Representation and Matching of Molecular Surfaces

Thomas Funkhouser

Princeton University

CS597A, Fall 2007

Outline

Molecular surfaces
- Definitions
- Representations

Surface matching
- Continuous surface mappings
- Discrete point correspondences

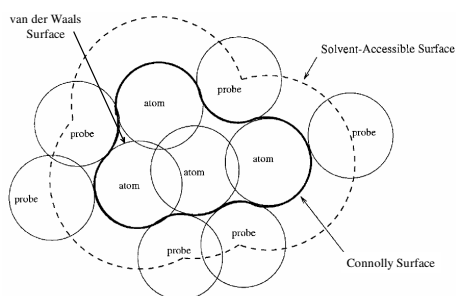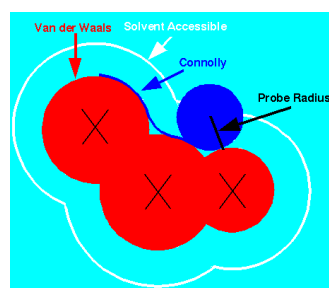Surface retrieval
- Shape descriptors
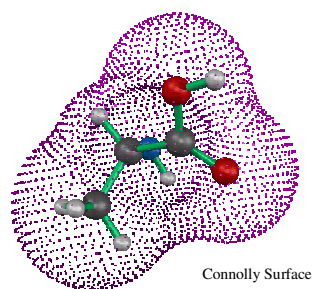
Results

Discussion

Molecular Surfaces

van der Waals Surface

Solvent-Accessible Surface

probe

atom

Connolly Surface

[Cai98]

Molecular Surfaces

Van der Waals

Solvent Accessible

Connolly

Probe Radius

http://www.simbiosys.ca/sprout/eccc/cangaroo.html

Molecular Surfaces

Connolly Surface

http://www.netsci.org/

Molecular Surfaces

van der Waals   Solvent accessible   Connolly

http://www.simbiosys.ca/sprout/eccc/cangaroo.html
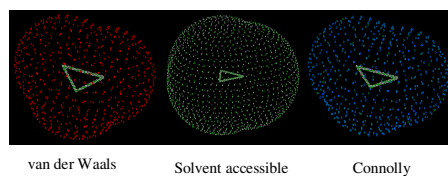
Molecular Surfaces
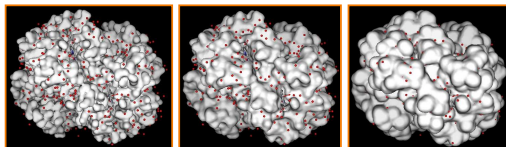
van der Waals   Connolly   Solvent accessible

Molecular Surfaces

van der Waals   Connolly
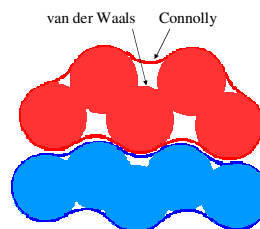
Connolly surfaces of bound molecules are complementary

Outline

Molecular surfaces
- Definitions
- ØRepresentations

Surface matching
- Continuous surface mappings
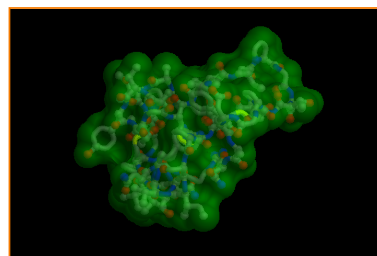- Discrete point correspondences

Surface retrieval
- Shape descriptors
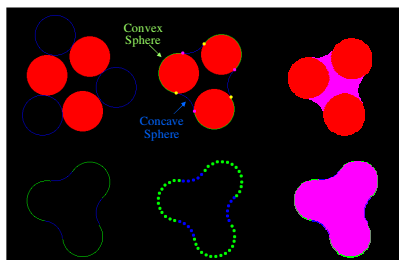
Results

Discussion

Molecular Surface Representation

How should we store the surface description in a computer?

