LINDERT
[research_group]$

THE OHIO STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

THE UNIVERSITY of MISSISSIPPI
SCHOOL OF PHARMACY

GlyCORE
Glycoscience Center of Research Excellence
AT THE UNIVERSITY OF MISSISSIPPI

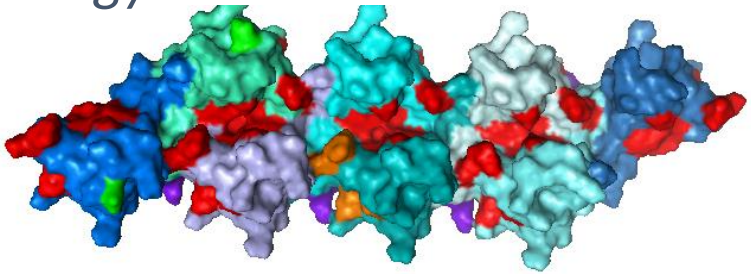
Validated determination of a protein structure by HR-HRPF combined with computational modeling

Niloofar Abolhasani Khaje, Alexander Eletsky, Sarah E. Biehn, Charles K. Mobley, Sushil Mishra, Robert J. Doerksen, Steffen Lindert, James Prestegard, Joshua S. Sharp

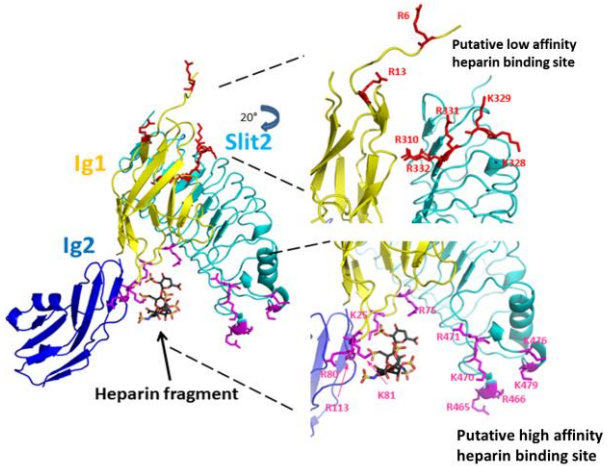
FCOI Statement: J.S.S. discloses a significant financial interest in GenNext Technologies, Inc., a growth-stage company seeking to commercialize benchtop HRPF to support the pharmaceutical industry

The Sharp Group

Developing new tools to address challenging problems in structural biology



Wang X. *et al* (2011) *Structure* **19:** 1138
Wang X. *et al.* (2013) *Progress in Molecular Biology and Translational Science*



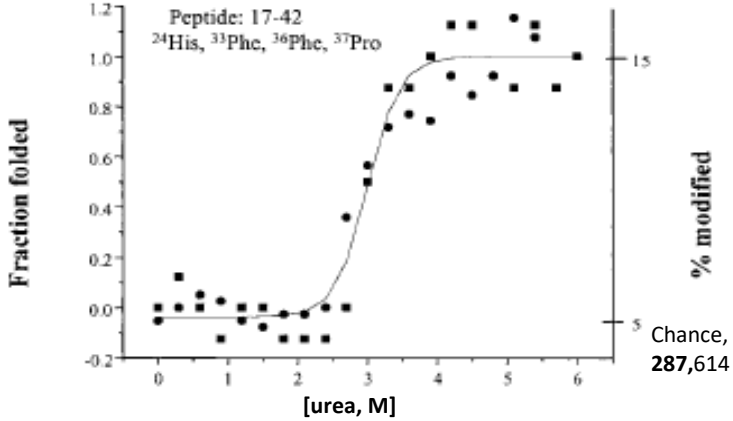
Li Z. *et al* (2015) *J Biol Chem* **290:** 10729
Zong C. *et al.* (2016) *J Am Chem Soc* **138:** 13059

- Structural biology is the backbone of our current molecular understanding of biology
- Numerous systems are highly challenging to study using traditional structural biology techniques
  - Dynamic systems
  - Heterogeneous systems
  - Large systems
  - Limited sample amounts
- Protein-carbohydrate complexes often fit into several of these categories simultaneously
  - Flexible
  - Dynamic oligomerization
  - Heterogeneous protein and/or carbohydrate ligand
  - Very large complexes
- The Sharp Lab focuses on developing and applying new tools to address these challenges in structural biology

