Can You Hear the Shape of a Drug?

Rachael Pirie

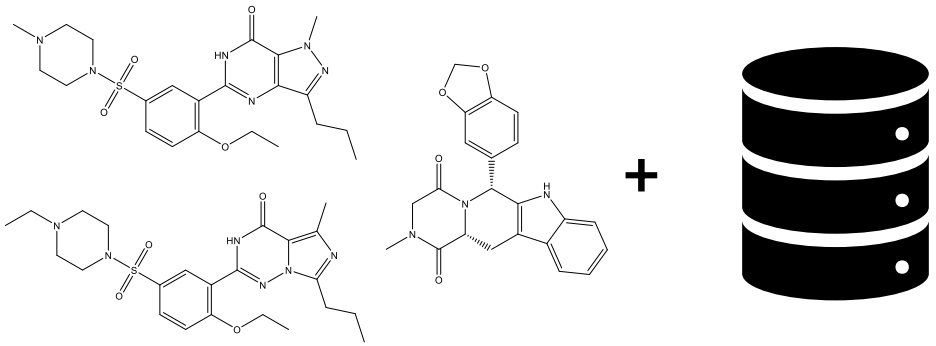
4th Year PhD Student, Cole Group

r.pirie2@newcastle.ac.uk / @rachaelpirie203

RDKit UGM 2022

Similar Property Principle

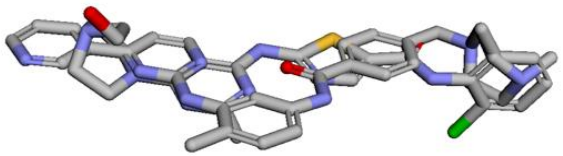
Alternative: Ligand-based Virtual Screening



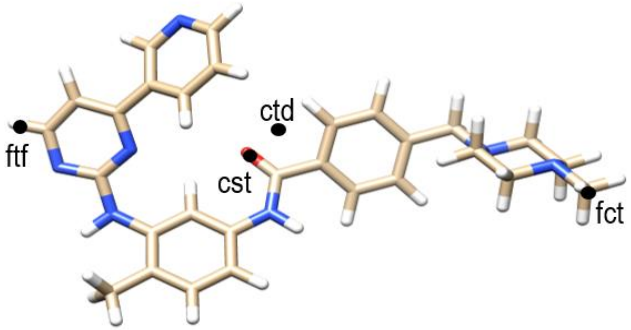
Hits suggested for further development

Similar molecules are likely to display similar properties, including biological activity – *Johnson and Maggiora, 1990*

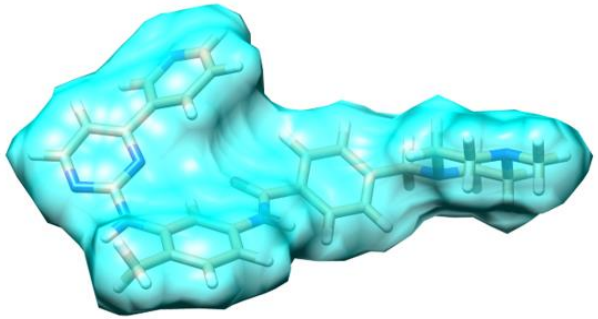
Molecular Shape Similarity



A
Overlap of Volumes
e.g. ROCS



B
Distribution of Atomic Distances
e.g. USR/USRCAT



C
Molecular Surfaces
e.g. Spherical Harmonics

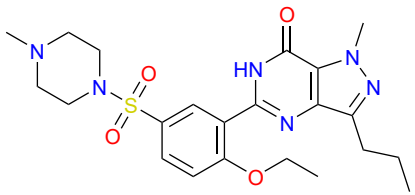
Hearing the Shape of a Drum

The frequencies of vibration (in the form of the eigenvalues of the Laplacian) of a drum determine its shape (Kac 1966)

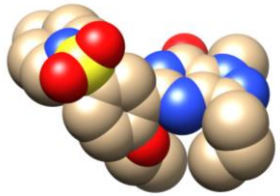
$$\Delta\varphi = \lambda\varphi$$

Can we hear the shape of a drug?

Approach 1: Riemannian Geometry for Molecular Surface Approximation (RGMolSA)



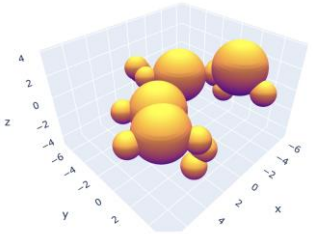
e.g. Sildenafil



Start with CPK Model

No. Atoms
3D Coordinates of Atoms
VDW Radii
Adjacency Matrix

Construct using RDKit



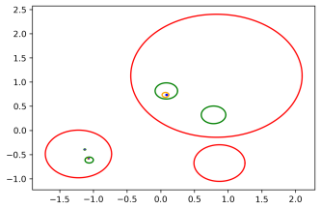
Rings → Spheres

```
1
[[0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [1 0 1 0 0 0 0 2 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 2]
 [0 0 0 1 0 1 0 0 0 2 2 0 0 1 0 0]
 [0 0 0 0 2 0 0 0 0 0 0 2 0 0 2 0]]
```

