

# A Broad View of Structure-Activity Relationship Performance for Alkanes and Haloalkanes

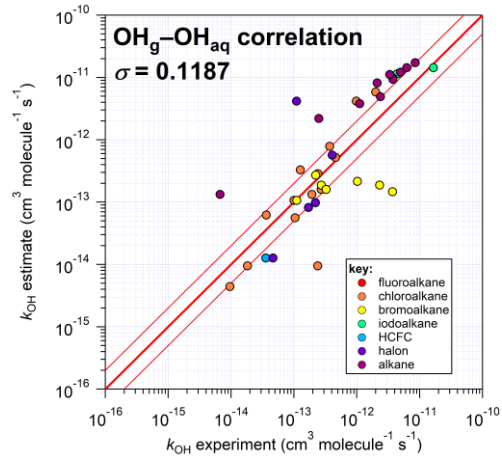
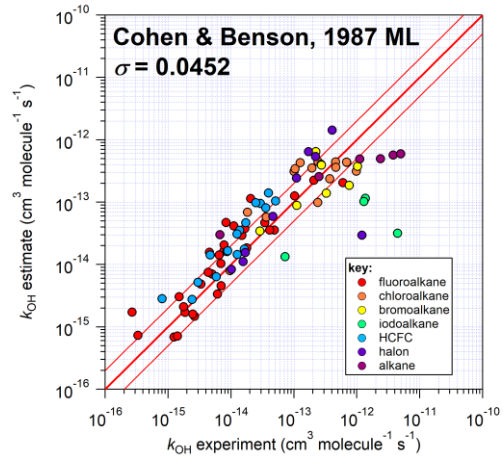
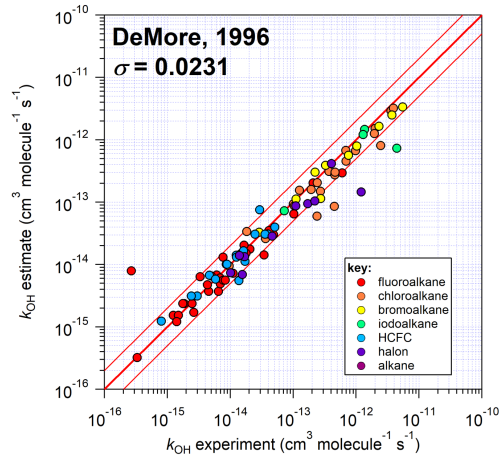
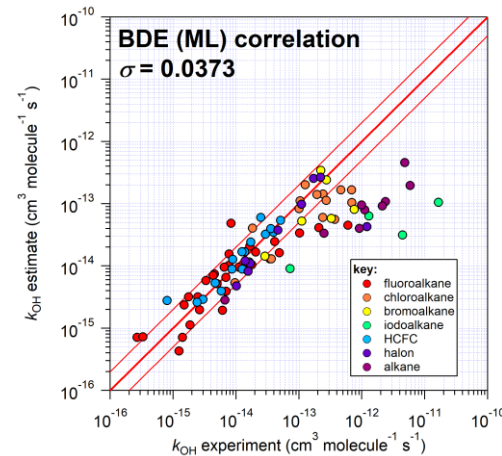
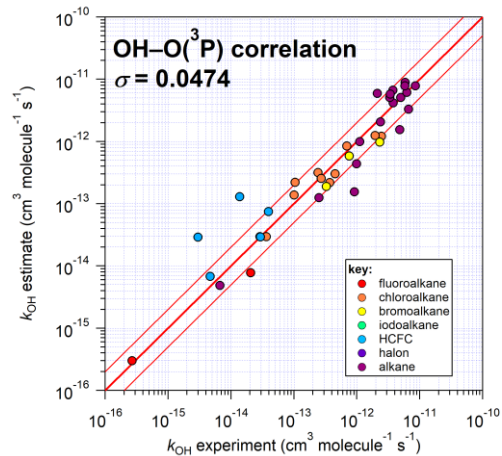
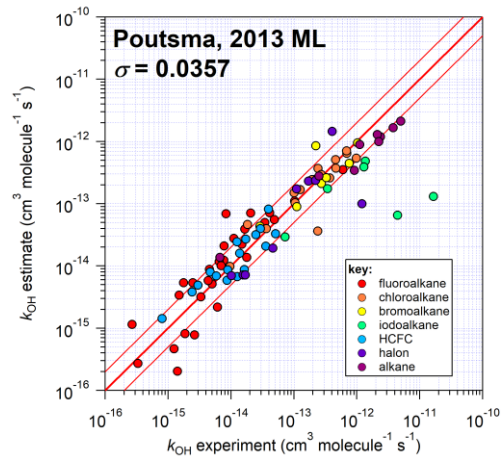
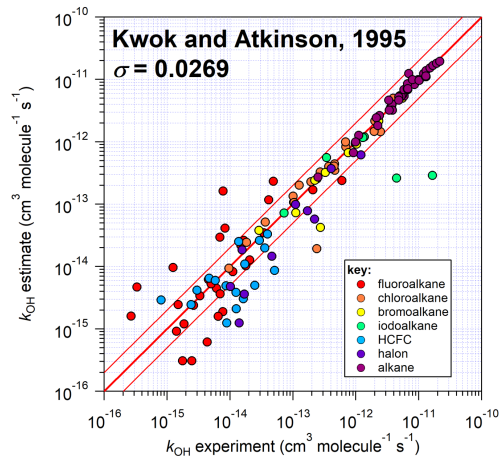
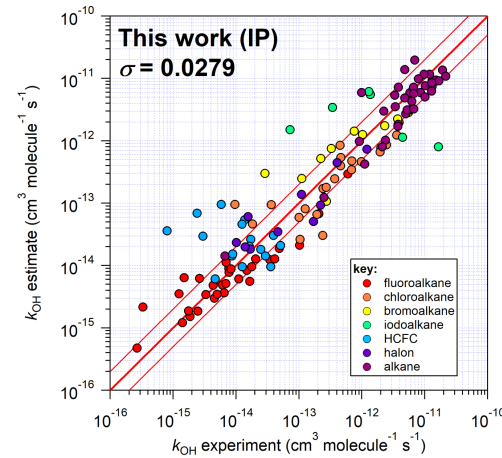
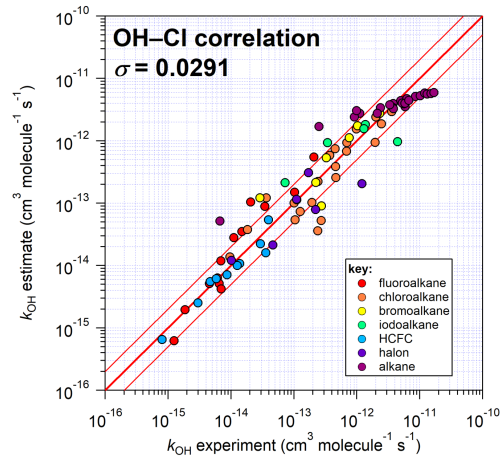
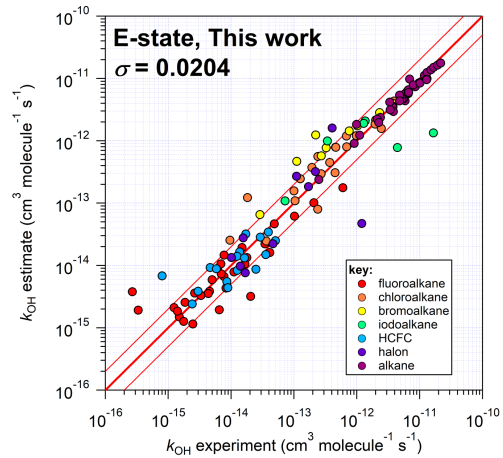
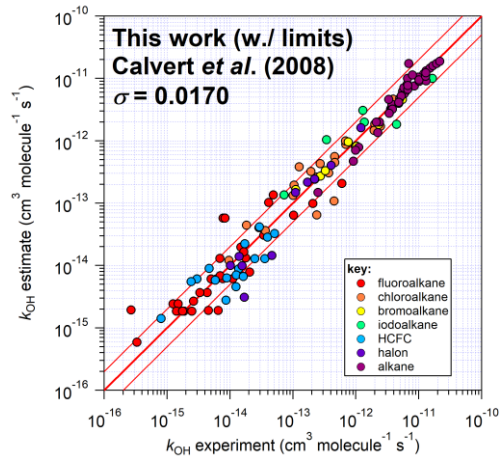