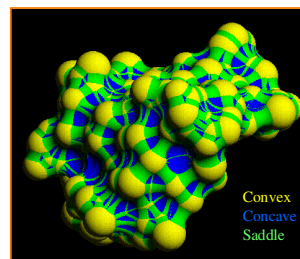## Molecular Surface Representations

Union of partial spheres and tori

Convex Sphere

Concave Sphere
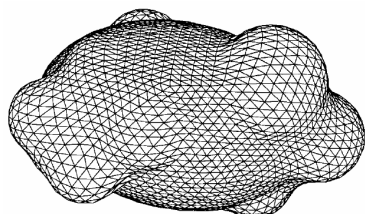
Molecular Surface Representations

Union of partial spheres and tori

Convex
Concave
Saddle

Molecular Surface Representations

Triangle mesh
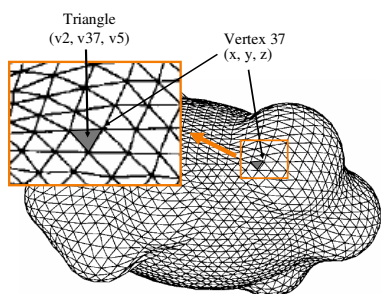

[Cai98]

Molecular Surface Representations
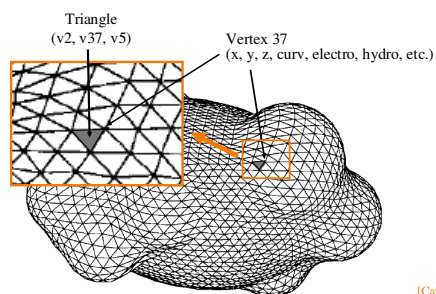
Triangle mesh

Vertex
(x, y, z)

[Cai98]

Molecular Surface Representations

Triangle mesh

Triangle
(v2, v37, v5)

Vertex 37
(x, y, z)

[Cai98]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

Triangle
(v2, v37, v5)

Vertex 37
(x, y, z, curv, electro, hydro, etc.)

[Cai98]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

Maximum Curvature   Minimum Curvature

[Duncan93]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

Maximum Curvature   Minimum Curvature

[Kinoshita03]
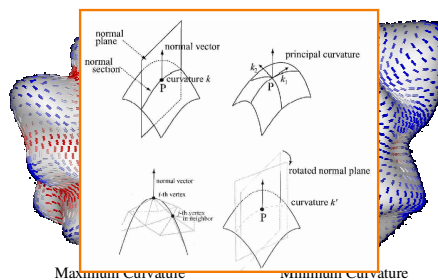
3

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

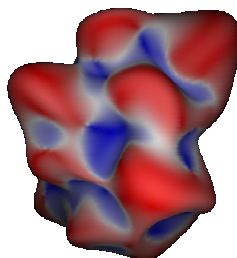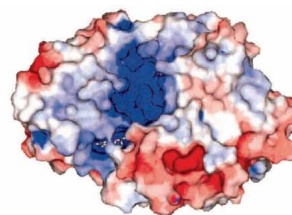Mean Curvature

[Duncan93]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

Electrostatic Potential
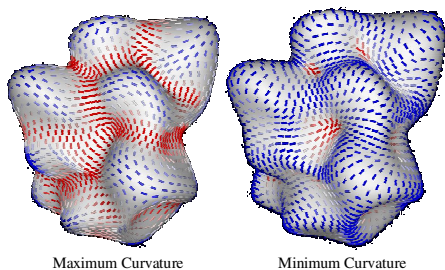
[Kinoshita03]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

Hydrophobicity

N
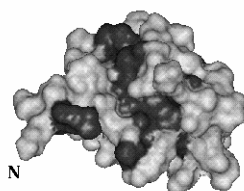
Exposure of hydrophobic surface before binding of calcium to calmodulin.

[CaBP Data Library]

Molecular Surface Representations

Triangle mesh (with properties at every vertex)

(x, y, z, curv, electro, hydro, etc.)
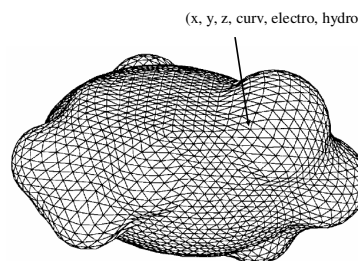
[Cai98]

Outline

Molecular surfaces
- Definitions
- Representations

Surface matching ←
- Continuous surface mappings
- Discrete point correspondences

Surface retrieval
- Shape descriptors

Results

Discussion

Surface Matching

Goal: compute the similarity between two molecular surfaces with attributes at vertices

(x, y, z, c, e, h, etc.)