Pick Base Sphere

$$A = 2\pi \sum_i (2r_i^2 - r_i \sum_j |r_i - \lambda_{ij}|)$$

Calculate Surface Area



Map into Complex Plane

$$g = \begin{cases} \frac{4r_B^2}{(1+|z|^2)^2}(dx^2+dy^2) & \text{if } z \in \mathcal{C} \\ \frac{C_1}{(|z-A_1|^2+B_1)^2}(dx^2+dy^2) & \text{if } z \in \mathbb{D}(a_1, R_1) \\ \vdots & \vdots \\ \frac{C_{N-1}}{(|z-A_{N-1}|^2+B_{N-1})^2}(dx^2+dy^2) & \text{if } z \in \mathbb{D}(a_{N-1}, R_{N-1}) \end{cases}$$

Riemannian Metric (Symmetric Form)

$$\Delta f := -\frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x_i}\left(\sqrt{|g|} g^{ij} \frac{\partial f}{\partial x_j}\right)$$

Spectrum of the Laplacian

$$v_{S,k} = \left({}^{10^4}/_{A}, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_9\right)$$

9-Element Descriptor

$$d(u,v) = 1 - \frac{\sum |u_i - v_i|}{\sum |u_i + v_i|}$$

Bray-Curtis Distance

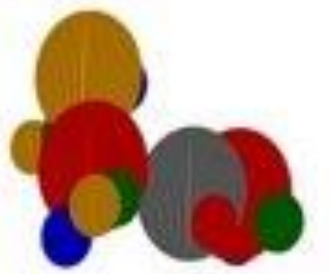
Approach 1: Riemannian Geometry for Molecular Surface Approximation (RGMolSA)

Pros:

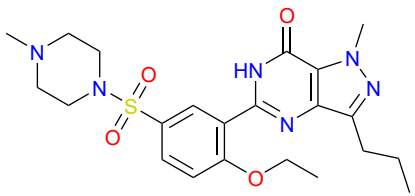
- Rotation and Translation Invariant
- Alignment Free
- Quick to Calculate
- Insensitive to small deformations of the surface

Cons:

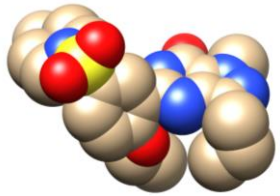
- Dependent on the initial choice of starting atom
- For long chain molecules: introduces numerical error
- Cannot describe Macrocycles



Approach 2: Kähler Quantisation for Molecular Surface Approximation (KQMolSA)



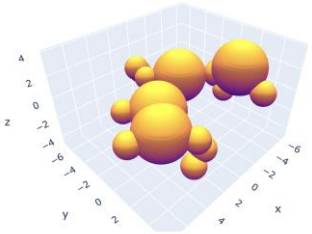
e.g. Sildenafil



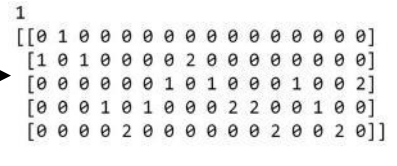
Start with CPK Model

No. Atoms
3D Coordinates of Atoms
VDW Radii
Adjacency Matrix

Construct using RDKit



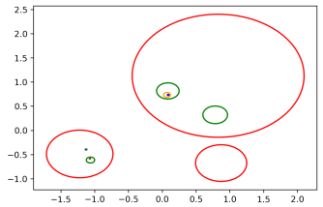
Rings → Spheres



Pick Base Sphere

$$A = 2\pi \sum_i (2r_i^2 - r_i \sum_j |r_i - \lambda_{ij}|)$$

Calculate Surface Area



Map into Complex Plane

$$F(z) = \begin{cases} \frac{2r_B^2}{(1+|z|^2)^2} & \text{if } z \in \mathcal{C} \\ \frac{C_1}{(|z-A_1|^2+B_1)^2} & \text{if } z \in \mathbb{D}(a_1, R_1) \\ \vdots & \vdots \\ \frac{C_{N-1}}{(|z-A_{N-1}|^2+B_{N-1})^2} & \text{if } z \in \mathbb{D}(a_{N-1}, R_{N-1}) \end{cases}$$

Riemannian Metric (Antisymmetric Form)

$$\mathbb{M}_{ij} = \iint_{\mathbb{C}} z^i \bar{z}^j e^{-k\varphi} F(z) \sqrt{-1} dz \wedge d\bar{z}$$

Construct Hermitian Matrix

$$\mathbb{M} = \begin{bmatrix} 2.159 + 0j & -2.435 + 0.855j & 2.407 - 1.928j \\ -2.435 - 0.855j & 3.085 + 0j & -3.479 + 1.221j \\ 2.407 + 1.928j & -3.4786 - 1.221j & 4.407 + 0j \end{bmatrix}$$

