small group group additivity

based on a presentation at GEOPIG meeting 2011-05-10

Jeffrey M. Dick

School of Earth and Space Exploration
Arizona State University

August 22, 2011

small group group additivity

- Additivity is useful for some tasks
  - Group contributions come from known properties of model compounds
  - Use them to estimate thermodynamic properties of compounds lacking experimental data
  - Can it be used for properties of interesting molecules, e.g. metabolites?
  - Challenge: Apply group additivity to highly substituted molecules

Scheme 1: Big Groups

ethane	2[-CH3]			
propane	2[-CH3]	+[-CH2-]		
ethanol	[-CH3]		+[-CH2OH]	
acetic acid	[-CH3]			+[COOH]

- These structural groups were used by Amend and Helgeson [1997] for predictions of properties of homologous series with carbon number ≥ 2.

Equations = Unknowns (Determined System)

A	$[-\mathrm{CH3}]$	$[-\mathrm{CH2}-]$	$[-\mathrm{CH2OH}]$	$[-\mathrm{COOH}]$		X		B
ethane	2					$X_{[-\mathrm{CH3}]}$		88.3
propane	2	1			$\times$	$X_{[-\mathrm{CH2}-]}$	$=$	110.6
ethanol	1		1			$X_{[-\mathrm{CH2OH}]}$		62.2
acetic acid	1			1		$X_{[-\mathrm{COOH}]}$		40.56

- Matrix (A) represents a group contribution model.
- Vector of knowns (B) here contains values of C_P° (standard molal heat capacity in cal K^{-1} mol^{-1}) for aqueous species.
- Known values are based on experiments (those here are available in thermodynamic database of CHNOSZ).
- Solve for X (in $A \times X = B$) to get group contribution values.

Equations = Unknowns (Determined System)

- Solve the system in R:

```
> A <- matrix(c(2, 2, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0,
    0, 0, 0, 1), 4)
> b <- c(88.3, 110.6, 62.2, 40.56)
> x <- solve(A, b)
> names(x) <- c("CH3", "CH2", "CH2OH", "COOH")
> x
```

```
  CH3   CH2 CH2OH  COOH
44.15 22.30 18.05 -3.59
```

- Compare with literature values [Amend and Helgeson, 1997] obtained from consideration of many more model compounds in homologous series.

```
> Cp.AH97 <- c(CH3 = 47, CH2 = 20.7, CH2OH = 17.28,
    COOH = -5.94, CHCH3 = 39.92)
> Cp.AH97
```

```
  CH3   CH2 CH2OH  COOH CHCH3
47.00 20.70 17.28 -5.94 39.92
```

Equations > Unknowns (Overdetermined System)

- Consider more species, as well as an additional group: ketone group, [-CO-].
- Heat capacities are reported values (mostly from experiments) in cal K^{-1} mol^{-1}.

	Cp	[-CH3]	[-CH2-]	[-CH2OH]	[-CO-]	[-COOH]
ethane	88.30	2				
propane	110.60	2	1			
butane	133.90	2	2			
ethanol	62.20	1		1		
1-propanol	84.30	1	1	1		
1-butanol	104.40	1	2	1		
1-pentanol	125.20	1	3	1		
acetone	57.70	2			1	
butanone	80.40	2	1		1	
3-pentanone	102.37	2	2		1	
2-heptanone	144.00	2	4		1	
acetic acid	40.56	1				1
propanoic acid	60.50	1	1			1
butanoic acid	80.50	1	2			1

Approximate Solutions for Overdetermined Systems

- Least-squares solution is possible with QR decomposition.
- Example below takes after help page for `qr.solve` in base R.
- Four equations, three unknowns

```
> set.seed(24)
> print(A <- matrix(runif(12), 4))
```

```
          [,1]      [,2]      [,3]
[1,] 0.2925740 0.6626196 0.8016306
[2,] 0.2248911 0.9204438 0.2547251
[3,] 0.7042230 0.2797356 0.6048889
[4,] 0.5188971 0.7638205 0.3707349
```

```
> b <- 1:4
> x <- qr.solve(A, b)
> as.numeric(A %*% x)
```

```
[1] 0.9023479 2.3515328 3.2438201 3.5718037
```

- $A \times X$ gives us an approximation of B.

