All electron *ab initio* molecular simulations
Status, successes, and some computational challenges

Volker Blum

Fritz-Haber-Institut der Max-Planck-Gesellschaft
Berlin

100 YEARS OF PHYSICAL CHEMISTRY IN BERLIN—DAHLEM
100 YEARS

The ab initio molecular simulations package
FHI aims

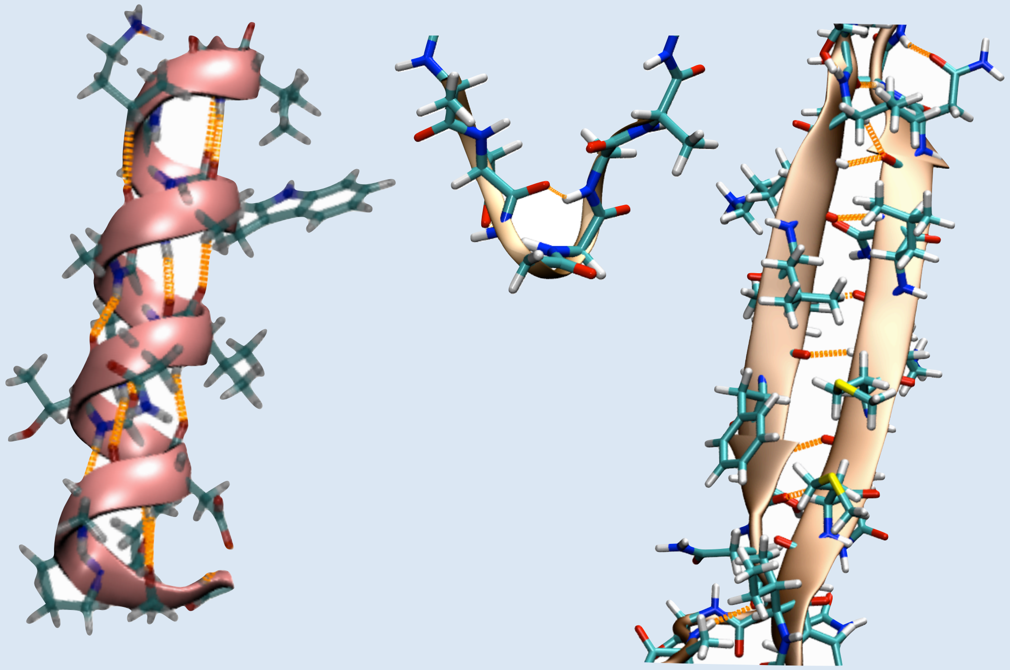
IPAM Workshop "Challenges in Warm Dense Matter" - IPAM, May 24, 2012

Electronic structure theory for real materials

Enormous successes:

Electronic structure theory for real materials

Enormous successes:

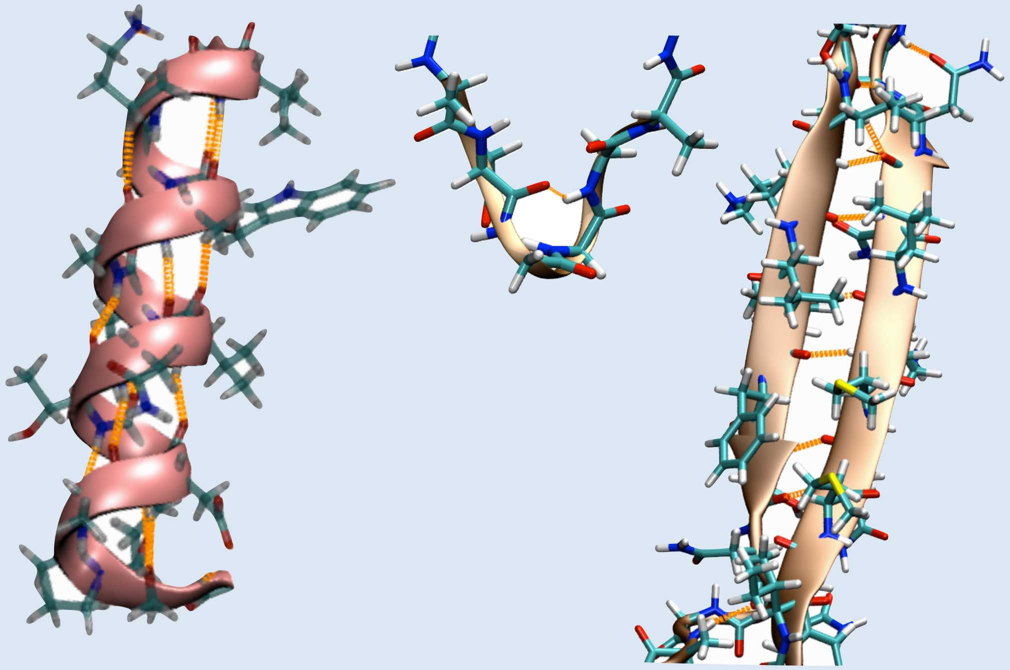


(Bio)molecular matter

- Structural complexity
- statistical averages & dynamics
- “weak” interactions critical

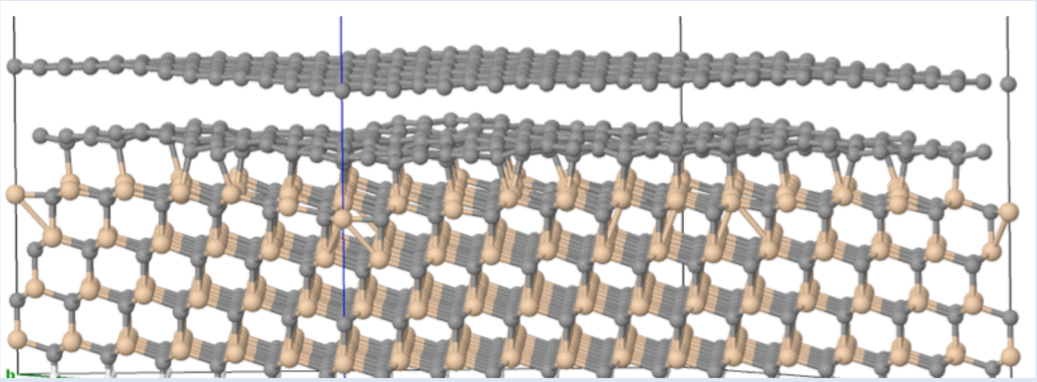
Electronic structure theory for real materials

Enormous successes:

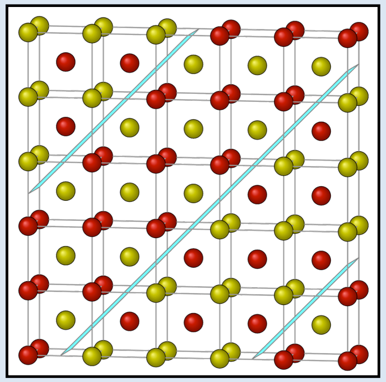


(Bio)molecular matter

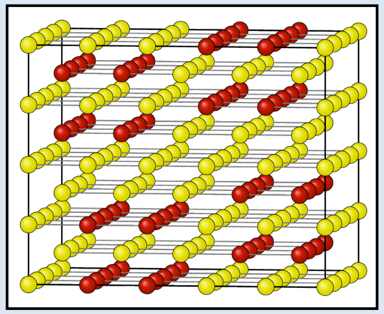
- Structural complexity
- statistical averages & dynamics
- “weak” interactions critical



Graphene / SiC



Ta_3W_3



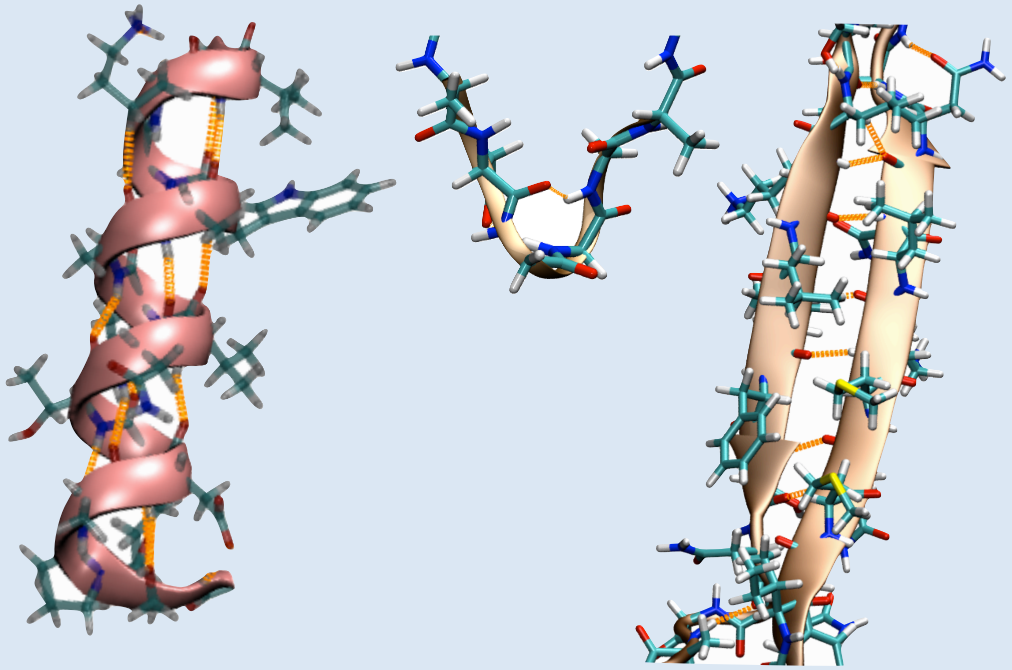
Ta_4W_9

Condensed phases

(solids, surfaces, ...)

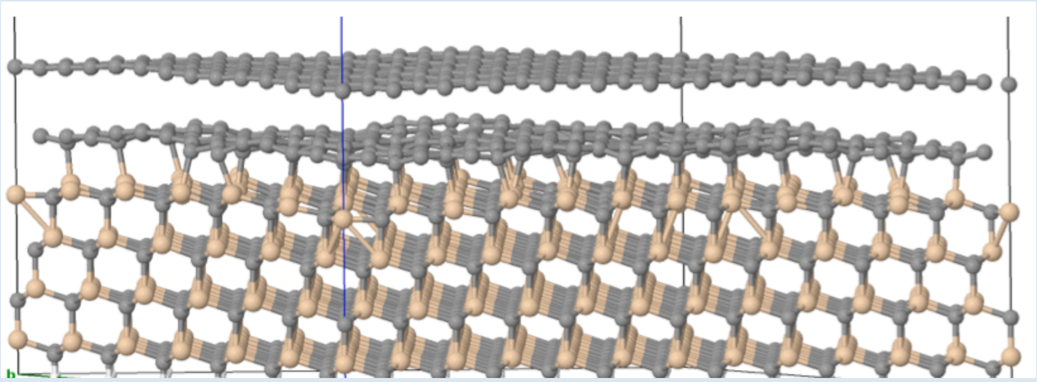
- Structure!
- Stability, free energies
- electronic, mechanical, optical, ... properties

Electronic structure theory for real materials

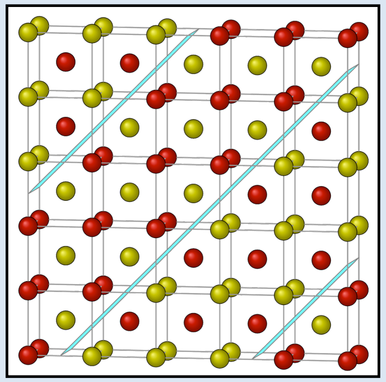


(Bio)molecular matter

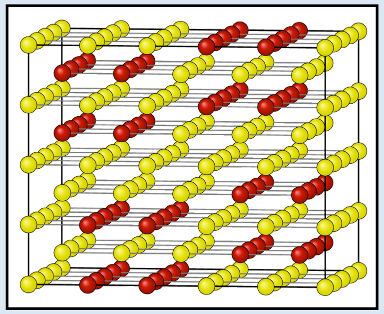
- Structural complexity
- statistical averages & dynamics
- “weak” interactions critical



Graphene / SiC



Ta_3W_3



Ta_4W_9

Condensed phases

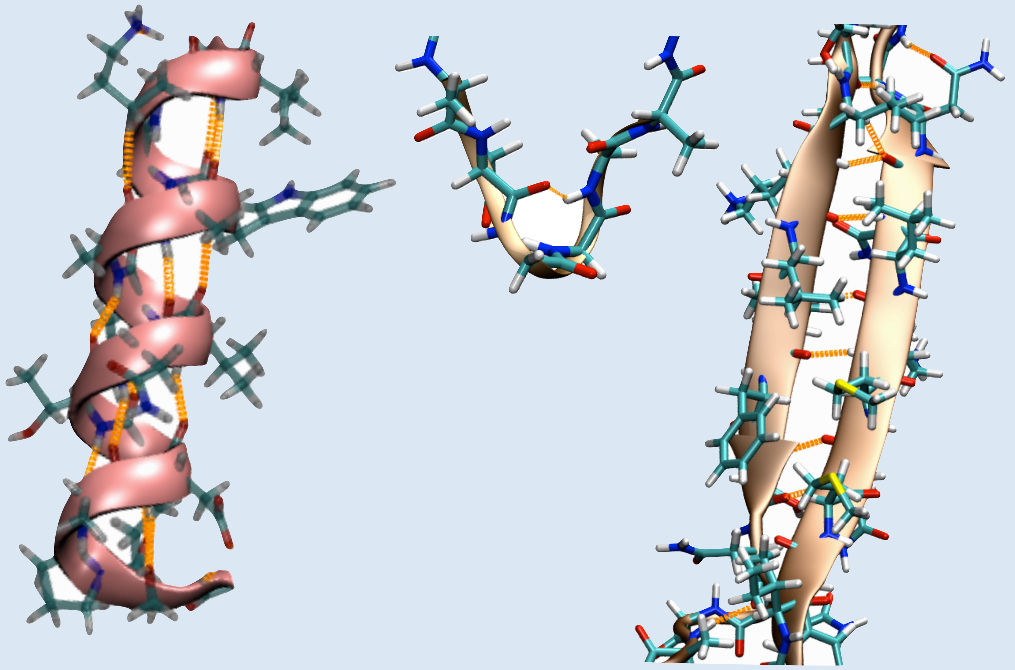
(solids, surfaces, ...)

- Structure!
- Stability, free energies
- electronic, mechanical, optical, ... properties

Matter at extreme conditions

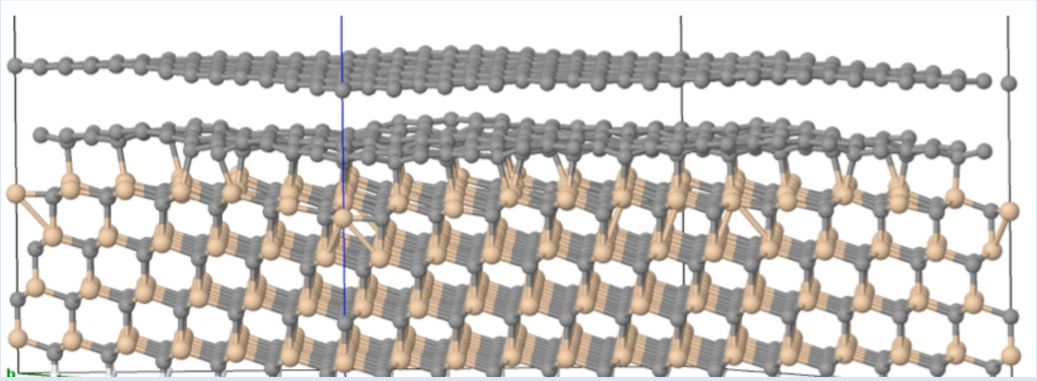
- “electron gas + protons”; high-pressure compounds, transitions
- (Born-Oppenheimer) molecular dynamics, classical nuclei
- Quantum nuclei? (PIMD)

Electronic structure theory for real materials

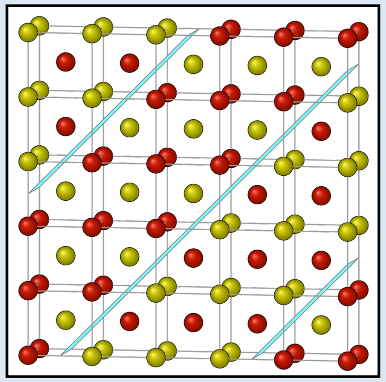


(Bio)molecular matter

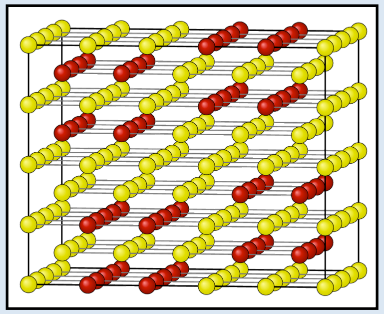
- Structural complexity
- statistical averages & dynamics
- “weak” interactions critical



Graphene / SiC



Ta_3W_3



Ta_4W_9

Condensed phases

(solids, surfaces, ...)

- Structure!
- Stability, free energies
- electronic, mechanical, optical, ... properties

Today: “Mostly density-functional theory”, plenty of flavors

Matter at extreme conditions

- “electron gas + protons”; high-pressure compounds, transitions
- (Born-Oppenheimer) molecular dynamics, classical nuclei
- Quantum nuclei? (PIMD)

Electronic structure theory for real materials

... but we still face some challenges:

Electronic structure theory for real materials

... but we still face some challenges:

- Methods. Are we computing the right thing?
  - Current DFT (LDA/GGA *and beyond*) may qualitatively fail with *or without* warning for much of the interesting space, even for "structure"
  - Other numerical approximations? (grids, cutoffs, pseudoization, ...)
  - "Classical" vs. "quantum" nuclei? Born-Oppenheimer?

Electronic structure theory for real materials

... but we still face some challenges:

- Methods. Are we computing the right thing?
  - Current DFT (LDA/GGA *and beyond*) may qualitatively fail with *or without* warning for much of the interesting space, even for "structure"
  - Other numerical approximations? (grids, cutoffs, pseudoization, ...)
  - "Classical" vs. "quantum" nuclei? Born-Oppenheimer?

- Algorithms. Can we compute the right thing?
  - Realistically sized systems to capture "reality"
  - Statistical averages, dynamics, combinatorial complexity of "structure"?
  - Simply, hardware vs. software - utilize *available* hardware effectively

Outline

An approach to all-electron "density functional theory and beyond": FHI-aims [1]

- Numeric atom-centered (localized) basis sets
- Scalability (1,000s of atoms, 1(0),000s of CPUs)
- Pushing towards "better" functionals (→P. Rinke)

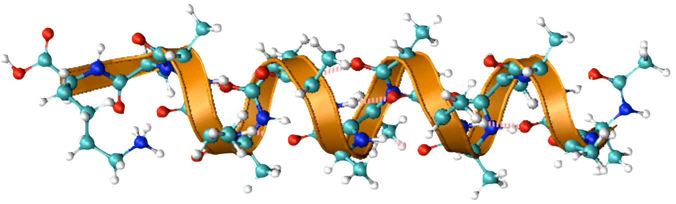
---

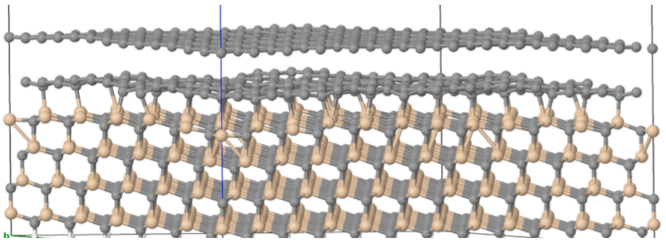
[1] The Fritz Haber Institute ab initio molecular simulations suite (FHI-aims)
V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
Computer Physics Communications ***180****, 2175-2196 (2009) - http://www.fhi-berlin.mpg.de/aims/*

Outline

An approach to all-electron "density functional theory and beyond": FHI-aims [1]

- Numeric atom-centered (localized) basis sets
- Scalability (1,000s of atoms, 1(0),000s of CPUs)
- Pushing towards "better" functionals (→ P. Rinke)





Where do "we" come from?

- (Bio)molecular structure and spectroscopy
- Nanostructured inorganic surfaces

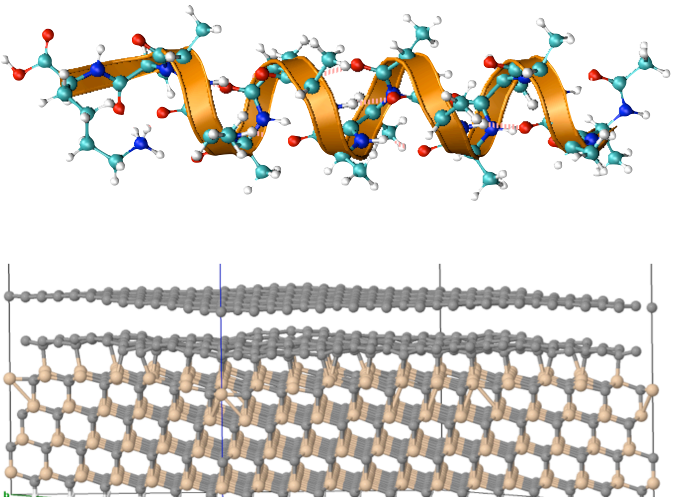
---

[1] *The Fritz Haber Institute ab initio molecular simulations suite (FHI-aims)*
V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
Computer Physics Communications ***180****, 2175-2196 (2009) - http://www.fhi-berlin.mpg.de/aims/*

Outline

An approach to all-electron "density functional theory and beyond": FHI-aims [1]

- Numeric atom-centered (localized) basis sets
- Scalability (1,000s of atoms, 1(0),000s of CPUs)
- Pushing towards "better" functionals (→P. Rinke)



Where do "we" come from?