A

$Similarity(A, B)$

(x, y, z, c, e, h, etc.)

B

[Cai98]

Surface Matching

Challenge: find a consistent parameterization

$P_A(u,v) = (x, y, z, c, e, h, \text{etc.})$

A

v

u

$P_B(u,v) = (x, y, z, c, e, h, \text{etc.})$

B

$$Similarity(A,B) = \int_{u,v} (P_A(u,v) - P_B(u,v))^2 \, du dv$$

[Cai98]

Continuous Surface Mappings

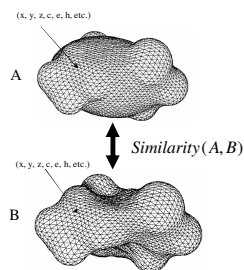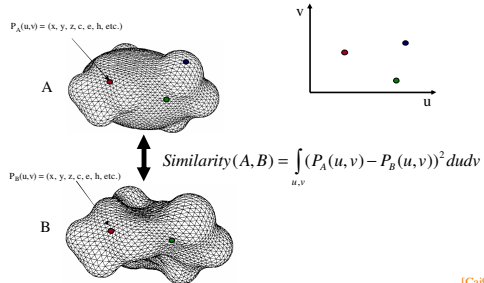Map one surface onto the other, and measure the cost of the mapping
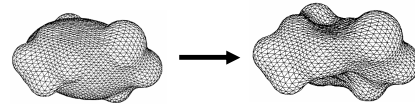
[Cai98]

Continuous Surface Mappings

Earth mover's distance
- Amount of "work" required to deform A onto B
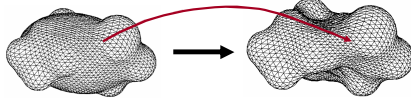
[Rubner00] [Cai98]

Continuous Surface Mappings

Map all surfaces to planar parameterization

$\mathcal{U}_{(x,y,z)}$

$\mathcal{X}_{(u,v)}$

u

$\mathbb{R}^3$

S

$\mathbb{R}^2$

Ω v
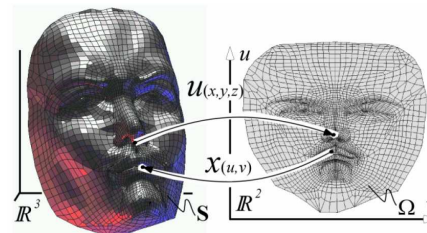
[Sheffer]

Continuous Surface Mappings

Map all surfaces to planar parameterization
- Difficult without adding lots of distortion

[Sorkine02]

Continuous Surface Mappings

Map all surfaces to cubic parameterization

cube parametrization   geometry image (6x105x105)   flat-shaded remesh (64,523 non-hole vertices)

[Praun03]

Continuous Surface Mappings

Map all surfaces to spherical parameterization
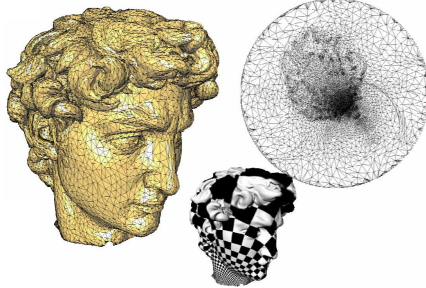
[Sheffer]

Continuous Surface Mappings

Harmonic map to spherical parameterization
- Minimizes

$$\frac{d^2 P_i}{dt^2} = -\gamma \frac{dP}{dt} + \vec{F}_{ext} + \vec{F}_{int}$$
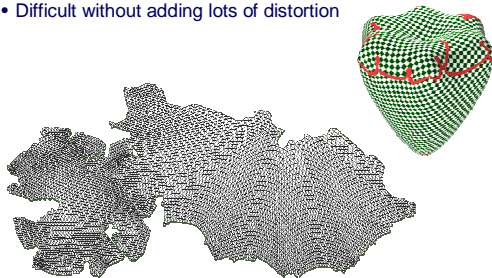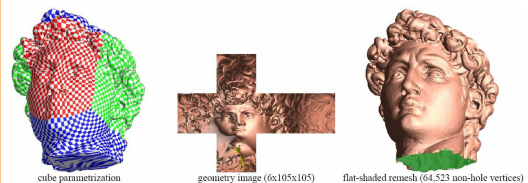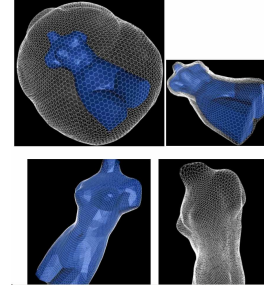

