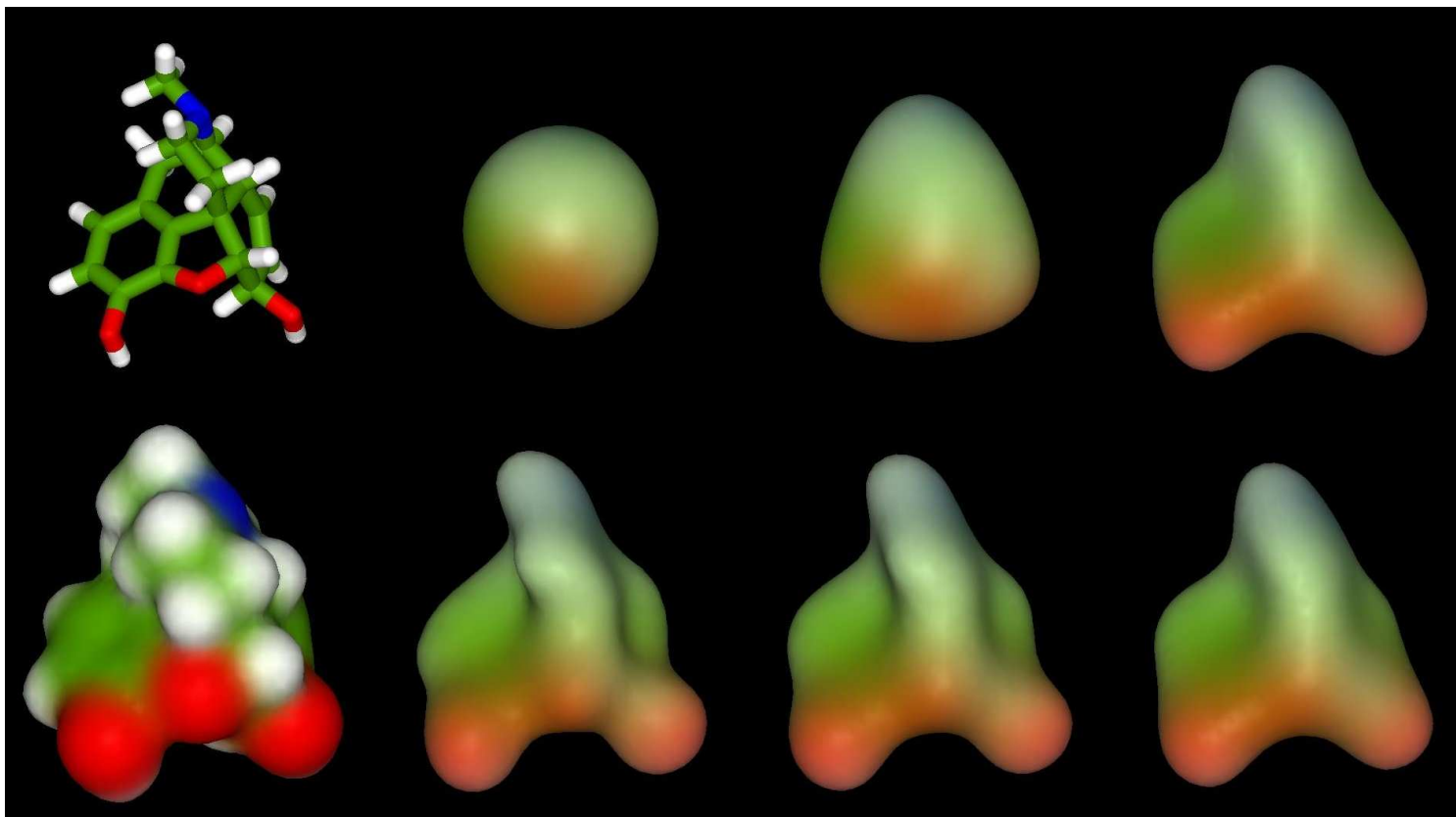


Protein Docking and Molecular Shape Recognition Using Polar Fourier Correlations



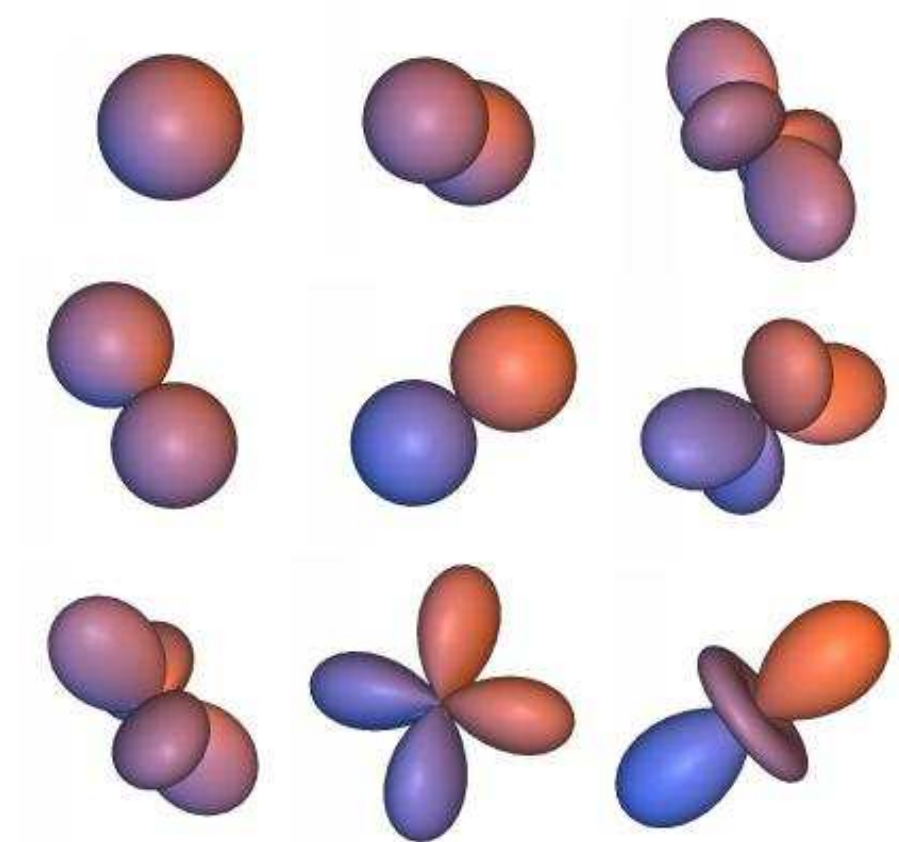
Dave Ritchie
Department of Computing Science
King's College, University of Aberdeen

Protein Docking and Molecular Shape Recognition

Contents

- **Polar Fourier Protein Shape Representation**
- **Application to Protein Docking (CAPRI blind docking trial)**
- **Molecular Shape Recognition & 3D Database Search**
- **Semi-Empirical QM Surface Properties for ComFA/QSAR?**
- **Future Prospects & Conclusions**

