

# Molecule properties and compound developability – what's new?

Dr Tim Ritchie

TJR-Chem, Ranco (VA), Italy

<http://www.linkedin.com/in/timritchiechemistry>

# Dr Tim Ritchie - Background



1989 – 2005

Novartis Institute, London UK  
Medicinal chemist, neuroscience area:  
*BKB1, BKB2, nAChR $\alpha$ 7, NK1, TRPV1*



2005 – 2008

GlaxoSmithKline, Stevenage UK  
Medicinal chemistry design expert, respiratory area:  
*Medchem-Compchem interface*  
*CCR4, DP1, PI3K, Syk*



2008 – date

**TJR-Chem, Ranco (VA), Italy**  
Independent chemistry consultant:  
*Data mining, structure-property analysis (rings),*  
*ligand-based drug design, phys-chem tutorials*

# Outline of talk

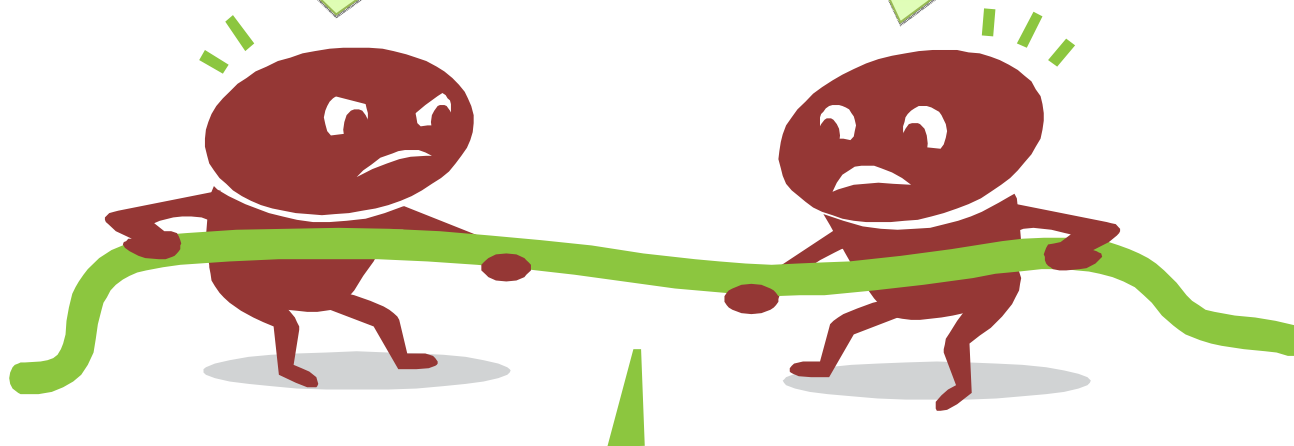
- Molecule properties
- Developability
- Analysis of compounds in different regions of clogP/MW space
  - Developability score from solubility, permeability, protein binding and P450 inhibition data
  - Which properties can best differentiate high/low developability?
- Summary
- Acknowledgements



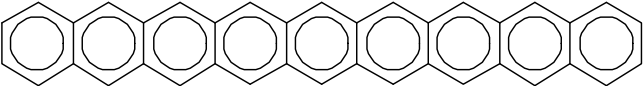
# A balanced view of molecule properties?

Rules are rules: a clogP of 4.99 is OK but not 5.01!

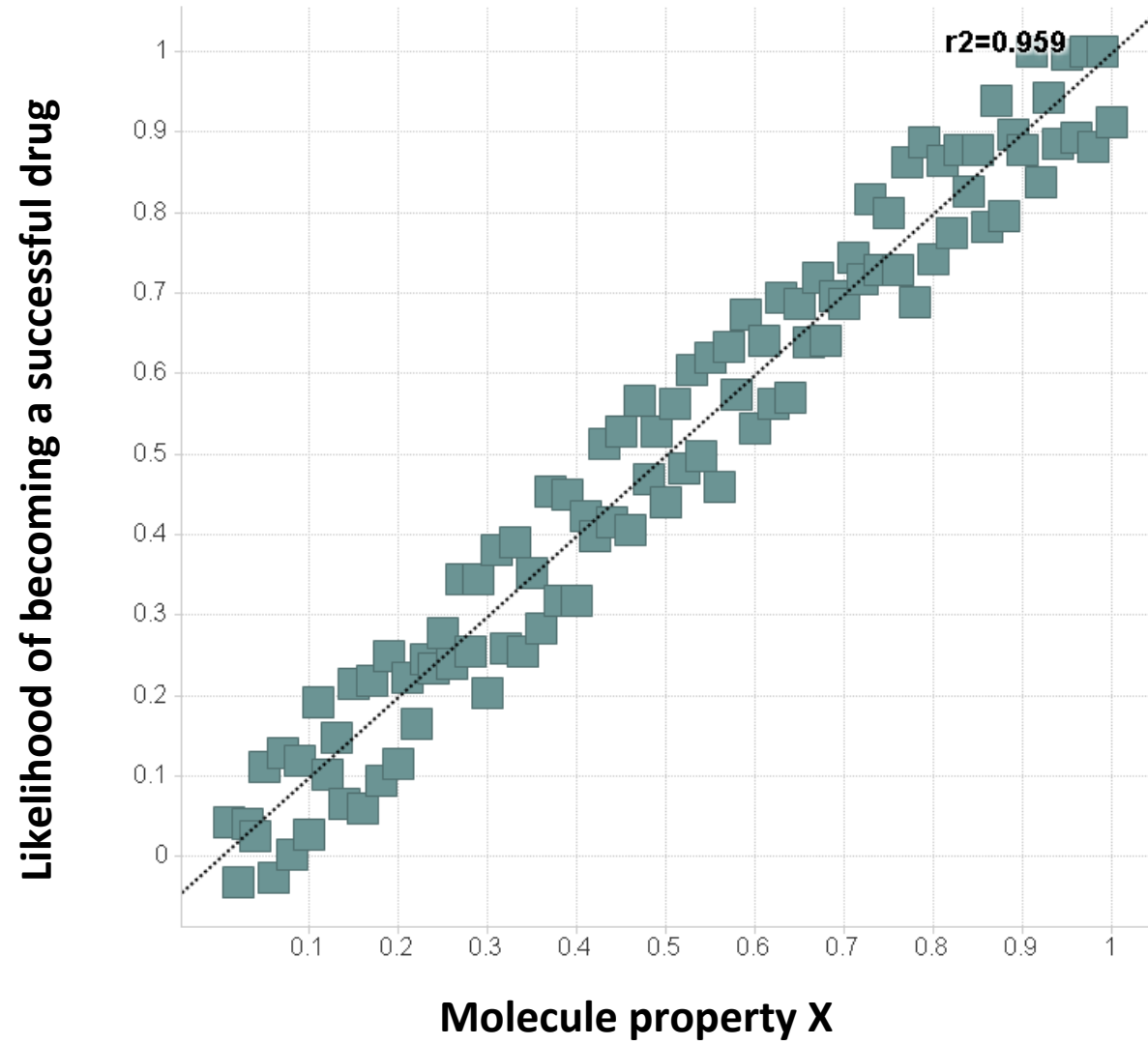
Properties? The ADMET assays tell me what's OK!



