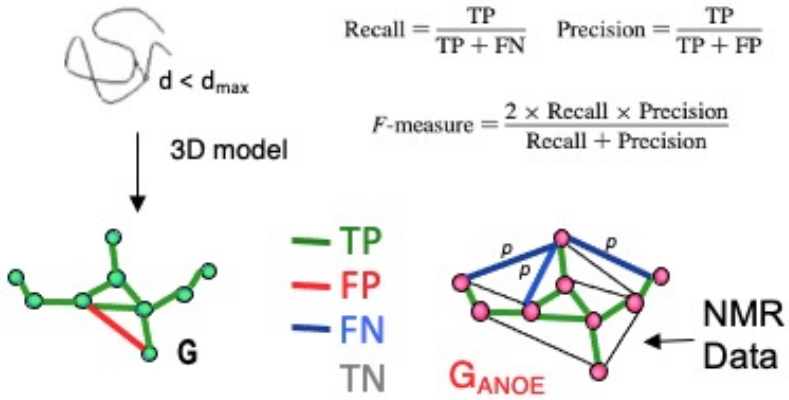
well-defined region:
residues 310-428

<RMSD> = 0.97 Å
GDT = 0.90

For AF9 ensemble well-defined region based on convergence and AF2 confidence scores

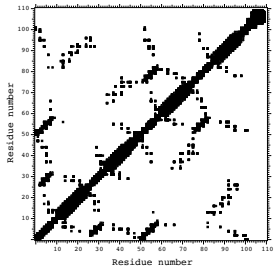
Huang YJ, et al **Proteins** 89.12 (2021): 1959-1976.
Bersch B, et al. JMB 433.13 (2021): 167009.

RPF-DP Score – a measure of how well the NOESY peak list matches to the structural model; "NMR R factor"

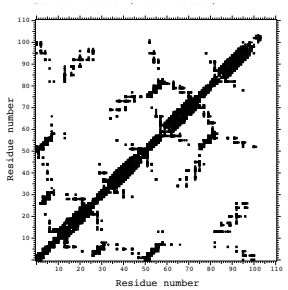


$$\text{Recall} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

$$F\text{-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$



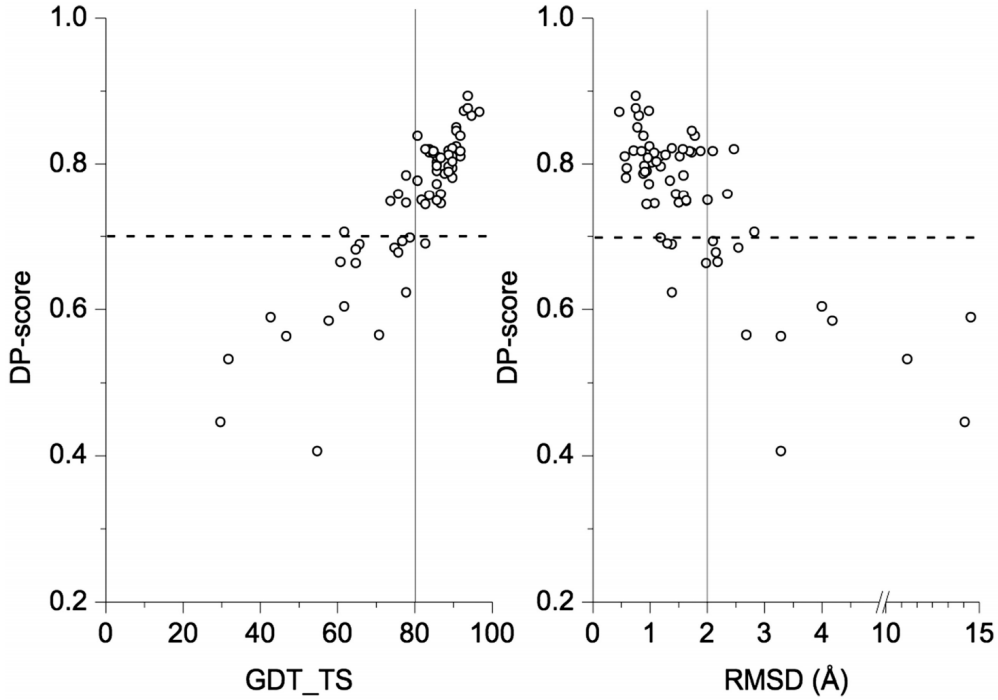
By comparing the differences between the two graphs $\bar{G}$ (derived from the structure) and G_{ANOE} (derived from the peaklists), a global measure of the goodness-of-fit of the query structures with the original peaklists can be formulated.