Fitting 14 Species Using 5 Big Groups

- Read the group contribution matrix. Only take selected groups (iuse) at this time.

```
> file.big <- system.file("extdata/thermo/groups_big.csv",
    package = "CHNOSZ")
> A.big <- read.csv(file.big, check.names = FALSE,
    row.names = 1)
> iuse <- c("ethane", "propane", "butane", "ethanol",
    "1-propanol", "1-butanol", "1-pentanol", "acetone",
    "butanone", "3-pentanone", "2-heptanone", "acetic acid",
    "propanoic acid", "butanoic acid")
> A.big <- A.big[iuse, ]
> A.big[is.na(A.big)] <- 0
```

- Get the values of heat capacity of the aqueous model species. CHNOSZ warns about some inconsistencies between heat capacities listed in the database and values calculated using equations-of-state parameters.

```
> ispecies <- info(rownames(A.big), quiet = TRUE)
> cp.species <- info(ispecies)$Cp
```

```
checkEOS: Cp of ethane aq differs by -9.4 from tabulated value
checkEOS: Cp of propane aq differs by -15.14 from tabulated value
checkEOS: V of propane aq differs by -1.45 from tabulated value
checkEOS: Cp of propanoic acid aq differs by 1.42 from tabulated value
```

Performance of "Big Groups" Model

- Calculate least-squares solution.

```
> cp.big <- qr.solve(A.big, cp.species)
> cp.big
```

```
    [-CH3]     [-CH2-]    [-CH2OH]      [-CO-]     [-COOH]
 44.807300  21.318732  17.239601 -30.804883  -5.606033
```

- What is the root mean square deviation (RMSD) between predicted and known values? (Note: `rmsd` used here is a simple function in CHNOSZ, not base R.)

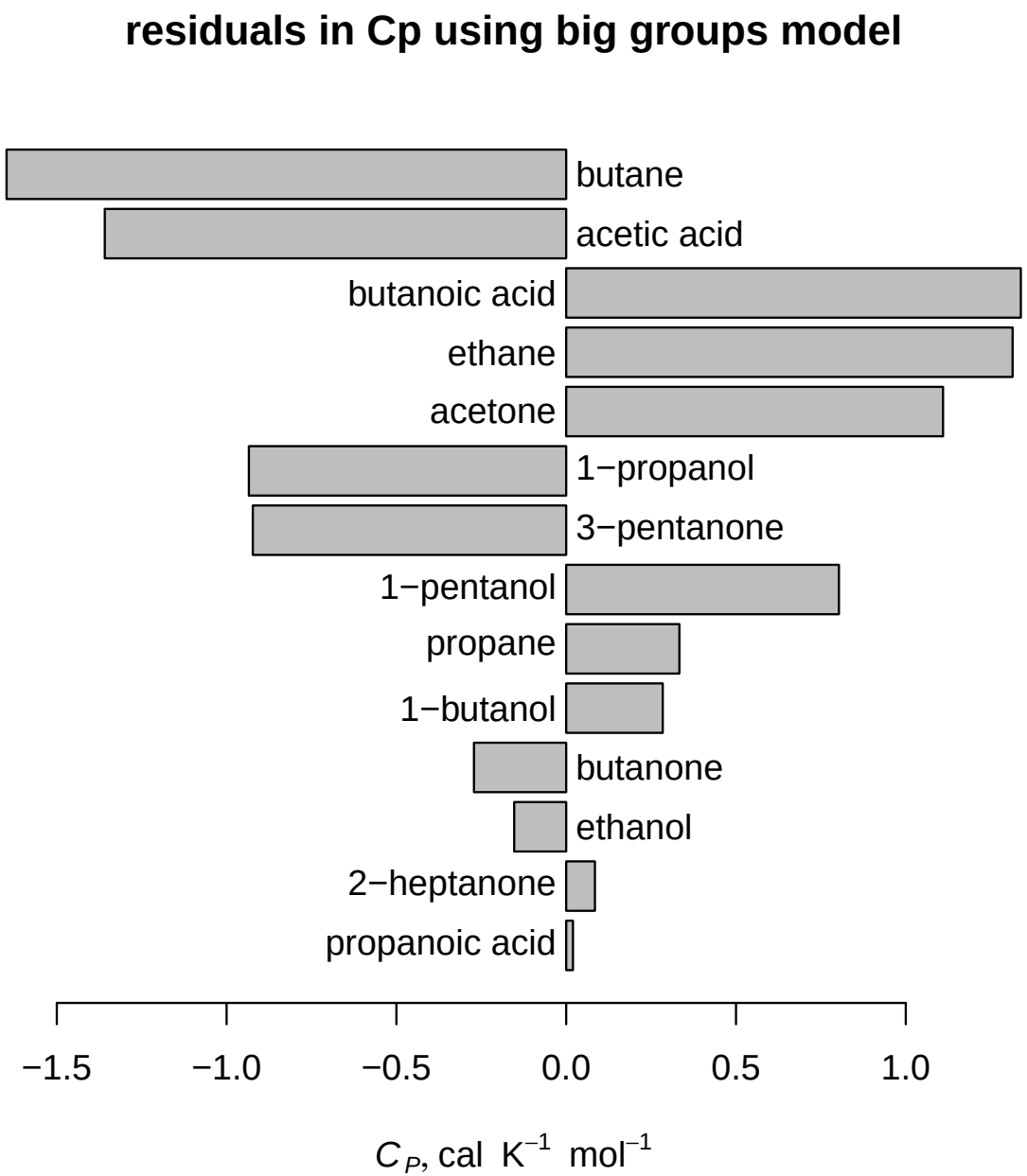
```
> pred.big <- as.matrix(A.big) %*% cp.big
> rmsd <- rmsd(pred.big, cp.species)
> rmsd
```

```
[1] 0.9250488
```

