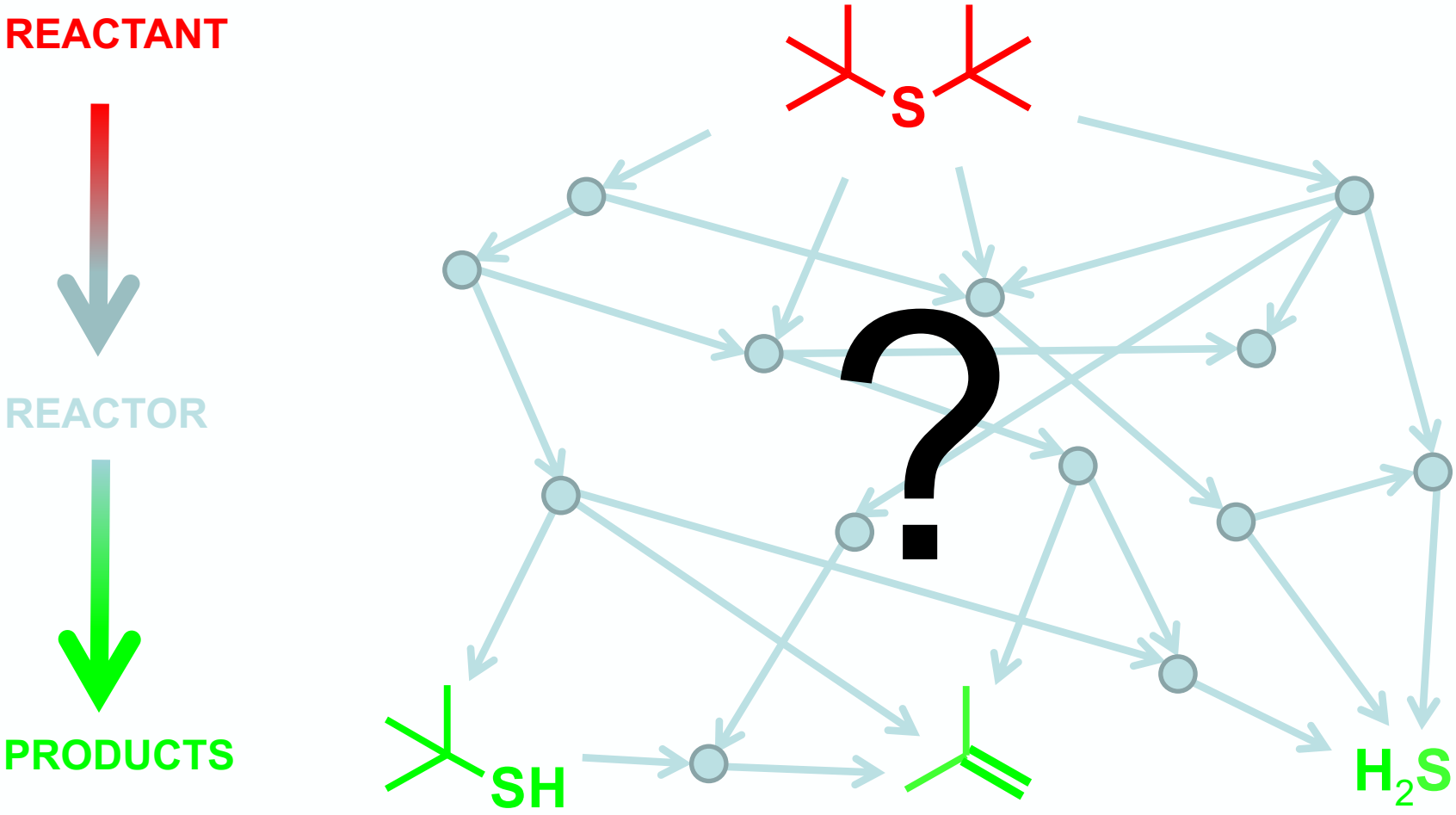
# Thermo in RMG

# Group additivity

# The experimentalist's problem...




**What are the dominant reaction paths? What are the intermediates?**

# Modeling reaction mechanisms

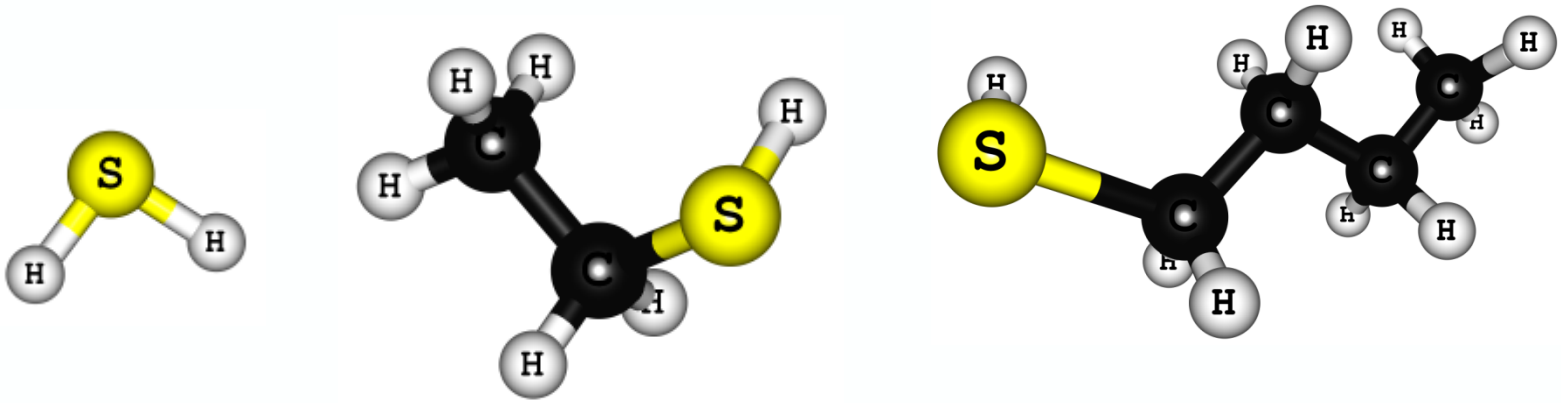
- Required information
  - How fast does each reaction proceed? (rate coefficients *k*)
  - Energy balance: during reactions energy is exchanged between the environment and the compounds (*H*, *S* and $c_p$)
- What is available?
  - If we are lucky:
    - experimental data (NIST)
    - (empirical) correlations between kinetic and thermodynamic parameters, e.g. Evans-Polanyi
  - Structural information: atoms and their connectivity (bonds)