- (Bio)molecular structure and spectroscopy
- Nanostructured inorganic surfaces

... and some challenges (towards WDM)

[1] *The Fritz Haber Institute ab initio molecular simulations suite (FHI-aims)*
V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
Computer Physics Communications ***180****, 2175-2196 (2009)* - *http://www.fhi-berlin.mpg.de/aims/*

People behind these projects

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

People behind these projects

Matthias
Scheffler

People behind these projects

Scalability

Matthias
Scheffler

Ville Havu
(FHI/Helsinki)

Rainer Johanni
(Munich)

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

People behind these projects

Matthias
Scheffler

Ville Havu
(FHI/Helsinki)

Scalability

Rainer Johanni
(Munich)

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

People behind these projects

Matthias
Scheffler

Scalability

Ville Havu
(FHI/Helsinki)

Rainer Johanni
(Munich)

"Beyond LDA / GGA"

Xinguo Ren
(FHI)

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

People behind these projects

Matthias
Scheffler

Scalability

Ville Havu
(FHI/Helsinki)

Rainer Johanni
(Munich)

“Beyond LDA / GGA”

Xinguo Ren
(FHI)

Biomolecular simulations

Mariana
Rossi

Franziska
Schubert

Alex
Tkatchenko

Carsten
Baldauf

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

People behind these projects

Matthias
Scheffler

Scalability

Ville Havu
(FHI/Helsinki)

Rainer Johanni
(Munich)

"Beyond LDA / GGA"

Xinguo Ren
(FHI)

Biomolecular simulations

Mariana
Rossi

Franziska
Schubert

Alex
Tkatchenko

Carsten
Baldauf

... FHI-aims - support from many more:

Karsten Reuter, Patrick Rinke, Ralf Gehrke, Paula Havu, Andreas Dolfen, Felix Hanke, Stefan Gutzeit, Andrea Sanfilippo, Luca Ghiringhelli, Sergey Levchenko, Matthias Gramzow, Mina Yoon, Christian Carbogno, Norbert Nemec, Jörg Meyer, Fabio Caruso, Sucismita Chutia, Jürgen Wieferink, Simiam Ghan, Viktor Atalla, Matti Ropo, Ferdinand Evers, Alex Bagrets, Heiko Appel, Daniel Berger, Oliver Hofmann, ...

Fritz Haber Institute, Berlin

[Richard-Willstätter-Haus]

The electronic structure “wishlist”

The electronic structure “wishlist”

- Cover (essentially) the entirety of chemistry / materials:

The electronic structure “wishlist”

- Cover (essentially) the entirety of chemistry / materials:
    - first/second row elements

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
    - first/second row elements
    - 3*d* transition metals (magnetism)
    - 4*d*/5*d* elements (relativity)
    - *f*-electron systems
    - ...

The electronic structure “wishlist”

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems
  - ...
- Periodic, cluster systems on equal footing

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems
  - ...
- Periodic, cluster systems on equal footing
- "Properties" (structure, dynamics, spectroscopy, ...)

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems
  - ...
- Periodic, cluster systems on equal footing
- "Properties" (structure, dynamics, spectroscopy, ...)
- Path "beyond" DFT-LDA/GGA (HF, hybrids, RPA, MP2, *GW*, ...)

The electronic structure “wishlist”

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems
  - ...
- Periodic, cluster systems on equal footing
- “Properties” (structure, dynamics, spectroscopy, ...)
- Path “beyond” DFT-LDA/GGA (HF, hybrids, RPA, MP2, *GW*, ...)
- (Massively) parallel scalability

The electronic structure "wishlist"

- Cover (essentially) the entirety of chemistry / materials:
  - first/second row elements
  - 3*d* transition metals (magnetism)
  - 4*d*/5*d* elements (relativity)
  - *f*-electron systems
  - ...
- Periodic, cluster systems on equal footing
- "Properties" (structure, dynamics, spectroscopy, ...)
- Path "beyond" DFT-LDA/GGA (HF, hybrids, RPA, MP2, *GW*, ...)
- (Massively) parallel scalability

Our goal: *Efficient* method, but do not sacrifice accuracy

Numerical approximations (including all electrons) should be reliably convergable for the *actual* problem of interest!

Central decision: the basis set

$$\left[-\frac{\nabla^2}{2} + v_{\text{ext}}(\boldsymbol{r}) + v_{\text{es}}(\boldsymbol{r}) + v_{\text{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r}) = \epsilon_k\psi_k(\boldsymbol{r})$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r}) = \sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2}+v_{\text{ext}}(\boldsymbol{r})+v_{\text{es}}(\boldsymbol{r})+v_{\text{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r})=\epsilon_k\psi_k(\boldsymbol{r})}$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r})=\sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

↓

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k=\epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij}=\langle\varphi_i|\hat{h}_{\text{KS}}|\varphi_j\rangle$$

$$s_{ij}=\langle\varphi_i|\varphi_j\rangle$$

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2}+v_{\mathrm{ext}}(\boldsymbol{r})+v_{\mathrm{es}}(\boldsymbol{r})+v_{\mathrm{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r})=\epsilon_k\psi_k(\boldsymbol{r})}$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r})=\sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

↓

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k=\epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij}=\langle\varphi_i|\hat{h}_{\mathrm{KS}}|\varphi_j\rangle$$

$$s_{ij}=\langle\varphi_i|\varphi_j\rangle$$

<u>Many good options:</u>

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2} + v_{\text{ext}}(\boldsymbol{r}) + v_{\text{es}}(\boldsymbol{r}) + v_{\text{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r}) = \epsilon_k \psi_k(\boldsymbol{r})}$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r}) = \sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

↓

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k = \epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij} = \langle\varphi_i|\hat{h}_{\text{KS}}|\varphi_j\rangle$$

$$s_{ij} = \langle\varphi_i|\varphi_j\rangle$$

<u>Many good options:</u>

- Plane waves

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2}+v_{\text{ext}}(\boldsymbol{r})+v_{\text{es}}(\boldsymbol{r})+v_{\text{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r})=\epsilon_k\psi_k(\boldsymbol{r})}$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r})=\sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

↓

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k=\epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij}=\langle\varphi_i|\hat{h}_{\text{KS}}|\varphi_j\rangle$$
$$s_{ij}=\langle\varphi_i|\varphi_j\rangle$$

<u>Many good options:</u>

- Plane waves
- *Augmented* plane waves (*Slater 1937; Andersen 1975; etc.*)

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2}+v_{\mathrm{ext}}(\boldsymbol{r})+v_{\mathrm{es}}(\boldsymbol{r})+v_{\mathrm{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r})=\epsilon_k\psi_k(\boldsymbol{r})}$$

Kohn-Sham Equations

Basis set:

$$\psi_k(\boldsymbol{r})=\sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

$\downarrow$

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k=\epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij}=\langle\varphi_i|\hat{h}_{\mathrm{KS}}|\varphi_j\rangle$$

$$s_{ij}=\langle\varphi_i|\varphi_j\rangle$$

<u>Many good options:</u>

- Plane waves
- *Augmented* plane waves
  (*Slater 1937; Andersen 1975; etc.*)
- Gaussian-type orbitals

Central decision: the basis set

$$\boxed{\left[-\frac{\nabla^2}{2}+v_{\text{ext}}(\boldsymbol{r})+v_{\text{es}}(\boldsymbol{r})+v_{\text{xc}}(\boldsymbol{r})\right]\psi_k(\boldsymbol{r})=\epsilon_k\psi_k(\boldsymbol{r})}$$

Kohn-Sham
Equations

Basis set:

$$\psi_k(\boldsymbol{r})=\sum_i c_{ki}\varphi_i(\boldsymbol{r})$$

↓

Generalized eigenvalue problem:

$$\boxed{\underline{\underline{h}}\,\underline{c}_k=\epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

$$h_{ij}=\langle\varphi_i|\hat{h}_{\text{KS}}|\varphi_j\rangle$$
$$s_{ij}=\langle\varphi_i|\varphi_j\rangle$$

Many good options:

- Plane waves
- *Augmented* plane waves (*Slater 1937; Andersen 1975; etc.*)
- Gaussian-type orbitals
- *Many* others: (L)MTO, "real-space", numeric atom-centered functions, ...

Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

- $u_i(r)$: Flexible choice - “Anything you like.”

<u>Many popular implementations:</u>
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

- $u_i(r)$: Flexible choice - "Anything you like."

Many popular implementations:
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

$$\boxed{\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r)\right] u_i(r) = \epsilon_i u_i(r)}$$

Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

- $u_i(r)$: Flexible choice - "Anything you like."

<u>Many popular implementations:</u>
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

$$\boxed{\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r)\right] u_i(r) = \epsilon_i u_i(r)}$$

- free-atom like: $v_i(r) = v^{\text{DFT}}_{\text{free atom}}(r)$

- Hydrogen-like: $v_i(r) = z/r$

- free ions, harm. osc. (Gaussians), ...

Our choice: Numeric atom-centered basis sets

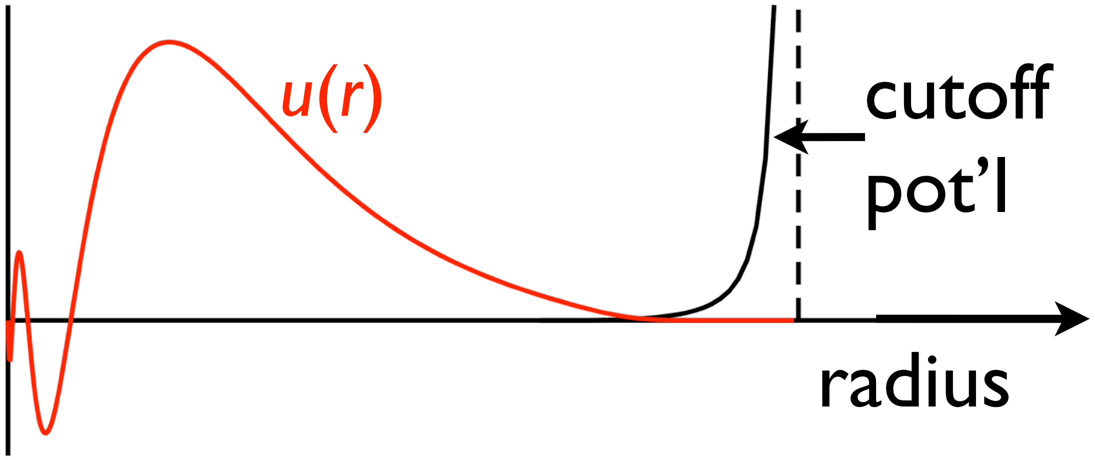
$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

- $u_i(r)$: Flexible choice - “Anything you like.”

Many popular implementations:
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

$$\boxed{\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r)\right] u_i(r) = \epsilon_i u_i(r)}$$

- free-atom like: $v_i(r) = v^{\text{DFT}}_{\text{free atom}}(r)$
- Hydrogen-like: $v_i(r) = z/r$
- free ions, harm. osc. (Gaussians), ...



Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

- $u_i(r)$: Flexible choice - “Anything you like.”

→ Localized; ”naturally” all-electron

<u>Many popular implementations:</u>
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
“Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals”,
Computer Physics Communications ***180****, 2175-2196 (2009)*

Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

<u>Many popular implementations:</u>
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

- $u_i(r)$: Flexible choice - "Anything you like."

→ Localized; "naturally" all-electron

→ The choice of <u>efficient</u> and of <u>enough</u> radial functions is obviously important

V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
"Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals",
Computer Physics Communications ***180****, 2175-2196 (2009)*

Our choice: Numeric atom-centered basis sets

$$\boxed{\varphi_{i[lm]}(\boldsymbol{r}) = \frac{u_i(r)}{r} \cdot Y_{lm}(\Omega)}$$

<u>Many popular implementations:</u>
DMol^3 (Delley), FPLO (Eschrig *et al.*), PLATO (Horsfield *et al.*), PAOs (Siesta, Conquest, OpenMX^2, Fireball, ...)

- $u_i(r)$: Flexible choice - “Anything you like.”

→ Localized; ”naturally” all-electron

→ The choice of <u>efficient</u> and of <u>enough</u> radial functions is obviously important

→ We have a basis set library for all elements (1-102), from fast qualitative to meV-converged (total energy, LDA/GGA) calculations - <u>efficient and accurate approach</u>

V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
“*Ab Initio* Molecular Simulations with Numeric Atom-Centered Orbitals”,
Computer Physics Communications ***180***, *2175-2196 (2009)*

How to find good basis sets - and how far do they carry?

Lots of shapes $u(r)$ available:
Confined free atoms, hydrogen-like, free ions, ...

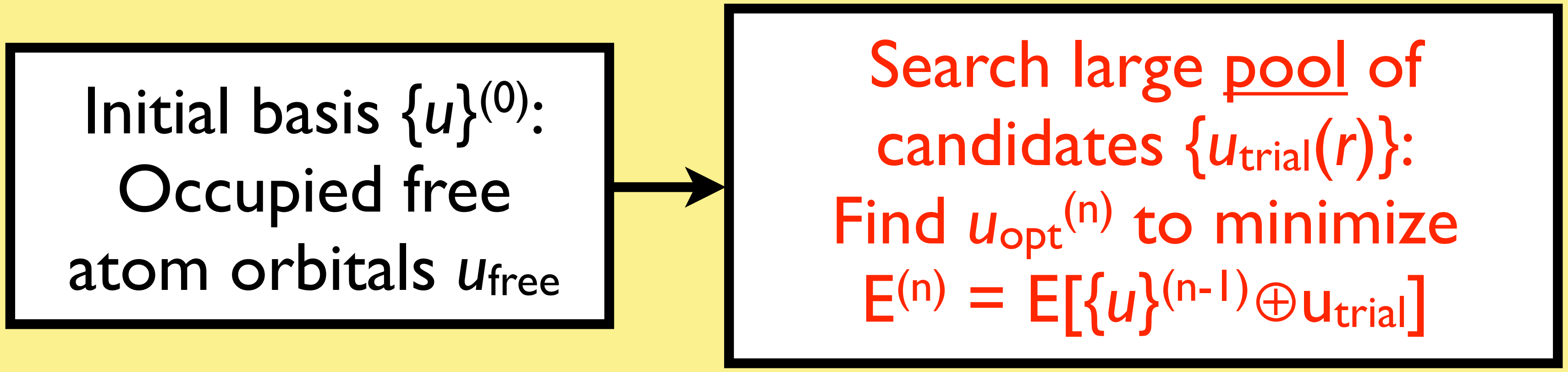
Simple robust selection strategy:

Initial basis $\{u\}^{(0)}$:
Occupied free
atom orbitals u_{free}

How to find good basis sets - and how far do they carry?

Lots of shapes $u(r)$ available:
Confined free atoms, hydrogen-like, free ions, ...

Simple robust selection strategy:



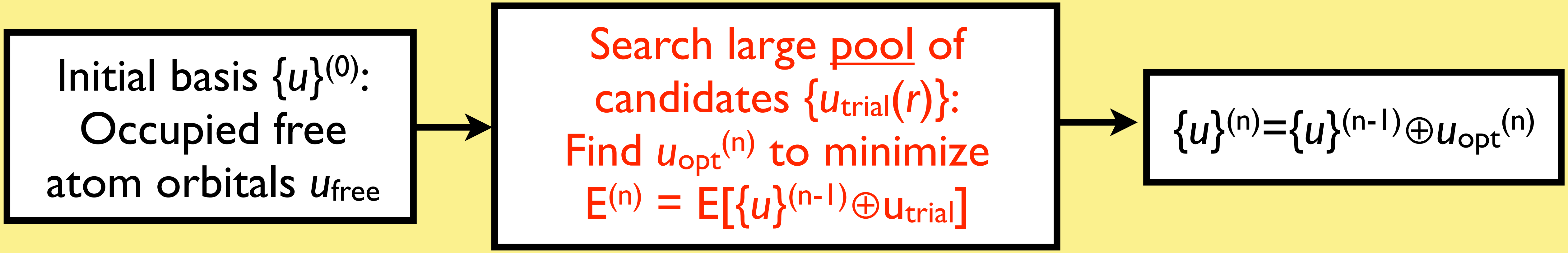
Initial basis $\{u\}^{(0)}$:
Occupied free
atom orbitals u_{free}

Search large pool of
candidates $\{u_{\text{trial}}(r)\}$:
Find $u_{\text{opt}}^{(n)}$ to minimize
$E^{(n)} = E[\{u\}^{(n-1)} \oplus u_{\text{trial}}]$

How to find good basis sets - and how far do they carry?

Lots of shapes $u(r)$ available:
Confined free atoms, hydrogen-like, free ions, ...

Simple robust selection strategy:



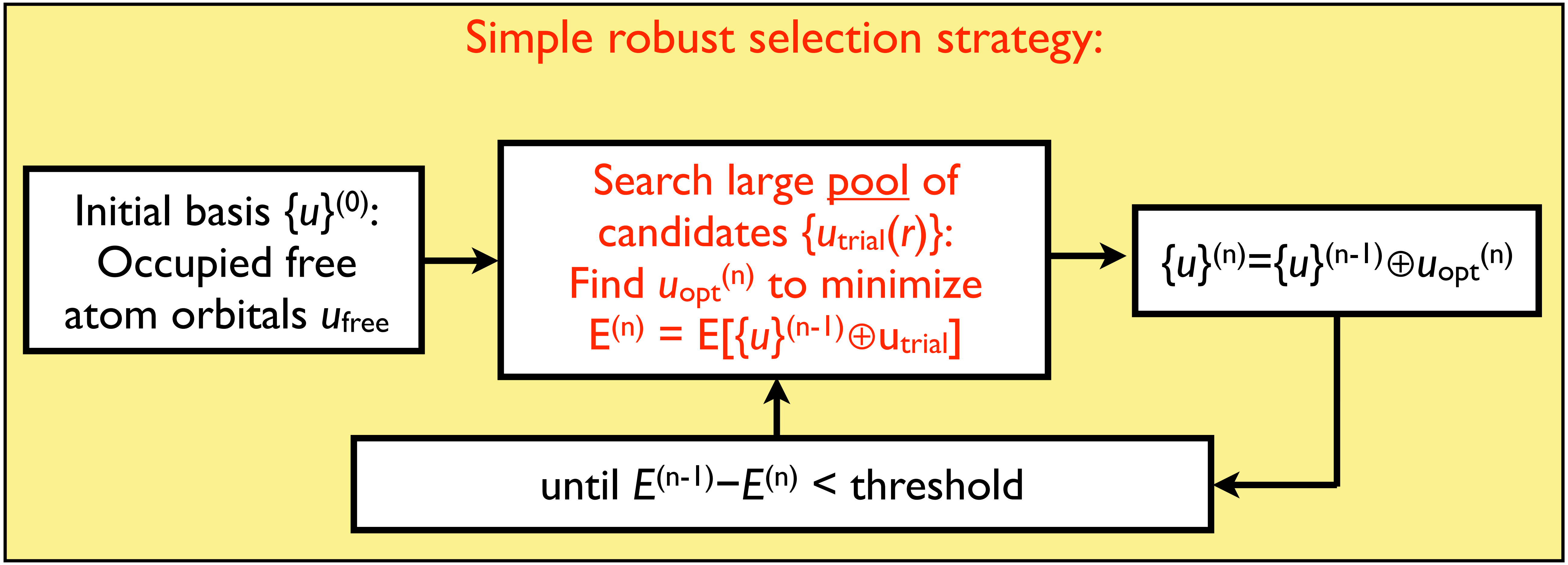
Initial basis $\{u\}^{(0)}$: Occupied free atom orbitals u_{free}

Search large <u>pool</u> of candidates $\{u_{\text{trial}}(r)\}$:
Find $u_{\text{opt}}^{(\text{n})}$ to minimize
$E^{(\text{n})} = E[\{u\}^{(\text{n-1})} \oplus u_{\text{trial}}]$

$\{u\}^{(\text{n})} = \{u\}^{(\text{n-1})} \oplus u_{\text{opt}}^{(\text{n})}$

How to find good basis sets - and how far do they carry?

Lots of shapes $u(r)$ available:
Confined free atoms, hydrogen-like, free ions, ...