- We can also calculate and plot the residuals. `residualsplot` in CHNOSZ uses R's `barchart` function and adds labels and a title.

```
> residuals <- pred.big - cp.species
> names(residuals) <- rownames(A.big)
> residuals <- residuals
> residualsplot(residuals, "Cp", "big groups (added)")
```

Big Groups ... when added



small group group additivity

- Additivity is useful for some tasks
- Big groups have their place
  - Large structural groups; no nearest-neighbor effects
  - Can we apply them to a wider variety of compounds?

Equations > Unknowns (Overdetermined System)

- Now more species. For some of them we have to use negative group contributions.

	Cp	[-CH3]	[-CH2-]	[-CH2OH]	[-CO-]	[-COOH]
(1) ethane	88.30	2				
(2) propane	110.60	2	1			
(3) butane	133.90	2	2			
(4) methanol	37.80	1	-1	1		
(5) ethanol	62.20	1		1		
(6) 1-propanol	84.30	1	1	1		
(7) 1-butanol	104.40	1	2	1		
(8) 1-pentanol	125.20	1	3	1		
(9) 3-pentanol	130.21	1	3	1		
(10) 2-propanol	86.28	1	1	1		
(11) 2-butanol	107.34	1	2	1		
(12) 2-pentanol	131.17	1	3	1		
(13) acetone	57.70	2			1	
(14) butanone	80.40	2	1		1	
(15) 3-pentanone	102.37	2	2		1	
(16) 2-heptanone	144.00	2	4		1	
(17) acetaldehyde	34.90	2	-1		1	
(18) acetic acid	40.56	1				1
(19) propanoic acid	60.50	1	1			1
(20) butanoic acid	80.50	1	2			1
(21) 2-methylpropanoic acid	79.83	3	-1			1
(22) 2,4-dimethyl-3-pentanone	100.86	6	-2		1	
(23) lactic acid	66.70		1	1		1
(24) citric acid	73.47	-2	4	1		3

Fitting 24 Species Using 5 Big Groups

- Read the group contribution matrix.

```
> file.big <- system.file("extdata/thermo/groups_big.csv",
    package = "CHNOSZ")
> A.big <- read.csv(file.big, check.names = FALSE,
    row.names = 1)
> A.big[is.na(A.big)] <- 0
```

- Get the values of heat capacity of the aqueous model species.

```
> ispecies <- info(rownames(A.big), quiet = TRUE)
> cp.species <- info(ispecies, quiet = TRUE)$Cp
```

