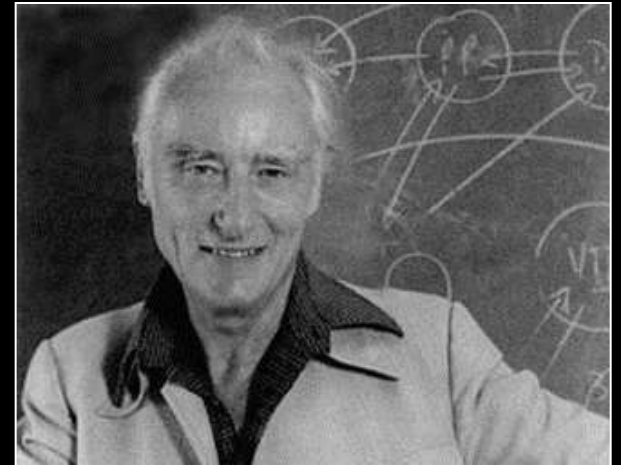
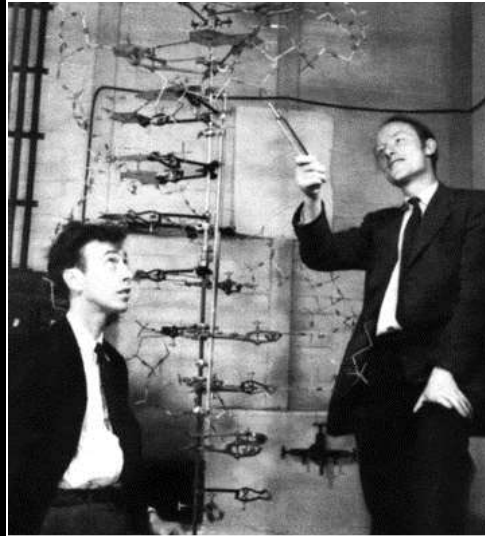


A Quantitative Universe

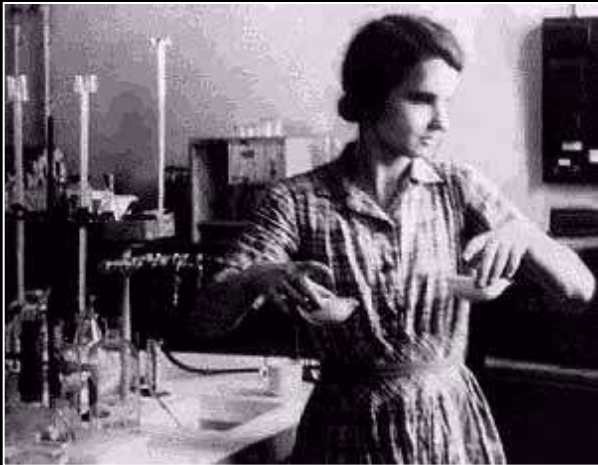
Brian Y. Chen



Francis Crick



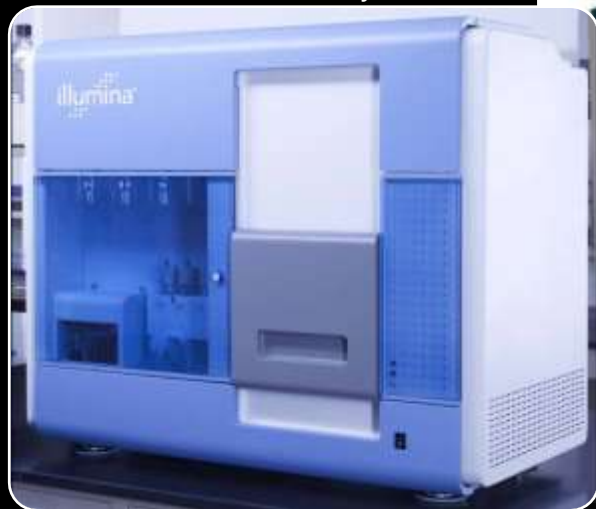
James Watson



Rosalind Franklin

SOLEIL beamline diagram, Paris

Illumina Genome Analyzer II



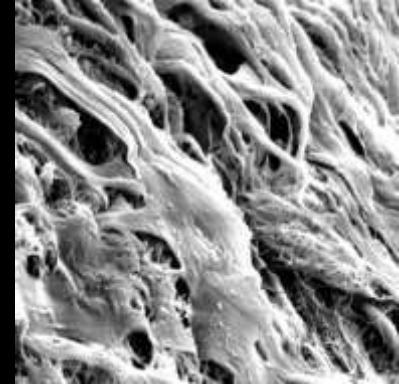
Drug Design



Biofuels



Biomaterials



Diagnostics



Agriculture



Synthetic Biology

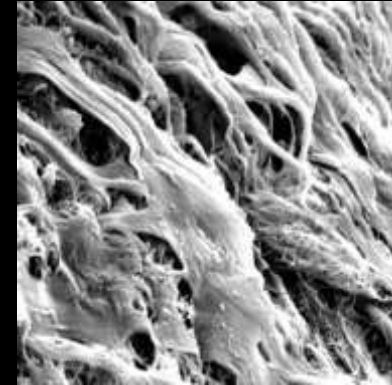
Drug Design



Biofuels



Biomaterials



How do these biological systems work?



Diagnostics

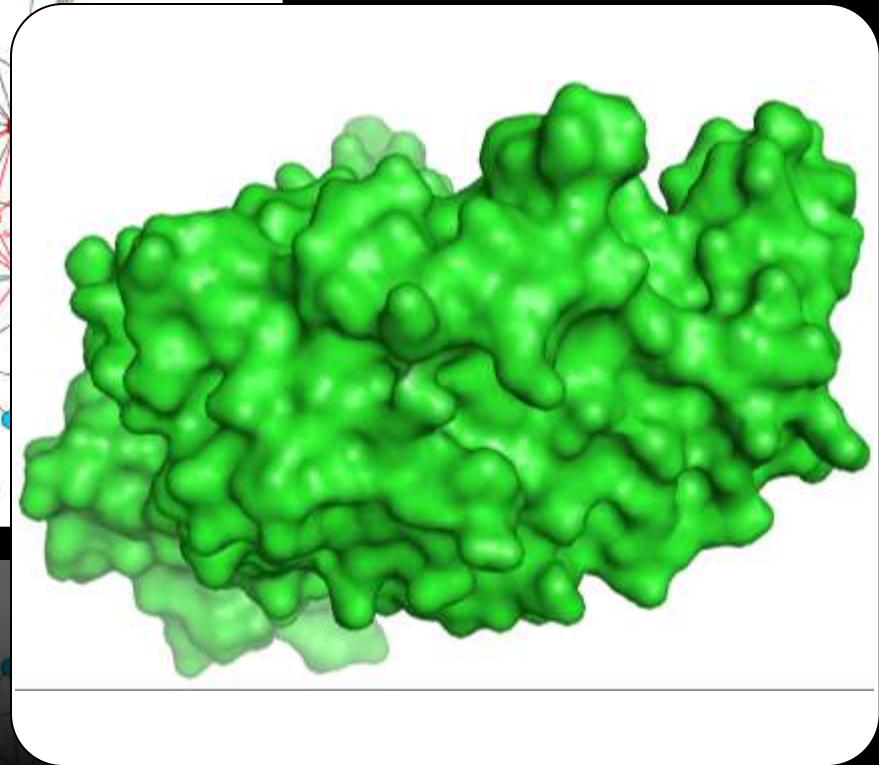
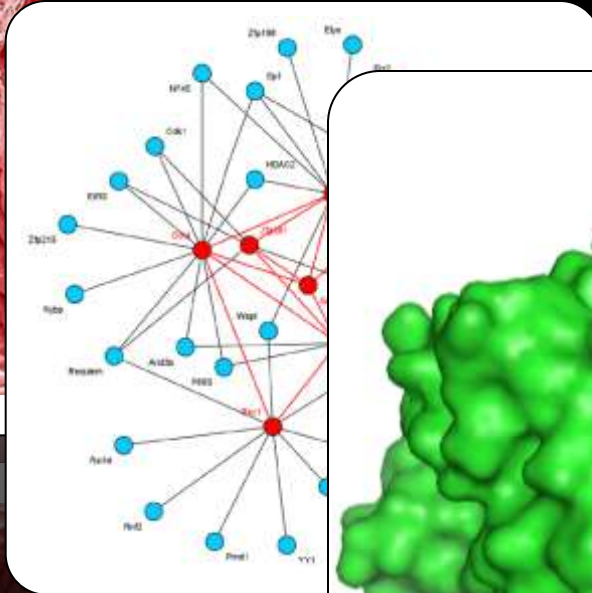
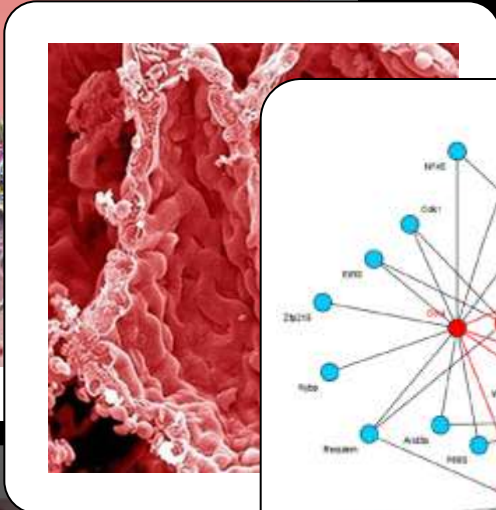
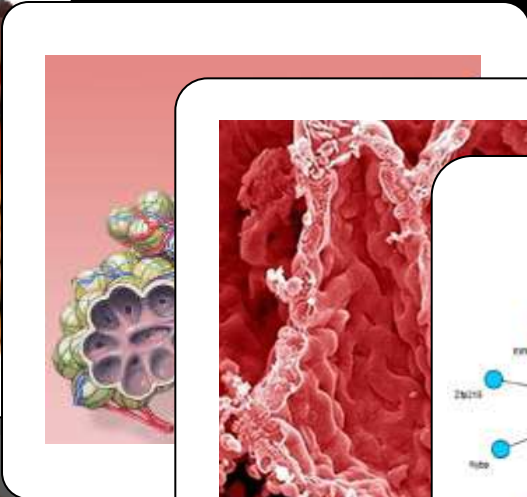


Agriculture

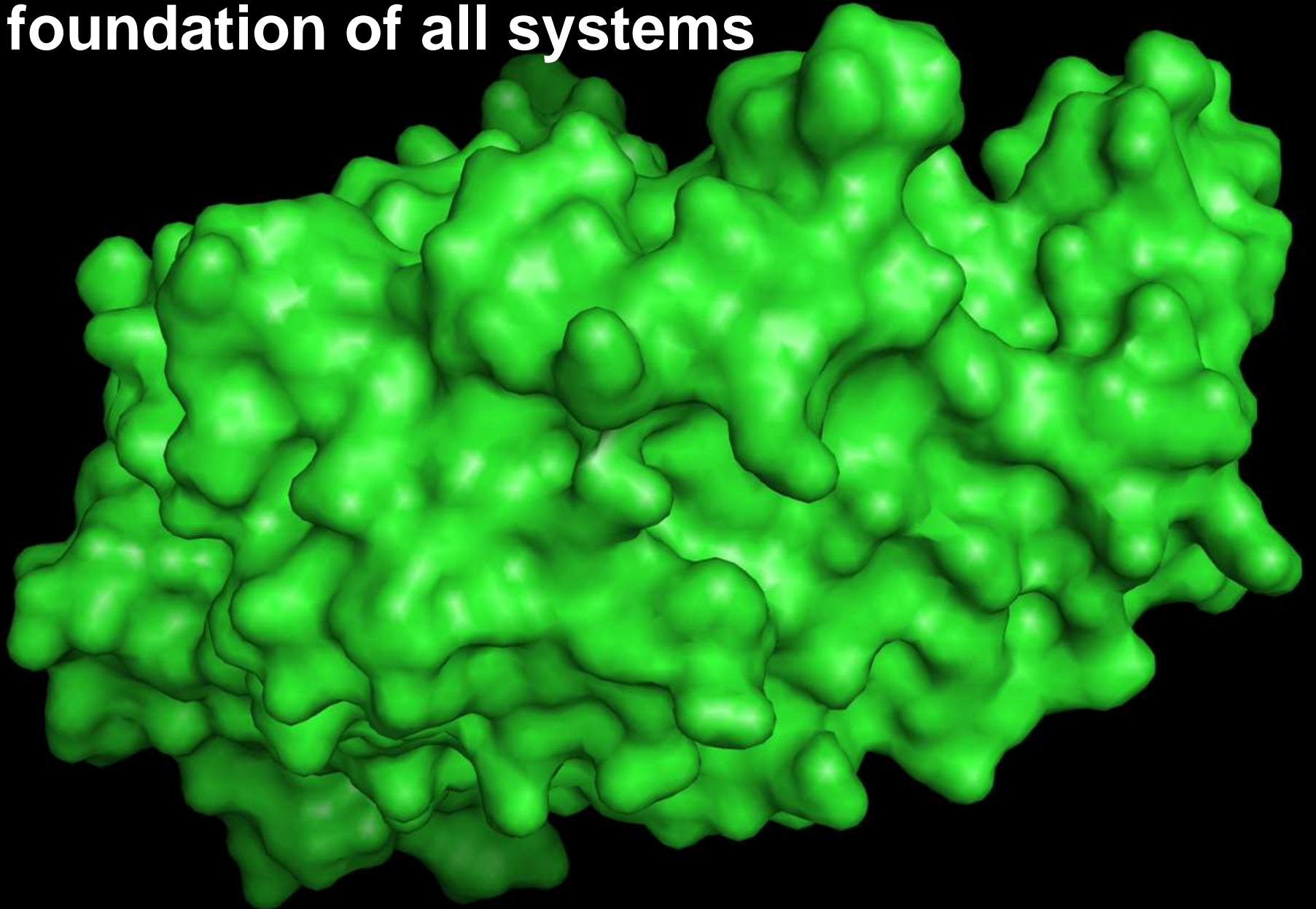


Synthetic Biology

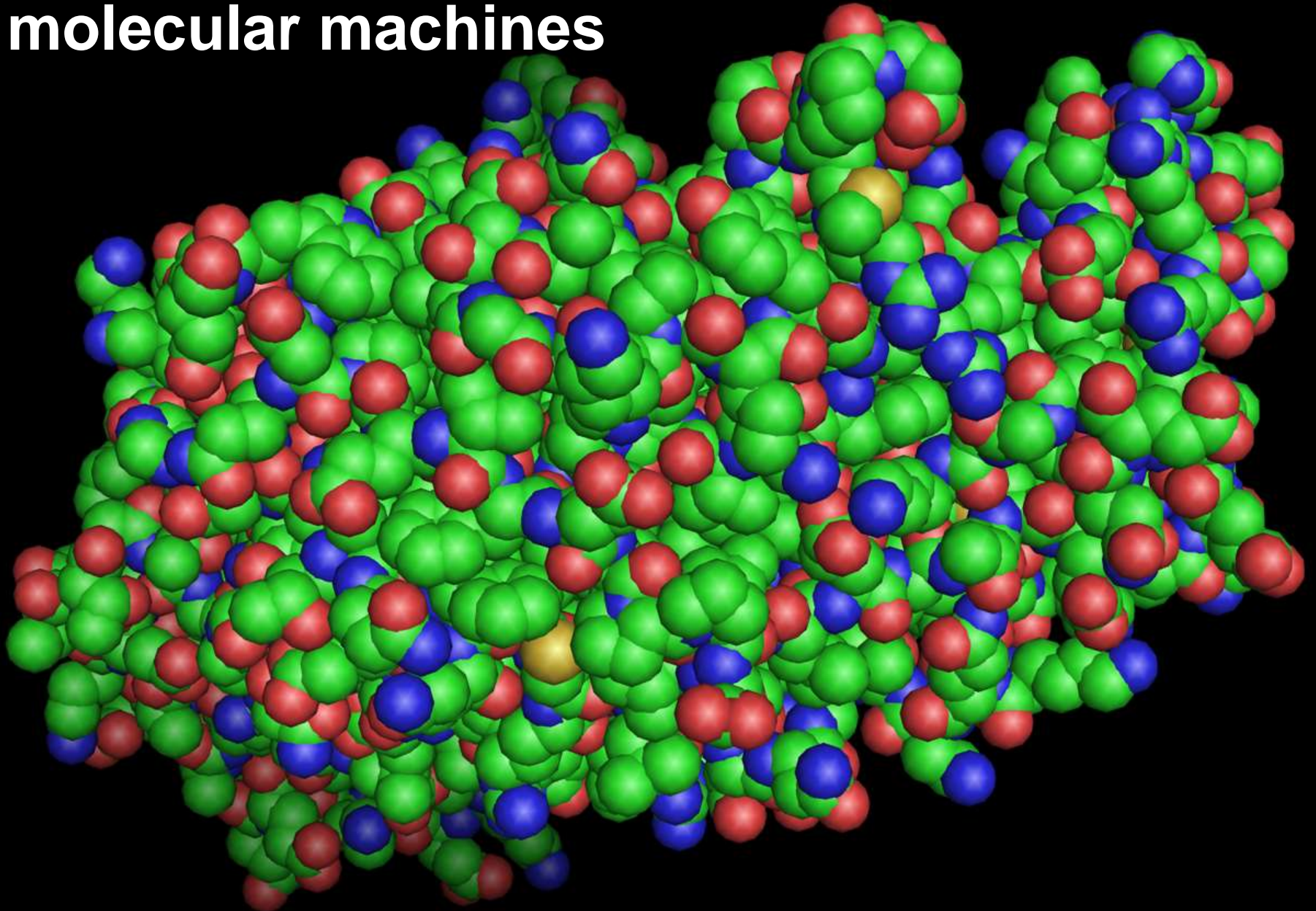
Biological systems are nested and interacting machines



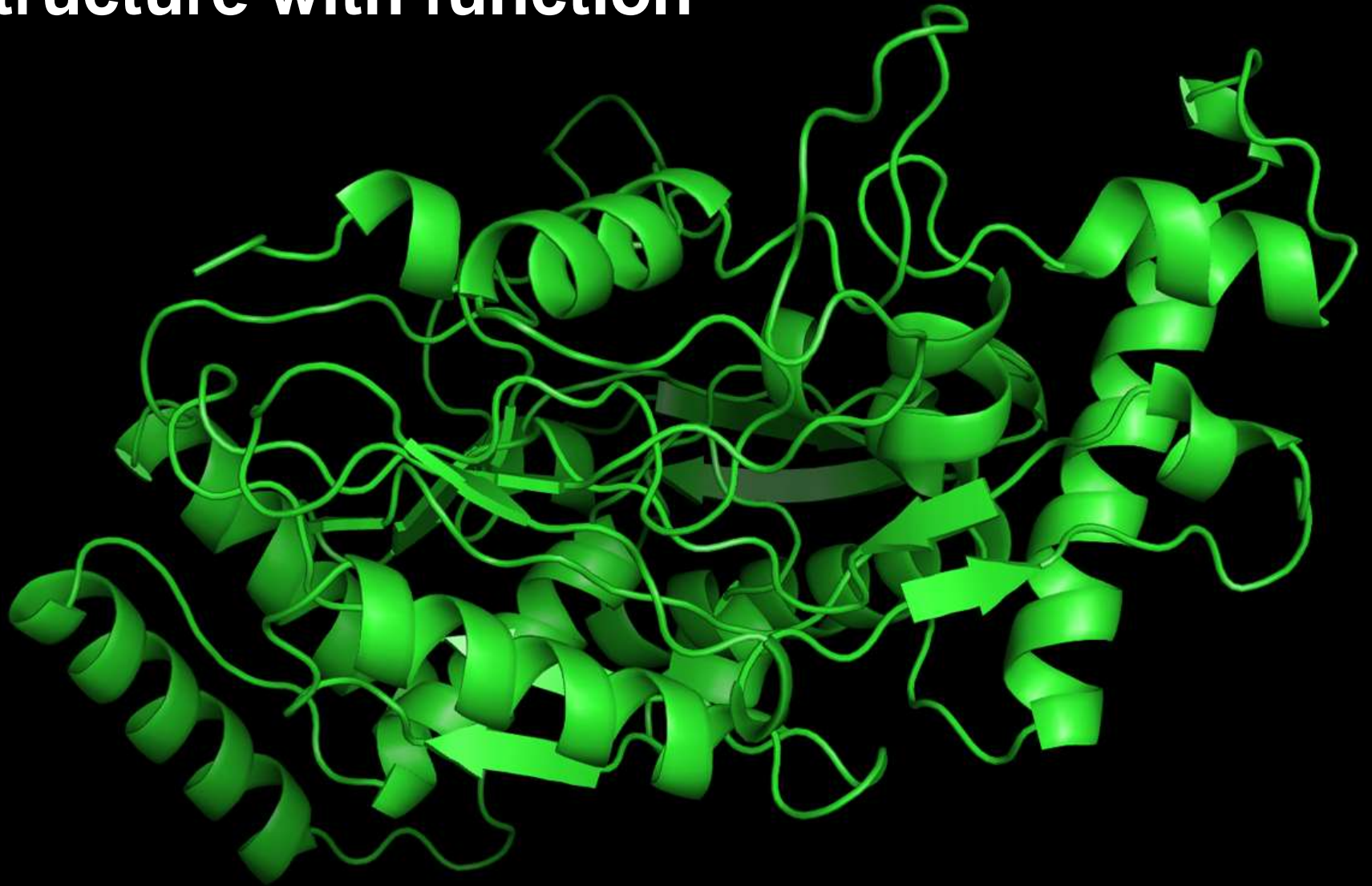
**Individual molecules are the
foundation of all systems**



Structural biology studies molecular machines



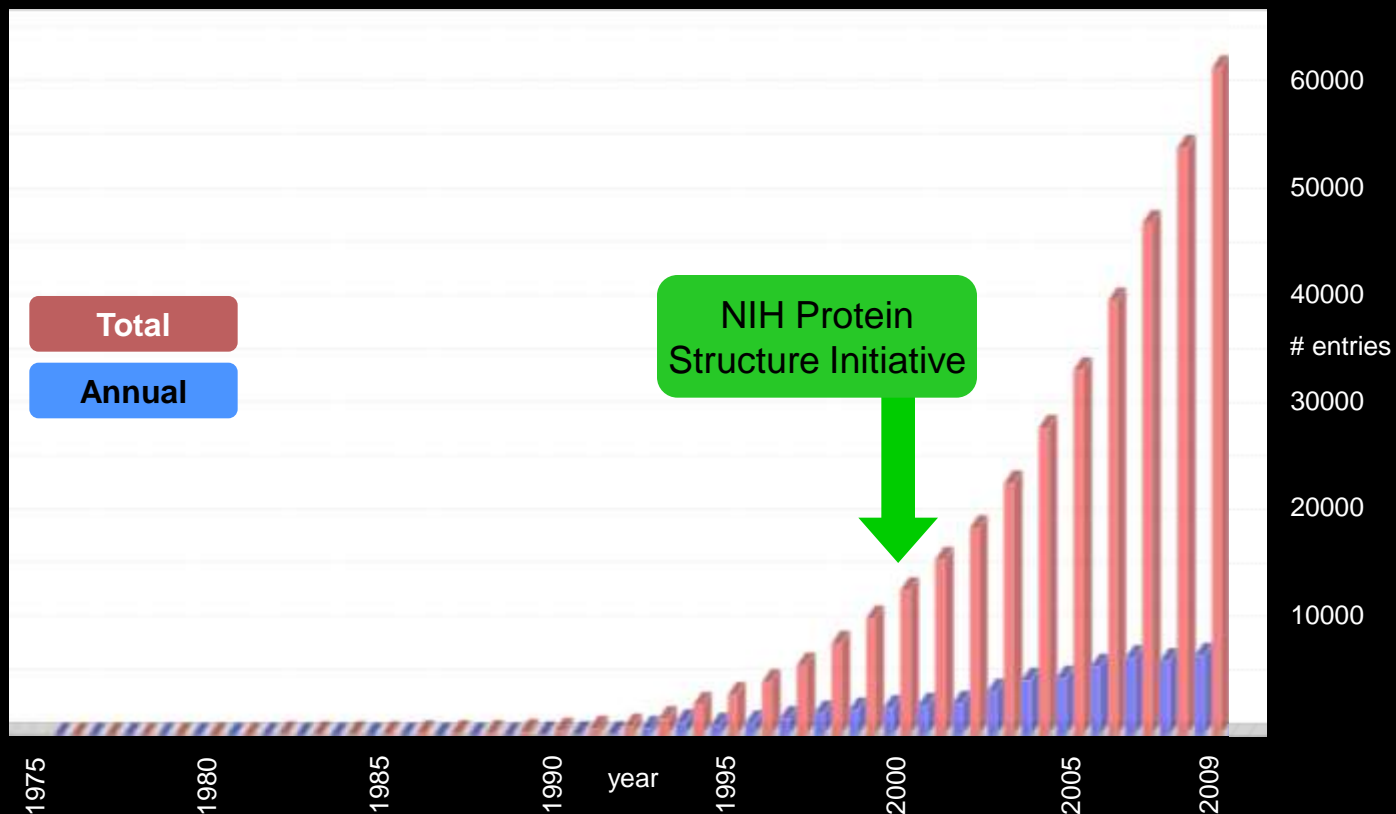
Structural biology connects structure with function



