

# *Statistical Theory and Learning from Molecular Simulations*

Lawrence R. Pratt<sup>1</sup> and Susan B. Rempe<sup>2</sup>

<sup>1</sup>Department of Chemical & Biomolecular Engineering



<sup>2</sup>Center for Biological and Material Sciences, Sandia National Laboratories,  
Albuquerque

CECAM workshop "Free energy calculations: From theory to applications"

June 5, 2012

# Molecular Stat-Mech from La Nouvelle-Orléans



## Abstract

- 1 A role for molecular theory in the age of simulation.

*“[Do not] put too much confidence in experimental results until they have been confirmed by theory.”*  
(A. S. Eddington)

- 2 General molecular theory for solutions:  
quasi-chemical theory.

# Perspective on coarse-graining

- Coarse-graining  $\equiv$  stat-mech
  - 1 Example: McMillan-Mayer theory for electrolyte solutions.\*
  - 2 Do not expect to get absolute free energies fully. The goal is to get well-defined relative free energies accurately.
  - 3 Crucial tests are  $T$ -dependence of parameters of coarse-grained models.
- "Fine-graining" is an equally important challenge.
  - 1 Is SPC/E the coarse-grained model of water, from which we might obtain a legitimate fine-grained model?

---

\*W. G. McMillan Jr, and J. E. Mayer, *J. Chem. Phys.* **13**, 276–305 (1945): "The statistical thermodynamics of multicomponent systems"

# Molecular Stat-Mech from La Nouvelle-Orléans



## Abstract

- 1 A role for molecular theory in the age of simulation.

*“[Do not] put too much confidence in experimental results until they have been confirmed by theory.”  
(A. S. Eddington)*

- 2 General molecular theory for solutions:  
quasi-chemical theory.

# Learning from simulation

Consider a simple example: molecular theory of primitive hydrophobic effects.\*

For example, it is widely, but not universally, agreed that the hydrogen bonding interactions between water molecules are a key to understanding hydrophobic effects. Conventional molecular simulation calculations with widely accepted molecular interaction models for small hydrophobic species in water broadly agree with experimental results on such systems (26–134). In this sense, everything is

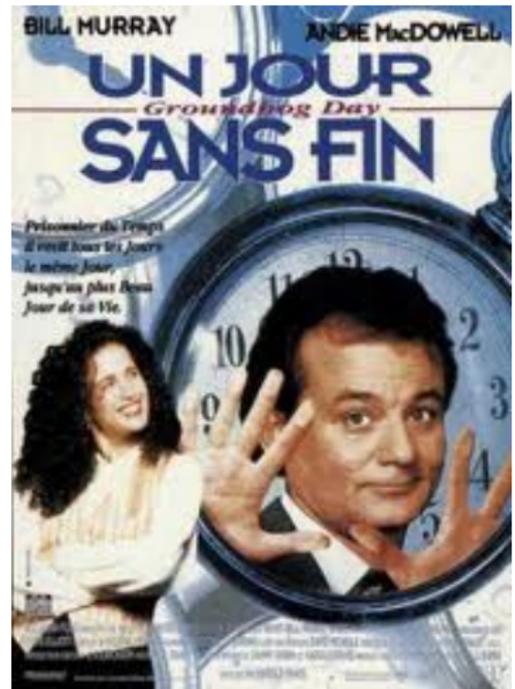
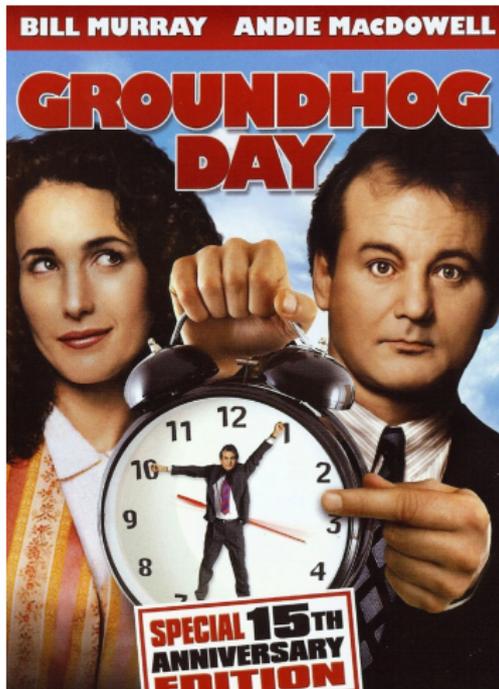
Notice the piling-up of  $134-25=109$  simulations cited here.

- Evidently, something is missing.

