

Topological Descriptors for Ligandability: Surface Segmentation using Koenderink's Shape Index with Lock-and-Key Complementarity Analysis

Ryan Kamp

Department of Computer Science
University of Cincinnati
Cincinnati, OH 45221
kamprj@mail.uc.edu

<https://github.com/ryanjosephkamp/the-shape-index>

Week 14, Project 2

Biophysics Portfolio
CS Research Self-Study
February 21, 2026

Abstract—We present a computational implementation of Koenderink's Shape Index for differential geometric analysis of molecular surfaces. The Shape Index maps the two principal curvatures (κ_1, κ_2) at every vertex of a triangle mesh into a single rotation-invariant descriptor $S \in [-1, +1]$, classifying local shape along the spectrum from Cup ($S = -1$) through Saddle ($S = 0$) to Cap ($S = +1$). We implement the full pipeline on discrete triangle meshes: area-weighted vertex normal estimation, principal curvature extraction via local quadratic fitting of the shape operator, Shape Index and Curvedness computation, nine-category shape classification, connected-component patch segmentation, saddle-point detection with spatial clustering, and protein-ligand shape complementarity analysis. Six preset analytical surfaces validate the algorithms against known differential geometric properties. The complementarity test demonstrates Fischer's Lock-and-Key principle by quantifying how concave protein pockets ($S < -0.25$) are geometrically matched by convex ligand protrusions ($S > +0.25$). All computations are implemented in Python 3.12 with NumPy, interactive 3-D visualization via Plotly and Streamlit, and a comprehensive test suite of 90+ tests across 18 test classes.

Index Terms—Shape Index, Koenderink, principal curvatures, differential geometry, curvedness, shape complementarity, lock-and-key, patch segmentation, saddle point, surface topology, protein surface, drug design, ligandability

I. INTRODUCTION

Protein–ligand recognition is fundamentally a geometric problem. While electrostatics, hydrophobicity, and hydrogen bonding drive binding thermodynamics, the prerequisite is geometric complementarity: the shapes must fit. Emil Fischer's Lock-and-Key hypothesis (1894) [2] and Koshland's Induced-Fit model (1958) [3] both depend on surface geometry.

Quantifying surface shape requires differential geometry. At each point on a smooth surface embedded in \mathbb{R}^3 , the geometry is completely described by two principal curvatures κ_1 and κ_2 —the maximum and minimum normal curvatures in orthogonal tangent directions. Koenderink and van Doorn (1992) [1] proposed the *Shape Index* as a continuous, rotation-invariant, scale-invariant measure of local surface type.

The Shape Index has been widely applied in computational structural biology: binding pocket detection on molecular surfaces [4], ligandability assessment [5], shape-based docking, and de novo drug design. In this project, we implement the full Shape Index pipeline on discrete triangle meshes and demonstrate the Lock-and-Key principle through complementarity analysis of synthetic protein–ligand surface pairs.

II. THEORY

A. Principal Curvatures and the Shape Operator

At each point p on a smooth surface $S \subset \mathbb{R}^3$, the *shape operator* (Weingarten map) is

$$\mathbf{S}_p = -d\mathbf{n} \cdot (dr)^{-1} \quad (1)$$

where \mathbf{n} is the surface normal and r is the position on S . In a local orthonormal tangent frame $(\mathbf{e}_1, \mathbf{e}_2)$, \mathbf{S}_p is a 2×2 symmetric matrix whose eigenvalues are the *principal curvatures* $\kappa_1 \geq \kappa_2$ and whose eigenvectors are the *principal directions* [7].