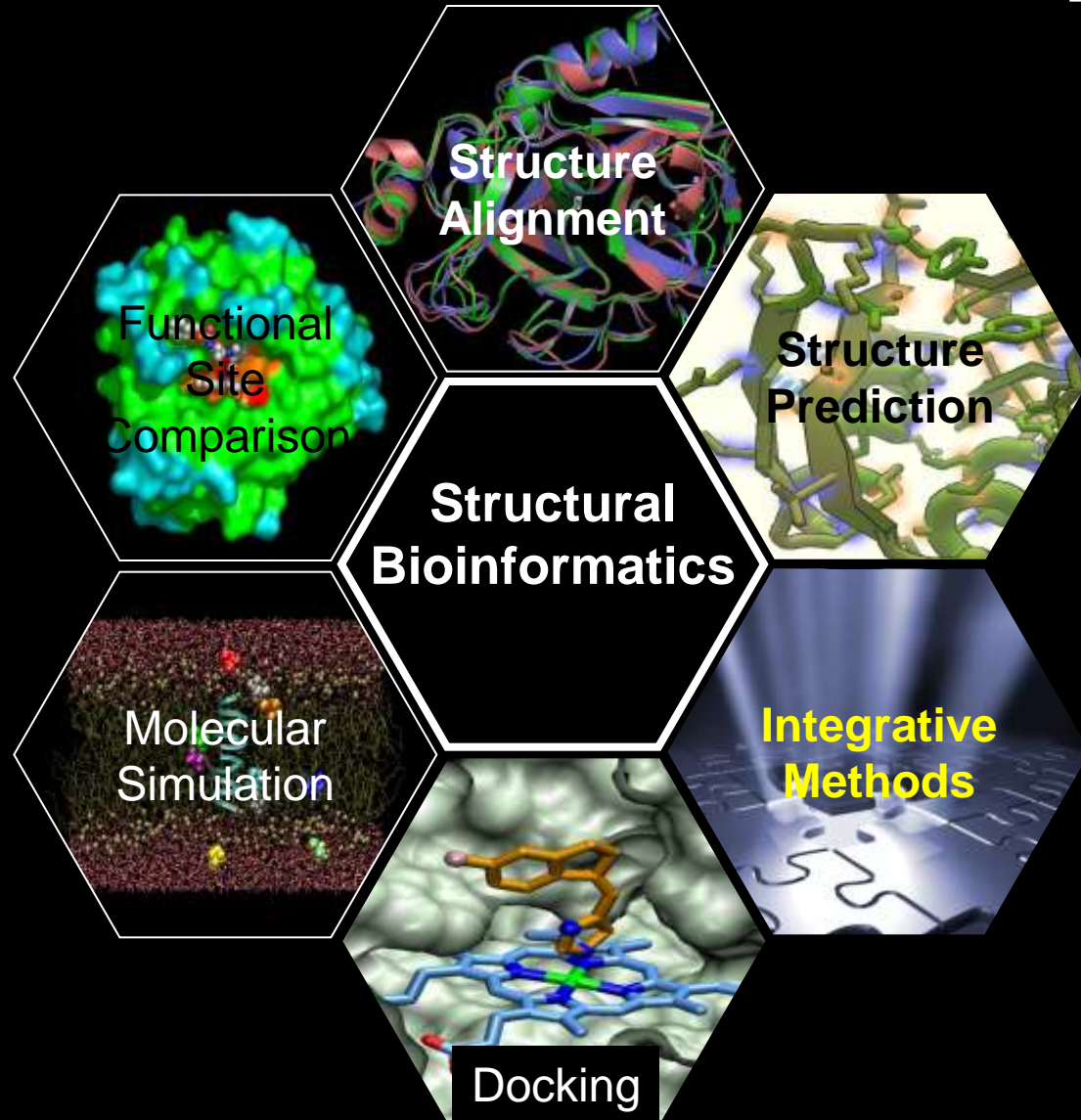
Structural biology has become a quantifiable universe



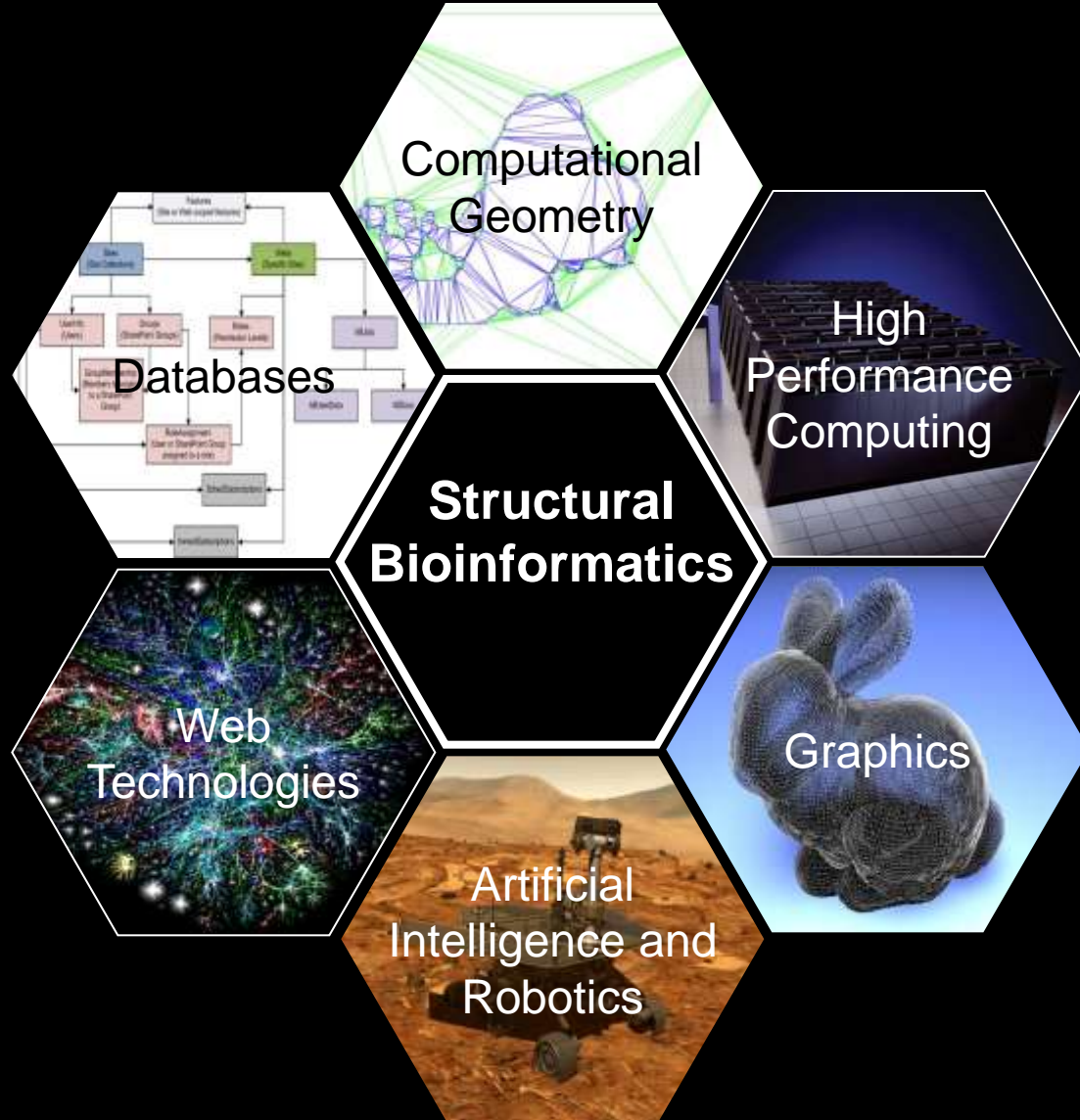
Number of Entries in the Protein Data Bank



Structural bioinformatics connects structure with function at scale and with precision



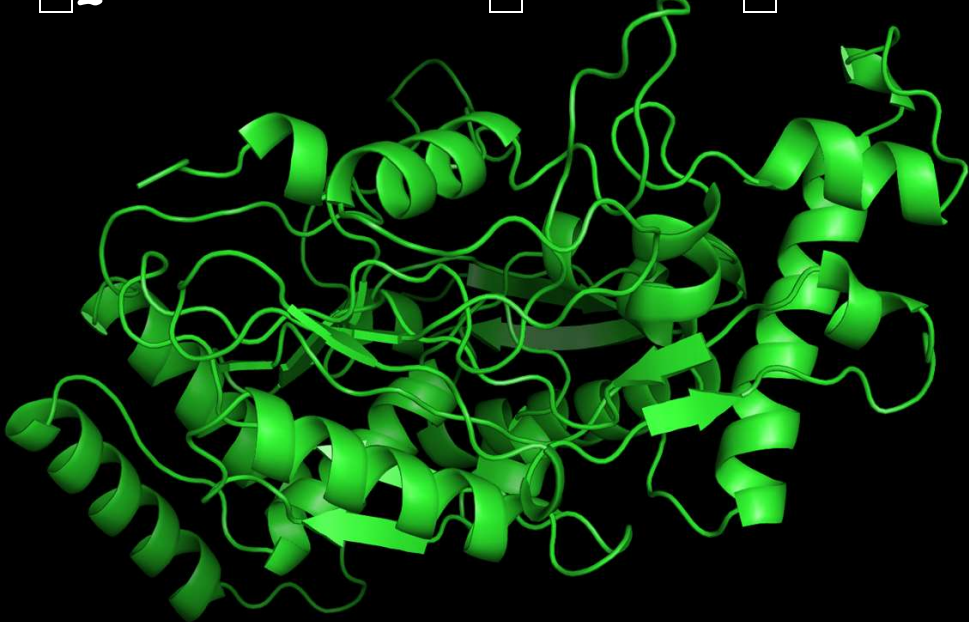
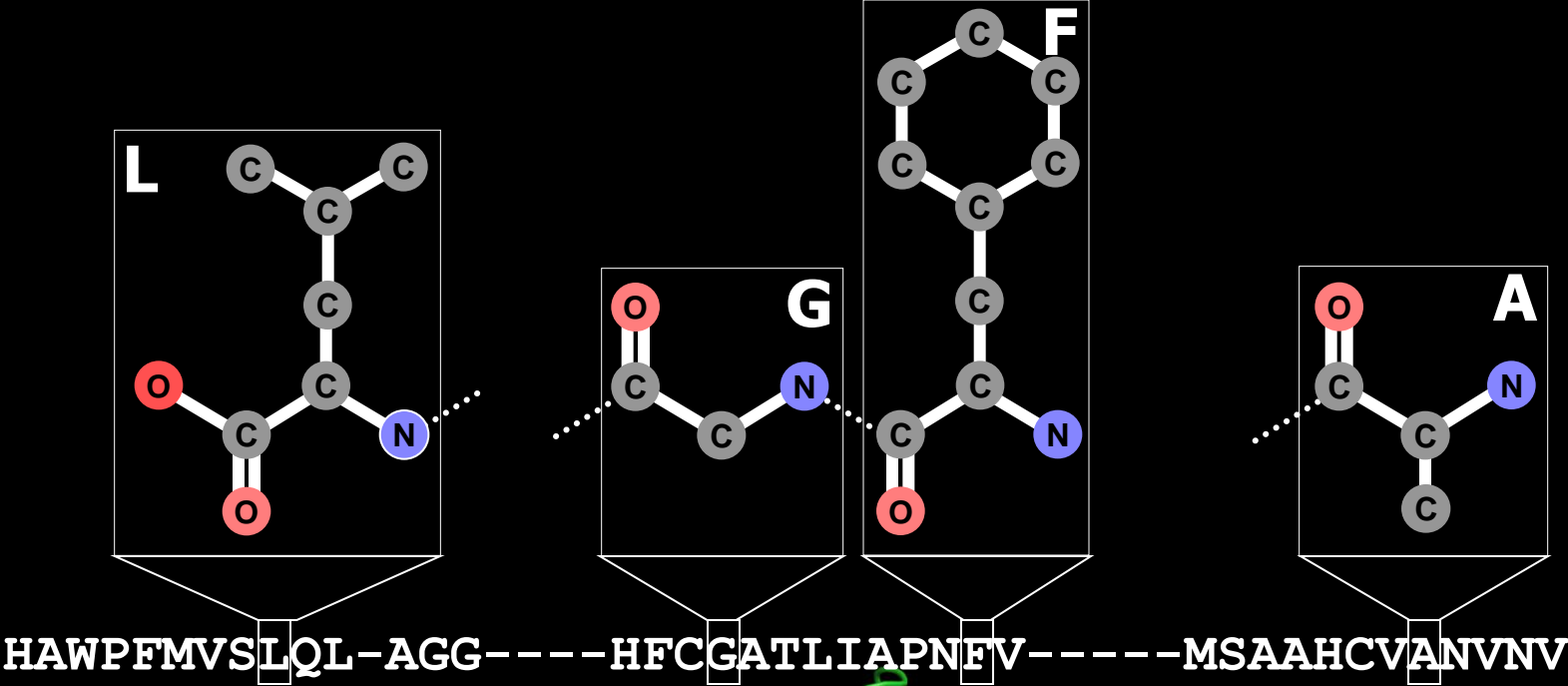
Structural bioinformatics draws from many computational fields



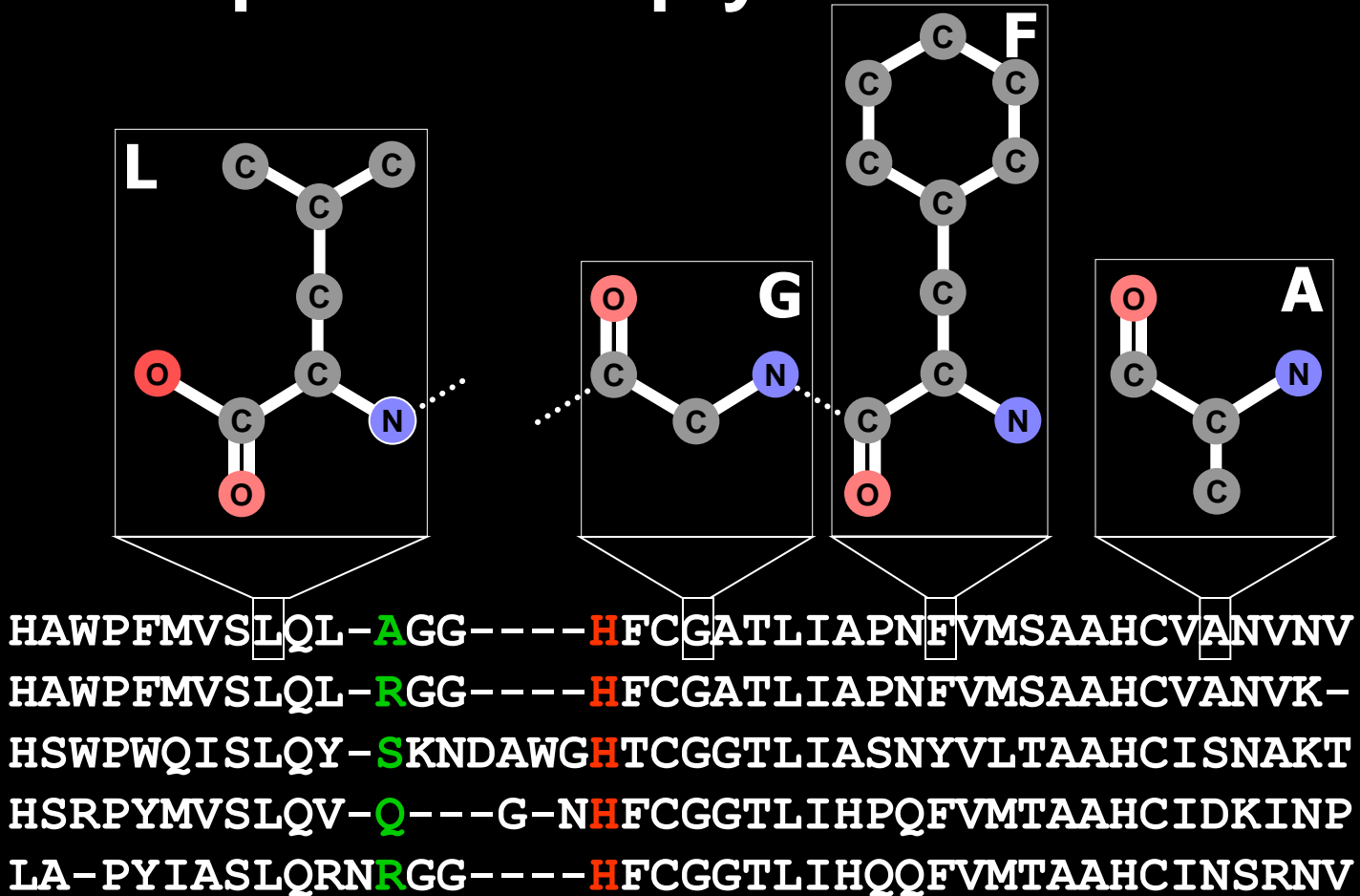
The General Problem:

**Gather, analyze, and integrate data that
can indicate function**

Proteins are chains of amino acids

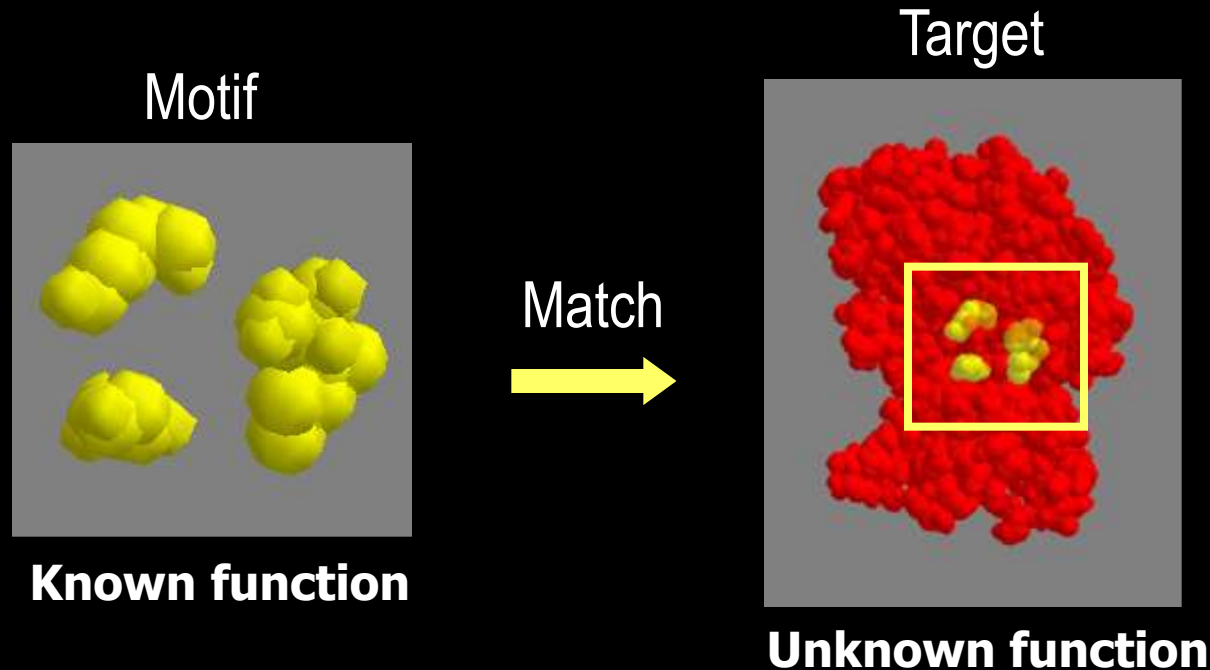


Similar sequences imply similar function



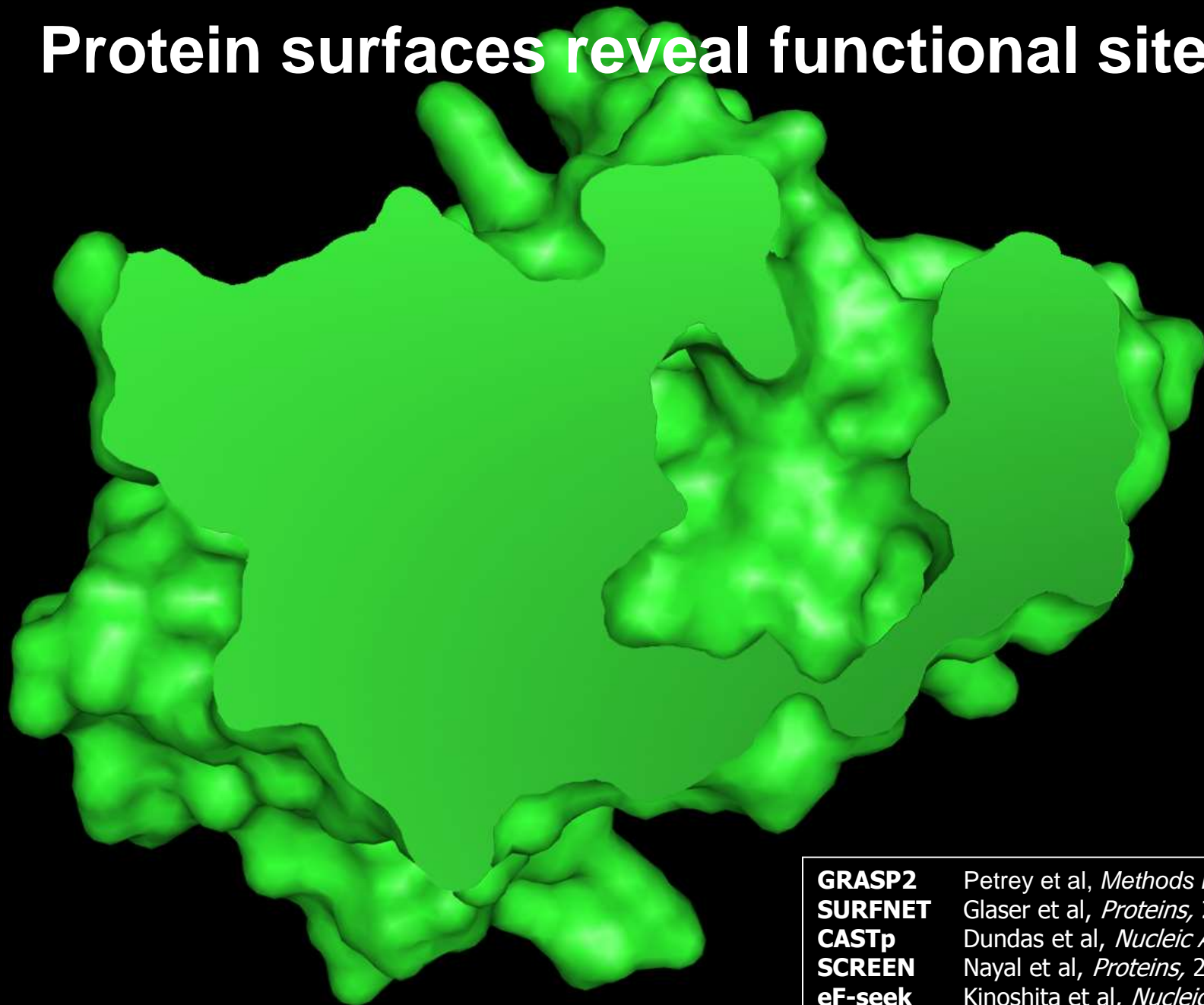
ConSurf Glaser, et al. *Bioinformatics*, 2003.
Evolutionary Trace Mihalek, et al. *Proteins*, 2006.
HMAP Tang, et al. *J. Mol. Biol.* 2003.
FASTA Mackey, et al. *Mol. Cell. Prot.* 2002.
CLUSTALW Larkin et al. *Bioinformatics.*, 2007.
BLAST Altschul et al. *Nuc. Acid. Res.* 1997.