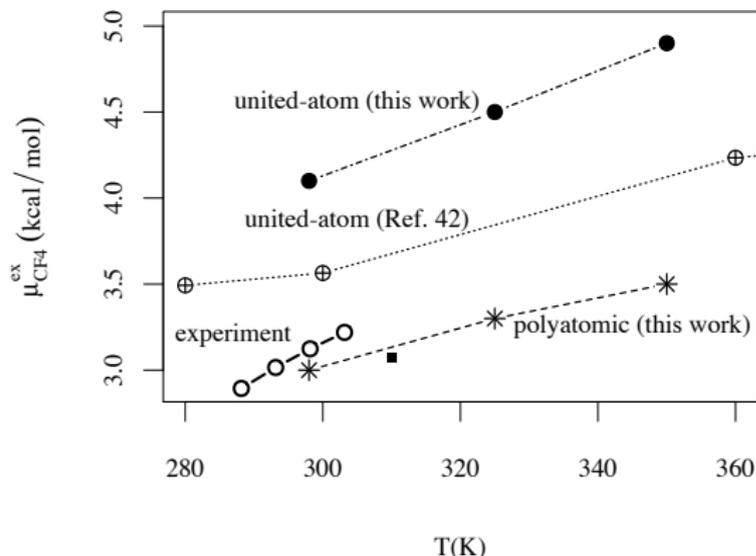
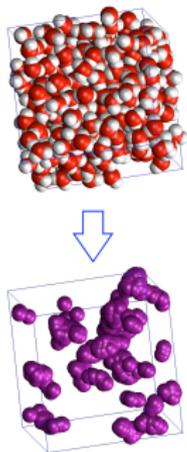
---

\*L. R. Pratt, *Ann. Rev. Phys. Chem.* **53**, 409-436(2002):  
“Molecular theory of hydrophobic effects: ‘She is too mean to have her name repeated.’ ”

# Learning from repetition



# Enabling idea for primitive hydrophobic effects: study cavities/occupancies

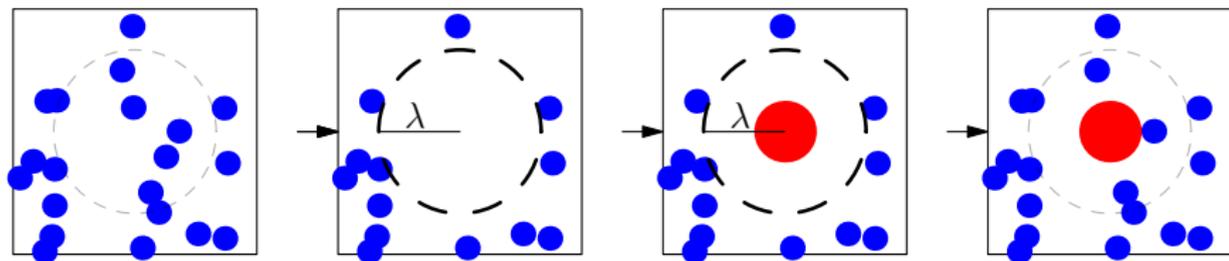


See Asthagiri, *et al.*\*

\*D. Asthagiri, H. S. Ashbaugh, A. Piryatinski, M. E. Paulaitis, and L. R. Pratt, *J. Am. Chem. Soc.* **129**, 10133 (2007): "Non-van der Waals Treatment of the Hydrophobic Solubilities of  $CF_4$ "

# Molecular quasi-chemical theory: the free energy

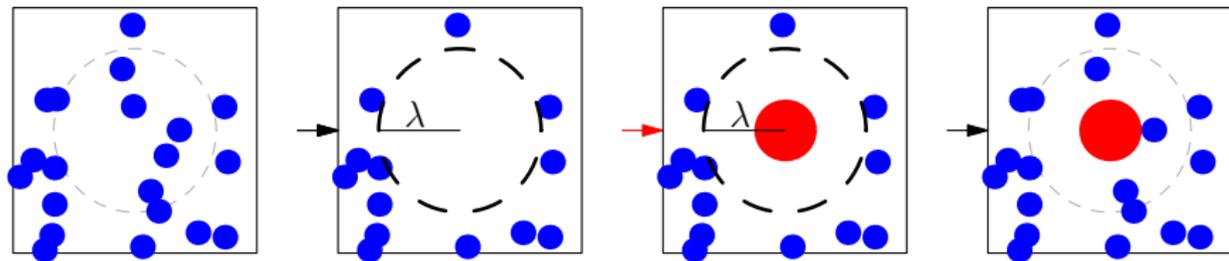
$$\beta\mu^{(\text{ex})} = -\ln p^{(0)}(n_\lambda = 0) + \ln \langle e^{\beta\varepsilon} | n_\lambda = 0 \rangle + \ln p(n_\lambda = 0)$$



- these contributions should be evaluated on the basis of data from simulation of the physical system.