Similar to lDDT developed later by Schwede and coworkers

Huang, Y J ; Powers, R ; Montelione, G T **J. Amer. Chem. Soc.** 2005, 127: 1665.
Huang, Y J ; Rosato, A ; Singh, G ; Montelione, G T **Nucleic Acids Research** 2012, 40:542

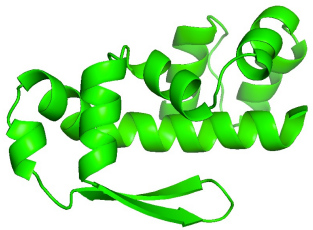
NMR DP scores correlate with structure accuracy



CASD-NMR

Comparative Recall Analysis: Compare two models against NOE data
Asks if there are NOE data that fit one mode

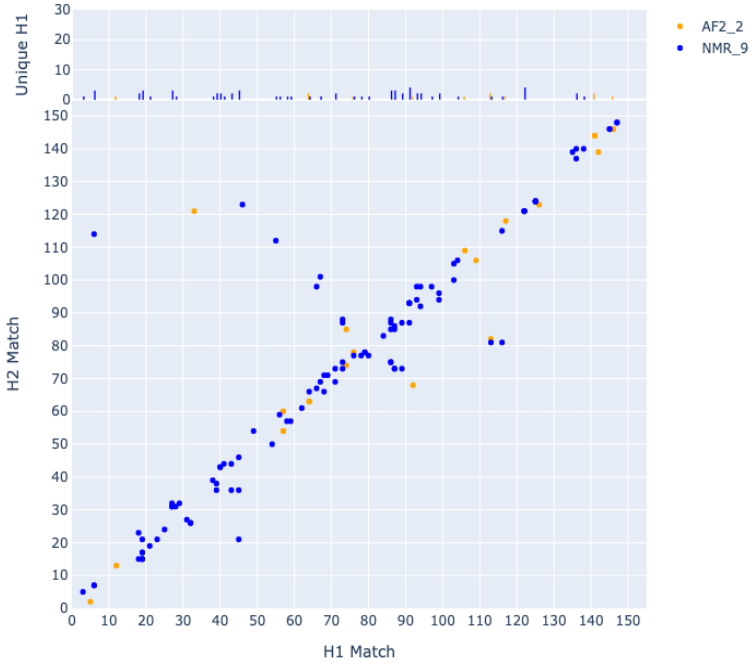
CR analysis reveals that the NMR data equally-well fit NMR, and AF models (actually slightly better fit to NMR model)