Similar functional sites imply similar function



MASH	Chen et al, <i>J. Comput. Biol.</i> , 2007
Combinatorial Extension	Jia et al, <i>J. Comput. Biol.</i> , 2004
Geometric Hashing	Nussinov et al, <i>Proteins</i> , 2001
pevoSOAR	Tseng et al, <i>J. Mol. Biol.</i> , 2009
Ska	Petrey et al, <i>Methods Enzymol.</i> 2003.
Geometric Sieving	Chen et al, <i>J. Bioinf. Comput. Biol.</i> , 2007
PINTS	Stark et al, <i>Nucleic Acids Res</i> , 2003.
JESS	Barker et al, <i>Bioinformatics</i> , 2003.
Dali	Holm et al, <i>Bioinformatics</i> , 2008.

Protein surfaces reveal functional sites



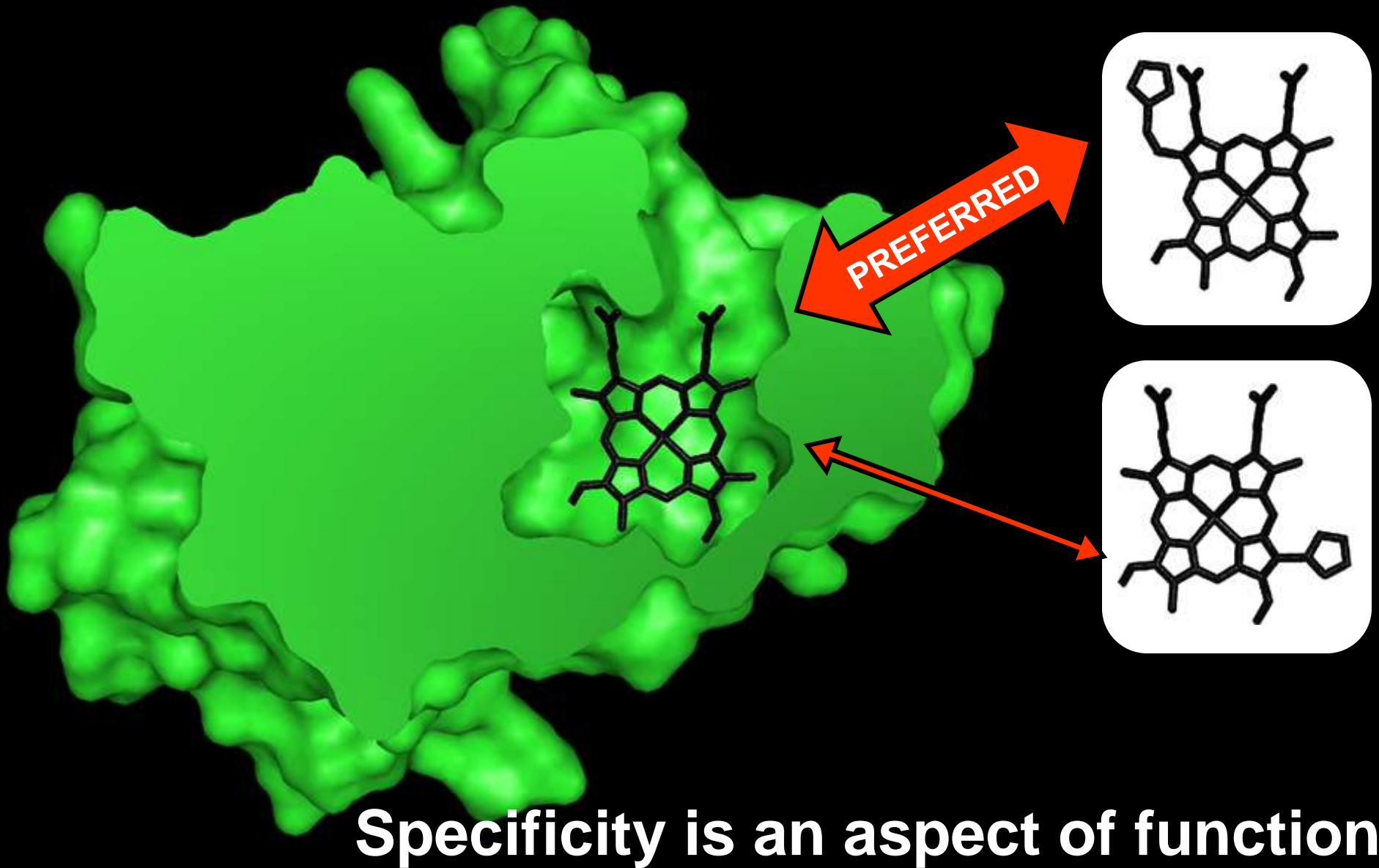
GRASP2	Petrey et al, <i>Methods Enzymol.</i> 2003.
SURFNET	Glaser et al, <i>Proteins</i> , 2006.
CASTp	Dundas et al, <i>Nucleic Acids Res.</i> 2006.
SCREEN	Nayal et al, <i>Proteins</i> , 2006
eF-seek	Kinoshita et al, <i>Nucleic Acids Res.</i> 2007
APROPOS	Peters et al, <i>J. Mol. Biol.</i> , 1996

Similarity doesn't tell us everything

How does this protein fit in the system?

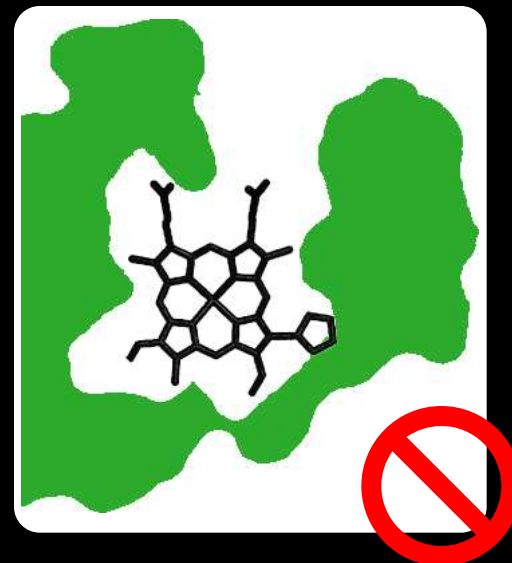
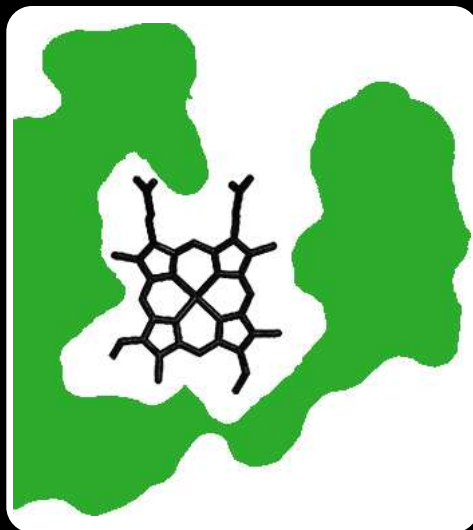
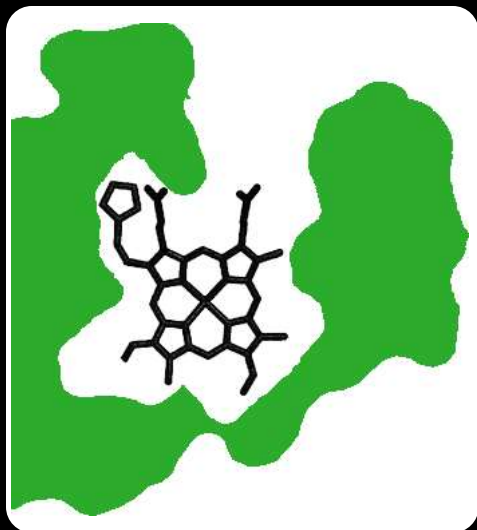
What parts of the protein make it work?

Specificity is preferential binding

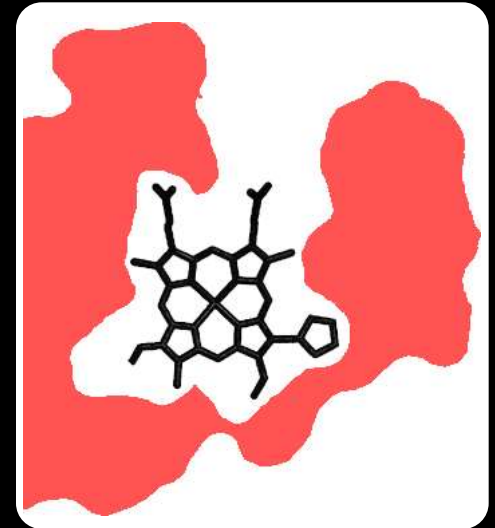
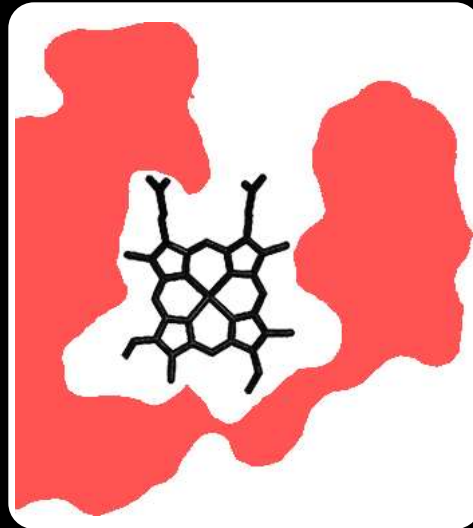
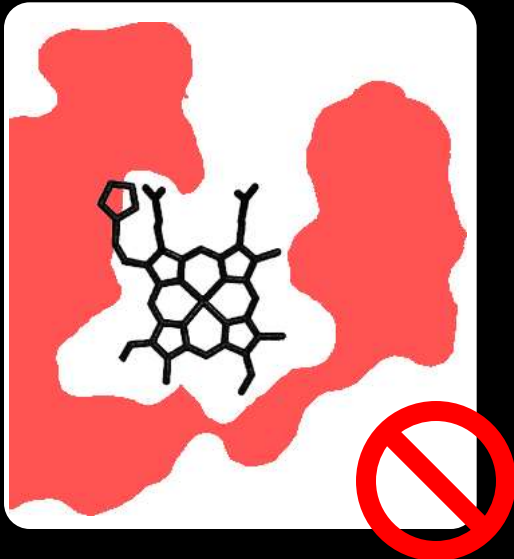
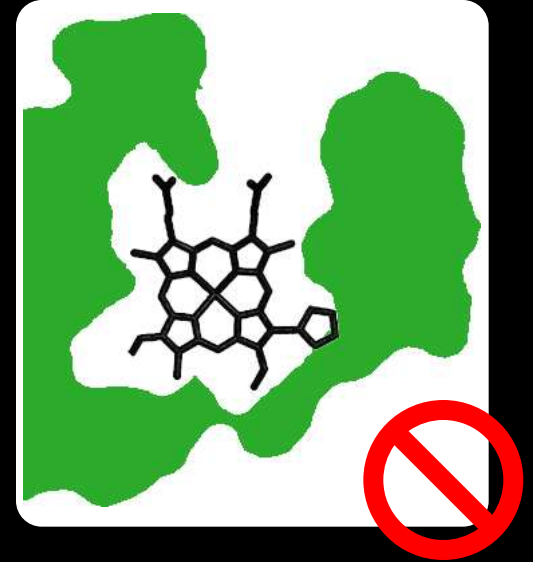
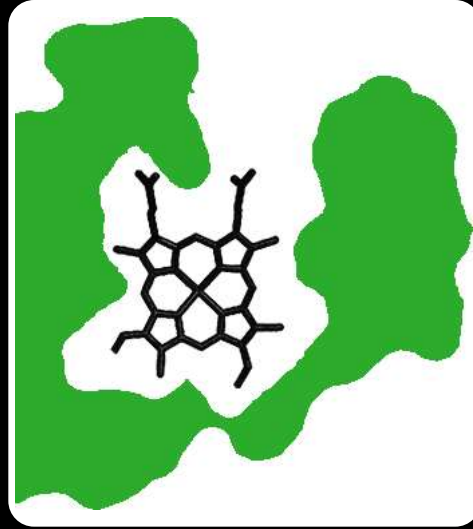
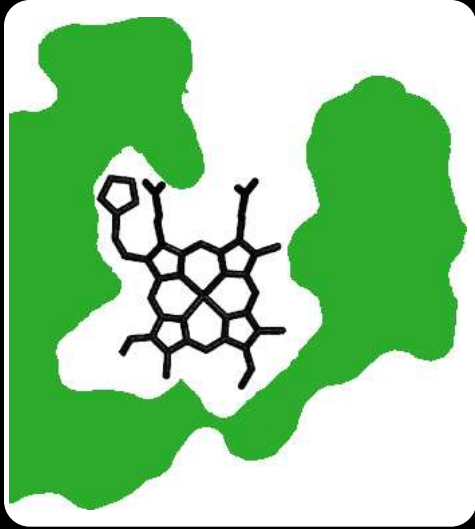


Specificity is an aspect of function

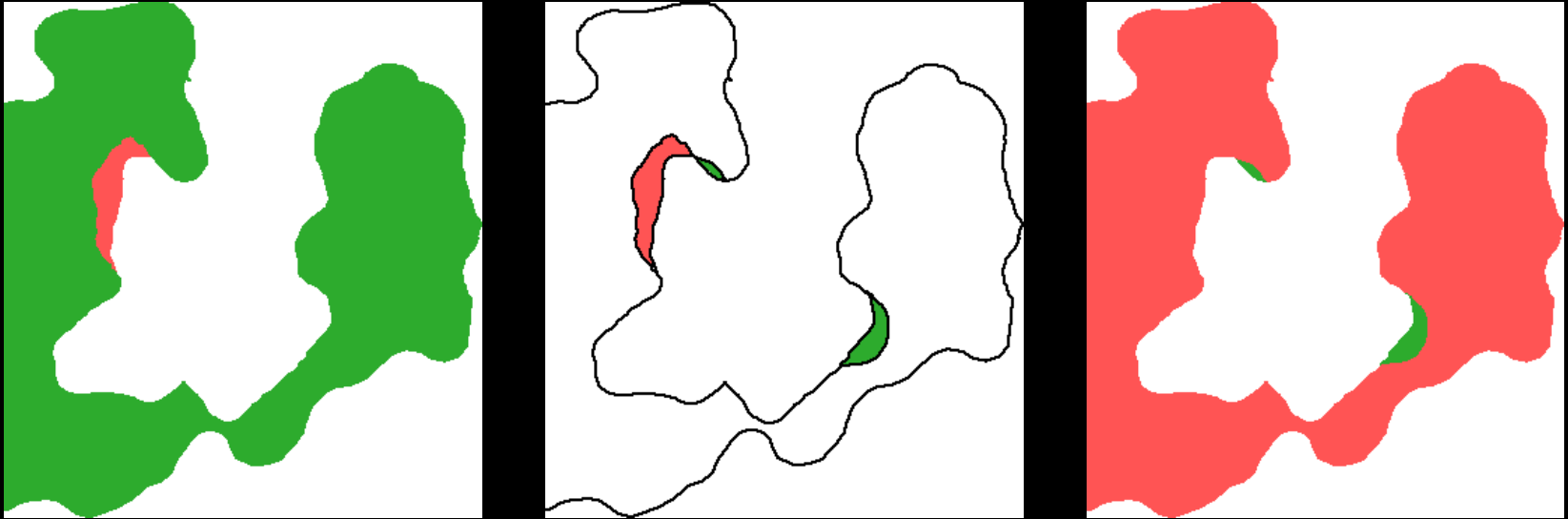
Cavity shape influences specificity



Proteins with the same function can have different specificity



VASP isolates differences in cavity shape

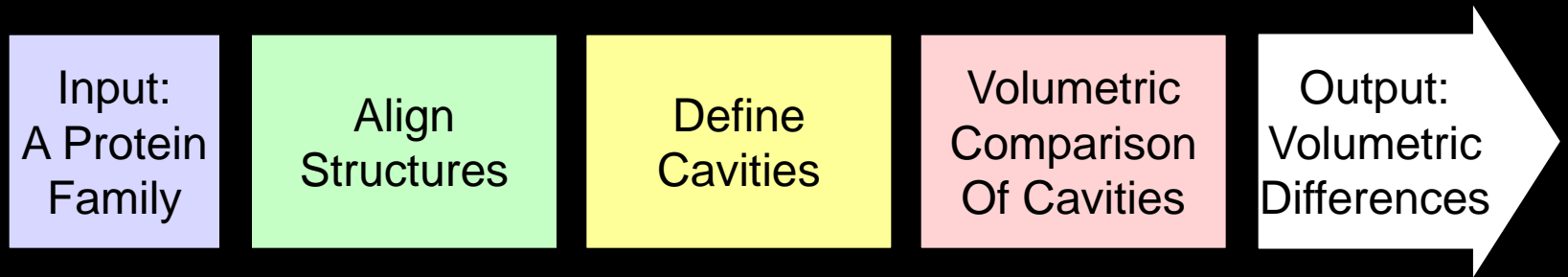


VASP: Volumetric Analysis of the Surfaces of Proteins

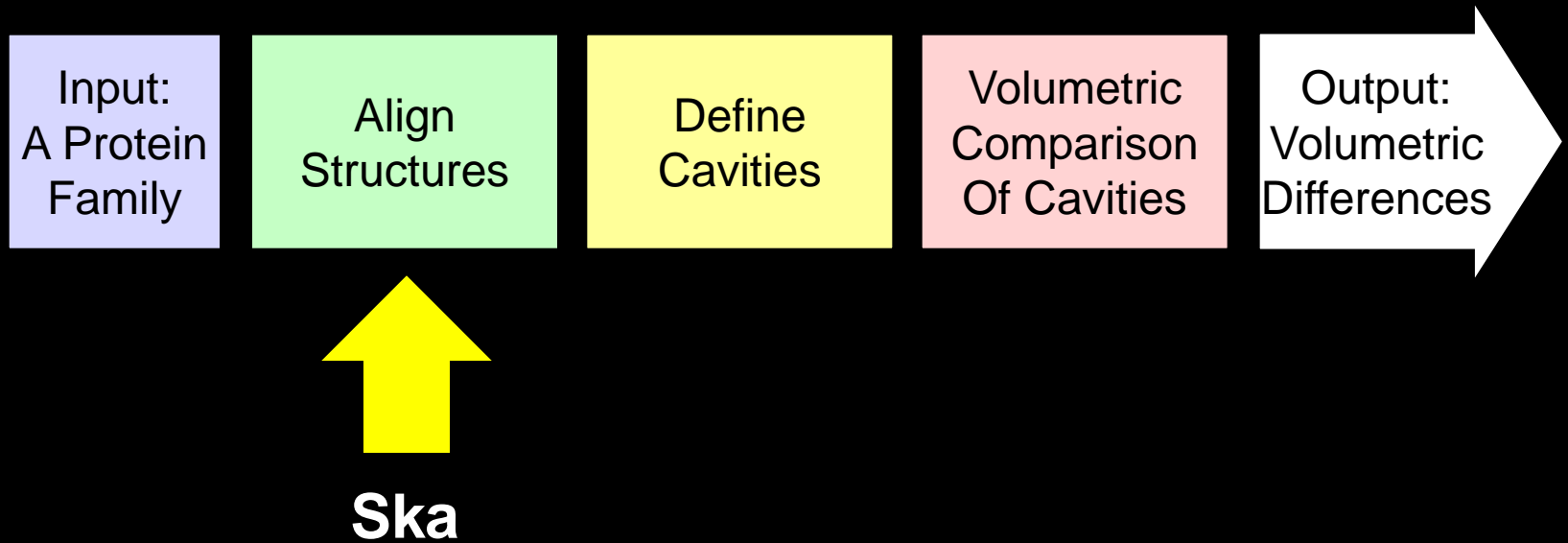
- Identify amino acids that alter cavity shape
- Identify subcavities that alter cavity shape

Results: VASP finds influences on specificity

The VASP procedure

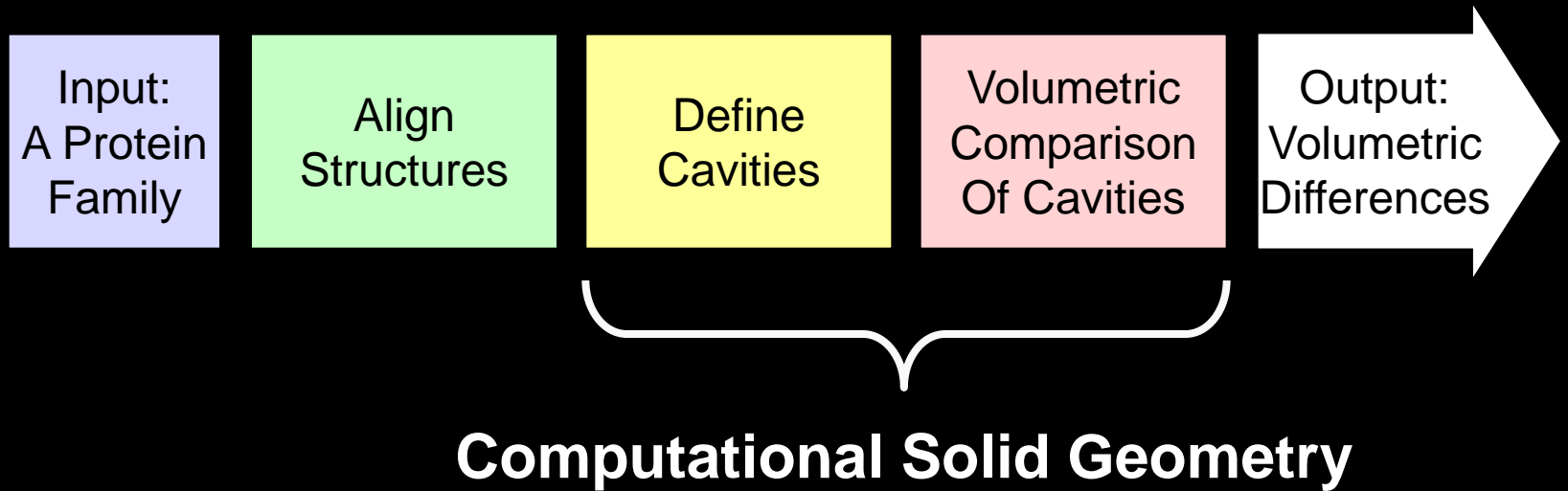


The VASP procedure

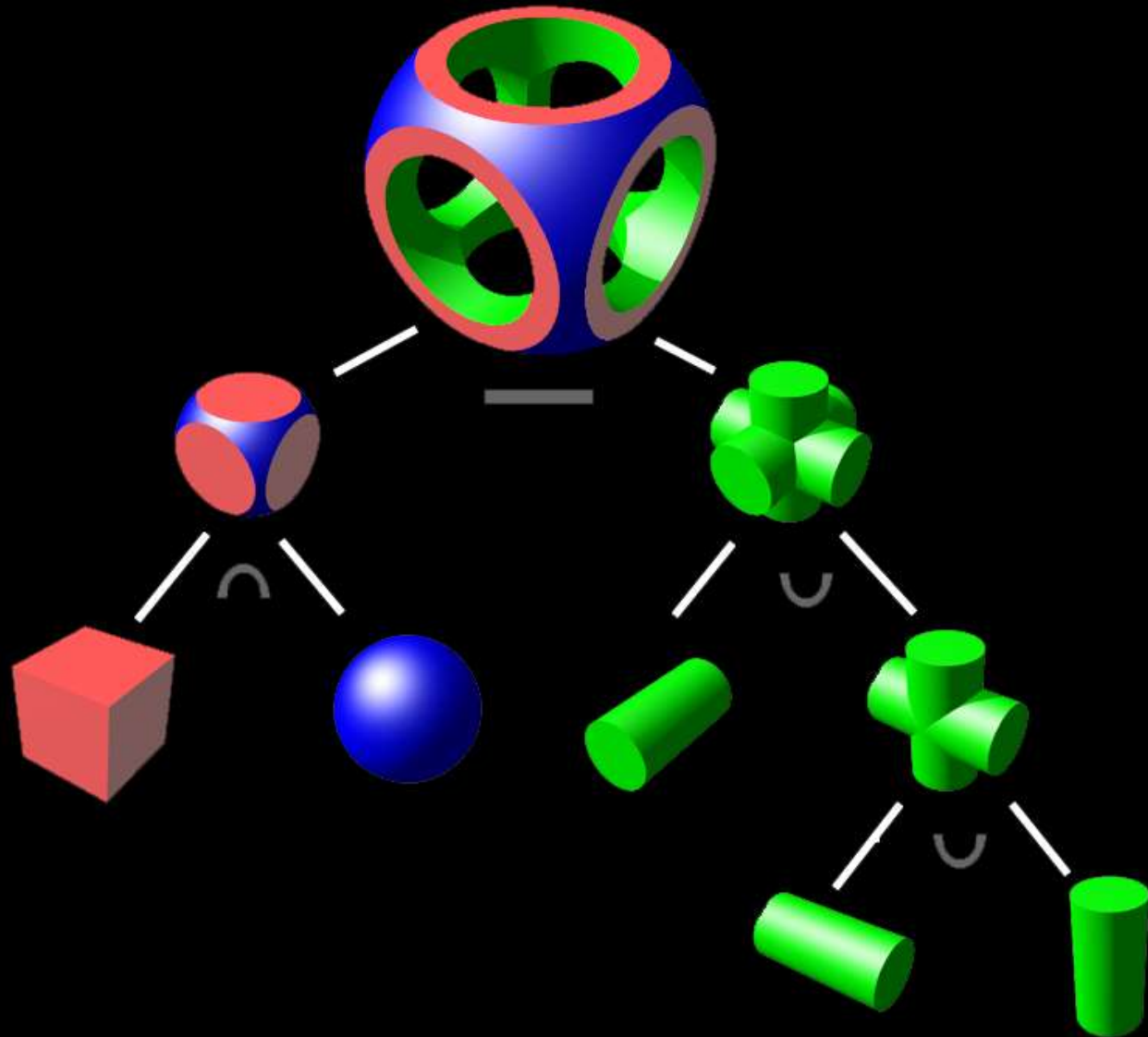


Petrey D, Honig B. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol.* 374:492-509. 2003.