3x3 Matrix Descriptor

$$S(\mathbb{M}_1, \mathbb{M}_2) = 0.3 \frac{A_{min}}{A_{max}} + 0.7 \frac{1}{1 + d(\mathbb{M}_1, \mathbb{M}_2)}$$

Similarity Score

Approach 2: Kähler Quantisation for Molecular Surface Approximation (KQMolSA)

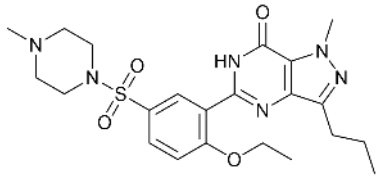
Pros:

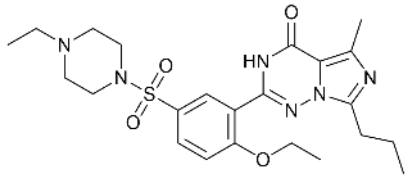
- Alignment Free
- Quick to Calculate
- Insensitive to small deformations of the surface
- Less dependent on choice of starting atom

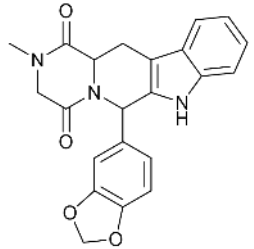
Cons:

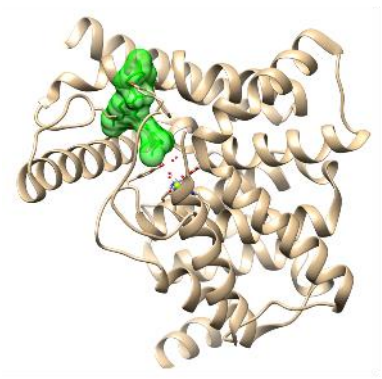
- Small dependence on initial position
- Slow to calculate distance between molecules
- Numerically Instable > k=2