# Outer-shell contribution

$$\beta\mu^{(\text{ex})} = -\ln p^{(0)}(n_\lambda = 0) + \ln \langle e^{\beta\varepsilon} | n_\lambda = 0 \rangle + \ln p(n_\lambda = 0)$$



$$\approx \beta \langle \varepsilon | n_\lambda = 0 \rangle + \beta^2 \langle \delta\varepsilon^2 | n_\lambda = 0 \rangle / 2$$

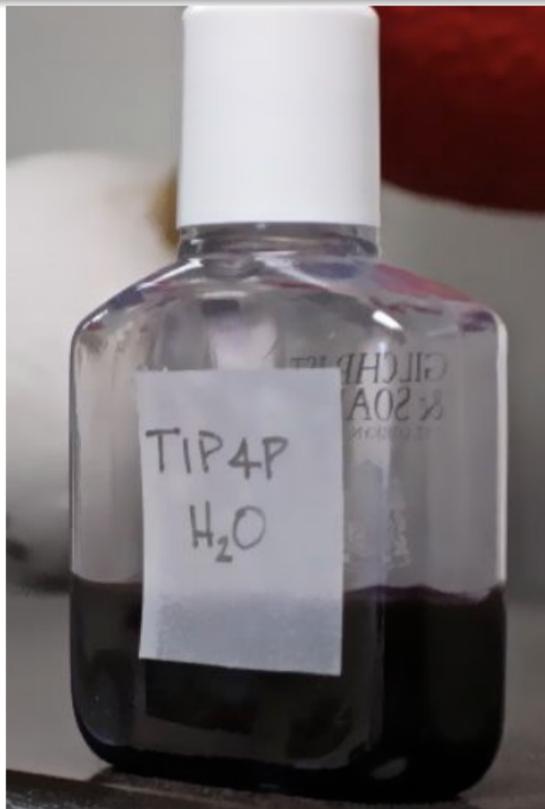
a **Gaussian (normal distribution) theory**.

- ok for  $\lambda \sim \infty$  because numerous and weakly correlated contributions add-up to make the outer-shell term.

# Meta-theory comments

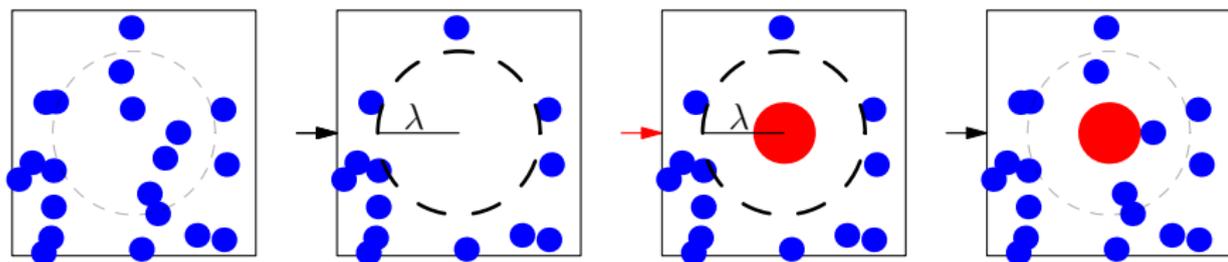
- 1 Implementation with simulation data.
- 2 Realizable.
- 3 For statistical problems 'good theories are either Gaussian or everything.'
- 4 This is general purpose theory, includes:
  - a van der Waals theory,
  - b Born 'models',
  - c Debye-Hückel theory,
  - d Network-liquid theory.

Example: liquid water. We do not have a theory of liquid water from molecular stat-mech.



# Quasi-chemical theory for liquid water

$$\beta\mu^{(\text{ex})} = -\ln p^{(0)}(n_\lambda = 0) + \ln \langle e^{\beta\varepsilon} | n_\lambda = 0 \rangle + \ln p(n_\lambda = 0)$$

