



Protein Structure Determination by Combining Structural Mass Spectrometry Data with Rosetta

Xiaojing Zheng

Supervisor: Mark R. Chance

Case Center for Proteomics & Bioinformatics

Case Western Reserve School of Medicine

Background

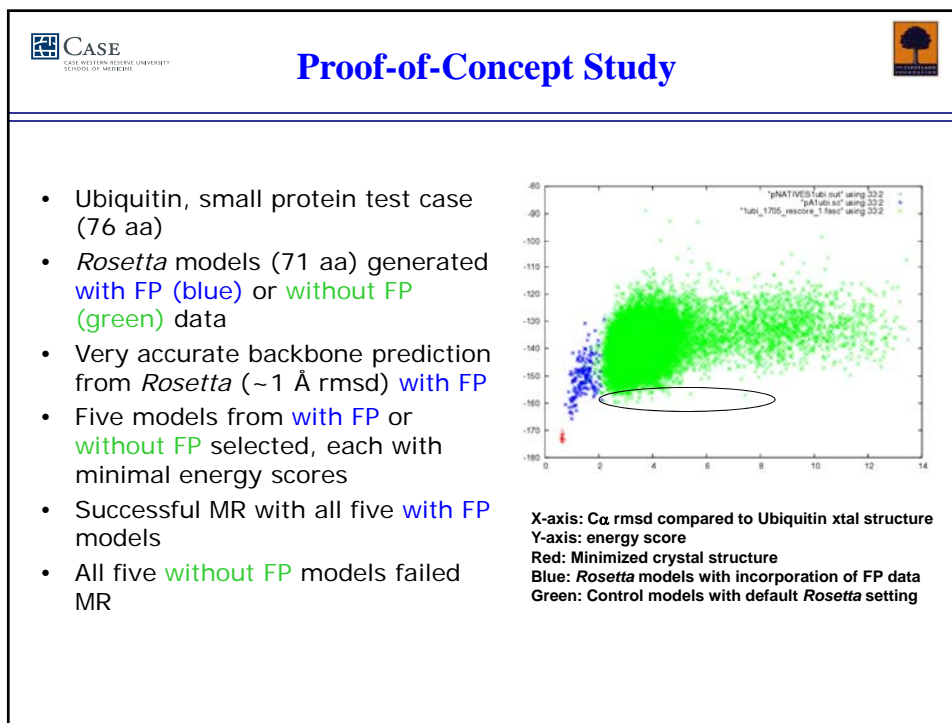
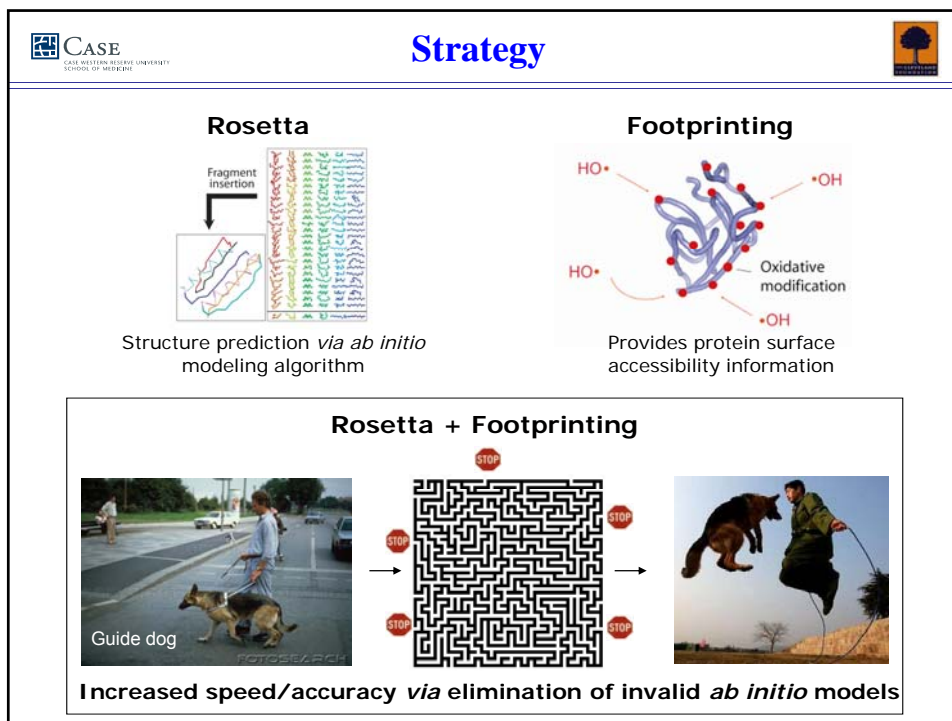


- Structural genomics problem
 - Total number of open reading frames (ORF) : 2-3 million
 - Total number of solved protein structures: 51,535 (~5000 nr)

million vs thousand !