Real Spherical Harmonics: $y_{lm}(\theta, \phi)$



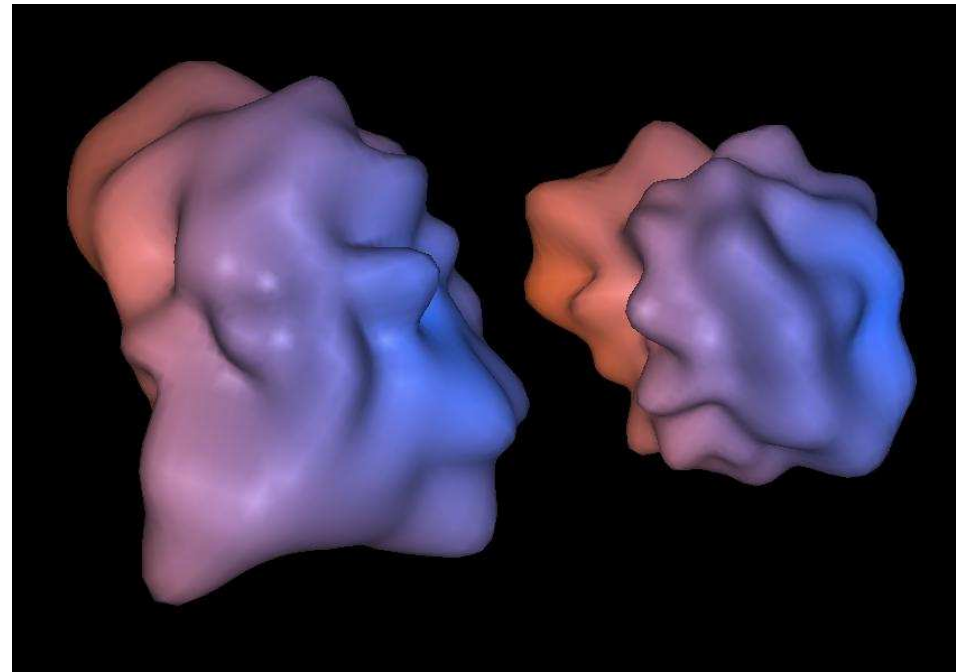
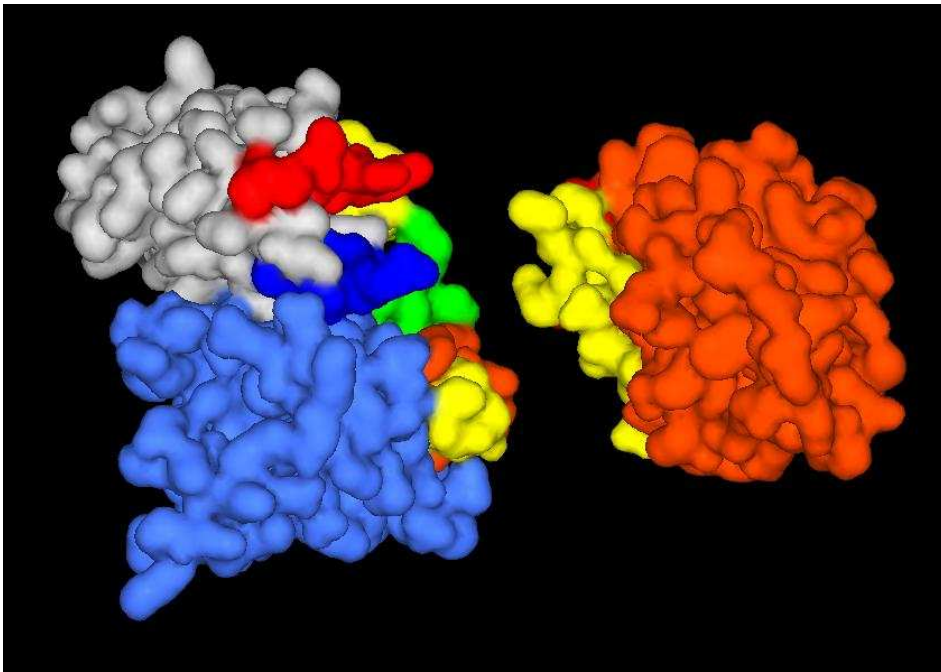
Orthogonality:
$$\int y_{lm}(\theta, \phi) y_{l'm'}(\theta, \phi) d\Omega = \delta_{ll'} \delta_{mm'}$$

Rotation:
$$y_{lm}(\theta', \phi') = \sum_{m'=-l}^l R_{m'm}^{(l)}(\alpha, \beta, \gamma) y_{lm'}(\theta, \phi)$$

Spherical Harmonic Surfaces

Example: 2D Radial Expansions (256 Basis Functions)

$$r(\theta, \phi) = \sum_{l=0}^{15} \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \phi)$$



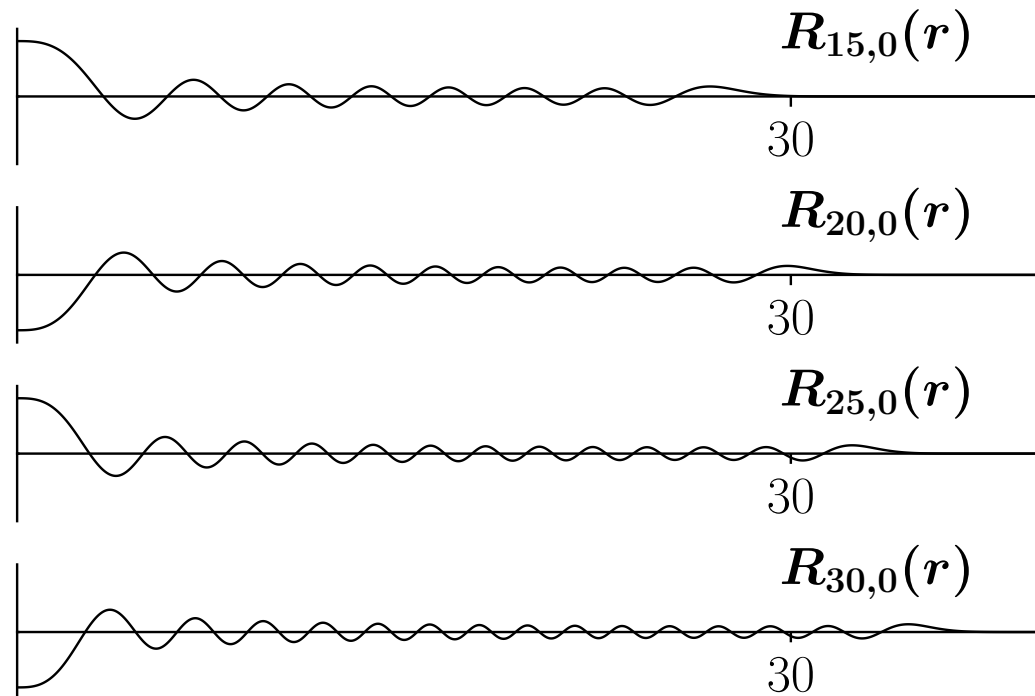
- Good for matching similar shapes, not so good for docking (complementary shapes) ...

Radial Basis Functions: $R_{nl}(r)$

HO-type (shape): $R_{nl}(r) = N_{nl}^{(q)} e^{-\rho/2} \rho^{l/2} L_{n-l-1}^{(l+1/2)}(\rho); \quad \rho = r^2/q, \quad q = 20.$

Coulomb (electro): $R_{nl}(r) = N_{nl}^{(\Lambda)} e^{-\rho/2} \rho^l L_{n-l-1}^{(2l+2)}(\rho); \quad \rho = 2\Lambda r, \quad \Lambda = 1/2.$

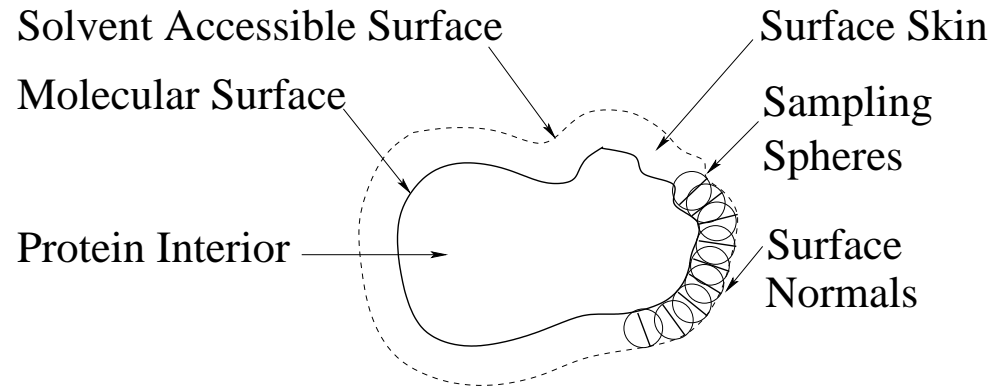
Orthogonality: $\int_0^\infty R_{nl}(r) R_{n'l}(r) r^2 dr = \delta_{nn'}$



3D Protein Shape Density Representations

(Ritchie & Kemp (2000) Proteins 39 178–194)

- Sample surface skins onto a $(0.75\text{\AA})^3$ grid...