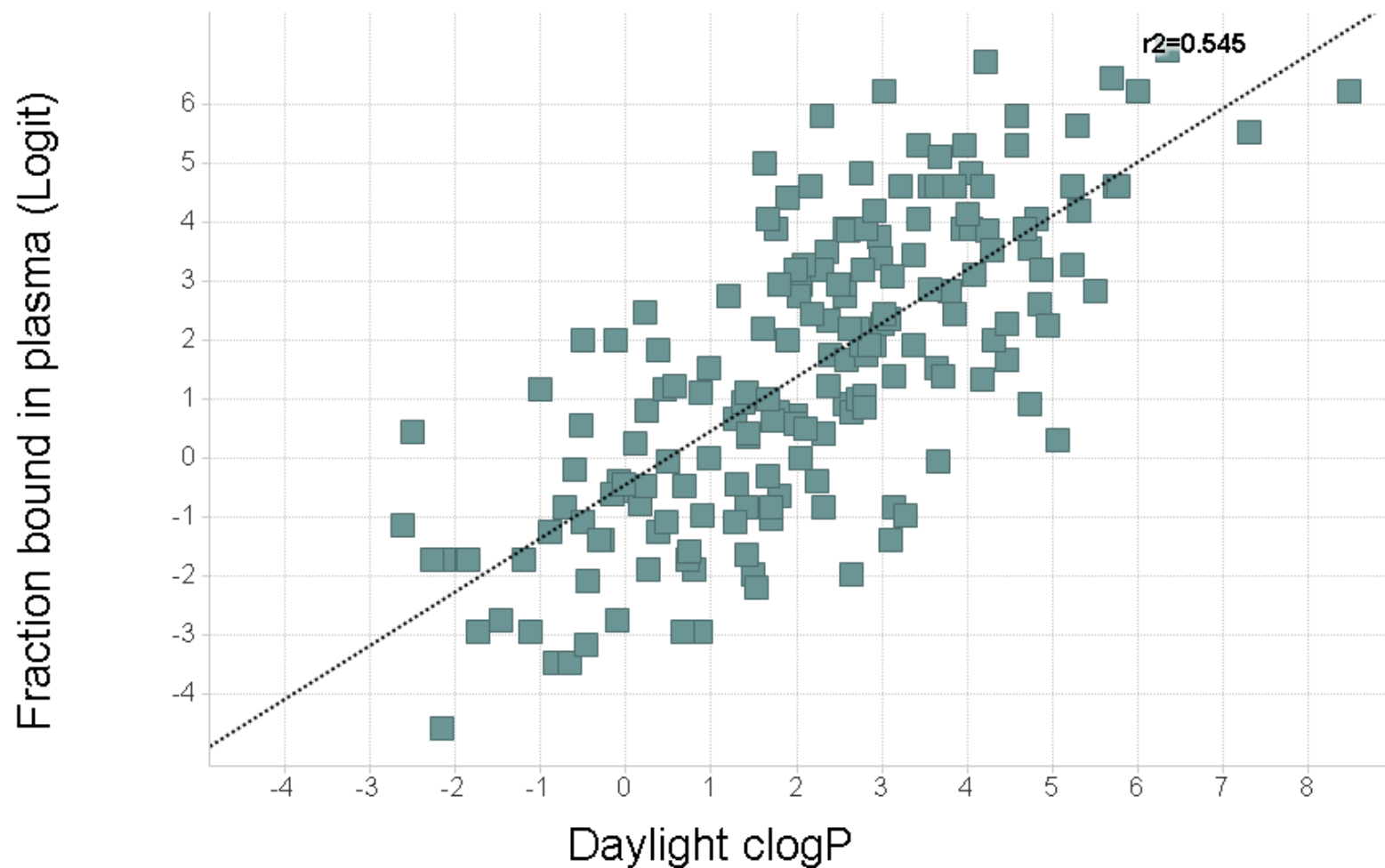
# Scepticism about molecule properties

- Rule-of-Five: help or hindrance?
  - Too lax? 
  - Too strict? – limits creativity, opportunities
  - Applied too widely? – Not a ‘drug-like’ score!
- Do property rules apply to all chemical space?
- Can general ‘global’ models compete with more specific ‘local’ models?
- Drug discovery is complicated: can simple properties really predict anything useful?

