Cryptic Binding Sites

Thomas Choi
Ramya Rangan
Anni Zhang
Protein cryptic binding sites

**Active Site** - Region of protein where substrate binds, catalyzes a reaction.

**Allosteric Site** - Separate binding site that interacts with the active site.

**Cryptic Binding Site** - Allosteric binding site that is hidden in native structure. Often visible when a ligand is in the active site.

[Diagram showing a protein with and without a cryptic site ligand]

Protein without cryptic site ligand  
Protein with cryptic site ligand
Allosteric binding site
Cryptic binding sites

Example: p38 MAP kinase has a cryptic binding pocket where octylglucoside binds
Motivation for finding cryptic binding sites

Expand druggable protein proteome

Diagram:
- Human genome: ~30,000
- Druggable genome: ~3,000
- Drug targets: ~600–1,500
- Disease-modifying genes: ~3,000
Using Ligand-Mapping Simulations to Design a Ligand Selectively Targeting a Cryptic Surface Pocket of Polo-Like Kinase 1

Yaw Sing Tan, Paweł Słedź, Steffen Lang, Christopher J. Stubbs, David R. Spring, Chris Abell, and Robert B. Best
Experiment Goals

Study the cryptic binding site of Polo-like Kinase-1 (Plk-1)

Identify all cryptic binding site conformations (known or novel) using ligand-mapping techniques

Synthesize a ligand using Structure-Based Drug Design (SBDD) to selectively bind to Plk-1 cryptic binding site
Experiment Overview

Experiment 1: Molecular Dynamics simulation without ligand

Experiment 2: Modified ligand-docking Molecular Dynamics simulation with Benzene

Experiment 3: Experimental protein synthesis and x-ray crystallography
Experiment 1: MD Simulation without Ligand

Run 50ns Molecular Dynamics simulation with unliganded protein in explicit water

Protein conformation comparison between MD simulation and crystal structure using Root-mean-square deviation (RMSD)

Analyze the spontaneous opening/closing of cryptic binding site
Side by Side View

Closed Cryptic Site
Tyr 481 in “closed” position

Open Cryptic Site
Tyr 481 in “open” position

MD Discovered Conformations

Reaction Coordinate for “Closed” and “Open” Conformations

Free Energy (ΔG)

“Closed”

“Open”

6 kcal/mol

4 kcal/mol

Reaction Coordinate
Highlighting problems with Unbiased MD Simulations

Rarely run simulation on protein conformation without a ligand sample

“Solvent mapping” requires binding site to be accessible
Experiment 2: Modified Ligand-Mapping MD Simulation

Incorporate low concentration of benzene into the simulation

Binding site affinity for phenyl moiety confirmed by crystal structure evidence

High concentration causes phase separation of water and benzene

10 independent 5 ns ligand-mapping MD simulations

Different initial distributions of benzene molecules
Modified-Ligand Mapping Results

Figure 2. Conformations of the hydrophobic binding pocket in Plk1 PBD. Percentage populations of the conformations observed in the single long, multiple short and ligand-mapping simulations are indicated in black, red and blue, respectively. a–c) Closed conformations observed in crystal structures. d) Closed conformation observed only in the single long MD simulation. e,f) Open conformations observed in crystal structures. g,h) New conformations observed in ligand-mapping MD simulations.
Benzene Stabilization

Experiment 3: Experimental Protein Synthesis

To experimentally confirm modified ligand-mapping simulation results

Designed ligand presents a phenyl ring to the pocket in the “half-open” conformation

Determine the $K_D$ and crystal structure of binding site
Experimental Protein Synthesis Results

Known Ligand vs. Synthesized Ligand

Major Takeaways

Success of Structure-Based Drug Design for targeting binding sites

New method of Molecular Dynamics probing expands accessible search space

Provides mechanism of studying different types of hydrophobic pockets
Limitations and Areas for Future Study

Needed to know substrate that typically binds to the pocket (known affinity)

Do not identify how the native ligand causes the pocket to open

Only used benzene (can use other hydrophobic molecules)


Discovery of multiple hidden allosteric sites by combining Markov state models and experiments

Gregory R. Bowman, Eric R. Bolin, Kathryn M. Hart, Brendan C. Maguire, and Susan Marquee; PNAS 2015
Overview

- Predict and verify cryptic binding sites
- Avoid first finding an cryptic pocket binding ligand
- System: TEM-1 β-lactamase
Markov models

Use Markov state model (MSM) to represent the conformations a protein takes
MSMBuilder

**Overview**

1. Run simulations (many in parallel)
2. Pick a feature to cluster on
3. Cluster the frames
4. Build out MSM

**The details**

1. 1,000 MD simulations with Folding@Home
2. RMSD between backbone atoms
3. K-centers until frames 1.2 Å away
MSMBuilder

Bowman et al. Cell Research 2010
Locating cryptic binding sites

In each cluster, check for cryptic binding sites. Look for regions that:

- Look like a pocket
- Correlated with the active site

Key assumption: pocket can be seen without the active site / cryptic site ligand

Bowman and Geisler, *PNAS*, 2012
Validating a cryptic binding site

Thiol binding:

Accessible

\[ \text{R}-\text{S} + \text{DTNB} \rightarrow \text{Fluorescence Readout} \]

1. Make a mutant protein that includes a cysteine (adds in an R-SH)
2. Add DNTB, check overall labeling rate - gives \( k_{\text{op}} \) when \( k_{\text{int}} \) is high

\[
\text{Closed} \xrightleftharpoons[k_{\text{op}}]{k_{\text{cl}}} \text{Open} \xrightarrow[k_{\text{ini}}]{\text{Labeled}}
\]
Thiol labeling on known binding site

- Thiol labeling works on known cryptic binding site!
- Means that drug-like molecule can fit
- Same thing on control residues shows no labeling
Checking for potential unfolding

A wrinkle: What if the cysteine substitution caused the protein to unfold?
Checking for potential unfolding

A wrinkle: What if the cysteine substitution caused the protein to unfold?