- Calculate least-squares solution.

```
> i.big <- 1:24
> cp.big <- qr.solve(A.big[i.big, ], cp.species[i.big])
> cp.big
```

```
    [-CH3]    [-CH2-]   [-CH2OH]     [-CO-]    [-COOH]
32.924314  25.624264  24.839122 -18.176013   4.627111
```

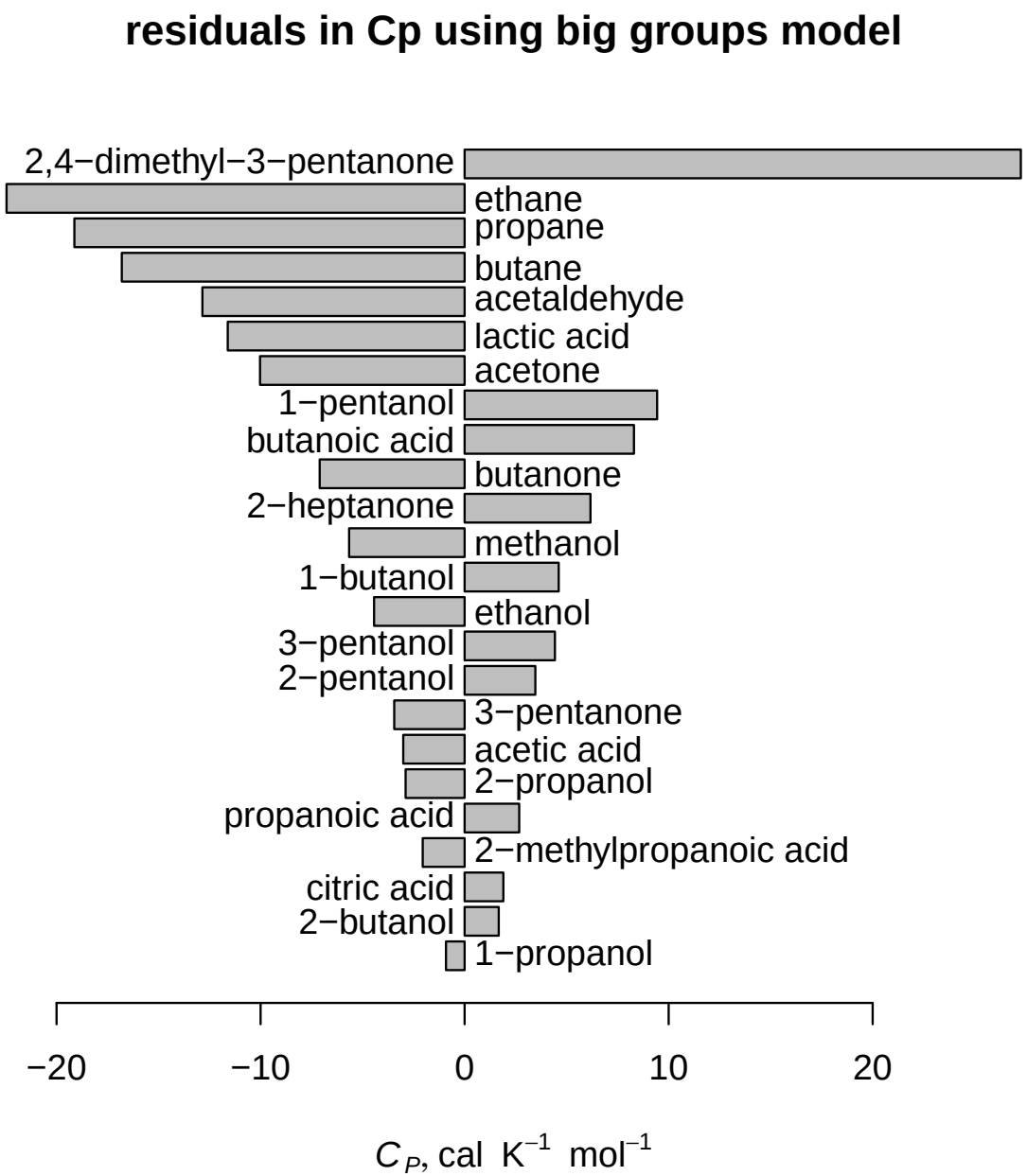
- Calculate the RMSD

```
> pred.big <- as.matrix(A.big) %*% cp.big
> rmsd <- rmsd(pred.big[i.big], cp.species[i.big])
> rmsd
```

```
[1] 10.59658
```

- Plot the residuals.