# The benefits of sizing down

**Chinese characters:**
50000+, each word one character

| | | | |
|---|---|---|---|
| anestri | asterin | eranist | nastier |
| ratines | resiant | restain | retains |
| retinas | retsina | sainter | stainer |
| | starnie | stearin | |

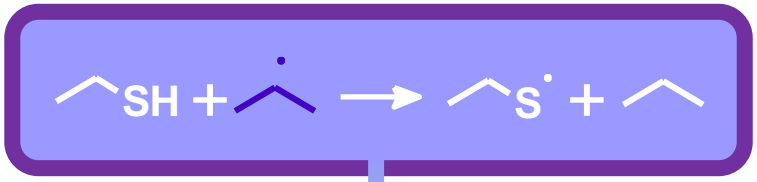
**Alphabet:**
26 letters to make all possible words




Property (molecule) = ∑ property (subsystem)
TRANSFERABILITY

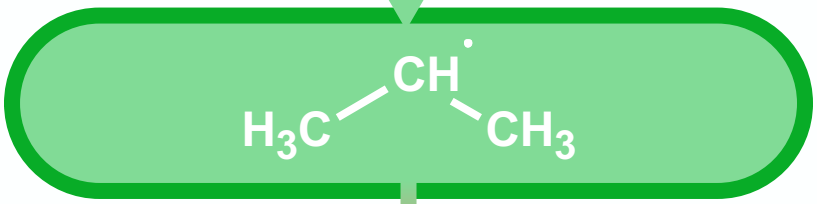
# Towards smaller entities

**REACTIONS**



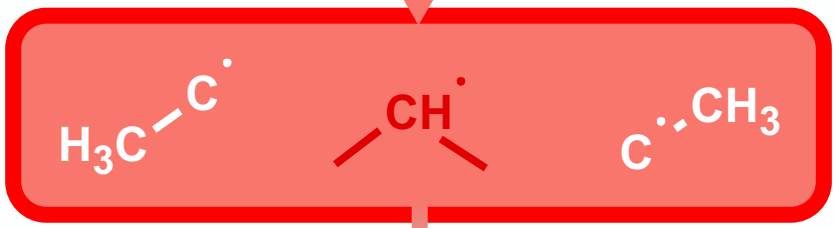

$$K_r = e^{-\frac{\Delta_r G^\circ}{RT}}$$

**COMPOUNDS**



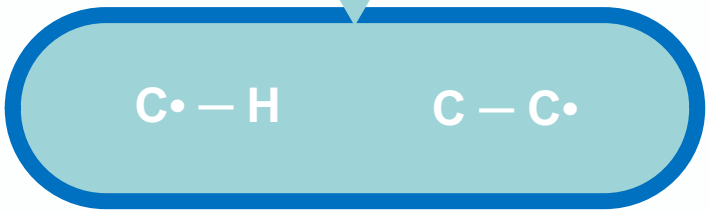

$$\Delta_r G^\circ = \sum_i \nu_i \, \Delta_f G^\circ(i)$$

**GROUPS**




$$\Delta_f G^\circ(i) = \sum_g \Delta_{gr} G^\circ(g)$$

**BONDS**

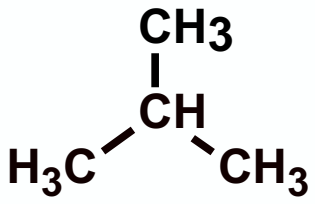



$$\Delta_{gr} G^\circ(g) = \sum_b \Delta_{bond} G^\circ(b)$$

# Group contribution methods for thermo

GA methods differ:

a) Definition of the groups




Joback : 1 × CH + 3 × CH3

Benson : 1 × C-(C)$_3$(H) + 3 × C-(C)(H)$_3$

b) The way properties are calculated from the GAVs

Joback :

$$\Delta_f H^\circ\ (298\ \mathrm{K}) = 68.3 + \sum_i \mathrm{GAV}_H$$

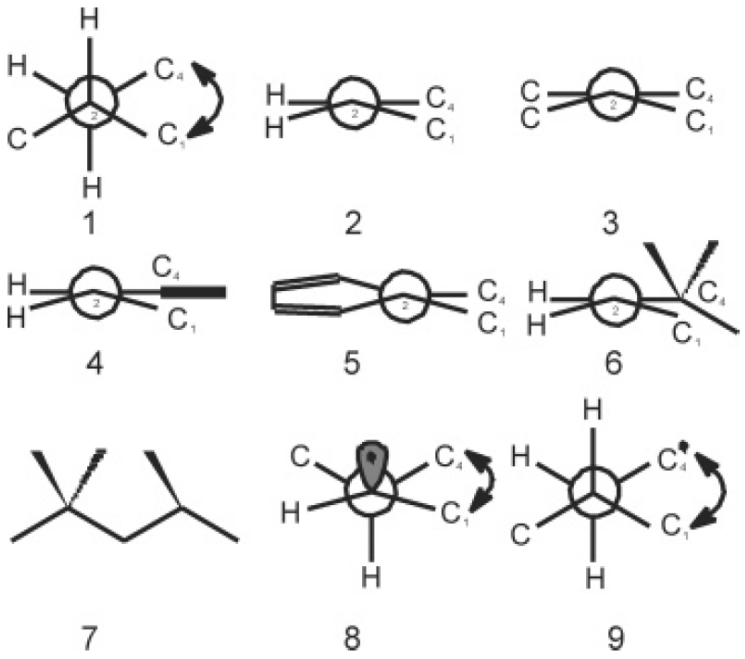
$$S^\circ\ (298\ \mathrm{K}) = (14.4 + \sum_i \mathrm{GAV}_{H-G})/298$$