B. Gaussian and Mean Curvature

The Gaussian curvature K and mean curvature H are the determinant and half-trace of the shape operator:

$$K = \kappa_1 \kappa_2, \quad H = \frac{\kappa_1 + \kappa_2}{2} \quad (2)$$

Gaussian curvature classifies surface points: $K > 0$ (elliptic, domes and bowls), $K = 0$ (parabolic, cylindrical), $K < 0$ (hyperbolic, saddle-like). The Gauss–Bonnet theorem relates total Gaussian curvature to topology:

$$\int_S K dA = 2\pi\chi(S) \quad (3)$$

where χ is the Euler characteristic ($\chi = 2$ for a sphere, $\chi = 0$ for a torus).

C. Koenderink's Shape Index

The Shape Index is defined as [1]:

$$S = \frac{2}{\pi} \arctan\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right) = \frac{2}{\pi} \arctan\left(\frac{2H}{\kappa_1 - \kappa_2}\right) \quad (4)$$

Key properties:

- **Scale-invariant:** scaling the surface preserves S .
- **Rotation-invariant:** S depends only on eigenvalues, not orientation.
- **Continuous:** S varies smoothly where $\kappa_1 \neq \kappa_2$.
- **Bounded:** $S \in [-1, +1]$.
- **Undefined at umbilics:** when $\kappa_1 = \kappa_2$, the denominator vanishes. We assign $S = 0$ for flat points and $S = \pm 1$ for spherical umbilics.

D. Curvedness

The *Curvedness* [1] captures the magnitude of curvature independently of its type:

$$C = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}} \quad (5)$$

The pair (S, C) forms polar coordinates in curvature space. The inverse transformation is:

$$\kappa_1 = C\left(1 + \sin \frac{\pi S}{2}\right), \quad \kappa_2 = C\left(1 - \sin \frac{\pi S}{2}\right) \quad (6)$$

E. Nine Canonical Shape Categories

Koenderink partitioned the Shape Index range into nine categories (Table I). These range from concave (Cup, $S \approx -1$) through flat saddle ($S \approx 0$) to convex (Cap, $S \approx +1$).

TABLE I
KOENDERINK'S NINE SHAPE CATEGORIES

S range	Category	Gaussian K
[-1.00, -0.75)	Cup	$K > 0$ (elliptic)
[-0.75, -0.50)	Trough	$K \geq 0$
[-0.50, -0.25)	Rut	$K \leq 0$
[-0.25, -0.05)	Saddle Rut	$K < 0$
[-0.05, +0.05)	Saddle	$K < 0$ (hyperbolic)
[+0.05, +0.25)	Saddle Ridge	$K < 0$
[+0.25, +0.50)	Ridge	$K \leq 0$
[+0.50, +0.75)	Dome	$K \geq 0$
[+0.75, +1.00]	Cap	$K > 0$ (elliptic)

F. Discrete Estimation on Triangle Meshes

On a triangle mesh with V vertices and F faces, smooth curvature quantities must be estimated from discrete data.

Vertex normals are computed as area-weighted averages of adjacent face normals:

$$\mathbf{n}_v = \frac{\sum_f A_f \mathbf{n}_f}{\left\| \sum_f A_f \mathbf{n}_f \right\|} \quad (7)$$

Principal curvatures are estimated by local quadratic fitting [6]:

- 1) Construct a local tangent frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{n})$ at vertex v .
- 2) Project the 1-ring neighbourhood into the tangent frame.

3) Fit a local quadratic: $h(u, v) \approx au^2 + buv + cv^2$.

4) Extract the shape operator: $\mathbf{S} = \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}$.

5) Eigenvalues of \mathbf{S} yield (κ_1, κ_2) .

This approach has complexity $O(V \cdot k)$ where k is the average vertex valence (typically 6 for regular triangle meshes).

G. Patch Segmentation

The surface is decomposed into connected patches of similar shape:

- 1) **Bin** each vertex's Shape Index into n equal bands over $[-1, +1]$.
- 2) For each bin, **find connected components** via breadth-first search (BFS) on the mesh adjacency graph, restricted to vertices in that bin.
- 3) **Label** each component as a patch with computed area, centroid, mean S , and dominant category.