Simple robust selection strategy:

Initial basis $\{u\}^{(0)}$: Occupied free atom orbitals u_{free}

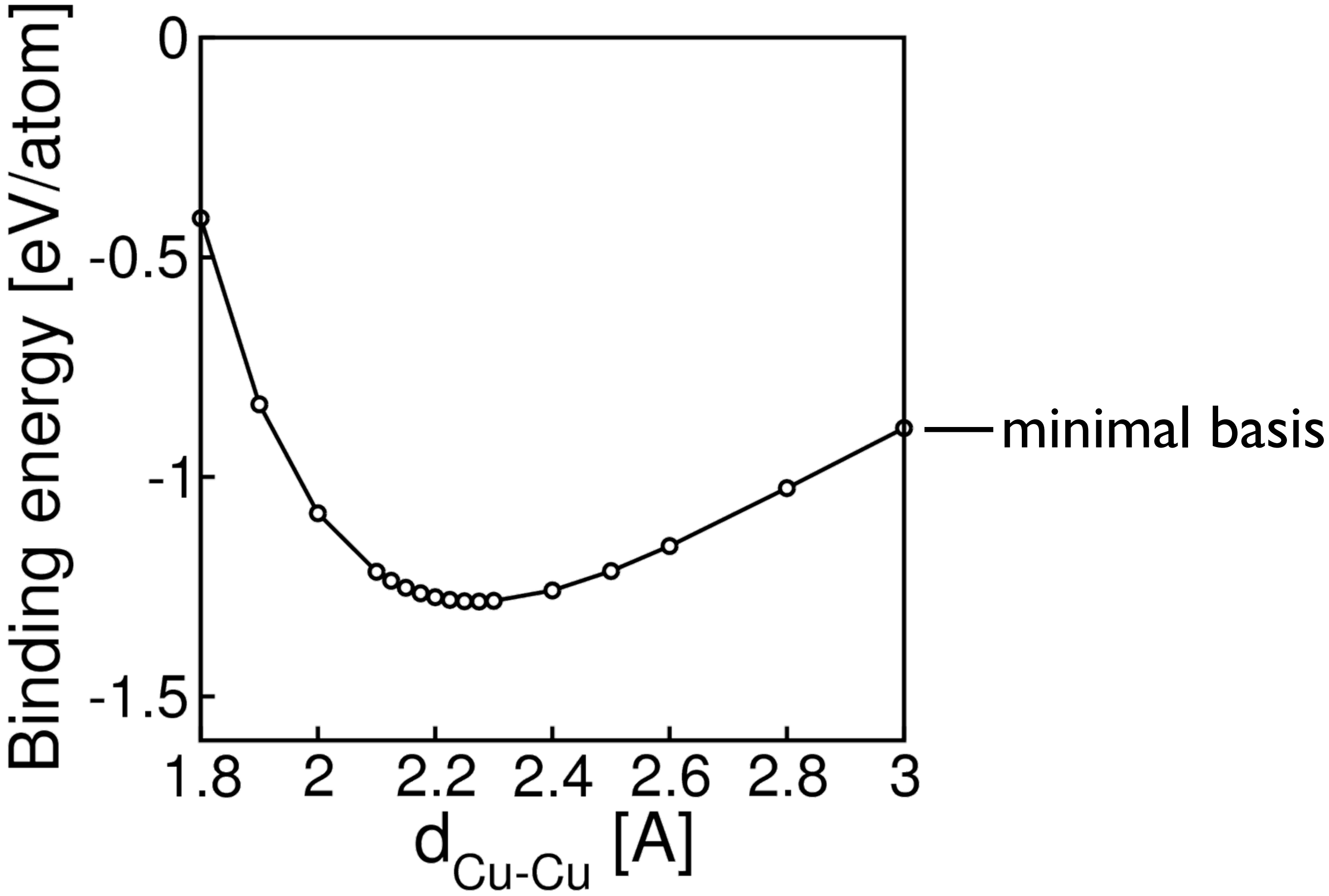
Search large pool of candidates $\{u_{\text{trial}}(r)\}$: Find $u_{\text{opt}}^{(n)}$ to minimize $E^{(n)} = E[\{u\}^{(n-1)} \oplus u_{\text{trial}}]$

$\{u\}^{(n)} = \{u\}^{(n-1)} \oplus u_{\text{opt}}^{(n)}$

until $E^{(n-1)} - E^{(n)} <$ threshold

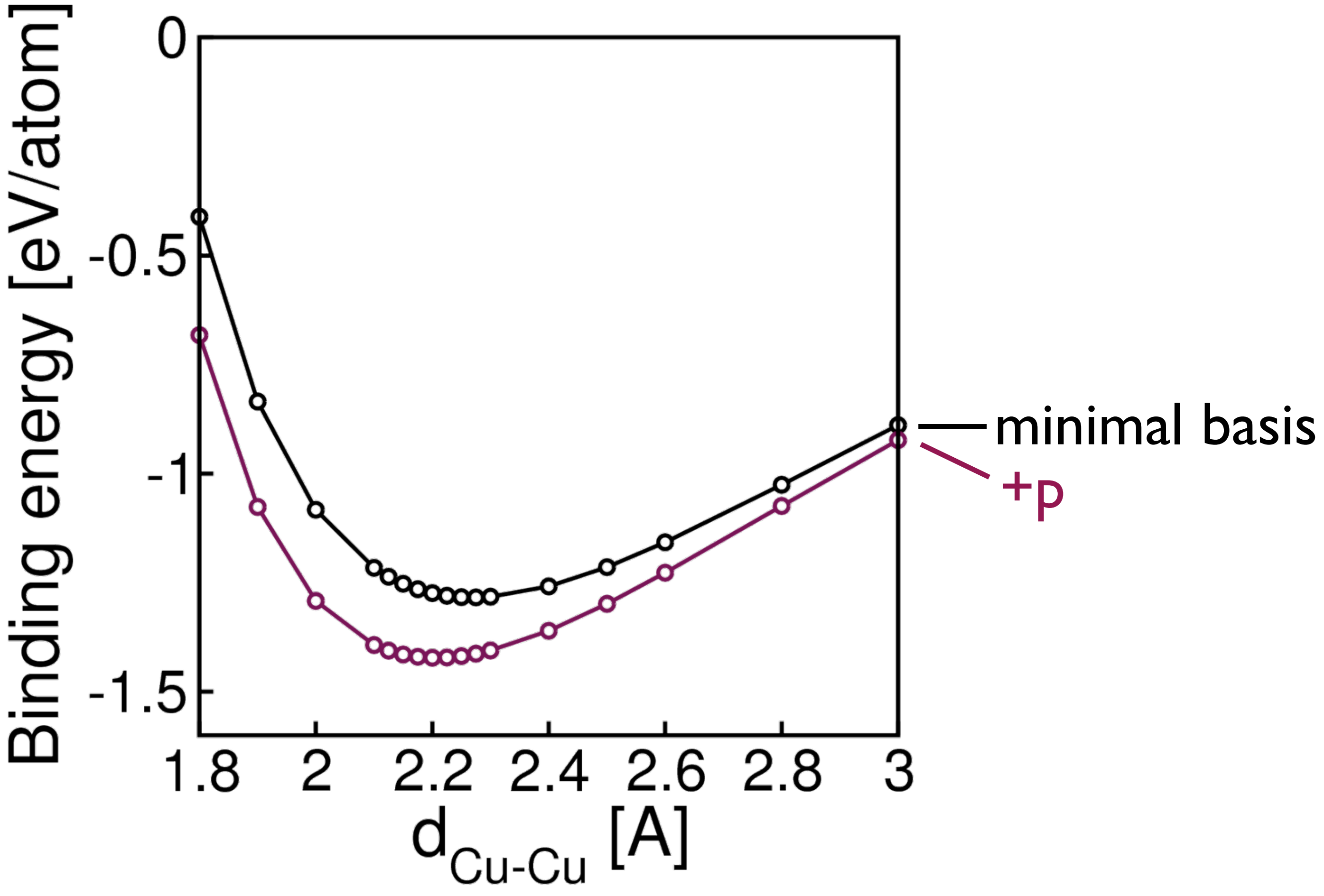
Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets



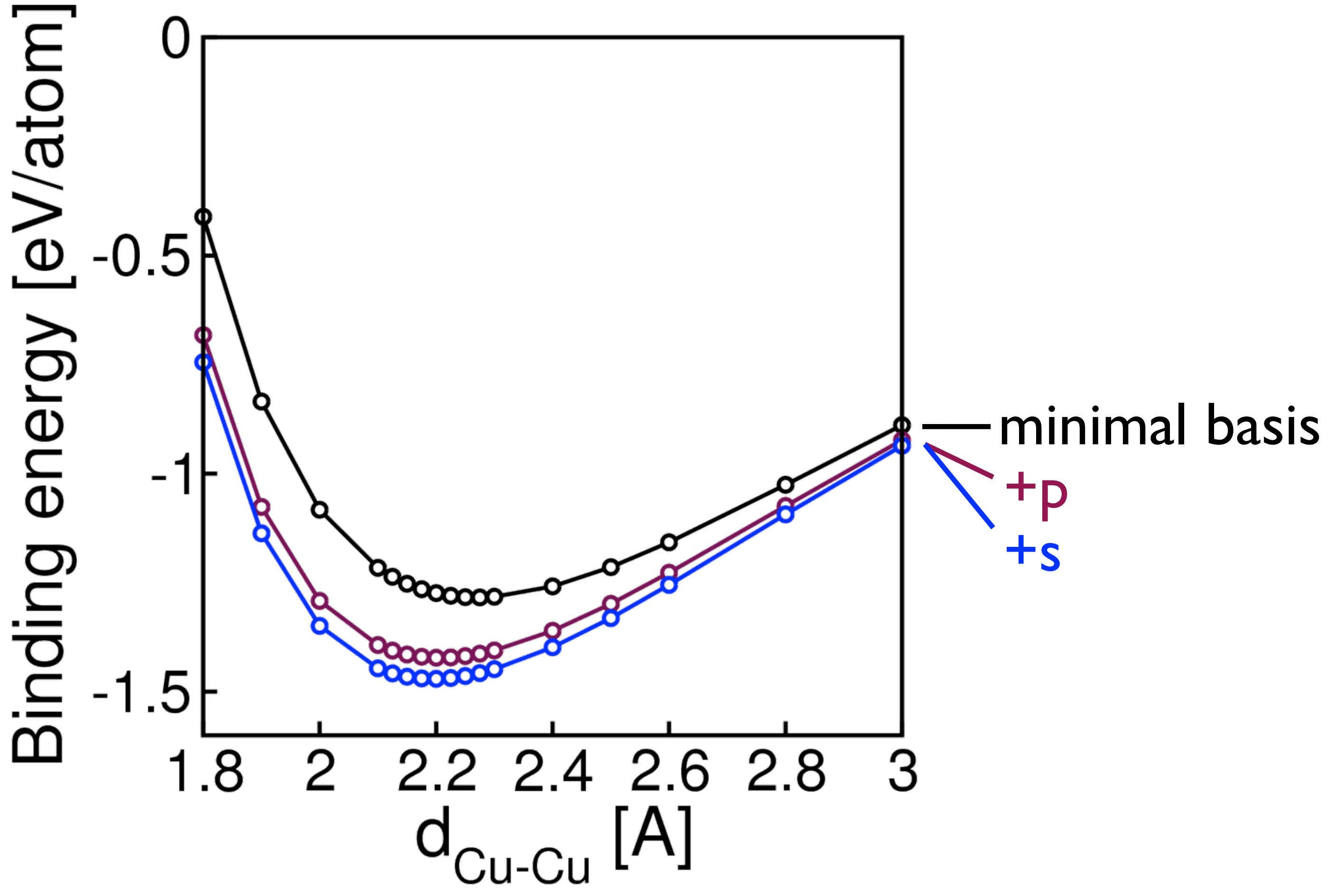
Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets



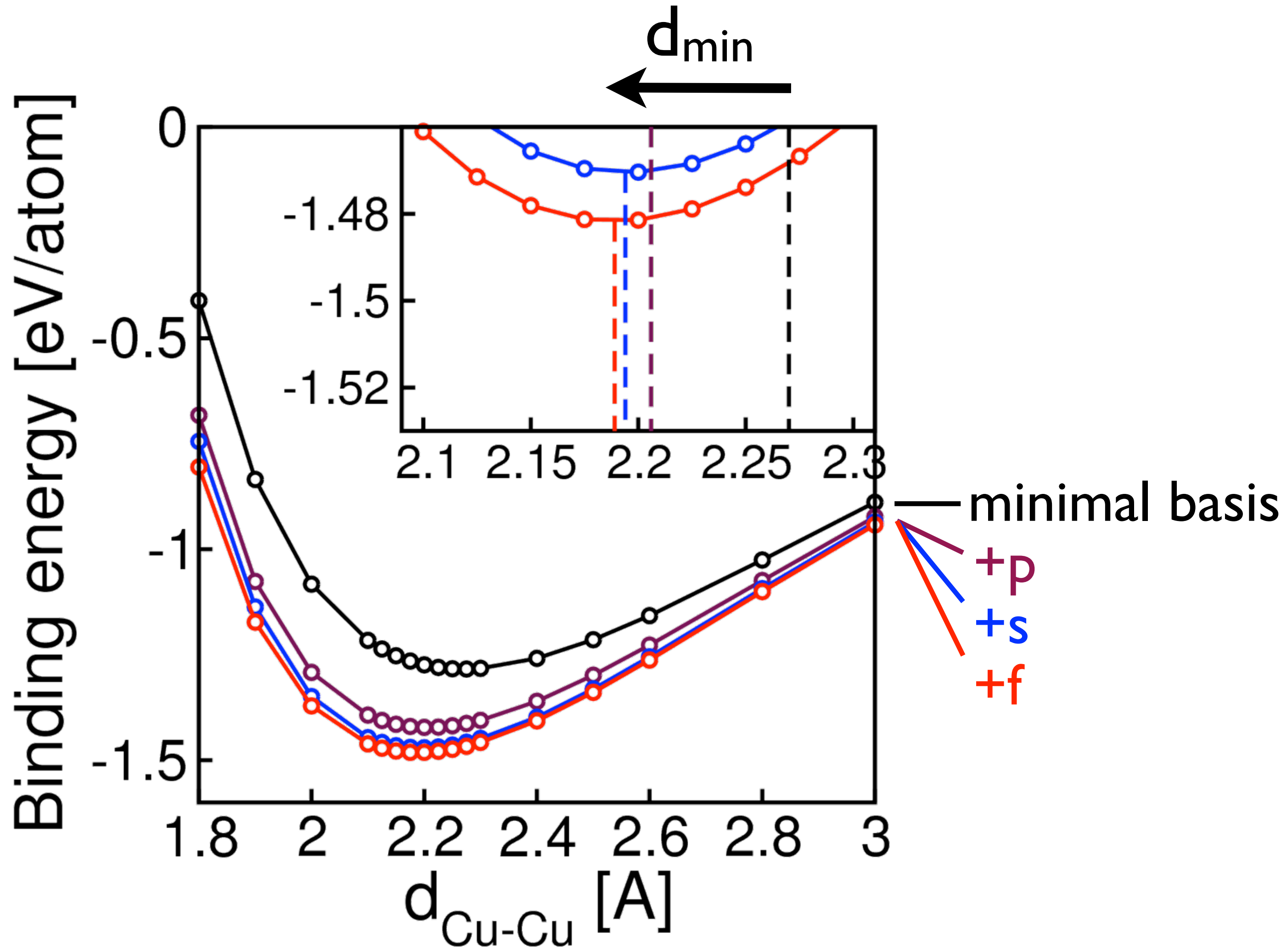
Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets



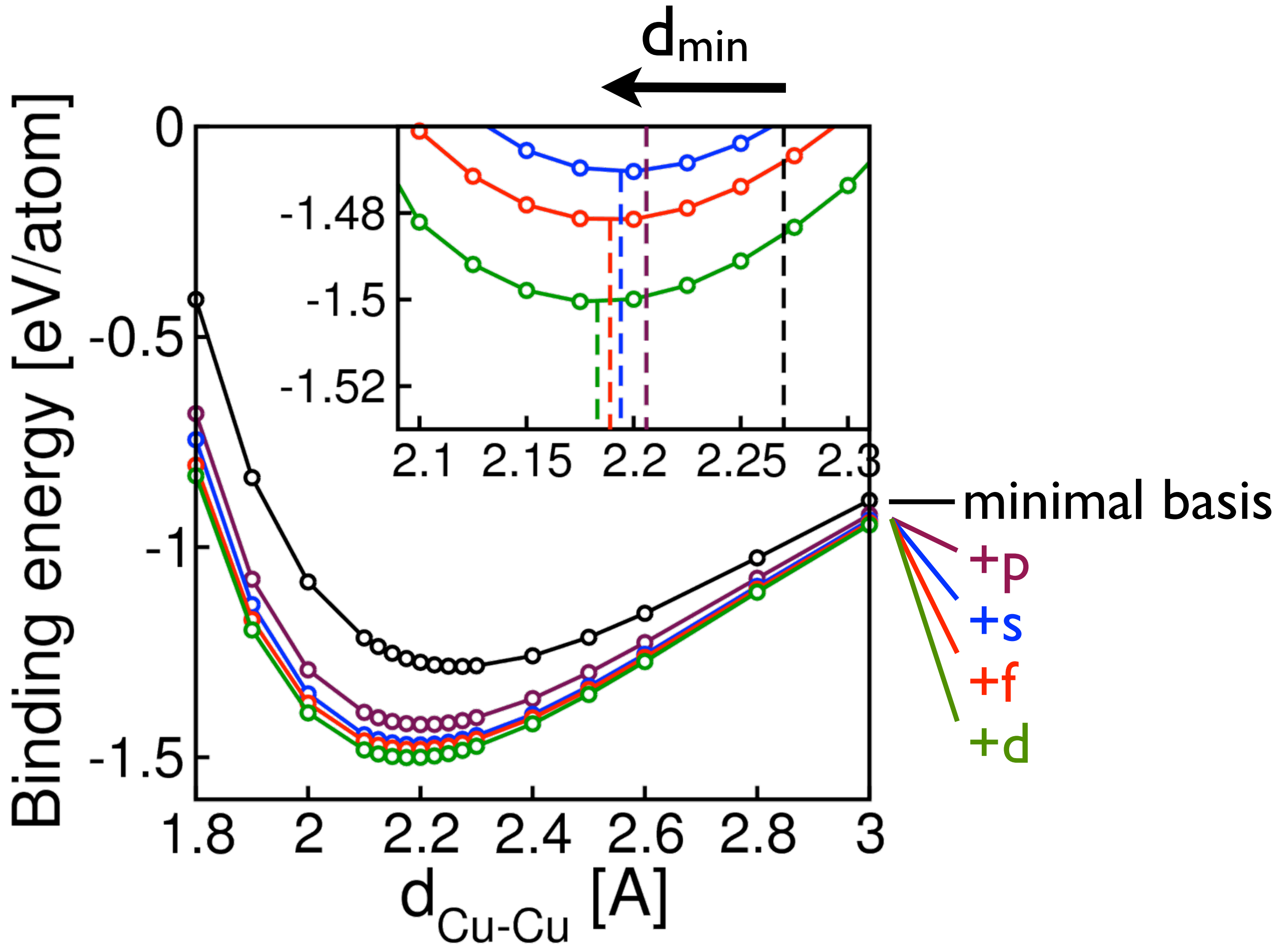
Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets



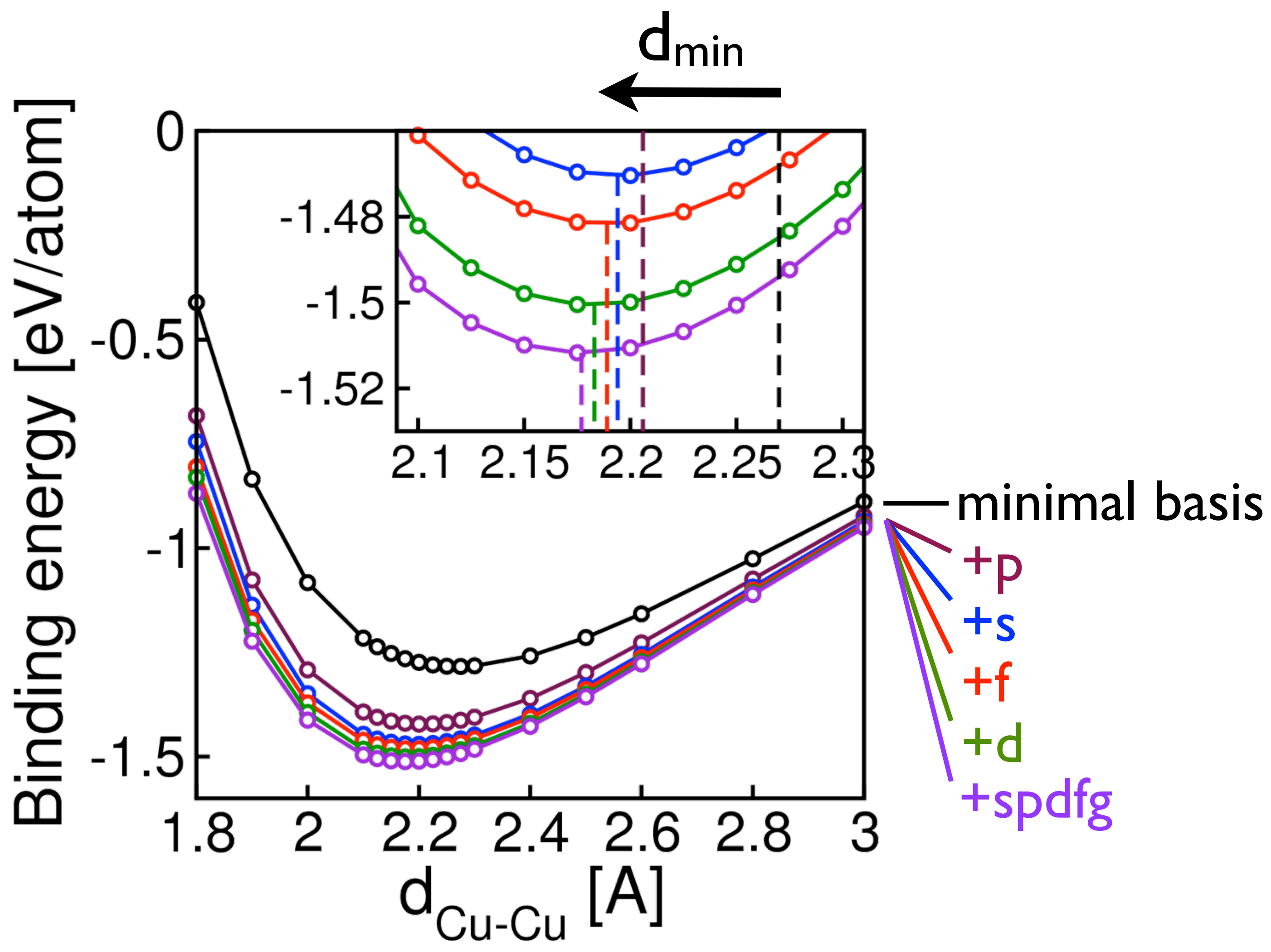
Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets



Dimer binding curves to optimize the basis

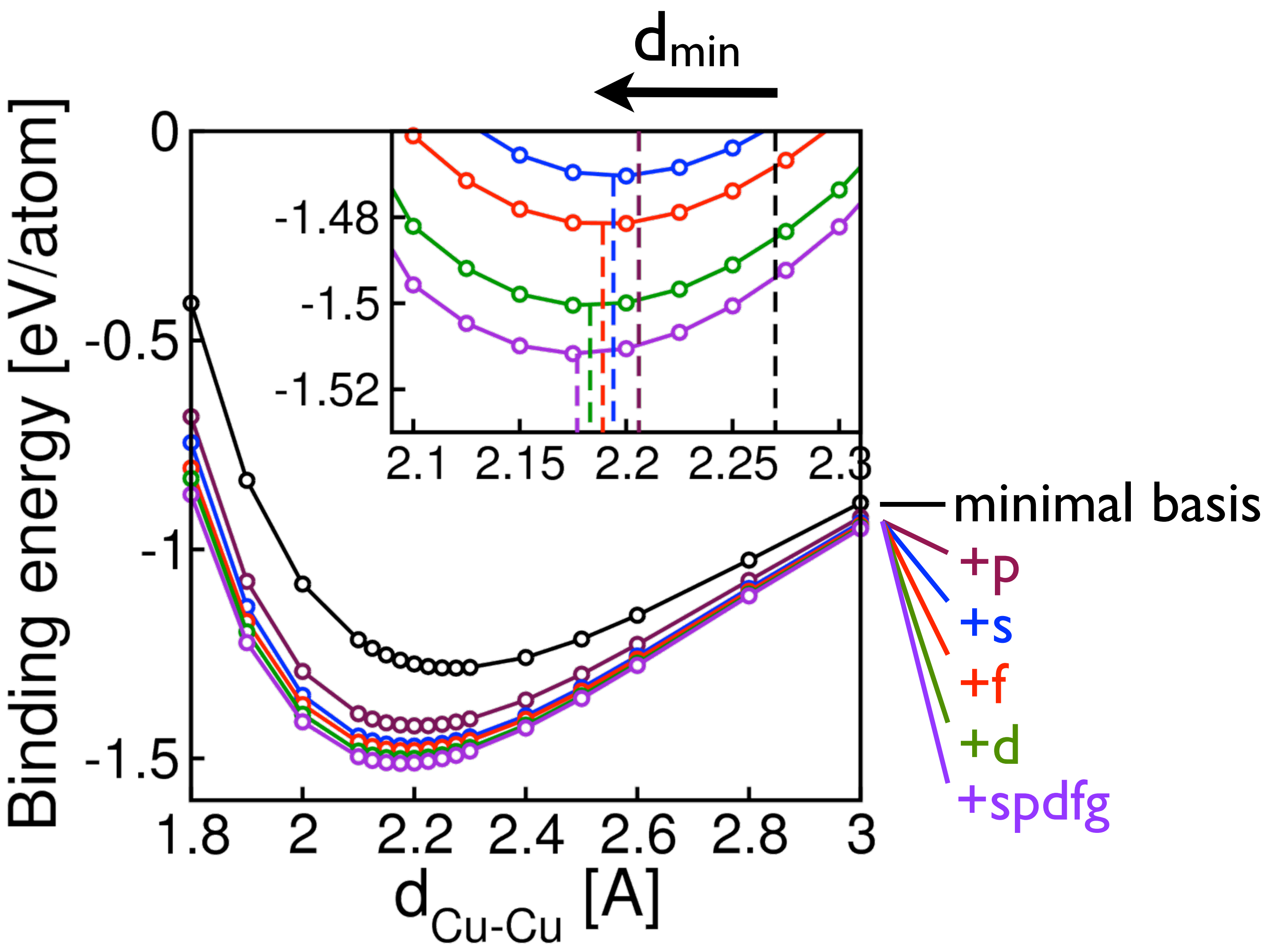
Example: Cu_2 binding curve for different basis sets

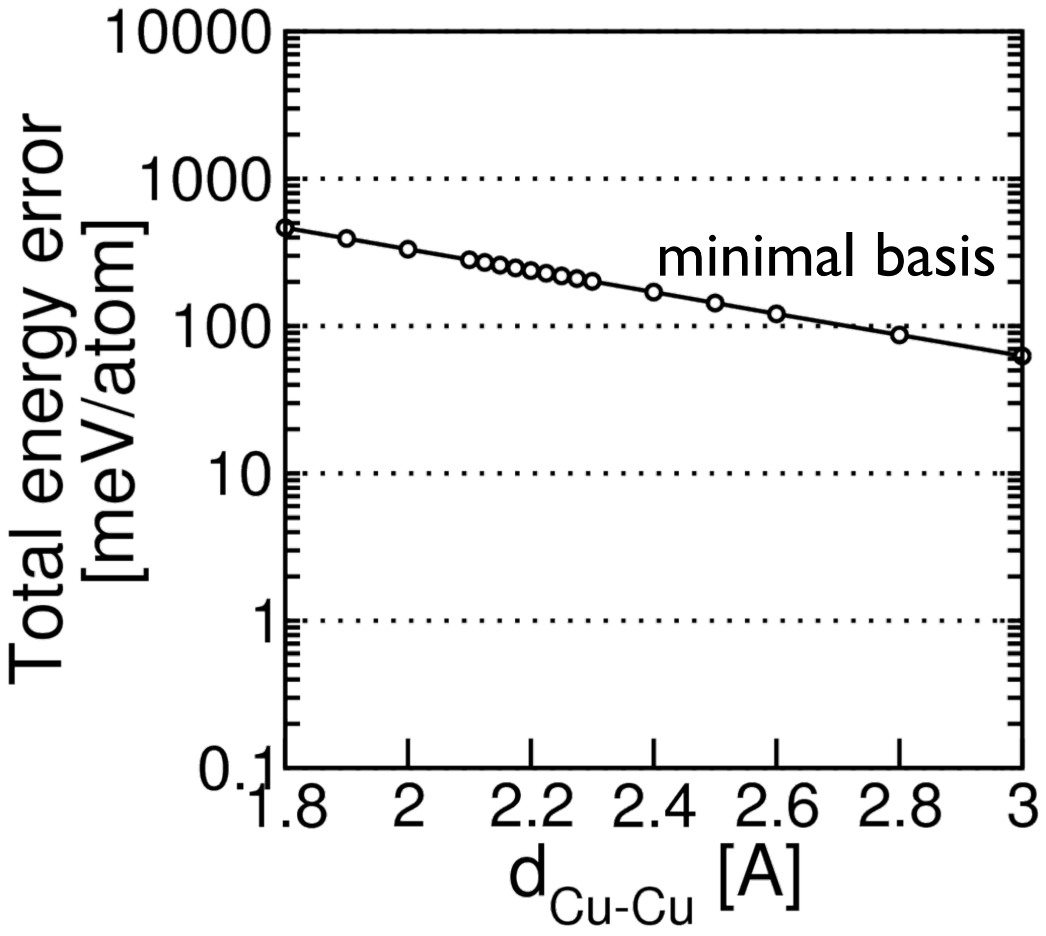


- Increasing basis set: clear drift of d$_{min}$ towards smaller values

Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets

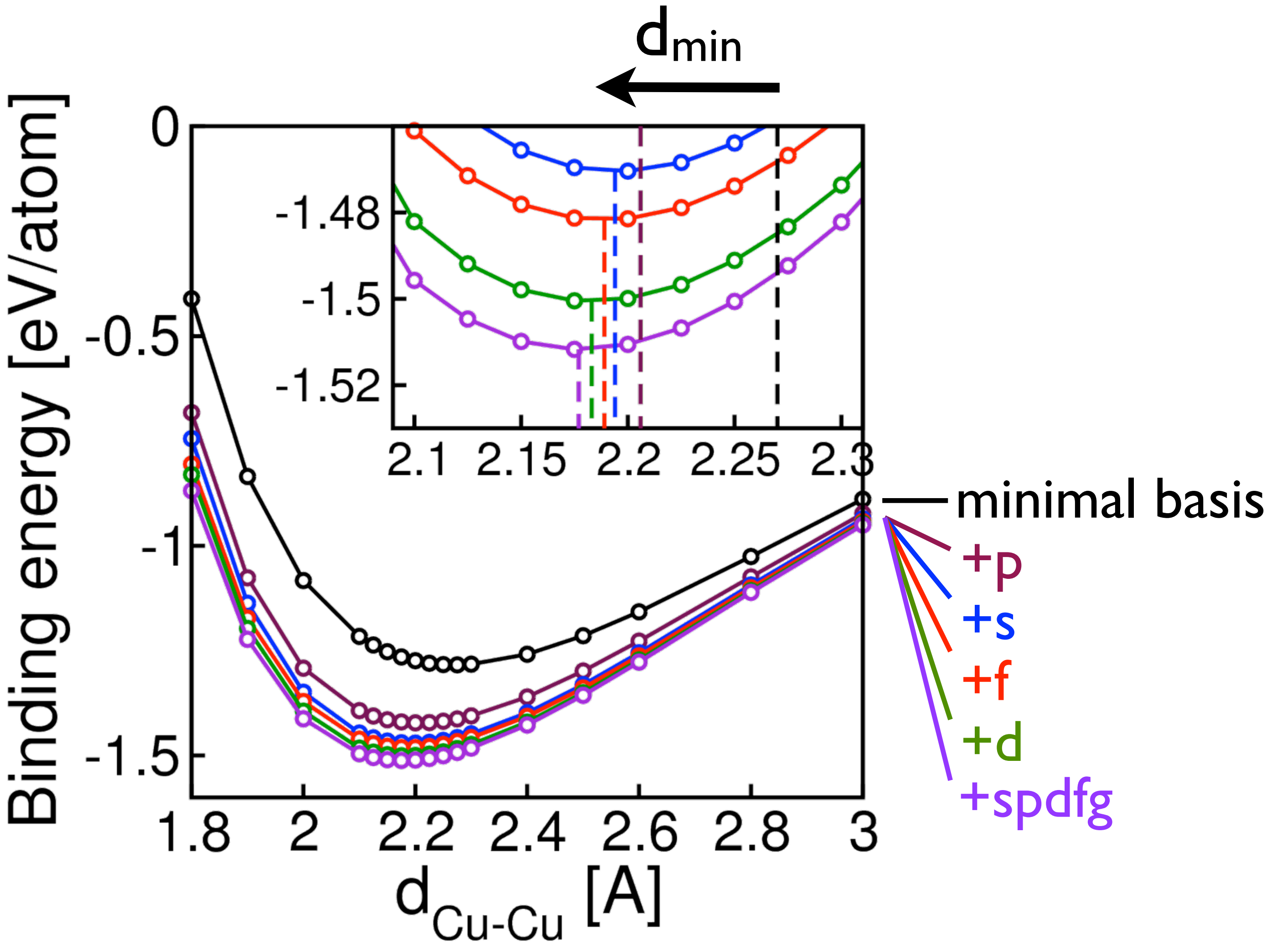


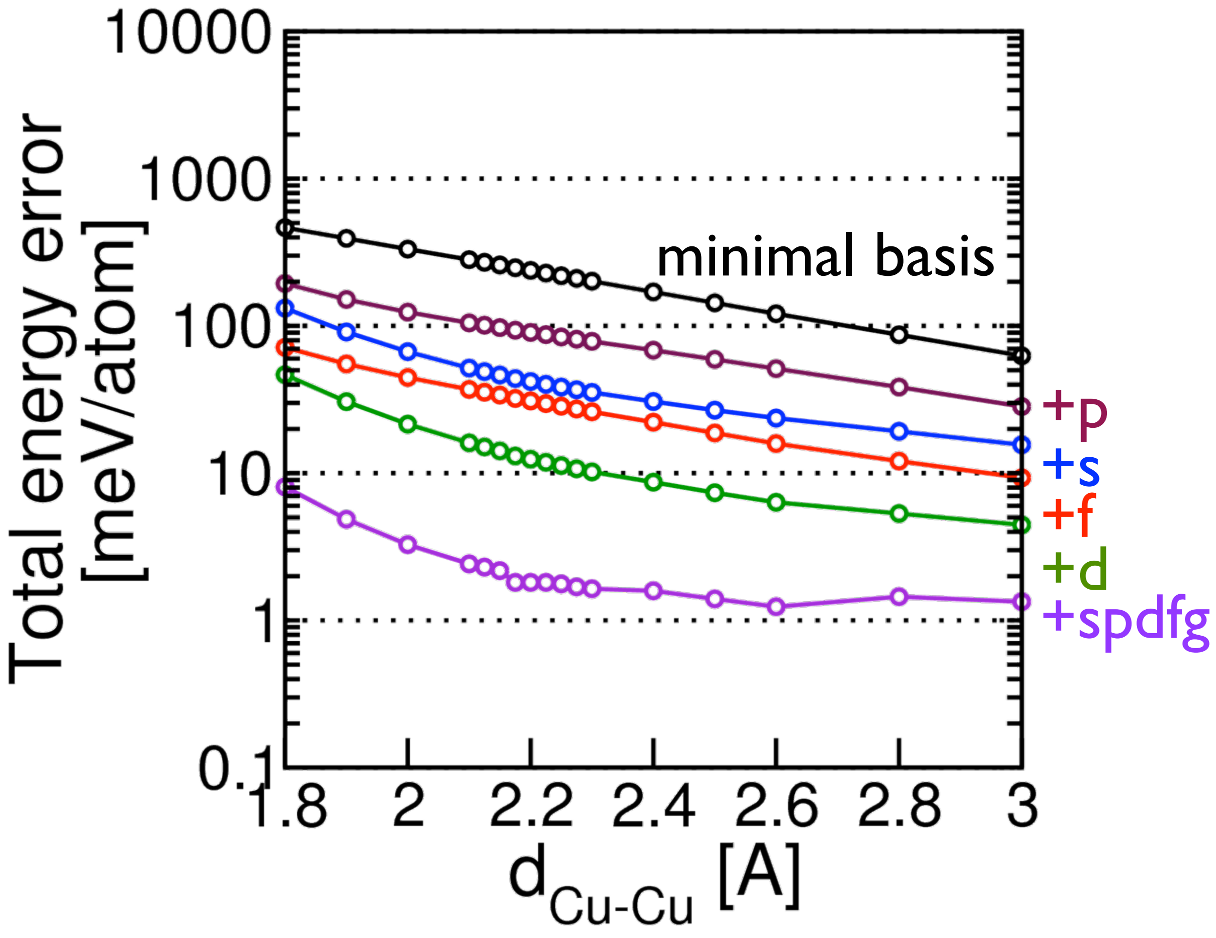


- Increasing basis set: clear drift of d_{min} towards smaller values
- Reason: Minimal basis → exponential basis set error as function of d

Dimer binding curves to optimize the basis

Example: Cu_2 binding curve for different basis sets

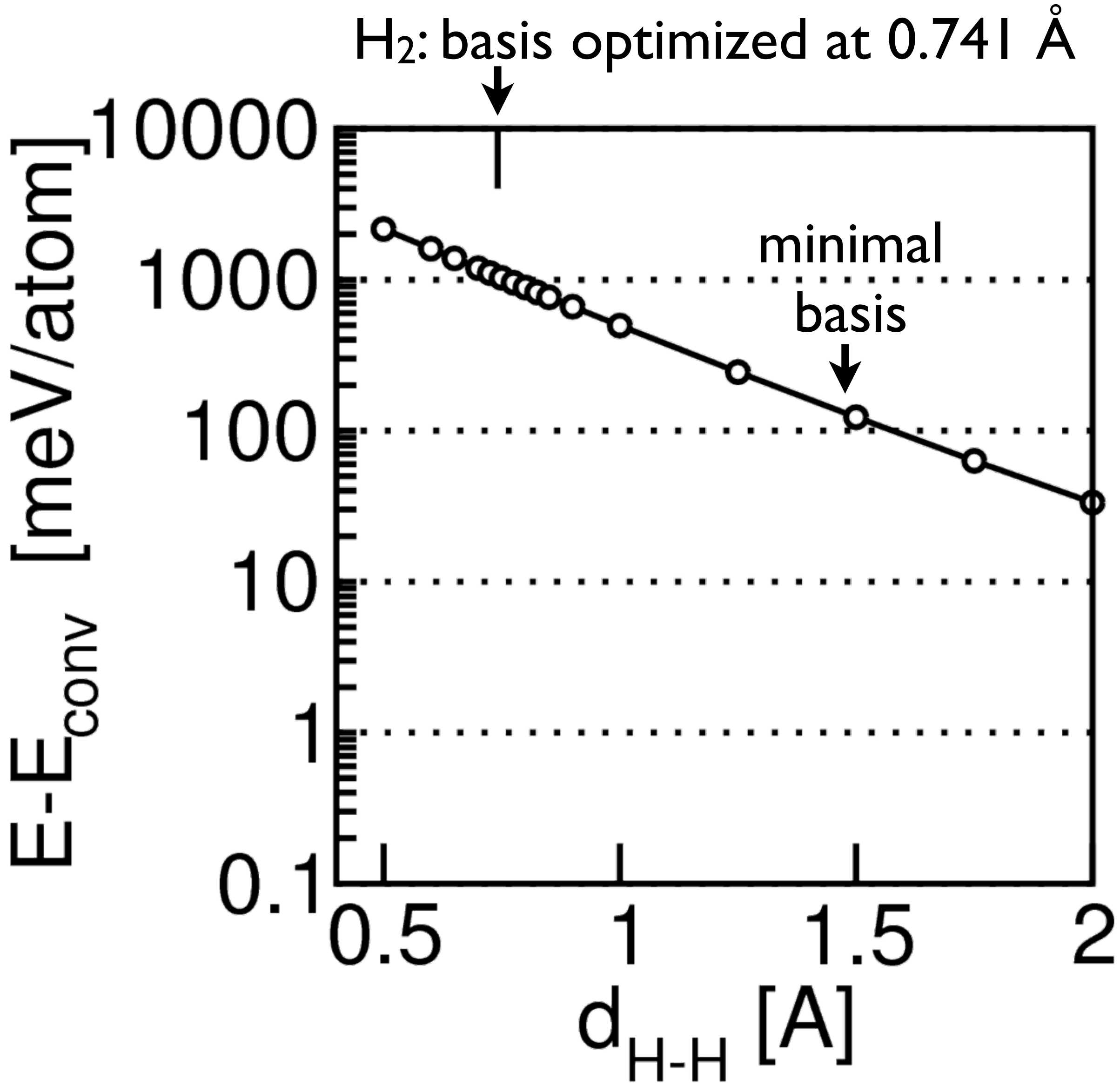




- Increasing basis set: clear drift of d_{min} towards smaller values
- Minimal basis → exponential basis set error as function of d
- meV-level accuracy for practical basis sets; small d require larger basis

Dimer binding curves to optimize the basis

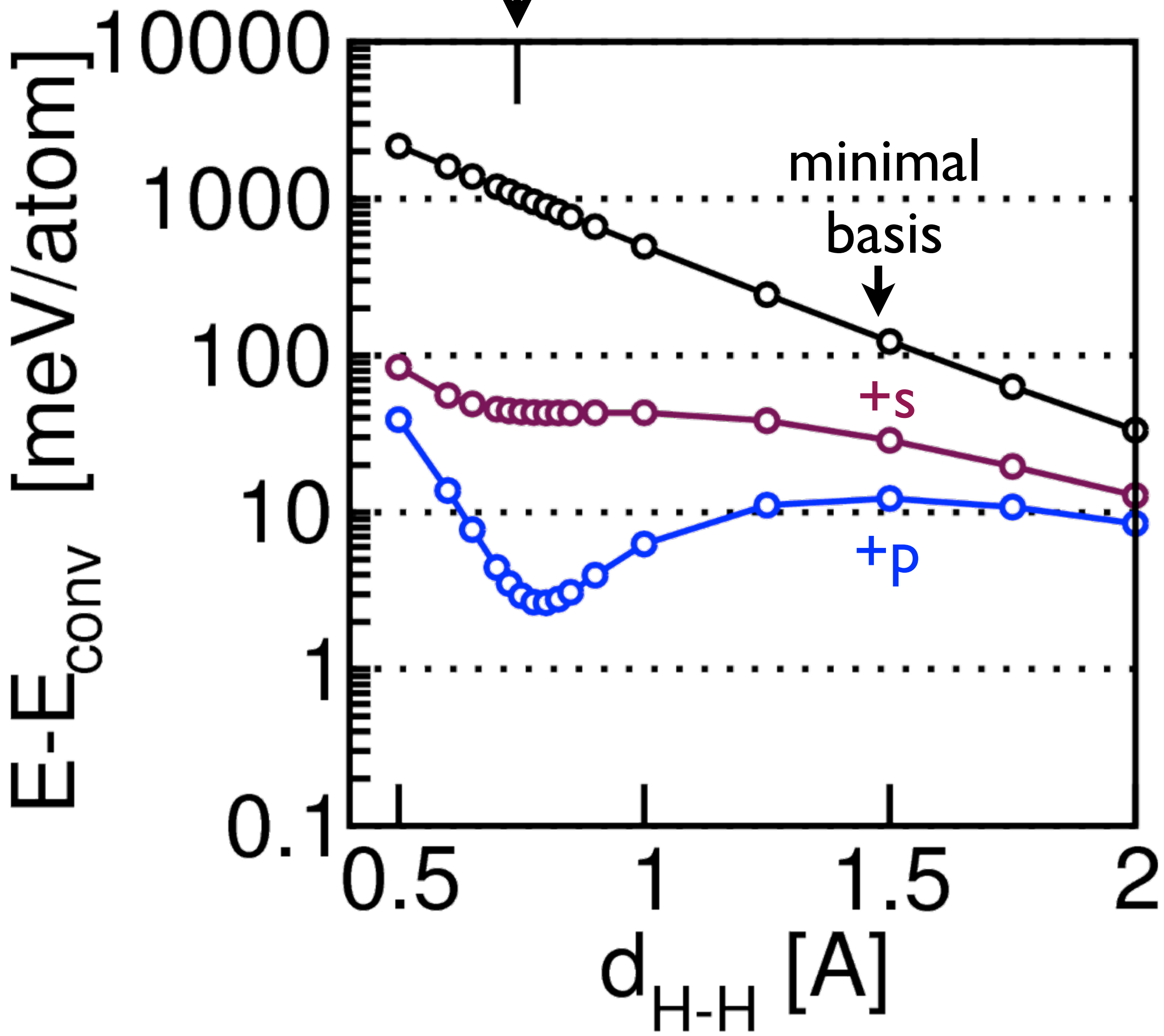
Extreme example: H_2 (only one occupied orbital)