Benson:

$$\Delta_f H^\circ\ (298\ \mathrm{K}) = \sum_i \mathrm{GAV}_H$$

$$S^\circ\ (298\ \mathrm{K}) = \sum_i \mathrm{GAV}_S + R \ln (n_{opt}/\sigma)$$

# Benson group additivity




- NNIs are often hard to recognize (e.g. Khan et al. J Phys Chem A 2009 113 5176-5194)
- Various java applets are available:

http://webbook.nist.gov/chemistry/grp-add/

http://rmg.mit.edu/molecule_search

# RMG thermo

# Overview of the thermo libraries

- Groups: Benson groups
- Radicals: hydrogen bond increments

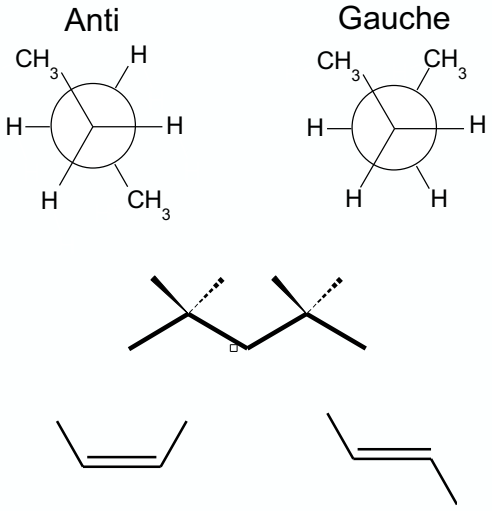
**WELL DEFINED!**

$$\mathrm{HBI}(\Delta_f H^\circ) = \Delta_f H^\circ_{\mathrm{AI}}(\mathrm{R}^\bullet) - \Delta_f H^\circ_{\mathrm{GA}}(\mathrm{RH})$$

$$\mathrm{HBI}(S^\circ_{\mathrm{int}}) = S^\circ_{\mathrm{int,AI}}(\mathrm{R}^\bullet) - S^\circ_{\mathrm{int,GA}}(\mathrm{RH})$$

$$\mathrm{HBI}(c^\circ_{\mathrm{p}}) = c^\circ_{\mathrm{p,AI}}(\mathrm{R}^\bullet) - c^\circ_{\mathrm{p,GA}}(\mathrm{RH})$$

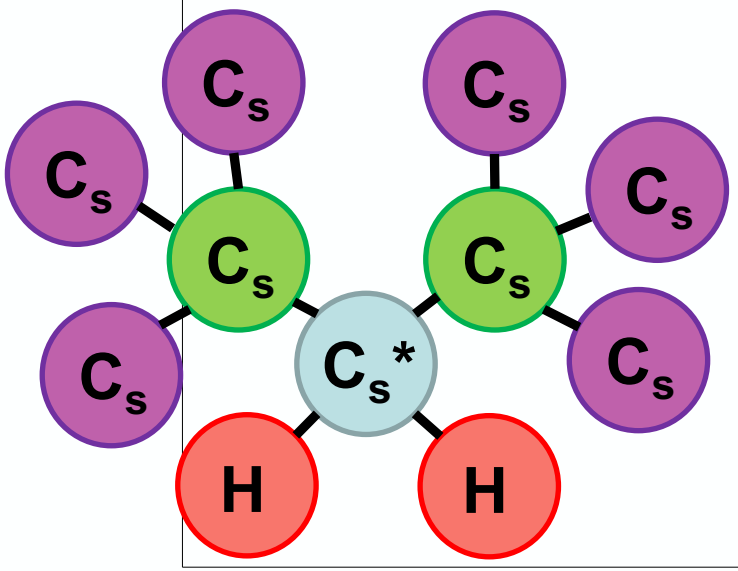
- Gauche
- 1,5
- Other: cis / trans
- Ring
- Polycyclic




**LESS WELL DEFINED…**

# Implementation

- Graph based recognition of (corrector) groups




2 Gauche Q - Q
1 * Cs 0 {2,S} {3,S} {4,S} {5,S}
2 Cs 0 {1,S} {6,S} {7,S} {8,S}
3 Cs 0 {1,S} {9,S} {10,S} {11,S}
4 {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {1,S}
5 {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {1,S}
6 Cs 0 {2,S}
7 Cs 0 {2,S}
8 Cs 0 {2,S}
9 Cs 0 {3,S}
10 Cs 0 {3,S}
11 Cs 0 {3,S}

# Group library

- Library – Dictionary – Tree files containing the group values, group definitions and dependencies respectively
- In contrast to kinetic libraries there is no automated filling of the tree
  - Every node needs a GA value or reference
  - Example:

Group + GAV(H) + GAV(S) + GAV($c_p$ 300 400 500 600 800 1000 1500)
dH dS dcp Comments

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1100 | S | **Ss-CsCs** | | | | | | |
| 1132 | Ss-HH | -5.37 | 50.52 | 8.15 | 8.48 | 8.85 | 9.26 | 10.08 … |
| 1133 | Ss-CsH | 5.05 | 33.68 | 6.17 | 6.22 | 6.40 | 6.65 | 7.18 … |
| 1134 | **Ss-CsCs** | 6.27 | 31.59 | 6.22 | 6.87 | 7.26 | 7.55 | 7.94 …. |

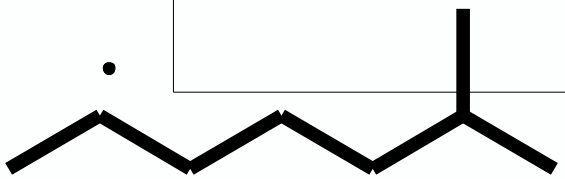
# Group library

- Changing this library should –in my opinion– be avoided to all cost
  - Linear dependencies
  - Corrections (gauche, cis/trans…) depend on the values assigned to the groups

  **Modifying the group values can have far reaching consequences!**
- Adding new groups (S, N, Cl, Br…) can be done relatively easy

# Radical library

- Hydrogen bond increments
  - Advantage over radical centered groups (e.g. C•-(C)(H)$_3$) in flexibility and interpretability
  - Describe the influence of the loss of a hydrogen atom on the enthalpy (related to BDE), entropy and $c_p°$
  - One drawback is their use for biradicals as they normally account for the spin multiplicity of 2