Big Groups ... when added and subtracted



small group group additivity

- Additivity is useful for some tasks
- Big groups have their place
  - Large structural groups; no nearest-neighbor effects
  - Can we apply them to a wider variety of compounds?
  - Have negative group contributions for some compounds of interest (e.g. citric acid)
  - Large error for this type of calculation. Keep experimental error in mind too!

scheme 2: small groups

ethane	2 x C-(H)3(C)
propane	2 x C-(H)3(C) + C-(H)2(C)2
ethanol	C-(H)3(C) + C-(H)2(O)(C) + O-(H)(C)
acetic acid	C-(H)3(C) + CO-(O)(C) + O-(H)(CO)

- 4 equations (model compounds), 6 unknowns (groups)
- The system above is underdetermined.

small groups definitions

group	class	group	class
C-(H)3C	methyl	CO-(HorO)(C)	acid, ester, **aldehyde**
C-(H)2(C)2	methylene	C-(H)3(CO)	ketone
CH3corr(tert)	tertiary	CO-(C)2	ketone
O-(H)(C)	alcohol	C-(HorO)(H)(CO)(C)	ketone, acid, **alcohol**
C-(H)2(O)(C)	alcohol, ester	O-(H)(CO)	acid
C-(H)(O)(C)2	alcohol, peroxide	C-(H)(CO)(C)2	acid

- Additivity scheme is adapted from Benson and Buss [1958] and Domalski and Hearing [1993]
  - CO-(HorO)(C) is merged from CO-(O)(C) [acids, esters] and CO-(H)(C) [aldehydes] (e.g., acetaldehyde).
  - C-(HorO)(H)(CO)(C) is merged from C-(H)2(CO)(C) [ketones, acids] and C-(H)(O)(CO)(C) [alcohols] (e.g., citric acid).
- These adaptations are necessary because for this example we are using a limited set of model compounds.
- Asterisks on next page indicate species added to database in CHNOSZ for this example, needed to make a non-singular matrix of group contributions.

small groups matrix (partially overdetermined)

	C-(H)3C	C-(H)2(C)2	O-(H)(C)	C-(H)2(O)(C)	C-(H)(O)(C)2	CH3corr(tert)	C-(H)3(CO)	CO-(C)2	C-(HorO)(H)(CO)(C)	CO-(HorO)(C)	O-(H)(CO)	C-(H)(CO)(C)2
ethane	2											
propane	2	1										
butane	2	2										
methanol	1		1									
ethanol	1		1	1								
1-propanol	1	1	1	1								
1-butanol	1	2	1	1								
1-pentanol	1	3	1	1								
3-pentanol *	2	2	1		1							
2-propanol *	2		1		1	2						
2-butanol *	2	1	1		1	1						
2-pentanol *	2	2	1		1	1						
acetone							2	1				
butanone	2							1	1			
3-pentanone *	2							1	2			
2-heptanone	2	2						1	2			
acetaldehyde	1									1		
acetic acid	1									1	1	
propanoic acid	1								1	1	1	
butanoic acid	1	1							1	1	1	
2-methylpropanoic acid *	2					2				1	1	1
2,4-dimethyl-3-pentanone *	4					4		1				2
lactic acid	1		1			1			1	1	1	
citric acid			1				1			3	3	2
isocitric acid			1						2	3	3	1

fitting 24 species using 12 small groups

- Read the group contribution matrix.

```
> file.small <- system.file("extdata/thermo/groups_small.csv",
    package = "CHNOSZ")
> A.small <- read.csv(file.small, check.names = FALSE,
    row.names = 1)
> A.small[is.na(A.small)] <- 0
```