Dimer binding curves to optimize the basis

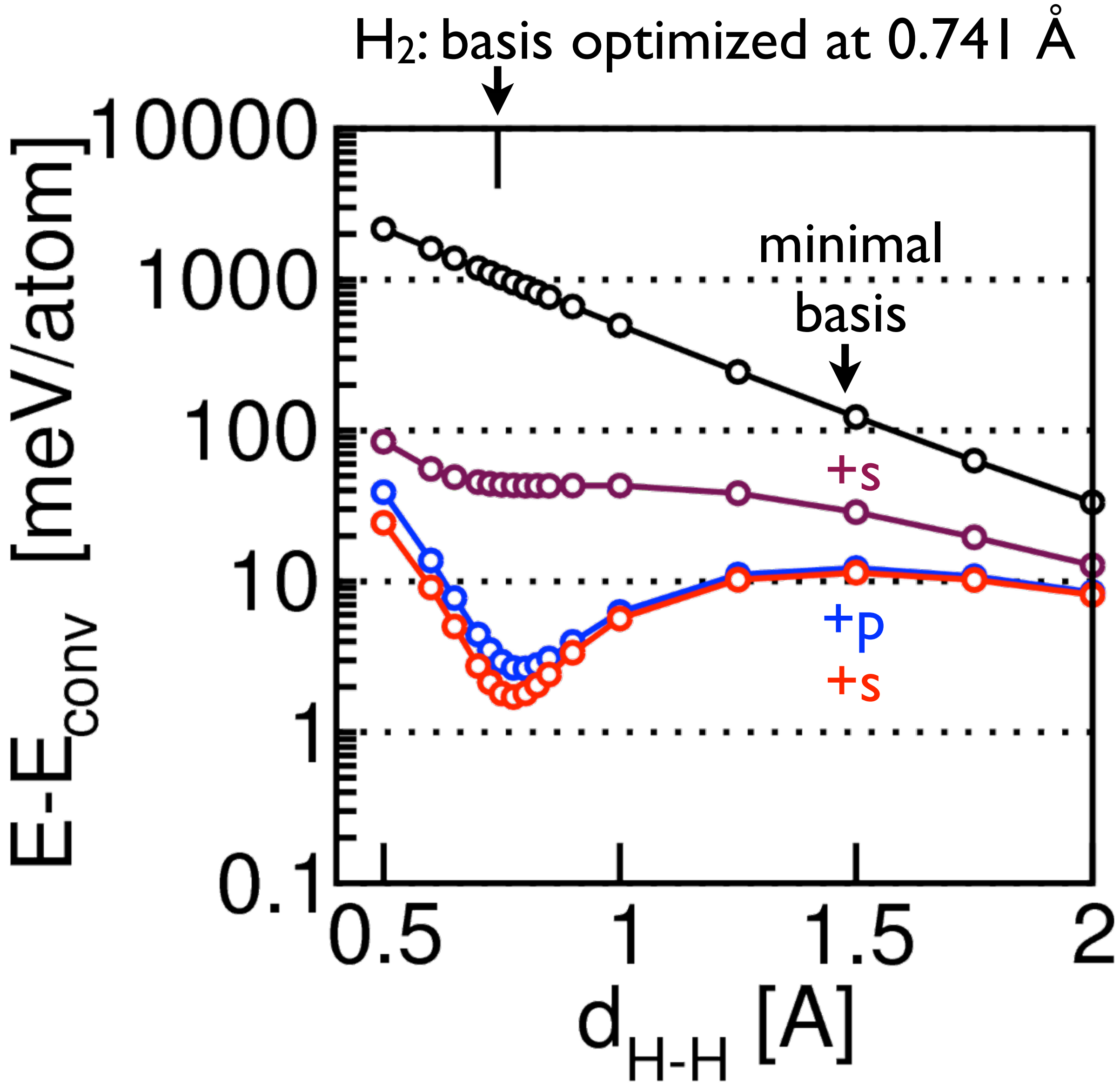
Extreme example: H_2 (only one occupied orbital)

H_2: basis optimized at 0.741 Å



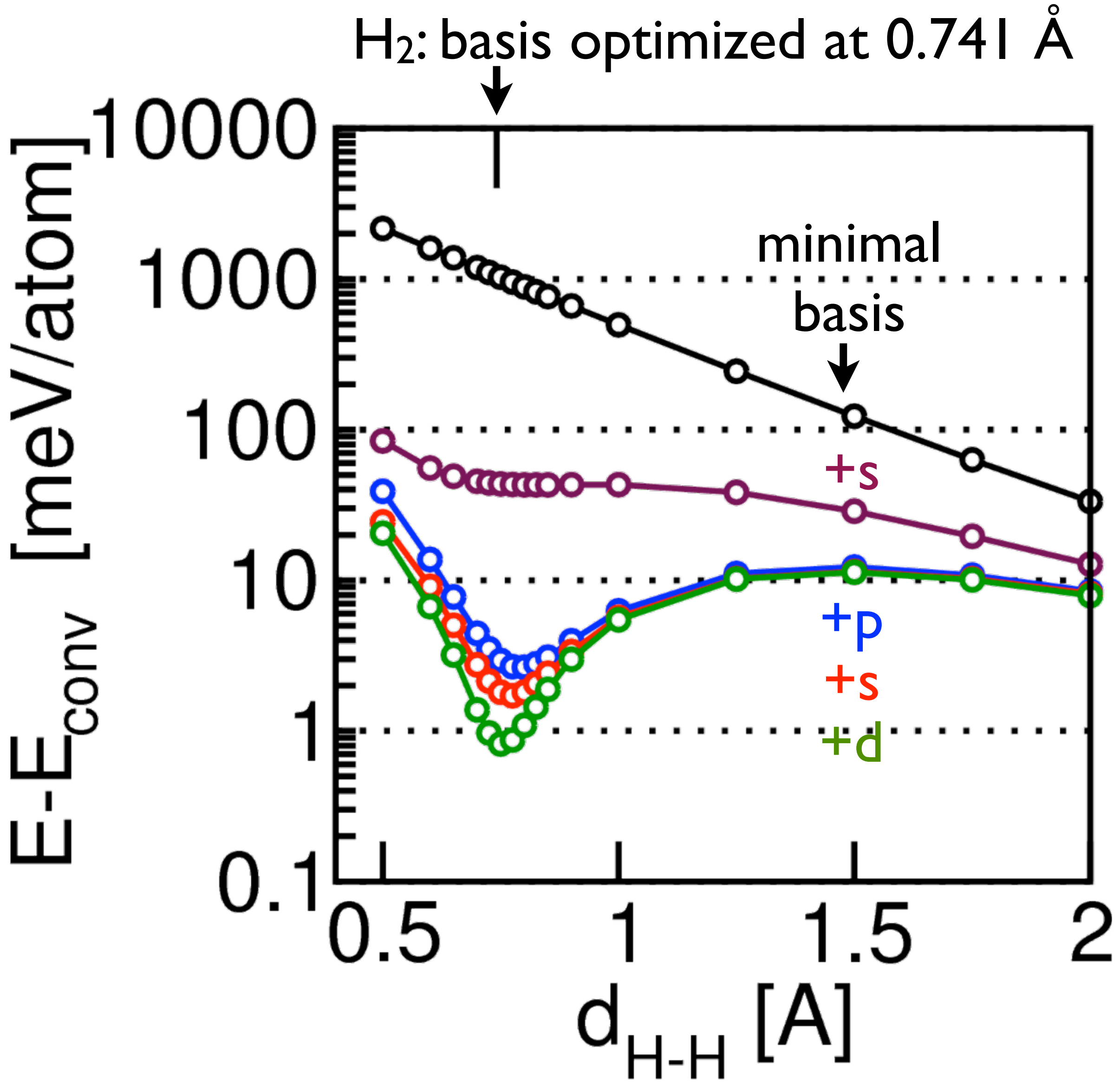
Dimer binding curves to optimize the basis

Extreme example: H_2 (only one occupied orbital)



Dimer binding curves to optimize the basis

Extreme example: H_2 (only one occupied orbital)

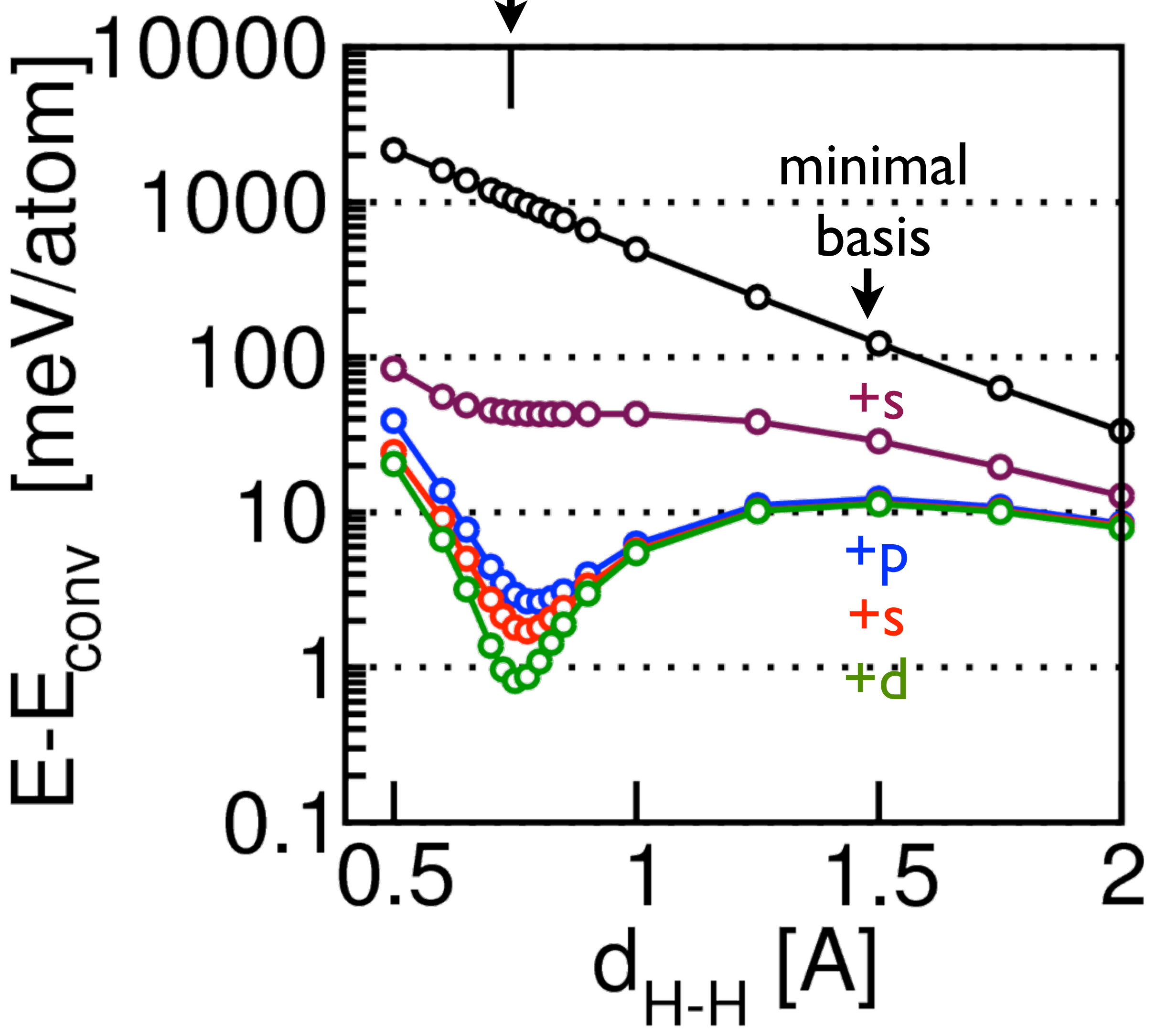


Light elements (particularly H):
Basis optimization for single geometry not enough!

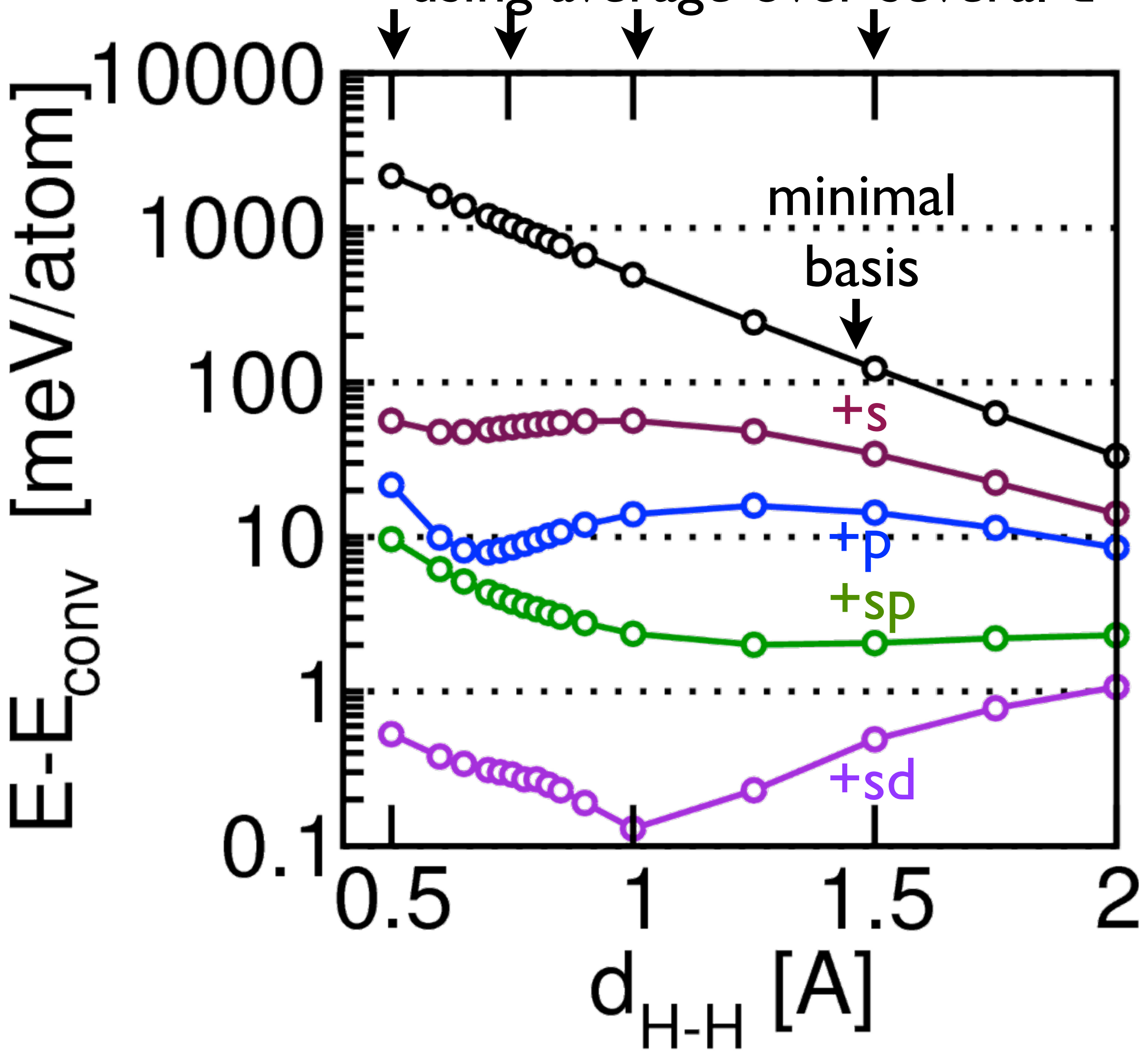
Dimer binding curves to optimize the basis

Extreme example: H_2 (only one occupied orbital)

H_2: basis optimized at 0.741 Å



H_2: basis optimized using average over several d



Light elements (particularly H):

Basis optimization for single geometry not enough!

Robust basis construction:

Optimize average of several dimer bond distances

Iterative selection of NAO basis functions

"Pool" of trial basis functions:
2+ ionic *u(r)*
Hydrogen-like *u(r)* for *Z*=0.1-20

Optimization target:
Non-selfconsistent symmetric dimers, averaged for different *d*

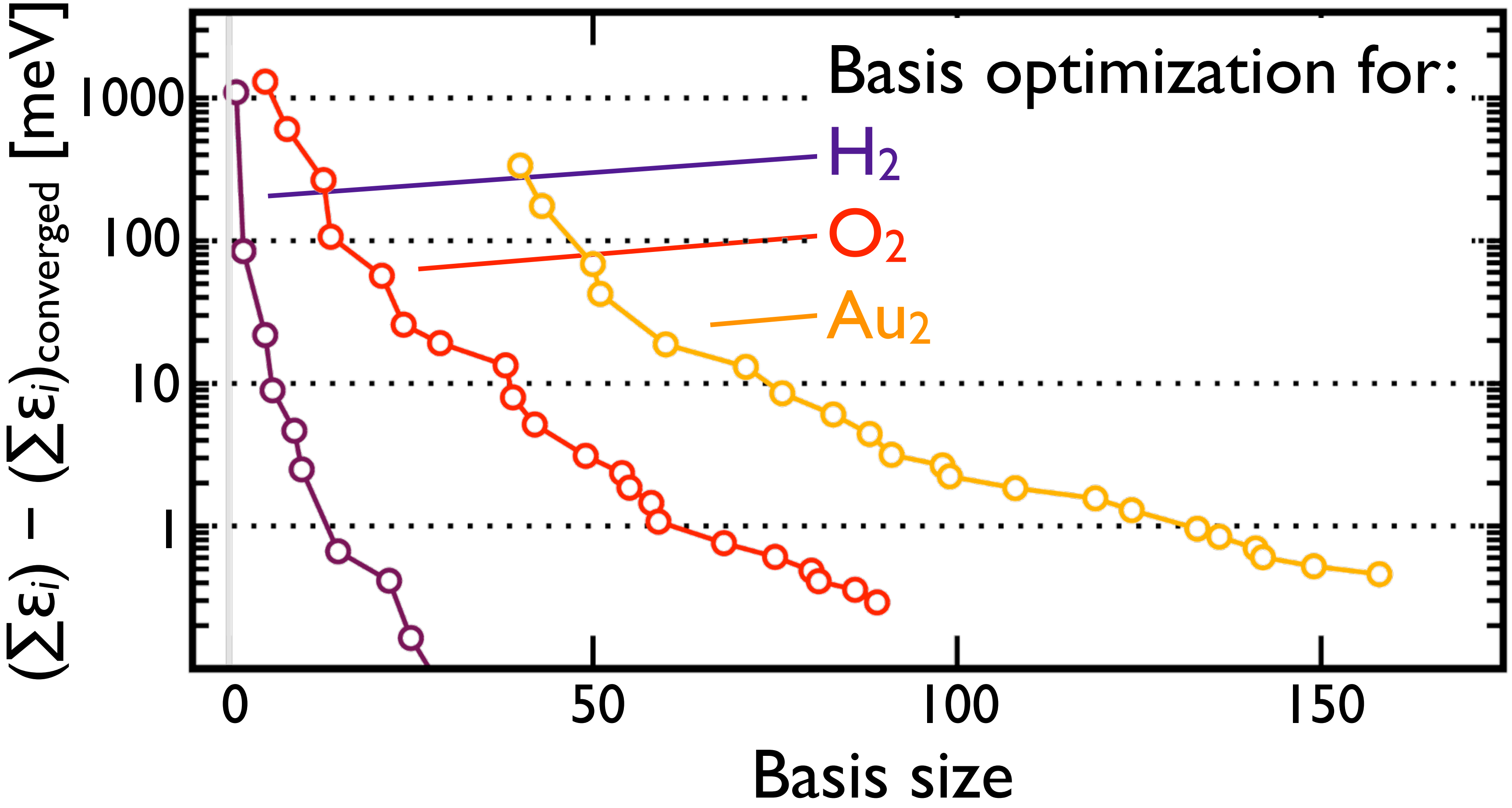
Pick basis functions one by one: T*otal energy* convergence

Iterative selection of NAO basis functions

<u>"Pool" of trial basis functions:</u>
2+ ionic *u(r)*
Hydrogen-like *u(r)* for *Z*=0.1-20

<u>Optimization target:</u>
Non-selfconsistent symmetric dimers, averaged for different *d*

Pick basis functions one by one: *Total energy* convergence



In practice: Hierarchical basis set library for all elements

	H	C	O	Au
minimal	1s	[He]+2s2p	[He]+2s2p	[Xe]+6s5d4f
Tier 1	H(2s,2.1)	H(2p,1.7)	H(2p,1.8)	Au^{2+}(6p)
	H(2p,3.5)	H(3d,6.0)	H(3d,7.6)	H(4f,7.4)
		H(2s,4.9)	H(3s,6.4)	Au^{2+}(6s)
				H(5g,10)
				H(6h,12.8)
				H(3d,2.5)
Tier 2	H(1s,0.85)	H(4f,9.8)	H(4f,11.6)	H(5f,14.8)
	H(2p,3.7)	H(3p,5.2)	H(3p,6.2)	H(4d,3.9)
	H(2s,1.2)	H(3s,4.3)	H(3d,5.6)	H(3p,3.3)
	H(3d,7.0)	H(5g,14.4)	H(5g,17.6)	H(1s,0.45)
		H(3d,6.2)	H(1s,0.75)	H(5g,16.4)
				H(6h,13.6)
Tier 3	H(4f,11.2)	H(2p,5.6)	O^{2+}(2p)	H(4f,5.2)*
	H(3p,4.8)	H(2s,1.4)	H(4f,10.8)	H(4d,5.0)
	...	...	...	...