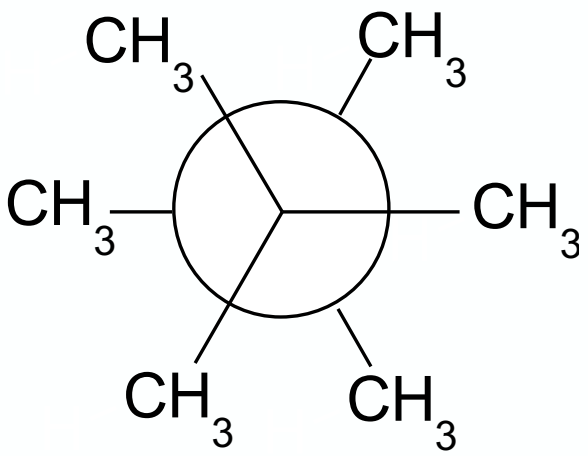
$$\mathrm{HBI}(\Delta_f H^\circ) = \Delta_f H^\circ_{\mathrm{AI}}(\mathrm{R}^\bullet) - \Delta_f H^\circ_{\mathrm{GA}}(\mathrm{RH})$$

$$\mathrm{HBI}(S^\circ_{\mathrm{int}}) = S^\circ_{\mathrm{int,AI}}(\mathrm{R}^\bullet) - S^\circ_{\mathrm{int,GA}}(\mathrm{RH})$$

$$\mathrm{HBI}(c^\circ_{\mathrm{p}}) = c^\circ_{\mathrm{p,AI}}(\mathrm{R}^\bullet) - c^\circ_{\mathrm{p,GA}}(\mathrm{RH})$$

# Gauche library

Read the manuscript and what the corrections account for!



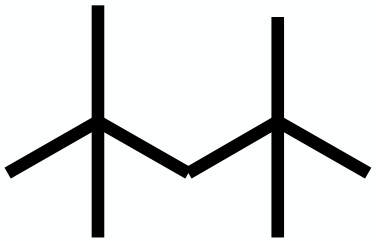

**RMG** **Benson** **Sabbe**

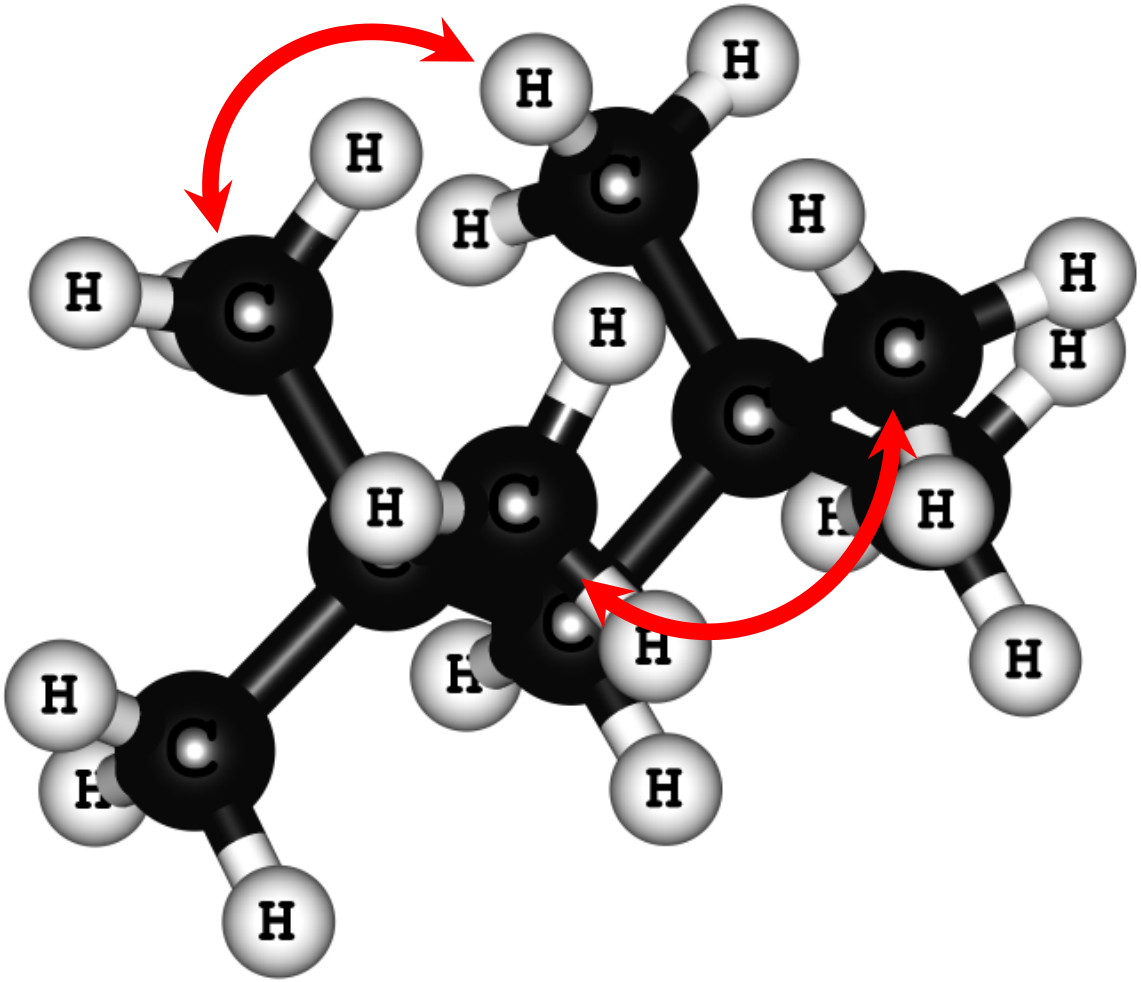
2 x Cs(Cs(CsCsCs)CsCsCs)

**= 6 GAUCHE CORR.** **= 6 GAUCHE CORR.** **= 8 GAUCHE CORR.**

# 1,5 library

- Crowded methyls





# Ring/Polyring library

- Ring corrections are added to account for the strain caused by a ring

$$\Delta_f H(\text{cyclopentane}) = 5\text{GAV}(\text{C}-(\text{C})_2(\text{H})_2) + \text{RSC}(5\text{ring})$$

- Other NNI´s are accounted for as they would appear in the open ring analogue

- Polyring tree deviates slightly from the other trees as priority can be assigned:

  When two hits are received RMG only uses the first hit!