The VASP procedure



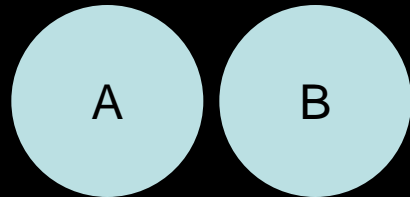
Computational Solid Geometry (CSG)



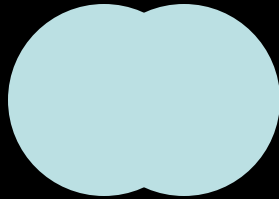
CSG was originally for modeling parts



Computational Solid Geometry (CSG)



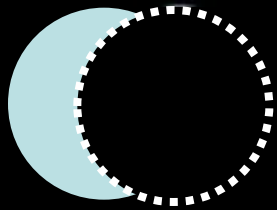
**Boolean Set
Operations**



Union

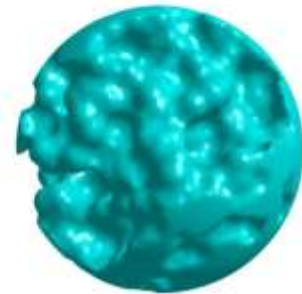
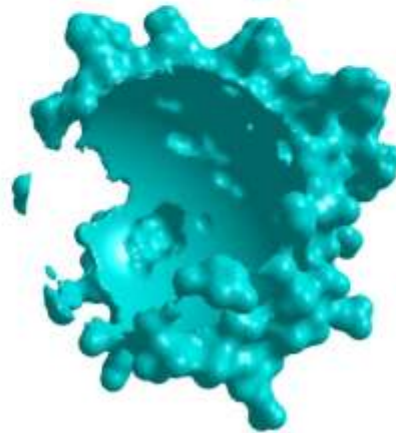
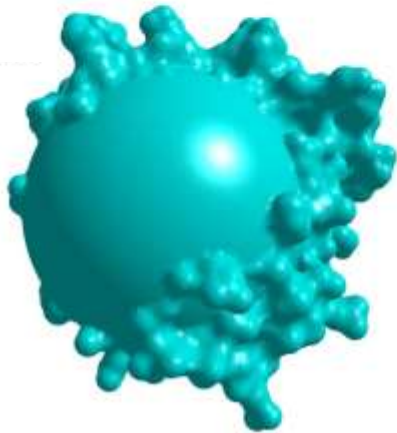
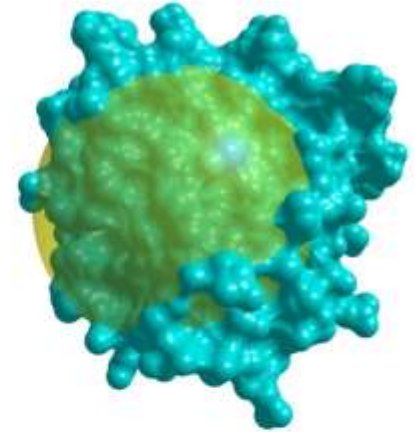
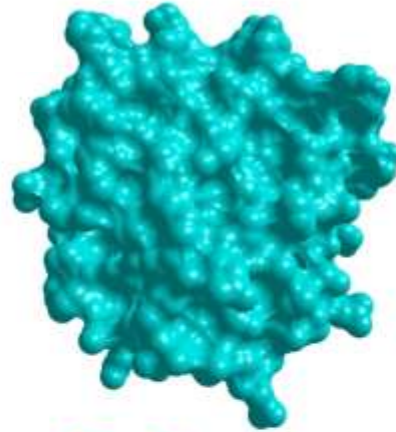


Intersection

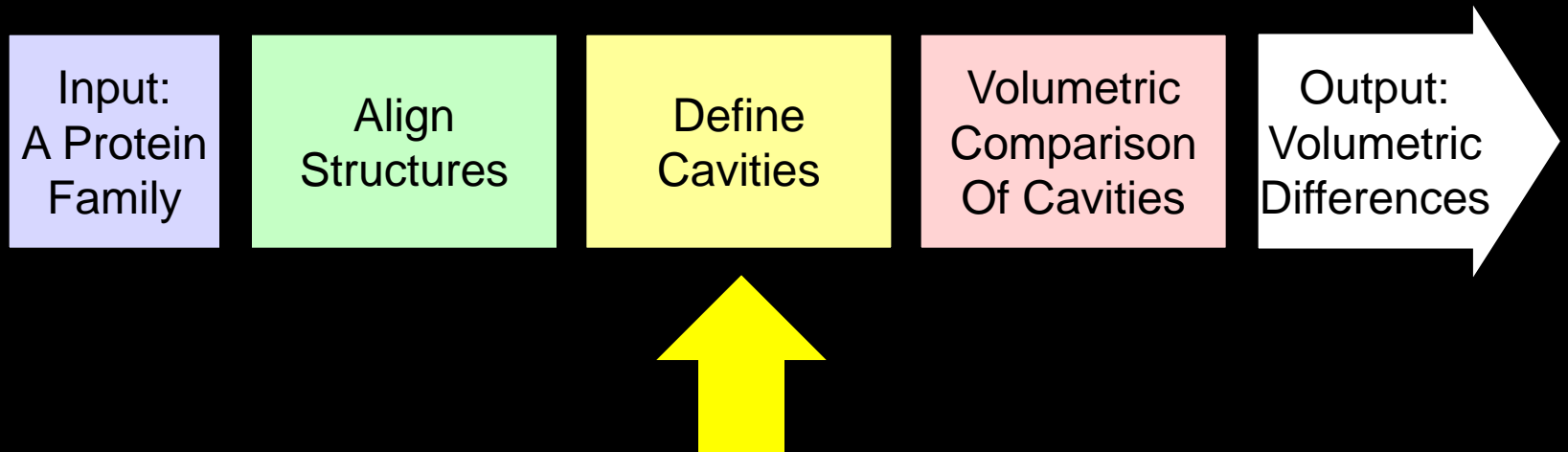


Difference

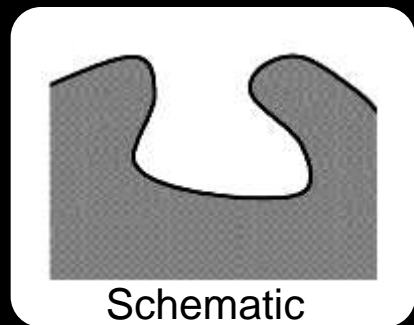
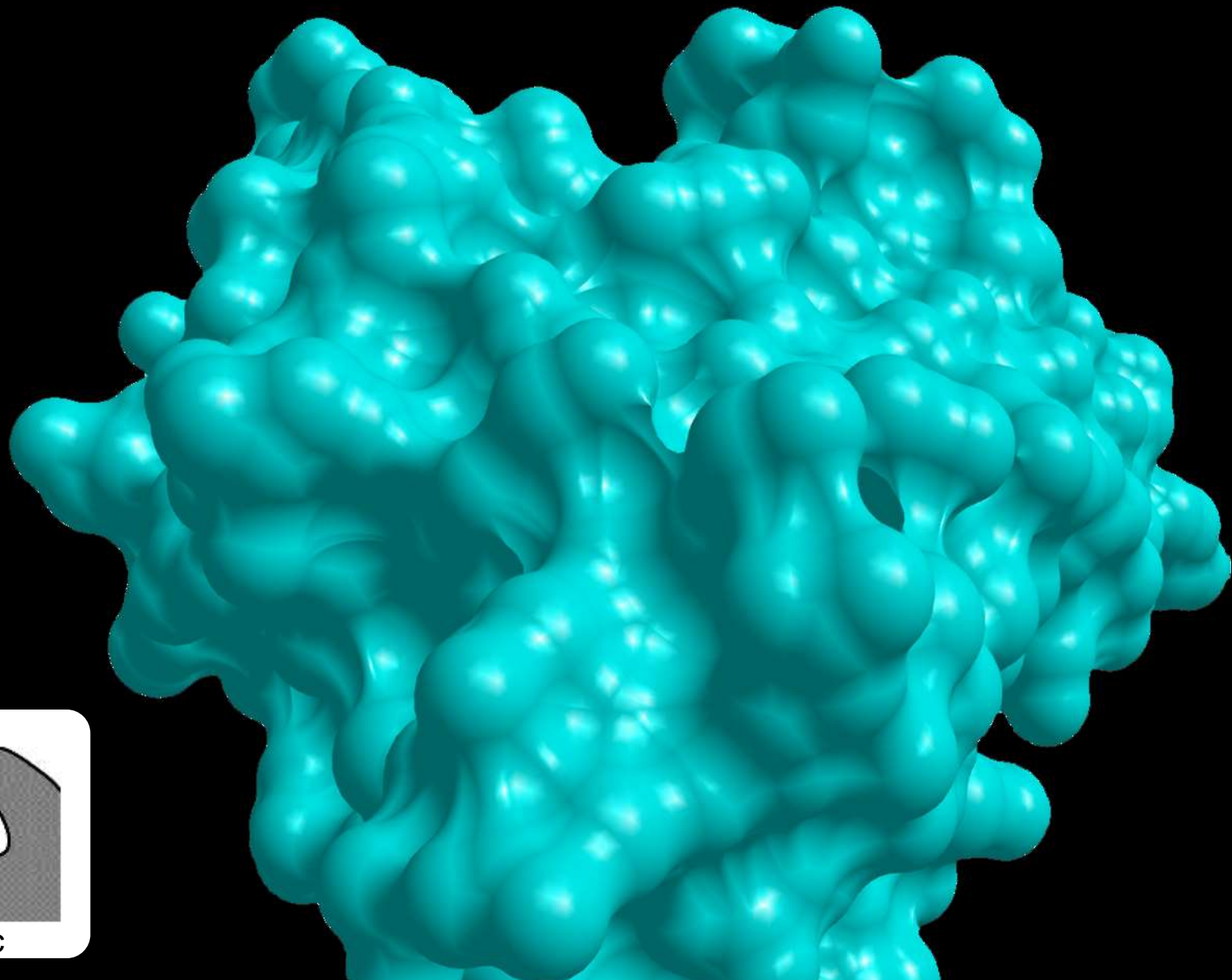
Using CSG with protein structures



The VASP procedure

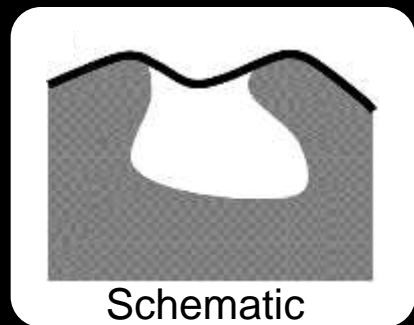
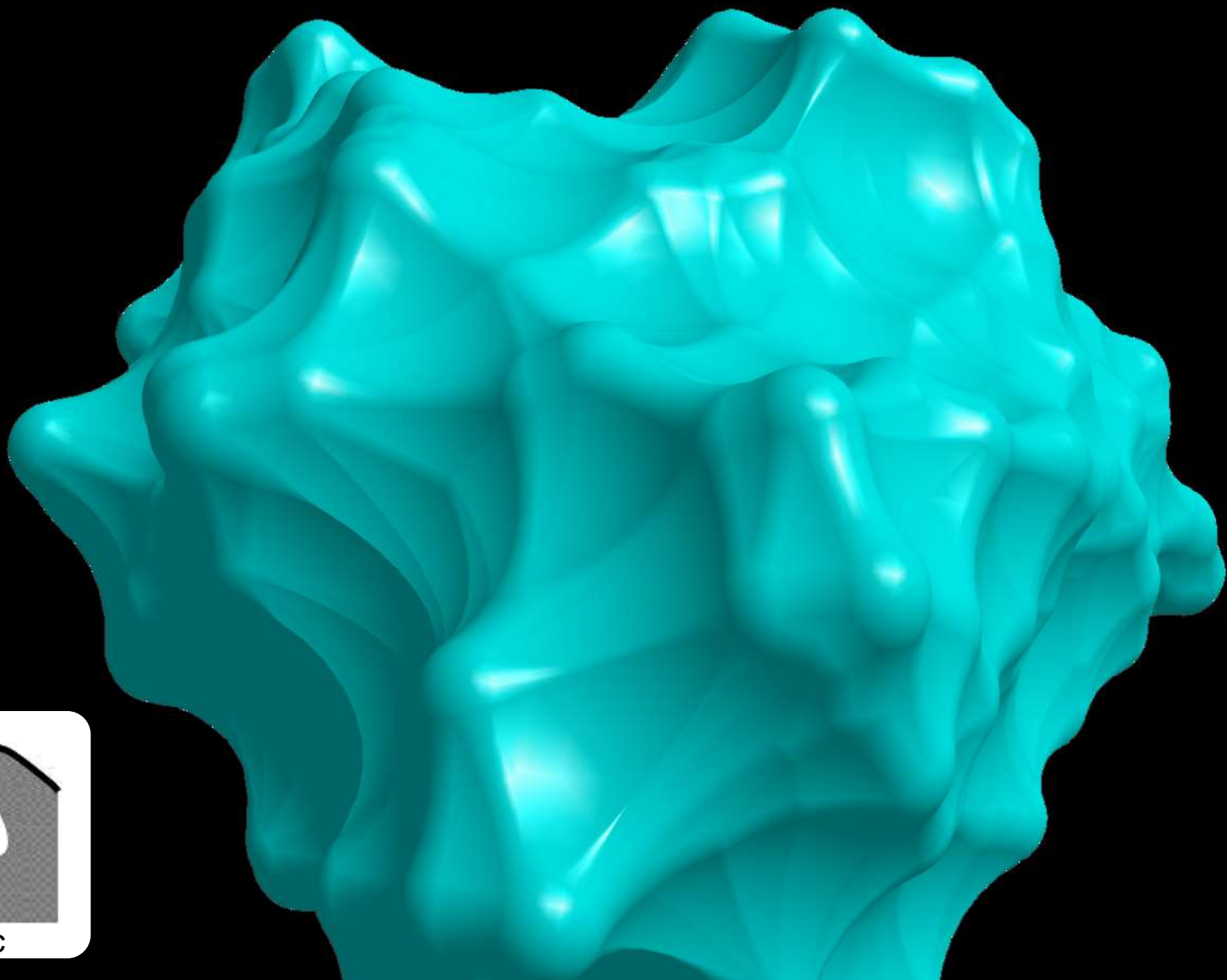