Systematic hierarchy of basis (sub)sets, iterative *automated* construction based on *dimers*

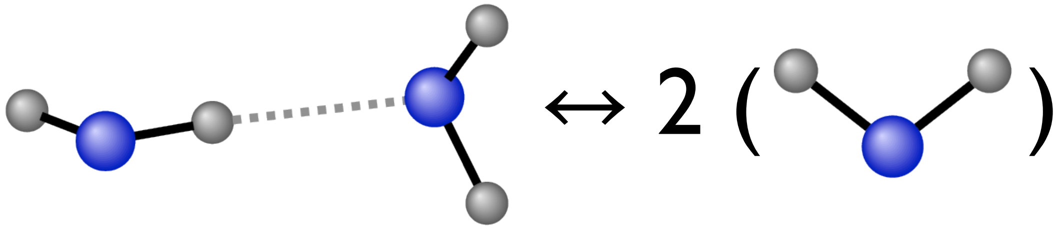
"First tier"

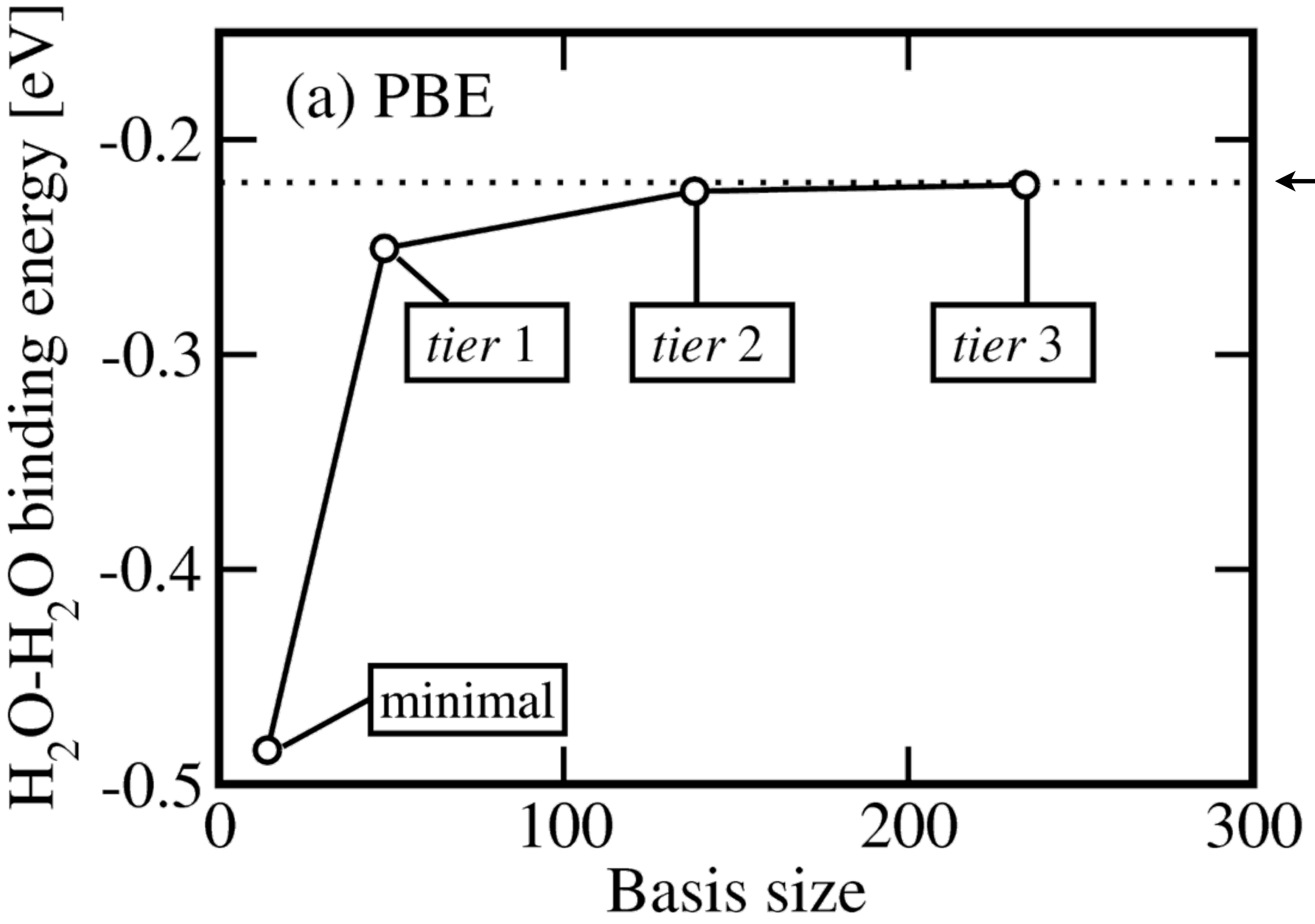
"Second tier"

"Third tier"

...

Transferability: $(H_2O)_2$ hydrogen bond energy ...





Basis set limit (independent):
$E_{Hb} = -219.8$ meV

... conformational energy hierarchy, large molecules ...

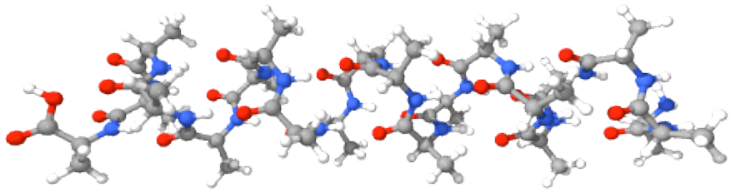
Ala$_{20}$ - fully extended structure

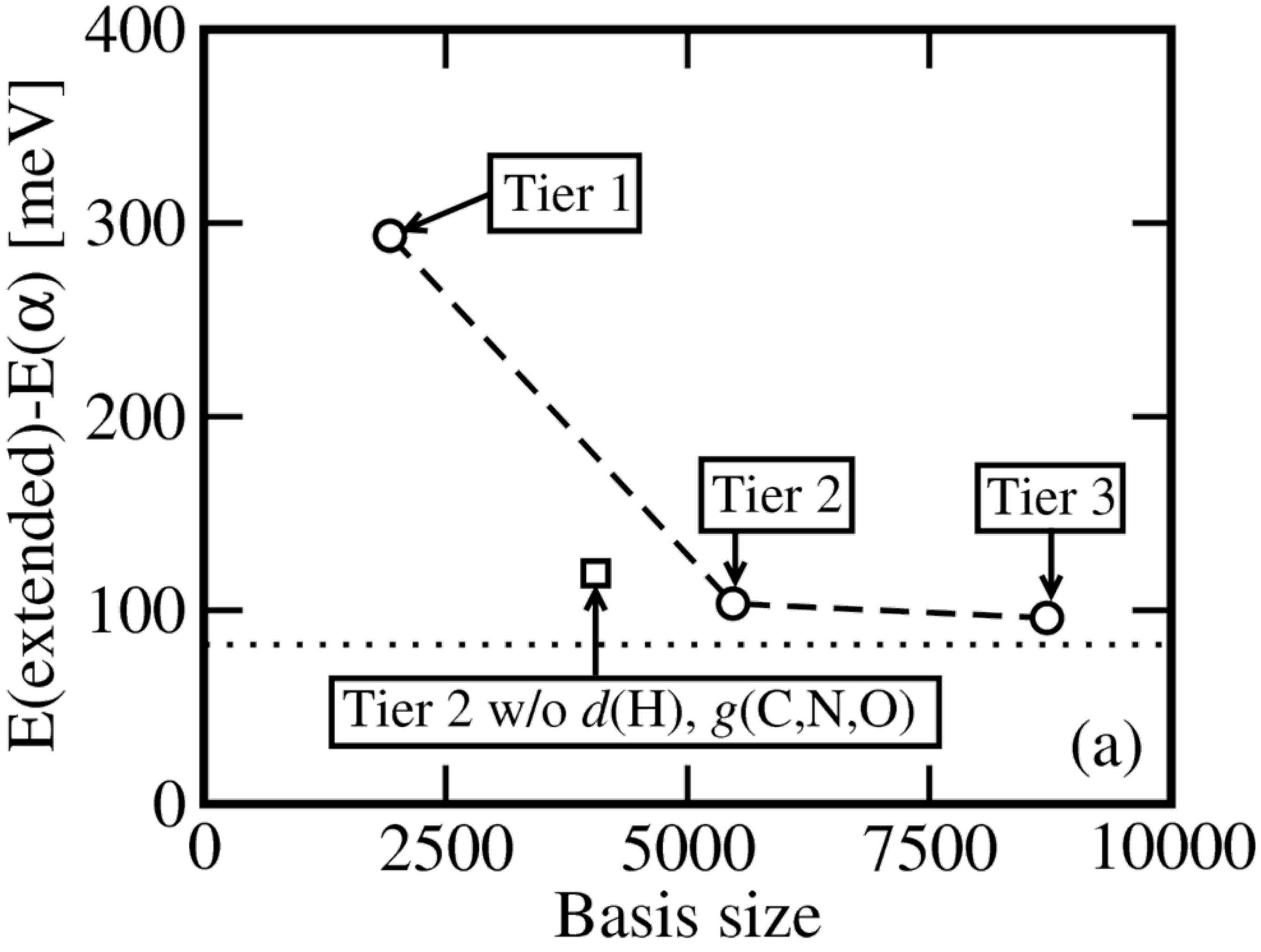
203 atoms

⟷

ΔE=0.08 eV

Ala$_{20}$ - α helix





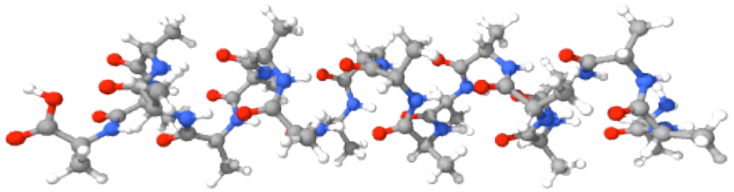
... conformational energy hierarchy, large molecules ...

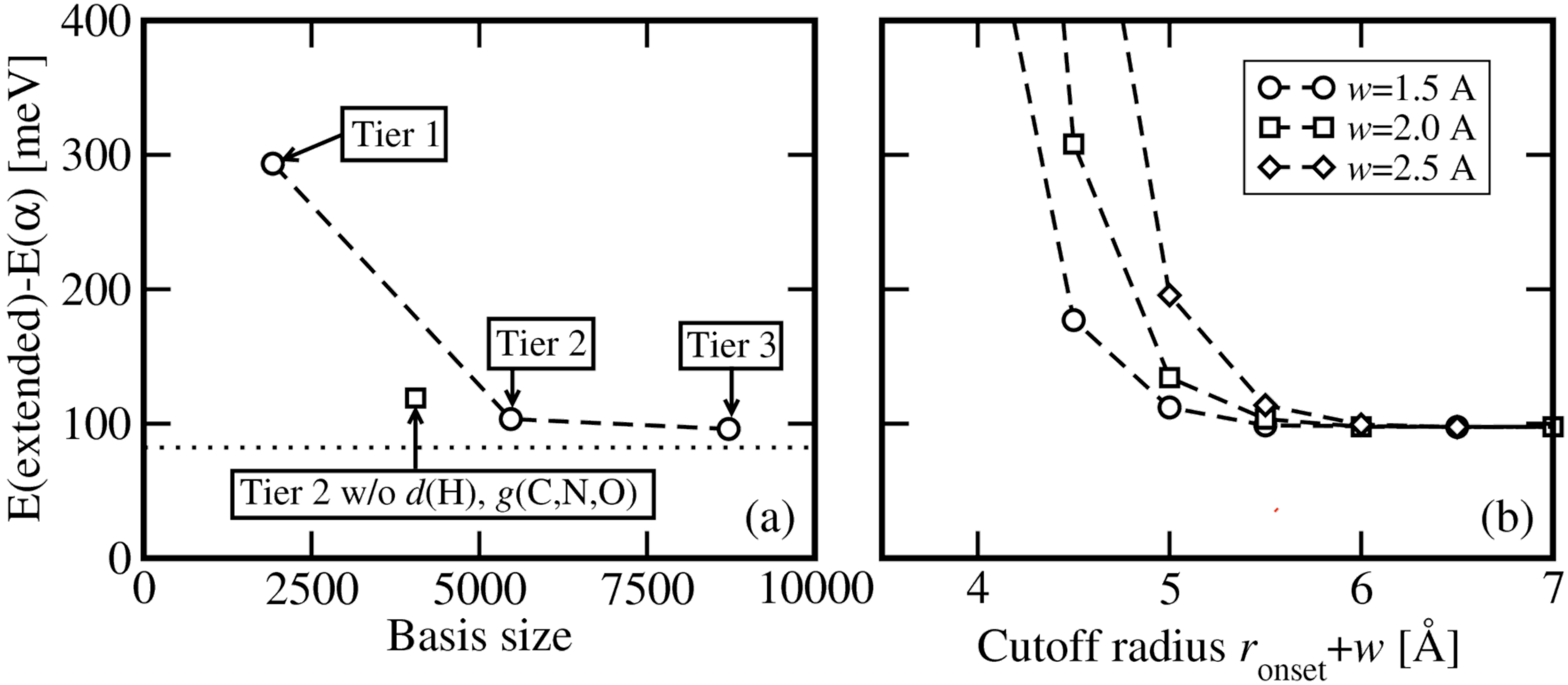
Ala$_{20}$ - fully extended structure

203 atoms ⟷

ΔE=0.08 eV

Ala$_{20}$ - α helix





... periodic hybrid functionals ...

Cohesive properties, bulk semiconductors

Si

PBE0	a [Å]	B_0 [Mbar]	E_{coh} [eV]
FHI-aims, *tight*	5.439	0.99	4.553
Ref. [1]	5.433	1.00	4.555
HSE06			
FHI-aims, *tight*	5.446	0.98	4.527
Ref. [2]	5.435	0.98	4.582

GaAs

HSE06			
FHI-aims, *tight*	5.695	0.71	3.150
Ref. [2]	5.687	0.71	3.149

Ge

HSE06			
FHI-aims, *tight*	5.700	0.71	3.761
Ref. [3]	5.703	0.73	n/a

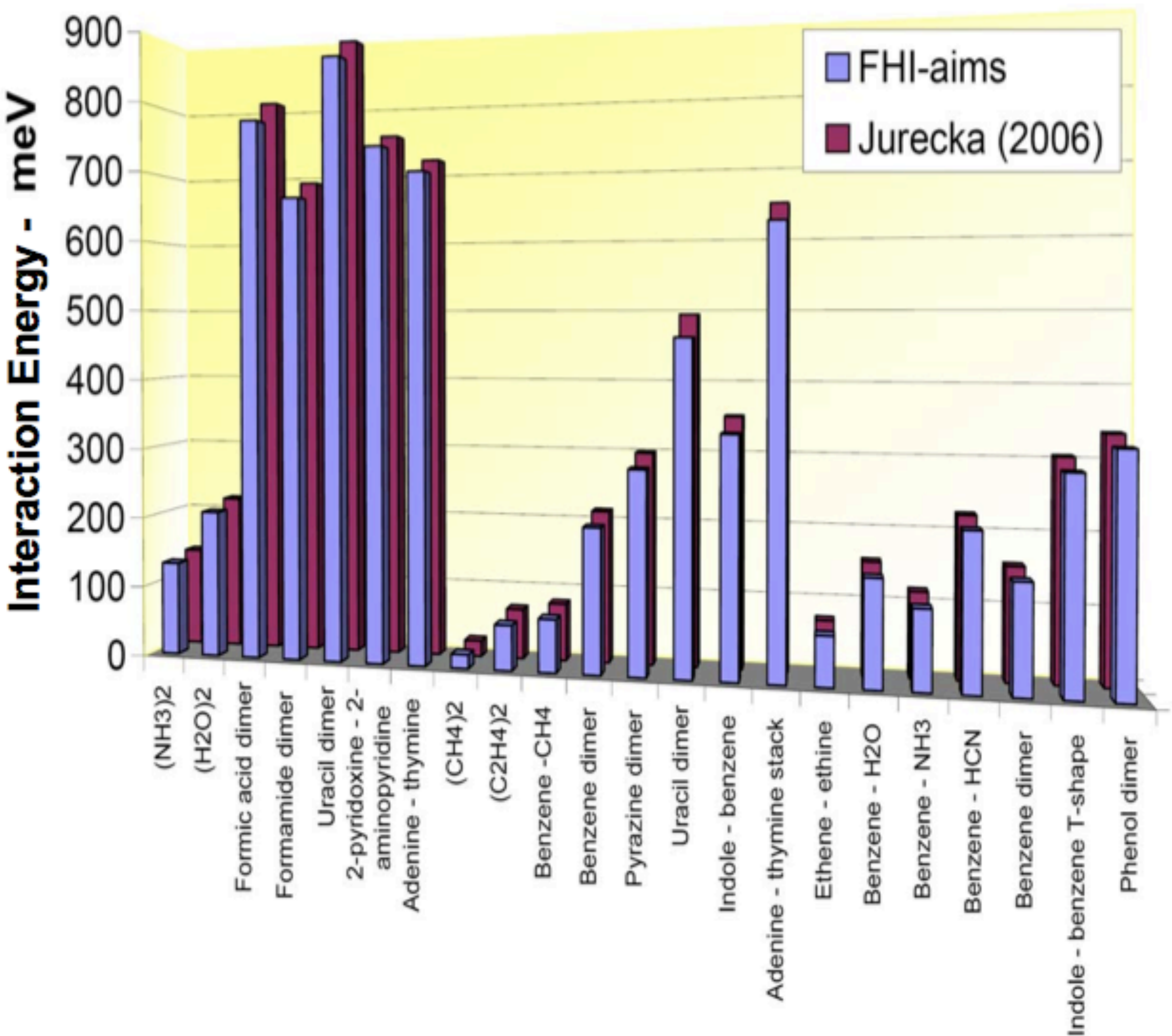
[1] J. Paier *et al.*, J. Chem. Phys. **124**, 154709 (2006).

[2] J. Paier *et al.*, J. Chem. Phys. **125**, 249901 (2006).

[3] A. Stroppa *et al.*, PRB **83**, 085201 (2011).

... and many-body perturbation theory - MP2, RPA, *GW*

*Ren, Rinke, Blum, Wieferink, Tkatchenko, Sanfilippo, Reuter, Scheffler, NJP **14**, 053020 (2012)*



NAO basis set:
CNO: min.+3*s*3*p*3*d*1*f*
H: min.+4*s*3*p*2*d*

Resolution of identity
for Coulomb operator

→ recover CBS limit
within 5%!

Perturbation theory:
Counterpoise correction
ESSENTIAL
for MP2, RPA

MP2: Complete basis set limit benchmark set "S22":
Jurecka et al., Phys. Chem. Chem. Phys 8, 1985 (2006)

So where are we at?

So where are we at?

- Sound, accurate basis sets, elements 1-102

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- “Properties”: Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- “Properties”: Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.
- LDA, GGA, van der Waals corrections, hybrid functionals, Hartree-Fock+MP2, RPA, *GW*, ...