# Thermo from AI

# How do theoreticians calculate thermo?

- Partition functions describe the statistical properties of a system and can be used to derive thermodynamic properties (enthalpy, entropy, heat capacity)

$$q = \sum_{i=1}^{n} \sum_{j=1}^{\infty} e^{-\frac{\varepsilon_{i,j}}{k_B T}}$$

*n* = number of modes

$$S = R \ln(q\, e) + RT \left( \frac{\partial \ln q}{\partial T} \right)_V$$

$$H = RT^2 \left( \frac{\partial \ln q}{\partial T} \right)_V + RT$$

$$C_p = \left( \frac{\partial H}{\partial T} \right)_p$$

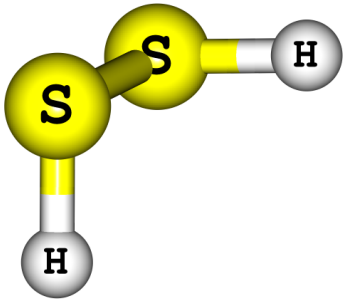
- But there is a snatch…

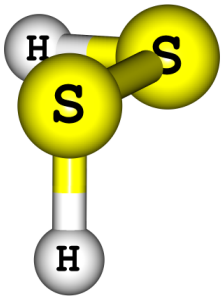
IUPAC:
*"Chemical species = an ensemble of chemically identical molecular entities that can explore the same set of molecular energy levels on the time scale of the experiment."*

# Species = f(*T*)

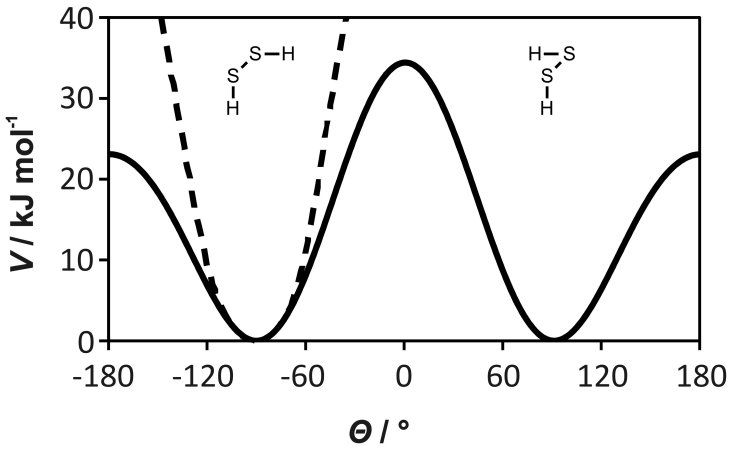
One species! Its *S*° = 258.2 J mol$^{-1}$ K$^{-1}$.

I detect two different species, their *S*° = 252.4 J mol$^{-1}$ K$^{-1}$.