[Shum][Kalnins00]

Continuous Surface Mappings

Conformal map to spherical parameterization
- Minimizes distortion to angles
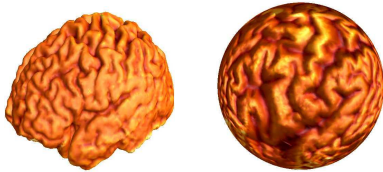
http://www.cise.ufl.edu/~gu/yau_talk/

Continuous Surface Mappings

Consistent spherical parameterization
- Guarantees feature point correspondences

Step 1
Step 2   Steps 3, 4   Step 5   Step 6

[Asirvatham05]

Continuous Surface Mappings

Consistent spherical parameterization
- Guarantees feature point correspondences

Average
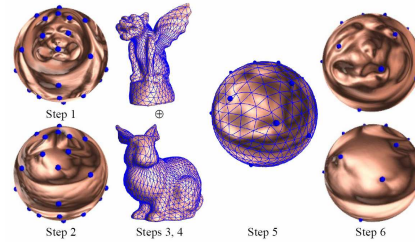
Principal Components

[Asirvatham05]

Continuous Surface Mappings

Partial matches don't fit well into this framework

A

$$Similarity(A, B) = \int_{u,v} (P_A(u,v) - P_B(u,v))^2 \, du \, dv$$

B

[Kinoshita03]

6

Piecewise Continuous Mappings

Use continuous surface matches for local patches

A

Patch

B

$$Similarity(A,B) = \sum_{Patches}\left(\int_{u,v}(P_A(u,v) - P_B(u,v))^2 dudv\right)$$
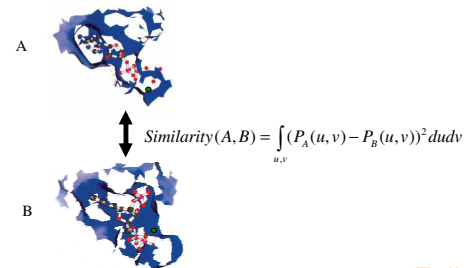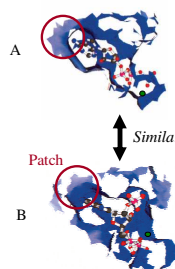
[Kinoshita03]

Piecewise Continuous Mappings
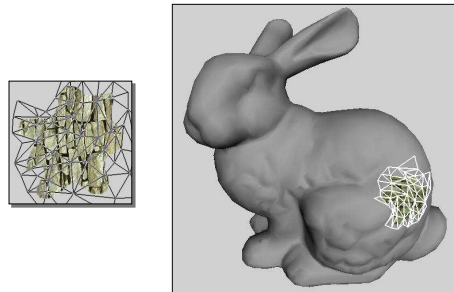
[Praun00]

Piecewise Continuous Mappings
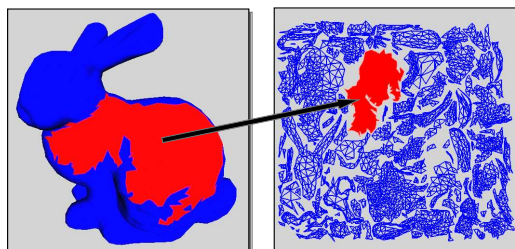
[Praun00]

Piecewise Continuous Mappings

Need to balance coverage vs. quality of match

coverage

z- score

50%-1P
70%-1P
90%-1P
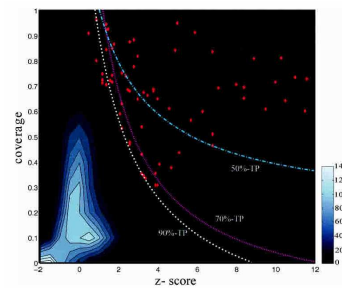
[Kinoshita05]