# Molecule properties – if only...



# clogP and drug fraction in human plasma

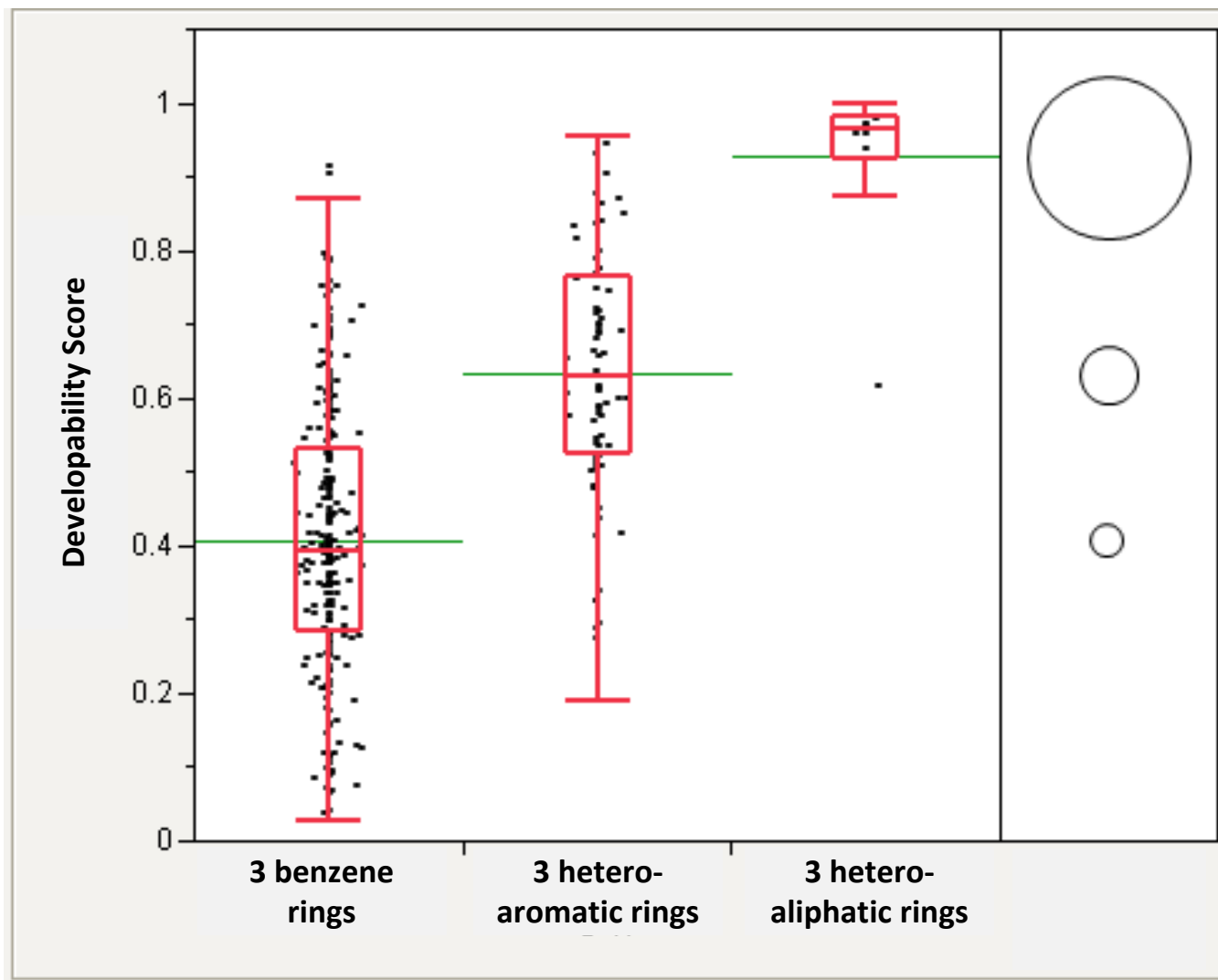


From data published in J Med Chem 2010 53 1098

Tim Ritchie Cambridge 2013

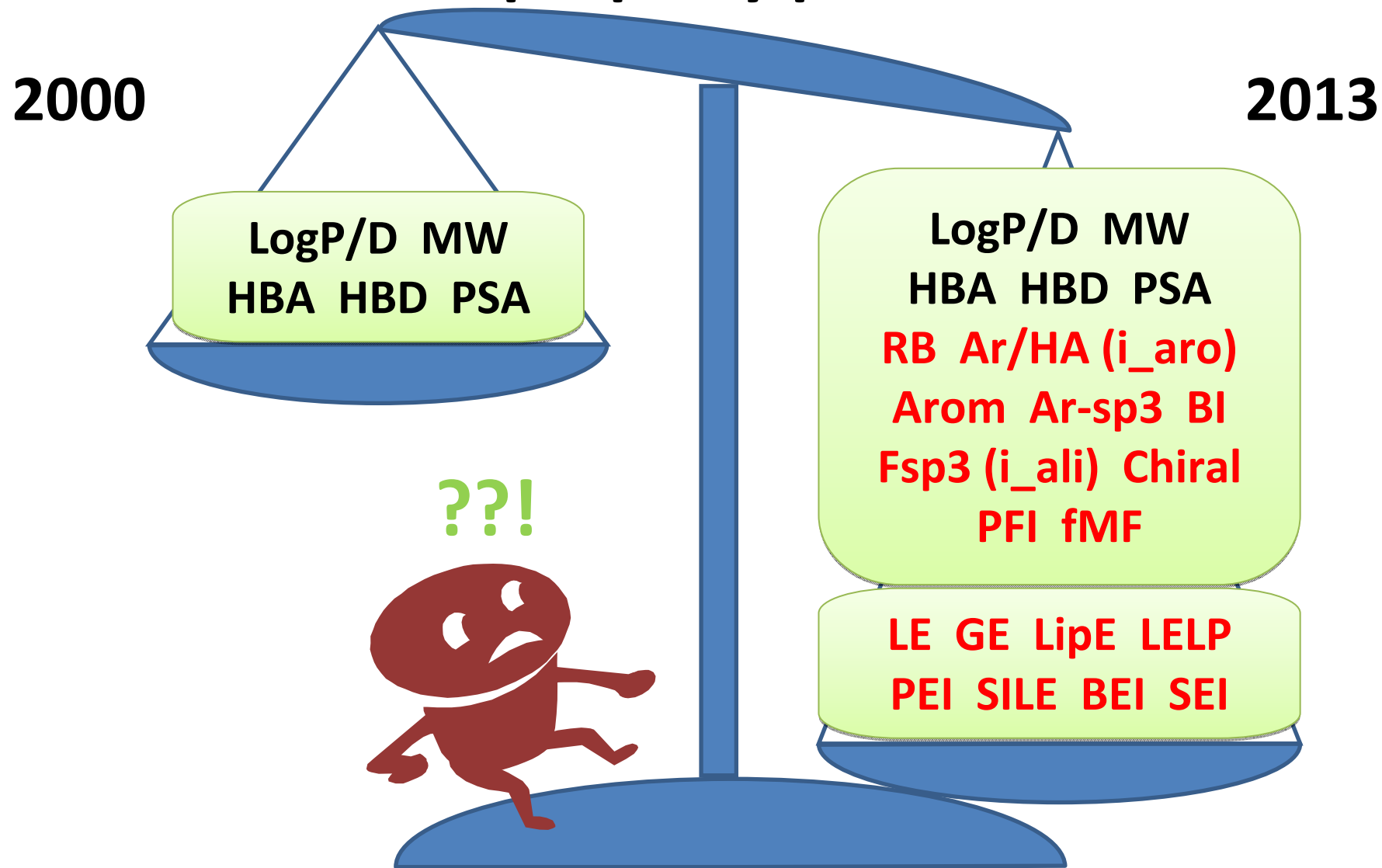
# 'Developability' and ring type

Sol/Perm/PB/3A4 profile for compounds with 3 rings





# Molecule property proliferation



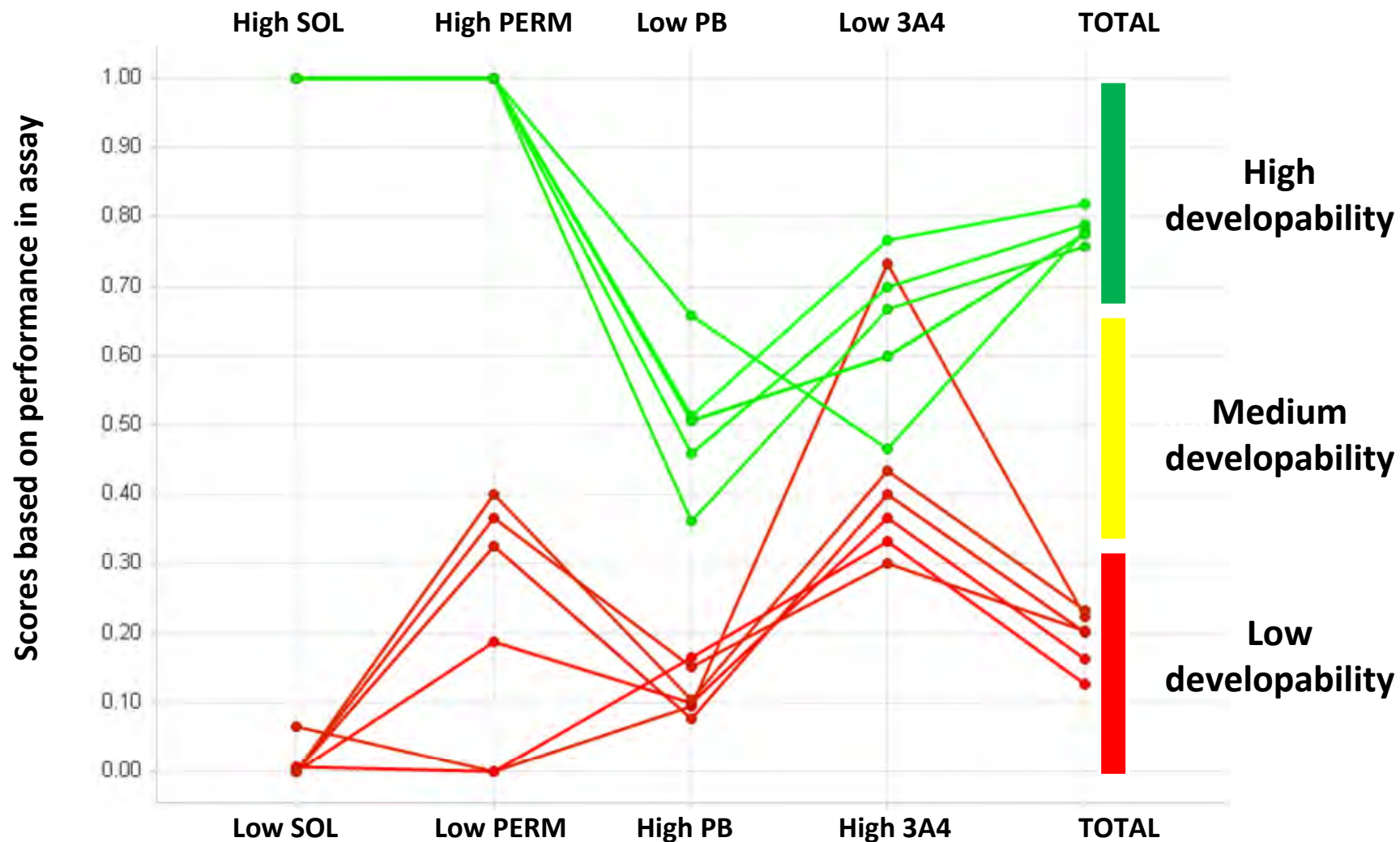
•Do we need all of them? Which are most important?

Tim Ritchie Cambridge 2013

# Compound developability

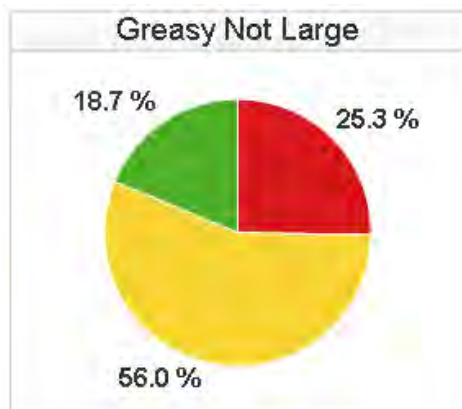
- ‘Developability’ is determined by how well a compound fares in ADME-related assays that are relevant to drug development
- High developability compounds are defined as having:
  - High solubility
  - High permeability
  - Low protein binding
  - Low CYP450 3A4 inhibition

# Compound scoring based on assays

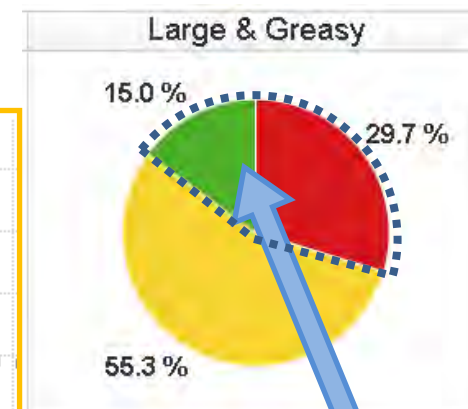
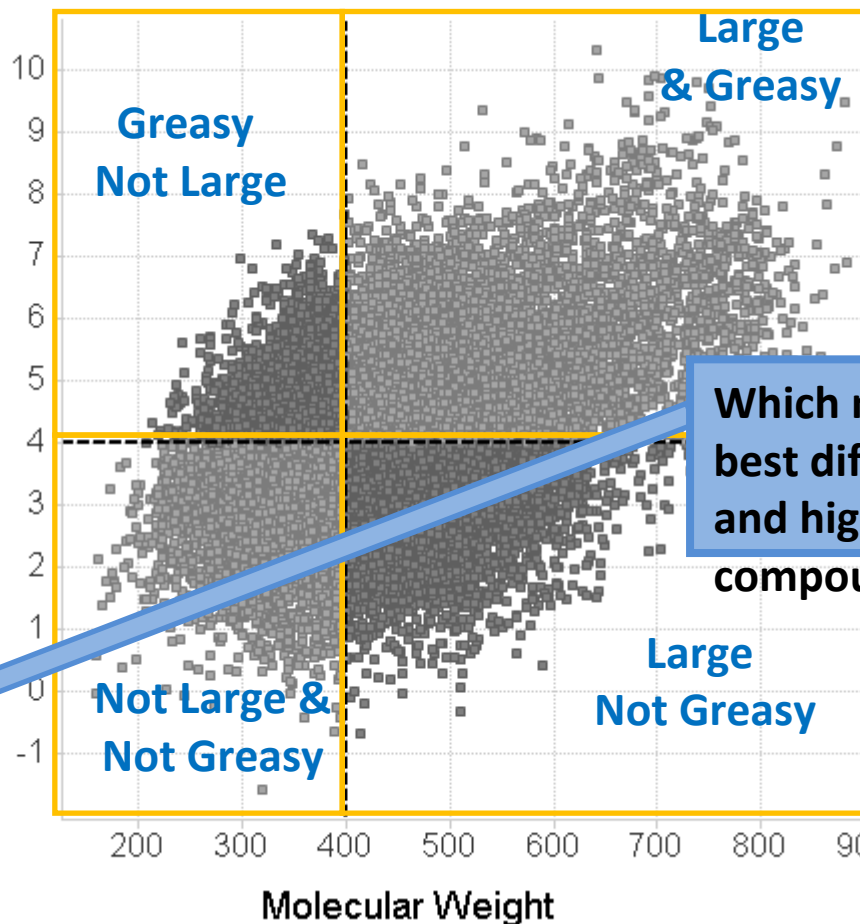


# Developability in clogP/MW quadrants

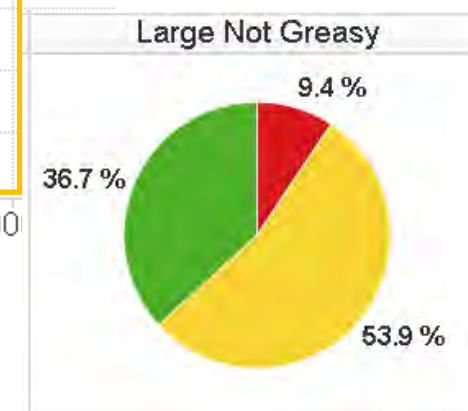
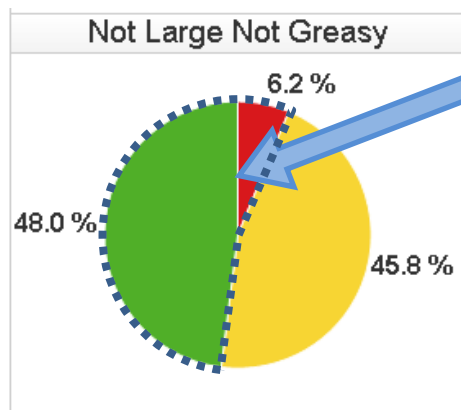
30,000 diverse GSK molecules