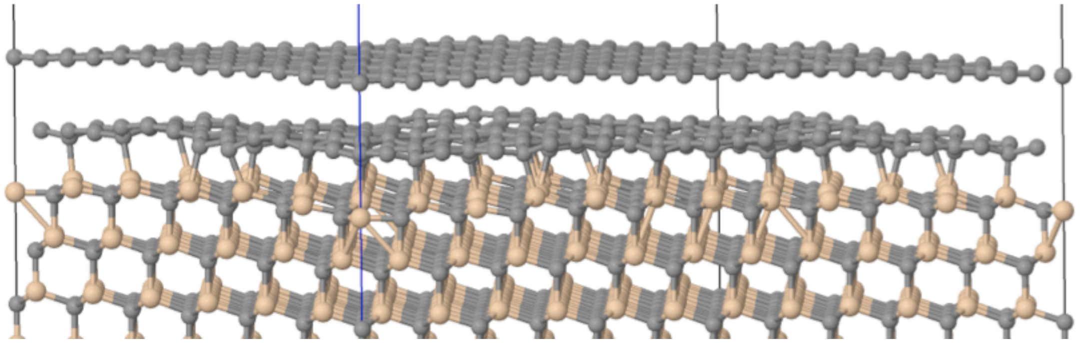
So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- “Properties”: Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.
- LDA, GGA, van der Waals corrections, hybrid functionals, Hartree-Fock+MP2, RPA, *GW*, ...
- Massively parallel - scalable eigensolver ELPA

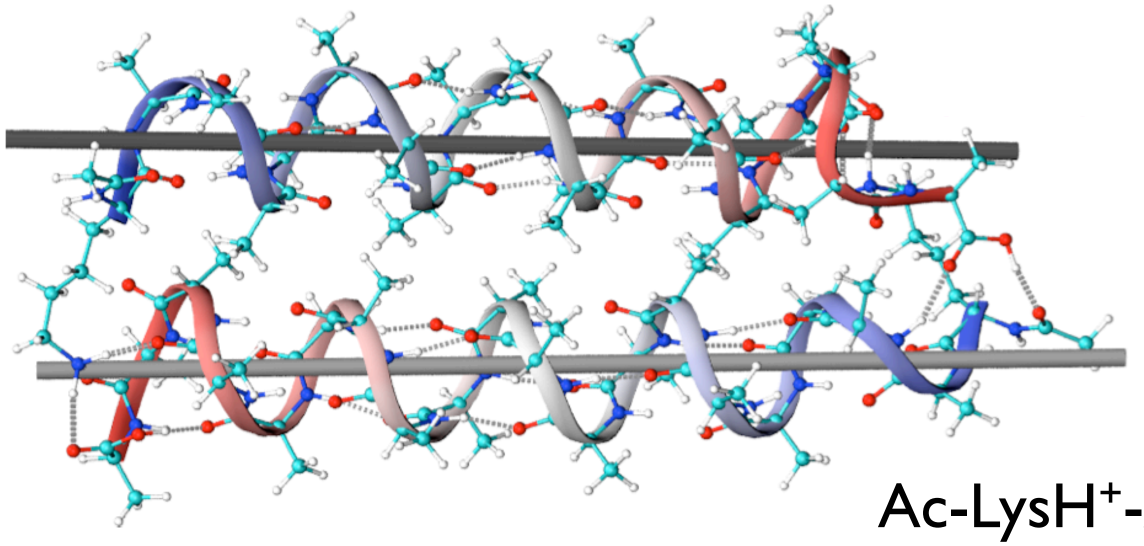
So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- “Properties”: Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.
- LDA, GGA, van der Waals corrections, hybrid functionals, Hartree-Fock+MP2, RPA, *GW*, ...
- Massively parallel - scalable eigensolver ELPA

Scalability: Real systems (Kohn-Sham DFT)



SiC(111)-(6√3x6√3) graphene-like monolayer
(216-338 atoms/layer ... want >2000 atoms total)

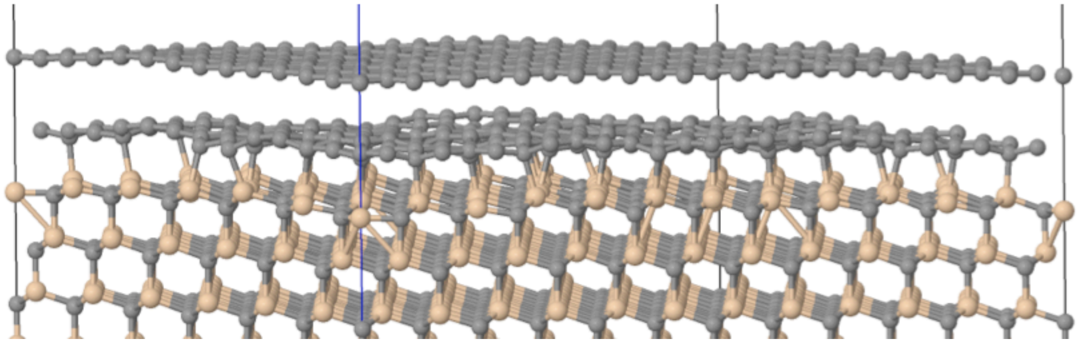


Ac-LysH$^+$-Ala$_{19}$
dimerized peptide

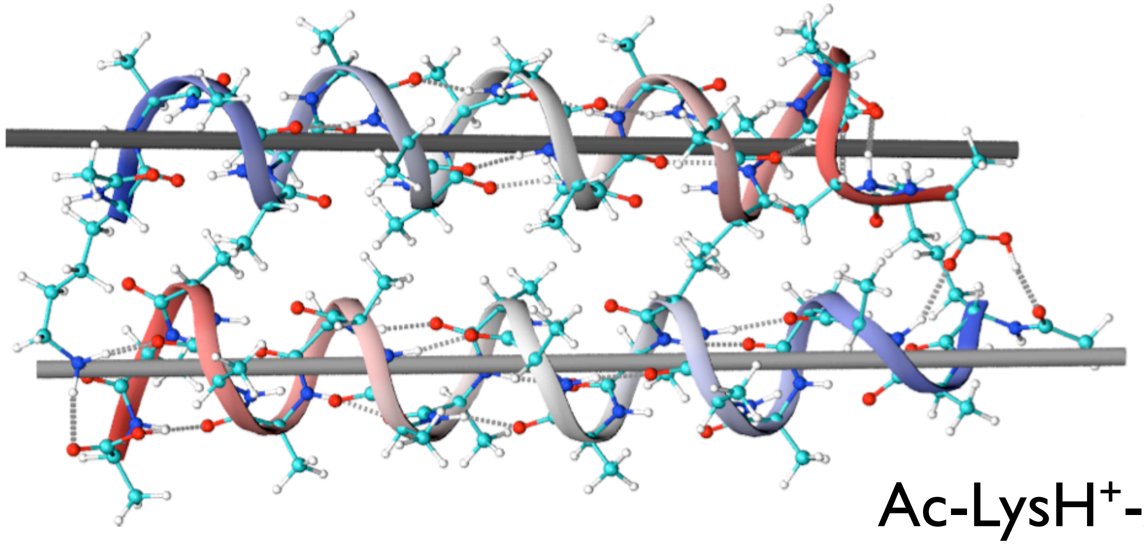
<u>What we would like to do routinely:</u>

- 1,000s of atoms, light or heavy
- 10-70 basis function
- Occupied eigenstates for Kohn-Sham DFT:
  1/3-1/6 of full basis size

Scalability: Real systems (Kohn-Sham DFT)



SiC(111)-(6√3x6√3) graphene-like monolayer
(216-338 atoms/layer ... want >2000 atoms total)



Ac-LysH$^+$-Ala$_{19}$
dimerized peptide

<u>What we would like to do routinely:</u>

- 1,000s of atoms, light or heavy
- 10-70 basis function
- Occupied eigenstates for Kohn-Sham DFT: 1/3-1/6 of full basis size

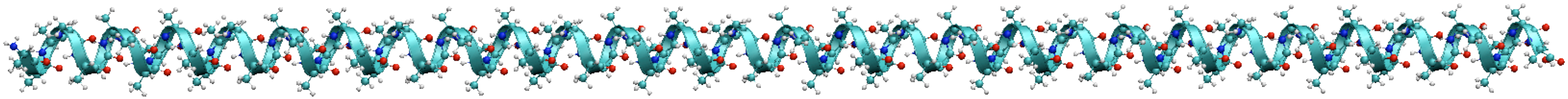
... and we have (now!) large computers:

IBM BlueGene (MPG, Garching)
16384 CPU cores

Where does the time go? (Kohn-Sham DFT)

IBM BlueGene (MPG, Garching)
16384 CPU cores

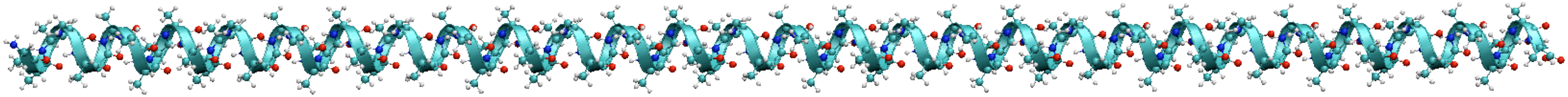
Where does the time go? (Kohn-Sham DFT)



α-helical Ala$_{100}$ (1000 atoms),
high accuracy, DFT-PBE

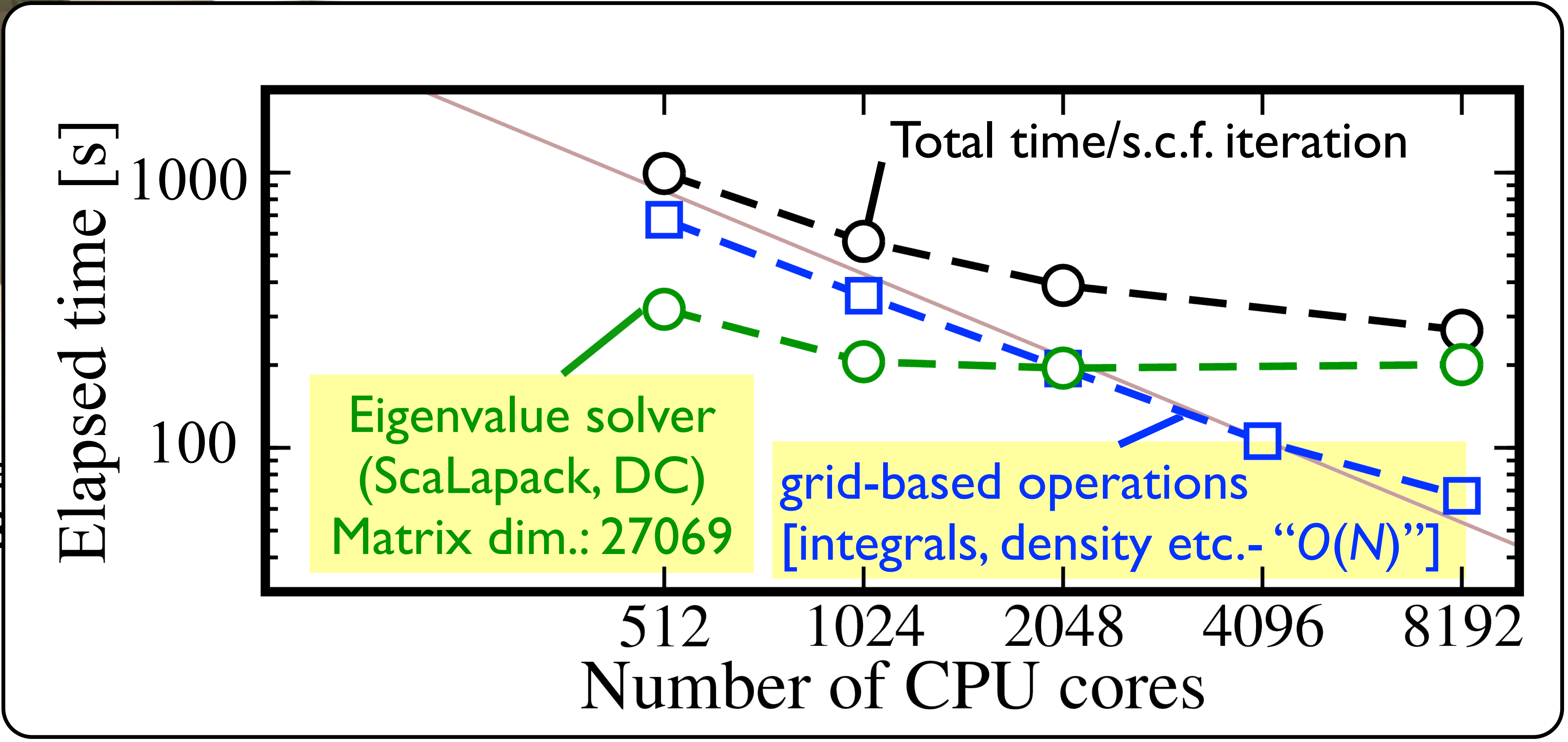
IBM BlueGene (MPG, Garching)
16384 CPU cores

Where does the time go? (Kohn-Sham DFT)



α-helical Ala_{100} (1000 atoms),
high accuracy, DFT-PBE

IBM BlueGe
1638



Taking apart the eigenproblem

$$\boxed{\underline{\underline{h}}\,\underline{c}_k = \epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

Generalized (non-orthogonal) eigenvalue problem:

- Transform to orthogonal form: $U^{-T}HU^{-1}$
- Transform orthogonal *H'* to *tridiagonal* form
- Solve *tridiagonal* eigenproblem
- Backtransform (1) solution to standard form
- Backtransform (2) standard to general form

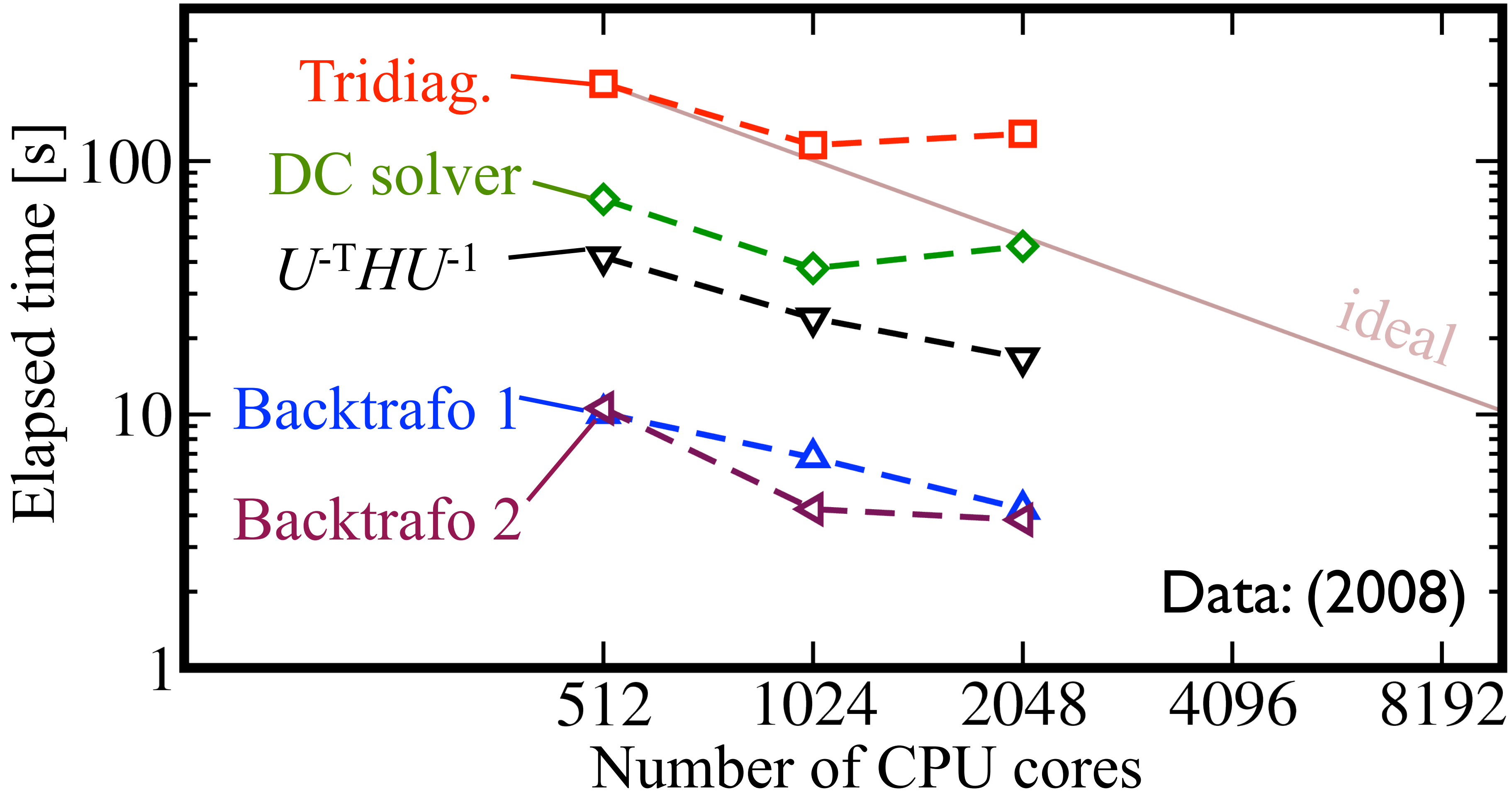
Data: (2008)

Taking apart the eigenproblem

$$\boxed{\underline{\underline{h}}\,\underline{c}_k = \epsilon_k\,\underline{\underline{s}}\,\underline{c}_k}$$

Generalized (non-orthogonal) eigenvalue problem:

- Transform to orthogonal form: $U^{-T}HU^{-1}$
- Transform orthogonal *H'* to *tridiagonal* form
- Solve *tridiagonal* eigenproblem
- Backtransform (1) solution to standard form
- Backtransform (2) standard to general form



α-helical Polyalanine Ala_{100}, BlueGene/P

Significant improvement: “ELPA” library

http://elpa.rzg.mpg.de
http://elpa-lib.fhi-berlin.mpg.de

$$\underline{\underline{h}}\,\underline{c}_k = \epsilon_k\,\underline{\underline{s}}\,\underline{c}_k$$

“EigensoLver for Petaflop Applications”

- “Drop-in enhancement” for ScaLapack solution (same layout)
- Rewrite of all communication, data handling etc. from scratch (retain only serial BLAS, Lapack)
- LGPL license - free to use with open or closed codes as long as modifications to ELPA library itself are open

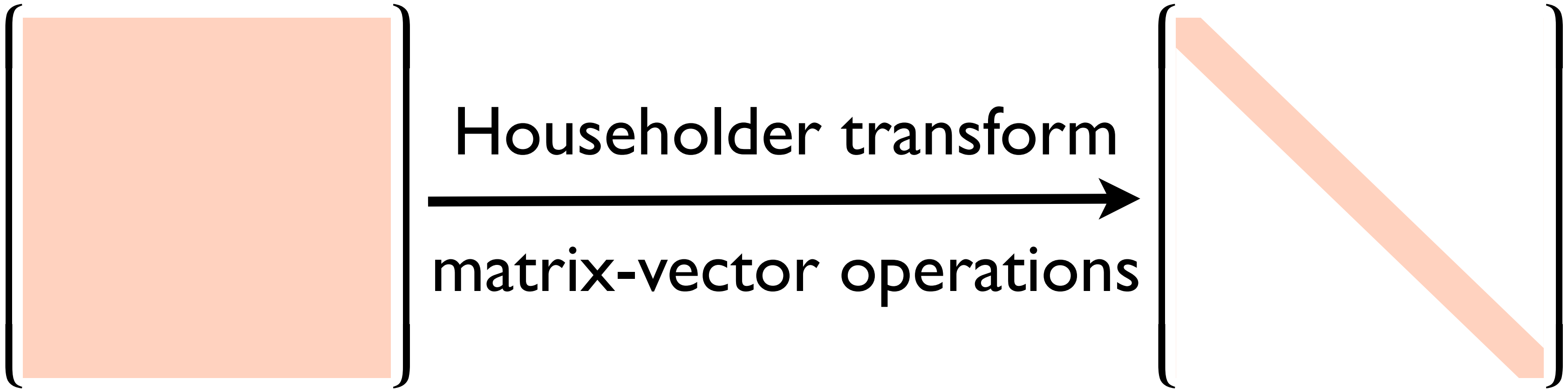
ELPA

German Ministry for Research and Education (BMBF) funded consortium, 2008-2011:
Garching Computing Center (Max Planck Society); Fritz Haber Institute;
Wuppertal University; Technical University Munich; MPI Mathematics in Science; IBM

Example: Reduction to tridiagonal form (just linear algebra)

Chief bottleneck: Tridiagonalization

“Conventional” reduction:



Householder transform

matrix-vector operations

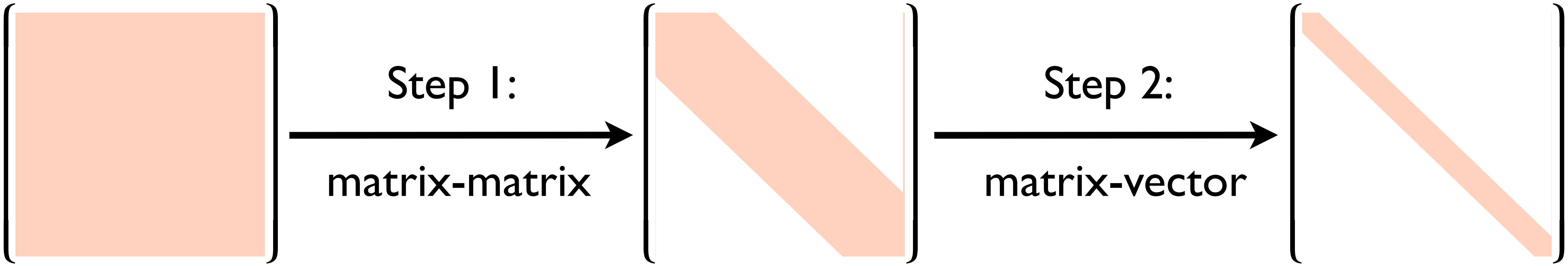
full matrix

tridiagonal matrix

Example: Reduction to tridiagonal form (just linear algebra)

Chief bottleneck: Tridiagonalization

"Two-step" reduction:
C. *Bischof, B. Lang, X. Sun, ACM Trans. Math. Software* ***26****, 581 (2000).*