Surface Skin: $\sigma(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{surface skin} \\ 0; & \text{otherwise} \end{cases}$ Interior: $\tau(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{protein atom} \\ 0; & \text{otherwise} \end{cases}$

Parametrise as: $\sigma(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\sigma} R_{nl}(\underline{r}) y_{lm}(\theta, \phi), \text{ etc.}$

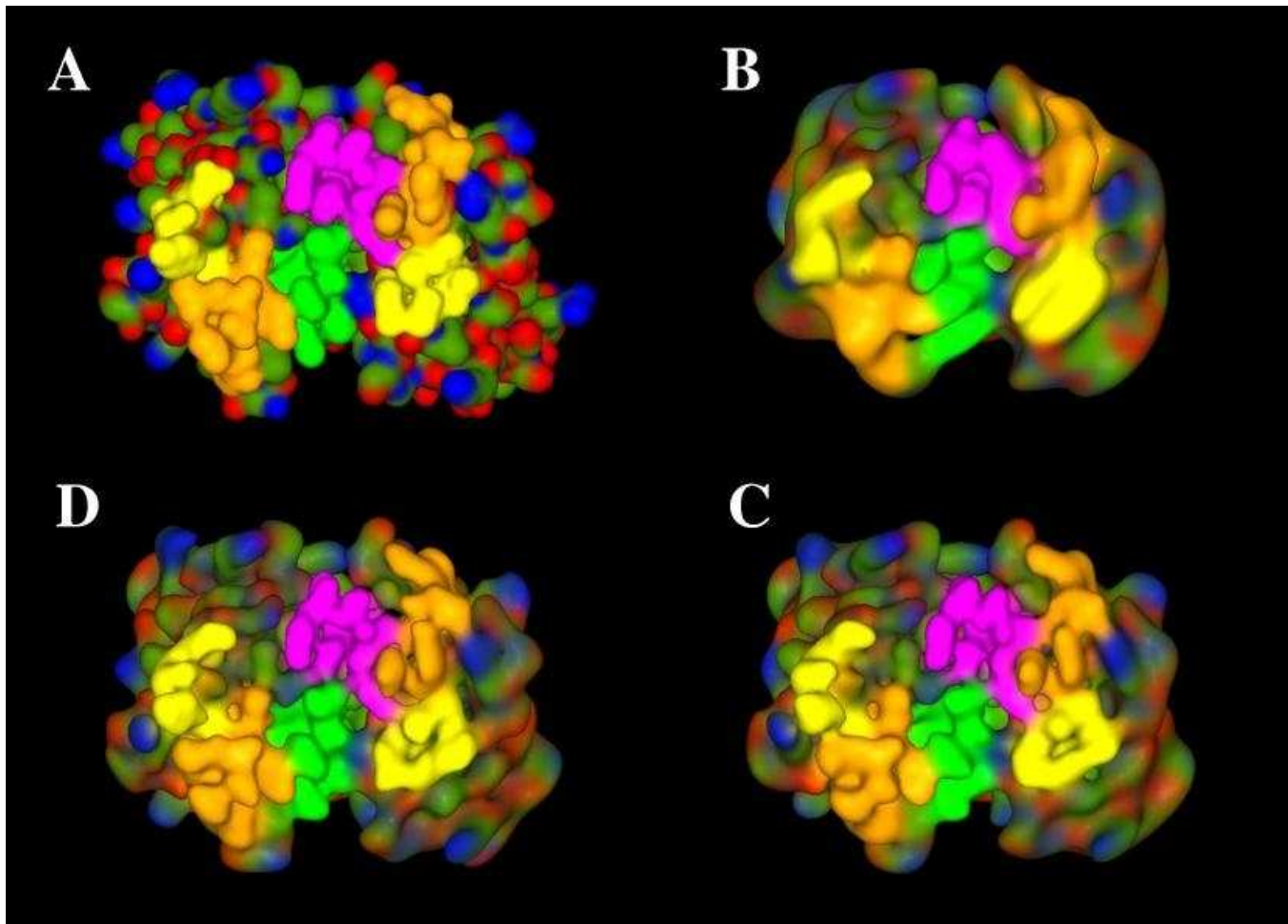
Estimate as: $a_{nlm}^{\sigma} \simeq \sum_c R_{nl}(r_c) y_{lm}(\theta_c, \phi_c) \Delta V$

- Only need to do this once for each protein...

3D Shape Density Reconstruction

(Ritchie (2003) Proteins 52 98–106)

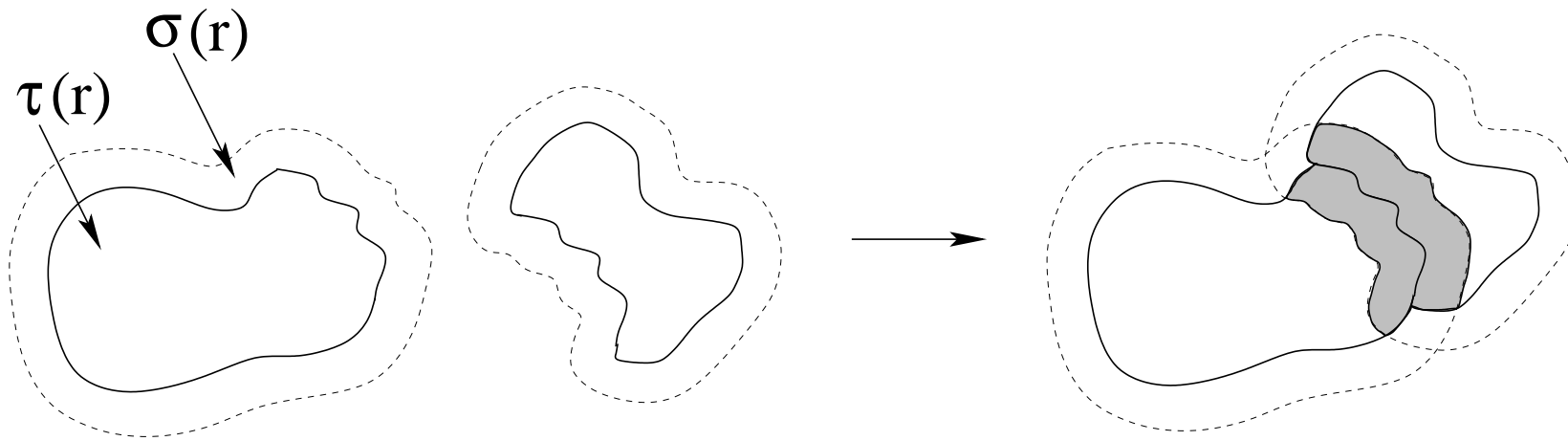
- Example – Looking at MCV Antibody CDR Loops:



$$\tau(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\tau} R_{nl}(r) y_{lm}(\theta, \phi)$$

Image	Expansion	Coefficients
A	Gaussians	-
B	N = 16	1,496
C	N = 25	5,525
D	N = 30	9,455

Protein Shape Complementarity Using Density Representations



Favourable:
$$\int (\sigma_A(\underline{r}_A)\tau_B(\underline{r}_B) + \tau_A(\underline{r}_A)\sigma_B(\underline{r}_B))dV$$

Unfavourable:
$$\int \tau_A(\underline{r}_A)\tau_B(\underline{r}_B)dV$$

Score:
$$S_{AB} = \int (\sigma_A\tau_B + \tau_A\sigma_B - Q\tau_A\tau_B)dV;$$
 Penalty Factor: $Q = 11$

Correlations – Overlap as a Function of Coordinate Operations

(Ritchie (2005) J Appl Cryst 38 808–818)

Basic Overlap:

$$\int \sigma_A(\underline{r})\tau_B(\underline{r})dV = \sum_{nlm}^N a_{nlm}^\sigma b_{nlm}^\tau$$

Rotation:

$$\hat{R}(\alpha, \beta, \gamma)\sigma_A(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\sigma'} R_{nl}(r) y_{lm}(\theta, \phi)$$

Rotated Coefficients:

$$a_{nlm}^{\sigma'} = \sum_{m'=-l}^l R_{mm'}^{(l)}(\alpha, \beta, \gamma) a_{nlm'}^\sigma$$

Translation:

$$\hat{T}_z(R)\sigma_A(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\sigma''} R_{nl}(r) y_{lm}(\theta, \phi)$$

Translated Coefficients:

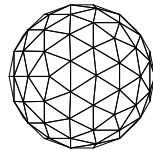
$$a_{nlm}^{\sigma''} = \sum_{n'l'}^N T_{nl,n'l'}^{(|m|)}(R) a_{n'l'm}^\sigma$$

Hence:

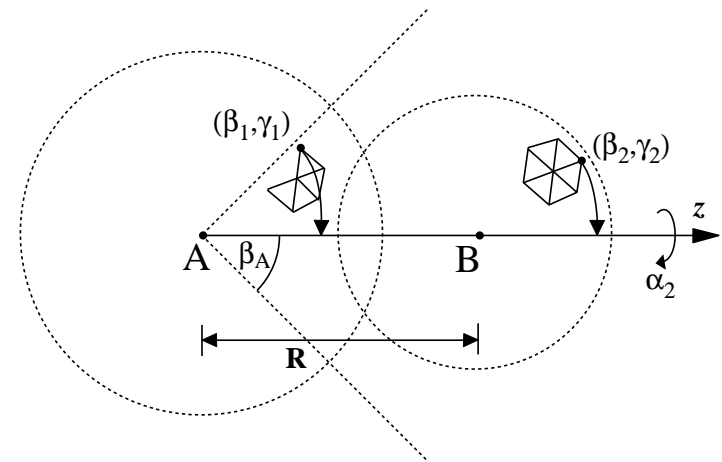
$$\int \sigma'_A(\underline{r})\tau''_B(\underline{r})dV = \sum_{nlm}^N a_{nlm}^{\sigma'} b_{nlm}^{\tau''} \text{ etc.}$$

6D Docking Search as a Nested Sequence of Transformations

Get 4 rotations from icosahedral tessellations ...