Daylight clog P



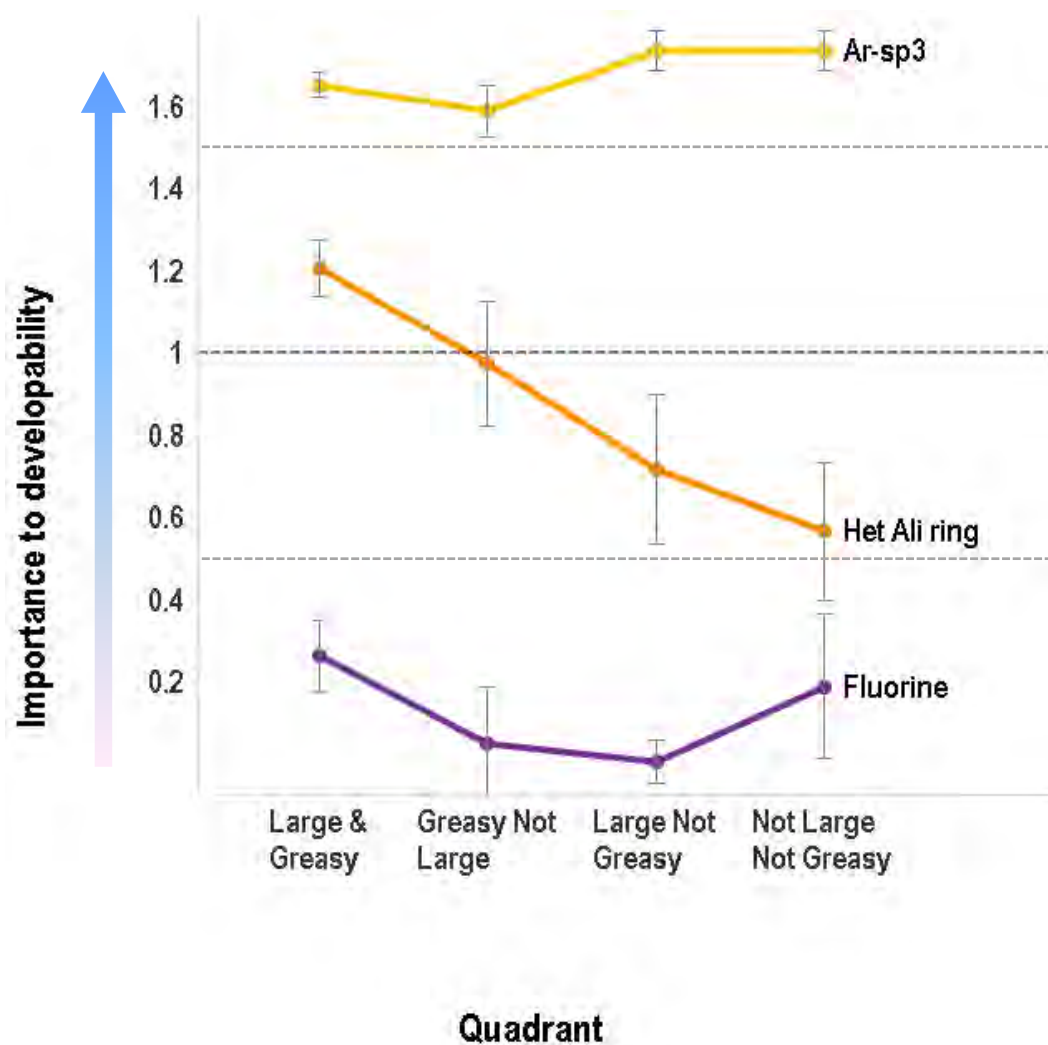
Which molecule properties best differentiate low-scoring and high-scoring compounds?



# Phys-chem properties and developability

- 40 molecule properties calculated
  - Aromatic, aliphatic, lipophilic, polar, ionisable, atom counts, flexibility, size
- *O*-PLS regression models generated for each clogP/MW quadrant
  - Separates variation in properties that is correlated with developability from variability that is uncorrelated (orthogonal)
  - Importance of each property in differentiating high/low developability determined for each quadrant
- **Results indicate that:**
  - **Some properties are important in all quadrants**
  - **Some properties vary in importance depending on the quadrant**
  - **Others are less important to developability**

# Examples of property importance\*



**Ar-sp3** has high importance to developability across all quadrants

**Heteroaliphatic ring count** importance to developability varies with quadrant

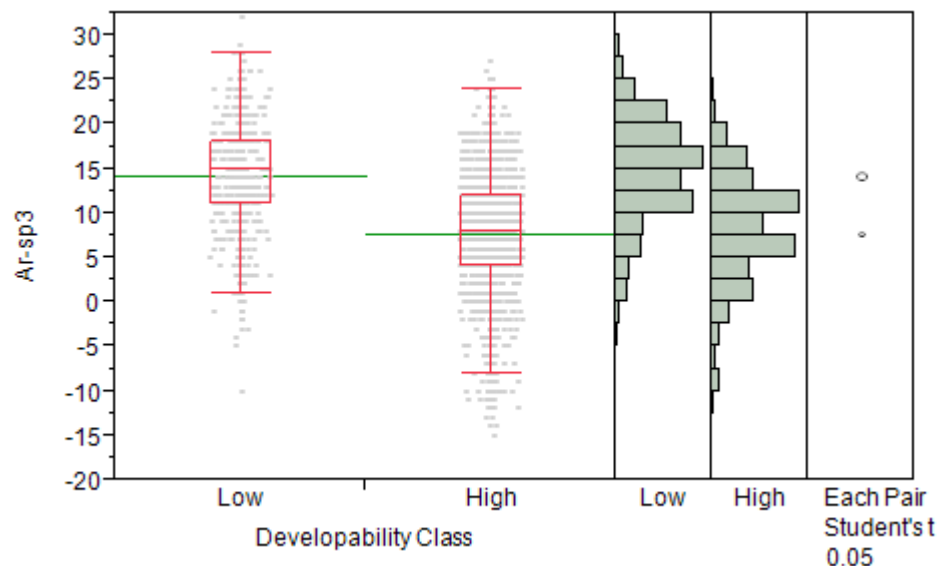
**Fluorine atom count** has little impact on developability

\*Values >1 are considered more important

# Importance of Ar-sp3 and fluorine count

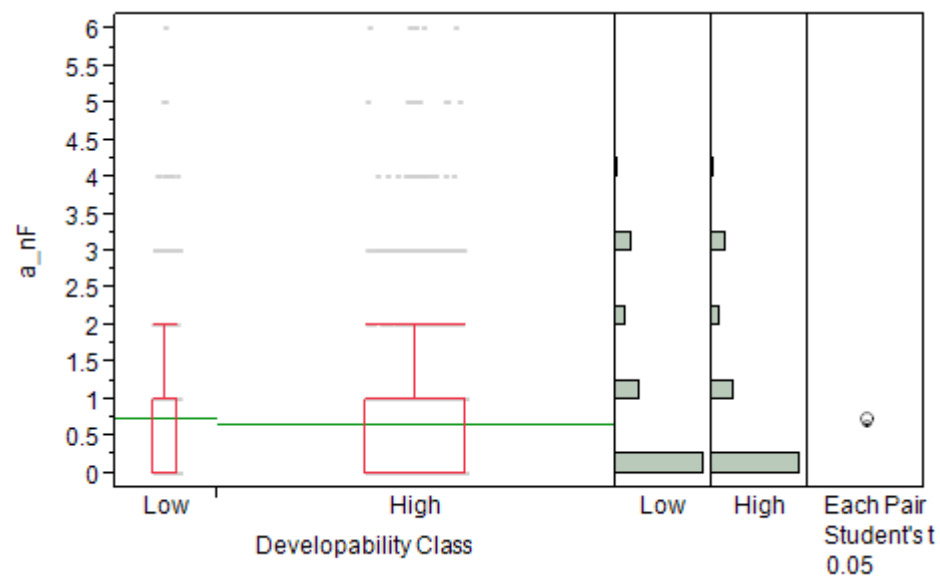
**Ar-sp3: high importance to developability:**

There is a significant difference in Ar-sp3 values between the low and high developability compounds