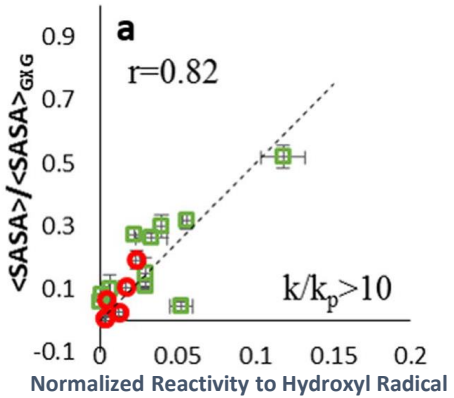
Foundation of Hydroxyl Radical Protein Footprinting

Why does it work?

- Rate of oxidation appears to primarily be a function of two factors
  - Chemical nature of oxidized residue
  - Average accessibility of oxidation target to the hydroxyl radical over the time of radical exposure
- General inherent reaction rates: sulfur-containing > aromatic > aliphatic > charged > polar > Ala > Gly
- Sequence context can influence inherent reactivity, especially for less reactive amino acids
- Changes in reaction rates in the same sequence reflects changes in average solvent accessibility

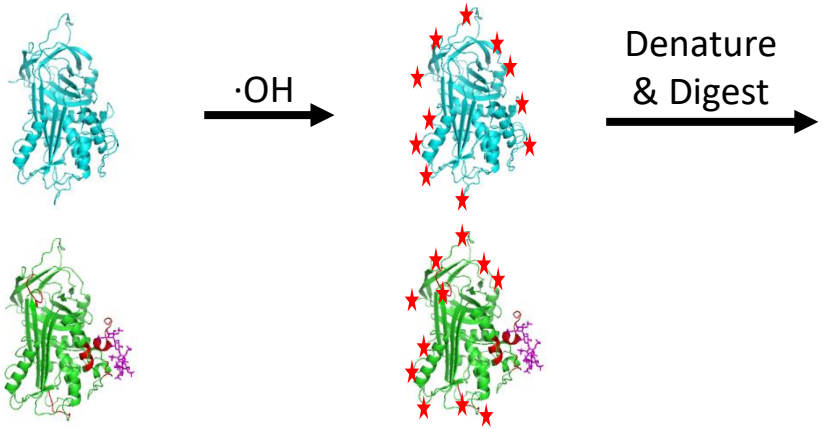


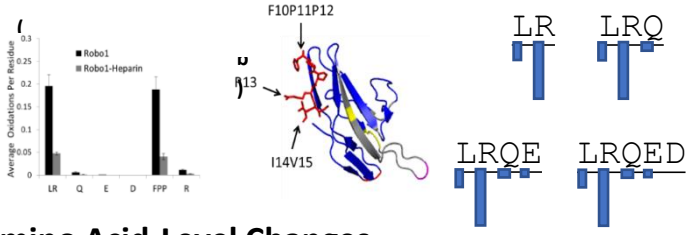
Chance, *Biochem Biophys Res Commun* **287,**614 (2001)



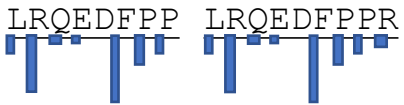
Xie *et al*, *Sci Rep* **7,**4552 (2017)