...



Rotate A ($\times 492$ @ 9.9°): $A'(\underline{r}) = \hat{R}(0, \beta_1, \gamma_1)A(\underline{r})$

Translate A ($\times 50$ @ 0.75\AA): $A''(\underline{r}) = \hat{T}_z(-R)A'(\underline{r})$

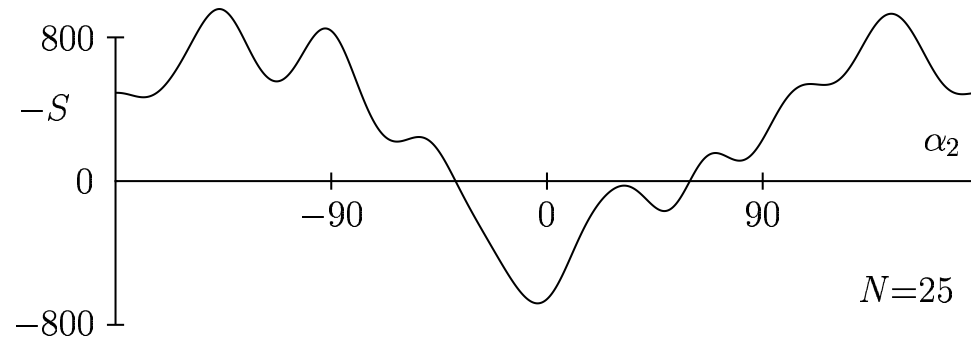
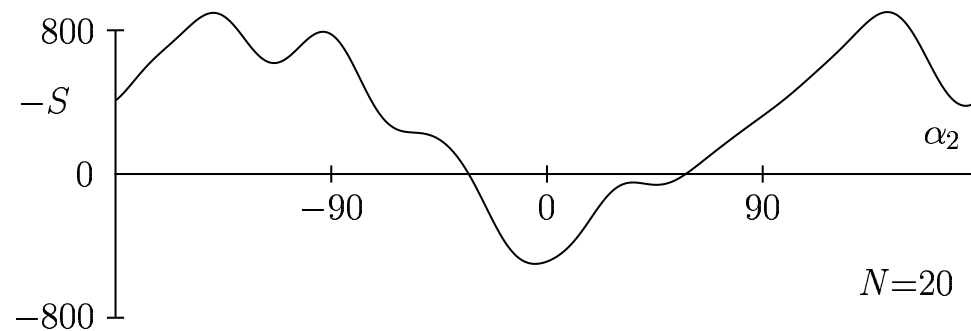
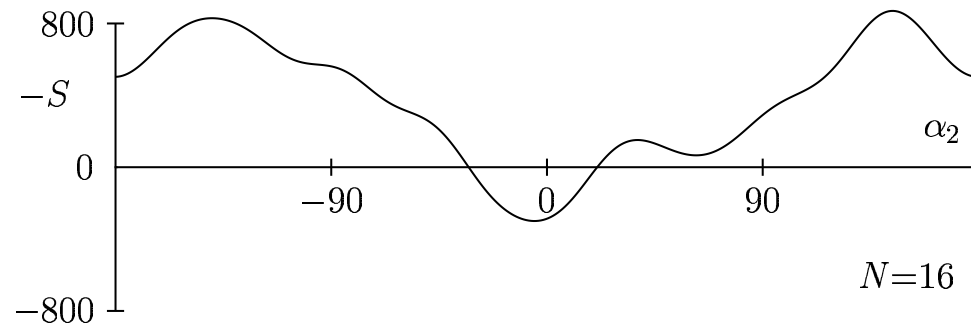
Rotate B ($\times 642$ @ 8.5°): $B'(\underline{r}) = \hat{R}(0, \beta_2, \gamma_2)B(\underline{r})$

Twist B ($\times 64$ @ 5.6°): $B''(\underline{r}) = \hat{R}(\alpha_2, 0, 0)B'(\underline{r})$

1D FFT:
$$S_{AB}(\alpha_2) = \sum_{m=1-N}^{N-1} P_m \cos m\alpha_2 + Q_m \sin \bar{m}\alpha_2$$

Search Space: $492 \times 50 \times 642 \times 64 \simeq 10^9$ ($\sim 10^6/s$ on a 1GHz PIII Xeon)

Shape Correlation Score as a Function of Twist Angle α_2 (Antibody HyHel-5/Lysozyme Complex)



Correlating Electrostatics

The classical electrostatic energy of a charge density, ρ , in a potential, ϕ , is given by:

$$E = \frac{1}{2} \int \rho(\underline{x}) \phi(\underline{x}) dV$$

For a pair of proteins, A and B:

$$\rho(\underline{r}) = \rho_A(\underline{r}_A) + \rho_B(\underline{r}_B)$$

$$\phi(\underline{r}) = \phi_A(\underline{r}_A) + \phi_B(\underline{r}_B)$$

Hence we obtain a polar Fourier expression:

$$E = \frac{1}{2} \sum_{nlm} \sum_{n'l'm'} \left(a'_{nlm} b'_{n'l'm'} + a'_{n'l'm'} b'_{nlm} \right) J_{nl,n'l'}^{(|m|)}(\mathbf{R}) \delta_{mm'}$$

Adding the shape and electrostatic components gives a total docking energy (KJ/mol units):

$$E_{\text{total}} = K_H S + \left(\frac{1391.4}{K_R} \right) E$$

K_H : hydrophobic factor, $K_H = -0.8 \text{ KJ/mol/\AA}^3$.

K_R : relative permittivity, $K_R = 8$.

Re-Docking Known Protein Complexes

Case	$N = 16$		$N = 20$		$N = 25$	
	Top	RMS	Top	RMS	Top	RMS
SIC	3,407	0.00	2	0.22	1	0.82
KAI	17	0.41	3	0.69	7	0.81
PTC	132	0.52	2	0.48	1	0.48
CGI	1	0.38	1	0.38	1	0.38
CHO	1	0.45	1	0.55	1	0.55
BGS	1	0.82	1	0.82	1	0.88
GGI	1	2.47	1	0.90	1	0.90
TET	5	1.48	1	1.16	1	1.03
FPT	102	1.04	1	0.42	1	0.42
IGF	3	0.71	1	0.77	1	0.77
JEL	4,867	0.81	1,060	0.81	2	0.81
BQL	524	1.85	12	0.96	1	0.39
HFL	318	1.01	5	1.00	1	1.00
HFM	7	2.19	27	1.09	10	1.09
VFB	8,344	1.49	216	0.20	9	0.20
MLC	1,401	0.00	116	0.00	187	0.84
MEL	9,898	1.03	27	1.03	3	1.03
JHL	385	0.62	8	0.38	1	1.08
FBI	14	1.09	1	1.09	1	0.38
NCA	68	1.53	1	0.32	1	0.32
NMB	160	2.43	1,630	1.39	1,009	1.39
NSN	19,992	1.11	716	0.75	1,130	2.29
IAI	1,381	1.48	111	0.37	20	1.39
DVF	11,145	0.00	88	1.38	49	0.44
KB5	140	0.34	1	0.34	78	1.38
IGC	1,328	1.74	269	0.81	1	0.34

CAPRI – Critical Assessment of Predicted Interactions

- Following from CASP: Janin, Wodak, et al.
- During 2001/2: seven target complexes made available...

Target	Receptor	Ligand	Type	Complex	Lab
1	HPr Kinase	HPr	U/U	Fioulaine et al.	Janin
2	Rotavirus VP6	MCV	U/B	Vaney et al.	Rey
3	Hemagglutinin	HC63	U/B	Barbey-Martin et al.	Knossow
4	α -Amylase	AMD10	U/B	Desmyter et al.	Cambillau
5	α -Amylase	AMB7	U/B	Desmyter et al.	Cambillau
6	α -Amylase	AMD9	U/B	Desmyter et al.	Cambillau
7	SpeA	TCR 14.3.D	U/U	Sundberg et al.	Mariuzza

- At least one docking partner presented in its unbound form
- Participants permitted 5 attempts for each target
- Any predictive approach allowed: homology/literature, etc.
- Initially, 19 groups & 7 targets (now 35 groups & 20 targets)...