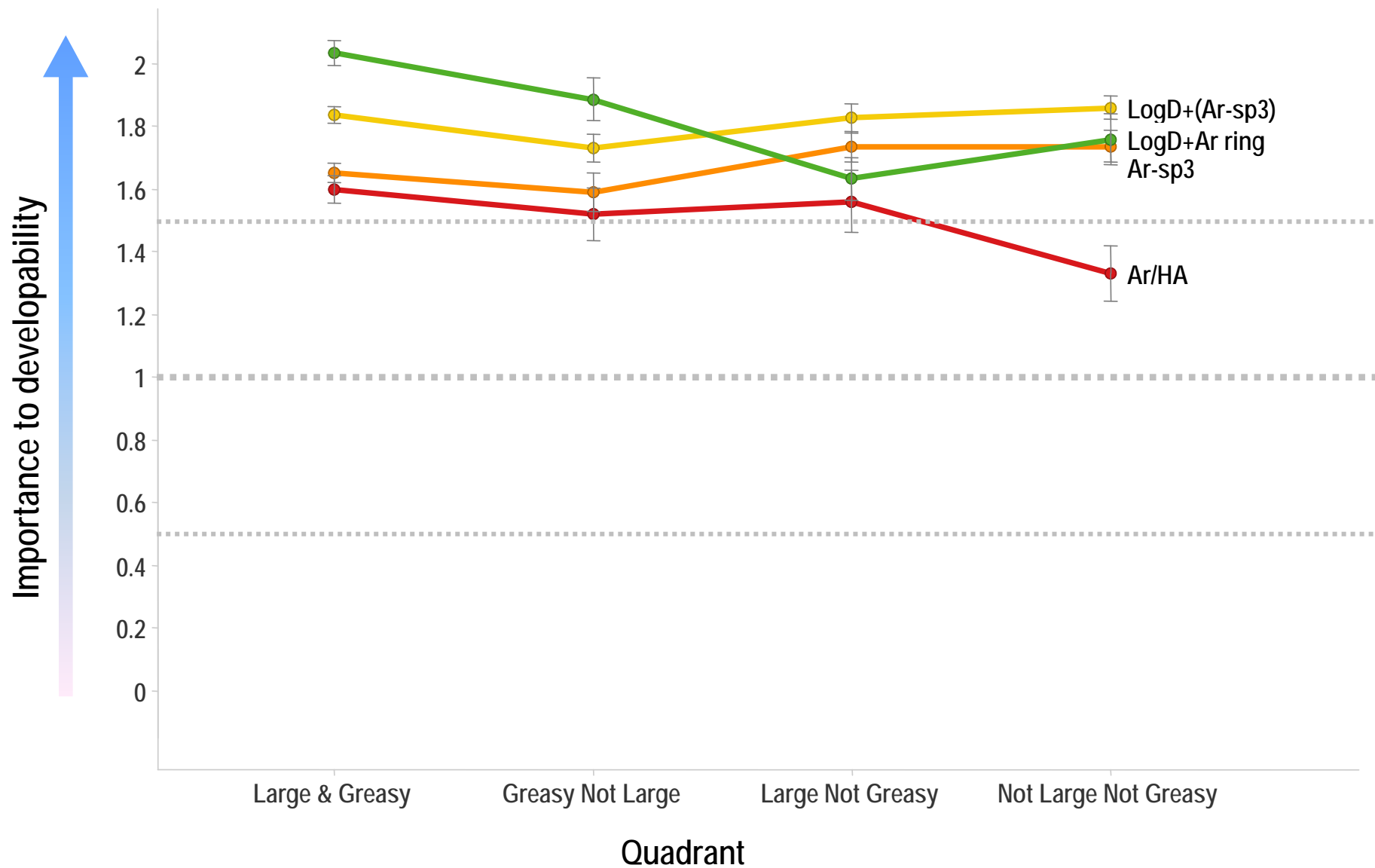


**Fluorine count: low importance to developability:**

No difference in fluorine atom count between high and low developability compounds

