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


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

- 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)

General HRPF Workflow

Amino Acid-Level Changes
Structural Models

Peptide-Level Changes

LRQEDFPPR
Protein Structure Determination by HR-HRPF: Highlights

- 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


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

Premegard 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

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 \textit{hrf\_dynamics} and mover models, the top 250 scoring Rosetta models are evenly distributed between $\sim$2 - 14 Å RMSD.

- Addition of \textit{hrf\_dynamics} scoring term and mover models causes models to cluster around $\sim$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