- The small groups matrix has an additional row for isocitrate, but its heat capacity is not available in the database.

```
> ispecies <- info(head(rownames(A.small), -1), quiet = TRUE)
> cp.species <- info(ispecies, quiet = TRUE)$Cp
> cp.species <- c(cp.species, NA)
```

- Calculate least-squares solution.

```
> i.small <- 1:24
> cp.small <- qr.solve(A.small[i.small, ], cp.species[i.small])
> cp.small
```

```
        C-(H)3C       C-(H)2(C)2           O-(H)(C)
      42.752115        20.740304         3.299958
  C-(H)2(O)(C)     C-(H)(O)(C)2    CH3corr(tert)
      16.862471         8.544542        -9.524111
    C-(H)3(CO)          CO-(C)2 C-(HorO)(H)(CO)(C)
      47.185790       -34.186490        24.205239
   CO-(HorO)(C)       O-(H)(CO)    C-(H)(CO)(C)2
      -7.852115        10.537020         4.979678
```

performance of "small groups" model

- Calculate predicted values and RMSD.

```
> pred.small <- as.matrix(A.small) %*% cp.small
> rmsd <- rmsd(pred.small[i.small], cp.species[i.small])
> rmsd
```

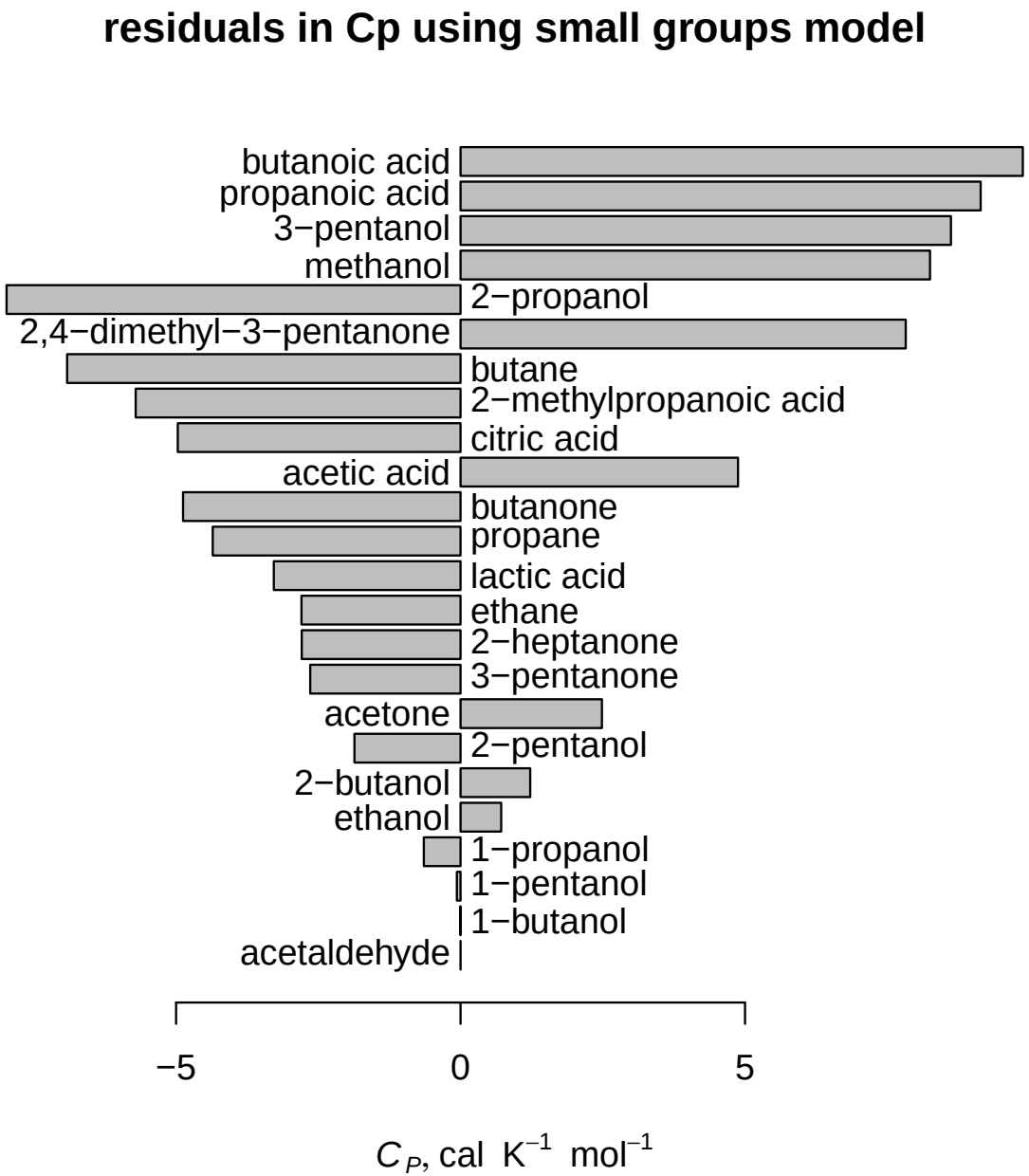
```
[1] 5.266396
```

- It's smaller than what we got using the big groups!
- Plot the residuals.

```
> residuals <- pred.small - cp.species
> names(residuals) <- rownames(A.small)
> residuals <- residuals[i.small]
> residualsplot(residuals, "Cp", "small groups")
```

- Residuals of 0 are where the system is not overdetermined (acetaldehyde in this model has its "own" group).

small groups ...



small group group additivity

- Additivity is useful for some tasks
- Big groups have their place
- small groups aren't perfect either
  - More detail about molecular structures – bond information