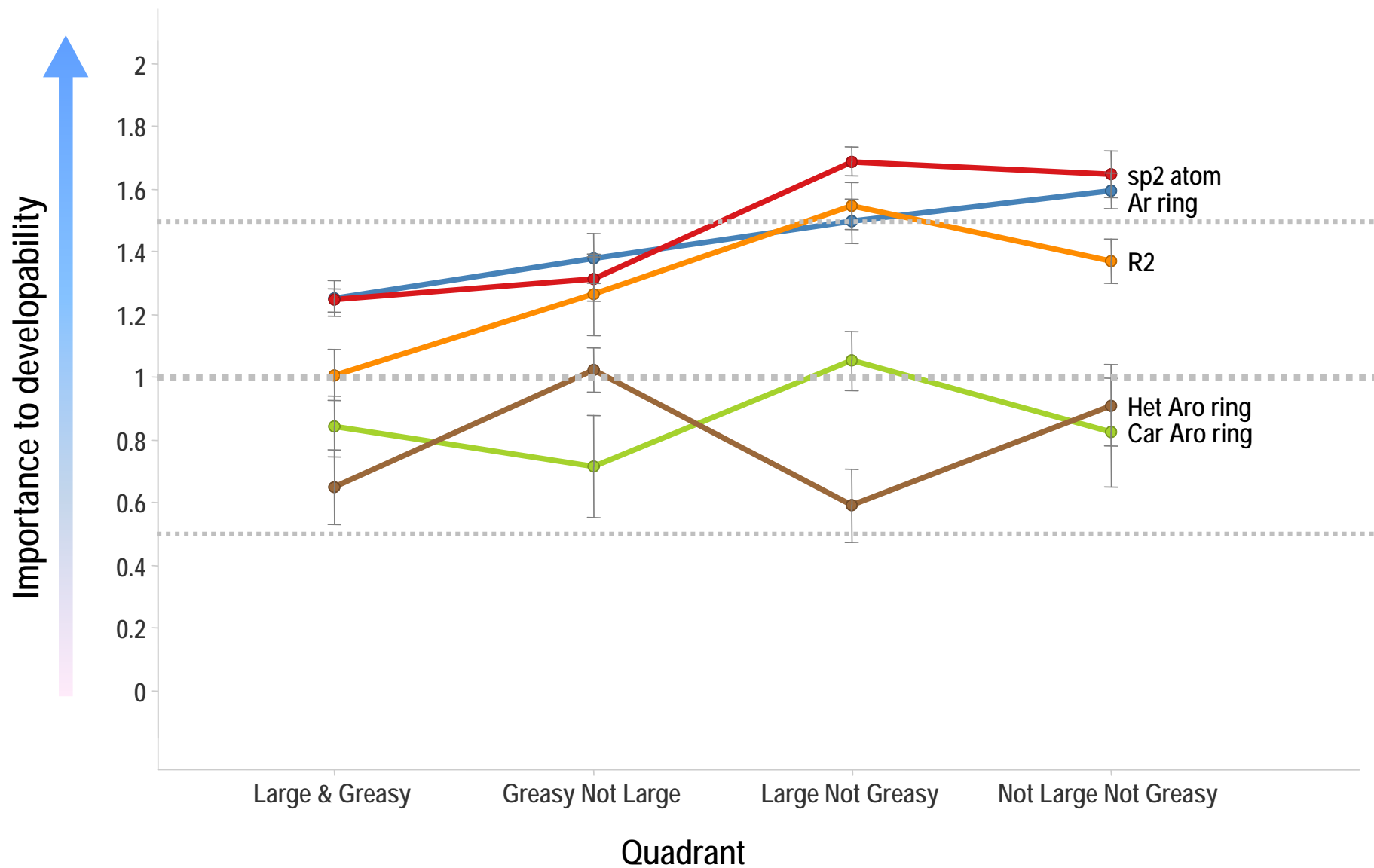


# Importance to developability: **Aromatic-composite**

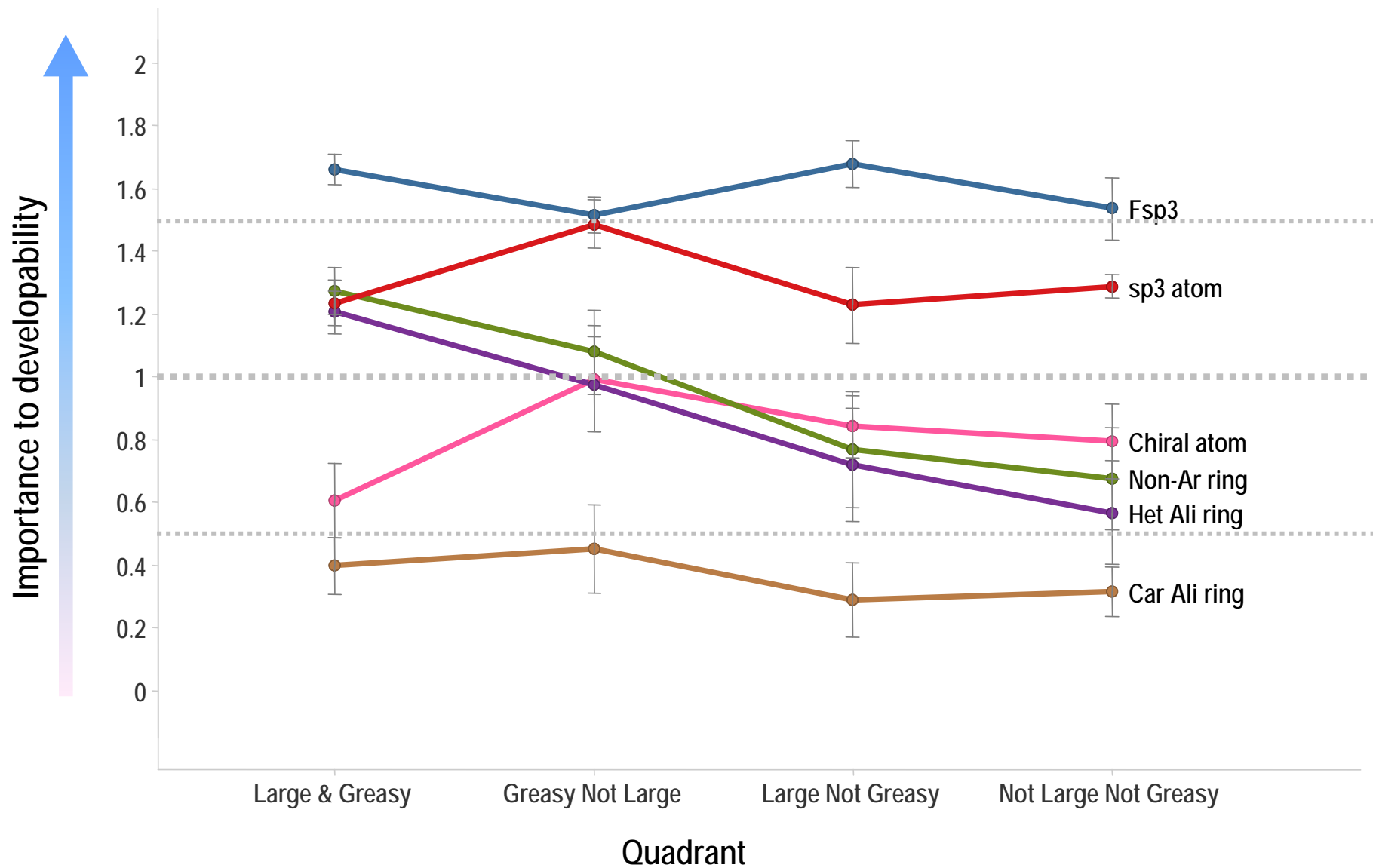




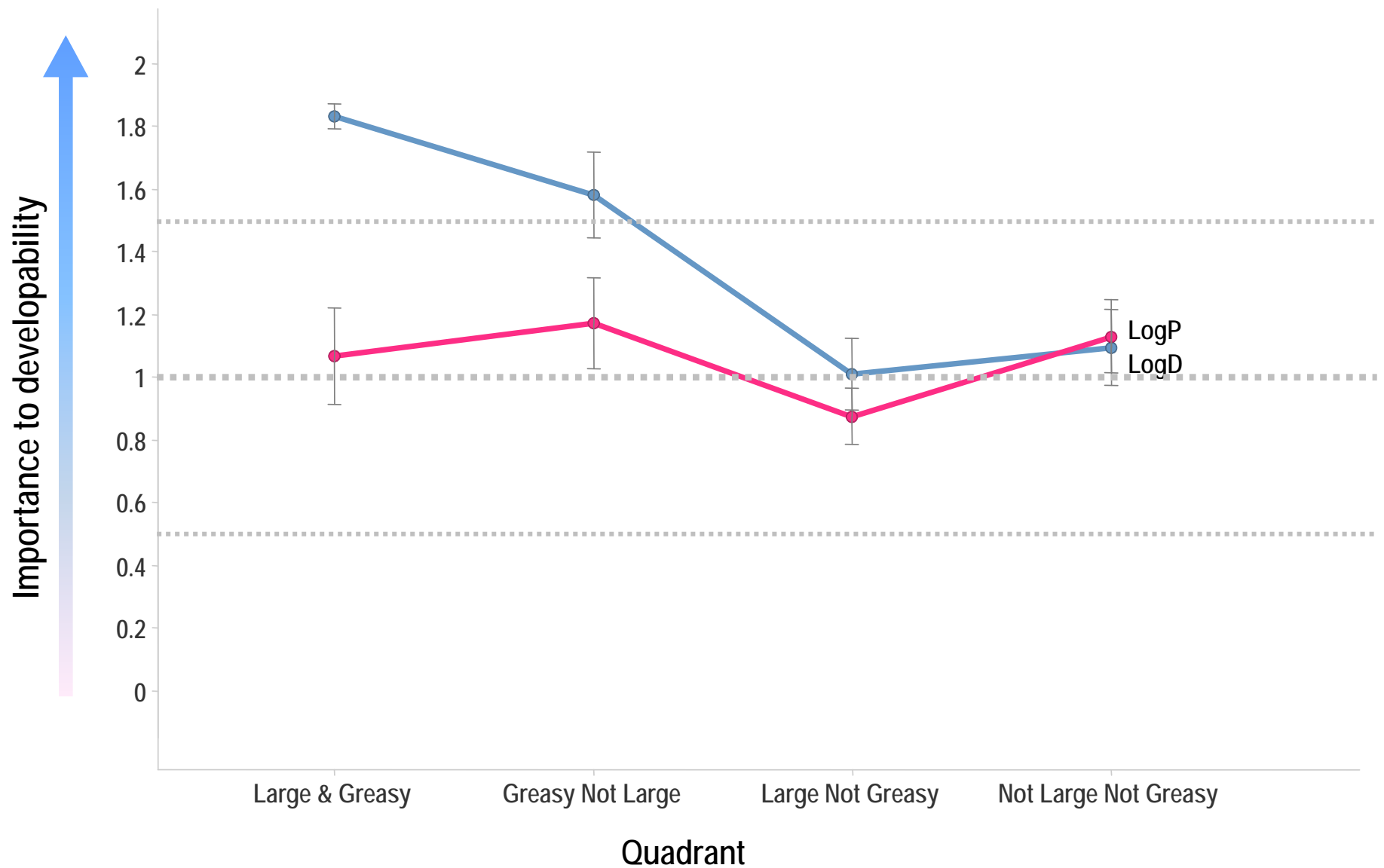
# Importance to developability: **Aromatic**



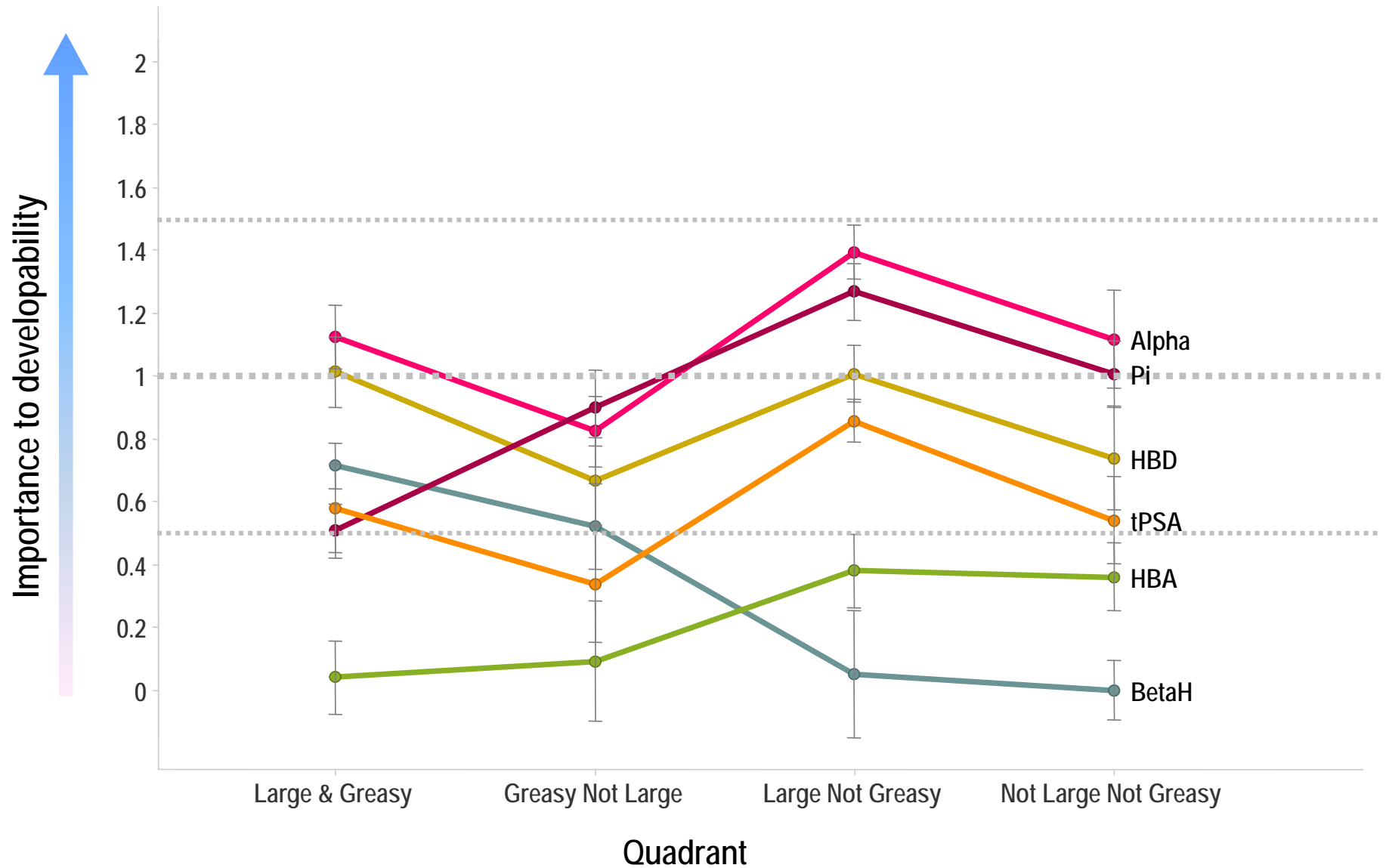
# Importance to developability: **Aliphatic**