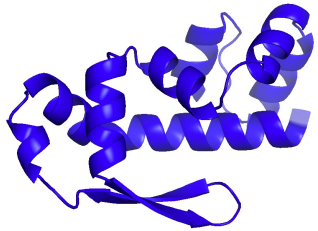


NMR Model

NOEs satisfied by AF2 but not NMR

NOEs satisfied by NMR but not AF2



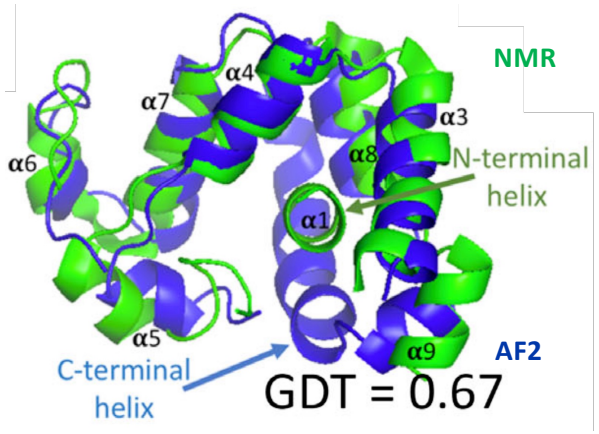


AF2 Model

YJ Huang: CASP15 Poster #2

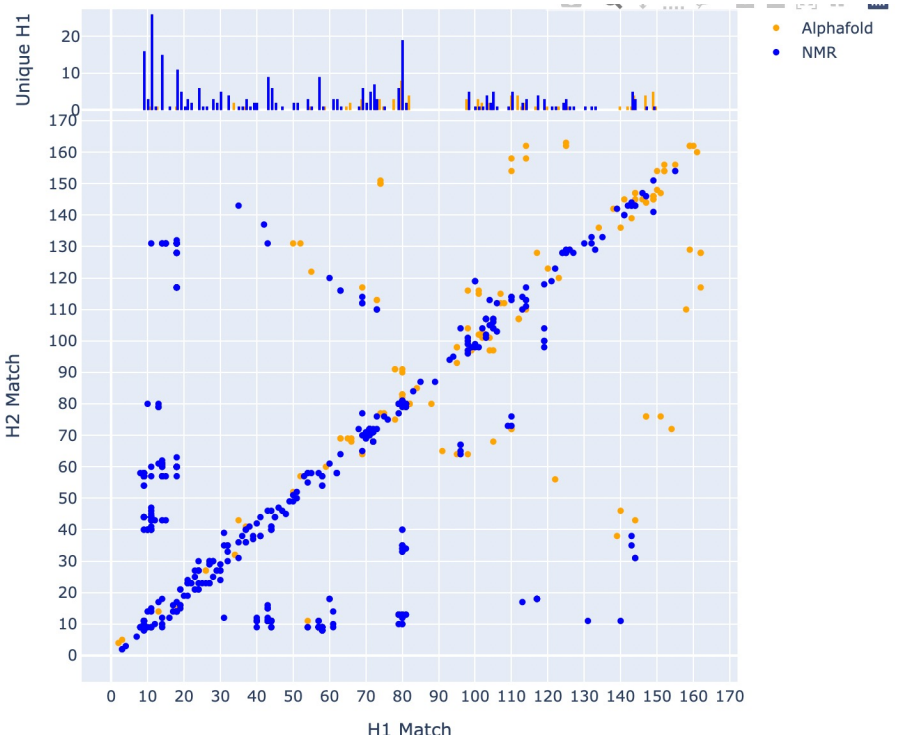
CASP14 Target T1055

Double Recall Violation Analysis



NOESY peaks consistent with NMR model; not explained by AF2 model

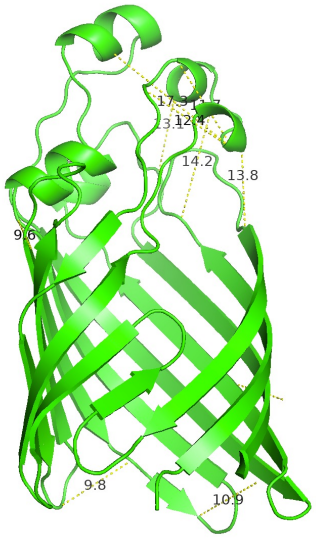
NOESY peaks consistent with AF2 model; not explained by NMR model



CASP14 Target T1027

AF appears to identify an alterative conformation present in solution.

Huang YJ, et al **Proteins** 89.12 (2021): 1959-1976**.**
PDB ID 7D2O –luciferase GLuc.
Wu N, Kobayashi N, Tsuda K, et al. **Sci Rep**. 2020, v10: 20069

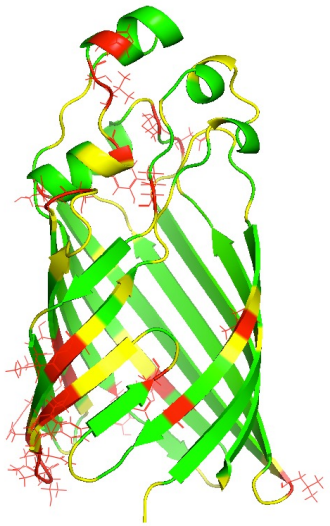


CASP14 Target T1088

Klebsiella pneumoniae MipA

many successful CASP14 methods predicted ideal β-barrel "open" form of MipA

AF2 MipA Model



Regions β2-β3-β4-β5 of AF2 models are inconsistent with NMR data

Dihedral Angle Constraints generated from TalosN from chemical shift data

Green – no violations
Yellow – no restraints from TalosN
Red – violations in >=3 models