Outline
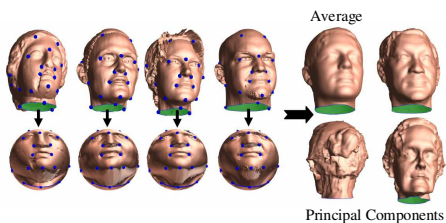
Molecular surfaces
- Definitions
- Representations

Surface matching
- Continuous surface mappings
- Ø Discrete point correspondences

Surface retrieval
- Shape descriptors

Results

Discussion

Surface Matching at Discrete Points

Sample the surfaces at discrete sets of points
- ØVertices
- Critical points
- Features
- etc.
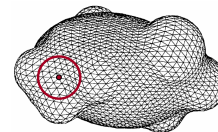
Match surface patches near sampled points
- Association graphs
- Geometric hashing
- Iterative closest points

Surface Matching at Discrete Points

Sample the surfaces at discrete sets of points
- Vertices

Ø Critical points
- Features
- etc.

Critical points

Match surface patches near sampled points
- Association graphs
- Geometric hashing
- Iterative closest points

Surface Matching at Discrete Points

Sample the surfaces at discrete sets of points
- Vertices

Ø Critical points
- Features
- etc.

Critical points
(without saddles)

Match surface patches near sampled points
- Association graphs
- Geometric hashing
- Iterative closest points

Surface Matching at Discrete Points

Sample the surfaces at discrete sets of points
- Vertices
- Critical points
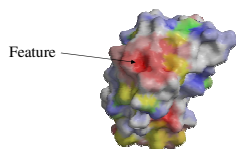
Ø Features
- etc.

Feature

Match surface patches near sampled points
- Association graphs
- Geometric hashing
- Iterative closest points

Surface Matching at Discrete Points

Differences from matching atoms/pseudo-centers?
- Point samples represent continuous set of points
- Parameterization
- Derivatives

Continuous surface

Surface Matching at Discrete Points

Differences from matching atoms/pseudo-centers?
- Point samples represent continuous set of points

Ø Parameterization
- Derivatives

Surface Matching at Discrete Points

How is it different than matching a set of points?
- Point samples represent continuous set of points
- Parameterization

Ø Derivatives

Normal Vector

Tangent Plane

Surface Matching at Discrete Points

Associate pairs of points only if they have the same:

- Intra-molecular distances
- Electrostatic potentials
- Lipophilic potentials
- Principal curvatures
- Harmonic shape images (geometric representation of local neighborhood shape)
- Relative orientations of harmonic shape images

Association Graph

[Hofbauer04]

Surface Matching at Discrete Points

Associate pairs of points only if they have the same:

- Intra-molecular distances
- Electrostatic potentials
- Lipophilic potentials
- Ø Principal curvatures
- Harmonic shape images (geometric representation of local neighborhood shape)
- Relative orientations of harmonic shape images

Principal Curvatures

k_2 k_1

P
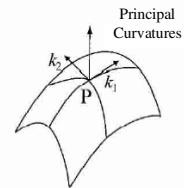
[Kinoshita03]

Surface Matching at Discrete Points

Associate pairs of points only if they have the same:

- Intra-molecular distances
- Electrostatic potentials
- Lipophilic potentials
- Principal curvatures
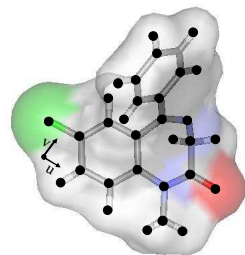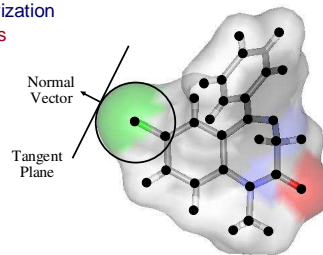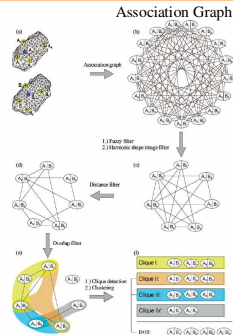- Ø Harmonic shape images (geometric representation of local neighborhood shape)
- Relative orientations of harmonic shape images

[Hofbauer04] [Shulman-Peleg04]

Surface Matching at Discrete Points

Associate pairs of points only if they have the same:

- Intra-molecular distances
- Electrostatic potentials
- Lipophilic potentials
- Principal curvatures
- Harmonic shape images (geometric representation of local neighborhood shape)
- Ø Relative orientations of harmonic shape images