Begin with the molecular surface



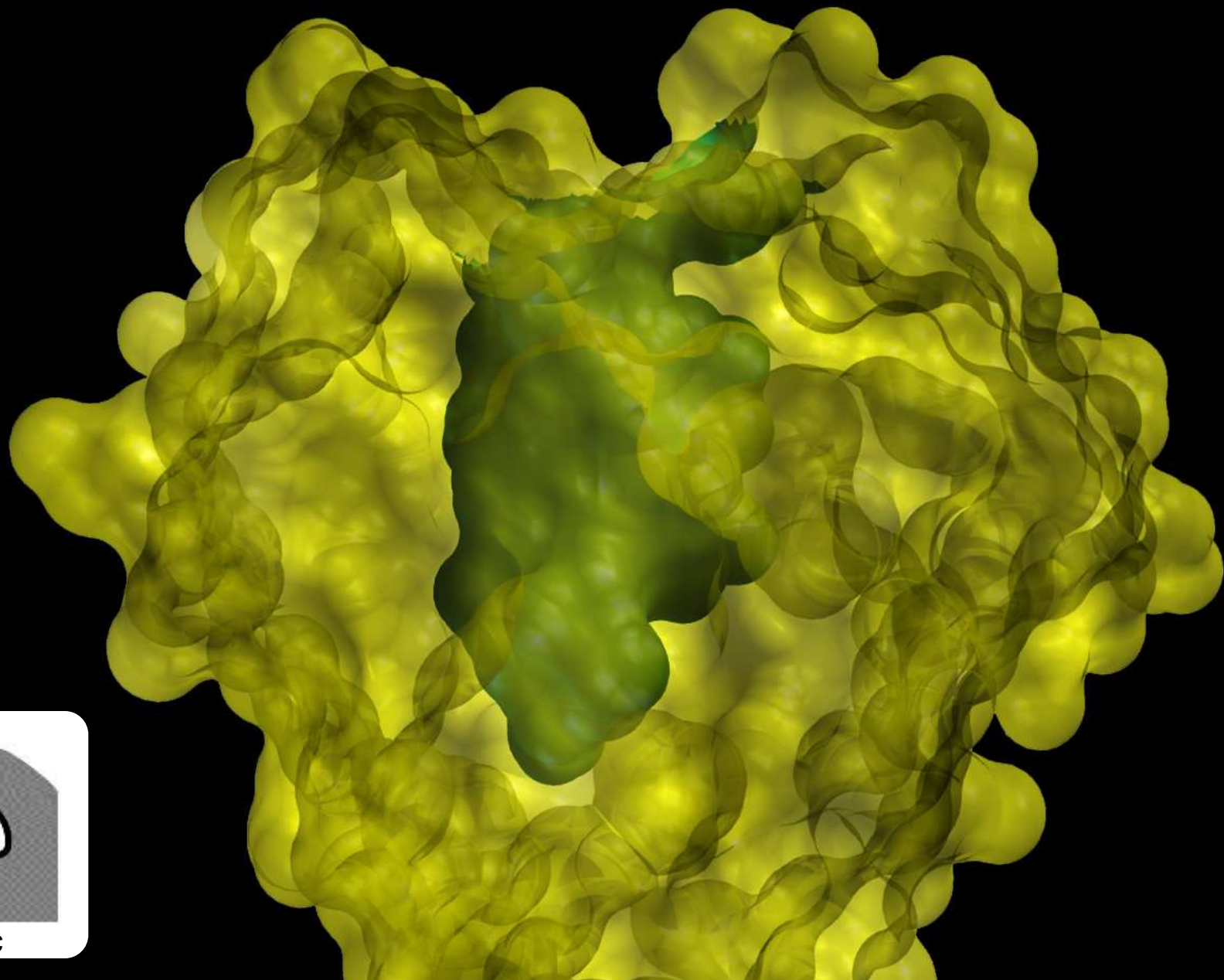
Schematic

Compute an envelope surface



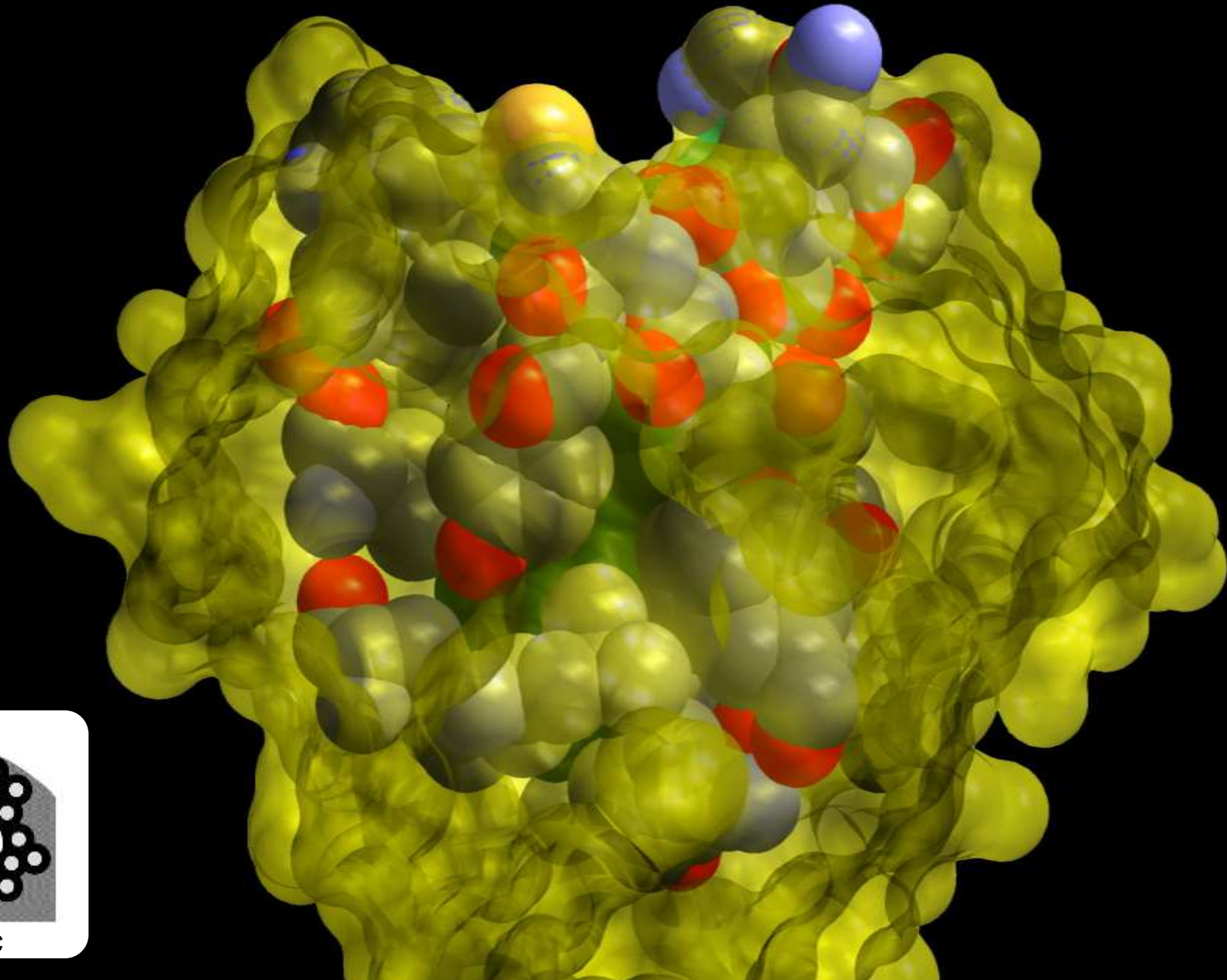
Schematic

Find the interior surface



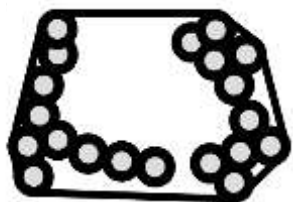
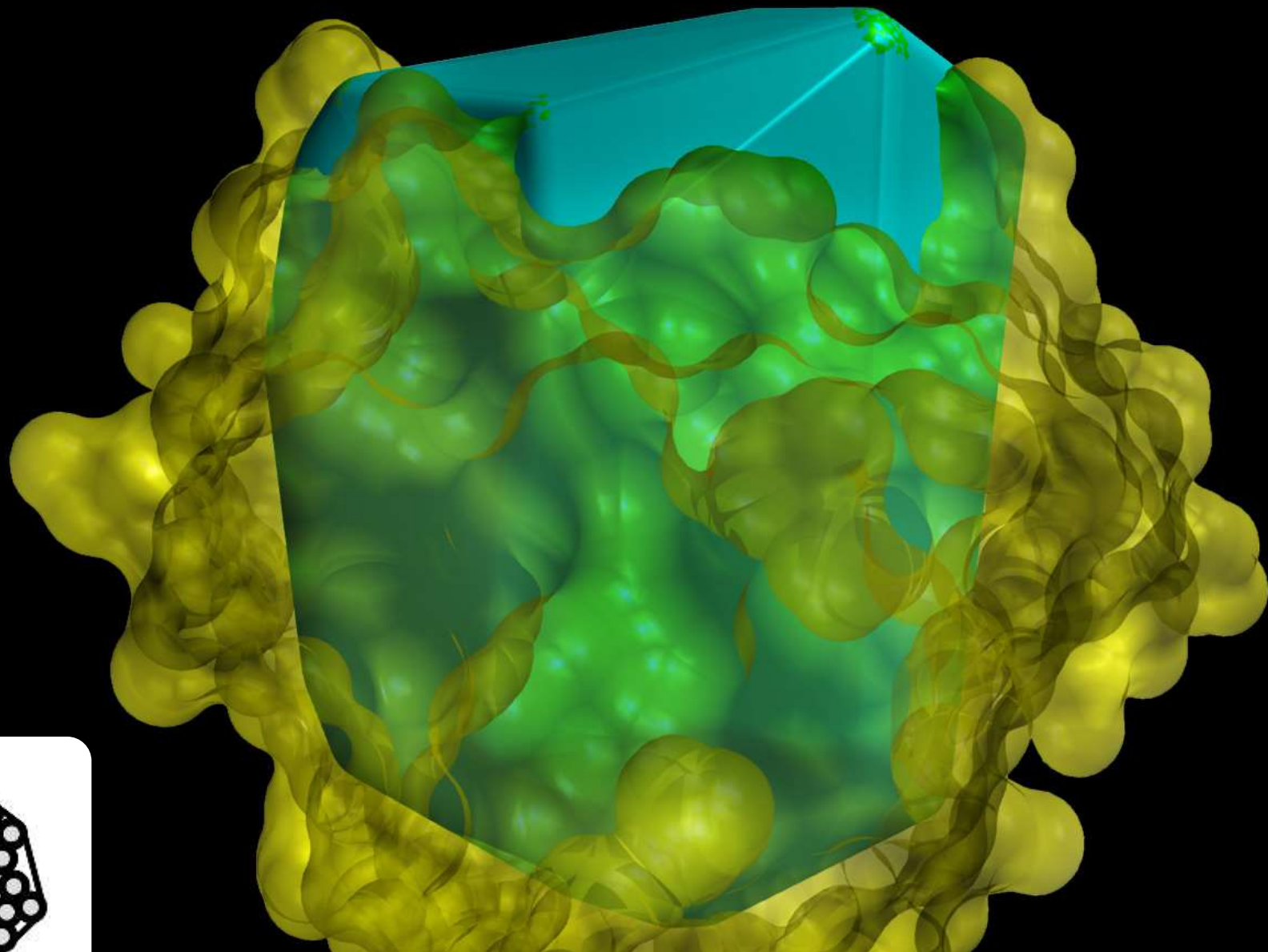
Schematic

Identify nearby amino acids



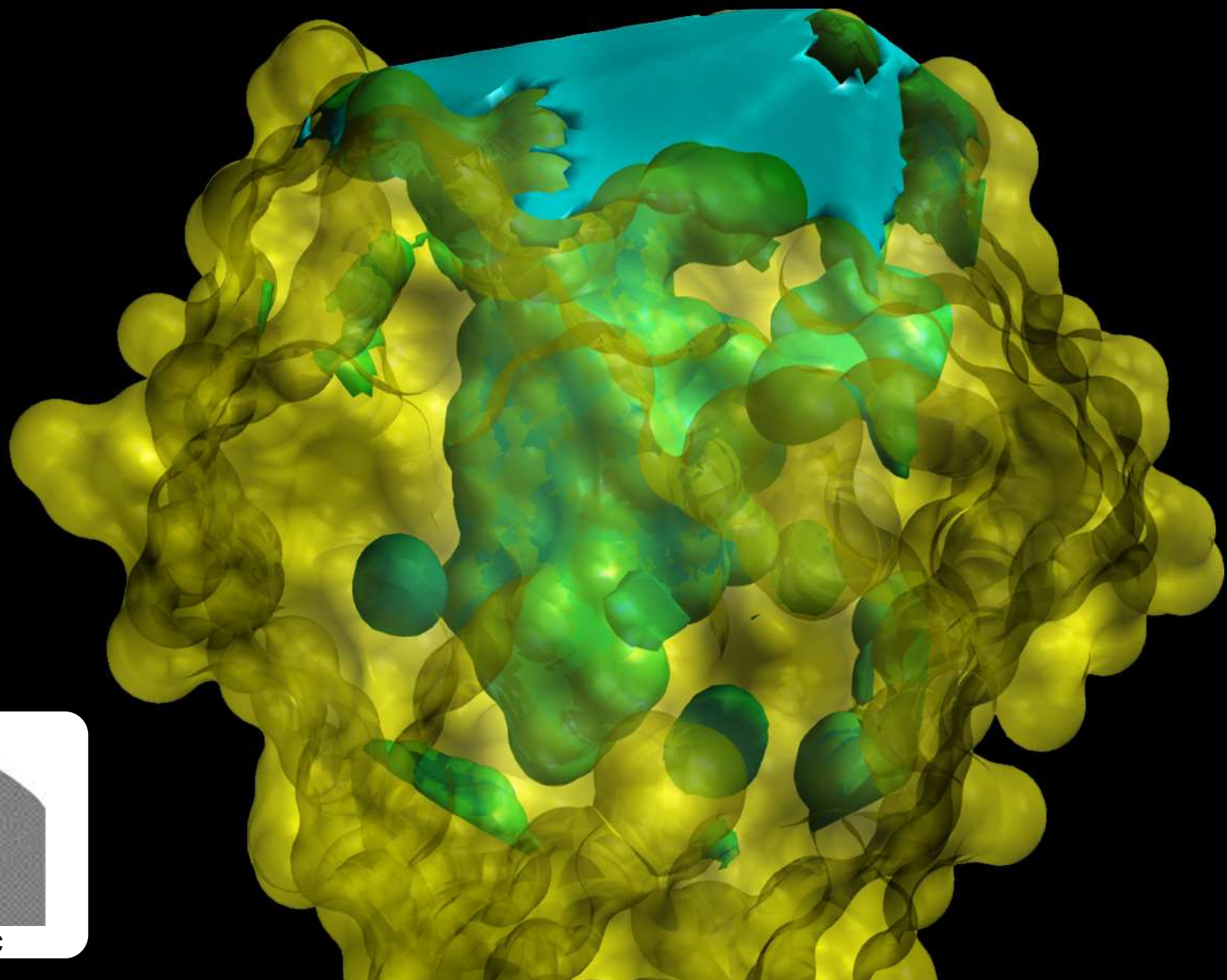
Schematic

Compute the convex hull



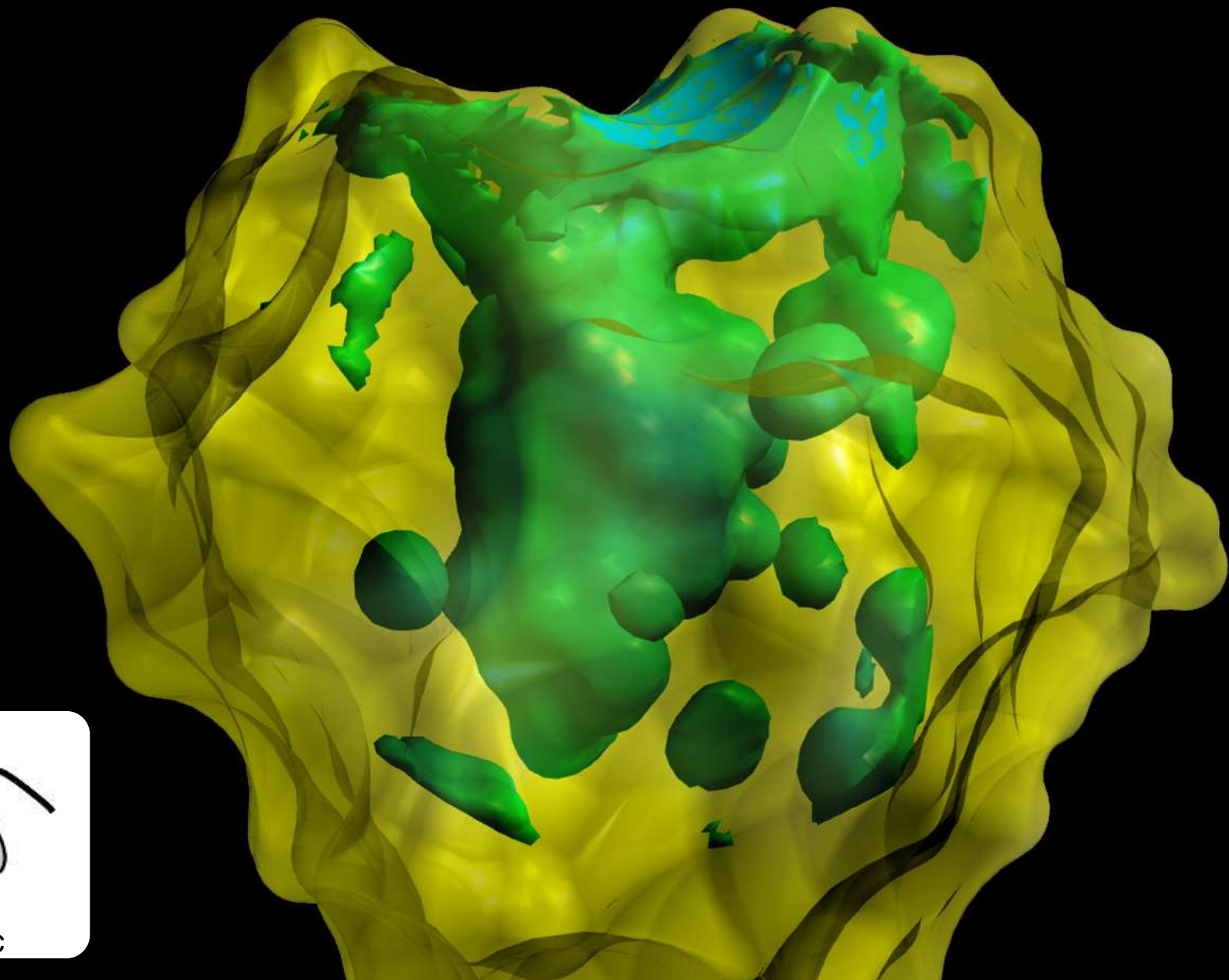
Schematic

CSG hull minus molecular surface



Schematic

CSG intersection with the envelope surface



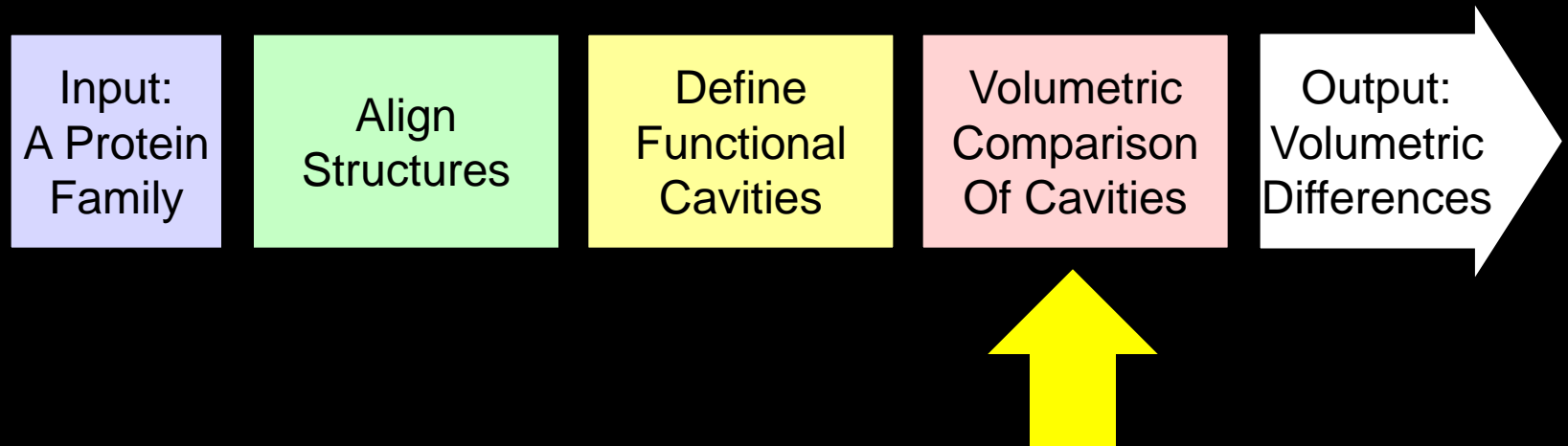
Schematic

Remove disconnected pieces



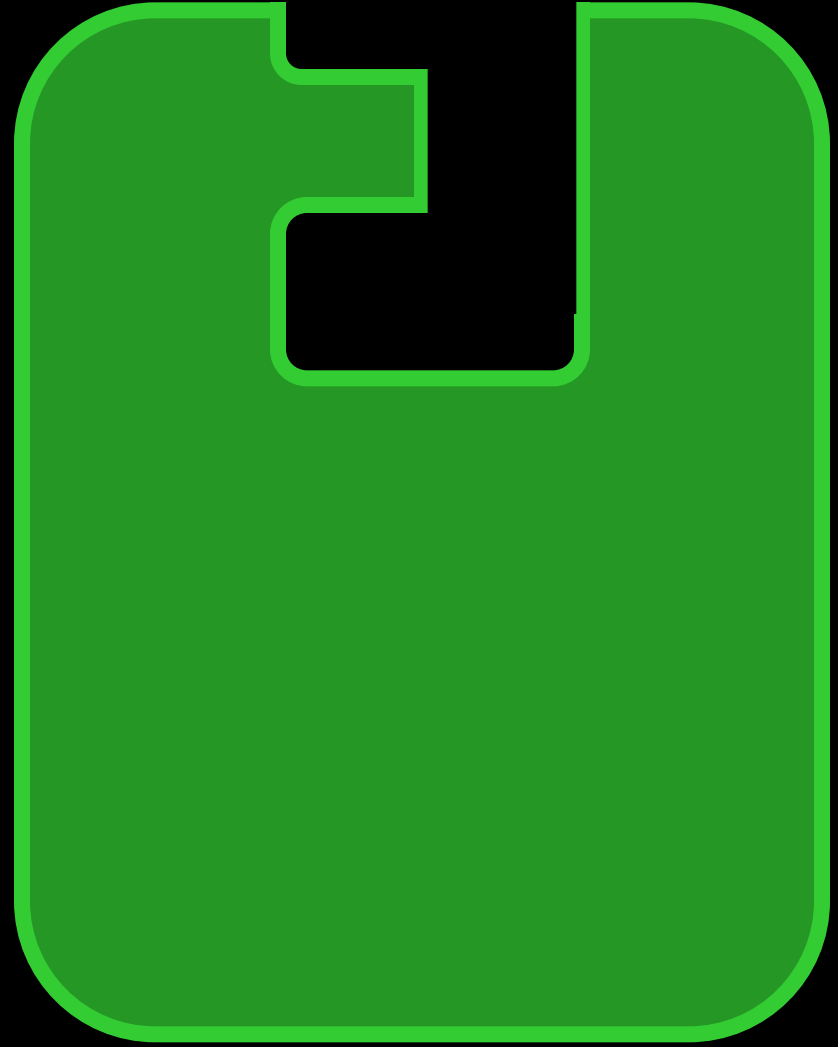
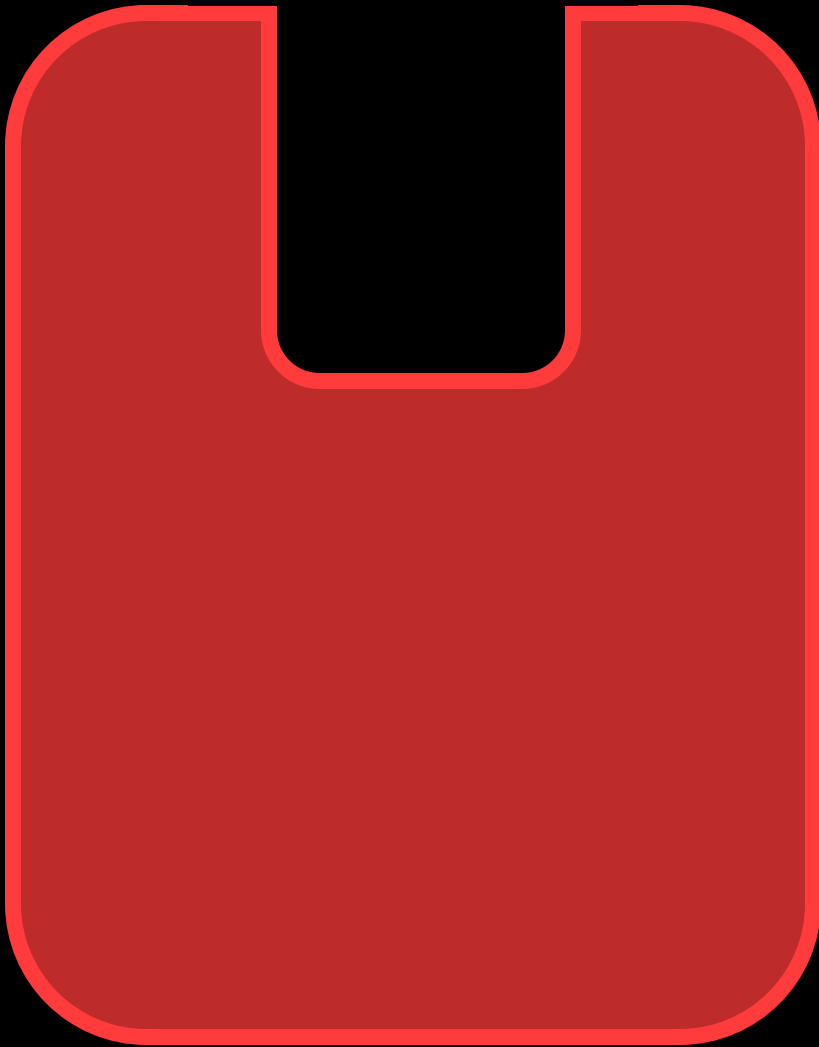
Schematic

The VASP procedure

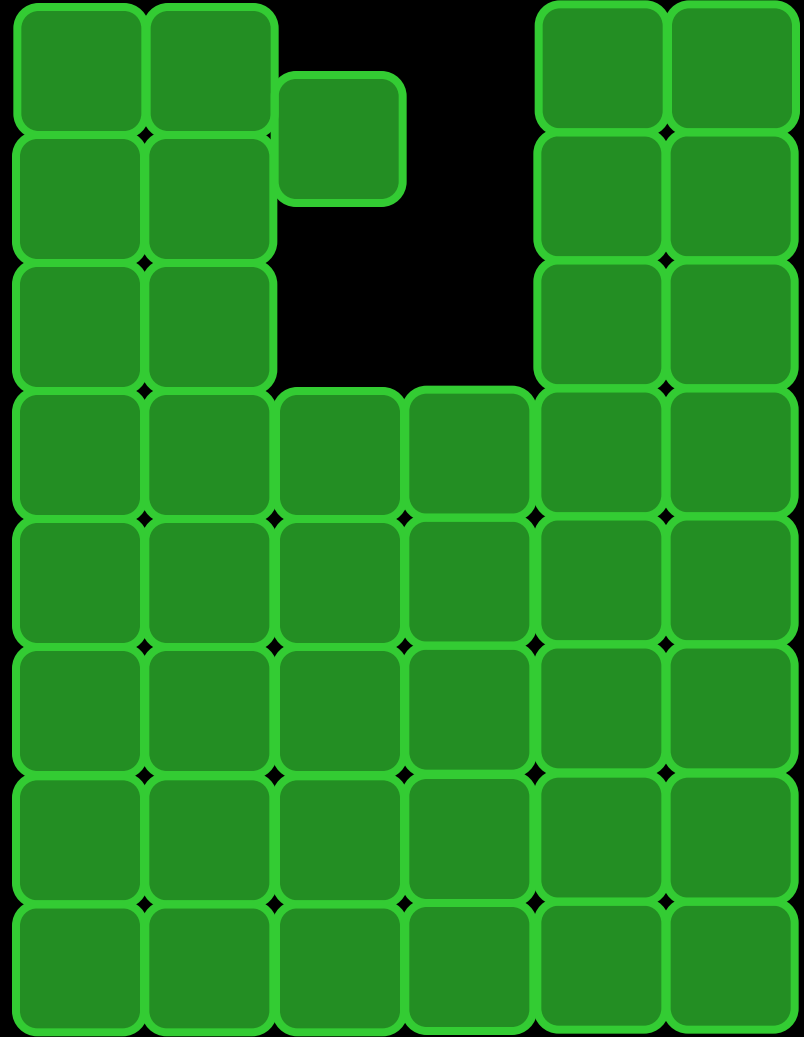
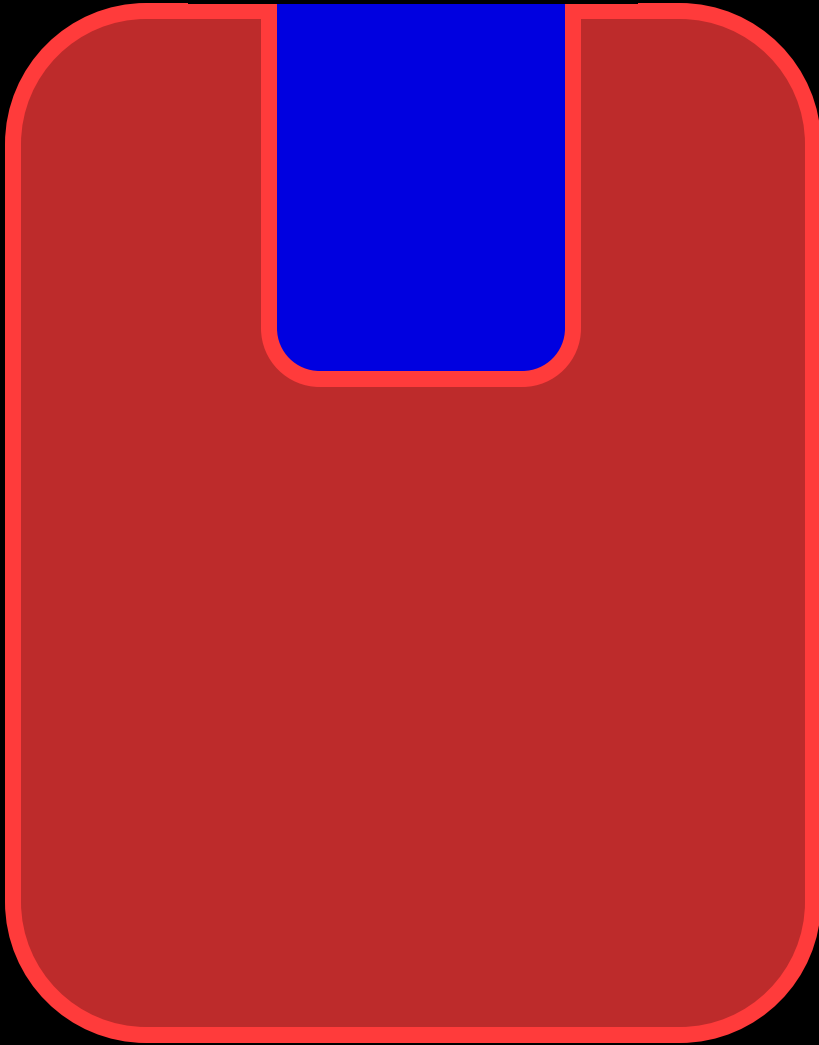