General HRPF Workflow

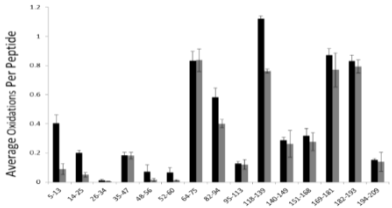




Amino Acid-Level Changes
Structural Models

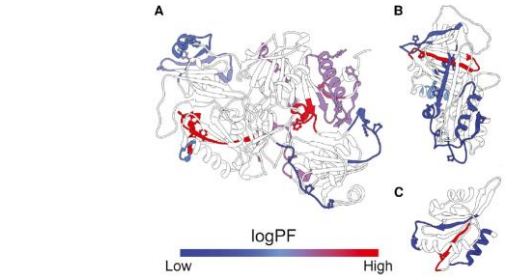


LC-MS

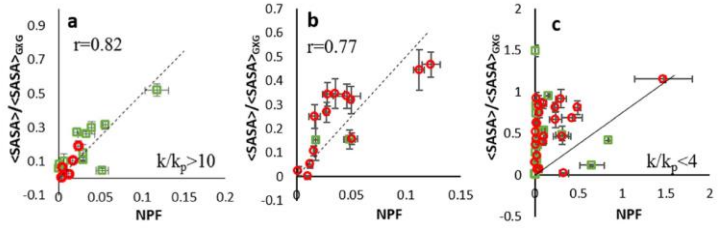


Peptide-Level Changes

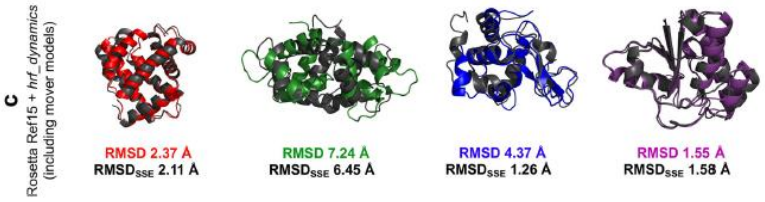
Protein Structure Determination by HR-HRPF: Highlights



Huang *et al* (2015), *Biophys J* **108:** 107; Kaur *et al* (2015), *Mol Cell Prot* **14:** 1159



Xie *et al* (2017), *Sci Rep* **7:** 4552



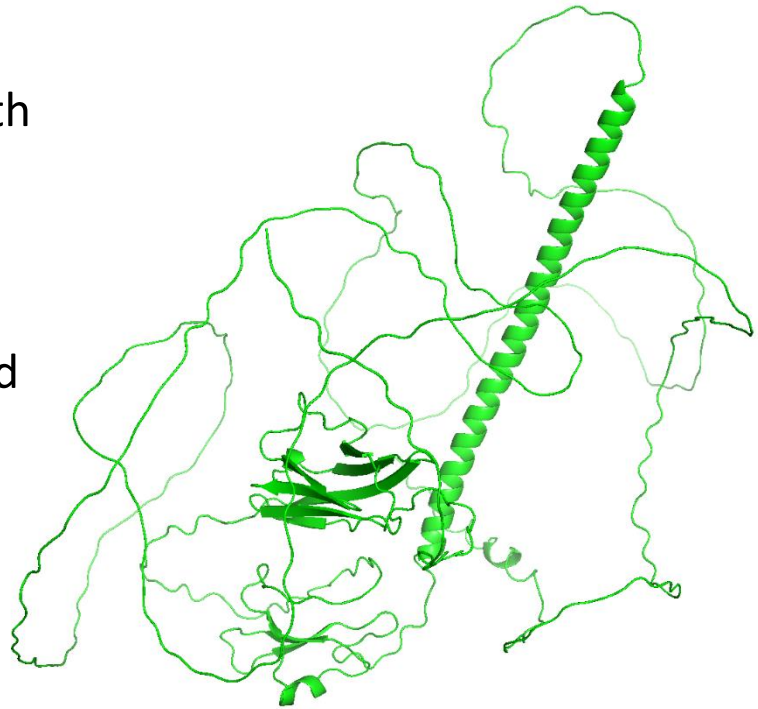
Aprahamian *et al* (2018), *Anal Chem* **90:** 7721; Biehn and Lindert (2021), *Nat Commun* **12:** 341

- Definition of protection factor (apparent oxidation rate normalized by the free amino acid reactivity with ·OH); correlation of protection factor with structural contacts and fractional SASA; demonstrated on proteins of known structure
- Observation that normalization by free amino acid reactivity is more accurate for the more reactive amino acids; development of denatured:native comparisons for determination of inherent reactivity rates; generation of empirical conversion factors to determine structure based on SASA; demonstrated ability to use SASA to select accurate homology models; examined MD simulations to include flexibility; demonstrated on proteins of known structure
- Found that most reactive amino acids (W,Y,F,H,L) gave most robust results; incorporated dynamics to improve predictions; used optimized conical neighbor count for cheap and robust correlations with HR-HRPF data; demonstrated on proteins of known structure