?

Computer modeling { Homology modeling
ab initio



Rosetta/FP of PSI targets



PSI: protein structural initiative targets-typically single domain proteins ~200 aa

Type I: 15 + 9 targets

- ❖ Selected from the target pool with crystal structure and with coordinates deposited in PDB during July 2006 - June 2007.
- ❖ For calibration and testing purpose in the modeling/FP method development.

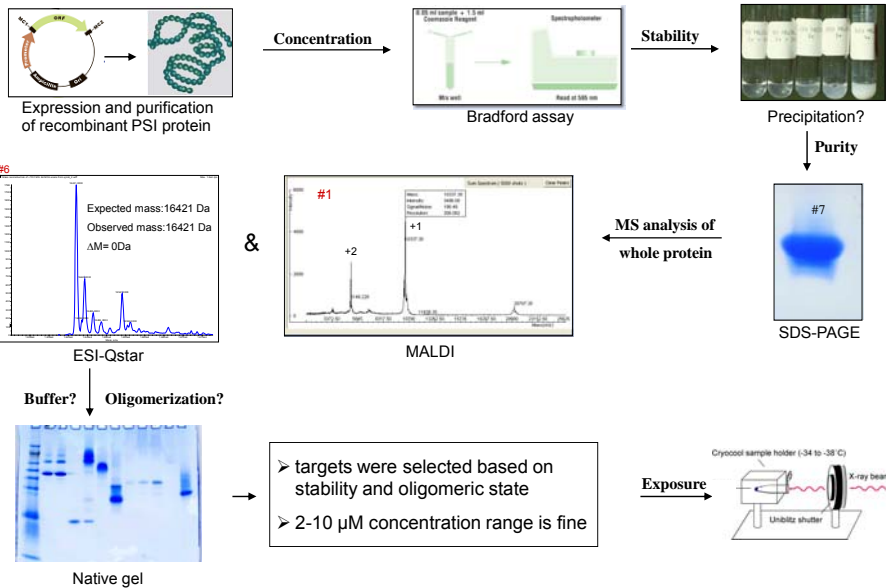
Type II: 1+7 targets

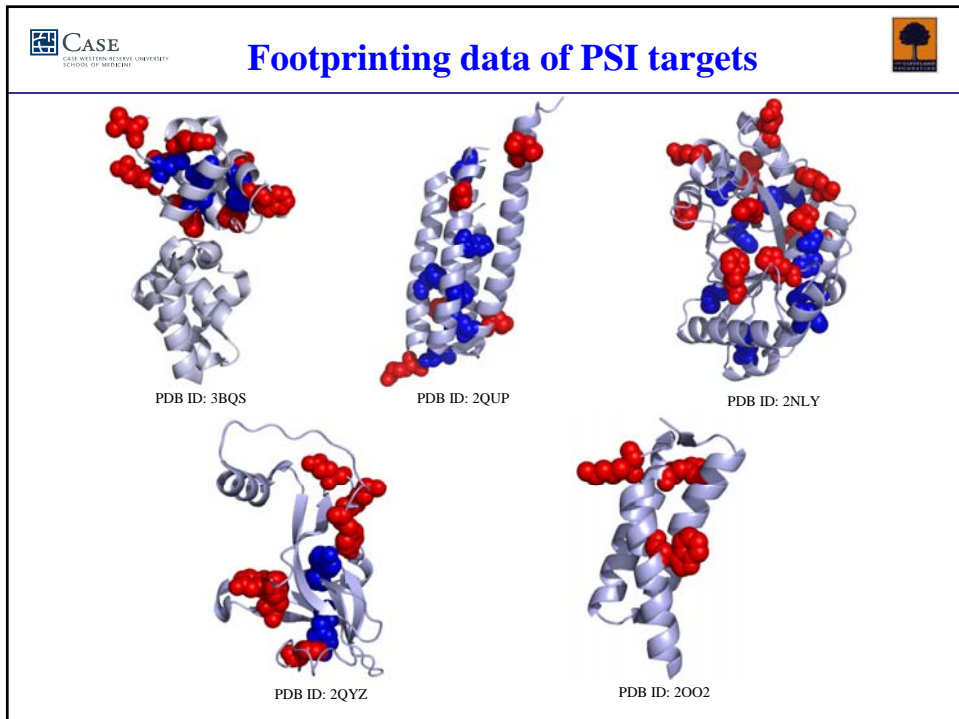
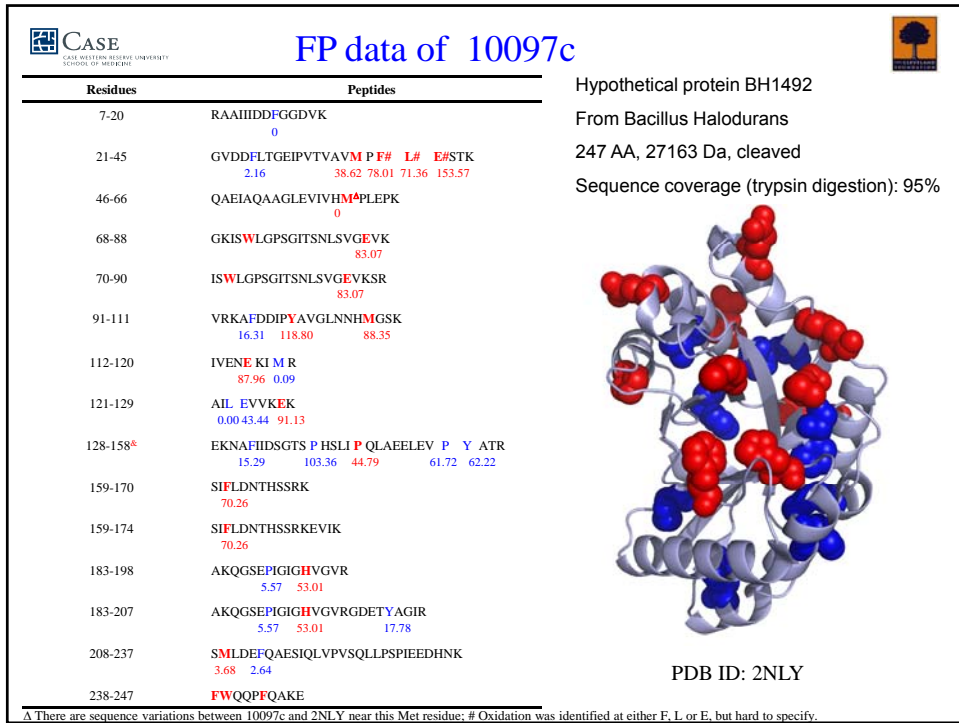
- ❖ Initially chosen from the target pool with diffraction data but not phased (no structure deposited in PDB).
- ❖ The model will be generated by *ab initio* modeling technique with incorporation of FP data and will be used as template for molecular replacement to solve the crystal structure.

Type III: not selected yet.

- ❖ Will be selected from target pool with no diffraction crystals to generate accurate *ab initio* model for the targets.
- ❖ Type III targets will be selected when the modeling/FP method is well developed and tested with Type I targets and proven to be successful with the Type II targets.

Biochemical/Biophysical Analysis for Optimal Target Selection







- Footprinting probes are well distributed along all PSI protein sequences. They provide information that is valuable to define exposed surface (modified residues) and buried hydrophobic core (unmodified reactive residues) of target proteins.