[Hofbauer04] [Shulman-Peleg04]

Surface Matching with ICP

Iterate until convergence:

1. Select source points (from one or both surfaces)
2. Correspond to closest points on other surface
3. Weight the correspondences
4. Reject outlier point pairs
5. Compute an error metric for the current transform
6. Minimize the error metric w.r.t. transformation
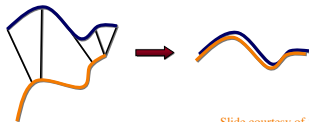
Slide courtesy of Szymon Rusinkiewicz

Surface Matching with ICP

Using point-to-plane distance for surfaces instead of point-to-point distance

- Lets flat regions slide along each other
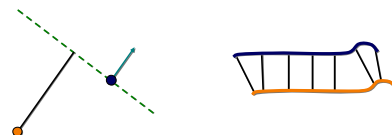
Slide courtesy of Szymon Rusinkiewicz

Outline

Molecular surfaces
- Definitions
- Representations

Surface matching
- Continuous surface mappings
- Discrete point correspondences

Surface retrieval
- Shape descriptors

Results

Discussion

Surface Retrieval

Goal: search a database of surfaces for the ones most similar to a query

Query

Database
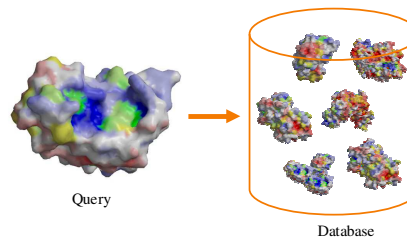
[Kinoshita03]

Surface Retrieval

Requirements:
- Should be fast (indexed)
- Should be conservative (don't miss any good matches)
- Can be approximate (check best matches in more detail)

Observation:
- Finding explicit correspondences/mappings for every pair of surfaces is too slow
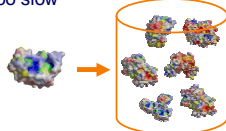
[Kinoshita03]

Surface Retrieval Pipeline

Database of Molecules

Query → Surface Retrieval

Candidates

Surface Matching

Best Matches

Indexed Search

Pairwise Matching
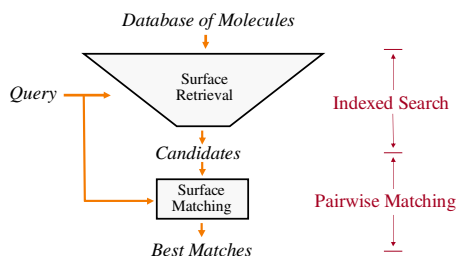
Surface Retrieval

General strategy:
- Compute shape descriptor for each surface
- Search for most similar shape descriptors

22 45 978 676 21 63 788 419

1 35 44 2 37 525 937 3 24

Similarity Measure
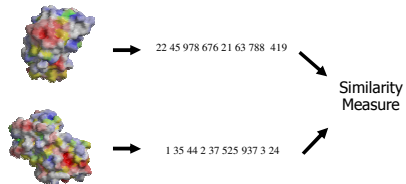
Surface Retrieval

Shape descriptor:
- Quick to compute
- Indexible
- Concise
- Invariant to translation
- Invariant to rotation
- Insensitive to small features
- Discriminating

Shape Descriptors

Avoid finding point correspondences
- Map surfaces to a feature space where correspondences are easy to find

Feature 1

Feature 2

Shape Descriptors

Avoid finding point correspondences
- Map surfaces to a feature space where correspondences are easy to find and classes are distinguishable

Feature 1

Feature 2

Shape Descriptors

Examples:
- Ø Set of geometric properties
- High-order moments
- Shape histograms
- etc.

Average Curvature

Average charge

Shape Descriptors

Examples:
- Set of geometric properties
- Ø High-order moments
- Shape histograms
- etc.

$$m_{pqr} = \int_{surface} x^p y^q z^r \, dx\,dy\,dz$$

Protein Surfaces   Moments

Similarity Measure

Shape Descriptors

Examples:
- Set of geometric properties
- Ø High-order moments
- Shape histograms
- etc.

$$m_{pqr} = \int_{surface} x^p y^q z^r \, dx\,dy\,dz$$

Align centers of mass

Protein Surfaces   Moments

Similarity Measure

Shape Descriptors

Examples:
- Set of geometric properties
- Ø High-order moments
- Shape histograms
- etc.

$$m_{pqr} = \int_{surface} x^p y^q z^r \, dx\,dy\,dz$$

Align centers of mass and principal axes

Protein Surfaces   Moments

Similarity Measure

Shape Descriptors

Examples:
- Set of geometric properties
- High-order moments
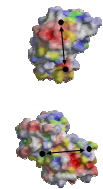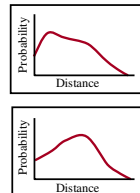
ØShape histograms
- etc.