Klebsiella pneumoniae MipA in detergent micelles

EC-NMR Structure

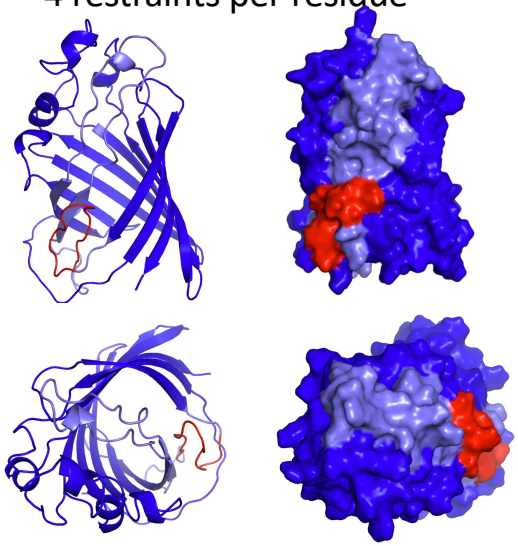
240 Residues

Perdeuterated Sparse Restraints

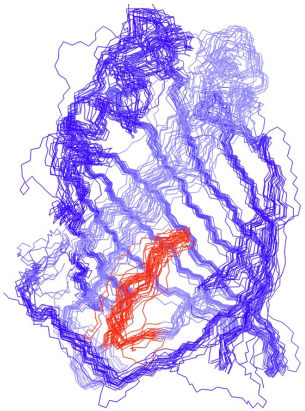
^{15}N-^{1}H , C^{α}, C^{β}, some CH_3

~ 1000 Conformational Restraints

~ 4 restraints per residue



NMR_10 Closed Conformer

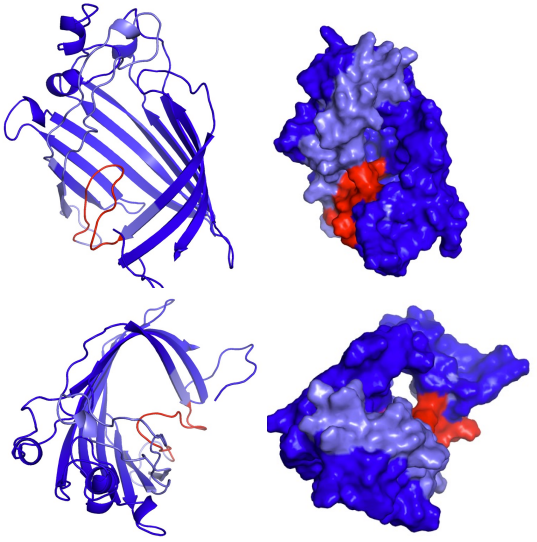


Structural variations in formation of strand β2, β3, β4, and β5

Antibiotic resistance factor

Extensive exchange broadening in red hairpin; do not see many HN-HN NOEs

Chemical shift data indicate β3-β4 hairpin is not beta strand -> rather indicate a dynamic local structure

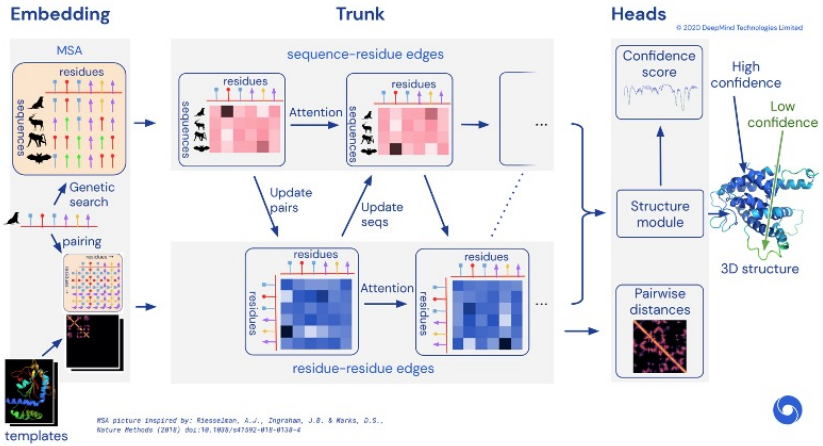


NMR_13 Open Conformer

Exploring AlphaFold to Predict Multiple Conformational States

attention-based machine learning

Use shallow MSAs to provide subsets of ECs



John Jumper
CASP14 Talk Dec 2020

eLife

NEWSLETTER ABOUT

HOME MAGAZINE COMMUNITY INNOVATION

Short Report
Structural Biology and Molecular Biophysics

Sampling alternative conformational states of transporters and receptors with AlphaFold2