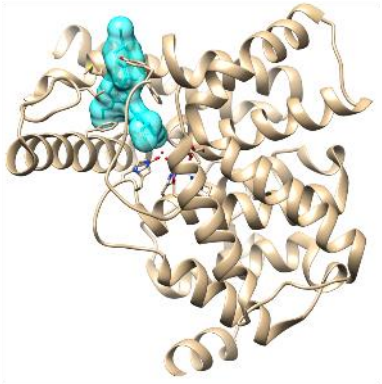
Case Study: PDE5 Inhibitors

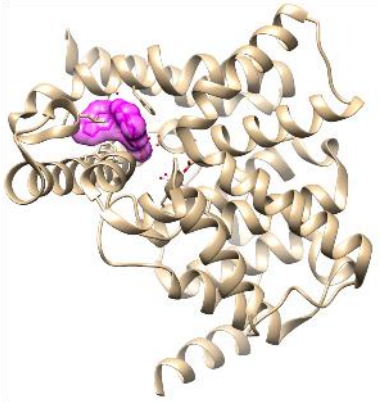












Sildenafil
Pfizer
1998

Vardenafil
Bayer/GSK
2003

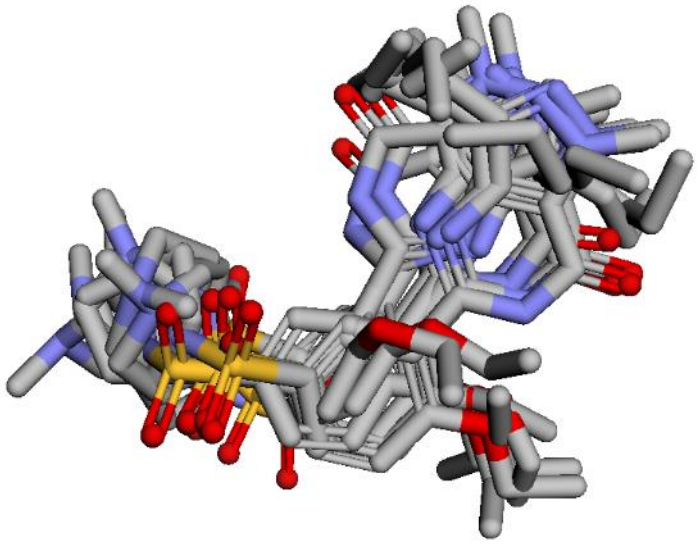
Tadalafil
Lilly
2003

Case Study: Comparison to Other Methods

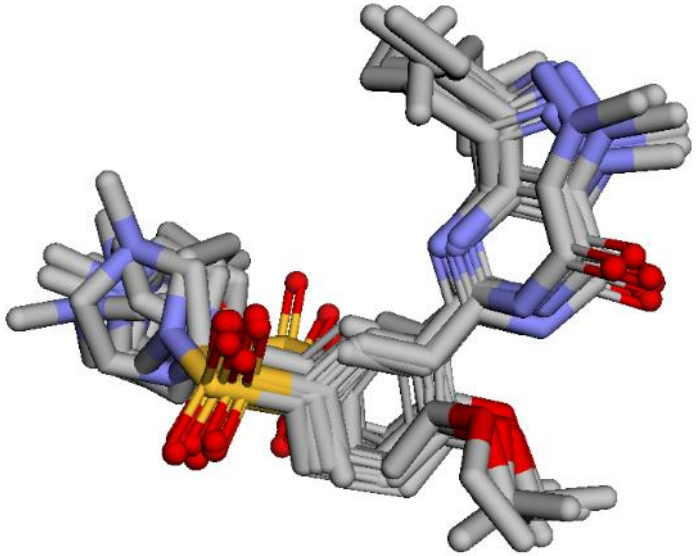
	RGMolSA	KQMolSA (k=1)	KQMolS (k=2)	USRCAT	Shape-It	MolSG
Sildenafil - Vardenafil	0.955	0.907	0.652	0.384	0.388	0.704
Sildenafil - Tadalafil	0.886	0.449	0.482	0.269	0.278	0.746
Vardenafil - Tadalafil	0.844	0.432	0.47	0.291	0.353	0.887

Scores: 0 (completely different) and 1 (identical); typical threshold >= 0.7 to be “similar”