These histograms are invariant under translations and rotations

Probability

Distance

Probability

Distance

Similarity Measure

Protein Surfaces   Shape Histogram
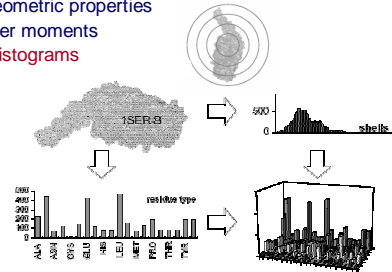
[Osada01]

Shape Descriptors

Examples:
- Set of geometric properties
- High-order moments

ØShape histograms
- etc.

1SER-3

shells

residue type

[Ankerst99]

Spherical Shape Descriptors

Examples:

ØShape histograms (sectors)
- Spherical extent function
- Extended Gaussian image
- Spherical attribute image
- etc.

Surface   Sectors

[Ankerst99]

Spherical Shape Descriptors

Examples:
- Shape histograms (sectors)

ØSpherical extent function
- Extended Gaussian image
- Spherical attribute image
- etc.

Surface   Spherical Extent Function

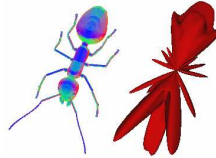## Spherical Shape Descriptors

Examples:
- Shape histograms (sectors)
- Spherical extent function

ØExtended Gaussian Image
- Spherical attribute image
- etc.

Surface   EGI
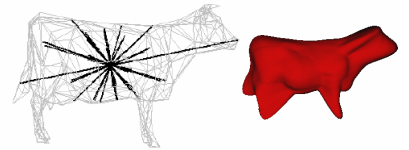
[Horn86]

Spherical Shape Descriptors

Examples:
- Shape histograms (sectors)
- Spherical extent function
- Extended Gaussian image

ØSpherical attribute image
- etc.

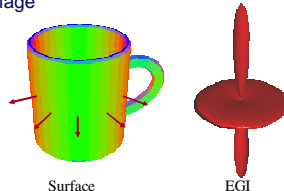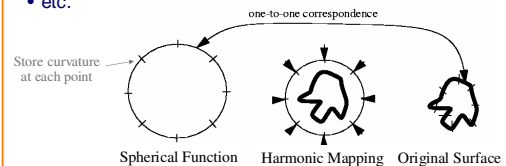one-to-one correspondence

Store curvature at each point

Spherical Function   Harmonic Mapping   Original Surface

[Ikeuchi95]

Outline

Molecular surfaces
- Definitions
- Representations

Surface matching
- Continuous surface mappings
- Discrete point correspondences

Surface retrieval
- Shape descriptors

Results ⟵

Discussion

Results

answer — prediction — similar binding site

[Kinoshita05]

Results

Query (1ayl)

[Kinoshita03]

Results

Query (1ayl)

Top Match from different SCOP fold (1fnc)
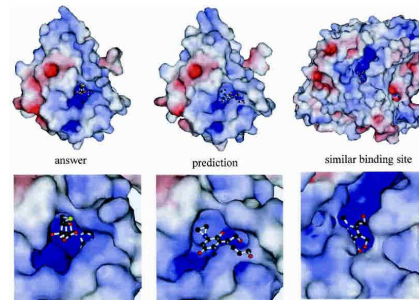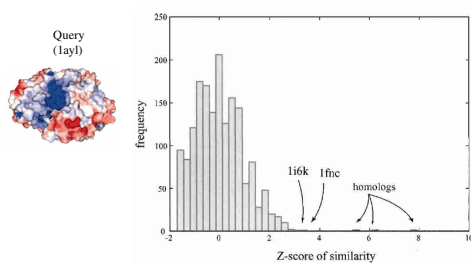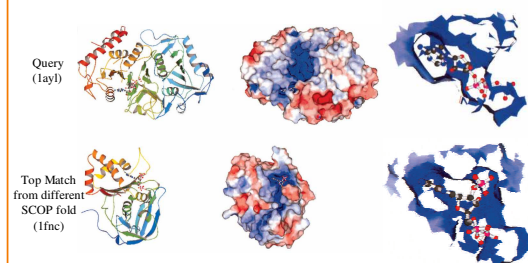
Similarities in adenosine base and sugar binding parts of site