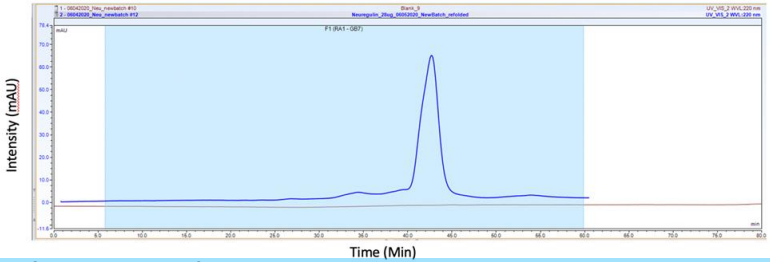
Immunoglobulin-like domain of human neuregulin 1

NRG1-Ig

- NRG1 is a signaling glycoprotein that interacts with tyrosine kinases
- Plays a key role in neuronal and cardiac development, regulation of synaptic plasticity
- Implicated in diseases including schizophrenia and some forms of cancer
- Many isoforms, including both soluble and membrane-bound
- Ig-like domain of NRG1 binds to heparan sulfate proteoglycans in the extracellular matrix
- ***Structure is currently unknown, and no near homologs of known structure***



AlphaFold model of full-length hNRG1
https://alphafold.ebi.ac.uk/entry/Q02297



Heterologous expression and purification of NRG-Ig

Sharp and Lindert Groups

HR-HRPF by multi-dose FPOP and LC-ETD MS/MS

Development and scoring of models by Rosetta with *hrf_dynamics* and mover dynamics

Prestegard Group

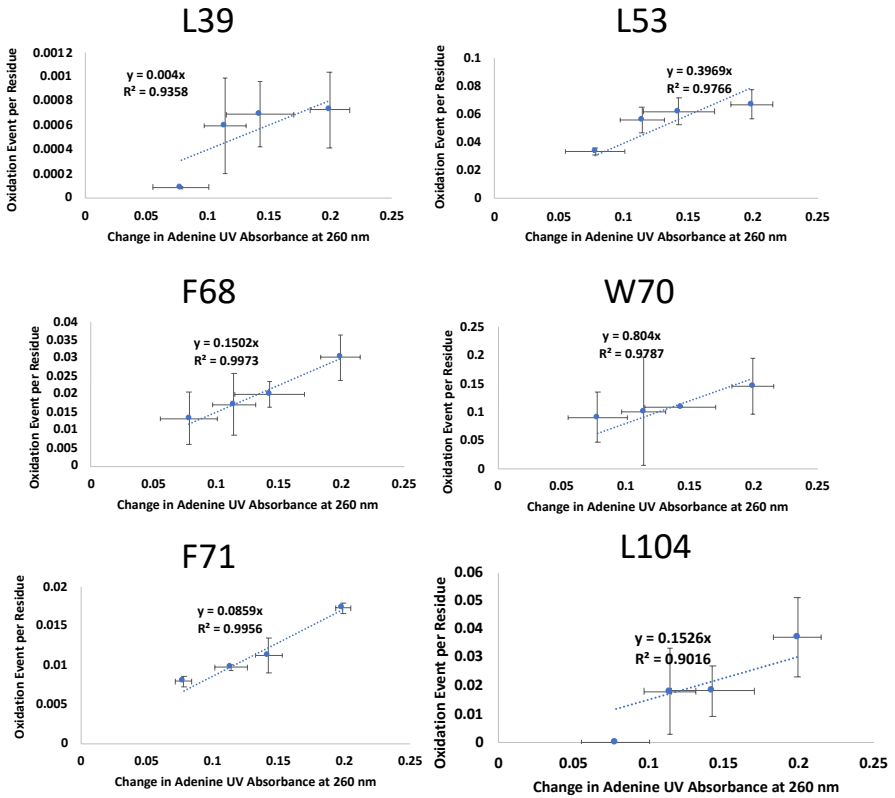
Multidimensional heteronuclear NMR

Resonance assignment and solution structure by NOE data

Comparison of independent models

Multi-Dose FPOP for HR-HRPF analysis

Measuring reactivity by FPOP



- HR-HRPF oxidation was measured at four different effective radical doses
- Reactivity measured by the slope of the regression of these data (b = 0)
- Reactivity measured for 20 amino acids in 118 amino acid construct
- R^2 of regression > 0.9 for all amino acids measured (mean R^2 = 0.971)
- Six amino acids used for *hrf_dynamics* scoring (W,Y,F,H,L)