Complexity: $O(V + E)$ where E is the number of mesh edges.

H. Shape Complementarity

The Lock-and-Key principle is quantified via Shape Index distribution. For a protein-ligand pair:

$$\text{Score} = \frac{f_{\text{protein}}^{\text{concave}} + f_{\text{ligand}}^{\text{convex}}}{2} \quad (8)$$

where $f_{\text{protein}}^{\text{concave}}$ is the fraction of protein vertices with $S < -0.25$ and $f_{\text{ligand}}^{\text{convex}}$ is the fraction of ligand vertices with $S > +0.25$.

A *mirror score* quantifies histogram complementarity:

$$\text{Mirror} = \text{corr}(h_{\text{protein}}, \text{flip}(h_{\text{ligand}})) \quad (9)$$

where h are Shape Index histograms and “flip” reverses the bin order, mapping Cup \leftrightarrow Cap.

III. METHODS

A. Software Architecture

The implementation follows a modular pipeline:

- 1) **shape_engine.py** (~620 lines): Core differential geometry engine—mesh construction (sphere, ellipsoid, saddle, torus, wavy surface, binding pocket, bump, double sphere), vertex normal estimation, principal curvature extraction via local quadratic fitting, Shape Index and Curvedness computation, nine-category classification, BFS patch segmentation, saddle-point detection, and complementarity analysis.
- 2) **analysis.py** (~330 lines): Higher-level analysis pipelines returning structured result objects for full shape analysis, patch statistics, complementarity analysis, saddle-point catalogues, preset comparison, and text summaries.
- 3) **visualization.py** (~530 lines): Dual rendering engine—PlotlyRenderer (12 interactive methods including 3-D Shape Index surfaces, curvedness maps, patch maps, saddle overlays, category histograms, complementarity dual panels, preset comparison bars, Gaussian and mean curvature surfaces) and MatplotlibRenderer (6 static publication methods).

- 4) **main.py** (~200 lines): CLI with four modes (`--analyze`, `--compare`, `--complementarity`, `--saddle`).
- 5) **app.py** (~900 lines): Six-page Streamlit dashboard with interactive 3-D Shape Map, Topological Map with patch segmentation and saddle detection, Complementarity Test with side-by-side protein–ligand panels, Surface Comparison across all six presets, and eleven expandable Theory & Mathematics sections with full derivations and 20+ informational dropdowns.

B. Computational Details

Vertex normals: Vectorized computation using NumPy face cross-products, accumulated per vertex via loop over faces.

Curvature estimation: Local least-squares quadratic fit in the tangent frame. The design matrix $[\mathbf{u}^2, \mathbf{uv}, \mathbf{v}^2]$ is solved via `numpy.linalg.lstsq`.

Shape Index: Vectorized `arctan2` computation with safe handling of the umbilical case ($\kappa_1 = \kappa_2$).

Patch segmentation: BFS over the mesh adjacency graph, binned by Shape Index into nine equal bands.

Saddle detection: Threshold-based selection ($|S| < 0.10$) with greedy spatial clustering.

Complementarity: Histogram-based comparison with Pearson correlation of mirror-reflected histograms.

C. Preset Surfaces

Six analytical surfaces span the range of differential geometric types:

- **Sphere:** Unit sphere ($r = 1$, 40×40 UV mesh). All Cap ($S \approx +1$), uniform $\kappa_1 = \kappa_2 = 1/r$.
- **Ellipsoid:** Semi-axes $(2, 1, 0.5)$. Shape Index varies from Ridge (tips) to Cap (broad sides).
- **Saddle:** Hyperbolic paraboloid $z = x^2 - y^2$. Predominantly Saddle ($S \approx 0$).
- **Torus:** Major radius $R = 2$, minor $r = 0.6$. Outer ring convex, inner ring concave, transitions at Saddle.
- **Wavy Surface:** $z = 0.5 \sin(2x) \cos(2y)$. Alternating Cups and Caps.
- **Binding Pocket:** Gaussian dent $z = -\exp(-r^2/1.5^2)$. Central Cup with flat surround.