- Amino Acids affecting cavity shape
- Subcavities affecting cavity shape

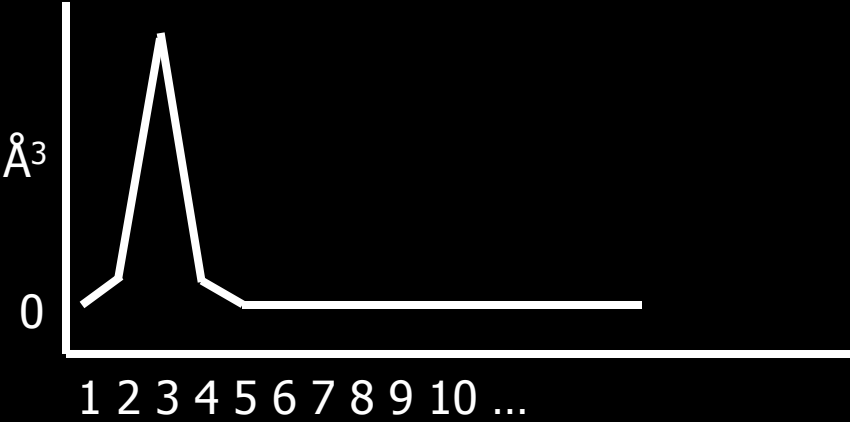
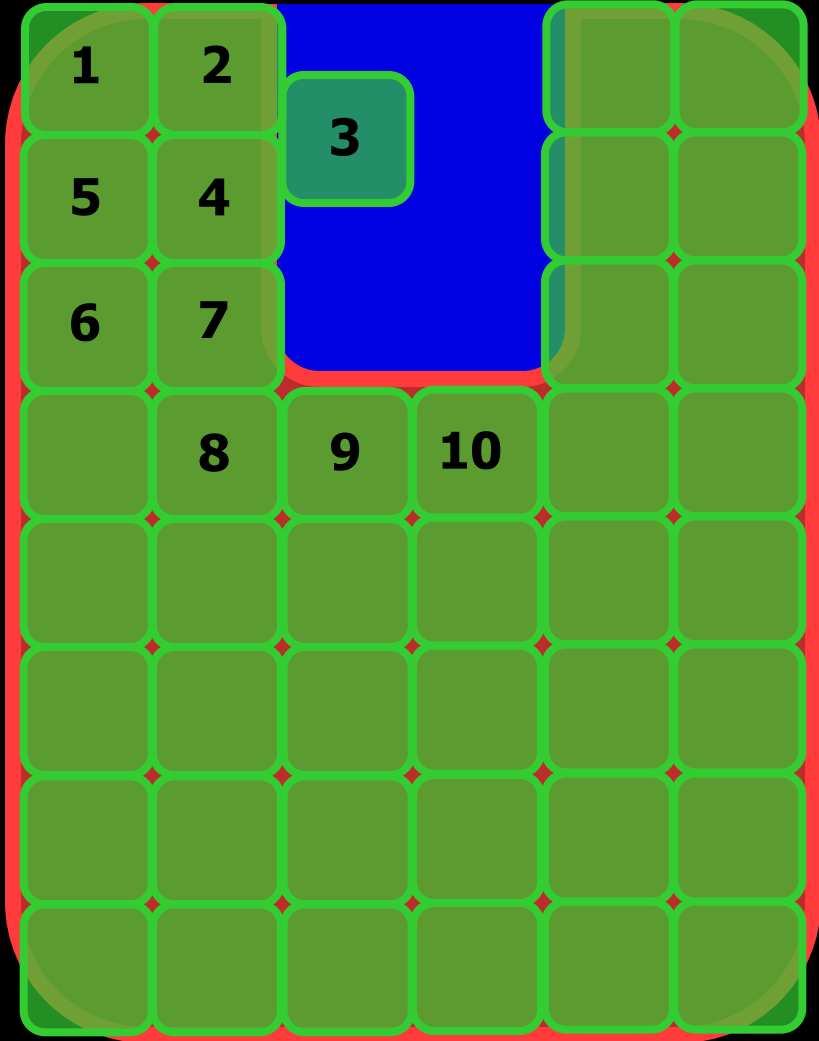
Finding amino acids that affect cavity shape



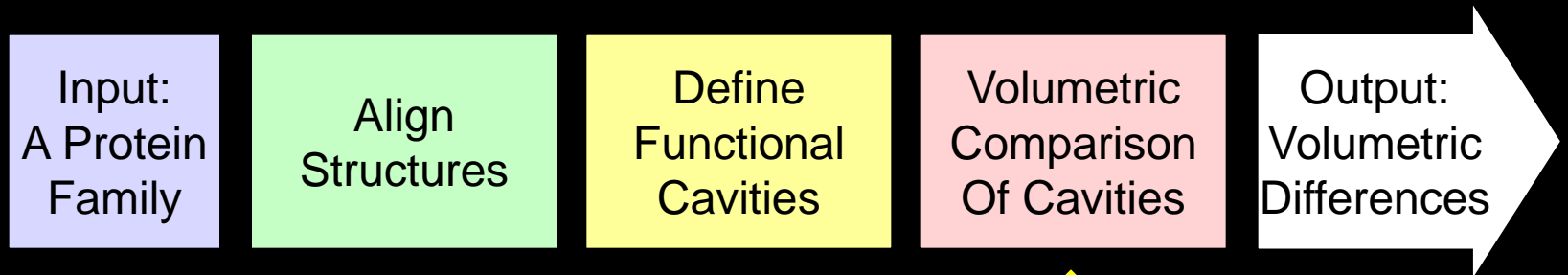
Finding amino acids that affect cavity shape



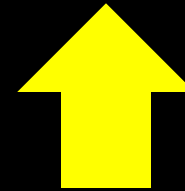
Finding amino acids that affect cavity shape



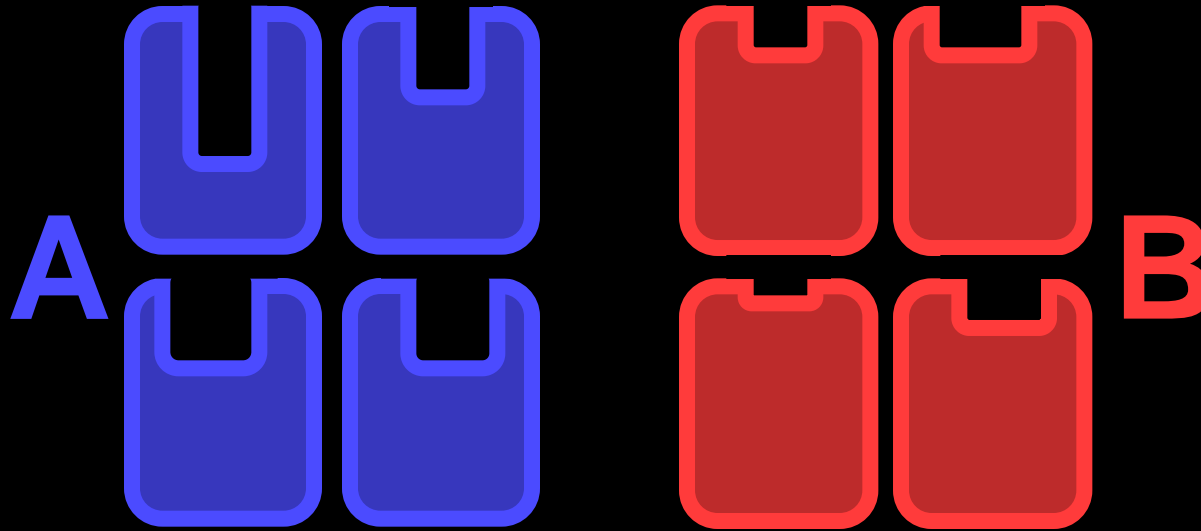
The VASP procedure



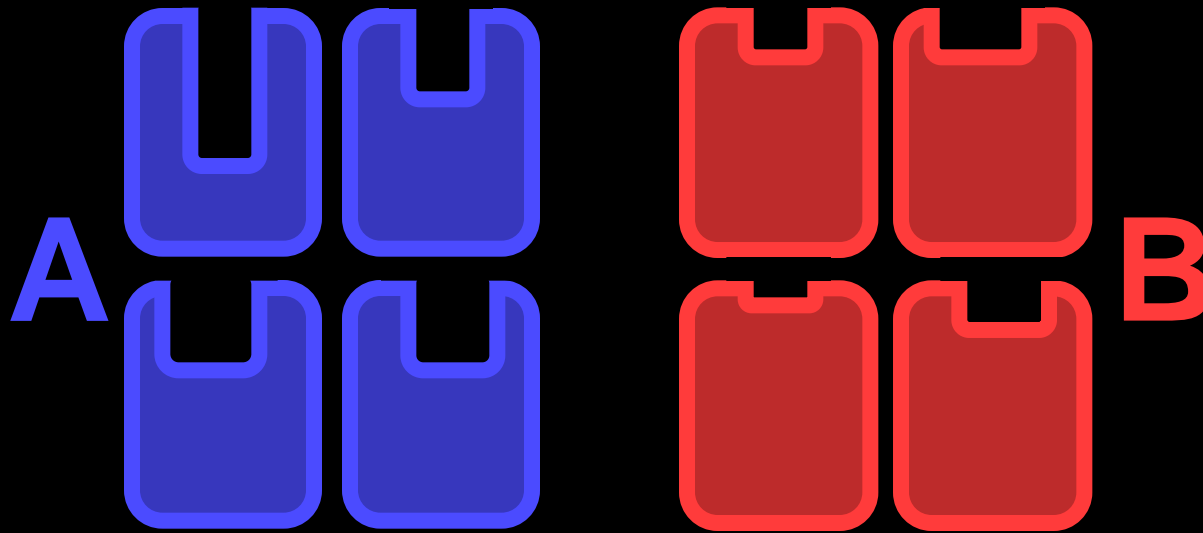
- Amino Acids affecting cavity shape
- Subcavities affecting cavity shape



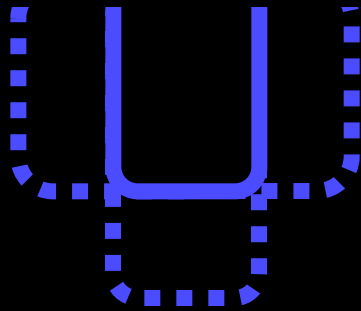
What makes A cavities different from B?



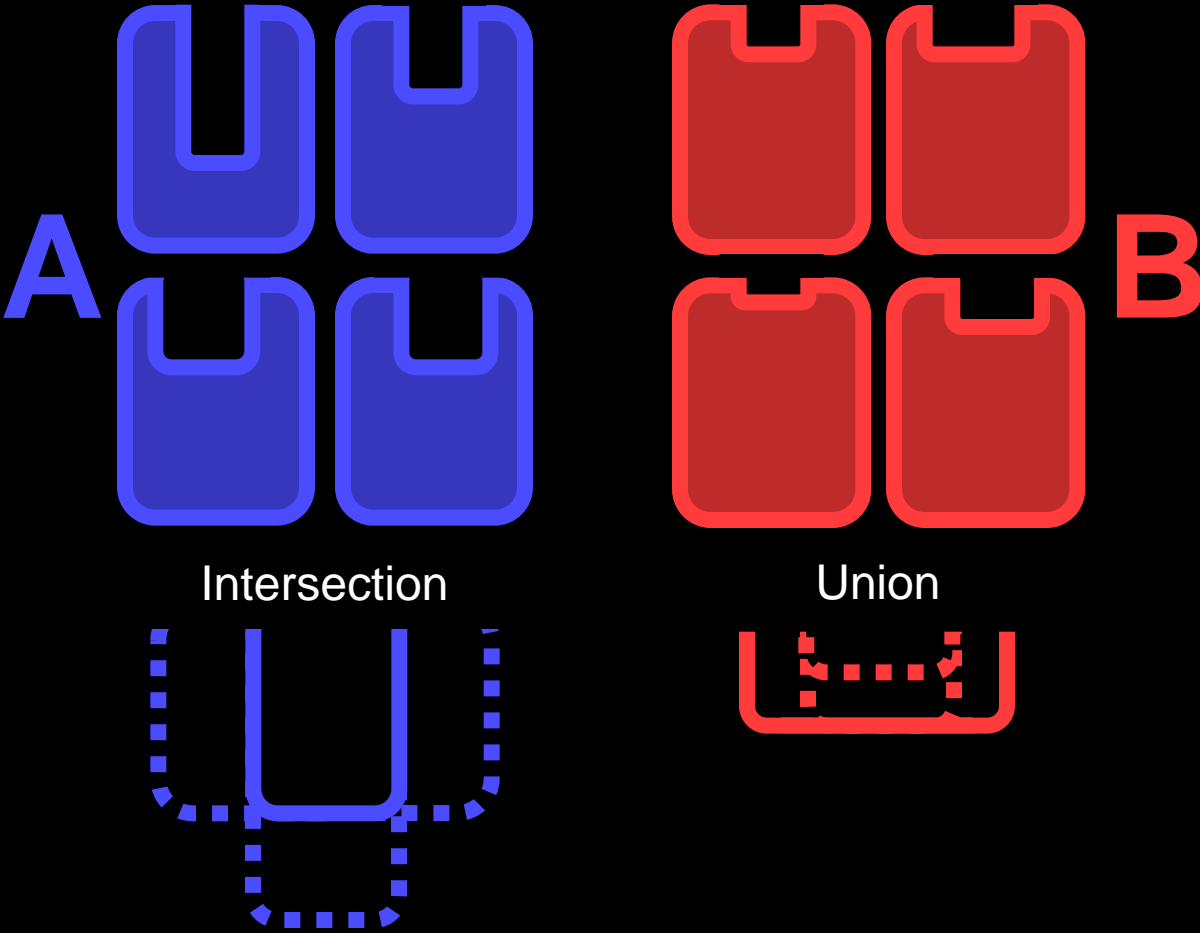
What is common in A?



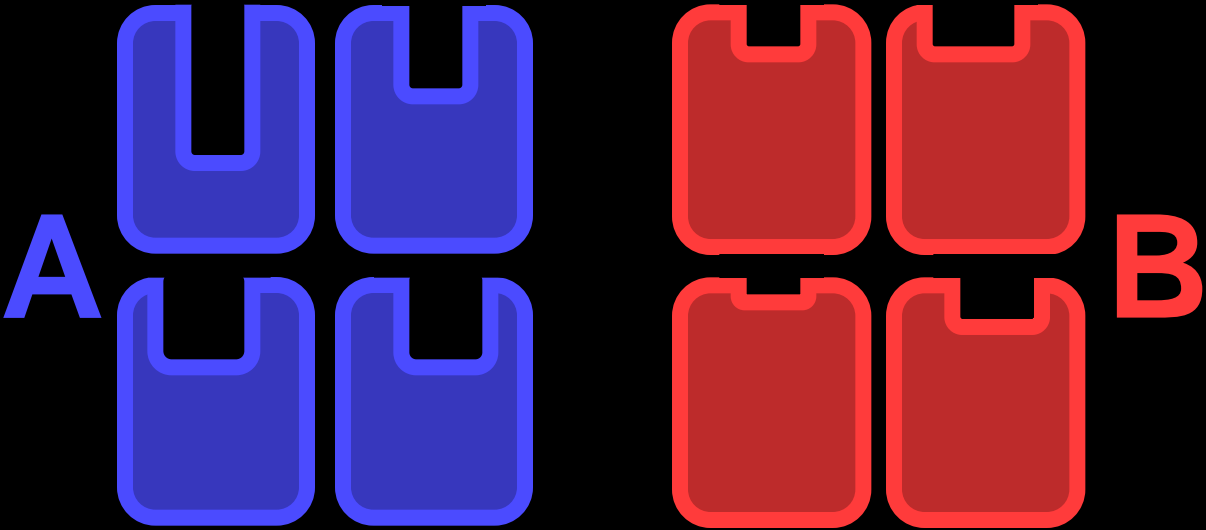
Intersection



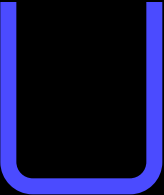
What is the maximum extent of B?



All parts of A that are not in any part of B



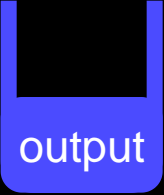
Intersection



Union



Difference



Results

- **Serine Proteases: Same function, different specificity**
 - Trypsins
 - Elastases
 - Chymotrypsins
- **Experiments**
 - VASP identifies amino acids that influence specificity
 - VASP identifies subcavities that influence specificity

The serine protease family

Serine Proteases

Chymotrypsin Clan

Catalytic Triad: His-Asp-Ser

Chymotrypsins

Trypsins

Elastases

Subtilisin Clan (Asp-His-Ser)

Subtilisins

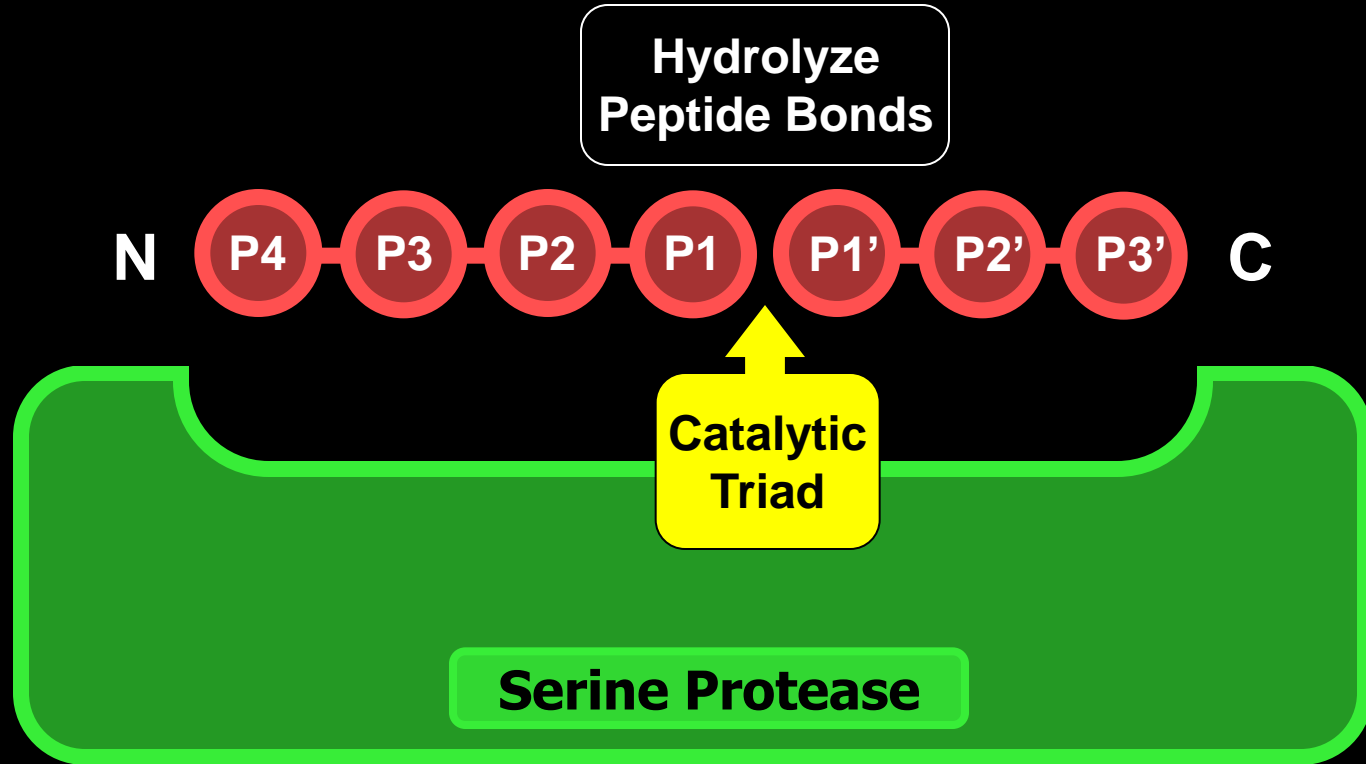
Other clans (not used)

Oligopeptidases (Asp-Ser-His)

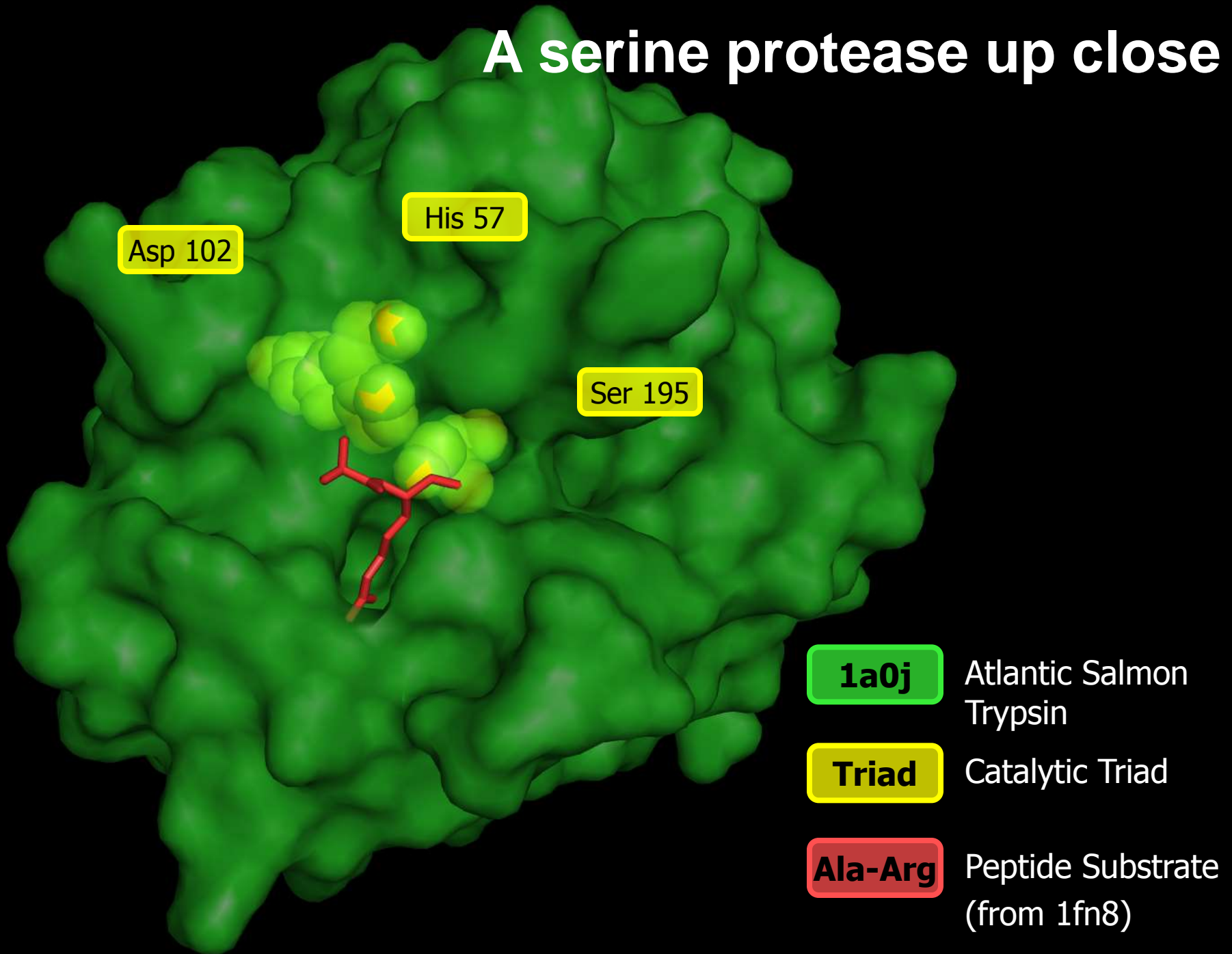
Carboxypeptidases (Ser-Asp-His)

Others..