Case Study: Consideration of Conformers



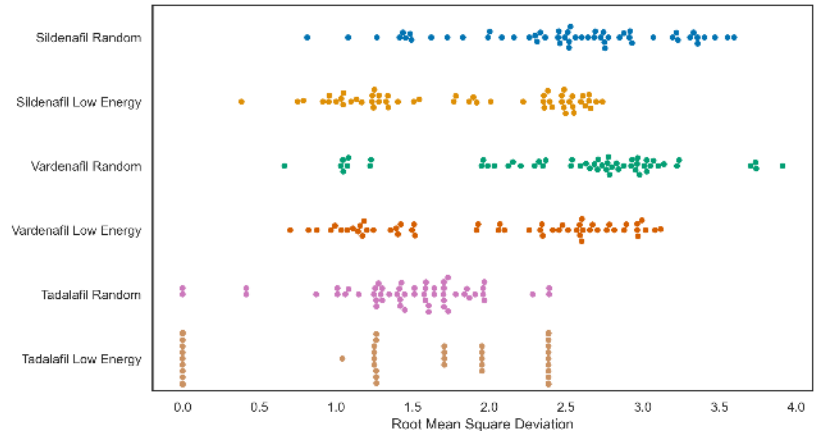
Sildenafil 10 Random Conformers

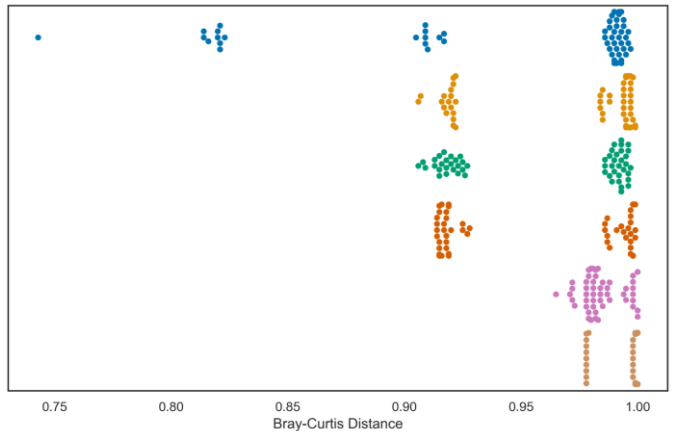


Sildenafil 10 Low Energy Conformers

Case Study: Consideration of Conformers

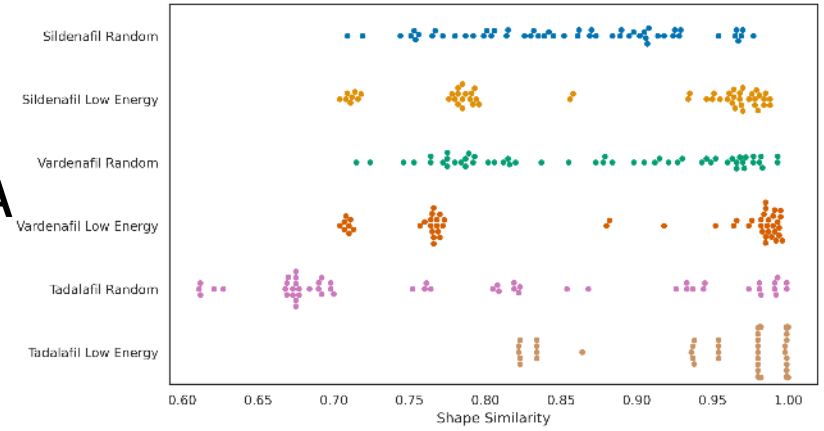
RMSD

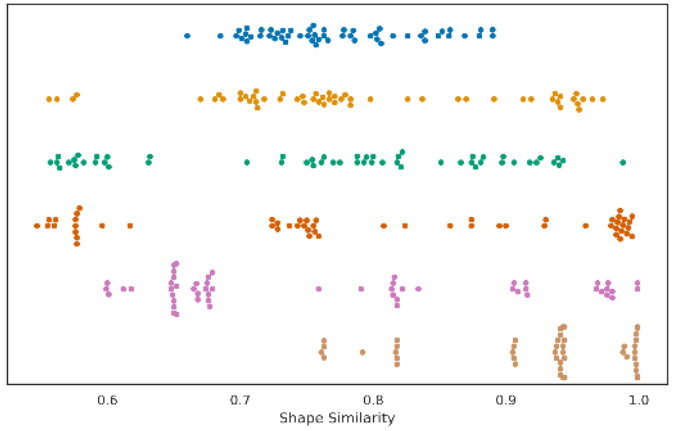




RGMolSA

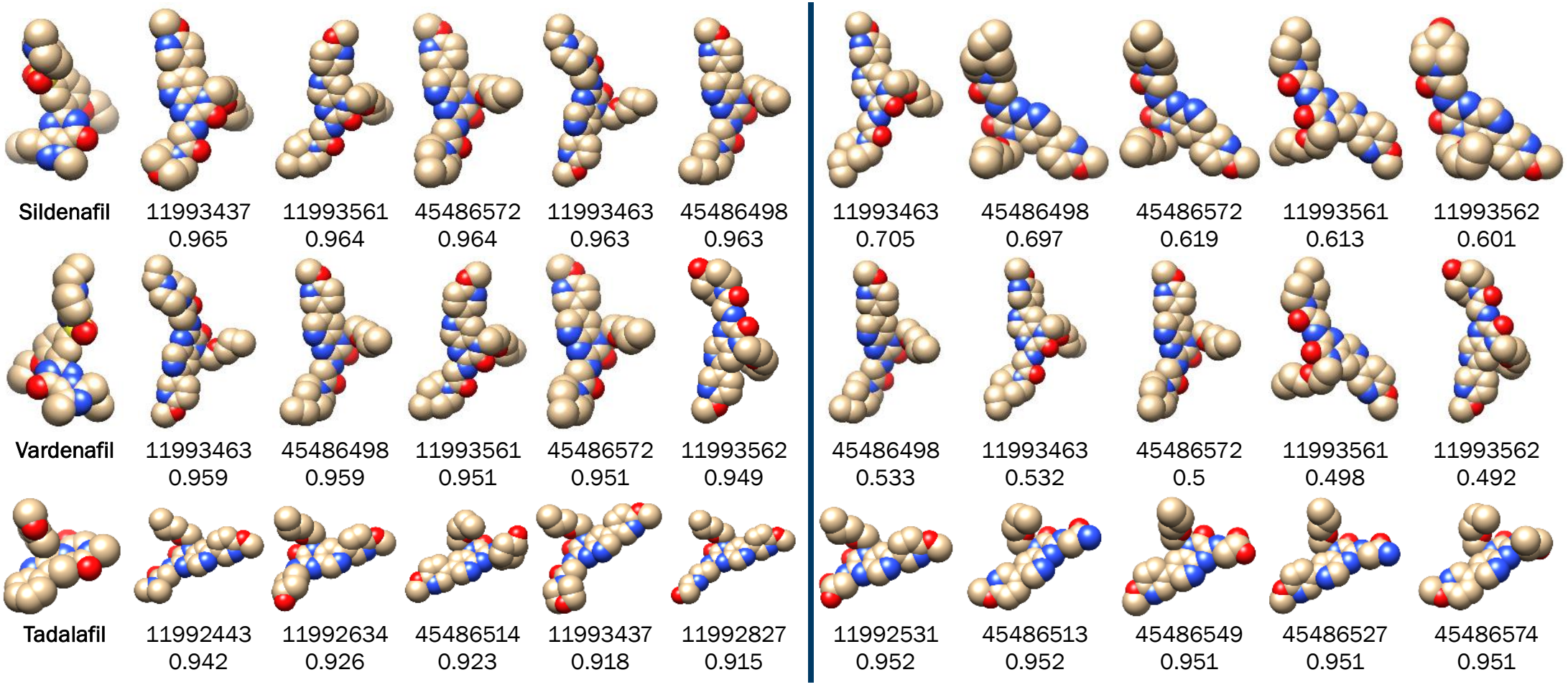
KQMolSA k=1





KQMolSA k=2

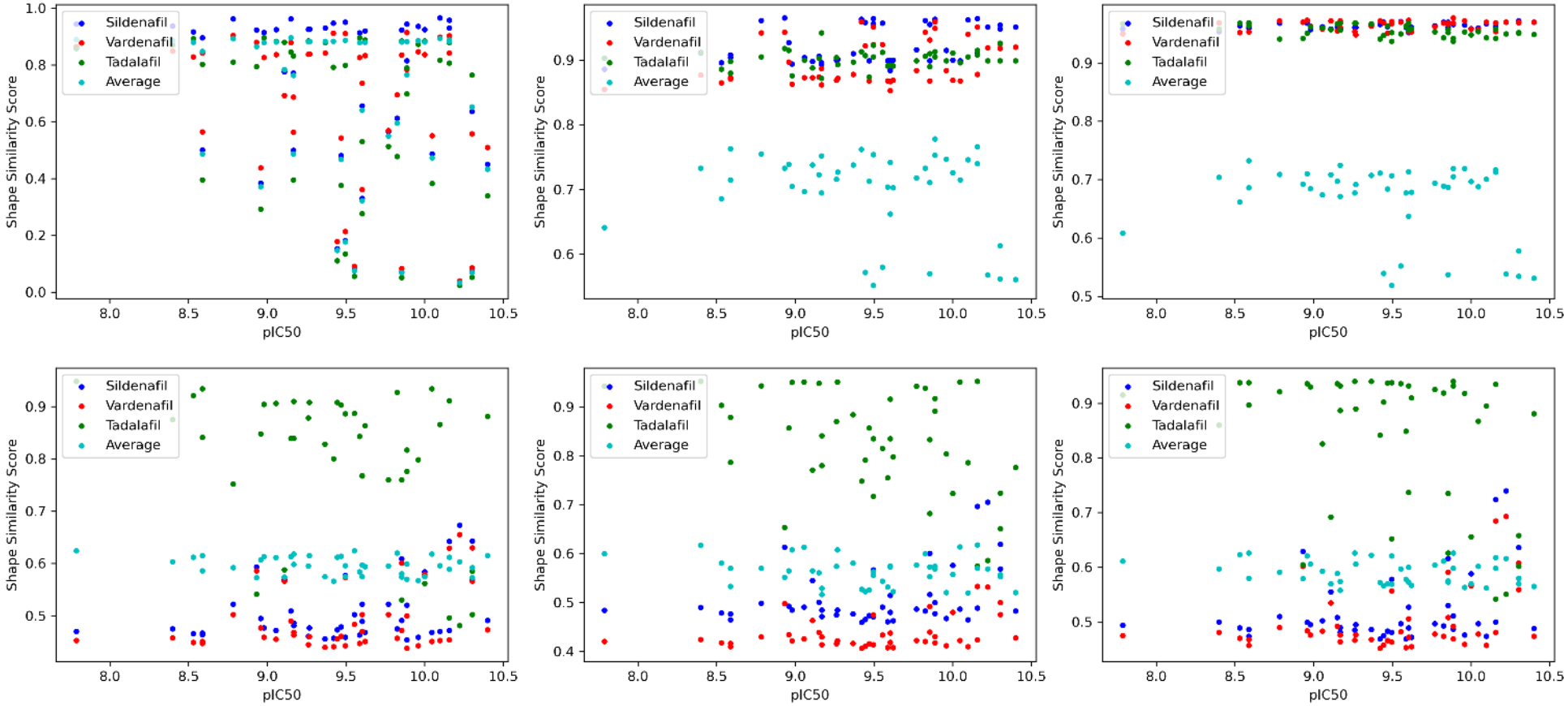
Case Study: Other PDE5 Inhibitors



Case Study: Other PDE5 Inhibitors

RGMolSA

KQMolSA

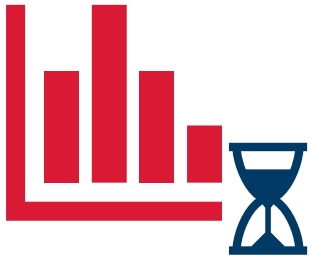


Single Random Conformer

Crystal Queries, Multiple Test Conformers