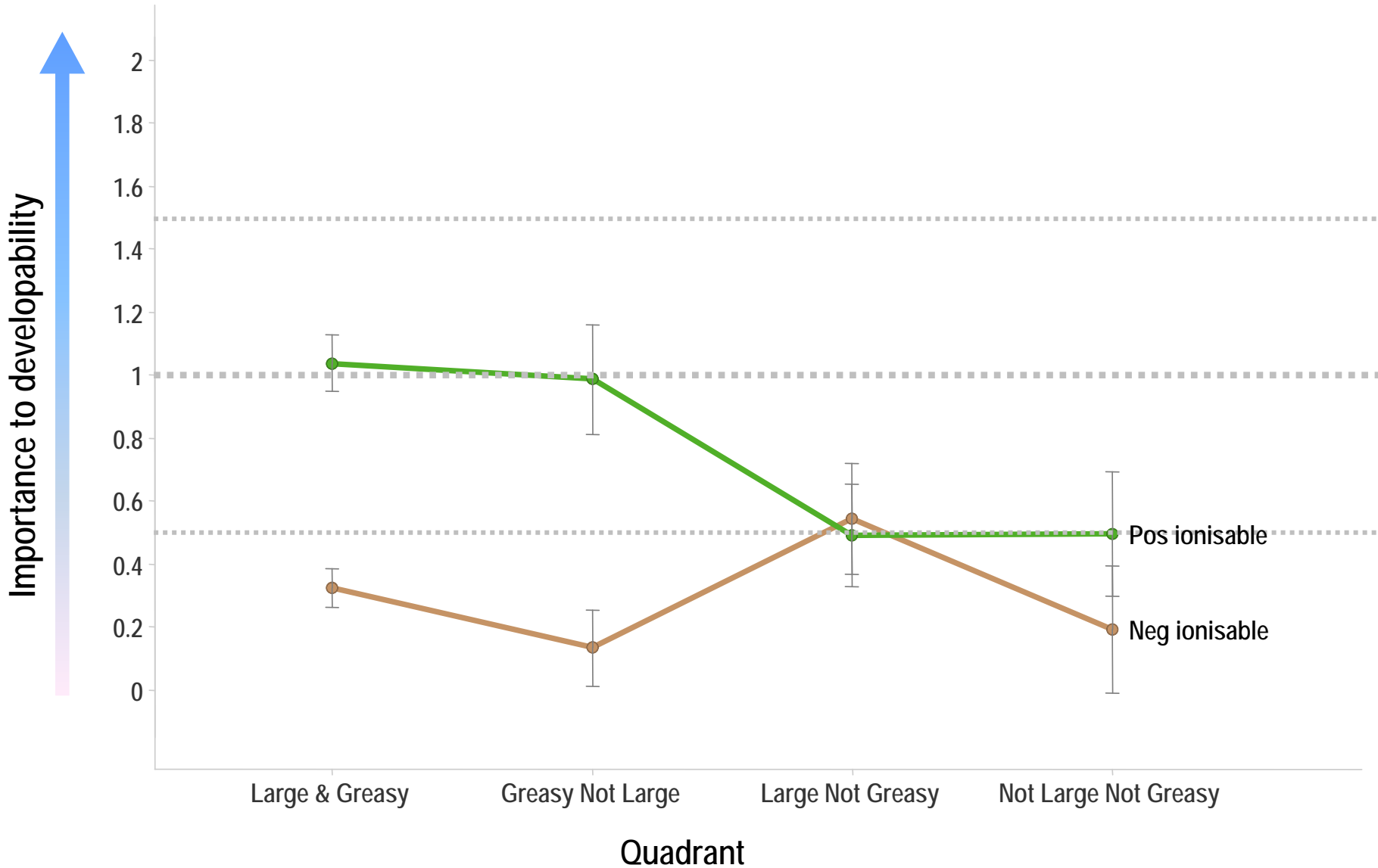
# Importance to developability: **Lipophilicity**



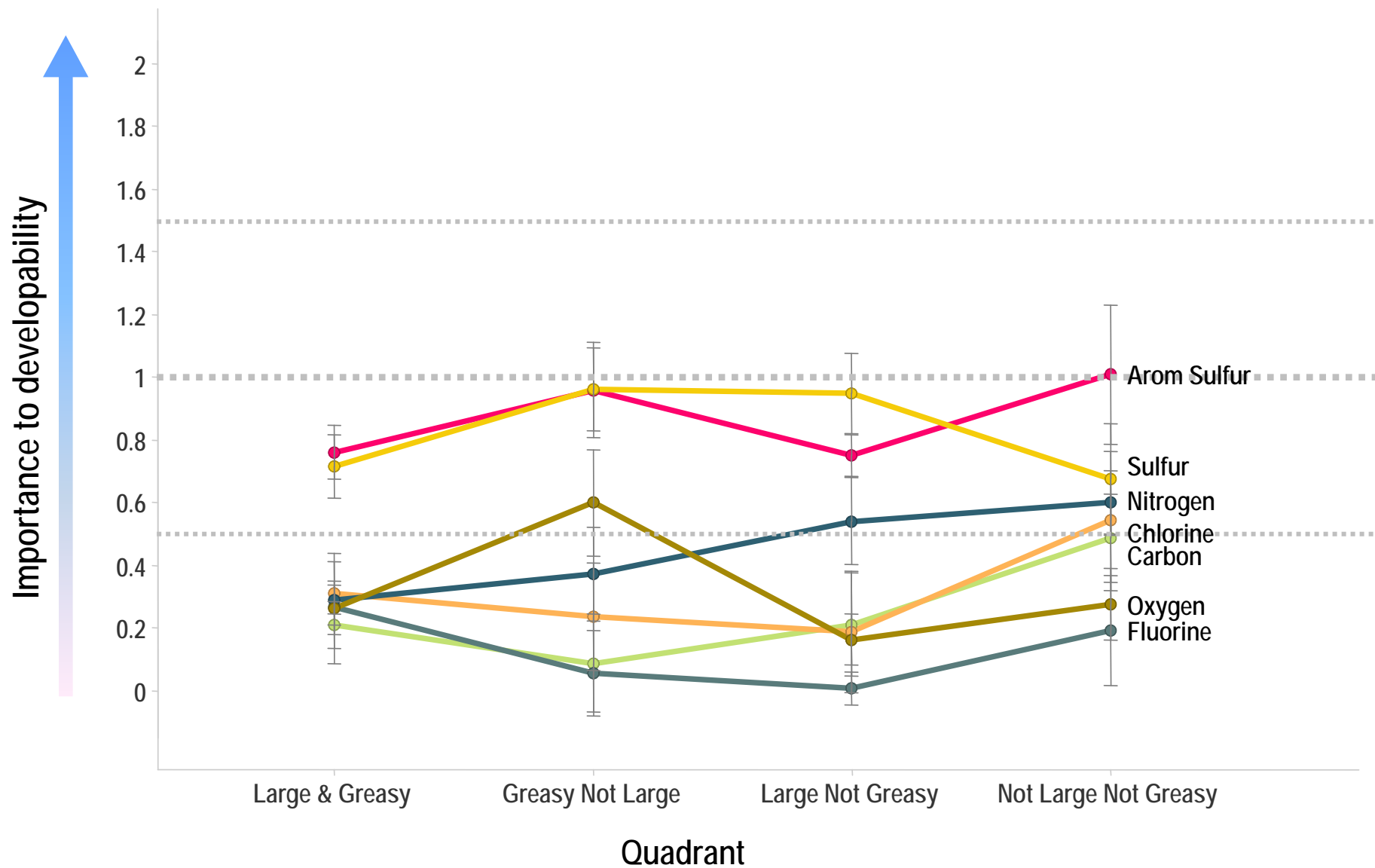
# Importance to developability: **Polarity**



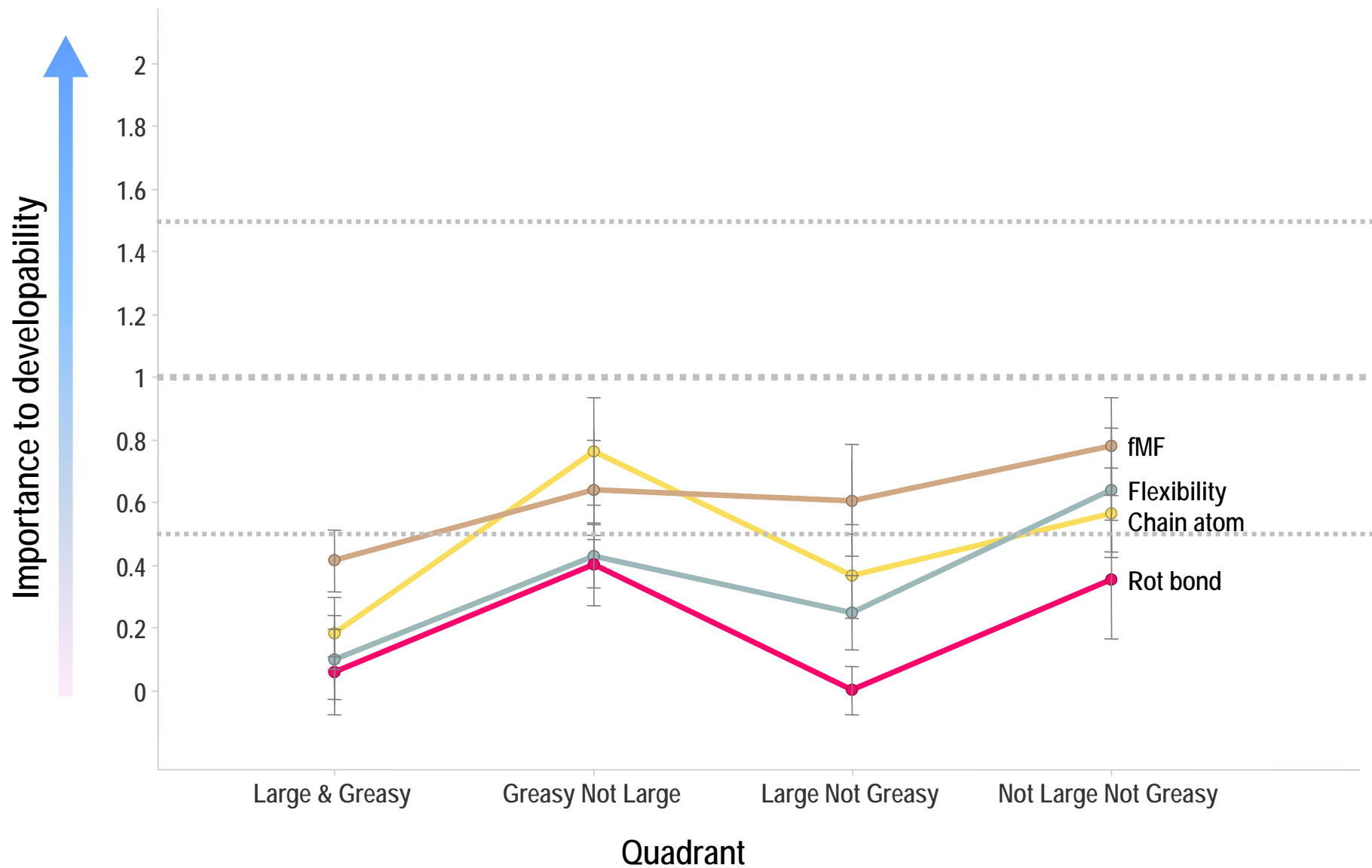
# Importance to developability: Ionisability