Diego del Alamo, Davide Sala, Hassane S Mchaourab, Jens Meiler

Received: 6 January 2022 | Revised: 7 April 2022 | Accepted: 26 April 2022

DOI: 10.1002/prot.26382

PROTEINS WILEY

RESEARCH ARTICLE

Multi-state modeling of G-protein coupled receptors at experimental accuracy

Lim Heo | Michael Feig

Prediction of multiple conformational states by combining sequence clustering with AlphaFold2

Hannah K. Wayment-Steele[1,2], Sergey Ovchinnikov[3], Lucy Colwell[4,5], Dorothee Kern[1*]

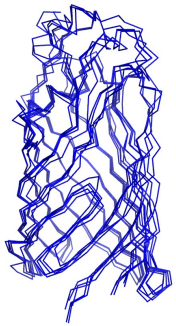
BioRxiv 2022

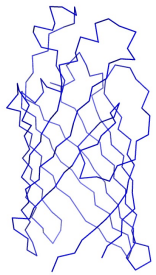
AF_ALT generates three clusters of models
structural variation in strands β2, β3, β4, β5 region

AF_ALT_1
cluster 1
DP max = 0.59

AF-ALT_2
cluster 2
DP_max = 0.61

AF_ALT_3
cluster 3
DP_max = 0.58



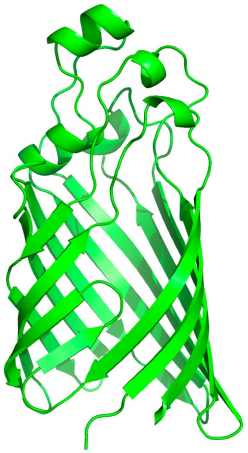


YJ Huang

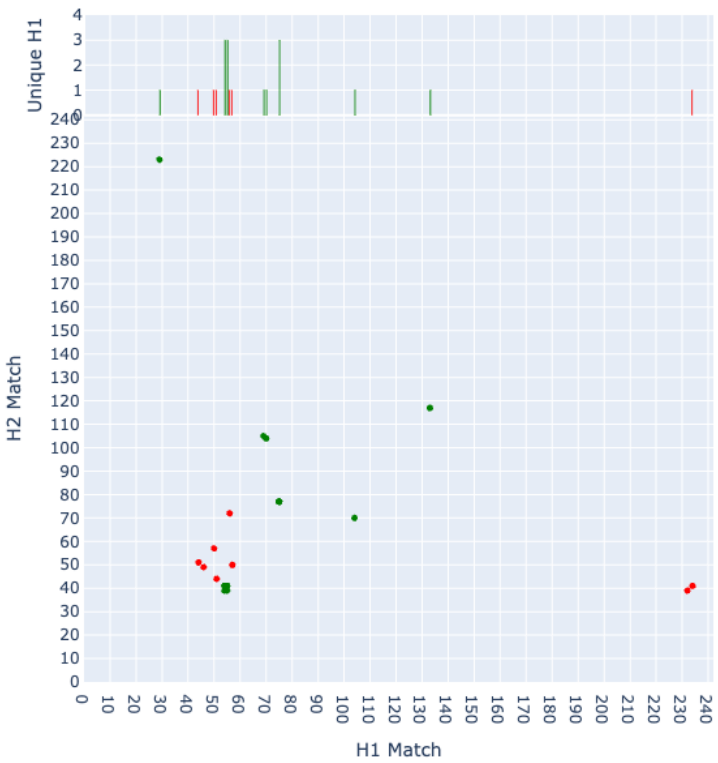
Comparative Recall Analysis

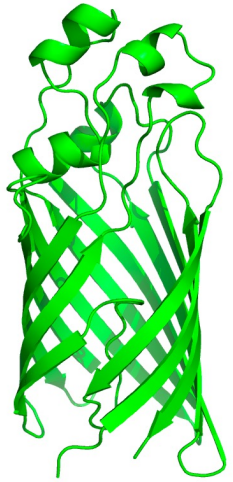
NOEs satisfied by ALT1 but not ALT2