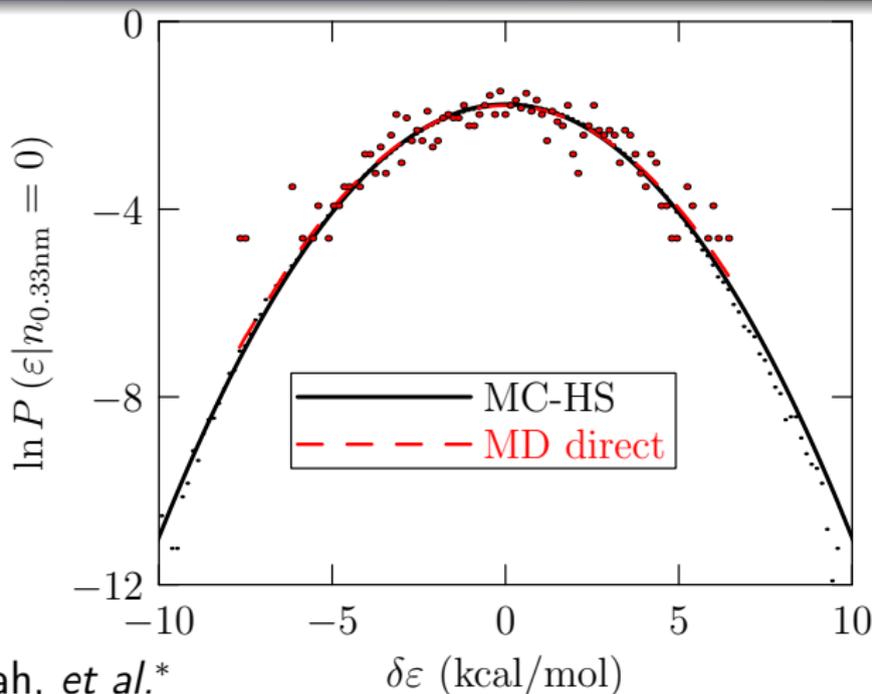


$$\approx \beta \langle \varepsilon | n_\lambda = 0 \rangle + \beta^2 \langle \delta\varepsilon^2 | n_\lambda = 0 \rangle / 2$$

a Gaussian (normal distribution) theory.

- only question is the accuracy of the assumed Gaussian distribution.

liquid water:  $T = 300\text{K}$ ,  $\rho = 1 \text{ atm}$  (TIP3P)



\*J. K. Shah, D. Asthagiri, L. R. Pratt, M. E. Paulaitis, *J. Chem. Phys.* **127**, 144508 (2007): "Balancing Local Order and Long-Ranged Interactions in the Molecular Theory of Liquid Water"

# What did you expect?

'Water Revisited'\*

(ST2)

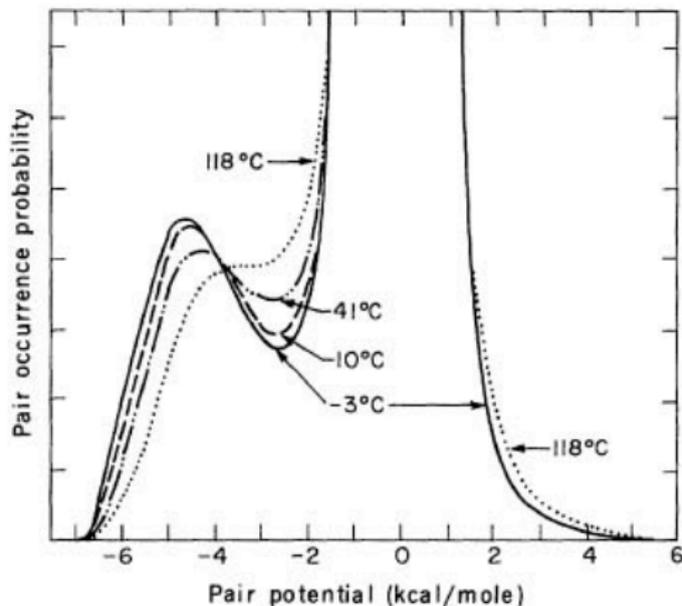


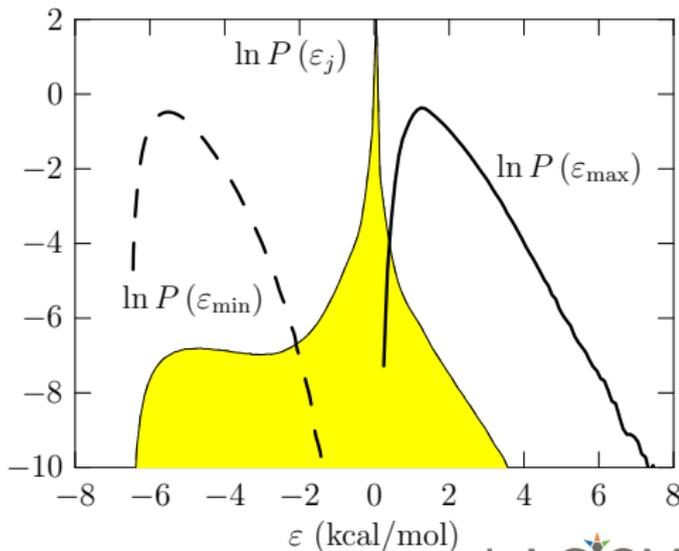
Fig. 4. Distribution function for effective pair interactions between molecules in liquid water (at 1 gram per cubic centimeter).

\*F. H. Stillinger *Science* **209**, 451-457 (1980)

# Explain!

- The 'Water Revisited' distribution is  $P(\epsilon_j)$ . As a distribution, it does not have a useful  $N \rightarrow \infty$  limit.
- $P(\epsilon)$  is much simpler, does not have an obvious signature of an H-bond energy, is broad, but a Gaussian model is valid.

$$\epsilon = \sum_{j=1}^N \epsilon_j$$



A dielectric model is correct when  $\lambda \approx 0.31$  nm.

$$\beta\mu^{(\text{ex})} = -\ln p^{(0)}(0) + \ln \langle e^{\beta\varepsilon} | 0 \rangle + \ln p(0) .$$