CAPRI Results: Targets 1–7

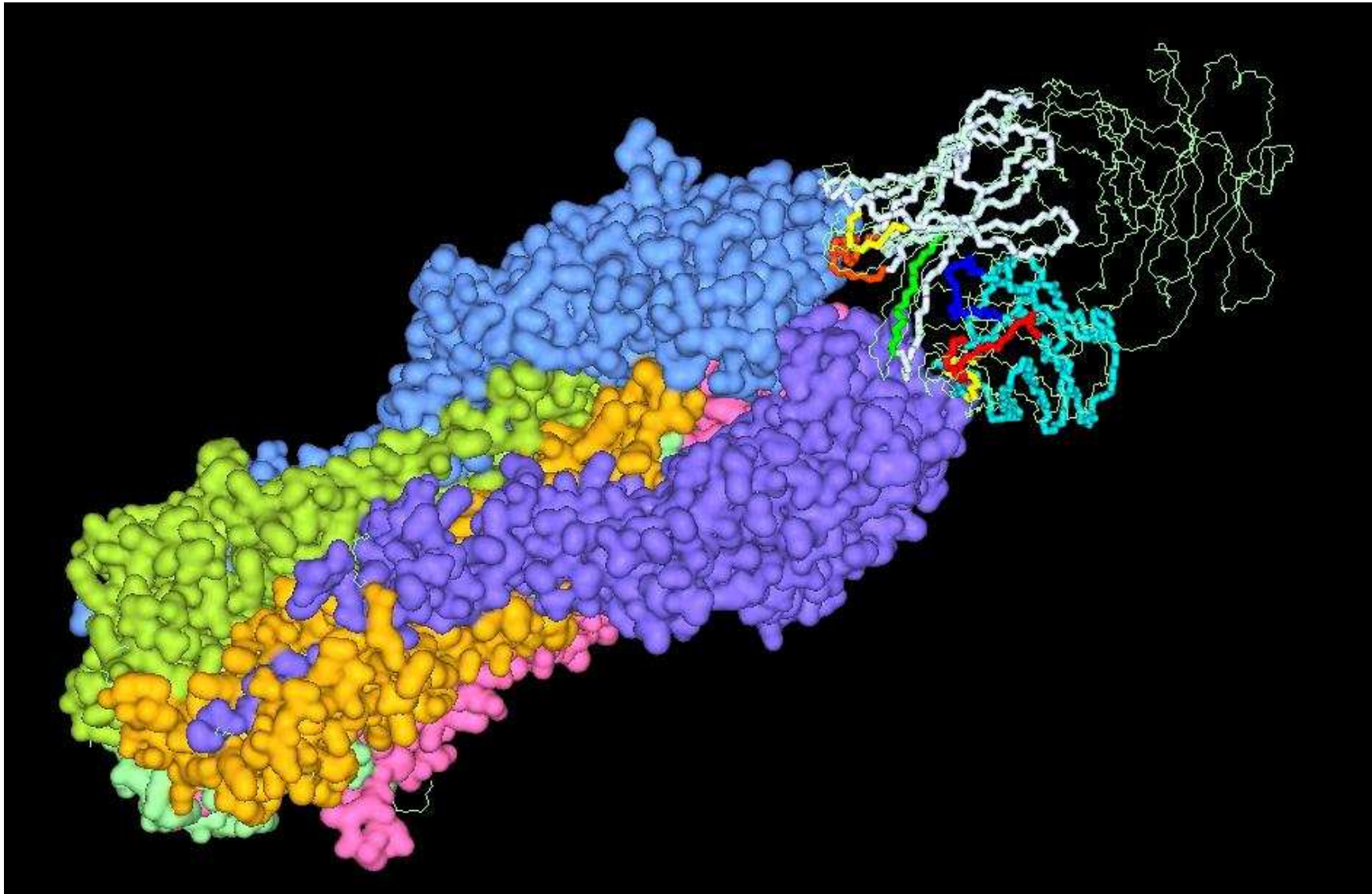
(Mendez et al. (2003) Proteins 52 51–67)

Predictor	Software	Algorithm	T1	T2	T3	T4	T5	T6	T7
Abagyan	ICM	FF			+			+	+
Bates	Guided Docking	FF							+
Camacho	CHARMM	FF	-					+	+
Eisenstein	MolFit	FFT	-	-					+
Gardiner	GAPDOCK	GA	-	-					
Gray		MC						+	+
Olson	Surfdock	SH	-						
Palma	BIGGER	GF						+	-
Ritchie	Hex	SPF			+			+	
Sternberg	FTDOCK	FFT	-					+	-
Ten Eyck	DOT	FFT	-	-				+	
Vakser	GRAMM	FFT		-					
Valencia		ANN	-						
Weng	ZDOCK	FFT		+					+
Wolfson	BUDDA/PPD	GH	-						+

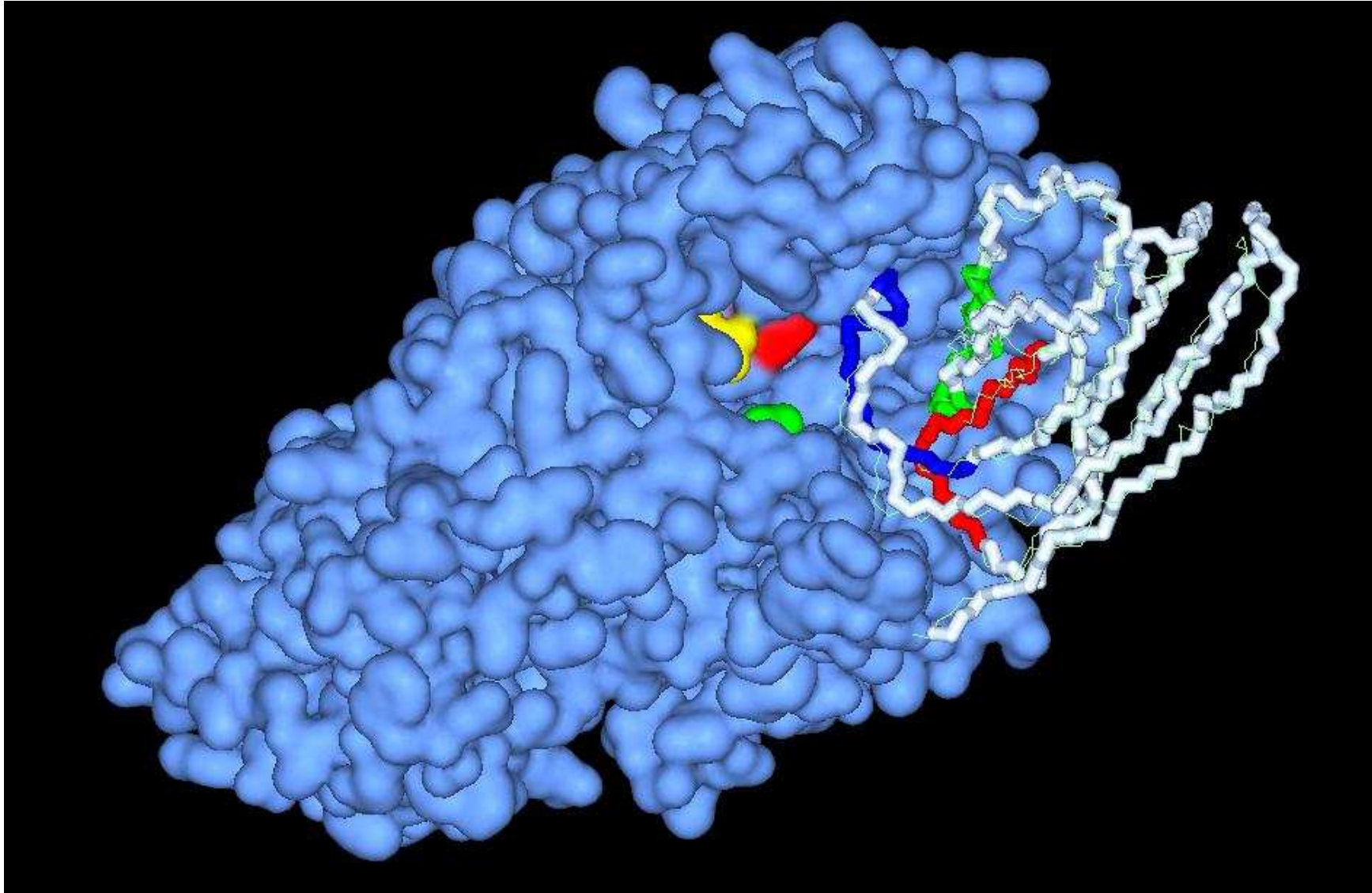
+ medium or high accuracy prediction; – low accuracy prediction