Max McGillen<sup>1</sup>, Lisa Michelat<sup>1</sup>, John Orlando<sup>2</sup>

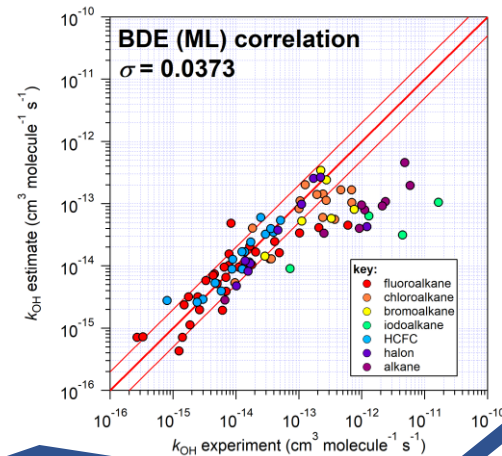
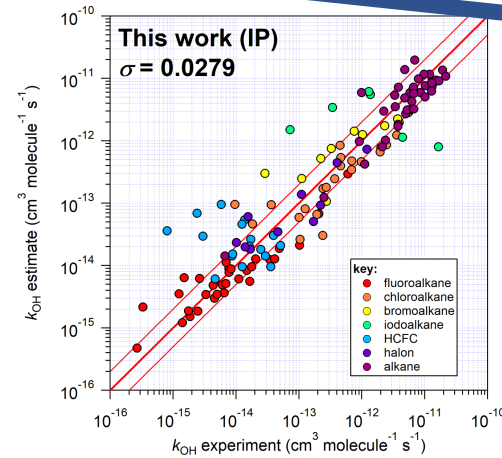
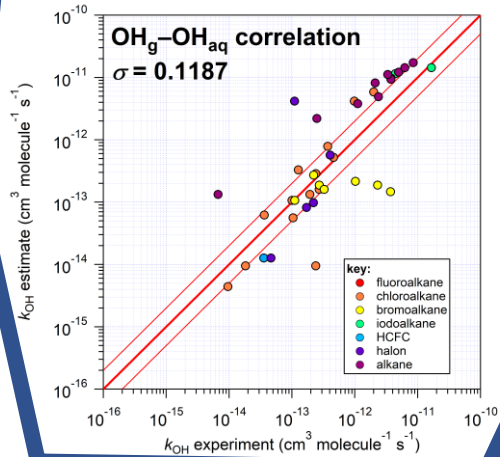
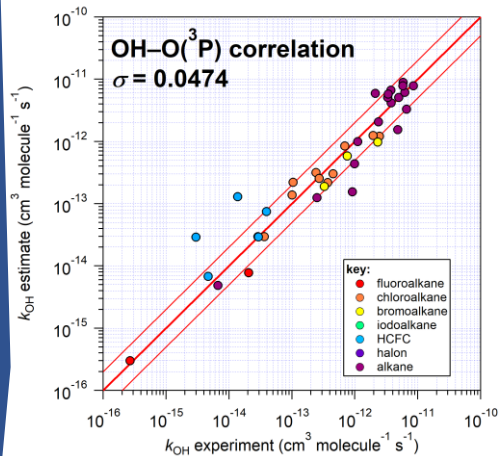
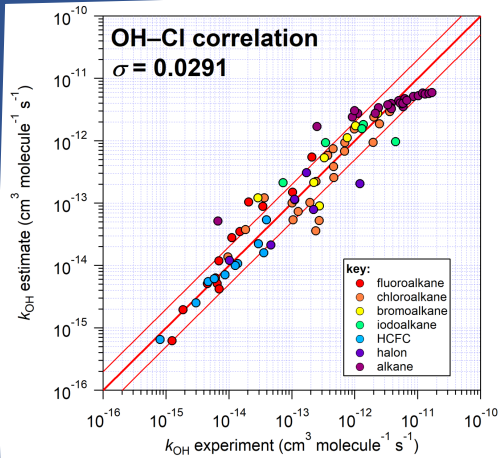
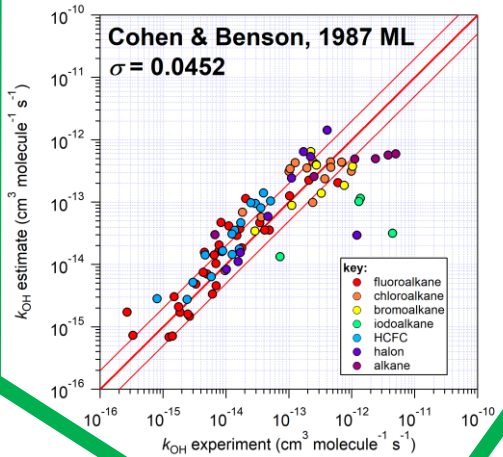
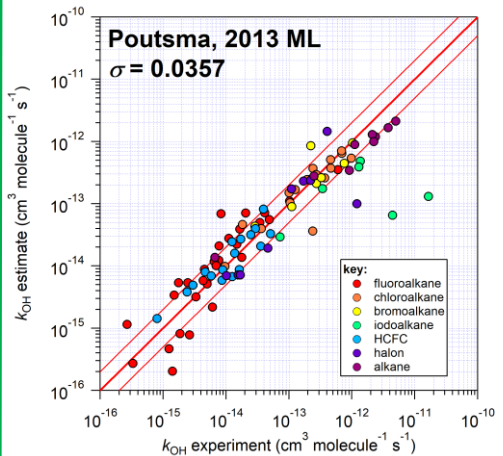
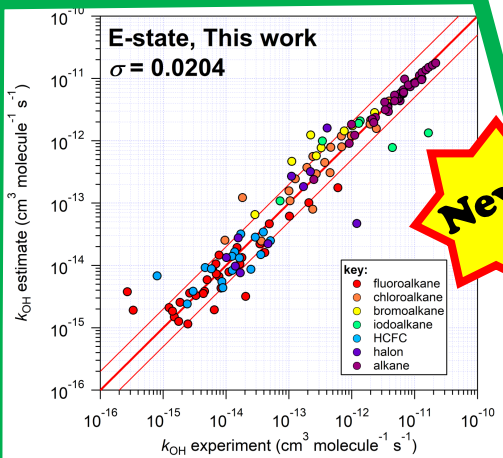
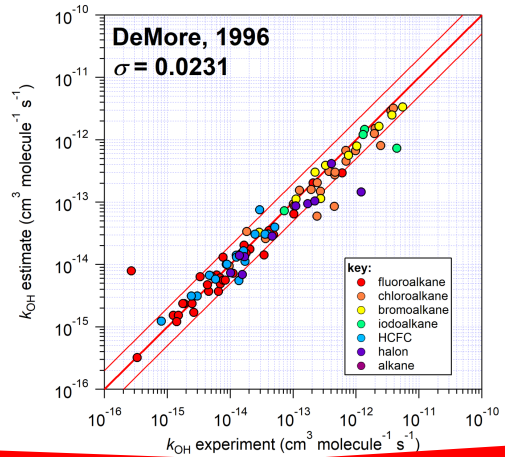
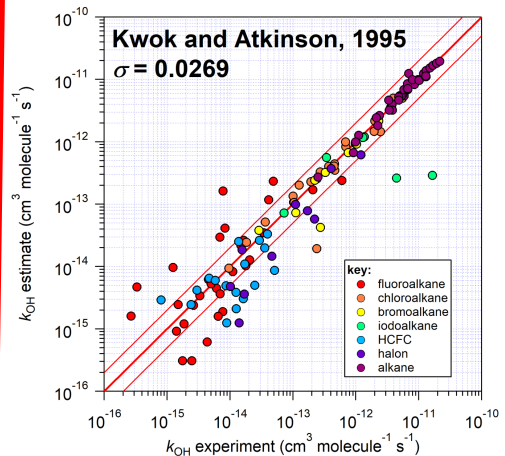
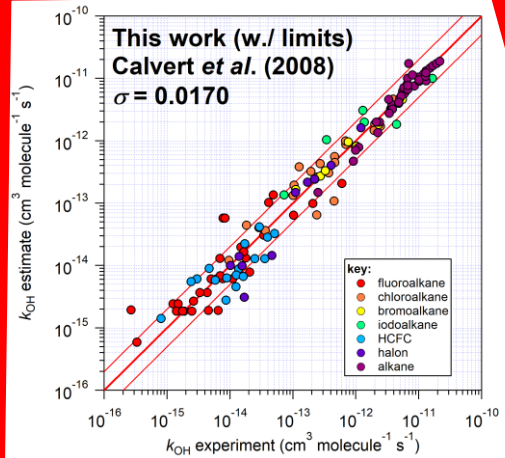


# Why study SARs of alkanes and haloalkanes?

- They represent a family of chemicals that we understand (relatively) well
  - **Mechanistically simple:** straightforward H-abstraction
  - **Range of reactivity:** from  $\text{CF}_3\text{H}$  ( $2.68 \times 10^{-16}$ ) to  $n\text{-C}_{16}\text{H}_{34}$  ( $2.16 \times 10^{-11}$ )
  - **Many experimental determinations:** 105 species, about 10% of all  $k_{\text{OH}}$  measurements; different oxidants, T-dep, repeat measurements, etc.
  - **Supporting data:** thermochemistry, bond-dissociation energies (measured and ML), Swain-Lupton coefficients, quantum chemical data
  - **Different techniques available for estimating  $k$ :** see later
- Therefore, they should be an ideal testing ground to assess state of the knowledge regarding estimation techniques for oxidation reactions
- What kind of estimation techniques do we have available for this class of compounds?