Labeling rate far too high to be explained by complete unfolding
Active site activity

Activities (nmol product / min) of labeled and unlabeled proteins

Check for allostery: Does binding in the cryptic pocket affect the active site?
Novel cryptic binding sites

- Predicted new cryptic binding sites
- Picked out accessible residues for Cys mutation
- Thiol labeling → again, positions could be labeled
- Again, interaction with active site
Takeaways / Next Steps

Takeaways:
- Tested an algorithm that pinpoints cryptic binding pockets
- Validated the solvent accessibility and interaction with active site

Next steps:
- How do binding pocket opening rates compare to simulation?
- Designing small molecules that fit the allosteric sites.
Strengths

- Pursued experimental validation for a challenging phenomenon
- Method requires no prior knowledge of allosteric ligands
- Made use of more simulation time via Markov models
Limitations

- Computationally intensive: 1,000 MD simulations totaling 81μs.

- Cys mutation effect? Decreased activity.

- More thorough checks of pocket formation could be useful.

- Requires knowledge of active site.

- Assumption that allosteric site visible in native ensemble.

- How many cryptic binding sites were predicted? How many validated?

Bowman and Geisler, *PNAS*, 2012
CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites

Peter Cimermancic, Patrick Weinkam, T. Justin Rettenmaier, Leon Bichmann, Daniel A. Keedy, Rahel A. Woldeyes, Dina Schneidman-Duhovny, Omar N. Demerdash, Julie C. Mitchell, James A. Wells, James S. Fraser and Andrej Sali
2016 J Mol Bio
Cryptic Binding Pocket Detection Overview

Are there current “undruggable” Proteins that actually have targetable cryptic sites?

Liu. et al. Applying Side-chain Flexibility in Motifs for Protein Docking.
Creating feature sets for proteins

Characterize sites based on

**Sequence** - protein sequence evolutionary conservation

**Structure** - protrusion, hydrophobicity, convexity

**Dynamics of individual residues and their neighbors** (from MD simulations): flexibility of residues

Feature Vector Set for the Machine Learning model

Comparative Analysis  ML Predictive Model Creation  Model Validation
Example Characteristics of Cryptic Sites

- Cryptic sites form from minor structural changes.
- Sequences are as evolutionarily conserved as binding pockets.
- Predominantly localizes at concave protein regions, but less concave than a binding pocket.
- Less hydrophobic than binding pockets.
- More flexible than a binding pocket.

All atom RMSD between apo and holo sites
Machine Learning Model Creation

Testing different ML algorithms, data pre-processing method and parameters
Supervised ML - Support Vector Machine

Black - cryptic sites residue

White - non cryptic sites residue

Finding hyperplane (Red Line) to classify cryptic sites that provides the farthest gap as possible

ML Model Validation

Greedy forward selection of characteristics

Comparison of CryptoSite to other binding pocket predictors

(a) ROC curves for different models:
- CryptoSite (AUC=0.83)
- FTFlex (AUC=0.77)
- Fpocket (AUC=0.63)
- Random (AUC=0.50)

(b) AUC plot for different characteristics: CNC_mean, CNC5_mean, PTM, Hn, HnN, D2S, DASn, NBG, AT4M, Res, CMC, ATMB, LCB, LS, Sn, ChRn
CryptoSite Output

Output of the algorithm produces a cryptic site score for each Residue.

Threshold for cryptic site residues is defined by user.
False positives and False Negatives

False Negative sites:

- Sites that had large conformational changes
- Pockets that were difficult to sample from MD
- Partial Sites that need another protein binding

False Positive sites:

- High scoring isolated residues
- Terminal regions of truncated proteins
- **Predicted cryptic sites that are true cryptic sites not annotated in the database**
TEM1 β-lactamase Experimental Validation

Used NMR to validate predicted cryptic site
Expanding the Druggable Proteome

Predicted cryptic sites yet to be discovered

(a)

- Cryptic sites only: 36.5%
- Pockets only: 6.4%
- Pockets & cryptic sites: 37.7%
- Undruggable: 19.4%

Human PDB (N = 4,421)

- Cryptic sites only: 38.0%
- Pockets only: 5.2%
- Disease-associated human PDB (N = 1,420)

- 34.5%
- 22.3%
Strengths

Open source and freely available on the internet

Decreased the computational time to find cryptic sites with use of simplified energy landscape MD

Found previously unknown/unannotated cryptic sites

Has experimental validation (NMR) for proposed cryptic site
Limitations

CryptoSite is only compared to methods that are optimized to find binding pockets, not cryptic sites

Best feature still uses MD simulations which are computationally expensive, to define dynamic features

Using greedy forward approach allows for local optimum but may not reach global optimum feature and may not choose features representatives of cryptic sites

Unable to find “outlier” cryptic sites that don’t conform to the proposed features

Druggable Proteome expansion calculations done without MD simulations, which represent some of the “best features”
https://modbase.compbio.ucsf.edu/cryptosite/