  - Improved overall fit compared to large groups
  - Does it help us get closer to lactic and citric acids, or others?

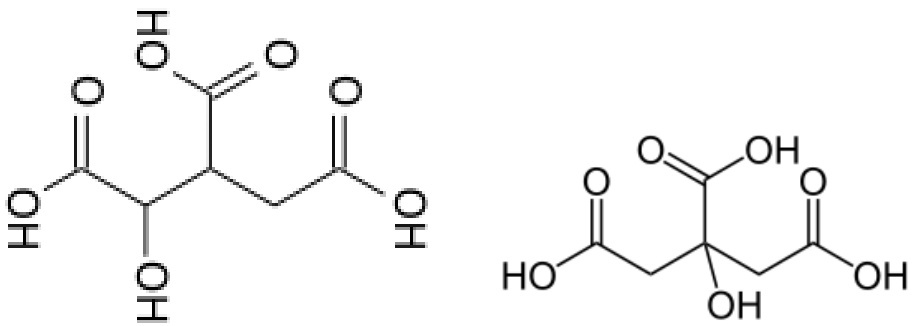
Lactic and Citric and Isocitric

- Heat capacities of the aqueous species (cal K^{-1} mol^{-1})

	database	big groups	small groups
lactic acid	66.7	55.09	63.42
citric acid	73.47	75.37	68.5
isocitric acid	NA	NA	64.74

- Big groups actually get pretty close to citric acid, but they give no hint about the properties of isomerization.
- We are modeling isomerization as disproportionation. In the pictures, the second carbon gains a hydrogen and the third carbon loses one.

Isocitric Acid $\rightarrow$ Citric Acid



C-(H)2(CO)(C) $\rightarrow$ C-(H)3(CO)
C-(H)2(CO)(C) $\rightarrow$ C-(H)(CO)(C)2

small group group additivity

- Additivity is useful for some tasks
- Big groups have their place
- small groups aren't perfect either
- Extend to other properties ... find more model compounds ...

references

J. P. Amend and H. C. Helgeson. Group additivity equations of state for calculating the standard molal thermodynamic properties of aqueous organic species at elevated temperatures and pressures. *Geochim. Cosmochim. Acta*, 61:11 – 46, 1997. doi: 10.1016/S0016-7037(96)00306-7.

S. W. Benson and J. H. Buss. Additivity rules for the estimation of molecular properties. Thermodynamic properties. *J. Chem. Phys.*, 29: 546 – 572, 1958. doi: 10.1063/1.1744539.

E. S. Domalski and E. D. Hearing. Estimation of the thermodynamic properties of C-H-N-O-S-Halogen compounds at 298.15 K. *J. Phys. Chem. Ref. Data*, 22:805 – 1159, 1993. doi: 10.1063/1.555927.