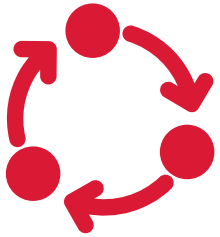
Multiple Conformers

Future Work



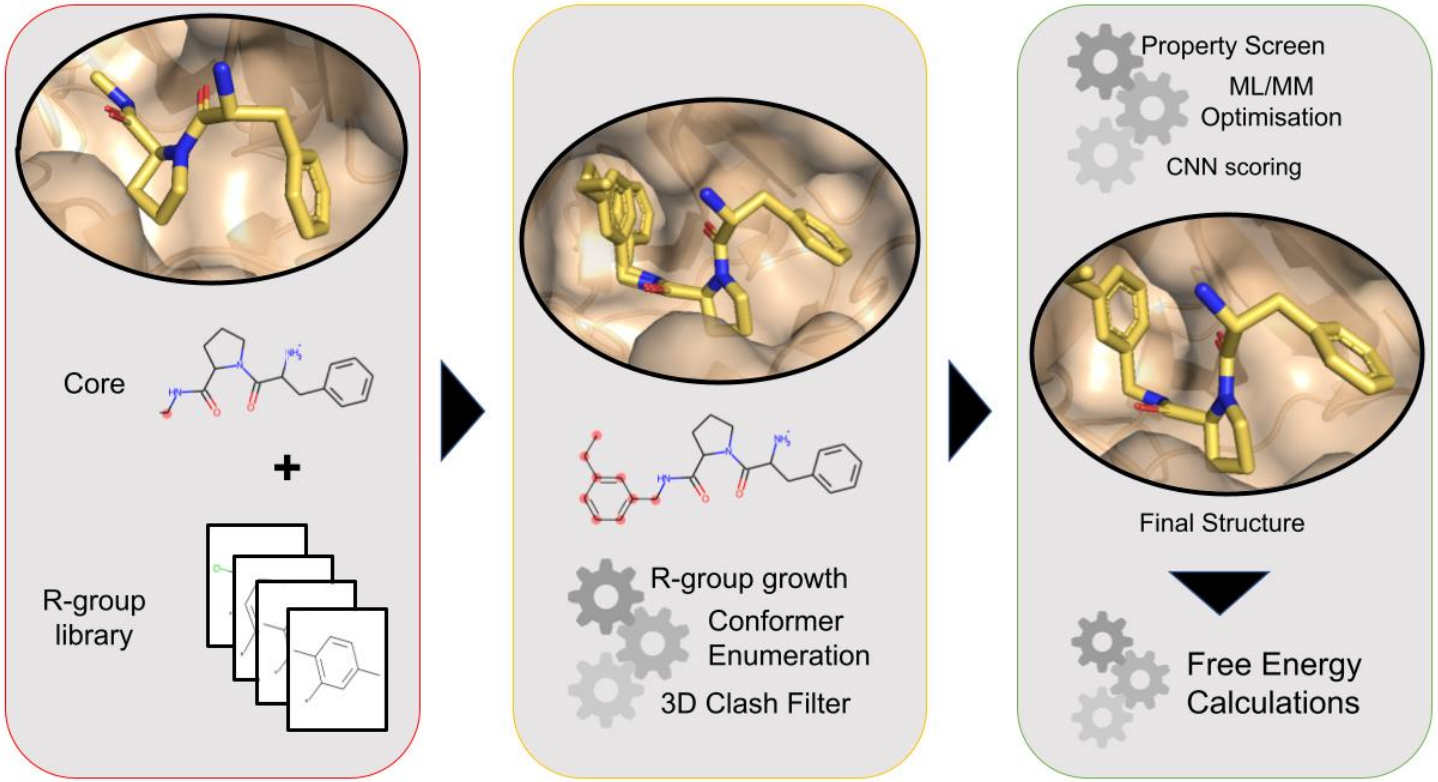
Full Validation and Benchmarking

Application to Real Drug Discovery Projects

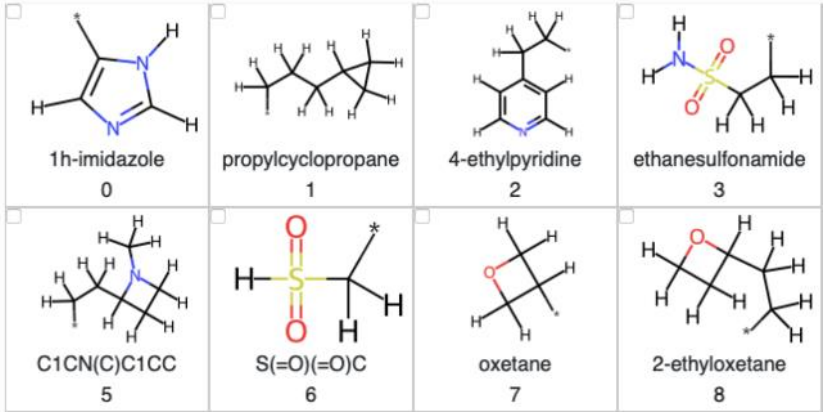


Consideration of Pharmacophoric Features

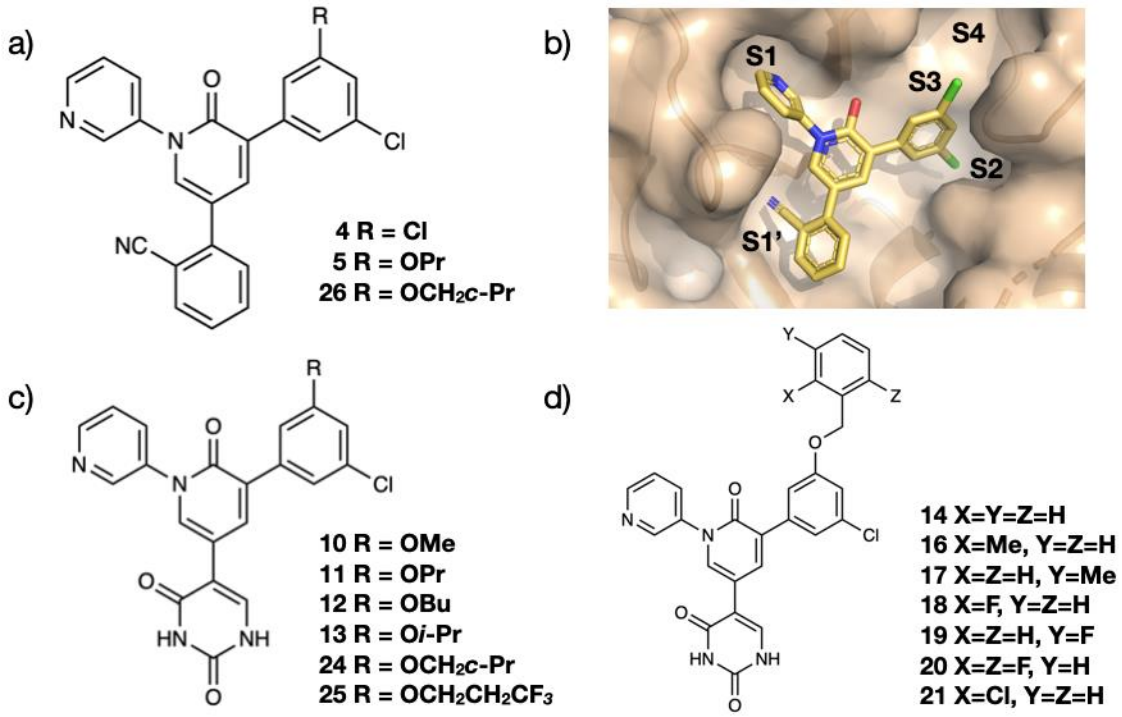
FEGrow

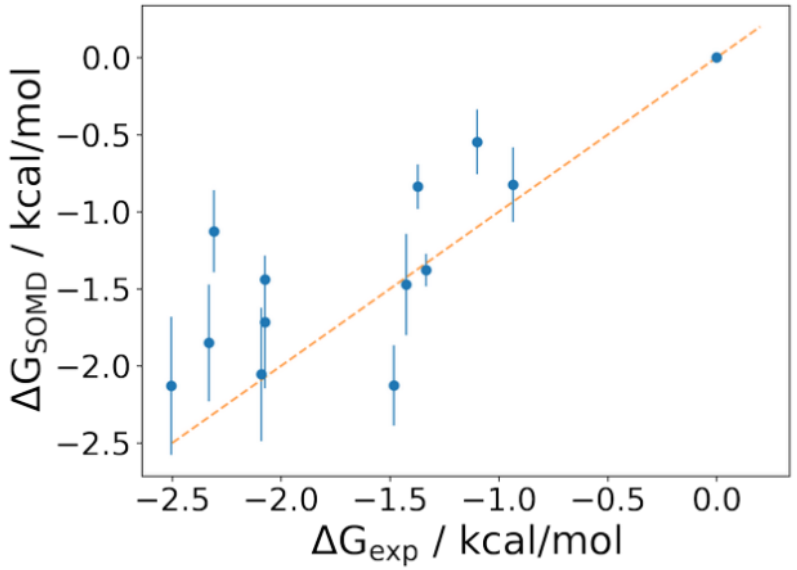


	MW	HBA	HBD	LogP	Pass_Ro5	has_pains	has_unwanted_subs	has_prob_fgs	synthetic_accessibility
0	413.093104	5	0	5.10018	True	False	False	True	7.223150
0	427.108754	5	1	4.62638	True	False	False	True	7.501563
0	423.113840	4	0	5.96898	True	False	False	True	7.482970