**A lot!**  
 (and this isn't a complete list)

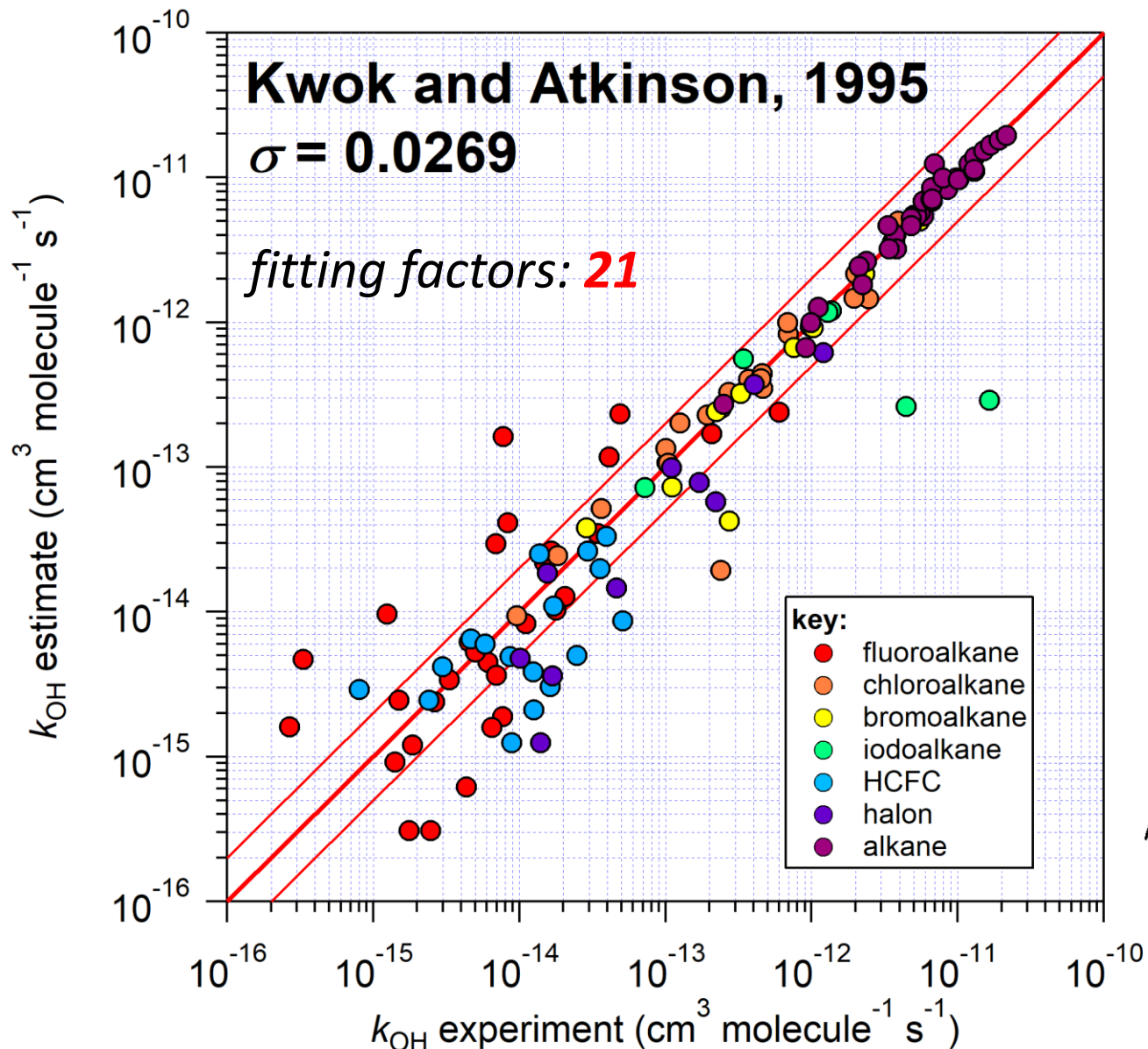


- pure group-additivity
- Group-additivity/linear free-energy relationship
- Pure linear free-energy relationship

# Some examples of estimation methods

- In the interests of time (and to make this less repetitive) I won't go through all the examples here, instead I will focus on a few of them:
- **Kwok and Atkinson** (because of its popularity)
- **DeMore** (because it's "better" than Kwok and Atkinson)
- **IP** (uses measurements/ calculations instead of pure fitting)
- **OH-Cl correlation** (uses measurements instead of pure fitting)
- **Electrotopological state** (because it's my new method)

# Group-additivity method: Kwok and Atkinson



- This method (and its derivatives) is the go-to group-additivity method for atmospheric chemistry
- **Despite this, it's not very effective for haloalkanes**

$$k(\text{CH}_3\text{-X}) = k_{\text{prim}} F(\text{X})$$

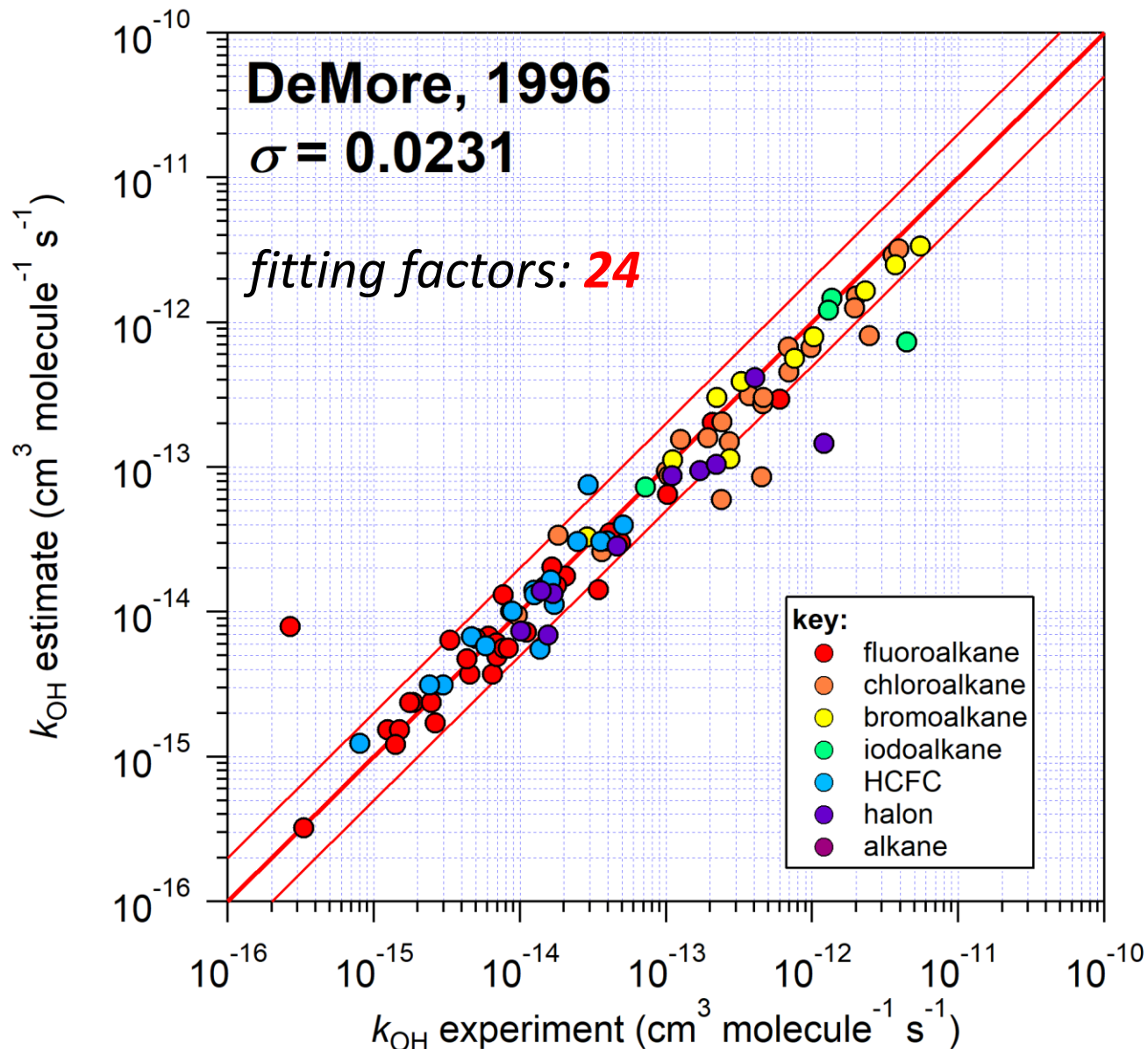
$$k(\text{X-CH}_2\text{-Y}) = k_{\text{sec}} F(\text{X}) F(\text{Y})$$

$$k\left(\text{X-CH} \begin{array}{l} \swarrow \text{Y} \\ \searrow \text{Z} \\ \backslash \text{X} \end{array}\right) = k_{\text{tert}} F(\text{X}) F(\text{Y}) F(\text{Z})$$

$$k_{\text{total}} = \sum [k_{\text{prim}}^{\circ} F(\text{X})] + \sum [k_{\text{sec}}^{\circ} F(\text{X}) F(\text{Y})] + \sum [k_{\text{tert}}^{\circ} F(\text{X}) F(\text{Y}) F(\text{Z})]$$

- Kwok, E. & Atkinson, R. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship: An update. *Atmos. Environ.* **29**, 1685–1695 (1995).