- For that value the *network-liquid* contributions balance:

$$-\ln p^{(0)}(0) + \ln p(0) \approx 0 .$$

⇒ For the free energy

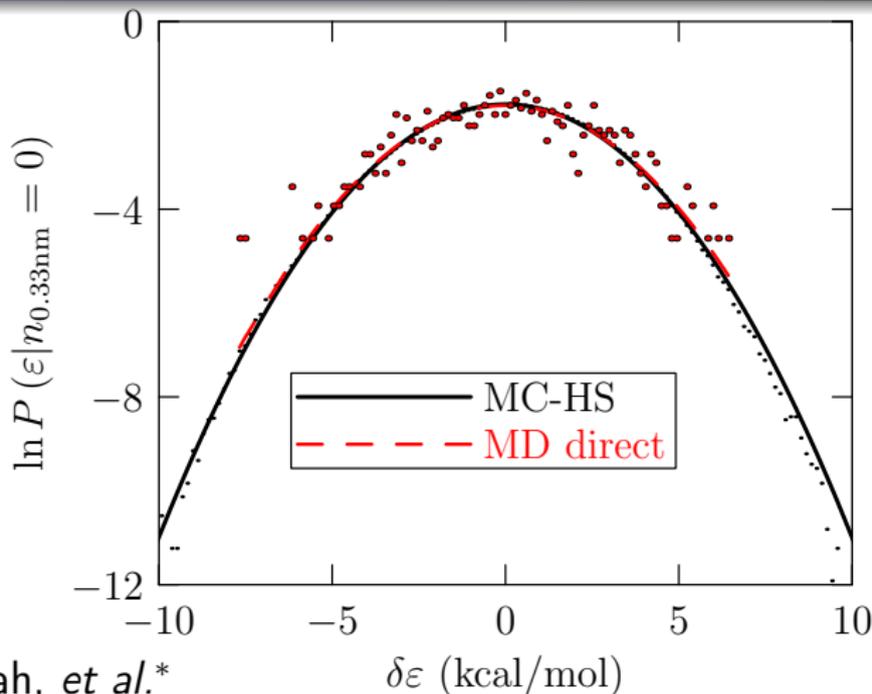
$$\mu_{\text{W}}^{(\text{ex})} \approx \langle \varepsilon | 0 \rangle + \beta \langle \delta\varepsilon^2 | 0 \rangle / 2 .$$

⇒ A network-liquid model that gets inner-shell occupancies  $p_{\text{W}}(n_{\lambda})$  right in comparison with experiment, and that gets hydrophobic things  $[-\ln p^{(0)}(0)]$  right, does **not** get the free energy of liquid water right.

# Explain!

- Interactions at second-shell range are strong. Big, well-recognized contributions come from outer shells.
- The number of strong-interaction partners of a water molecule is large enough that a Gaussian distribution of binding energies distributions can be satisfactory.
- **The biggest difficulty** is that these binding energy distributions are very broad.

liquid water:  $T = 300\text{K}$ ,  $\rho = 1 \text{ atm}$  (TIP3P)



See Shah, *et al.*\*

\*J. K. Shah, D. Asthagiri, L. R. Pratt, M. E. Paulaitis, *J. Chem. Phys.* **127**, 144508 (2007): "Balancing Local Order and Long-Ranged Interactions in the Molecular Theory of Liquid Water"

# Quasi-chemical theory: wait there's more!

Quasi-chemical theory has been applied to :

- Statistical thermodynamics of the hard-sphere fluid, in a self-consistent field implementation.\* It is as accurate as the most accurate previous physical theory (the scaled-particle theory).
- Hydrophobic hydration of  $\text{CF}_4$ .† Drying is not an essential consideration for the hydrophobic solubilities of  $\text{CF}_4$ , or of  $\text{C}(\text{CH}_3)_4$ .

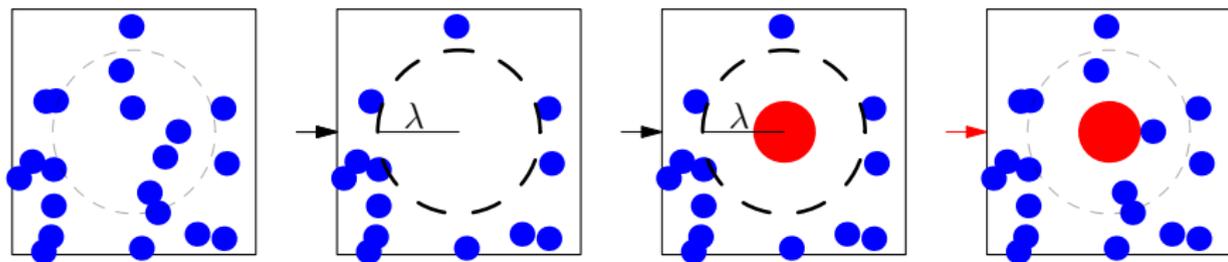
---

\*L. R. Pratt and H. S. Ashbaugh, *Phys. Rev. E* **68**, 021505 (2003):  
"Self-consistent molecular field theory for packing in classical liquids"