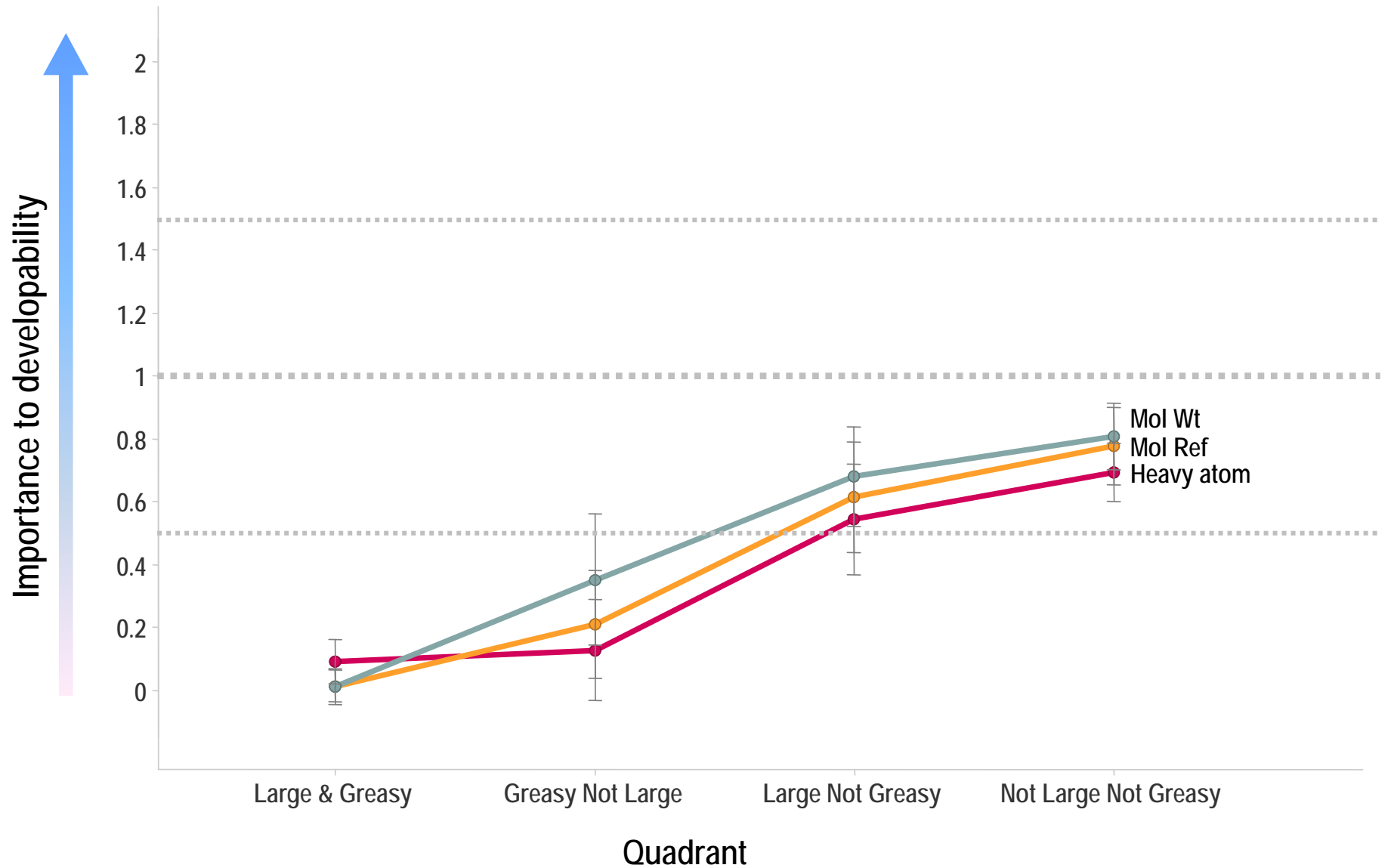
# Importance to developability: **Atom counts**



# Importance to developability: Flexibility



# Importance to developability: **Size**





# Properties & developability: Summary

## 'Always important'

- clogD+Ar ring ↓
- clogD+(Ar-sp3) ↓
- Ar-sp3 ↓
- Ar/HA ↓
- Fsp3 ↑

## 'Important some of the time'

- Ar rings ↓
- clogD ↓
- Positive ionisable groups ↑
- Heteroaliphatic ring count ↑
- HBDs ↓
- Pi ↓

## 'Less important'

- Atom counts ↓
- Flexibility ↑
- Negative ionisable groups ↓
- Size ↓
- HBAs

↓ Lower values increase developability

↑ Higher values increase developability

# Final points

- High clogP has a stronger negative impact on developability than high MW
- Aromatic-composite descriptors appear most important with respect to high/low developability
  - Large & Greasy compounds with less aromatic character / lower clogD = higher developability
  - Not Large & Not Greasy compounds with more aromatic character / higher clogD = lower developability
- N.B. This is a holistic view of GSK compounds
  - Is it representative of 'Pharma' structures?
  - Property importance may vary between different chemical classes / structural frameworks

# Acknowledgements

- GlaxoSmithKline UK
  - Simon Macdonald
  - Simon Peace
  - Stephen Pickett
  - Chris Luscombe

# Molecule properties and compound developability – what's new?

Dr Tim Ritchie

TJR-Chem, Ranco (VA), Italy

<http://www.linkedin.com/in/timritchiechemistry>

# List of descriptors used in analysis

- **log D + (Ar-sp3)** = ChemAxon clog D7.4 + (Ar-sp3)
- **log D + aring** = ChemAxon clog D7.4 + aromatic ring count
- **Ar-sp3** = aromatic atom count minus sp3 carbon atom count
- **Ar/HA** = aromatic atom count/heavy atom count
- **sp2 atom** = sp2-hybridised atom count
- **Ar ring** = aromatic ring count
- **R2** = Abraham's R2 (E) (excess molar refraction) descriptor
- **Het Aro ring** = heteroaromatic ring count
- **Car Aro ring** = carboaromatic ring count
- **Fsp3** = sp3 carbon atom count/total carbon atom count
- **sp3 atom** = sp3-hybridised atom count
- **Chiral atom** = chiral atom count
- **Non-Ar ring** = non-aromatic ring count
- **Het Ali ring** = heteroaliphatic ring count
- **Car Ali ring** = carboaliphatic ring count.
- **log D** = ChemAxon clog D7.4
- **log P** = daylight clog P
- **Alpha** = Abraham's Alpha (A) (H-bond acidity) descriptor
- **Pi** = Abraham's Pi (S) (combined dipolarity/polarisability) descriptor
- **HBD** = H-bond donor count
- **tPSA** = topological polar surface area
- **HBA** = H-bond acceptor count
- **BetaH** = Abraham's BetaH (B) (H-bond basicity) descriptor
- **Pos ionisable** = positive ionisable group count
- **Neg ionisable** = negative ionisable group count
- **Arom sulfur** = aromatic sulfur atom count
- **Ring atom count** = total ring atom count
- **Sulfur** = sulfur atom count
- **Chlorine** = chlorine atom count
- **Nitrogen** = nitrogen atom count
- **Oxygen** = oxygen atom count
- **Carbon** = carbon atom count
- **Fluorine** = fluorine atom count
- **fMF** = fraction of atoms in the molecular framework
- **Flexibility** = (100 rotatable bond count)/total bond count
- **Chain atom** = chain atom count
- **Rot bond** = rotatable bond count
- **Mol Wt** = molecular weight
- **Mol Ref** = molecular refractivity
- **Heavy atom** = heavy atom count

# O-PLS regression models

**Table 1** OPLS models summary

---

Quadrant	n(<0.33) <sup>a</sup>	n(>0.66) <sup>b</sup>	R2X <sup>c</sup>	R2Y <sup>d</sup>	Q2Y <sup>e</sup>
Large & Greasy	3726	1880	0.711	0.440	0.435
Greasy Not Large	873	647	0.564	0.510	0.497
Large Not Greasy	752	2952	0.422	0.329	0.322
Not Large Not Greasy	378	2910	0.471	0.271	0.261

<sup>a</sup> The number of compounds with a developability score of <0.33. <sup>b</sup> The number of compounds with a developability score of >0.66. <sup>c</sup> The cumulative fraction of X variation modelled by the predictive component. <sup>d</sup> The cumulative fraction of Y variation modelled by the predictive component. <sup>e</sup> The cumulative fraction of Y variation predicted by the X model, by cross-validation ('leave-many-out', sevenfold).

---