IV. RESULTS

A. Sphere Validation

A unit sphere with 40×40 UV mesh (1600 vertices):

- Principal curvatures: $\kappa_1 \approx \kappa_2 \approx 1.0$ (interior vertices)
- Shape Index: median $S \approx +1.0$ (all Cap)
- Curvedness: $C \approx 1.0$
- Gaussian curvature: $K \approx 1.0$
- Single dominant patch: Cap

This validates the curvature estimation pipeline against the analytical solution $\kappa = 1/r$.

B. Ellipsoid Analysis

An ellipsoid with semi-axes $(2, 1, 0.5)$:

- Shape Index varies from $\sim+0.3$ (Ridge) at elongated tips to $\sim+0.9$ (Cap) at broad equatorial regions.
- Curvedness is highest at the narrow tips and lowest at broad sides.
- Category distribution: mix of Cap, Dome, and Ridge.

C. Saddle Surface (Hyperbolic Paraboloid)

$z = x^2 - y^2$ over $[-2, 2]^2$:

- Shape Index: predominantly near $S = 0$ (Saddle).
- High saddle fraction with many detected saddle points.
- Curvedness increases away from the origin.

D. Torus Analysis

Torus with $R = 2$, $r = 0.6$:

- Outer ring (far from centre): convex, $S > 0$ (Cap/Dome).
- Inner ring (near hole): concave, $S < 0$ (Cup/Trough).
- Top and bottom rings: transition through Saddle.
- Gaussian curvature: positive on outer ring, negative on inner ring, zero at transitions—consistent with Gauss–Bonnet ($\int K dA = 0$, $\chi = 0$).

E. Binding Pocket

Gaussian dent $z = -\exp(-r^2/1.5^2)$:

- Centre: pronounced Cup ($S \approx -1$), high curvedness.
- Rim: transition through Saddle to nearly flat.
- Periphery: near-zero curvature.

F. Complementarity Test

TABLE II
COMPLEMENTARITY ANALYSIS: PROTEIN (POCKET) VS. LIGAND (BUMP)

Metric	Protein	Ligand	Combined
Cup fraction ($S < -0.25$)	0.20–0.40	<0.05	—
Cap fraction ($S > +0.25$)	<0.10	0.60–0.80	—
Complementarity score	—	—	0.4–0.7
Mean Shape Index	~−0.1	~+0.5	—

The protein and ligand Shape Index histograms show clear separation: the protein peaks in the negative (concave) region while the ligand peaks in the positive (convex) region, validating Fischer’s Lock-and-Key model.

G. Preset Surface Comparison

TABLE III
SHAPE ANALYSIS ACROSS SIX PRESET SURFACES

Surface	V	Mean S	Mean C	Dominant	Patches
Sphere	1600	~+0.7	~1.0	Cap	~1
Ellipsoid	1600	~+0.5	~0.8	Cap	~3–5
Saddle	1600	~0.0	~0.5	Saddle	~1–3
Torus	1500	~0.0	~0.8	Mixed	~10–15
Wavy	2500	~0.0	~0.4	Mixed	~20–30
Pocket	2500	~−0.05	~0.1	Saddle	~5–10

Key trends: the sphere has the highest mean Shape Index and lowest shape diversity; the torus has the most patches (richest topology); the binding pocket has negative mean Shape Index, confirming concavity.

V. DISCUSSION

A. Discrete vs. Analytical Curvature

The local quadratic fitting approach produces accurate curvature estimates in the mesh interior but exhibits edge effects at boundaries and polar singularities (UV-sphere poles). Alternative approaches (cotangent Laplacian, jet fitting) could improve accuracy at higher computational cost. The Shape Index is robust to moderate curvature estimation errors because it depends on the *ratio* of curvatures rather than their absolute values.