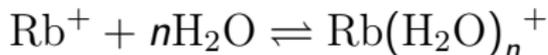
†D. Asthagiri, H. S. Ashbaugh, A. Piryatinski, M. E. Paulaitis, and L. R. Pratt, *J. Am. Chem. Soc.* **129**, 10133 (2007): "Non-van der Waals Treatment of the Hydrophobic Solubilities of  $\text{CF}_4$ "

# The chemical contribution and ionic hydration

$$\beta\mu^{(\text{ex})} = -\ln p^{(0)}(n_\lambda = 0) + \ln \langle e^{\beta\varepsilon} | n_\lambda = 0 \rangle + \ln p(n_\lambda = 0)$$



Consider the chemical equilibria:



$$\boxed{(p_{\text{Rb}^+}(0)\rho^n) \times K_n = p_{\text{Rb}^+}(n)}$$

No approximation.

# Isolating an electronic structure contribution

$$\textcircled{1} \quad -\ln K_n = -\ln K_n^{(0)} + \beta\mu_{\text{Rb}(\text{H}_2\text{O})_n^+}^{(\text{ex})} - n\beta\mu_{\text{H}_2\text{O}}^{(\text{ex})} - \beta\mu_{\text{Rb}(\text{H}_2\text{O})_0^+}^{(\text{ex})}$$

$\ln K_n^{(0)}$  is a few-body problem, and a well-defined target for modern electronic structure calculations.

$$\textcircled{2} \quad \beta\mu_{\text{Rb}(\text{H}_2\text{O})_0^+}^{(\text{ex})} = -\ln p_X^{(0)}(n=0) + \ln \langle e^{\beta\varepsilon} | n=0 \rangle ,$$

exactly cancels in the original formula for the free energy.

$$\beta\mu_{\text{Rb}^+}^{(\text{ex})} = -\ln K_n^{(0)} \rho^n + \ln p_{\text{Rb}^+}(n) + \beta\mu_{\text{Rb}(\text{H}_2\text{O})_n^+}^{(\text{ex})} - n\beta\mu_{\text{H}_2\text{O}}^{(\text{ex})}$$

No approximation. True for any  $\lambda$ , and for any  $n$ .

# Primitive quasi-chemical approximation

- 1 For given  $\lambda$ , choose  $n = \bar{n}$ , the most probable value. Then  $p_{\text{Rb}^+}(\bar{n})$  is as close to one (1) as possible and  $\ln p_{\text{Rb}^+}(\bar{n})$  is as close to zero (0) as possible:\*

$$\beta\mu_{\text{Rb}^+}^{(\text{ex})} \approx -\ln K_{\bar{n}}^{(0)} \rho^{\bar{n}} + \ln p_{\text{Rb}^+}(\bar{n}) + \beta\mu_{\text{Rb}(\text{H}_2\text{O})_{\bar{n}}}^{(\text{ex})} - \bar{n}\beta\mu_{\text{H}_2\text{O}}^{(\text{ex})}$$

*(A red arrow points from the term  $\ln p_{\text{Rb}^+}(\bar{n})$  to a red  $\approx 0$  above it.)*

- 2 Exploit available models for the various pieces, remembering that the stat-thermo contributions are designed to balance.

---

\*S. B. Rempe, L. R. Pratt, G. Hummer, J. D. Kress, R. L. Martin, and A. Redondo, *J. Am. Chem. Soc.* **122**, 966 (2000): "The hydration number of  $\text{Li}^+$  in liquid water"

# Meta-theory comments

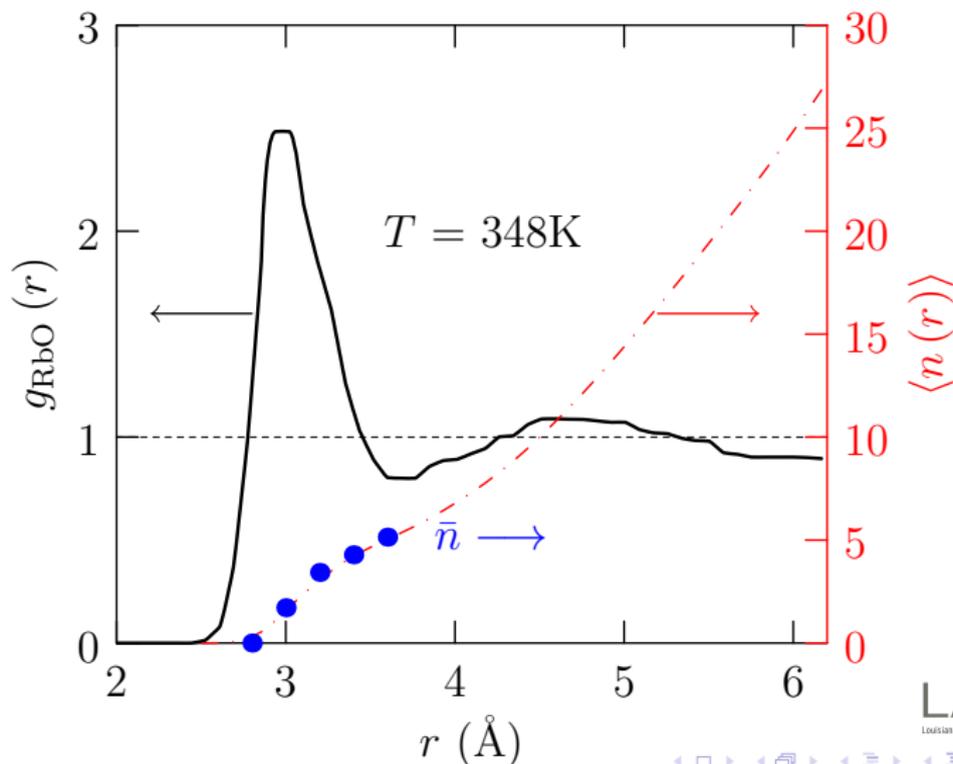
## I claim:

- 1 This is the only available stat-thermo theory that establishes a direct, practical role for *in situ* electronic structure computation of arbitrary accuracy. [But check out R. H. Wood, *et al.*<sup>†</sup>]
- 2 This is the only available stat-thermo theory that directly addresses obvious pathologies of dielectric continuum models, specifically arbitrariness and sensitivity to dielectric boundaries.
- 3 This approach treats variation of composition variables physically correctly, *i.e.*, it is not the ‘micro-solvation’ method of electronic structure specialists.

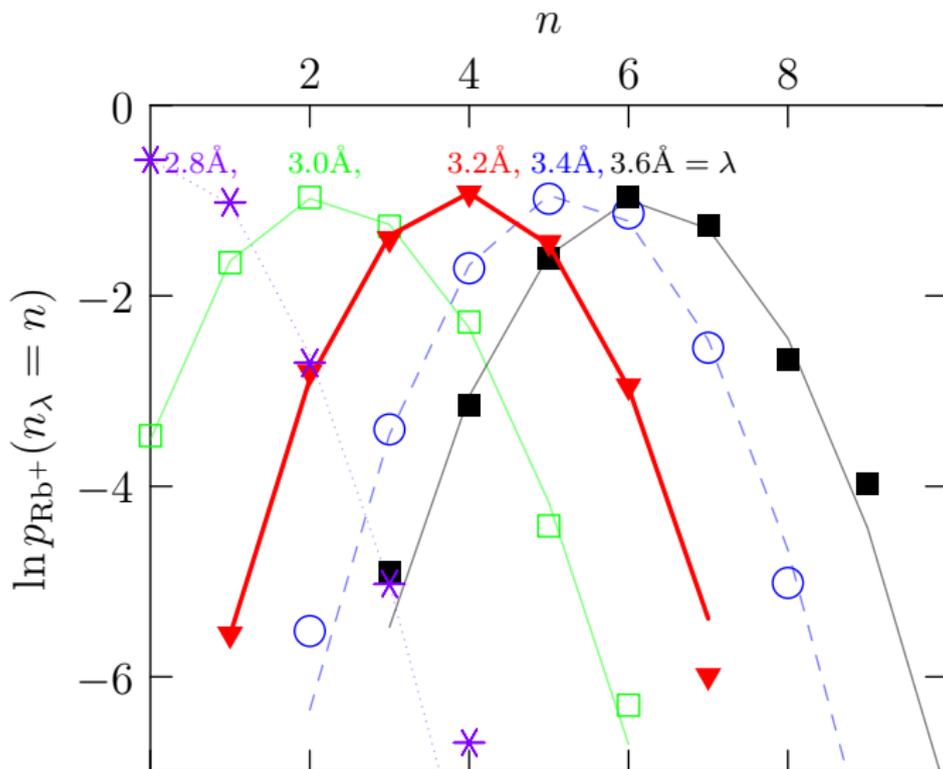
---

<sup>†</sup>R. H. Wood, W. Liu, and D. J. Doren, *J. Phys. Chem. A*, **106**, 6689 (2002): “Rapid Calculation of the Structures of Solutions with *ab Initio* Interaction Potentials”

Example:  $\text{Rb}^+(\text{aq})$   
(AIMD/VASP/PW91/ $N_{\text{D}_2\text{O}}=64$ )