Latest CAPRI paper: Mendez et al. (2005), Proteins 60 150-169

Docked Orientation for Target 3 (Hemagglutinin/HC63)



Docked Orientation for Target 6 (Amylase/AMD9)



3D Molecular Field Similarity Scores

- Define “distance” between two 3D functions:

$$D^2 = \int |\rho_A - \rho_B|^2 dV = \int \rho_A^2 + \rho_B^2 - 2\rho_A\rho_B dV$$

- Can normalise to give Hodgkin, Tanimoto, or Carbo index:

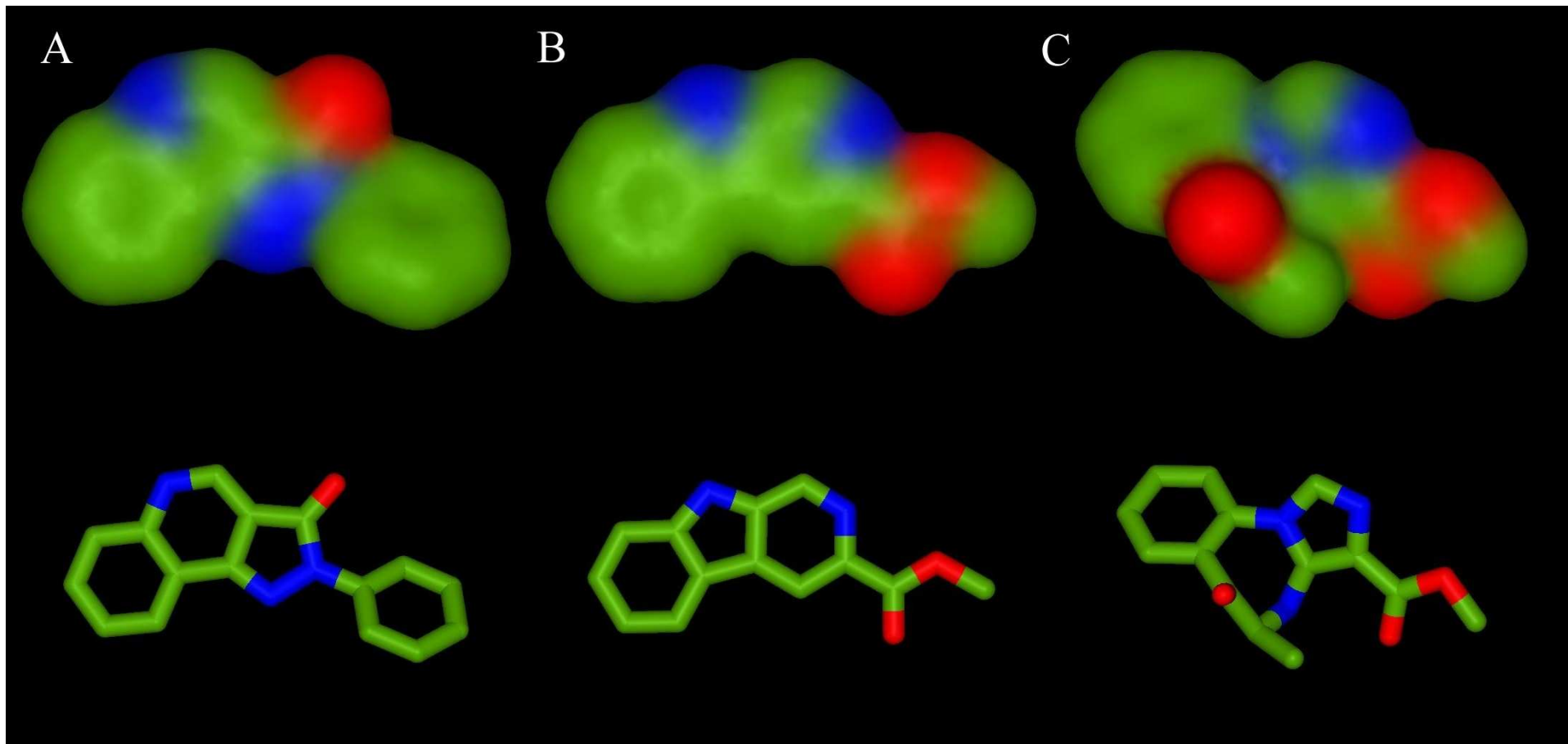
e.g. Carbo
$$S_{AB} = \frac{\int \rho_A \rho_B dV}{[\int \rho_A^2 dV \int \rho_B^2 dV]^{1/2}}$$

- So we wish to evaluate 6D correlations of the form:

$$S_{AB}(R, \beta_A, \gamma_A, \alpha_B, \beta_B, \gamma_B) = \int \rho_A(\underline{r})' \rho_B(\underline{r})'' dV$$

- With $N = 8$ (204 coeffs), get 5×10^6 /sec (1.8GHz PIII Xeon)