Rosetta with *hrf_dynamics* and mover

Calculated natural log of the protection factor (lnPF) using established relative intrinsic reactivity values and $slope_N$ values from Sharp group

↓

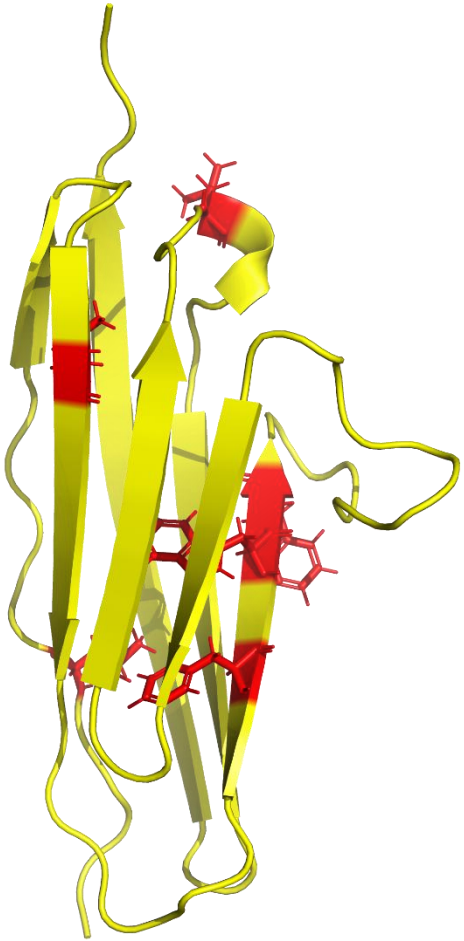
Generated 20,000 ab initio models + scored with Rosetta's score function named Ref15

↓

Scored models with our HRPF-guided score term, *hrf_dynamics*, then determined total score and identified top 20 scoring models