# Group-additivity method: DeMore



- This approach is less commonly used than the Atkinson approach
- **However, it is apparently better for haloalkanes, which may relate to its “third-group multiplier” factor**

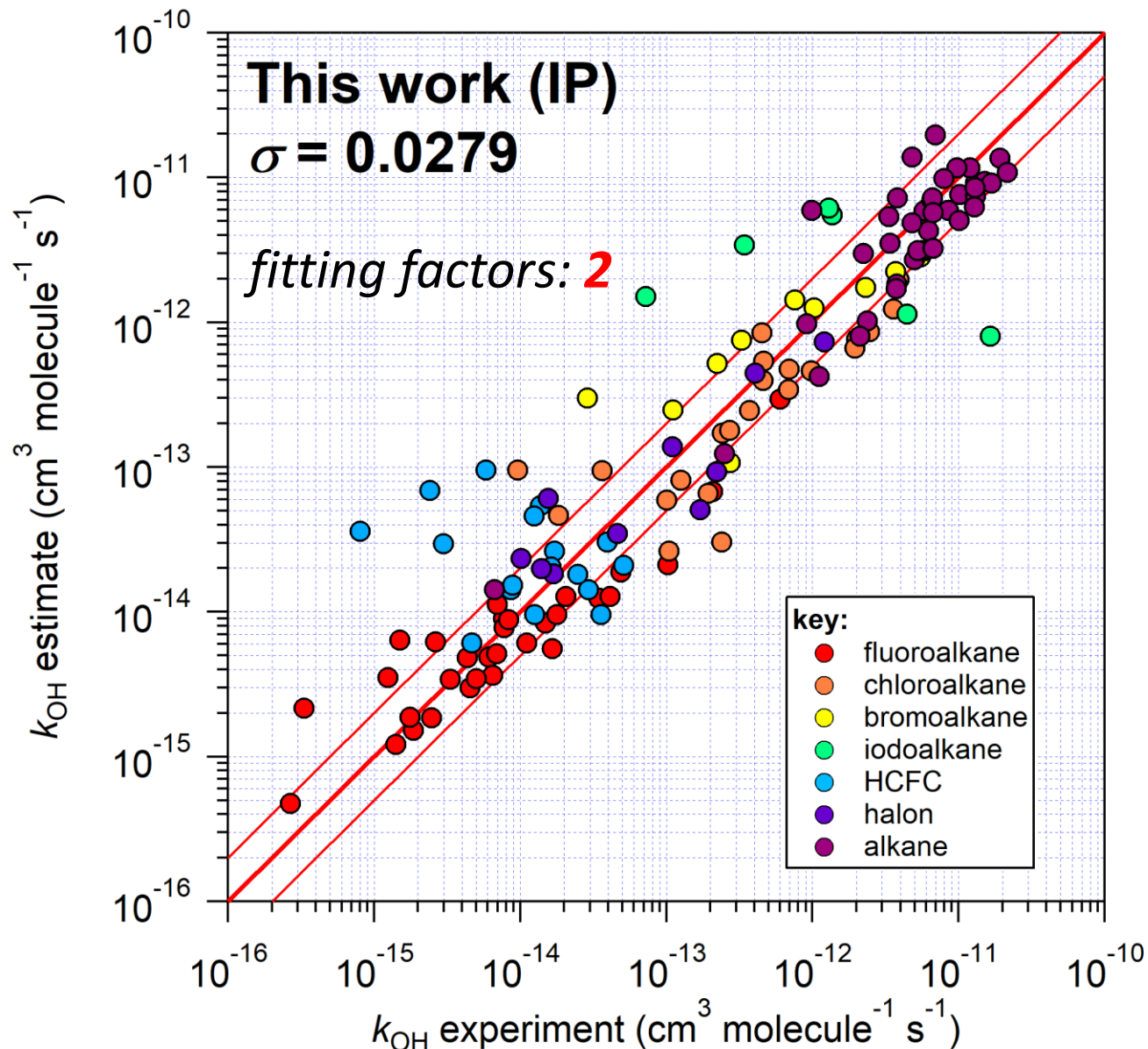
for the  $i^{\text{th}}$  reaction site:

$$\log k = \log k(\text{CH}_4) + G_1 \dots G_3$$

$$k_{\text{total}} = \sum_i k_i$$

- DeMore, W. B. Experimental and Estimated Rate Constants for the Reactions of Hydroxyl Radicals with Several Halocarbons. *J. Phys. Chem.* **100**, 5813–5820 (1996).
- **Update of fitting parameters: this work**

# Linear free-energy relationship method: IP



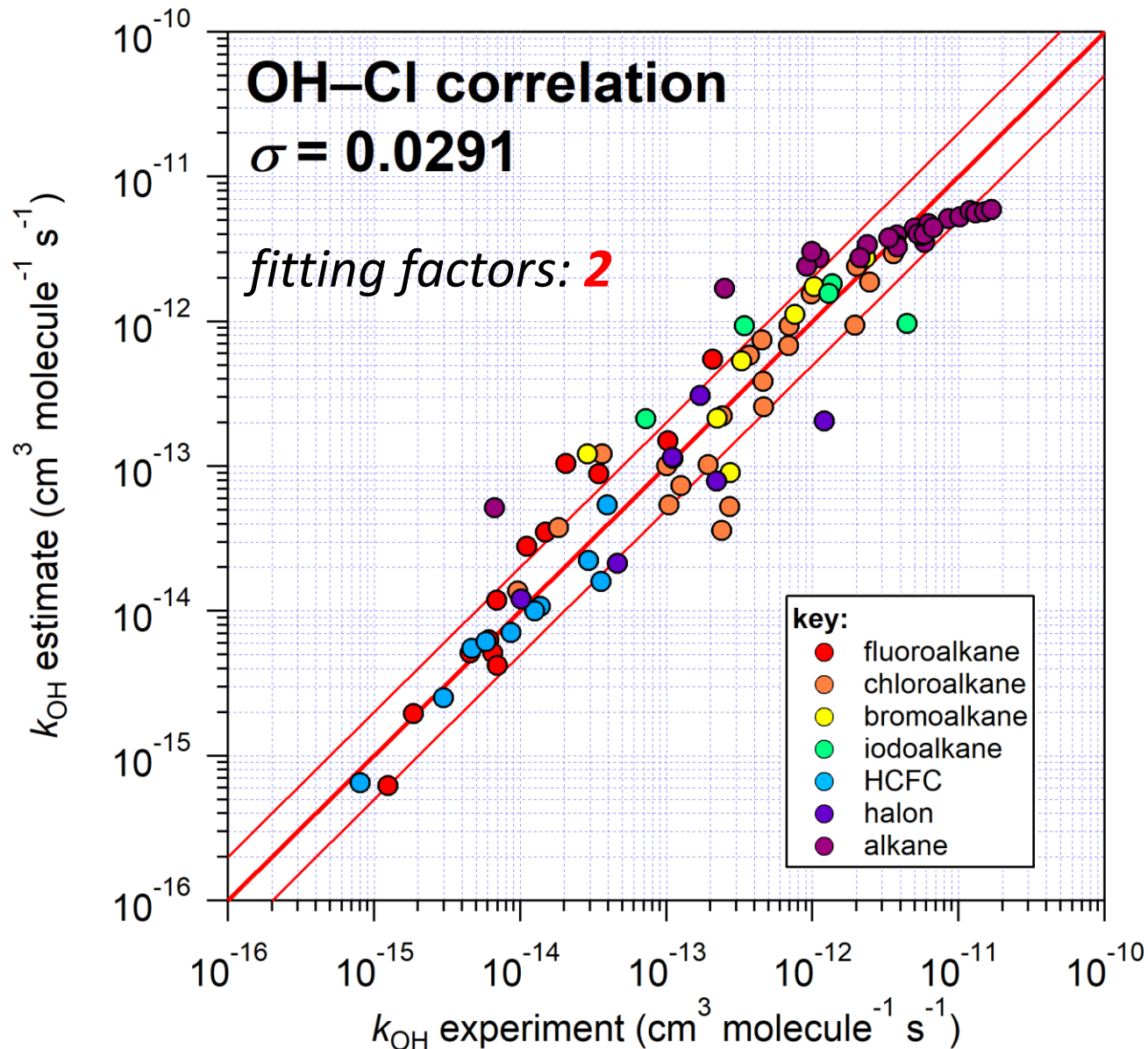
- Ionization potential (IP) has been used previously to correlate with  $k_{\text{OH}}$
- **IP calculated using PM6, and corrected against experimental values (where available)**

$$\frac{\log k}{n_{\text{H}}} = m \cdot \text{IP} + c$$

- Percival, C. J., Marston, G. & Wayne, R. P. Correlations between rate parameters and calculated molecular properties in the reactions of the hydroxyl radical with hydrofluorocarbons. *Atmospheric Environment* **29**, 305–311 (1995).
- Stewart, J. J. P. MOPAC2016.
- **Algorithm update, IP calculations: this work**