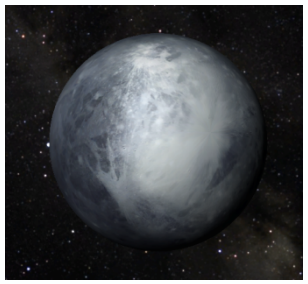



**273 K**



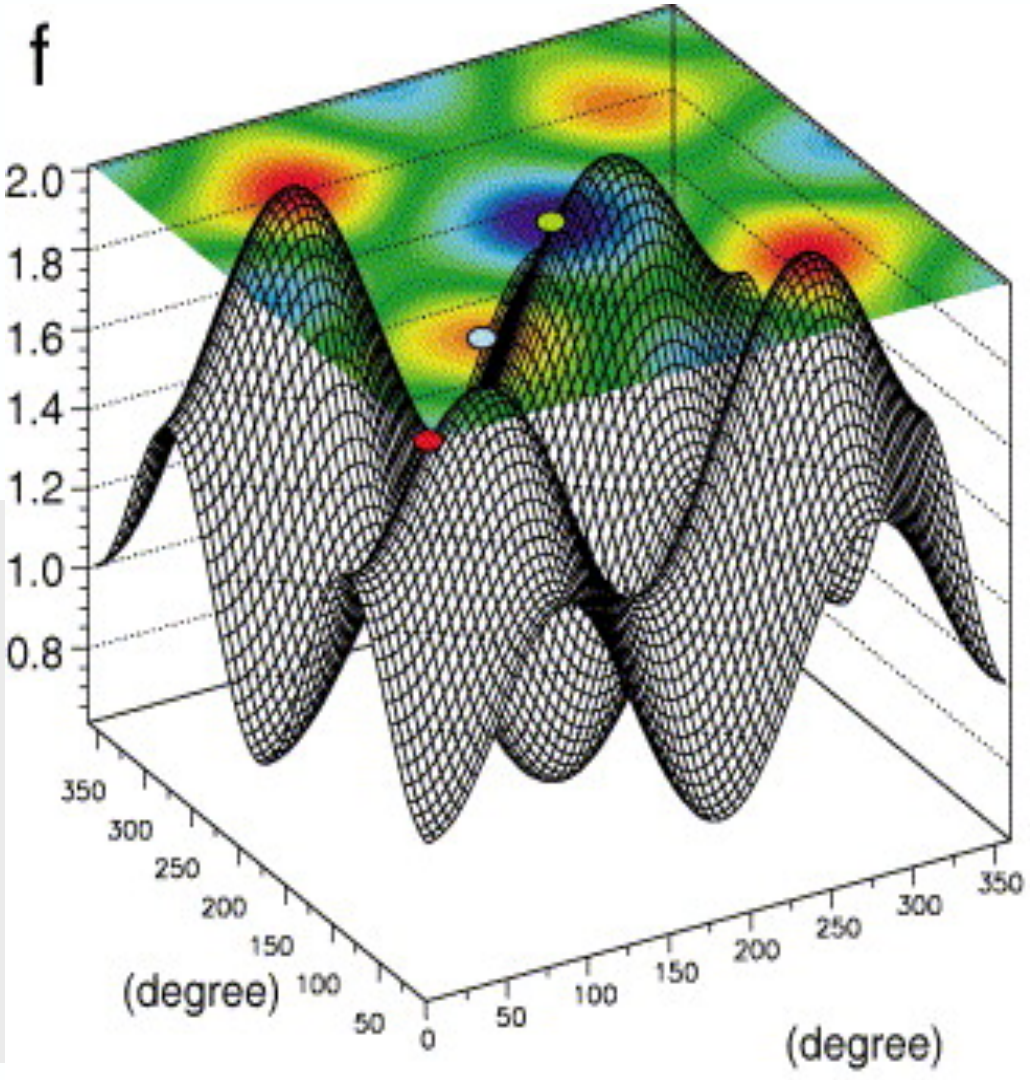

**44 K**

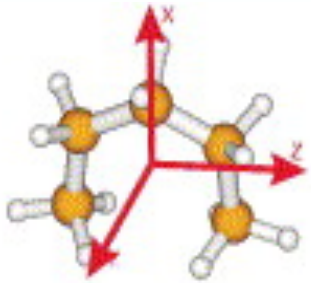


**τ = 1.7 ns**

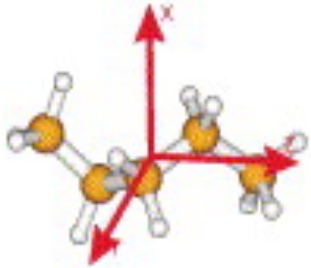
**τ = 9 days**

# What's in a name?

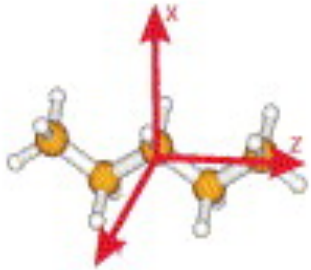



Ix=519,6 a.u.
Iy=747,2 a.u.
Iz=282,5 a.u.



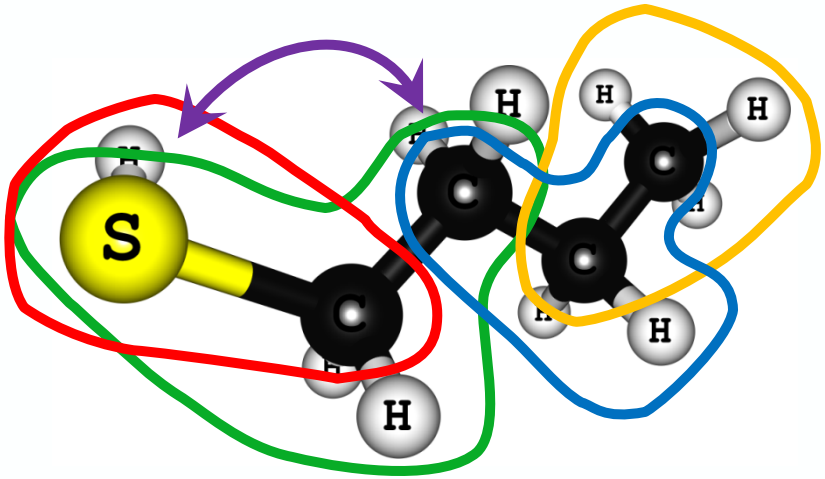

Ix=814,6 a.u.
Iy=827,2 a.u.
Iz=155,0 a.u.




Ix=937,9 a.u.
Iy=987,8 a.u.
Iz=105,5 a.u.

# Benson group additivity

- Benson divided molecules into (functional) groups




Group = central atom with all of its neighboring ligands

e.g. **S-(C)(H)**
**C-(S)(C)$(H)_2$**
**C-$(C)_2(H)_2$**
**C-(C)$(H)_3$**

+ non-nearest neighbor interactions (**NNIs**)

$$\Delta_f H°(298K) = \sum_i GAV_{\Delta_f H°}(X_i)$$

$$S°(298K) = \sum_i GAV_{S°}(X_i) - R\ln(\sigma / n_{opt})$$

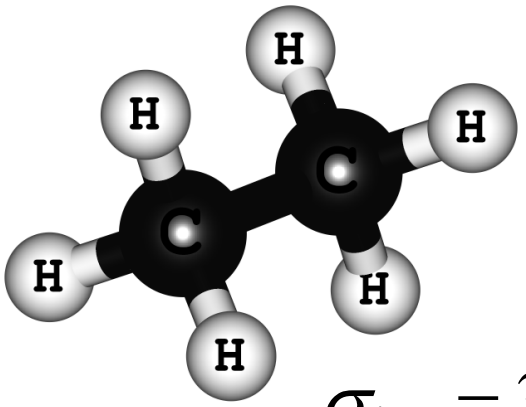
$$C_p°(T) = \sum_i GAV_{C_p°}(X_i)$$

# Symmetry numbers

- Symmetry numbers are used to avoid overcounting of equivalent energy levels

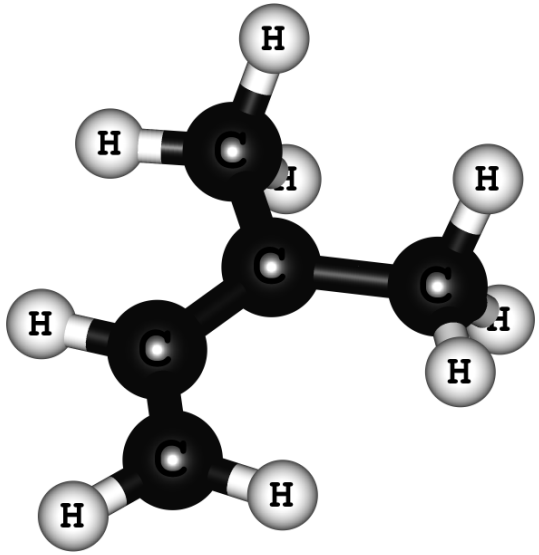
$$\sigma_{\text{tot}} = \sigma_{\text{int}}\sigma_{\text{ext}}$$