Serine proteases break up other proteins

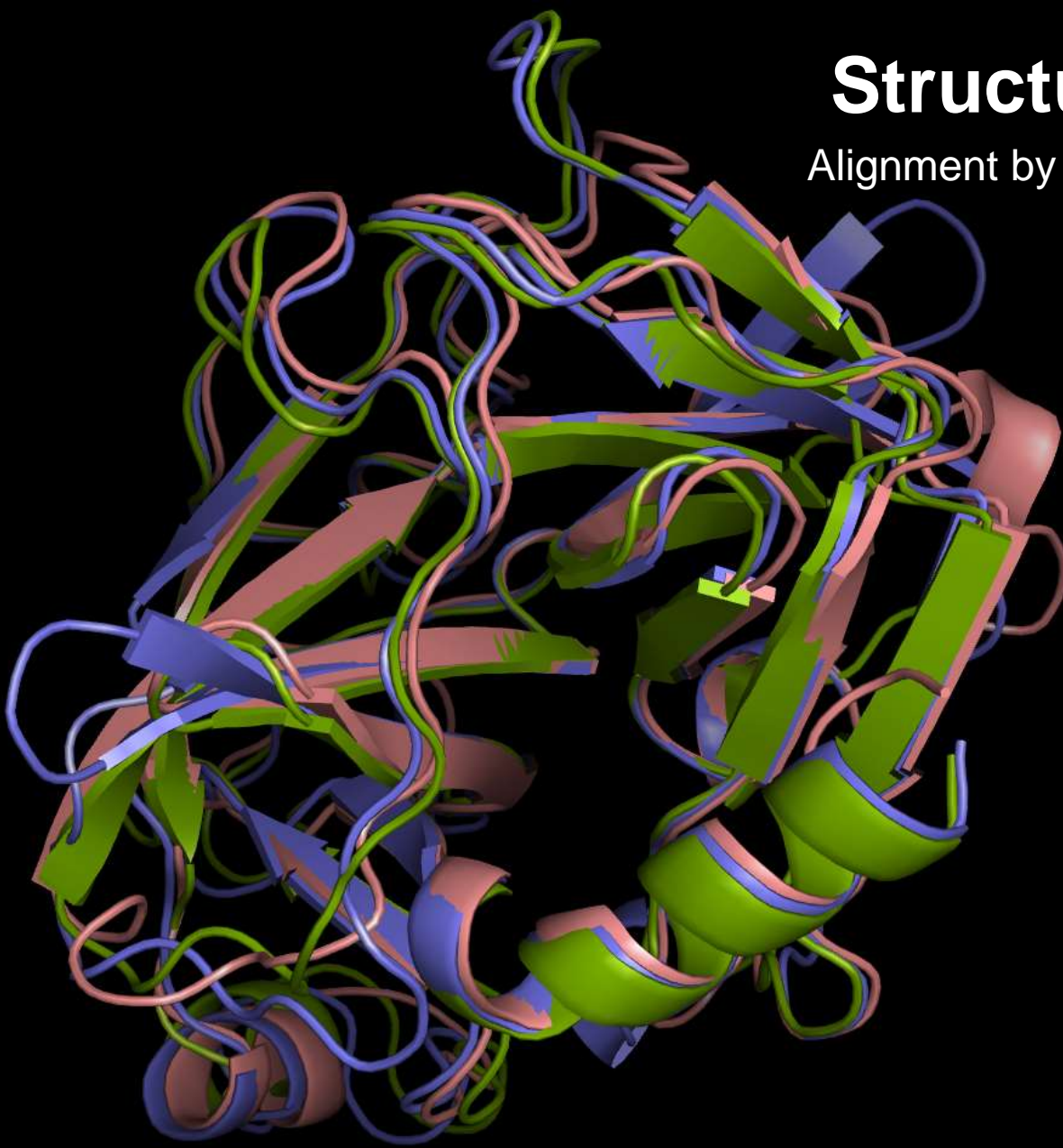


A serine protease up close



Structural Alignment

Alignment by Catalytic triad + S1 residue
(C α and C β atoms)



Chymotrypsins

3.4.21.1

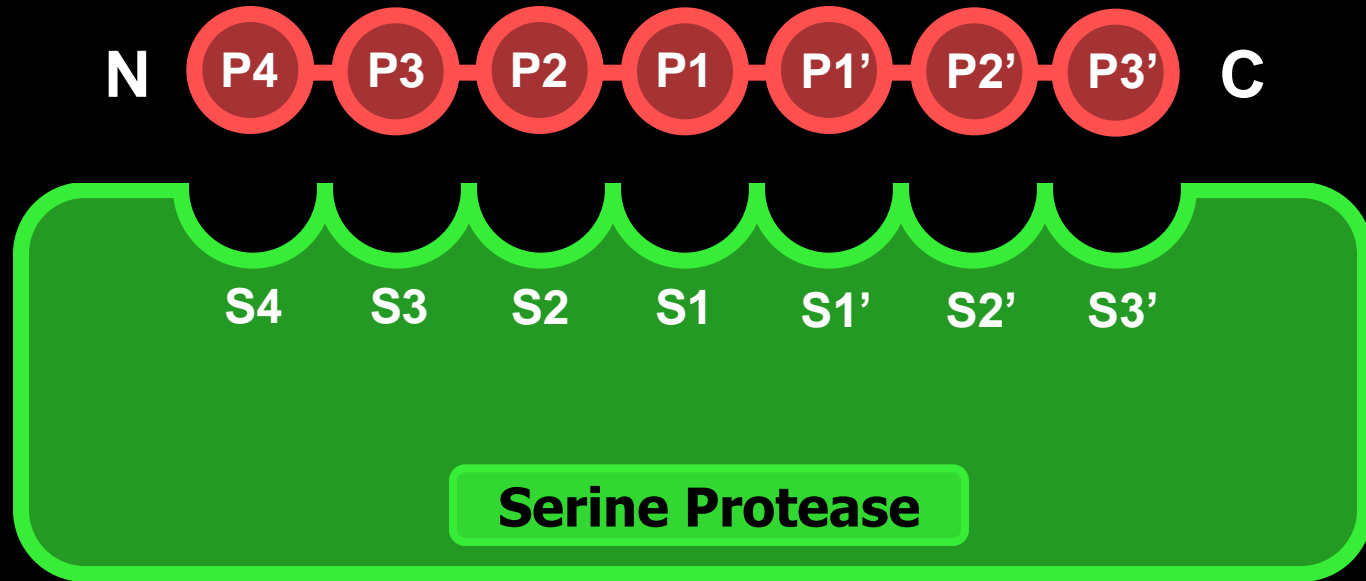
Trypsins

3.4.21.4

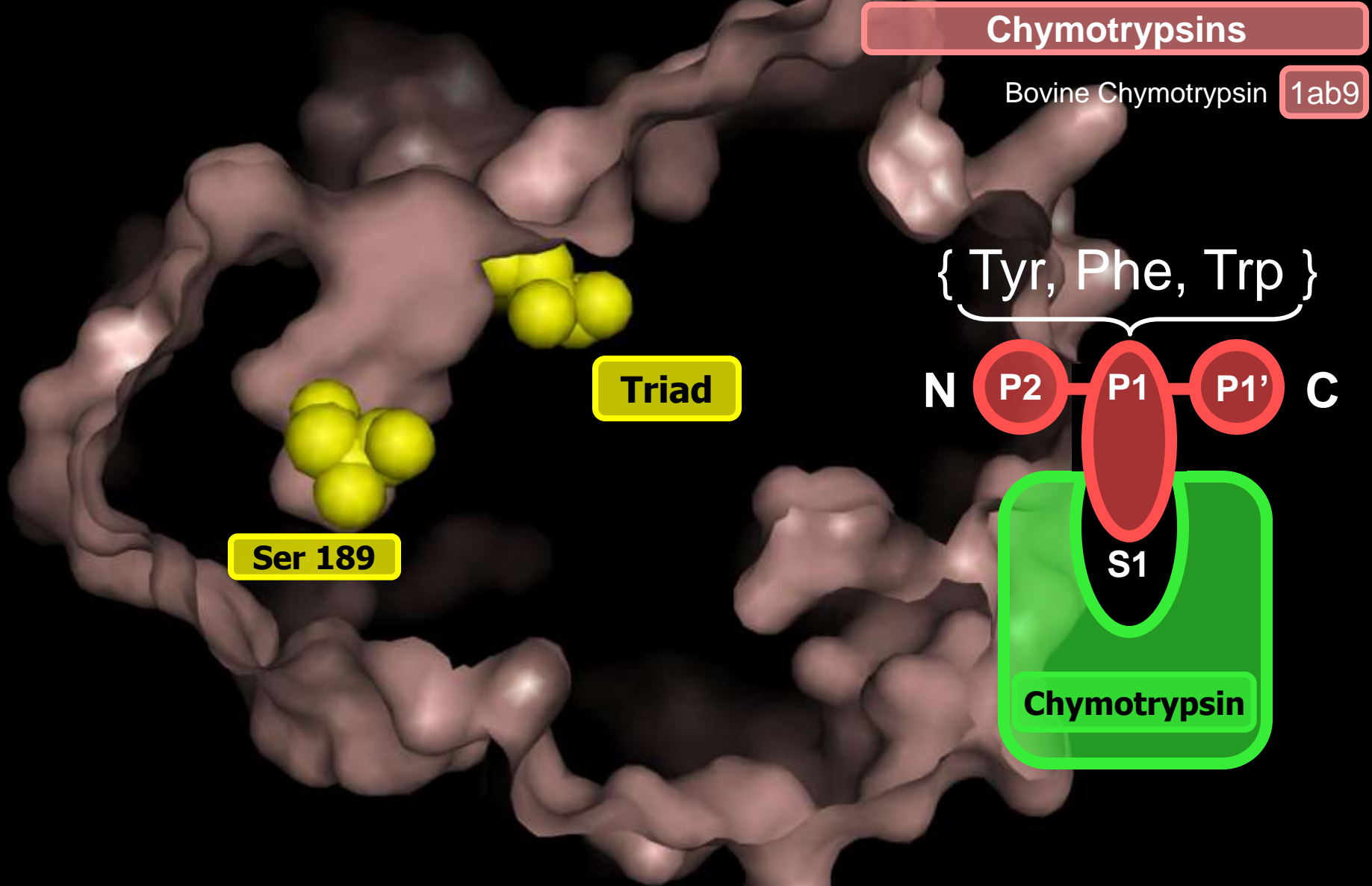
Elastases

3.4.21.36

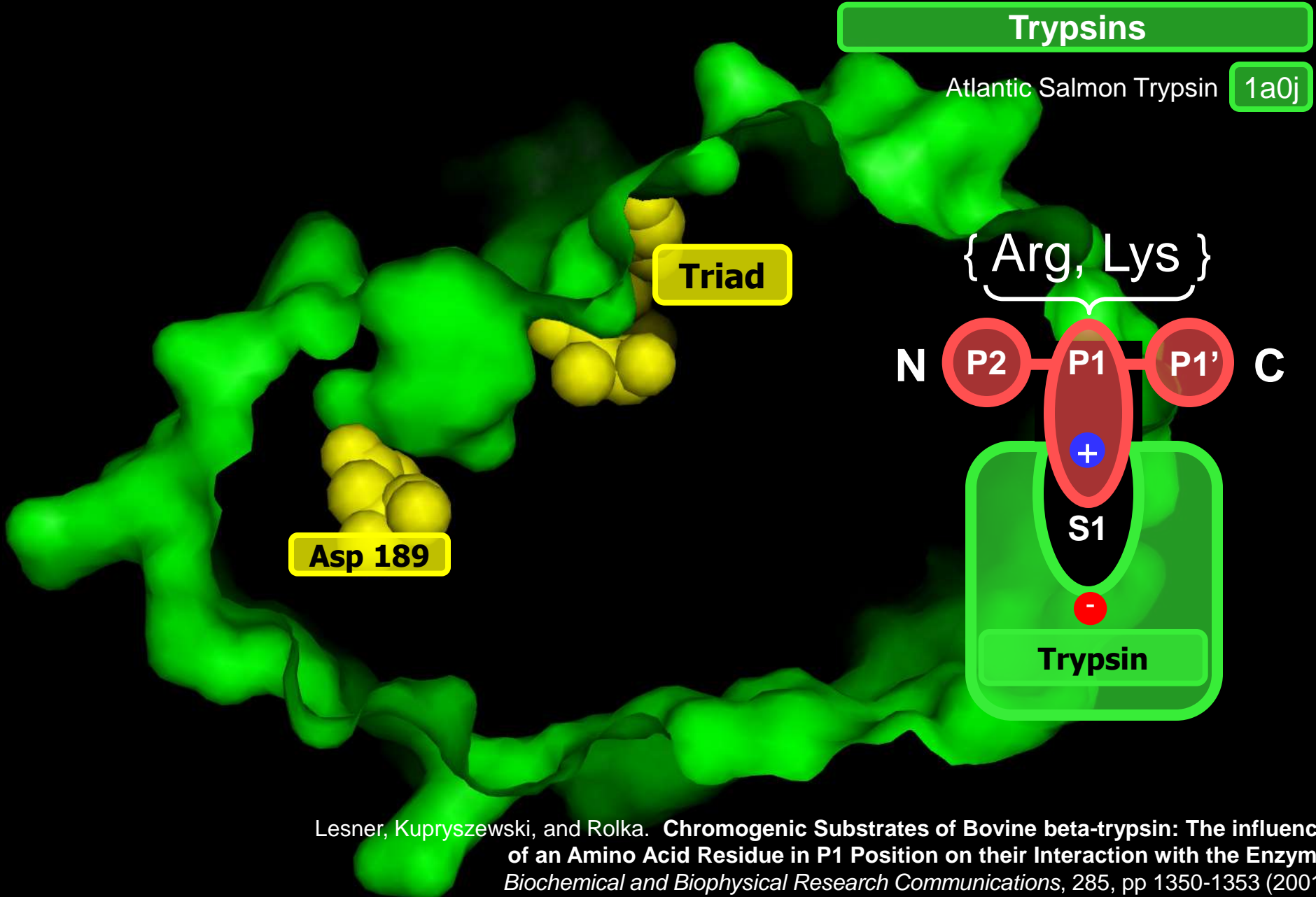
Serine proteases have specificity for different sequences of amino acids



Chymotrypsins prefer big amino acids



Trypsins bind positively charged residues

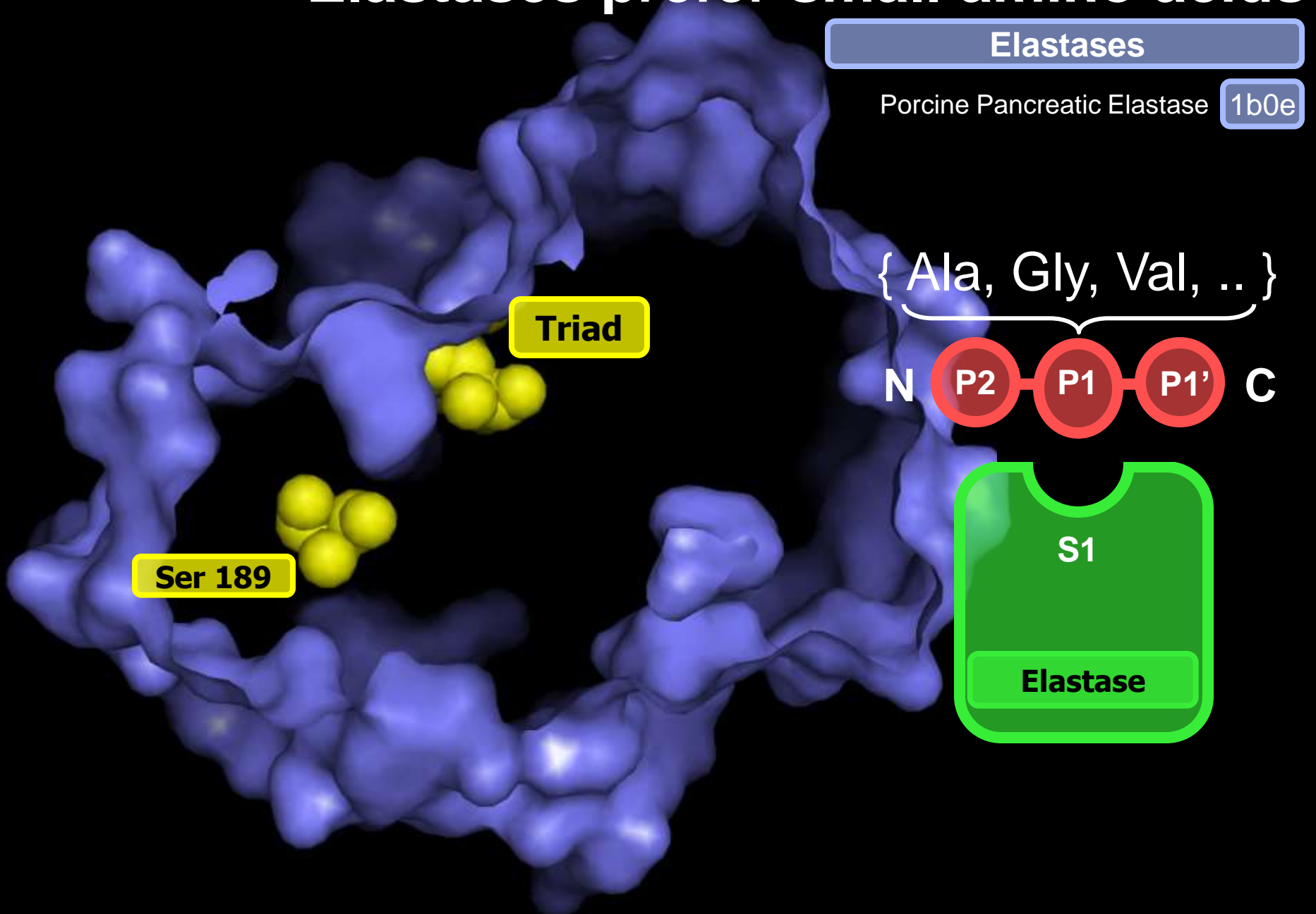


Lesner, Kupryszewski, and Rolka. **Chromogenic Substrates of Bovine beta-trypsin: The influence of an Amino Acid Residue in P1 Position on their Interaction with the Enzyme** *Biochemical and Biophysical Research Communications*, 285, pp 1350-1353 (2001)

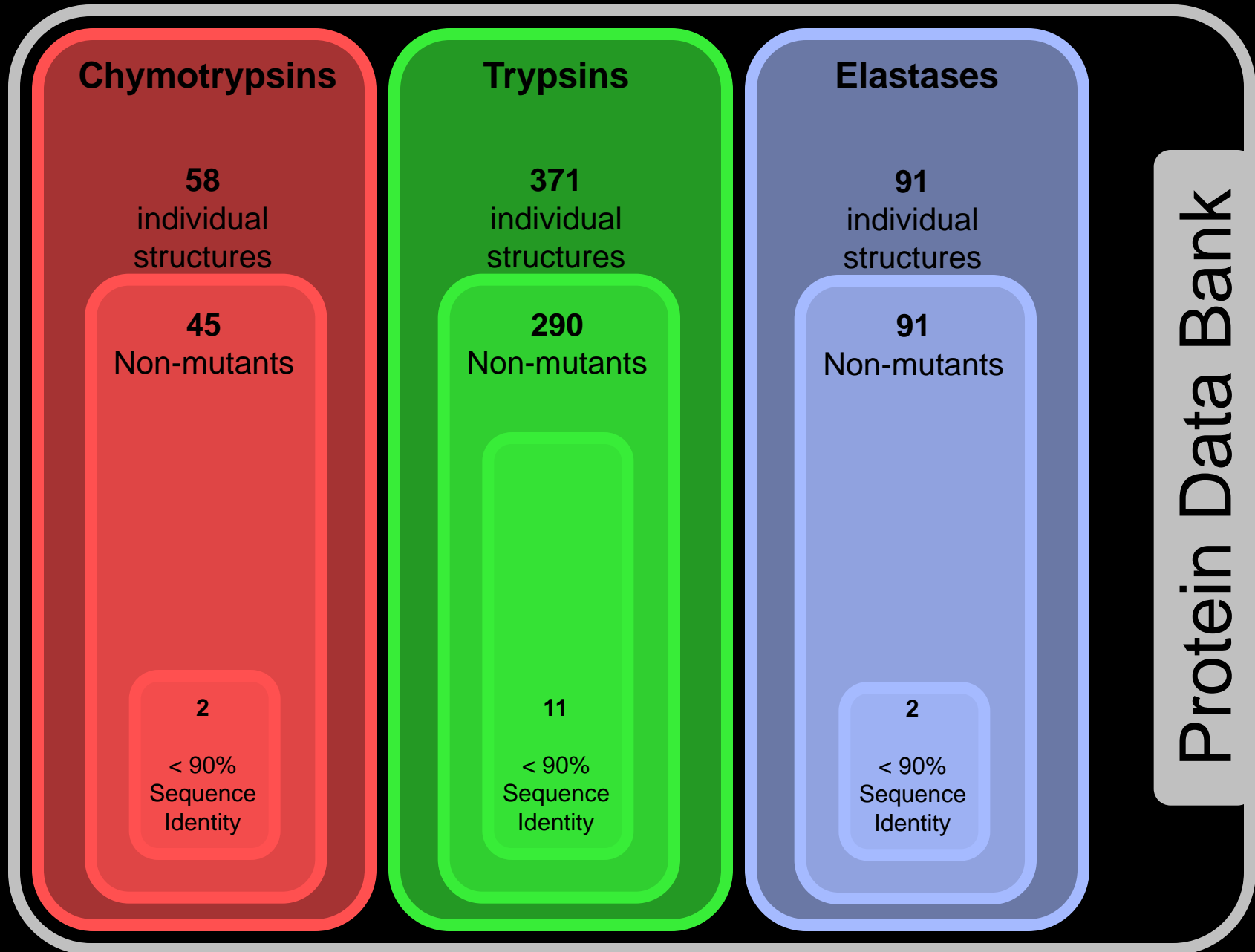
Elastases prefer small amino acids

Elastases

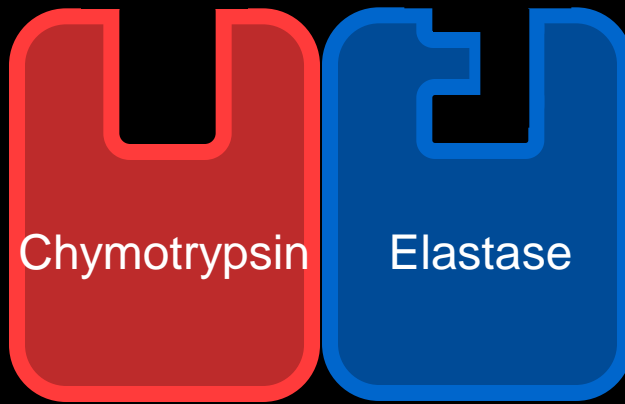
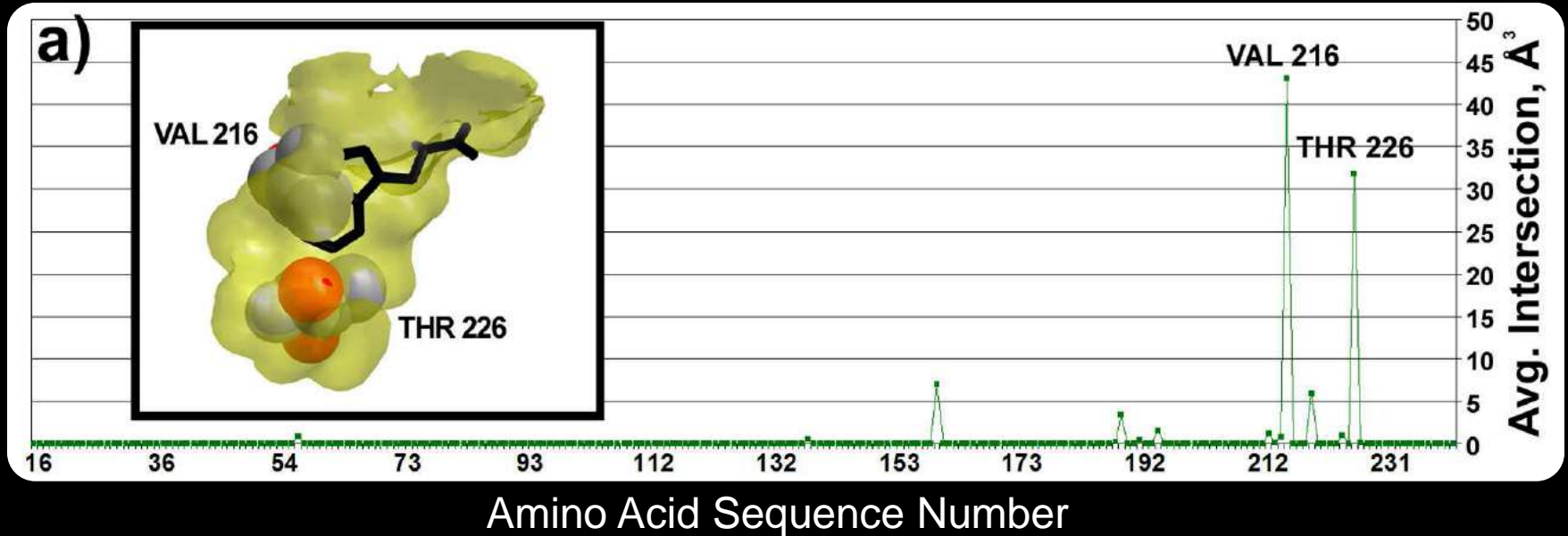
Porcine Pancreatic Elastase 1b0e