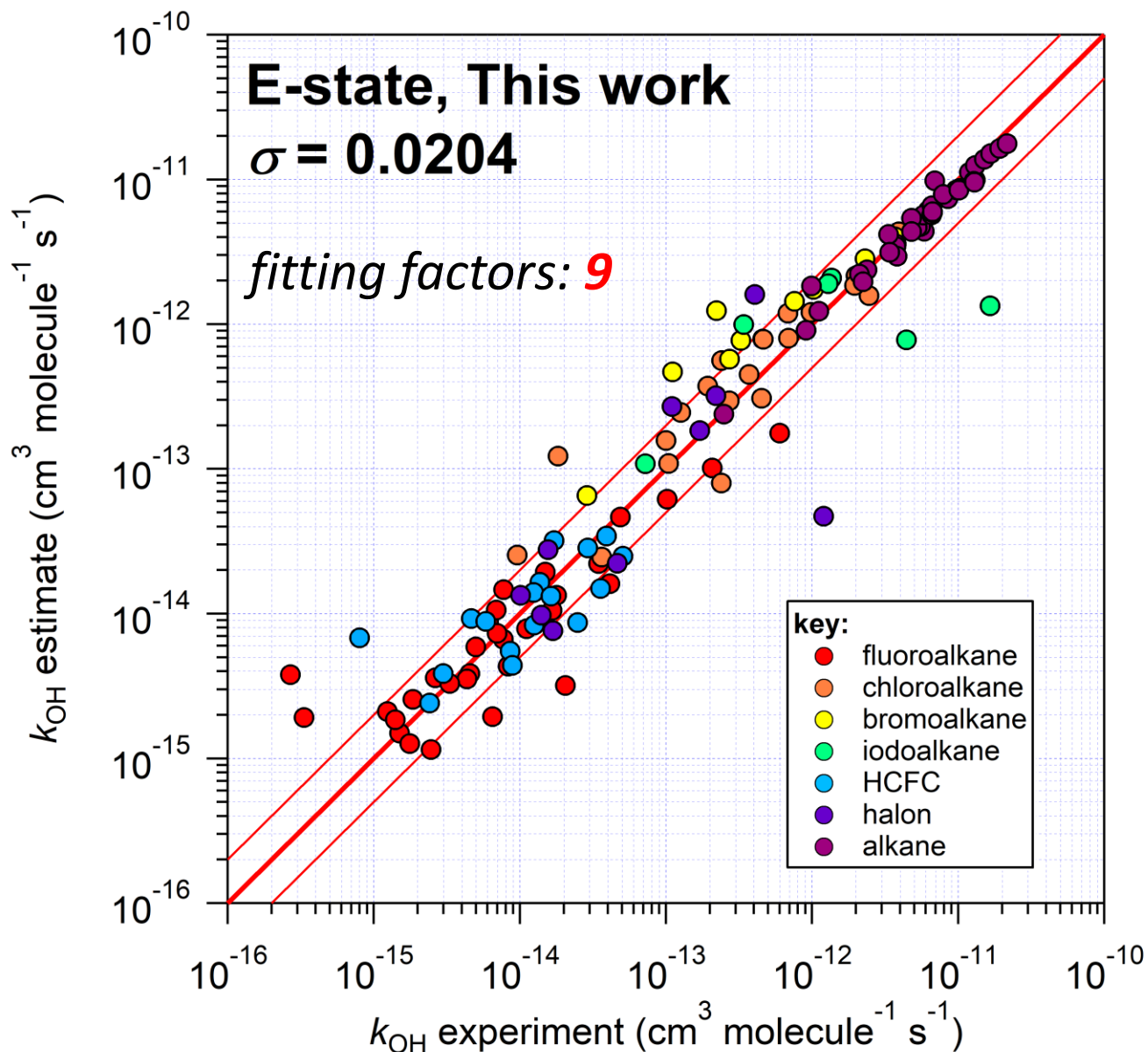
# Linear free-energy relationship method: $k_{\text{OH}}-k_{\text{Cl}}$



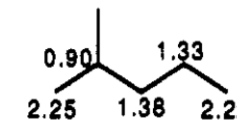
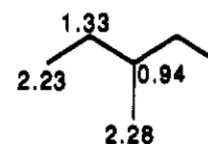
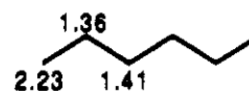
- The OH-Cl relationship has been observed to work well before
- It implies a similar mechanism, probably C-H bond strength
- **Although it is a reasonably good predictor, there are some problems with higher  $k$  values (which can be corrected)**

• T.J. Wallington and co-workers, various works, applied mainly to HFEs  
• **This work, focussing on a more complete alkanes and haloalkane dataset**

# Hybrid approach: the electrotopological state



- A new method has been developed, making use of the Kier-Hall index
- Describes the electronic character and the connectivity of each hydrogen environment
- **Approach yields site-specific rate coefficients at relatively high accuracy and a low number of fitting parameters**



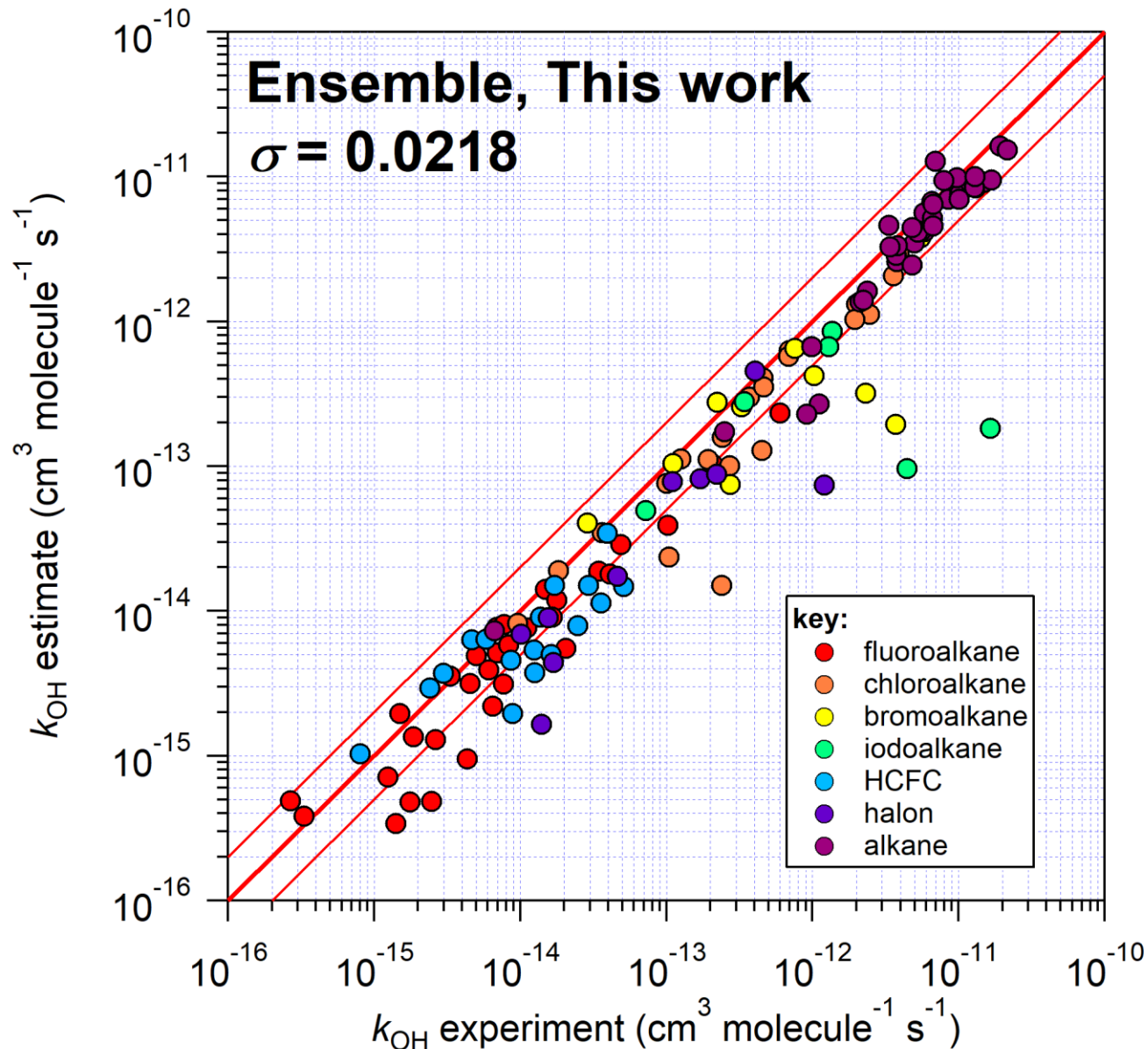
Hall, L. H., Mohney, B. & Kier, L. B. The electrotopological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Model.* **31**, 76–82 (1991).

# How can we combine all this information?

- You have seen already that some of these approaches are quite good at describing  $k_{\text{OH}}$
- But is there a way to connect these predictions together to provide a more informed estimate (to cover the estimation space more fully)?
- One simple idea is to treat each SAR estimate as though it were a separate measurement, with its own uncertainty
- Taking the methodology that we might use in the laboratory:

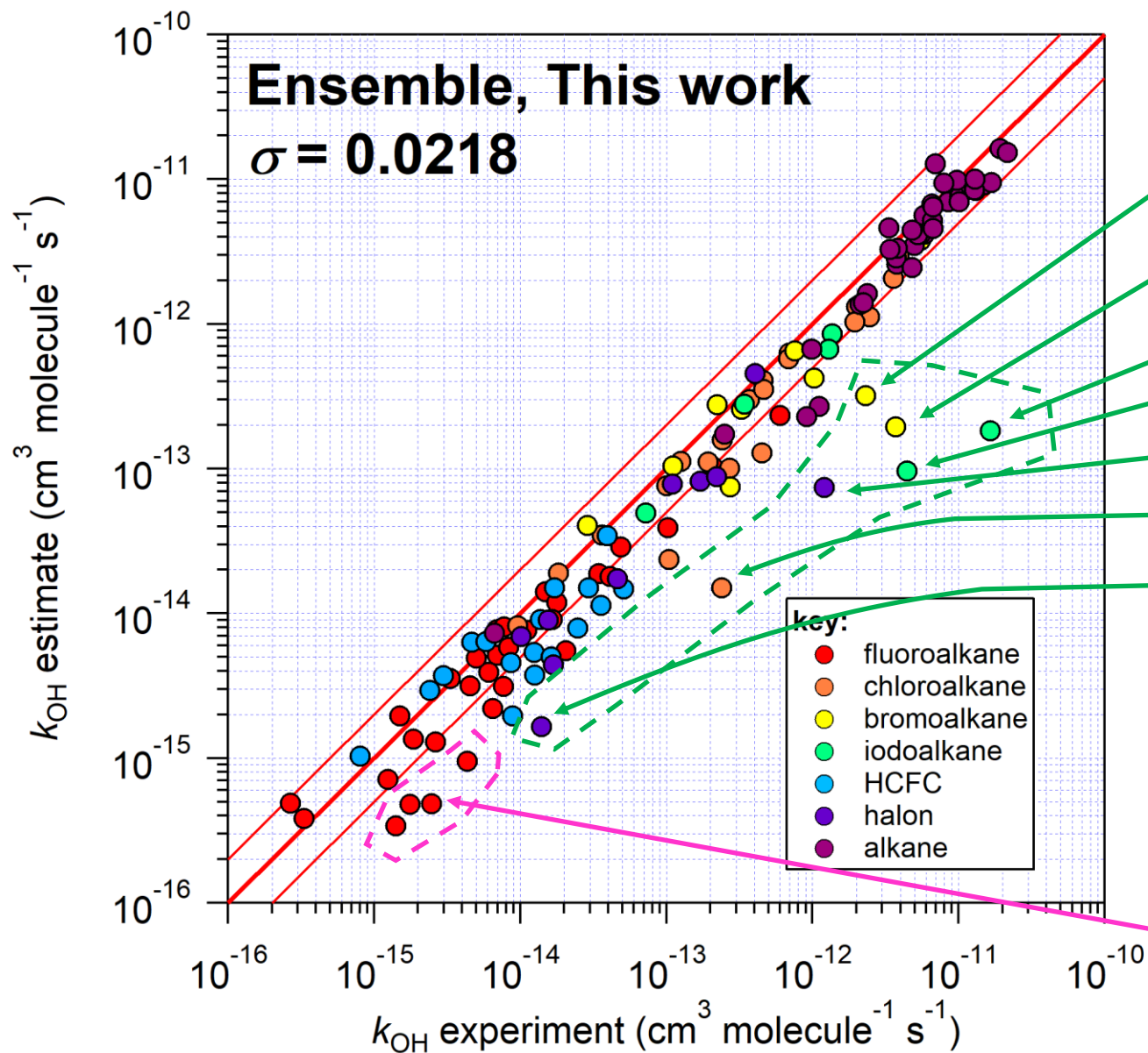
$$k_{\text{ensemble}} = \frac{\sum_{i=1}^N w_i k_i}{\sum_{i=1}^N w_i} \quad \text{where, } w_i = 1/\sigma_i^2$$

# Ensemble performance

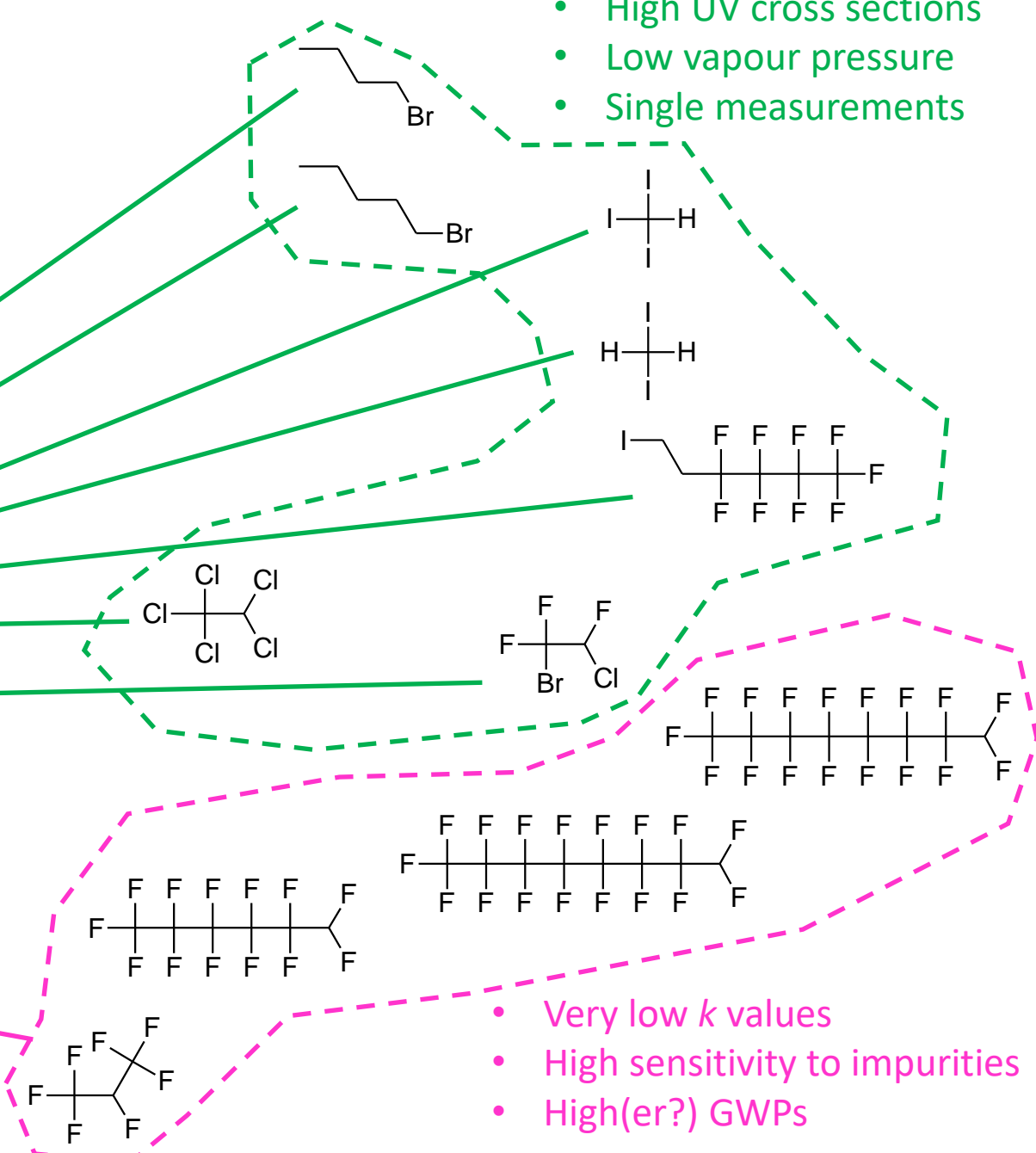


- The ensemble is less accurate at reproducing the experimental data than some individual approaches
- This was (initially) disappointing
- But, we can't assume that our experimental database is perfect
- **There is an interesting asymmetry to the errors in estimation space, with a tendency towards underprediction**