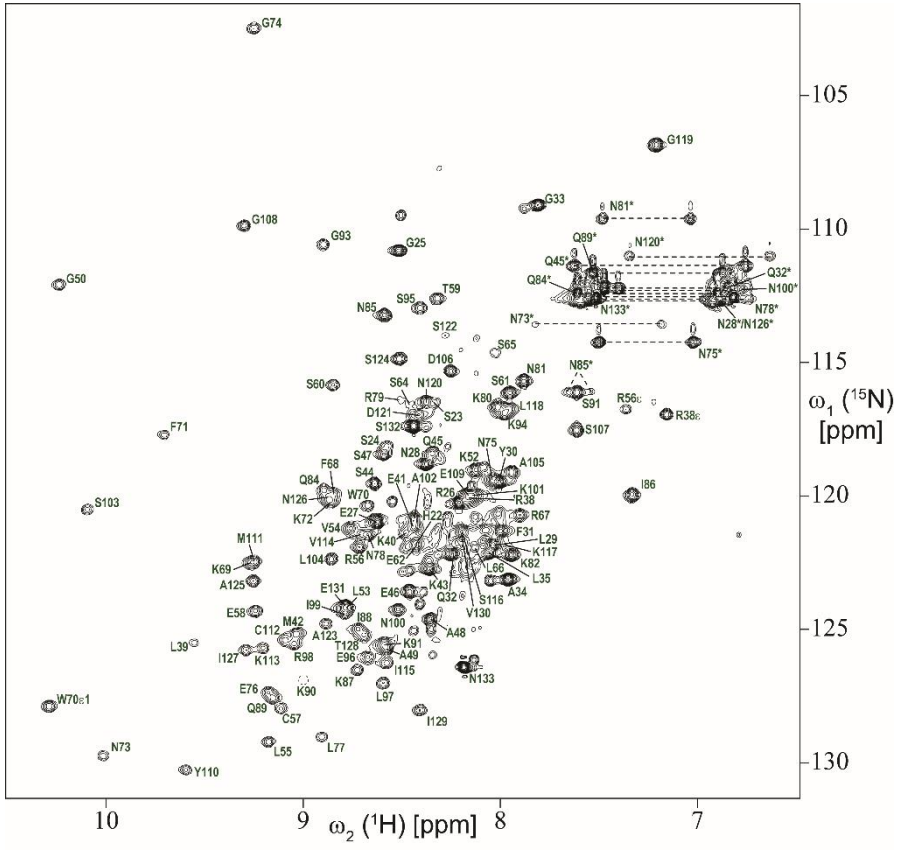
↓

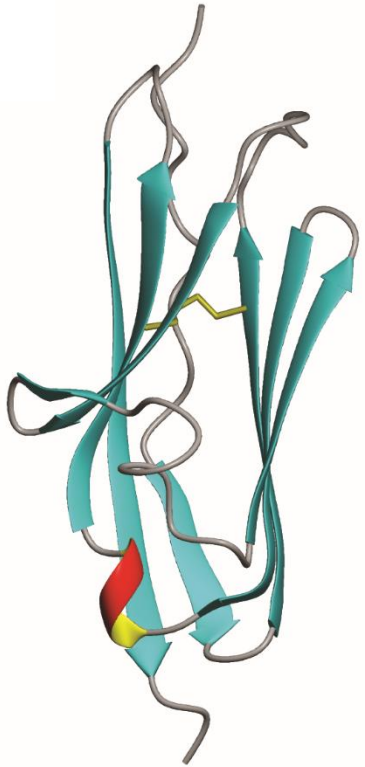
Generated 30 mover models for each of top 20 scoring models. Scored mover models with both Ref15 + *hrf_dynamics* and identified best scoring model from 20,600 models



Multidimensional NMR model

HSQC assignment





Comparison of HR-HRPF model and NMR structure

- After both groups independently determined their best model, groups shared information
- Models were judged by backbone RMSD from all amino acids, with the NMR model presumed to be accurate