NOEs satisfied by ALT2 but not ALT1



AF_ALT_1
DP = 0.59
Open Form



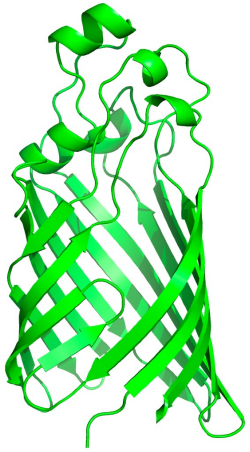


AF_ALT_2
DP = 0.61
Closed Form

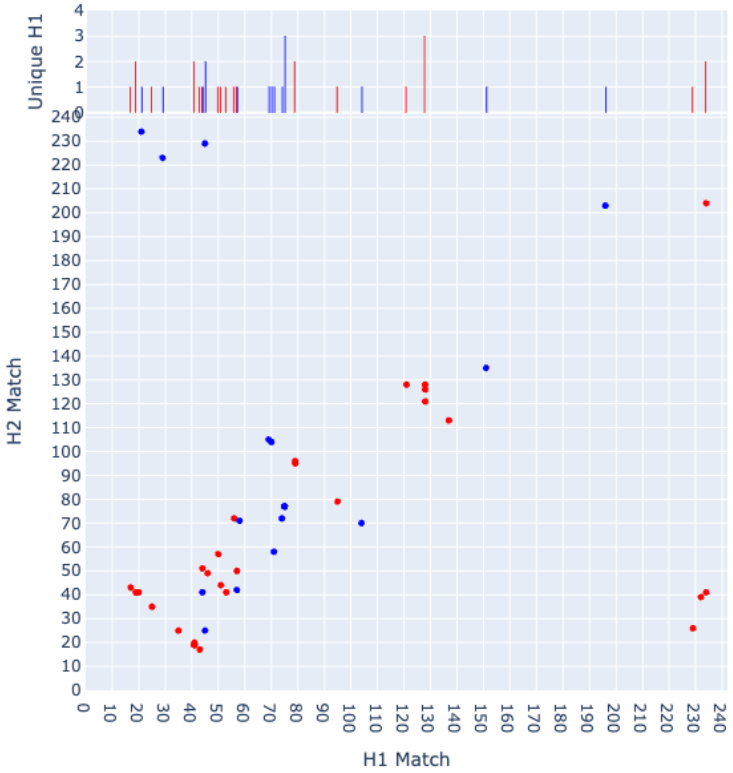
Comparative Recall Analysis

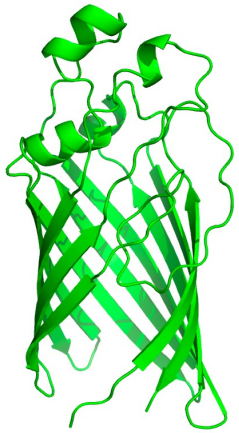
NOEs satisfied by ALT1 but not ALT3

NOEs satisfied by ALT3 but not ALT1



AF_ALT_1
DP = 0.59
Open Form



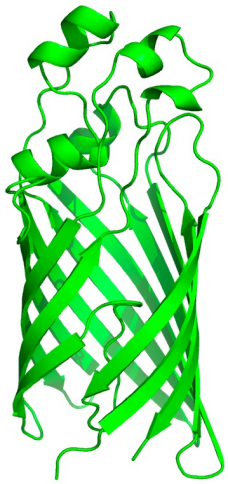


AF_ALT_3
DP = 0.58
Open Form

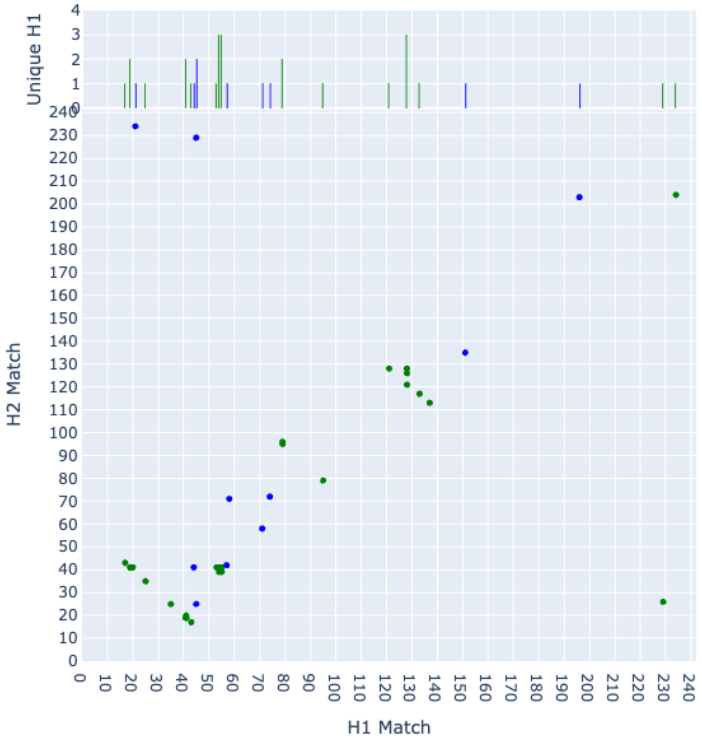
Comparative Recall Analysis

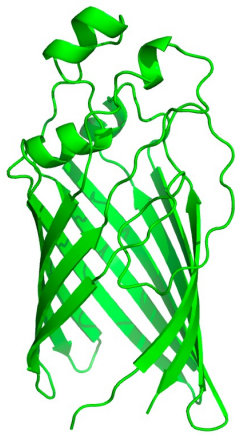
NOEs satisfied by ALT2 but not ALT3

NOEs satisfied by ALT3 but not ALT2



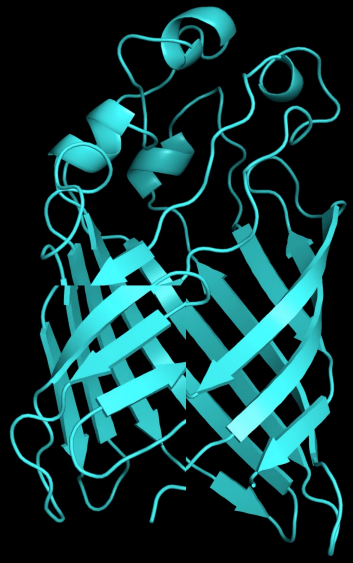
AF_ALT_2
DP = 0.61
Closed Form

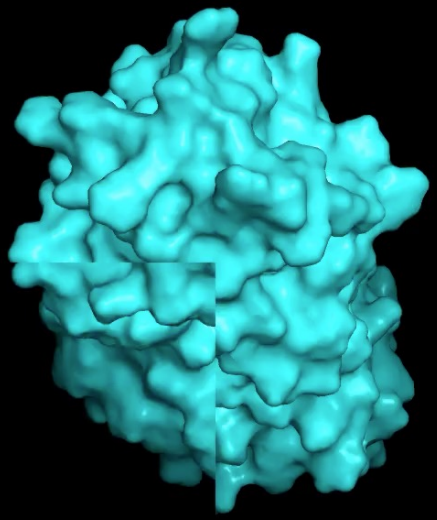




AF_ALT_3
DP = 0.58
Closed Form

Klebsiella pneumoniae MipA in detergent micelles





YJ Huang
G Liu
Y Ishida
GVT Swapna
S. McCallum

Conclusions

For CASP14 target T1055 the experimental data are about equally consistent for AF2 model or the experimentally-reported NMR model.

Similar results were obtained for 10 other proteins studied by AF2 and NMR

For CASP14 target T1027, the experimental data are not fully consistent with either the AF or experimentally-reported NMR model, but rather suggest a dynamic conformational exchange between these two conformations in solution.

For CASP14 target T1088, the NMR data are consistent with an equilibrium between multiple conformations generated by AF2-alt. All are significantly populated at 313 K

We introduce Comparative Recall Analysis for assessing how well pairs of models fit distance restraint data

Multiple Conformational Modeling in CASP

Yuanpeng Janet Huang
Roberto Tejero
Theresa Ramelot
GVT Swapna
Gaohua Liu

Kelly Brock
Chris Sander
Debbie Marks

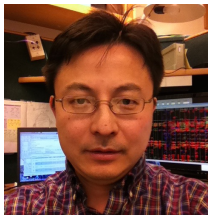

Yuanpeng Janet Huang


Roberto Tejero


Andriy Kryshtafovych


Theresa Ramelot


GVT Swapna


Gaohua Liu

G.T. Montelione
CASP15
Antalya, Turkey
Dec 12, 2022