- For type I structure solved targets, their side chain solvent accessibility information provided by our footprinting methods was very consistent with the SASA values calculated on X-ray structures.
- In particular burial of reactive residue and accessibility of un-reactive residue are predictions that can be made most confidently (Met vs Leu).
- Open question as to what level of error in data is optimal for Rosetta predictions.

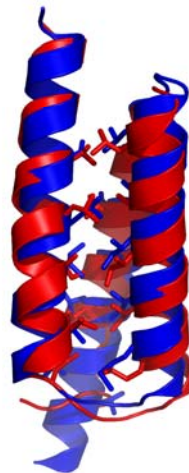


Example

- NYSGXRC target 10200c, hypothetical protein AF1782 from *Archaeoglobus fulgidus*
- *Rosetta* model (red)
- Molecular replacement using default *Rosetta* model was successful

Future

- Incorporation of footprinting data to improve the accuracy of MR search model
- Solve structure of type II targets
- High-throughput MR server (Phaser) is now available at UCSF to test 100-500 models for each study
- 10063a



Red: Rosetta model
Blue: PDB structure (2oo2)



Acknowledgements



Case Center for Proteomics & Bioinformatics

Mark R Chance

Keiji Takamoto

Janna Kiselar

Case Center for Synchrotron Biosciences

Sayan Gupta

Wuxian Shi

University of Washington

Robert M Vernon

Rhiju Das

David Baker

SGX Pharmaceuticals, Inc (Eli Lilly)

Michael Sauder

Stephen K Burley

Funding from NIGMS PSI Initiative