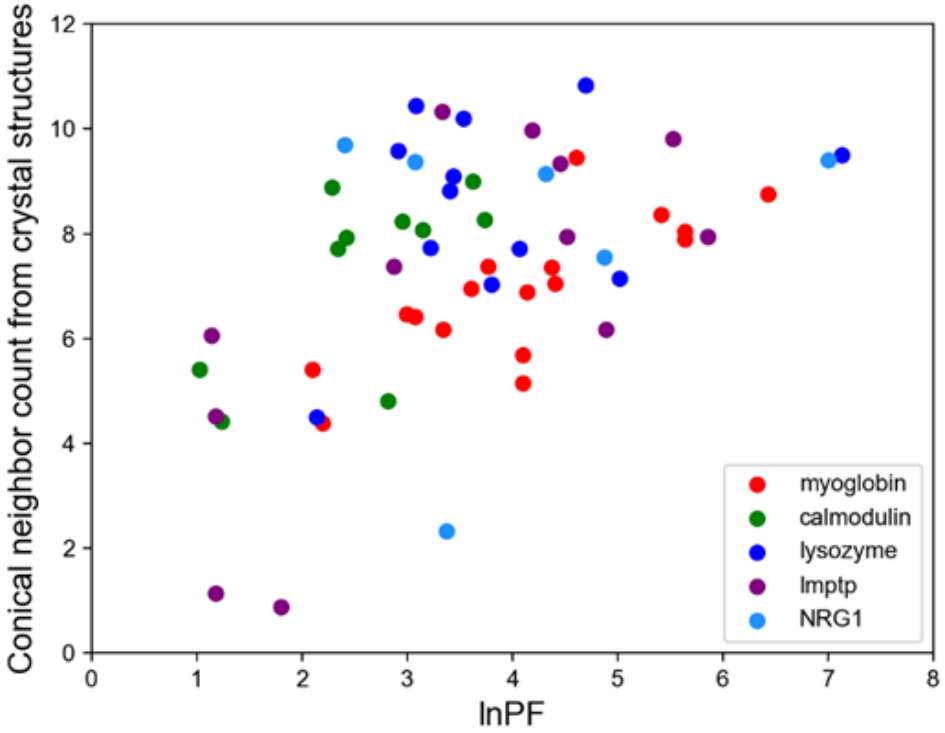
HR-HRPF

NMR

Backbone RMSD = 1.6 Å

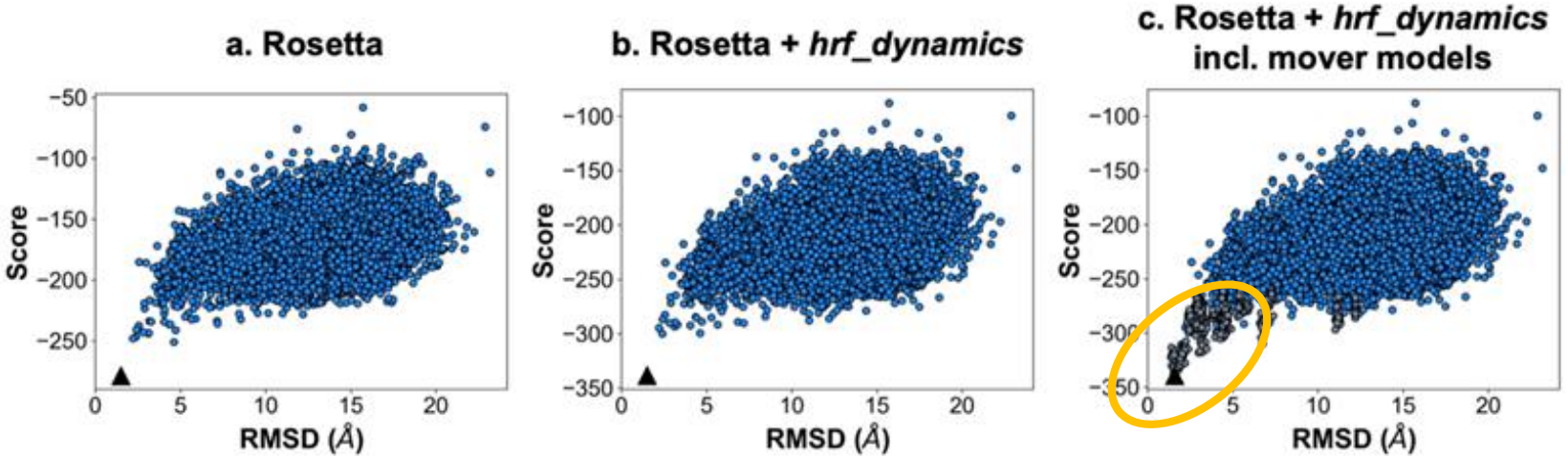
HR-HRPF correlation with conical neighbor count

Were data as reliable as those generated using proteins of known structure?



- Overlay of NRG1 data (cyan) with analysis of previously published data from Sharp (FPOP) and Kiselar/Chance (X-ray synchrotron) groups using proteins of known structure
- NRG1 data was comparably reliable to previously generated data
- No obvious evidence of bias in prior result reporting