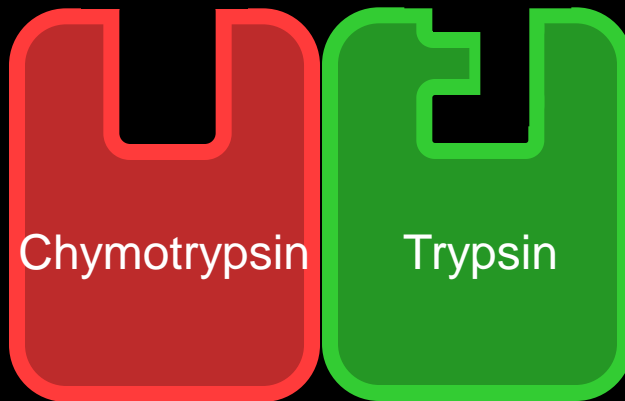
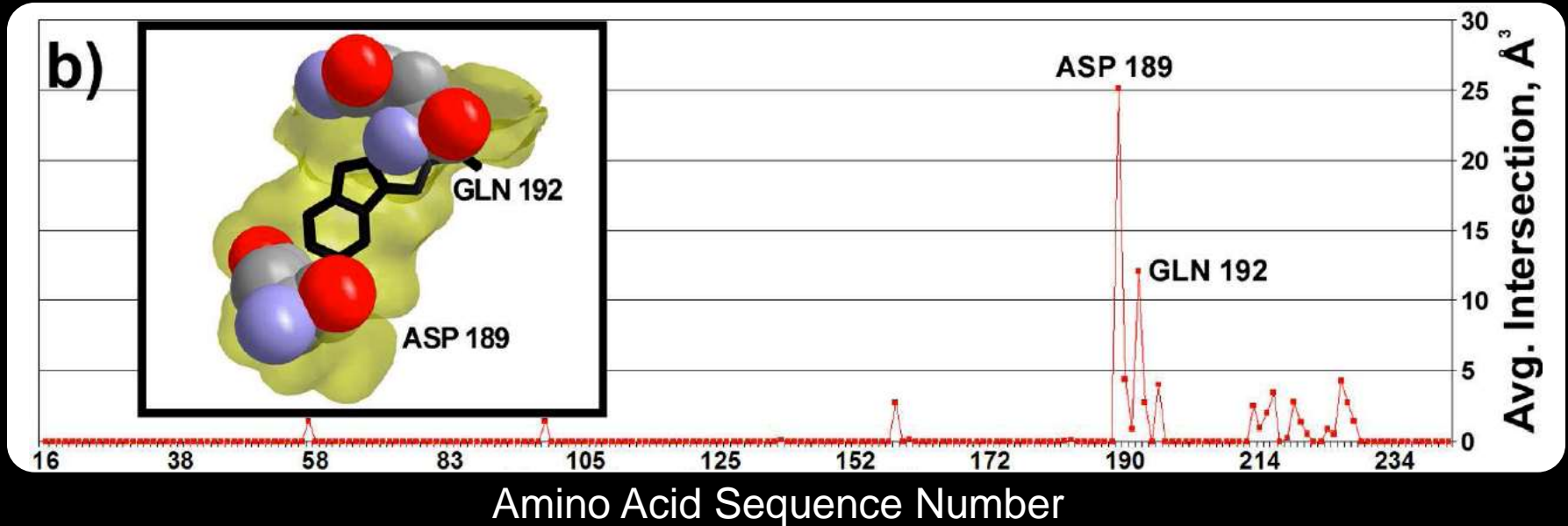
The data is filtered for noise and bias



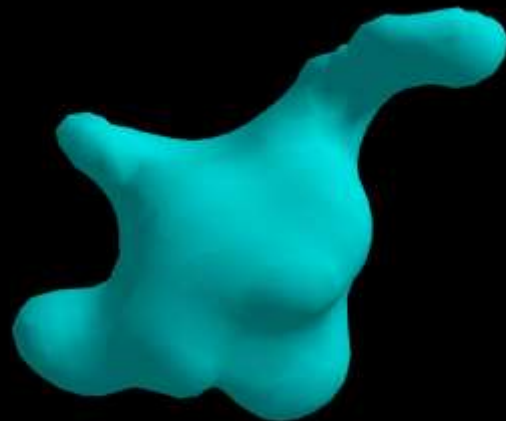
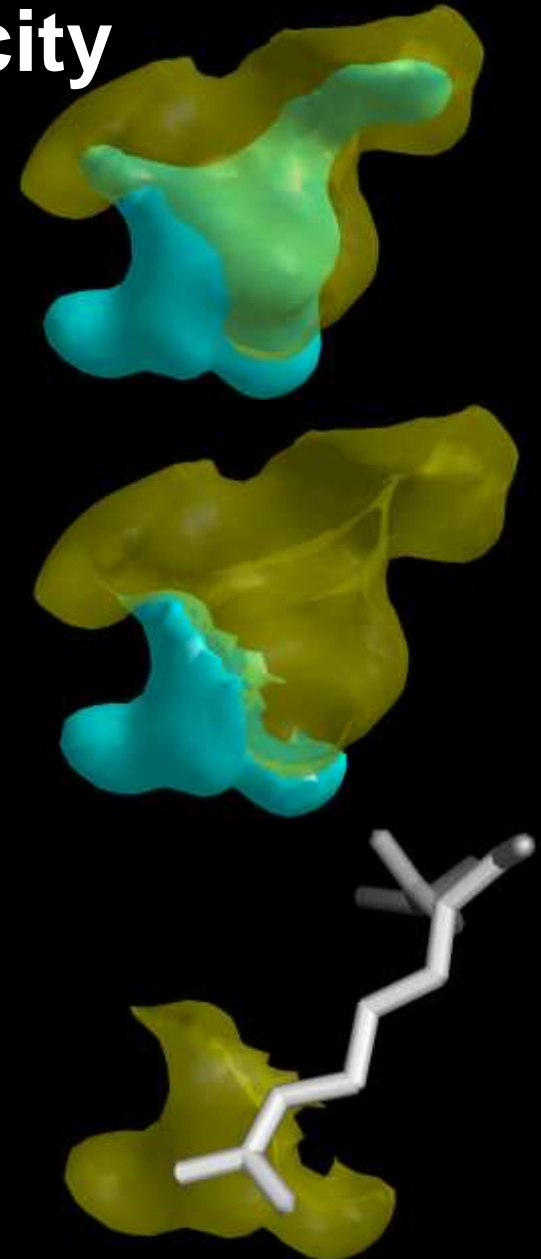
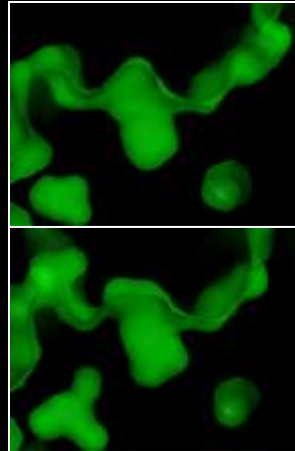
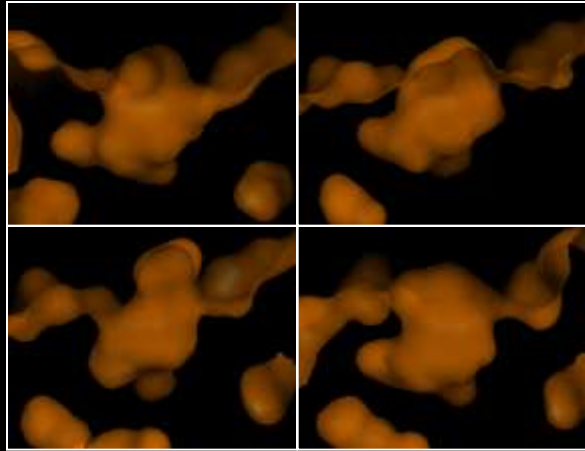
VASP finds amino acids in elastase that influence specificity



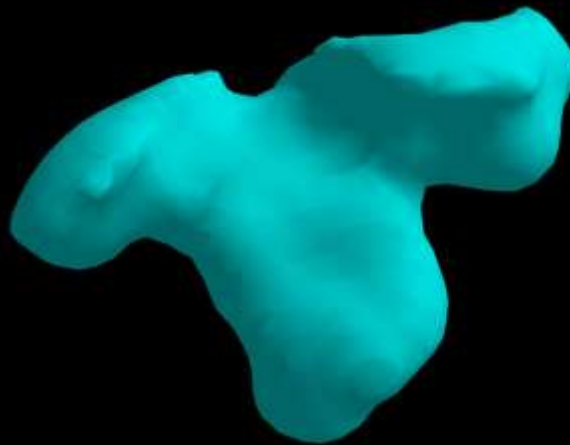
VASP finds amino acids in trypsins that influence specificity



VASP finds subcavities in trypsins and elastases that influence specificity



Trypsin Intersection



Elastase Union

Discussion

- VASP can identify:
 - Amino acids that influence specificity
 - Subcavities that influence specificity
- Contributions
 - The first unsupervised analysis of protein structures that identifies active components of functional sites
 - The first algorithm to isolate the basis for specificity in protein structures
 - The first representation of proteins using smooth solid volumes
- What can we use VASP for ?
 - Identify amino acids that might change specificity in drug resistance
 - Influential subcavities point to drug designs that bind more specifically, and thus reduce side effects

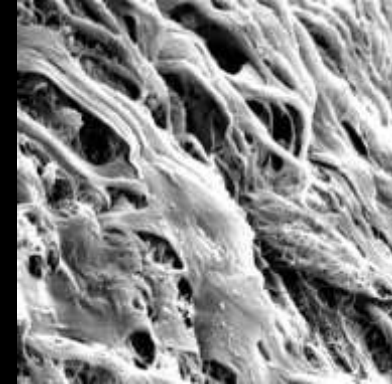
Drug Design



Biofuels



Biomaterials



Specificity is important in all systems



Diagnostics

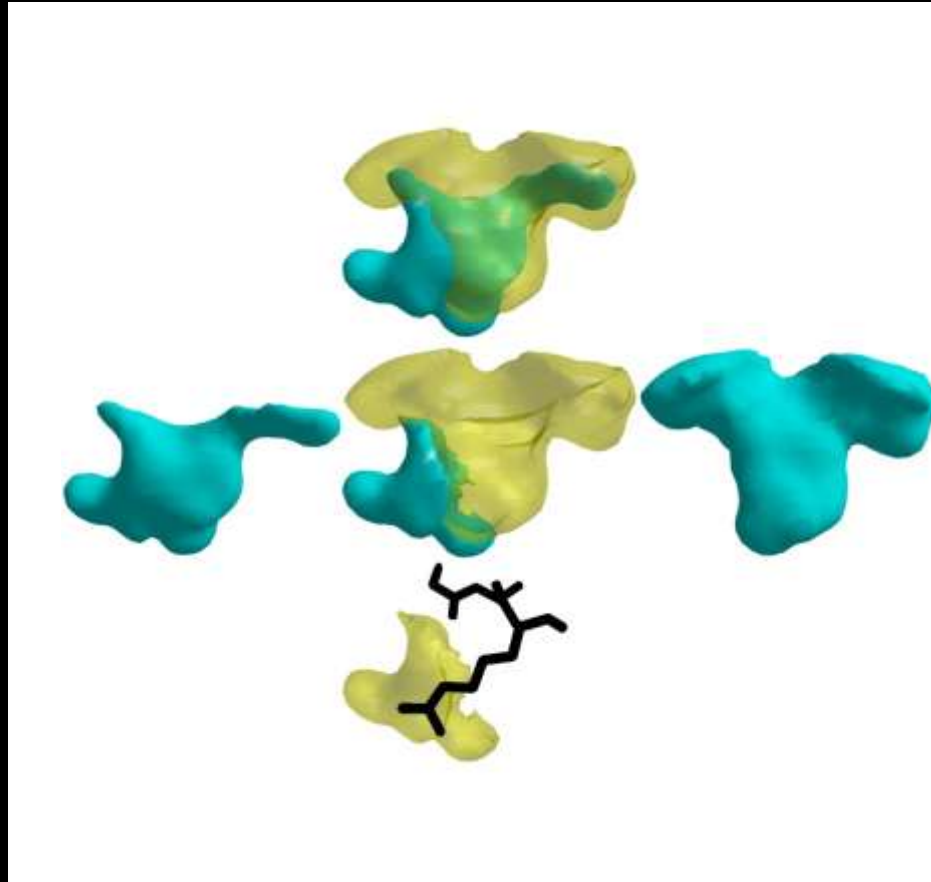


Agriculture



Synthetic Biology

VASP: A Volumetric Analysis of Surface Properties Yields Insights into Protein-Ligand Binding Specificity



Brian Chen and Barry Honig

PLoS Computational Biology. 6(8): e1000881. doi:10.1371/journal.pcbi.1000881.
(in print as of August 12, 2010)

VASP is only the beginning

CSE 350/450: Structural Bioinformatics

- Lectures: Structural Alignment, Finding functional sites, molecular surfaces, Protein-Protein Interactions
- Semester Project: Develop a module of a modern bioinformatics pipeline
- Learn how Biology, Computer Science, and Statistics work together to make biological observations
- Collaborate with students with diverse educational backgrounds
- No programming necessary for some modules

Special Thanks

Dan Lopresti
Jeanne Steinberg
Bryan Hodgson
Judy Frenick

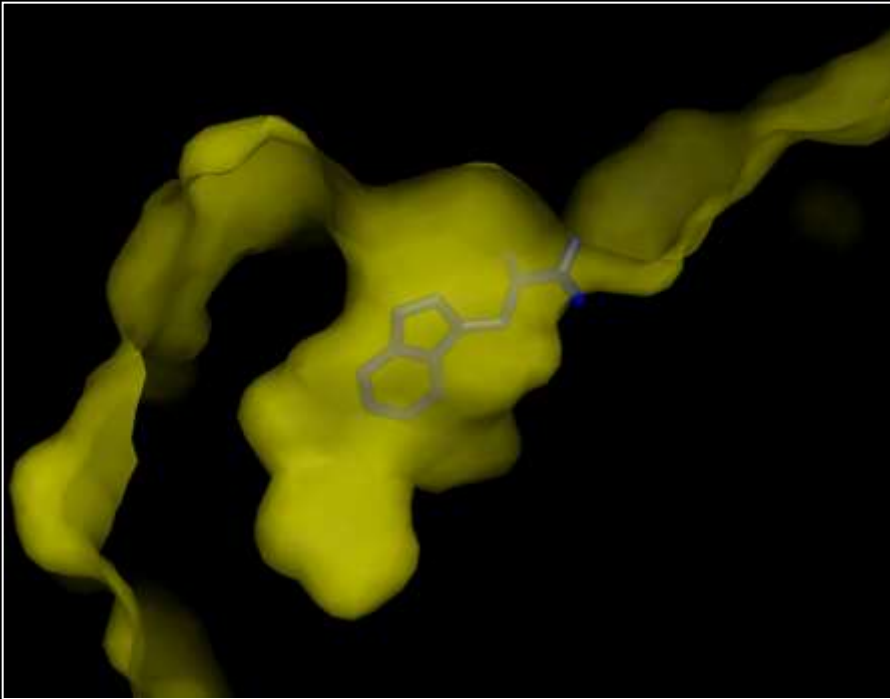
Acknowledgements

- Honig Lab
 - Barry Honig
 - Markus Fischer
 - Donald Petrey
 - Remo Rohs
 - Jeremie Vendome
 - Fabian Dey
 - Shana Posy
- Murray Lab
 - Diana Murray
 - Paul Murray
 - Hunjoong Li
 - Janey Zhang
 - Tonya Silkov
- Kavraki Group
 - Lydia Kavraki
 - Drew Bryant
 - Amanda Cruess
 - Brad Dodson
 - Joey Bylund
- Kimmel Group
 - Marek Kimmel
 - Viacheslav Y. Fofanov
- Lichtarge Lab
 - Olivier Lichtarge
 - David M. Kristensen
- Linusson Lab
 - Anna Linusson
 - David Andersson
- Funding (PhD)
 - NSF
 - NLM, NIGMS
 - HHMI
 - Whitaker Found.
 - Sloan Found.
 - AMD
- Organizations
 - C₂B²
 - Keck Center
- Fumiko Chino

Questions

Example: What makes these cavities different?

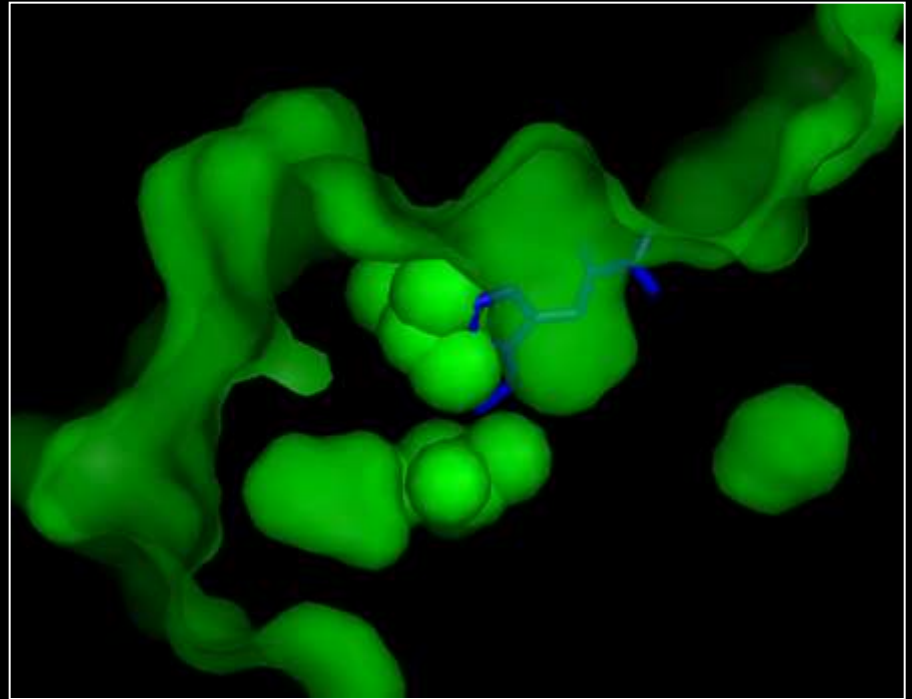
Prefer large hydrophobic amino acids



Bovine Chymotrypsin **8gch**

Tryptophan Ligand **8gch**

Prefer small hydrophobic amino acids



Salmon Elastase **1elt**

Salmon Elastase, V216, T226 **1elt**

Tryptophan Ligand **8gch**

Past Applications of Marching Cubes

- **Medical Imaging**

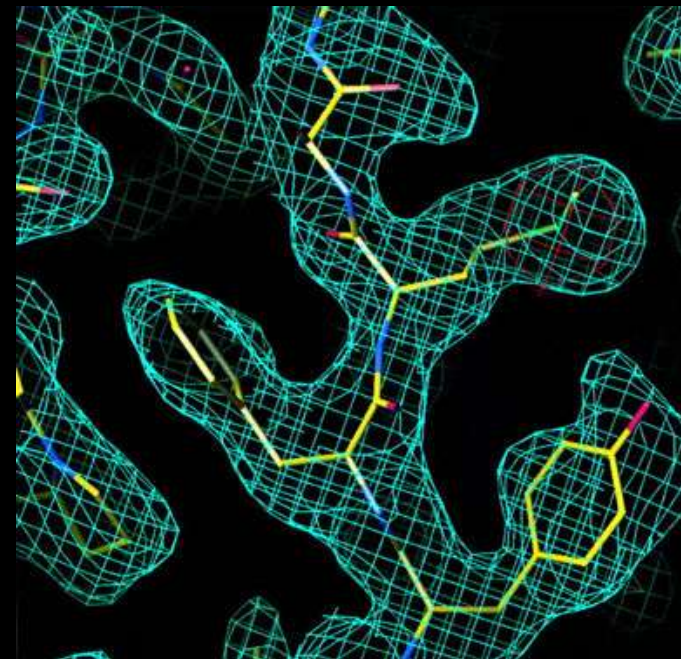
Clarke, LP et al. MRI segmentation: methods and application. *Magnetic Resonance Imaging*. 13(3), 1995.

- **Visualization for Crystallography**

Heiden, W, Goetze, T, Brickmann, J. *J. Comput. Chem.* 14(2), 1993.

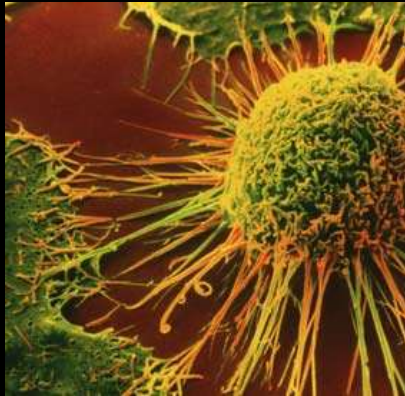
- **Visualization of Protein Surfaces**

Nicholls, A, Sharp, KA, Honig, B. *Proteins* 11:281, 1991.



Why is specificity important?

Cancer



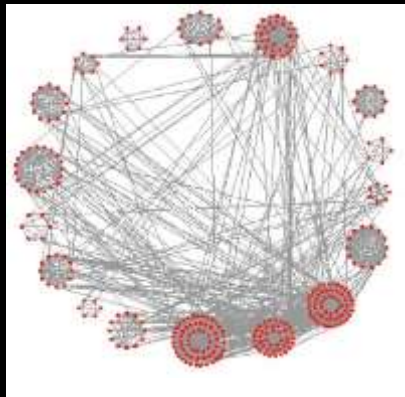
Biofuels



Synthetic Biology



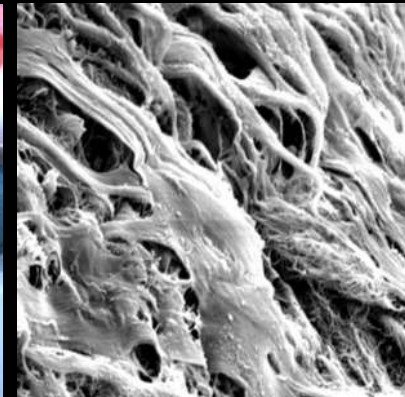
Specificity controls molecular interactions



Systems Biology



Drug Design



Biomaterials

Bloomberg
16:32 17 Déc.

Nouveau M (C)	MIB 30 (C)
3.033,74	43.669,00
▼ 2,91%	▼ 2,10%

DOW future	10.515,00	▼ 265,00
Nasdaq future	2.602,50	▼ 67,50
S&P 500 future	1.326,00	▼ 32,00

SRD (S) 13,00 ▼ 0,56 Equant 29,62 ▼ 1,78 Eram
 FTSE 100 (S) 5,00 ▲ 42,00 Bass 725,00 ▲ 53,00 BG 263,00

Allemagne: prix commerce gros en baisse de 0,2% en nov/oct., +6,4% sur un an