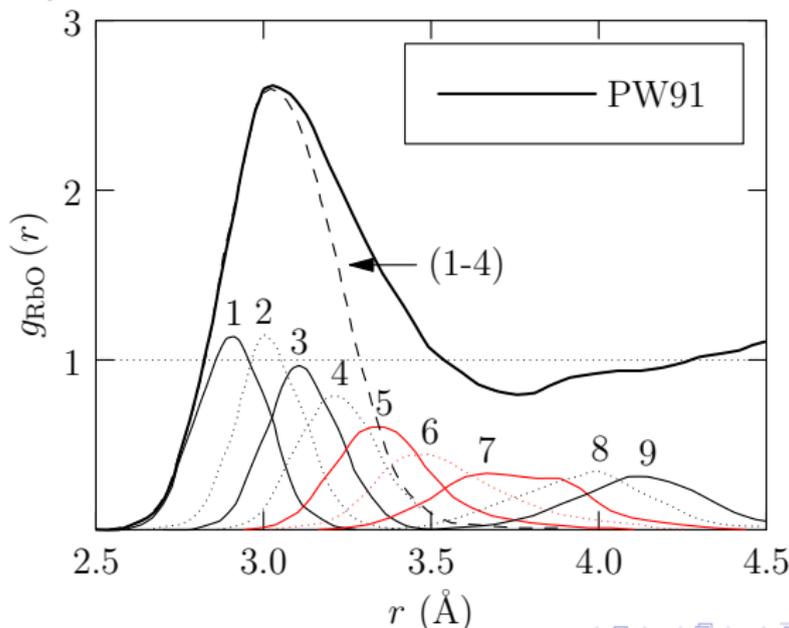


# Intrinsic error of $\ln p_{\text{Rb}^+}(\bar{n}) \approx 0$ : about $kT$



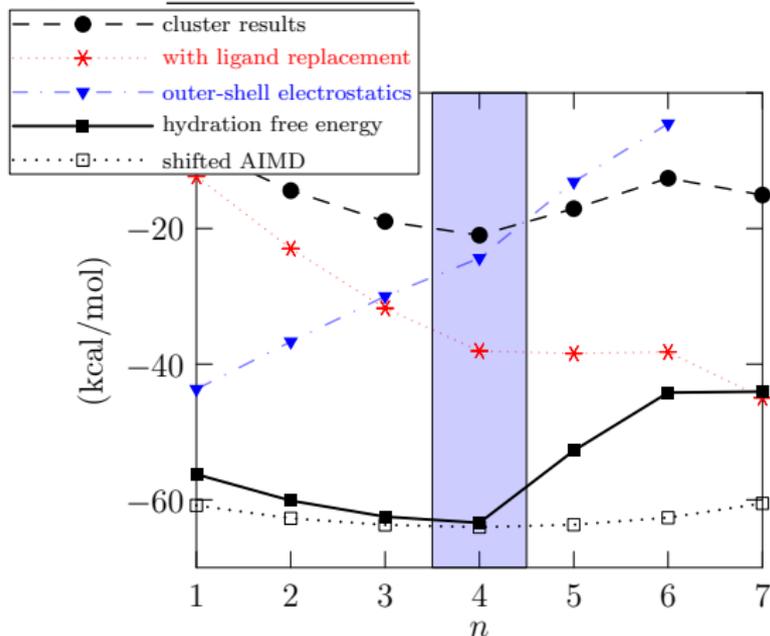
# What about the choice of $\lambda$ ?

- $\lambda \rightarrow$  large should give a more accurate theory? Wrong idea here!
- Choose  $\lambda$  to make the other contributions simple!
- Avoid split inner shells!



# Electronic structure implementation of the primitive theory

- The harmonic approximation adopted for  $\ln K_n^{(0)}$  is OK, if you obey the 'no split shells' rule.



# Primitive theory applied to $M^{z+}(\text{aq})$ works great.

- With  $\text{Li}^+(\text{aq})$ , this is presently the best attempt to calculate the absolute hydration free energy of a single ion.\*
- Has not been tested stringently for other ions/solutions.
- This is the only available stat-thermo theory that establishes a direct, practical role for in situ electronic structure computation of arbitrary accuracy. But check out R. H. Wood, *et al.*†

---

\*B. Roux and H. Yu, *J. Chem. Phys.* **132**, 234101(2011): "Assessing the accuracy of approximate treatments of ion hydration based on primitive quasichemical theory "

†R. H. Wood, W. Liu, and D. J. Doren, *J. Phys. Chem. A*, **106**, 6689 (2002): "Rapid Calculation of the Structures of Solutions with *ab Initio* Interaction Potentials"

# Some conclusions

- Quasi-chemical theory:\* general structure for statistical thermodynamical theories.
- Applications of quasi-chemical theory so far:
  - a van der Waals theory,
  - b Born 'models' → ion hydration
  - c Debye-Hückel theory,
  - d Network-liquid theory,
- Distributions of binding energies ( $\epsilon$ 's from simulations), and occupancies ( $n$ 's for inner-shells) are helpful in developing physical theories of solutions.

---

\*D. Asthagiri, P. D. Dixit, S. Merchant, M. E. Paulaitis, L. R. Pratt, S. B. Rempe, and S. Varma, *Chem. Phys. Letts.*, **485**, 1-7 (2010): "Ion selectivity from local configurations of ligands in solutions and ion channels"