# Ensemble performance

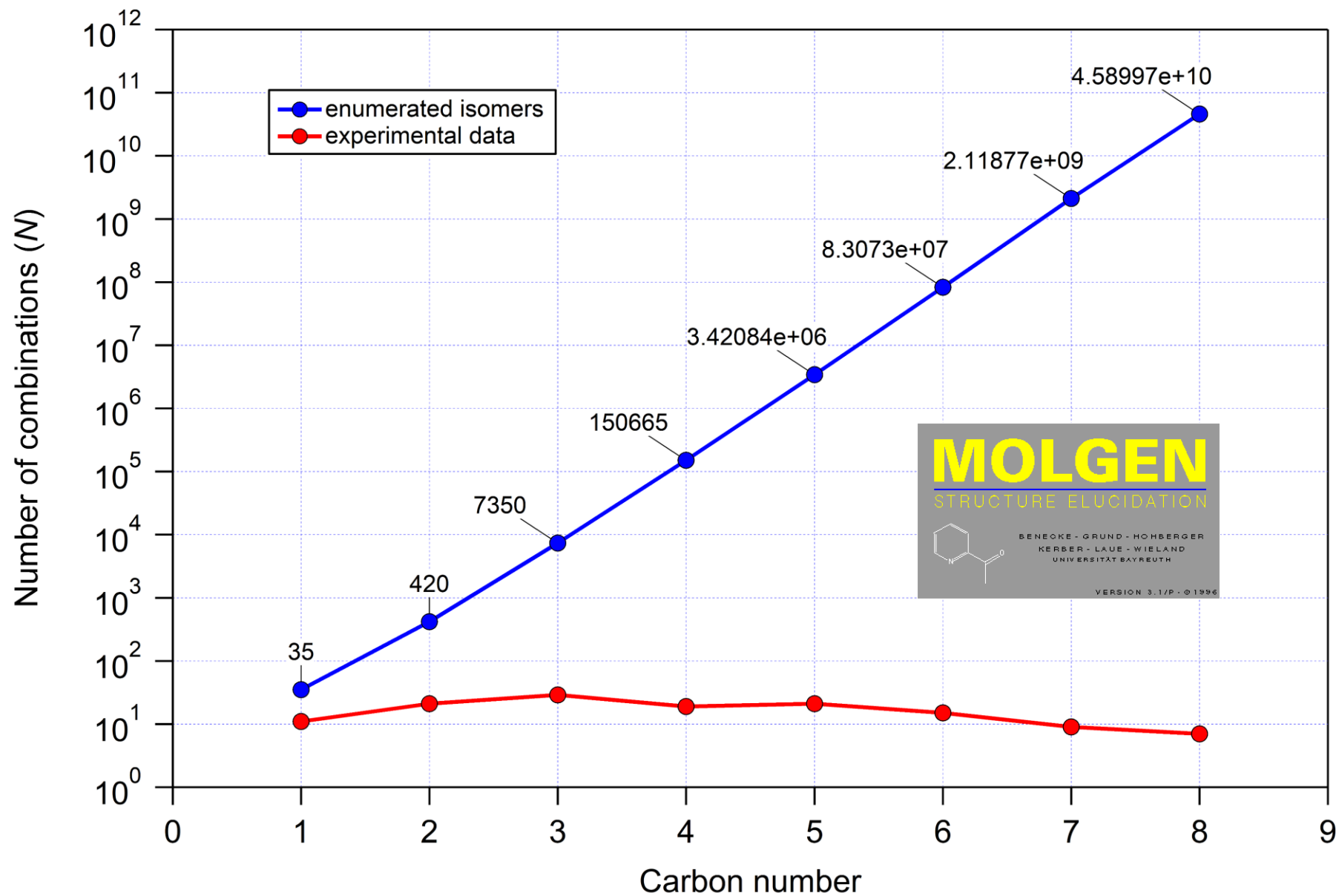


- High UV cross sections
- Low vapour pressure
- Single measurements

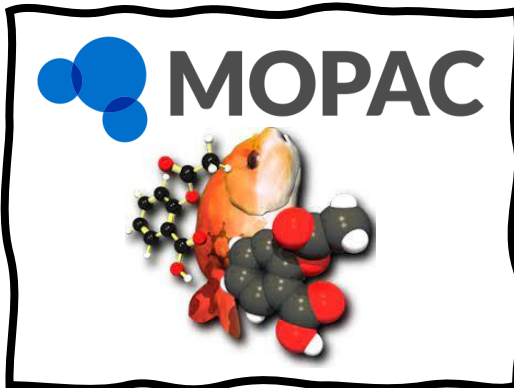
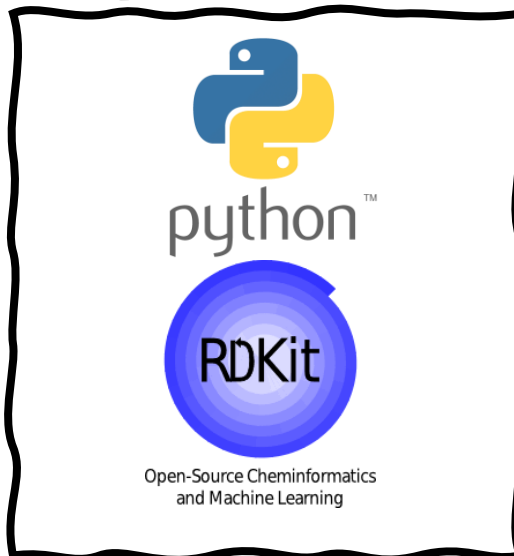
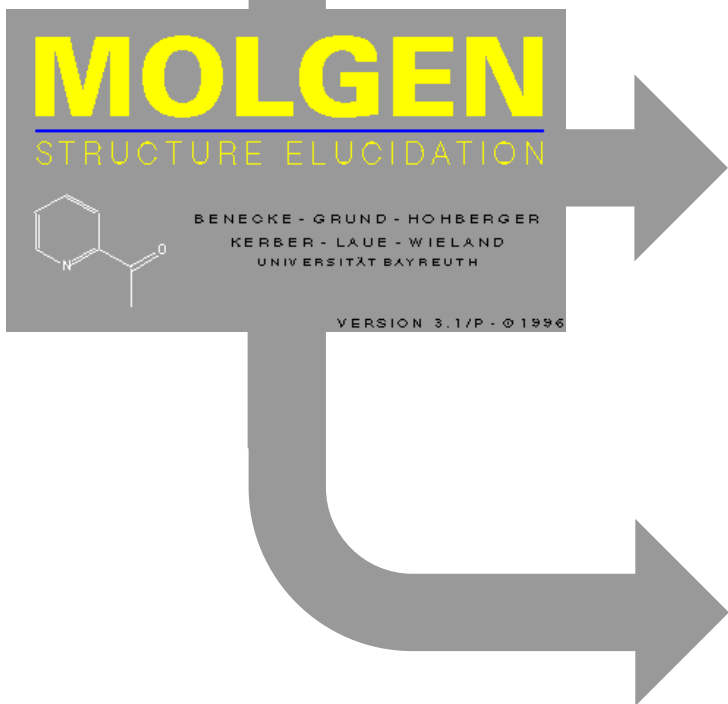


- Very low  $k$  values
- High sensitivity to impurities
- High(er?) GWPs

# Exploring the larger estimation-space

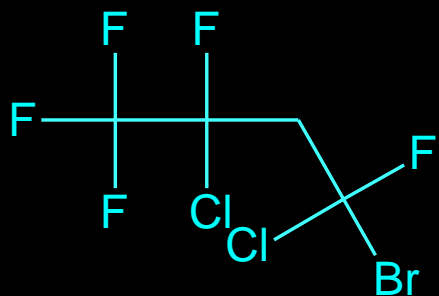
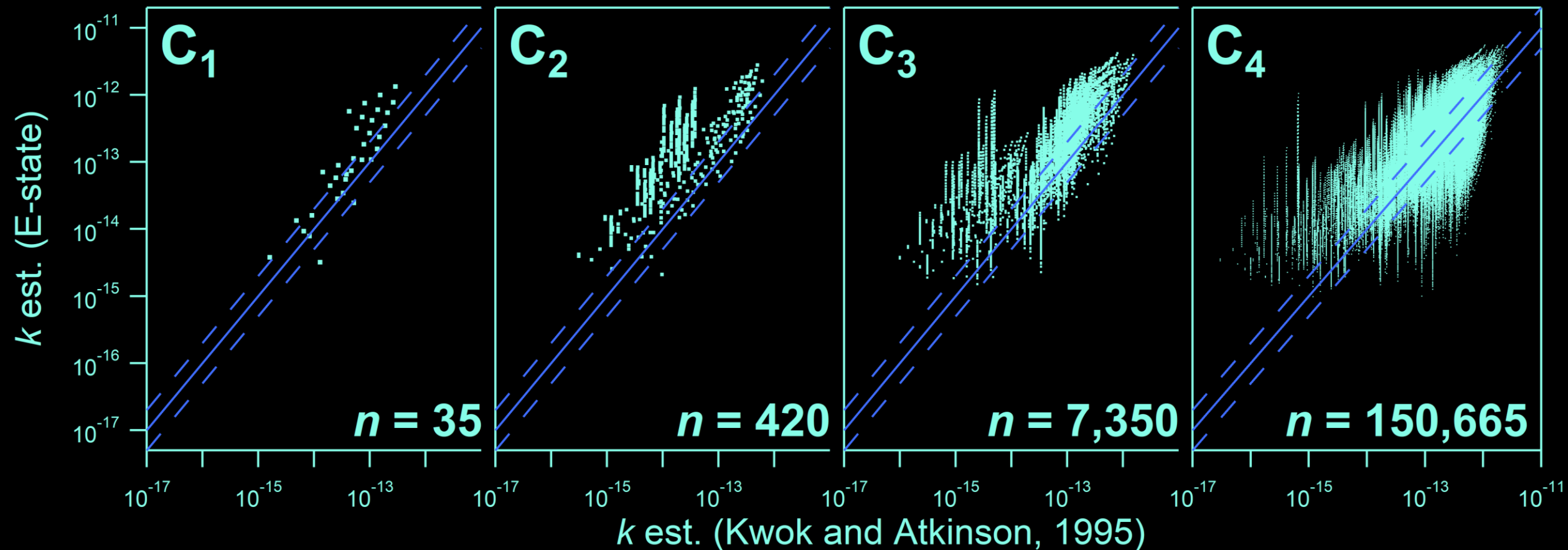


# Exploring the larger estimation-space

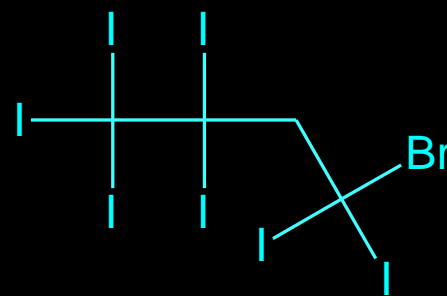


- Number of possible compounds presents several challenges
- Need for a fully enumerated list of haloalkanes (MOLGEN)
- Need to estimate large batches of molecules automatically + rapidly
- For Kwok and Atkinson, this can be done with EPI Suite™
- For E-states, this was done with some custom Python code + RDKit
- For IP this was done using some automated scripts in MOPAC2016™ and OpenBabel