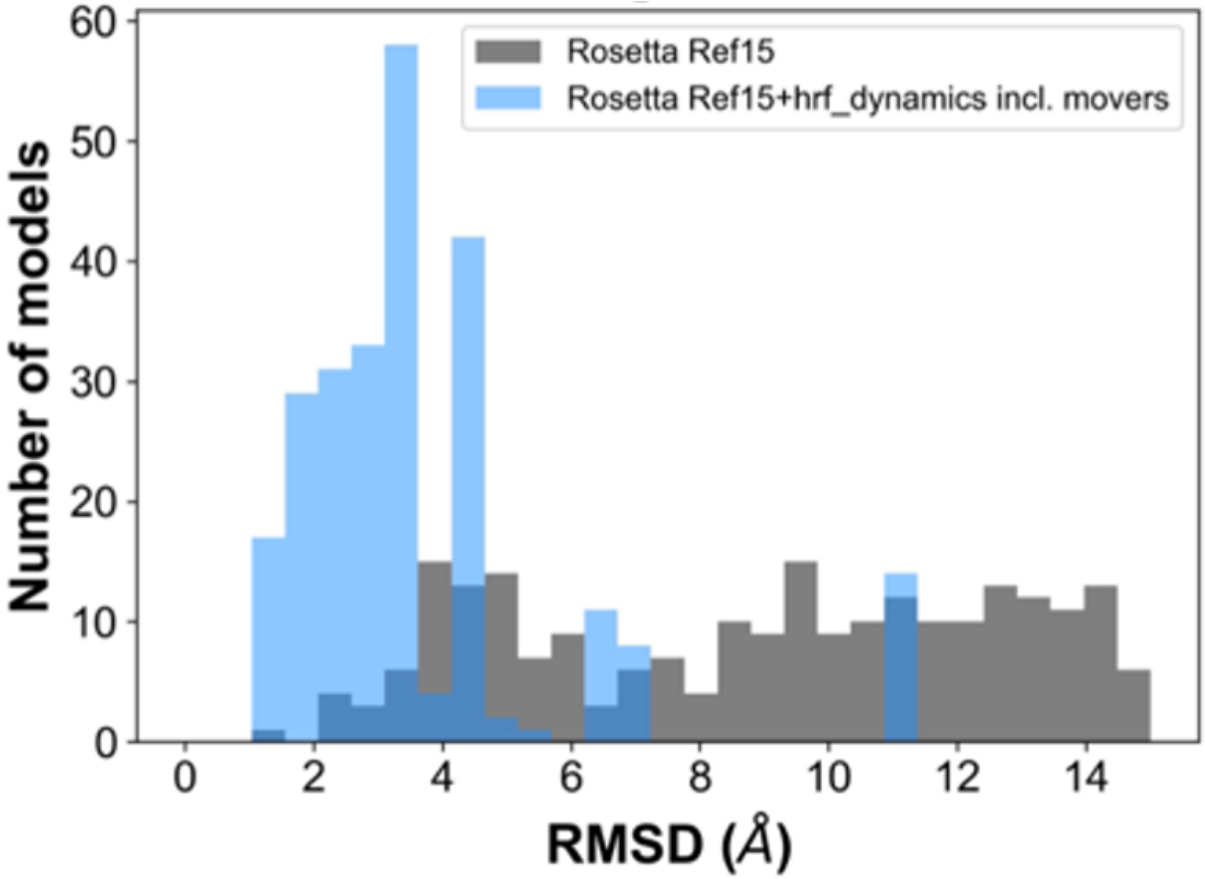
Improvements over Rosetta modeling alone



- Addition of *hrf_dynamics* term alone showed little improvement over Rosetta models alone
  - Score of best model improved, but accuracy of best model unchanged
- Inclusion of mover models scored with the *hrf_dynamics* term gave **great** improvement in robustness of modeling

RMSD distribution of top 250 scoring models

- Without *hrf_dynamics* and mover models, the top 250 scoring Rosetta models are evenly distributed between ~2 - 14 Å RMSD
- Addition of *hrf_dynamics* scoring term and mover models causes models to cluster around ~1.5 – 4.5 Å, with most below 3.5 Å RMSD



Conclusions

- First determination of unknown structure by covalent labeling mass spectrometry
- Determination of the structure by HR-HRPF and Rosetta modeling required much less sample, less time and no isotopic labeling
- Opportunities for use of HRPF-based modeling for larger multi-domain proteins of interest
- Work ongoing to investigate methods for including data from less reactive amino acids

Acknowledgements

Funding

NIGMS

R01 GM096049A
P20 GM130460
P41 GM111135

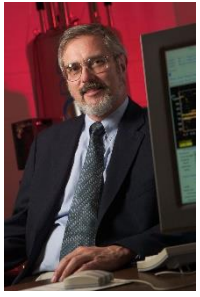
THE UNIVERSITY of MISSISSIPPI
SCHOOL OF PHARMACY

THE OHIO STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

The University of Georgia
Complex Carbohydrate Research Center

Dr. Niloofar Abolhasani Khaje

Prof. Steffen Lindert



Prof. James Prestegard

Dr. Sushil Mishra

Sarah Biehn

Dr. Alexander Eletsky