3D Shape Matching Example: Benzodiazepines



A: CGS-8216

B: Methyl- β -carboline-3-carboxylate

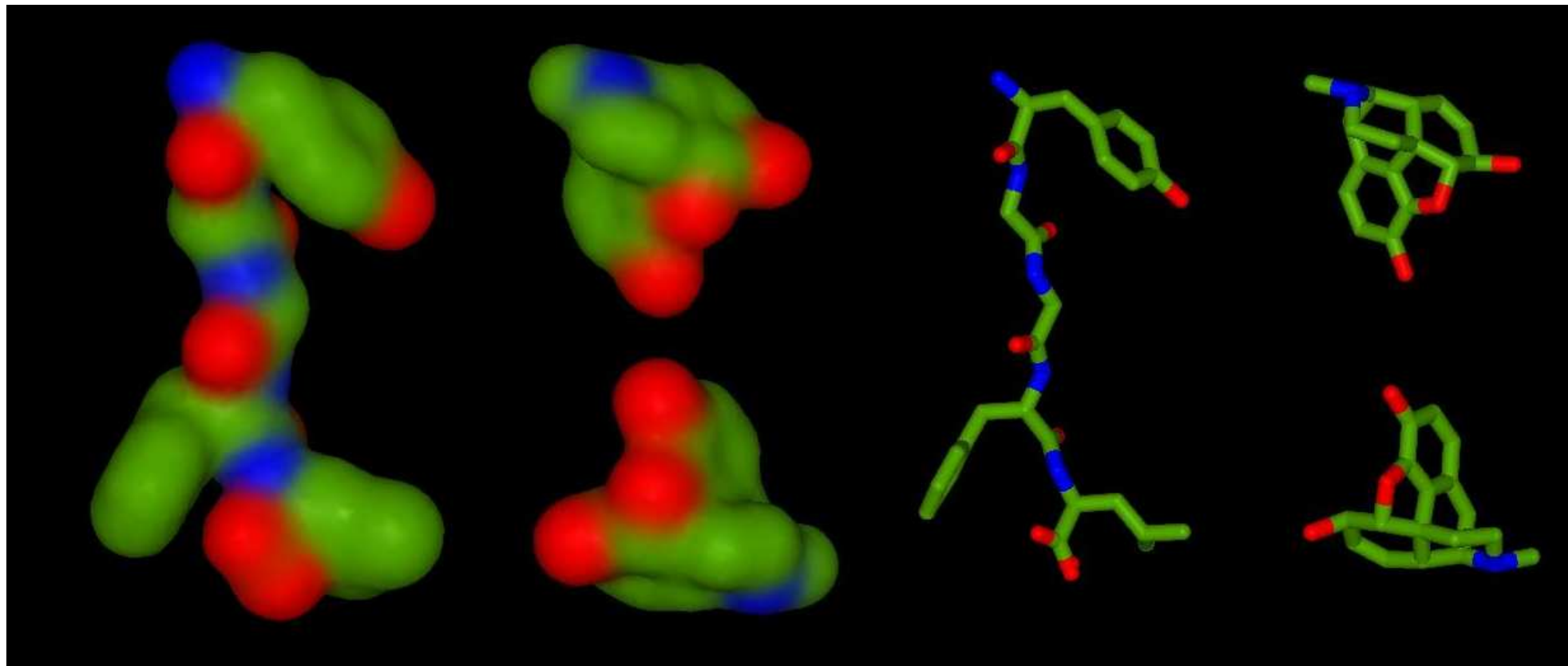
C: Ro15-1788

A-B: 0.95

B-C: 0.87

A-C: 0.85

3D Partial Shape Matching Example: Leu-Enkephalin/Morphine



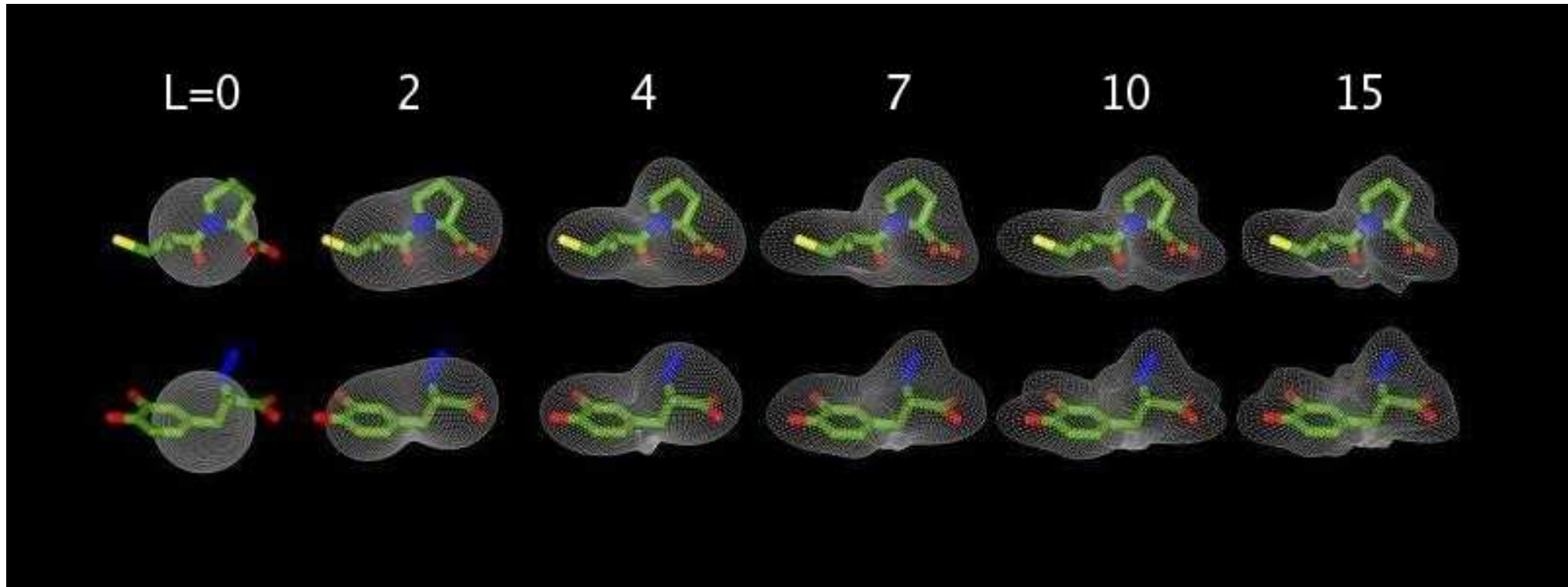
- Top (N-TYR): 0.57; Bottom (C-LEU): 0.60
- $\sim 10^8$ trial orientations: 25 seconds on 1.8GHz PIII Xeon

Fast 2D Surface Envelope Matching

(Ritchie & Kemp (1999) J Comp Chem 20 383–395)

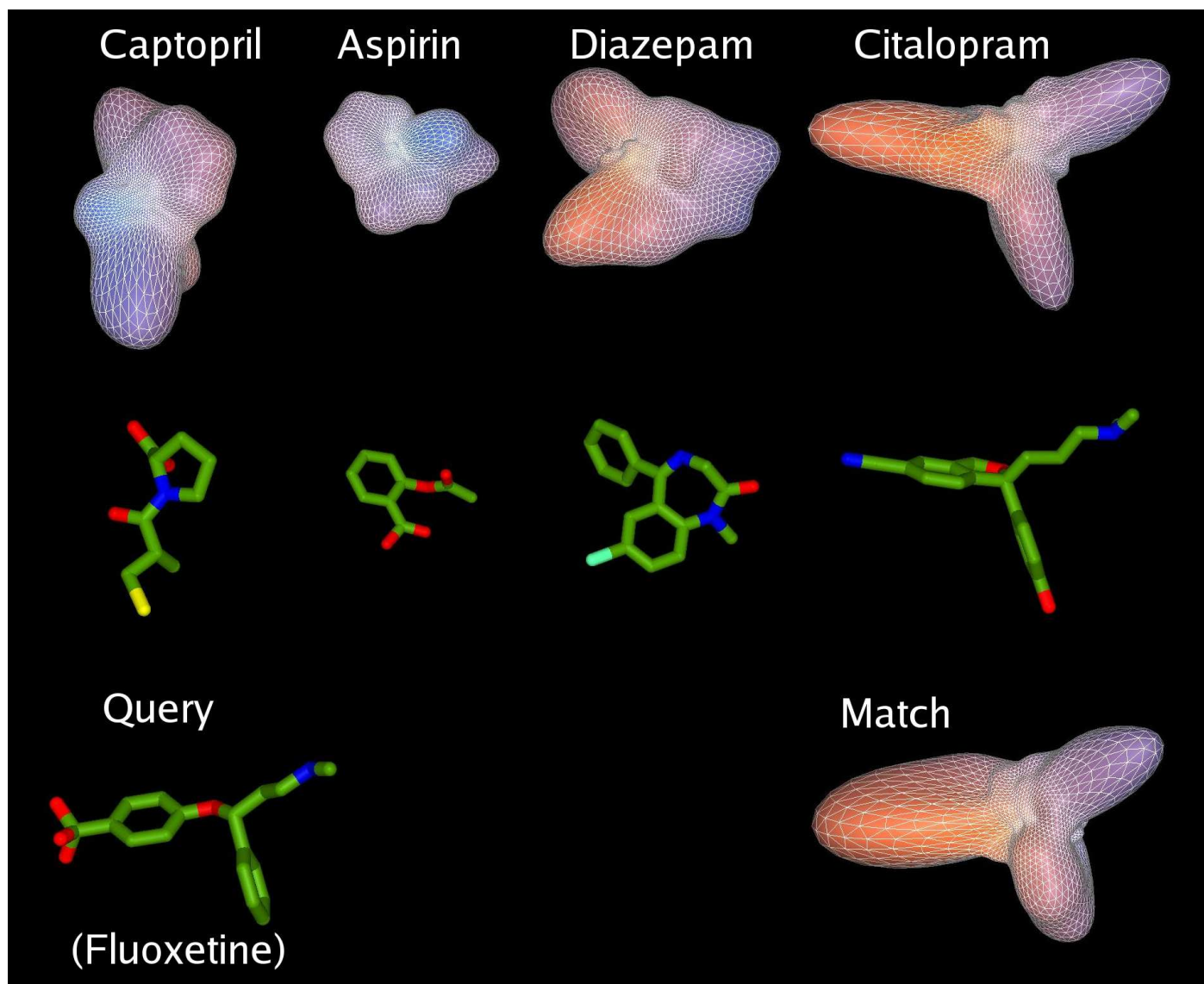
- 2D surface comparisons are much faster than 3D:

$$S_{AB} = \int |r_A(\theta, \phi) - r_B(\theta, \phi)|^2 d\Omega$$



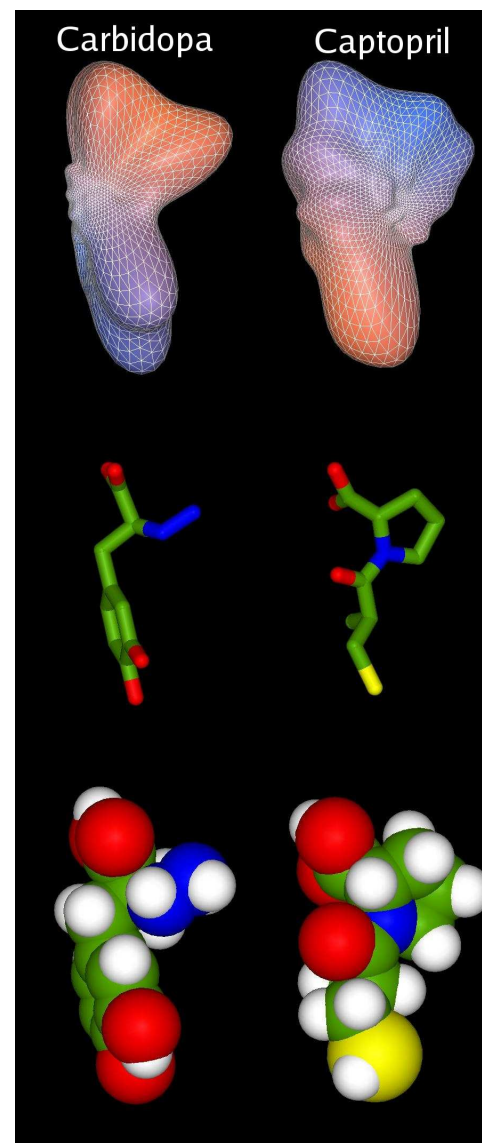
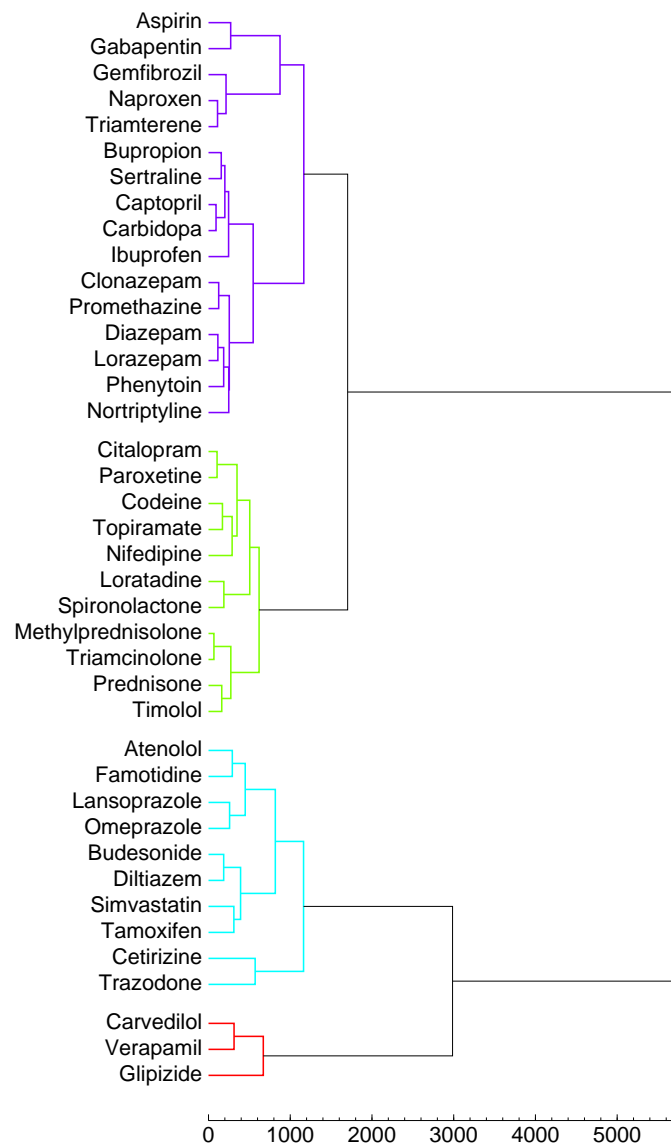
- Expansions to $L=7$ (64 coeffs) take ~ 0.05 s per superposition...

Database Screening by 2D Surface Matching



- Currently building a MySQL DB of $\sim 250,000$ ligands from NCI

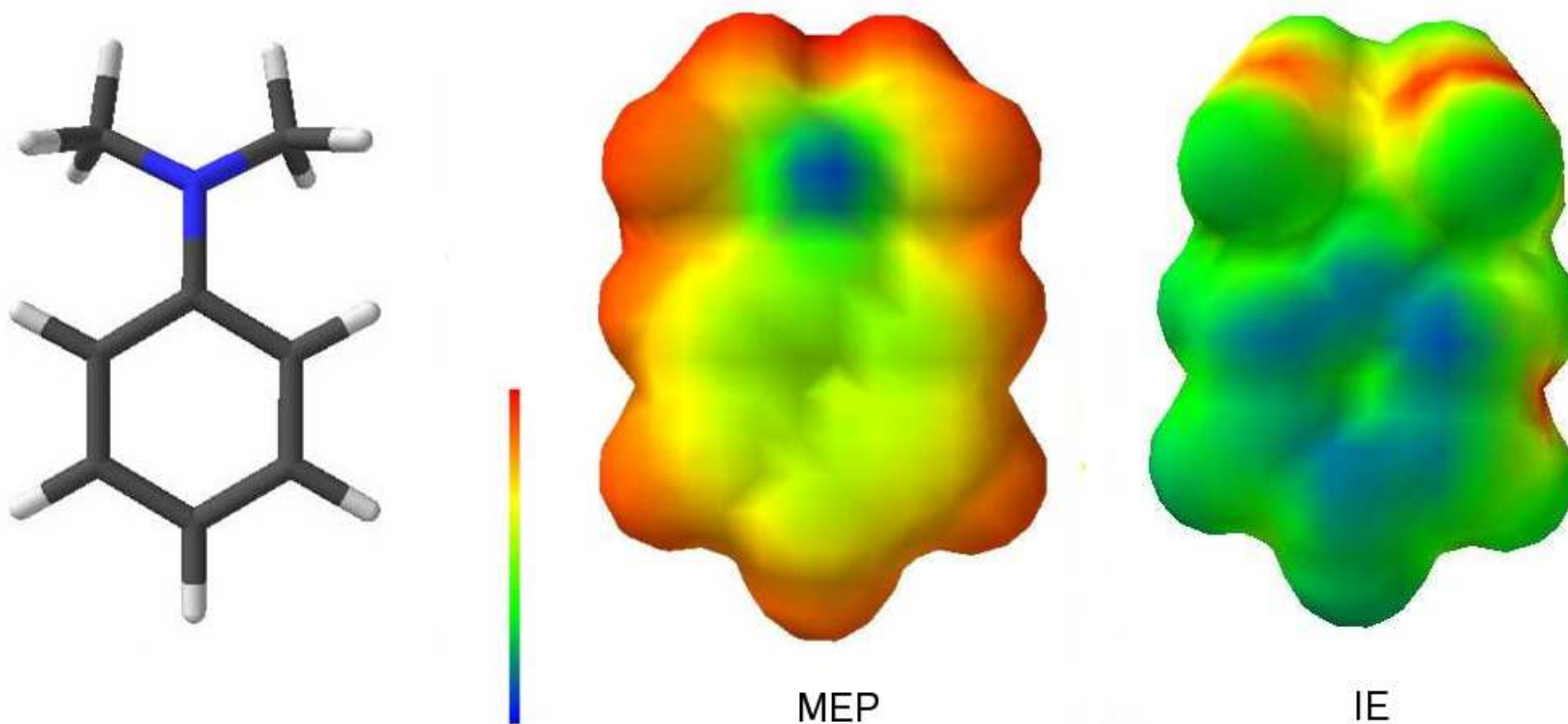
Clustering Drug Molecules by 2D Surface Shape



Spherical Harmonic Local Surface Properties

(Lin & Clark (2005) J Chem Inf Model 45 1010–1016; Clark (2004) J Mol graph 22 519–525)

- From semi-empirical QM (Mopac/Vamp), calculate 4 descriptors:
- MEP, IE_L , EA_L , α_L as expansions to $L=15$ (Parasurf & Parafit)



- Concise/convenient non-atomistic descriptors for ComFA/QSAR?

Future Prospects & Conclusions

- **Protein Docking (“Hex”):**
 - **Novel, fast, & fairly accurate docking algorithm**
- **Small-Molecule Applications:**
 - **Fast 3D shape/field matching...**
 - **Very fast 2D surface shape matching...**
 - **Fast 3D database screening now feasible...**
 - **Extensible to ComFA/QSAR...?**
 - **...and 3D pharmacophore searching & ligand docking...?**

Acknowledgments

Lazaros Mavridis
(EPSRC)

Brian Hudson
Tim Clark
Martyn Ford
Jon Essex

Software & Preprints: <http://www.csd.abdn.ac.uk/hex/>