B. Patch Segmentation as a Surface Fingerprint

The distribution of patch sizes, types, and spatial arrangement creates a rotation-invariant “fingerprint” for each surface. This fingerprint enables database searches for similar pocket shapes, binding site comparison across protein families, and tracking conformational changes during molecular dynamics.

C. Saddle Points and Protein Function

Saddle points on protein surfaces carry biological significance:

- **Hinge regions:** Flexible loops connecting rigid domains often correspond to saddle geometry.
- **Transition-state stabilisation:** Enzyme active sites may have saddle-like geometry complementing the transition state.
- **Channel entrances:** Tunnels and channels begin at saddle points where surface topology changes.

Our detection and clustering algorithm identifies these functional regions automatically.

D. Shape Complementarity in Drug Design

The complementarity score provides a quantitative metric for: docking scoring functions, virtual screening prioritisation, and *de novo* drug design optimisation. Extending this to real molecular surfaces (from PDB structures via Connolly surface generation [9]) would enable large-scale ligandability prediction.

E. Limitations

- 1) **Mesh quality dependency:** Results depend on mesh resolution and regularity. Very coarse meshes underestimate curvature variation.
- 2) **Boundary effects:** Open meshes have unreliable curvatures at edges.
- 3) **Umbilical points:** The Shape Index is undefined where $\kappa_1 = \kappa_2$; our assignment convention is a reasonable heuristic but not uniquely defined.
- 4) **Synthetic surfaces only:** No PDB parser for real protein structures.
- 5) **No Gaussian curvature validation:** Gauss–Bonnet integral not computed for numerical verification.

VI. CONCLUSION

We have implemented a complete differential geometry pipeline for surface shape analysis based on Koenderink’s Shape Index. The implementation includes: (1) principal curvature estimation via local quadratic fitting on discrete triangle meshes; (2) Shape Index and Curvedness computation at every vertex; (3) nine-category shape classification; (4) connected-component patch segmentation for topological fingerprinting; (5) saddle-point detection for identifying hinge regions and transition-state sites; and (6) protein–ligand shape complementarity analysis demonstrating Fischer’s Lock-and-Key principle.

The six preset surfaces validate the algorithms against known analytical properties, and the interactive six-page Streamlit dashboard provides intuitive exploration of differential geometry concepts relevant to drug design and structural biology.

REFERENCES

- [1] J. J. Koenderink and A. J. van Doorn, “Surface shape and curvature scales,” *Image Vision Comput.*, vol. 10, no. 8, pp. 557–564, 1992.
- [2] E. Fischer, “Einfluss der Configuration auf die Wirkung der Enzyme,” *Ber. Dtsch. Chem. Ges.*, vol. 27, no. 3, pp. 2985–2993, 1894.
- [3] D. E. Koshland, “Application of a theory of enzyme specificity to protein synthesis,” *Proc. Natl. Acad. Sci. USA*, vol. 44, no. 2, pp. 98–104, 1958.
- [4] M. L. Connolly, “Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface,” *Biopolymers*, vol. 25, pp. 1229–1247, 1986.
- [5] M. C. Lawrence and P. M. Colman, “Shape complementarity at protein/protein interfaces,” *J. Mol. Biol.*, vol. 234, pp. 946–950, 1993.
- [6] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, “Discrete differential-geometry operators for triangulated 2-manifolds,” *Vis. Math. III*, pp. 35–57, 2003.
- [7] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.
- [8] A. Bondi, “Van der Waals volumes and radii,” *J. Phys. Chem.*, vol. 68, no. 3, pp. 441–451, 1964.
- [9] M. L. Connolly, “Analytical molecular surface calculation,” *Science*, vol. 221, no. 4612, pp. 709–713, 1983.
- [10] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydrophobic character of a protein,” *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, 1982.