# Is there a consensus between methods?



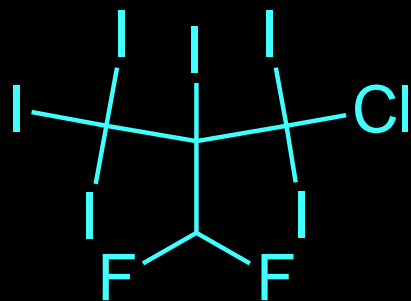
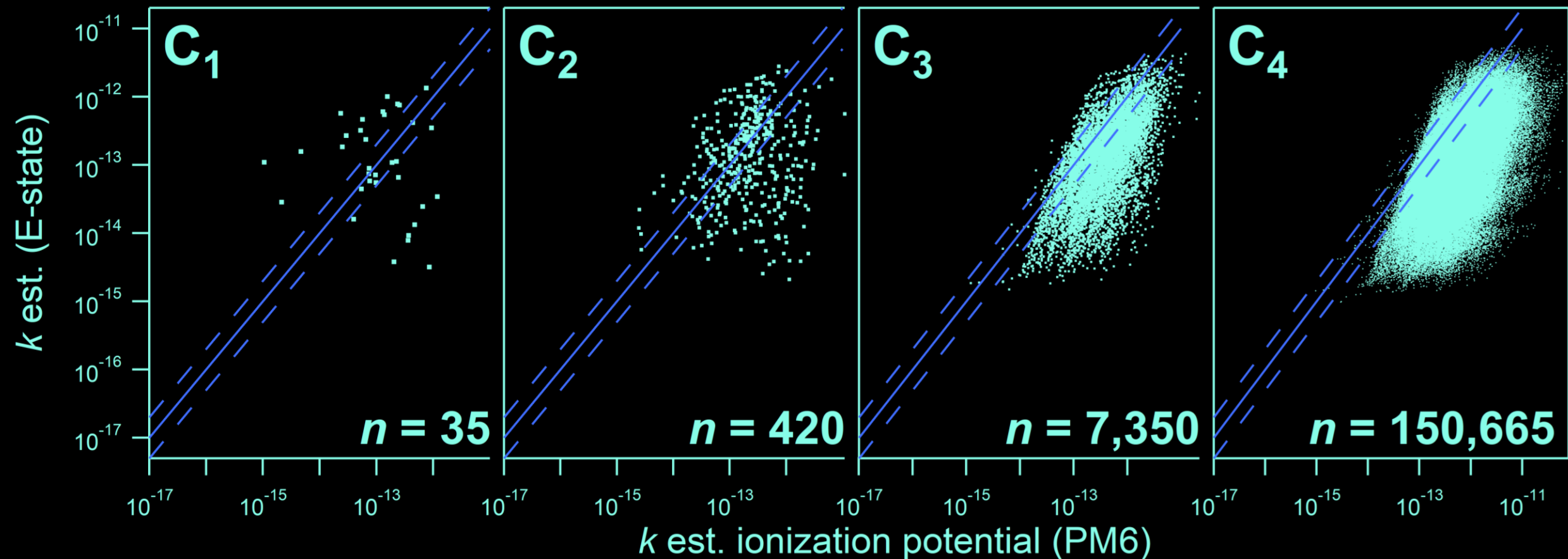
$$k_{\text{EPI Suite}} = 1.61 \times 10^{-14}$$
$$k_{\text{E-state}} = 8.10 \times 10^{-15}$$



$$k_{\text{EPI Suite}} = 1.61 \times 10^{-14}$$
$$k_{\text{E-state}} = 7.00 \times 10^{-13}$$



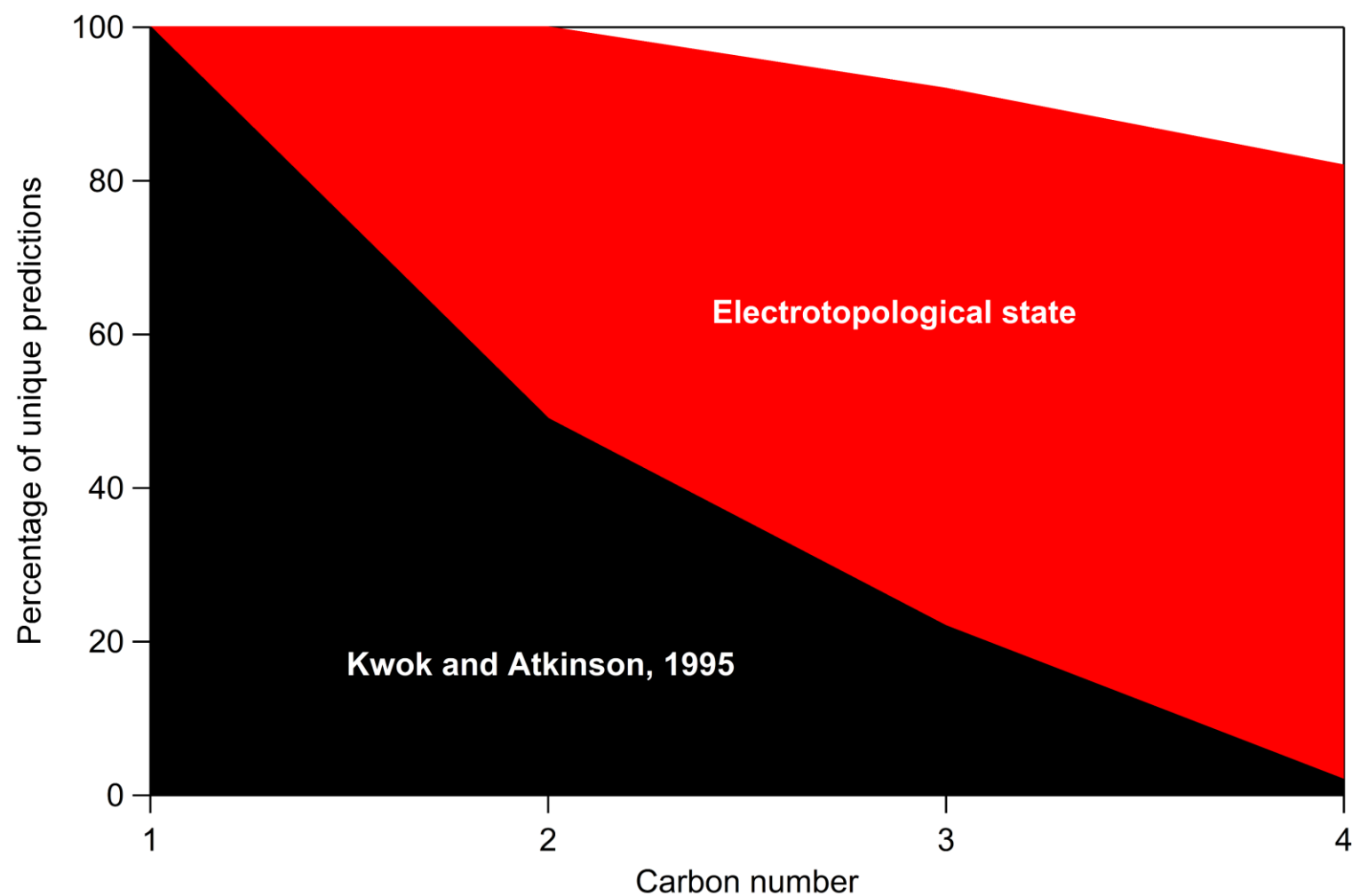
# Is there a consensus between methods?



$$k_{\text{IP}} = 1.66 \times 10^{-11}$$

$$k_{\text{E-state}} = 1.18 \times 10^{-14}$$

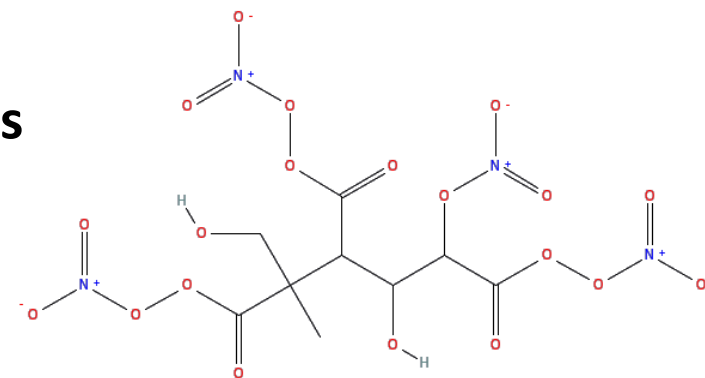
# Discriminating between structures



- Some algorithms are better at discriminating than others
- As the number of isomers grows so does the potential for degeneracy
- We can see that even at isomer counts (150,665), E-states produces a high % of unique predictions

# Perspectives for atmospheric mechanisms 1

- Detailed atmospheric chemical mechanisms contain **many** species
- The vast majority of these species have **no kinetic measurements**
- Therefore we need **estimation/ rapid computation of rate coefficients**
- Today, I showed you a **simple** case of a group of chemicals that I think we understand comparatively well and how estimation methods compare
- There is a reasonable consensus for the molecules that have been studied – leading to an interesting possibility for an ensemble
- **Yet, when we extend this approach to the broader estimation space we start to see bigger discrepancies – what does this mean for the more challenging molecules encountered in atmospheric chemistry?**



# Perspectives for atmospheric mechanisms 2

- But I think we can end on a positive note:
- We are working in a time where experimental data is more available
- There is more **information** than perhaps we realise
- New tools are available for performing calculations on large lists of molecules
- New sources of information are becoming available such as machine-learning
- We should extend our techniques to more atmospherically relevant species and continue to **explore the larger estimation space**
- **And we should not forget that where disagreements are encountered we should try to resolve these through experimental inquiry**

*Thank you for your attention,  
and thanks to my colleagues in the CRC SAR panel for helpful discussions*