0	477.135638	6	0	5.53078	True	False	False	True	7.719800
0	477.135638	6	0	5.53078	True	False	False	True	7.716155
0	494.187339	5	0	6.11838	True	False	False	True	7.685201
0	452.140389	5	0	5.29728	True	False	False	True	7.757132
0	463.119988	6	0	5.33638	True	False	False	True	7.523968



FEGrow: Case Study - SARS-CoV-2 Main Protease





Summary

Molecular Shape Similarity is a useful tool in the hit-identification stage of a drug discovery campaign.

The mathematical theory of Riemannian geometry can be applied to approximate the shape of molecules using either a 9-element vector (**RGMolSA**) or a 3x3 matrix (**KQMolSA**) of eigenvalues.

Both descriptors are quick to calculate and easy to compare. Initial case study shows promise compared to existing shape similarity approaches, in considering conformers and for virtual screening.

Code for all approaches discussed available on **GitHub.**

Acknowledgements

Supervisors

Dr Daniel Cole

Dr Stuart Hall

Dr Matthew Forshaw

Collaborators

Dr Thomas Murphy (CS Fullerton)

Funding

Engineering and Physical Sciences Research Council (grant: EP/R51309X/1)

Thank You For Listening!

@rachaelpirie203

www.linkedin.com/in/rmepirie

RPirie96

r.pirie2@newcastle.ac.uk

Contact and Publications