- $\sigma/n_{\text{opt}}$ can depend on the interpretation of the researcher




$$\sigma_{\text{int}} = 3$$

$$\sigma_{\text{ext}} = 6\,(\text{D}_{3\text{h}})$$




$$\sigma_{\text{int}} = 1 \text{ or } 2$$

$$\sigma_{\text{ext}} = 1$$

# Derivation of the GAV's from ab initio data

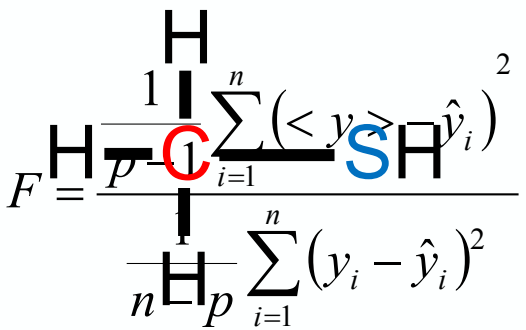
- Construction of a matrix **X** linking the GAV's with the molecules in the training set

| | $\Delta_f H°$ | S-(C)(H) | C-(S)(C)(H)$_2$ | C-(C)$_2$(H)$_2$ | C-(C)(H)$_3$ |
|---|---|---|---|---|---|
| $HSCH_2CH_3$ | -47.3 | 1 | 1 | 0 | 1 |
| $HSCH_2CH_2CH_3$ | -68.2 | 1 | 1 | 1 | 1 |
| $HSCH_2CH_2CH_2CH_3$ | -88.1 | 1 | 1 | 2 | 1 |

= set of linear equations

- Linear regression + statistics
  - Avoiding linear dependencies
  - Minimizing sum of squares: $\sum(y_i\text{-}\hat{y}_i)^2$

$$\overline{GAV} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\overline{y}$$

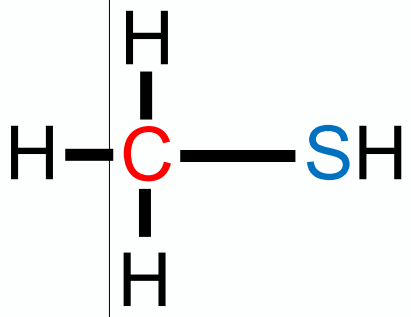


$$F = \frac{\frac{1}{p-1}\sum_{i=1}^{n}(\langle y\rangle - \hat{y}_i)^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- Significance of regression

# Linear dependent set of equations

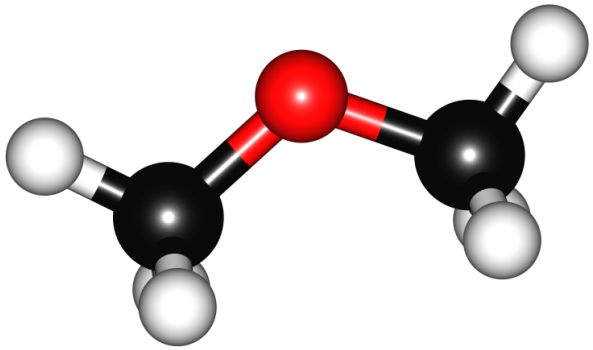
- Various groups contributions will be set to fixed values (of a similar group)



$C\text{-}(S)(H)_3$ $S\text{-}(C)(H)$

**a C-centered group having a S ligand will always be accompagnied by a S-(C) group and can not be solved independently!**

- When combining various sets of GAV`s care need to be taken that those fixed groups were assigned the same value

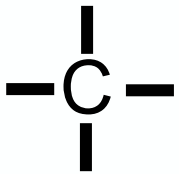


$C\text{-}(O)(H)_3 + O\text{-}(C)_2$

Khan et al. – 10.00 – 10.00 – 23.18 = –43.2 kcal mol$^{-1}$
Cohen et al. – 10.00 – 10.00 – 23.8 = –43.8 kcal mol$^{-1}$

# Benson group contribution method

- Hydrocarbons and organosulfur compounds
  - MAD($\Delta_f H°$) = 2 kJ mol$^{-1}$ Joback[1] and Gani[2] ~ 10 kJ mol$^{-1}$ [3]
  - MAD($S°$) = 5 J mol$^{-1}$ K$^{-1}$
  - MAD($c_p°$) = 3 J mol$^{-1}$ K$^{-1}$ ~ methods of Joback and Gani [3]
- Amount of groups increases fast with the inclusion of heteroelements



**H C $C_d$ $C_t$ $C_b$ = 70**
**+ O CO = 210**
**+ S CS = 495**

- What to do if groups are not available?

C-(S)(C)($H_2$) = -20.68 kJ mol$^{-1}$ C-(S)($C_2$)(H) = -8.29 kJ mol$^{-1}$

C-(S)($C_t$)($H_2$) = -11.27 kJ mol$^{-1}$ C-(S)(C)($C_t$)(H) = ? kJ mol$^{-1}$

**-8.29 +(-11.27 – (-20.68)) = 1.12 kJ mol$^{-1}$**

[1] Joback and Reid *Chem. Eng. Comm.* 57: 233-243 (1987), [2] Constantinou and Gani *Aiche J.* 40: 1697-170 (1994), [3] discussed in "The properties of gases and liquids" by Poling, Prausnitz and O'connell (2000)

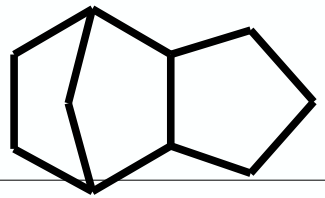
# What if I want to calculate 1 ring correction?