"The relationship between their biochemical functions and the observed similarities is not clear." [Kinoshita03]

Results

1THL (red) — 1TMN (blue)

Surfaces colored by electrostatic potential

[Hofbauer04]

Discussion

?

13

References

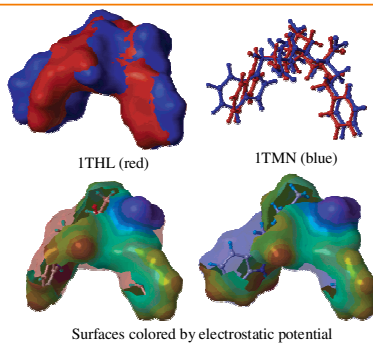[Ankerst99] Mihael Ankerst, Gabi Kastenmüller, Hans-Peter Kriegel, Thomas Seidl, "3D Shape Histograms for Similarity Search and Classification in Spatial Databases," SSD, 1999, pp. 207-226.

[Asirvatham05] A. Asirvatham, E. Praun, and H. Hoppe, "Consistent spherical parameterization", Computer Graphics and Geometric Modeling (CGGM) Workshop, 2005.

[Cai98] W. Cai, M. Zhang, B. Maigret, "New approach for representation of molecular surface," J. Comput. Chem, 19, 1998, pp. 1805-1815.

[Connolly83] M.L. Connolly, "Solvent-accessible surfaces of proteins and nucleic acids," Science, 221, 1983, pp. 709-713.

[Duncan93] B.S. Duncan, A.J. Olson, "Approximation and characterization of molecular surfaces," Biopolymers, 33, 1993, pp. 219-229.

[Hofbauer04] C. Hofbauer, H. Lohninger, A. Aszodi, "SURFCOMP: A Novel Graph-Based Approach to Molecular Surface Comparison," J. Chem. Inf. Comput. Sci., 44, 2004, pp. 837-847.

[Horn86] B.K.P. Horn, ``Extended Gaussian Images,'' Proceedings of the IEEE, Vol. 72, No. 12, December 1984, pp. 1671--1686.

[Ikeuchi95] K. Ikeuchi and M. Hebert. "Spherical Representations: from EGI to SAI," Tech. report CMU-CS-95-197, Computer Science Department, Carnegie Mellon University, October, 1995.

[Kinoshita05] K. Kinoshita,, H. Nakamura, "Identification of the ligand binding sites on the molecular surface of proteins," Protein Science, 14, 2005, pp. .

[Kinoshita03] K. Kinoshita,, H. Nakamura, "Identification of protein biochemical functions by similarity search using the molecular surface database eF-site," Protein Science, 12, 2003, pp. 1589-1595.

[Kinoshita02] K. Kinoshita, J. Furui, H. Nakamura, "Identification of Protein Functions from a Molecular Surface Database, eF-site," J. Struct. Func. Genomics, 2, 1, 2002, pp. 9-22.

[Lin94] S.L. Lin, R. Nussinov, D. Fischer, H.J. Wolfson, "Molecular-Surface Representations By Sparse Critical-Points," Proteins-Structure Function and Genetics, 18, 1994, pp. 94-101.

[Praun03] E. Praun and H. Hoppe, "Spherical Parameterization and Remeshing", SIGGRAPH, 2003.

[Rubner00] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance," International Journal of Computer Vision, 2000.

[Shulman-Peleg04] A. Shulman-Peleg, R. Nussinov, H.J. Wolfson, "Recognition of functional sites in protein structures," J Mol Biol, 339, 2004, pp. 607-633.

[Sorkine02] Olga Sorkine, Daniel Cohen-Or, Rony Goldenthal and Dani Lischinski, "Bounded-distortion Piecewise Mesh Parameterization", IEEE Visualization, 2002.