



University of Illinois  
at Chicago

## A Novel System for Discovery of Gene Function by Protein Surface Matching

*Spatial surface patterns of molecules can be used to quickly identify and classify potential drug targets.*

### Technology Reference CV40

#### Contact

Lanny Feder  
312-413-5496 voice  
312-413-5846 fax  
[Lfeder@uic.edu](mailto:Lfeder@uic.edu)

#### Inventors

Dr. Jie Liang  
T. Andrew Binkowski  
Dr. Larisa Adamian

#### Field

Bioinformatics  
Proteomics  
Drug Discovery

#### Key Words

Structural genomics  
Functional proteomics  
Molecular modeling

#### Stage of Development

Implemented in software  
package

#### Patent Status

Utility patent filed

#### Status

Seeking Licensees  
Seeking Licensing  
Partner

### Applications

This technology provides a molecular classification system useful to generate comparisons between molecules and allows determination of similarities and differences for predicting functional properties of biomolecules.

### Previous Knowledge

Proteins function by interacting with other molecules such as substrates and drugs, but the functional surfaces often involve only a small number of key residues. These residues are dispersed in diverse regions of the primary sequences and are difficult to detect by sequence alignment or full backbone structural based alignment methods. Identifying spatial motifs of proteins that are functionally relevant is therefore an essential task of bioinformatics, and will have wide applications in identifying novel drug targets through structural genomics, and facilitate drug discovery.

### New Discovery

This invention provides sensitive and powerful methods for detecting similarity patterns of surface motifs of molecular sequences previously unknown. Since protein functional surfaces are frequently associated with surface regions of prominent concavity, the focus on surface pockets and voids of a protein structure can provide important information about function. Using alpha shape as implemented in CASTp, 910,379 surface pockets and interior voids have been identified and characterized exhaustively on 12,177 protein structures.

After concatenating residues located in the same surface pocket from primary sequences, it is found that very short concatenated sequences of key residues derived automatically can be discriminating in identifying proteins with related functions. The technology uses both sequence and structural based methods to discover similar surfaces. Protein surface motifs can be examined by comparing a query protein against a database of surface motifs.

The methodology includes rigorous statistics to ensure biological significance and guarantee high value results. The technology is high throughput, allowing for rapid proteome-wide analysis and is scalable to match the increasing growth rate of solved protein structures.



University of Illinois  
at Chicago

## A Novel System for Discovery of Gene Function by Protein Surface Matching

*Spatial surface patterns of molecules can be used to quickly identify and classify potential drug targets.*

### Advantages and Benefits

The technology has many advantages over current technology:

- The surface identification is based on theoretical and algorithmic results of computational geometry from the CASTp database, resulting in an exact topological description. The advantages of which are pockets and cavities being identified analytically, the boundary between the bulk solvent and the pocket is defined precisely, all calculated parameters are rotationally invariant, and do not involve discretization.
- Complete surface motif patterns can be analyzed. Current existing technology is limited in the size of motif patterns that are analyzed. There is no limit to the motif size used in this method which can be expanded to include multiple surface motifs and whole protein surface analysis.
- The method ensures accurate, reproducible results over current manually curated analysis.

### Application Areas

The technology has many uses, some of which are, but not limited to:

- Gene discovery.
- Identification of novel drug targets
- Identification of alternative targets for existing drugs.
- Identifying proteins responsible for drug side effects.
- Design of screens for drug activity, for toxicity, and for other side effects.

### Status

The technology has been fully implemented in a software package for high throughput research. The software was licensed to a large pharmaceutical company for 3 months to validate large scale drug discovery research. The license has expired.

### References

Binkowski TA, Adamian L, Liang J. Inferring Protein Function from Matching Spatial Surface Patterns. Manuscript. 2002.  
Liang J, Binkowski TA. CASTp: Computed Atlas of Surface Topologies of Proteins. Web Server, <http://cast.engr.uic.edu>. 2001.  
Jie Liang, Herbert Edelsbrunner, and Clare Woodward. 1998. *Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design*. Protein Science, 7:1884-1897.