# Atomic/bond corr vs Reaction

- CBS-QB3 calculation:

  JP-10 H (298 K) = -389.872419 h

- Calculated energy contains errors in relativistic effects, diagonal BO corrections, extrapolation errors… but errors tend to be structure specific
- Error canceling transformations: <u>reaction enthalpies</u>



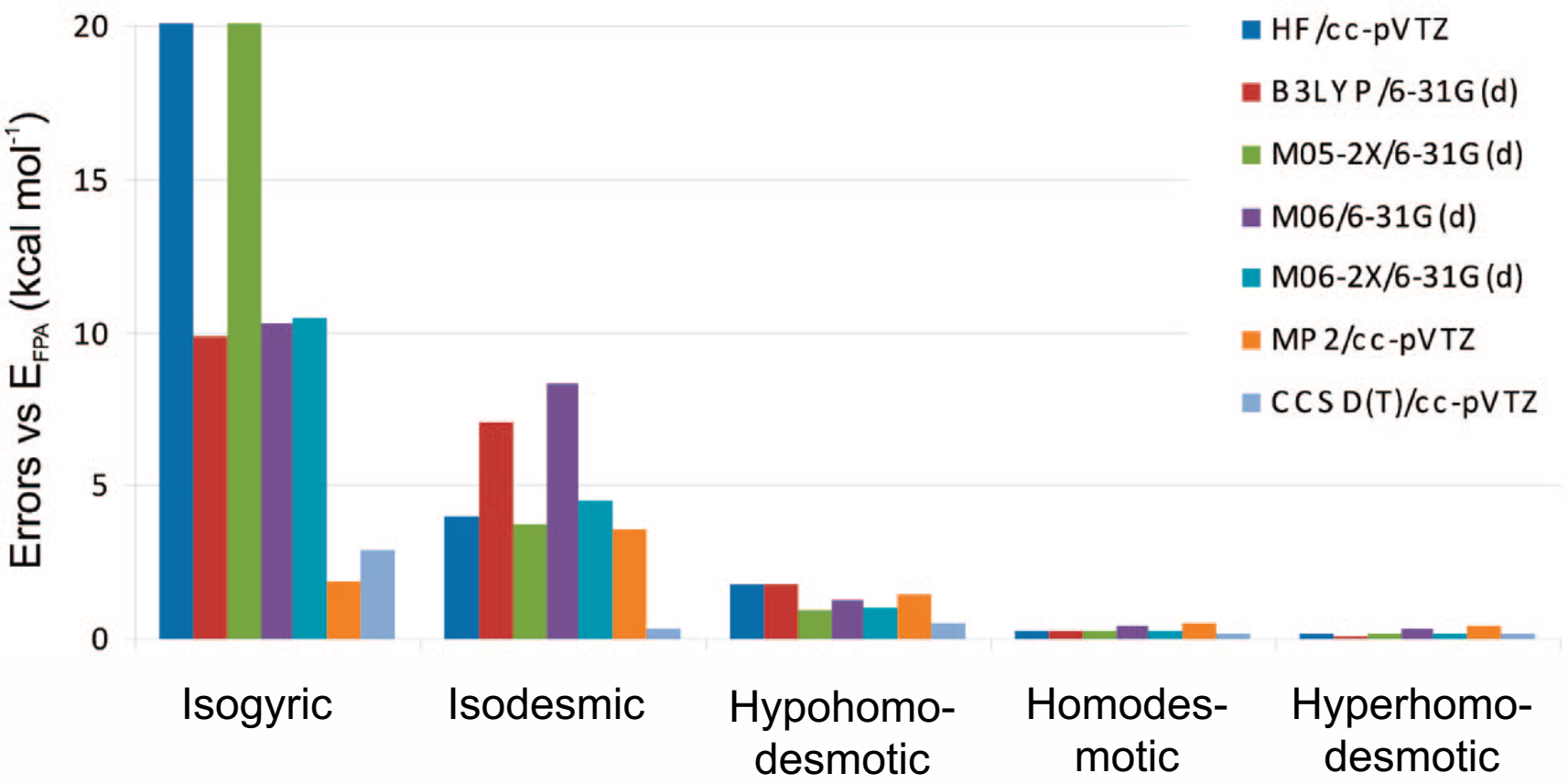
$$+ \ 11\,H_2 \longrightarrow 10\,CH_4$$

**Isogyric reaction : number of electron pairs is conserved**
**= AAC (atom additive corrections)**

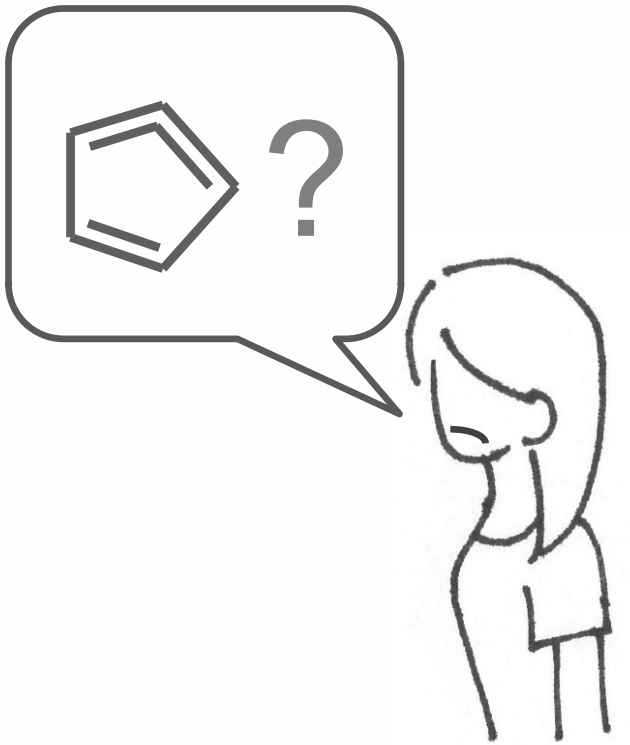
# Method selection

This image cannot currently be displayed.

**Isodesmic reaction : type of bonds is conserved**
**Hypohomodesmotic : same hybridization (eq sp, sp2 and sp3 C atoms)**
**Homodesmotic : bonds involving same hybridization**



# Routine




- Check if corrections are applied and correctly defined = ThermodataEstimator
- Ring correction = difference between RMG predicted value and the calculated values
- Add node to tree, library and dictionary
- Check if RMG estimate corresponds with calculated value

# Routine

## Isodesmic


| | | | | |
|---|---|---|---|---|
| EXP: | 190.90 | -10.8 ± 1.0 | ? | -74.6 ± 0.3 |

$\Delta_r H(298\ K) = 220.8\ kJ\ mol^{-1}$

$\Delta_f H(298\ K) = 550.1\ kJ\ mol^{-1}$

RSC (*H*) = 293.3 kJ mol$^{-1}$

What if RMG predicts a wrong symmetry number?

# Advantages of inclusion in groups