Step 1:

matrix-matrix

Step 2:

matrix-vector

full matrix

band matrix

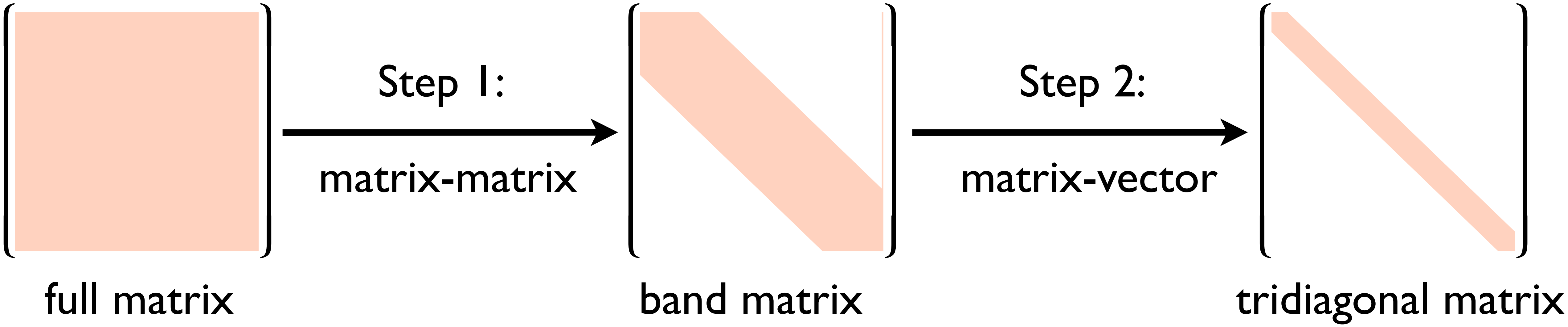
tridiagonal matrix

Auckenthaler, Blum, Bungartz, Huckle, Johanni, Krämer, Lang, Lederer, Willems,
Parallel Computing ***37****, 783-794 (2011).*

Example: Reduction to tridiagonal form (just linear algebra)

Chief bottleneck: Tridiagonalization

"Two-step" reduction:
C. *Bischof, B. Lang, X. Sun, ACM Trans. Math. Software* ***26****, 581 (2000).*



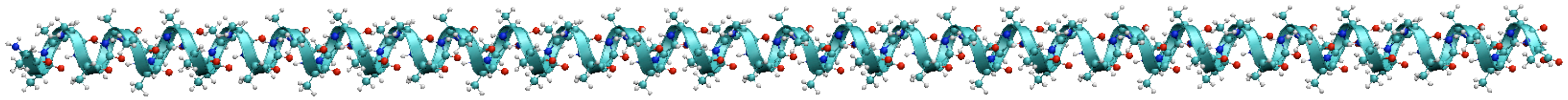
Need two eigenvector backtransformation steps instead of one

- Heavily optimized backtransform for eigenvectors (adaptive data layout, architecture-specific linear algebra kernels) to offset overhead

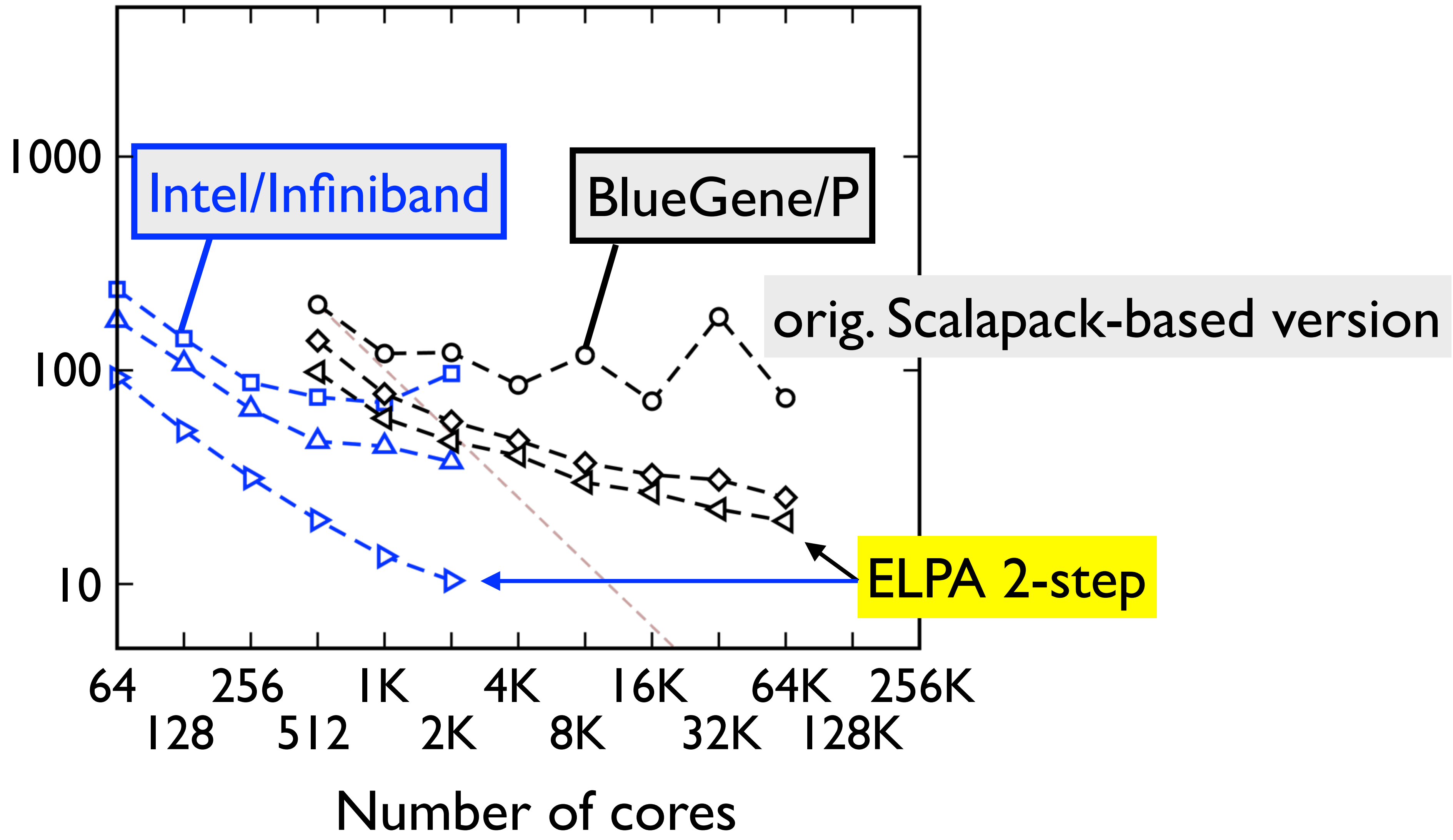
Auckenthaler, Blum, Bungartz, Huckle, Johanni, Krämer, Lang, Lederer, Willems,
Parallel Computing ***37****, 783-794 (2011).*

Our experience: Significant improvement

Open source (LGPL): http://elpa.rzg.mpg.de



α-helical Polyalanine Ala$_{100}$
Matrix: 27069, *States:* 3410



Intel/Infiniband

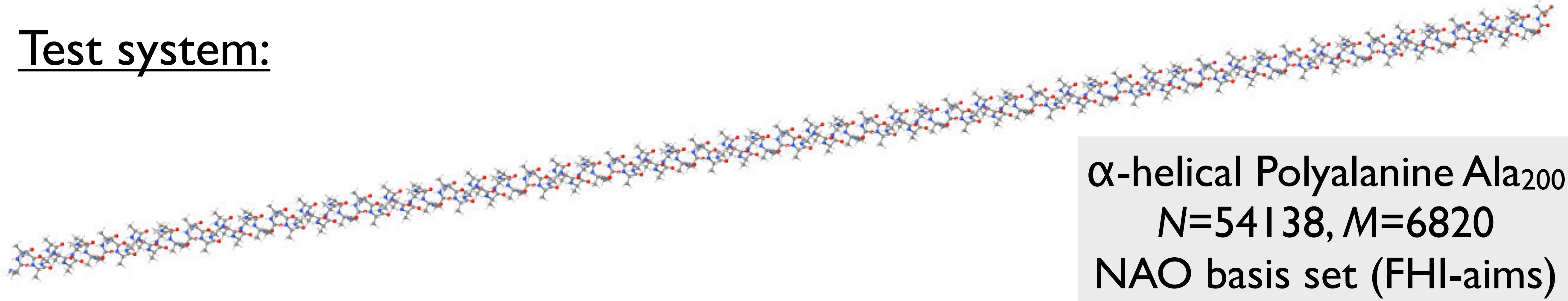
BlueGene/P

orig. Scalapack-based version

ELPA 2-step

... so what about that "petascale"?

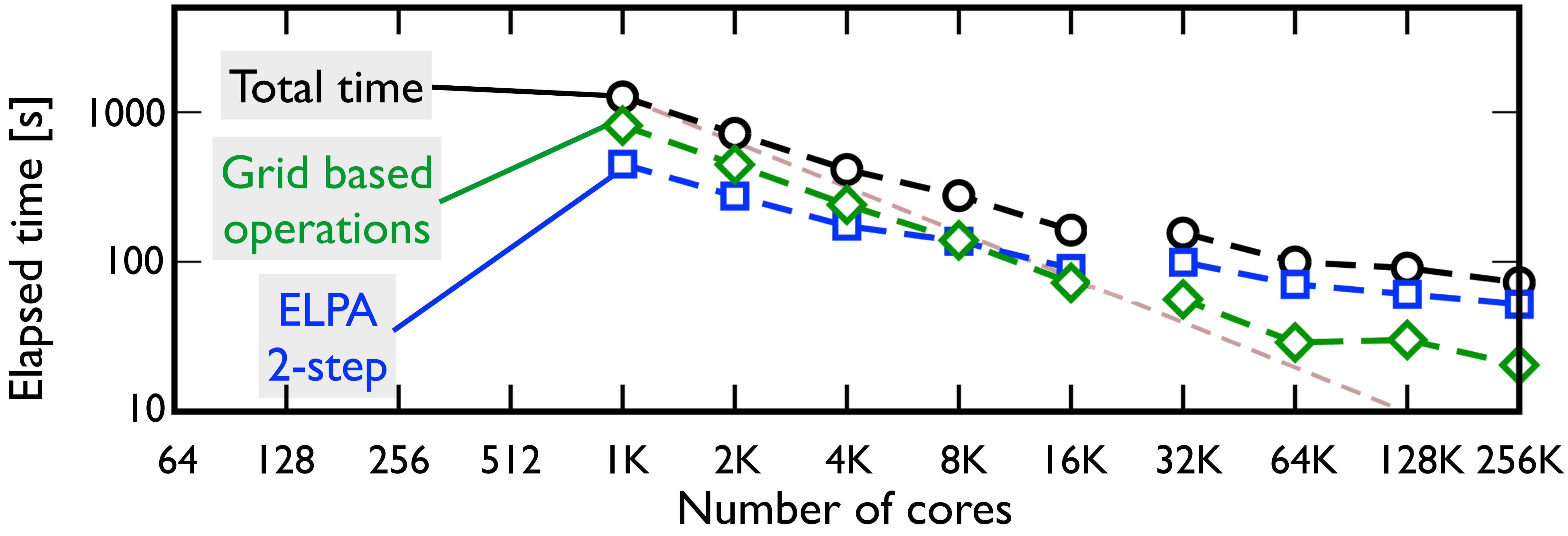
Test system:



α-helical Polyalanine Ala$_{200}$
N=54138, M=6820
NAO basis set (FHI-aims)

Time per FULL DFT-PBE s.c.f. iteration: BlueGene/P

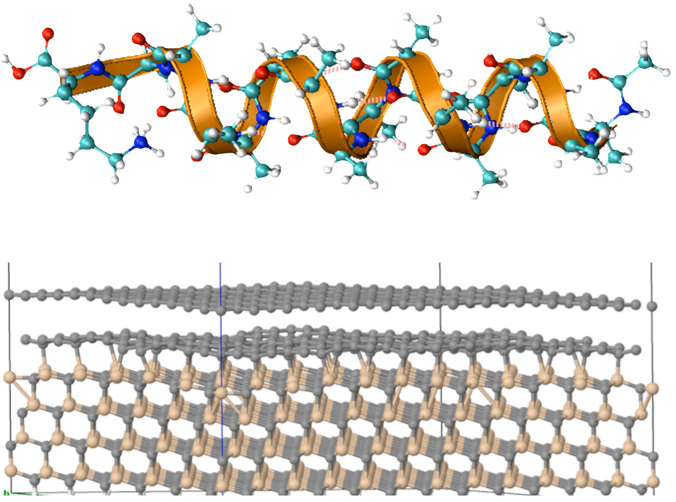
Timings: A. Marek, R. Johanni, Rechenzentrum Garching



Outline

An approach to all-electron "density functional theory and beyond": FHI-aims [1]

- Numeric atom-centered (localized) basis sets
- Scalability (1,000s of atoms, 1(0),000s of CPUs)
- Pushing towards "better" functionals (→P. Rinke)



Where do "we" come from?

- (Bio)molecular structure and spectroscopy
- Nanostructured inorganic surfaces

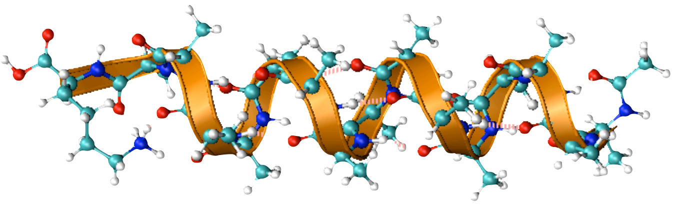
... and some challenges (towards WDM)

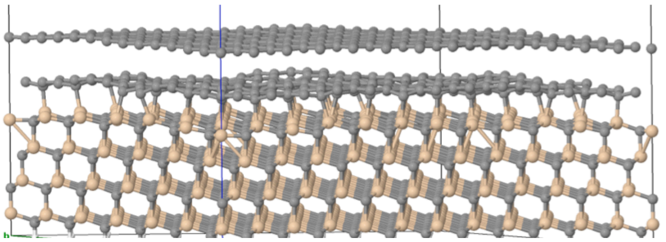
[1] The Fritz Haber Institute ab initio molecular simulations suite (FHI-aims)
V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
Computer Physics Communications ***180****, 2175-2196 (2009) - http://www.fhi-berlin.mpg.de/aims/*

Outline

An approach to all-electron "density functional theory and beyond": FHI-aims [1]

- Numeric atom-centered (localized) basis sets
- Scalability (1,000s of atoms, 1(0),000s of CPUs)
- Pushing towards "better" functionals (→ P. Rinke)





Where do "we" come from?

- (Bio)molecular structure and spectroscopy
- Nanostructured inorganic surfaces

... and some challenges (towards WDM)

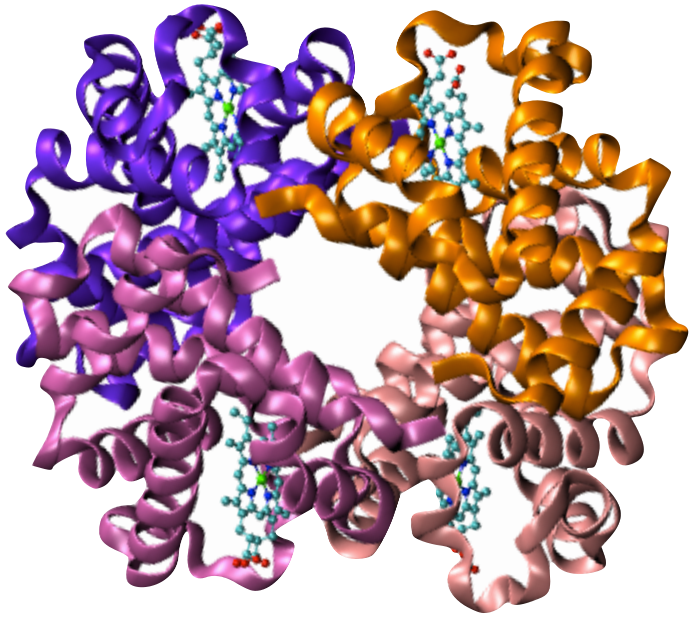
[1] The Fritz Haber Institute ab initio molecular simulations suite (FHI-aims)
V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler,
*Computer Physics Communications **180**, 2175-2196 (2009) - http://www.fhi-berlin.mpg.de/aims/*

Cold, not so dense world: Biomolecules

- Proteins: macromolecules that perform essential tasks inside living organisms
- ~60.000 different proteins in human organism, several billion per cell.
- *<u>The structure of a protein determines its function!</u>*

  This structure is not random, but tuned by the amino acid sequence.

Hemoglobin



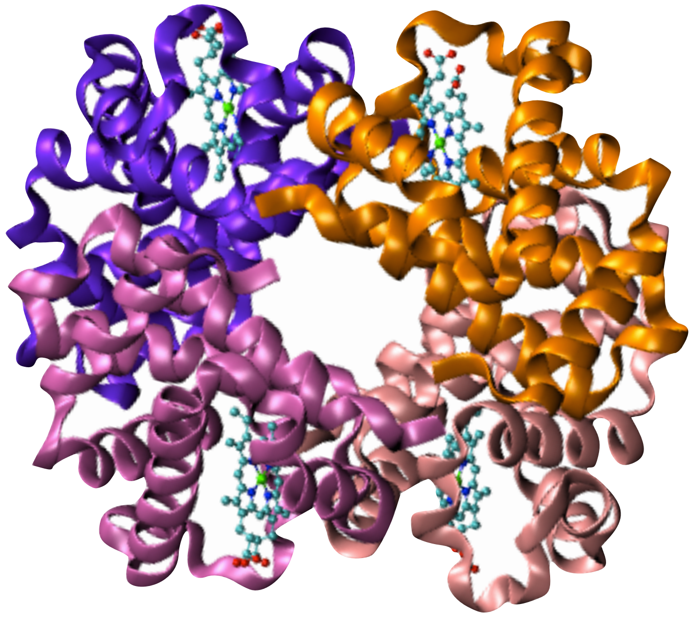
Cold, not so dense world: Biomolecules

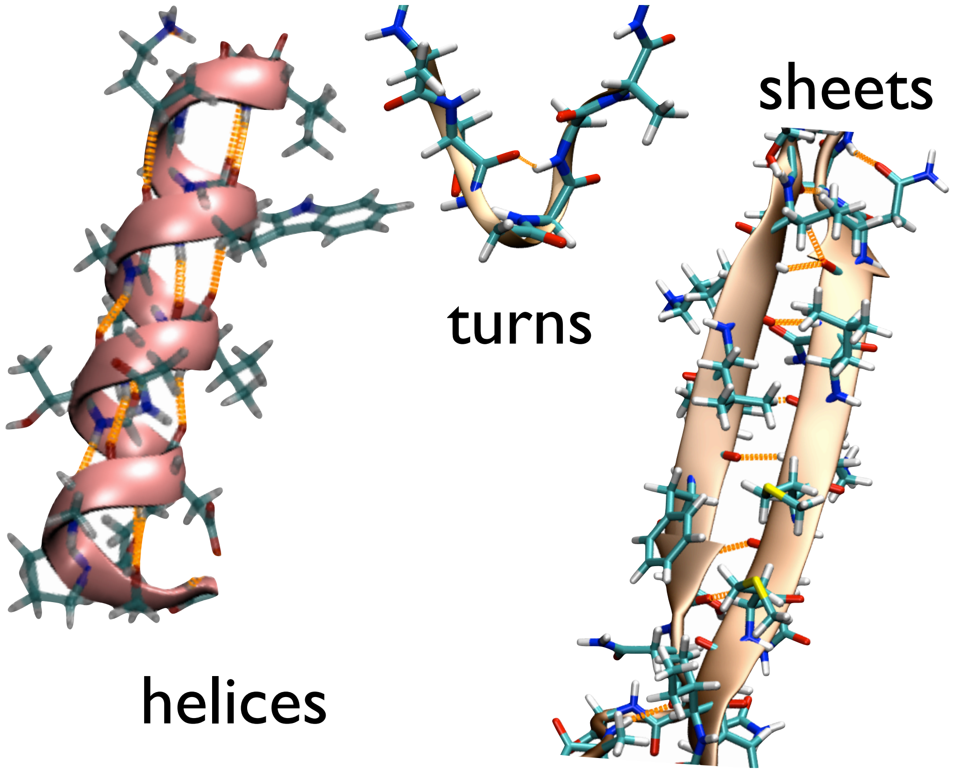
- Proteins: macromolecules that perform essential tasks inside living organisms
- ~60.000 different proteins in human organism, several billion per cell.
- *The structure of a protein determines its function!*

This structure is not random, but tuned by the amino acid sequence.

Typical "building blocks:"
Secondary structure

Hemoglobin





Cold, not so dense world: Biomolecules

- Proteins: macromolecules that perform essential tasks inside living organisms
- ~60.000 different proteins in human organism, several billion per cell.
- *The structure of a p*

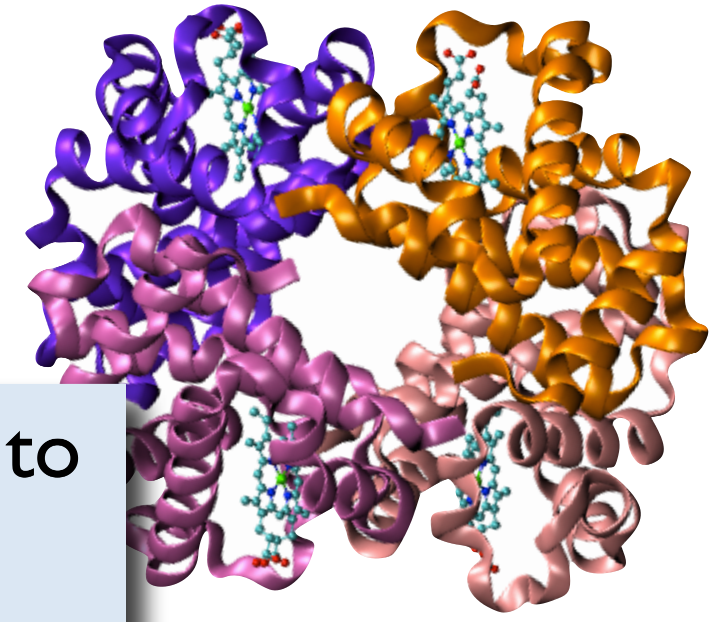
  function!

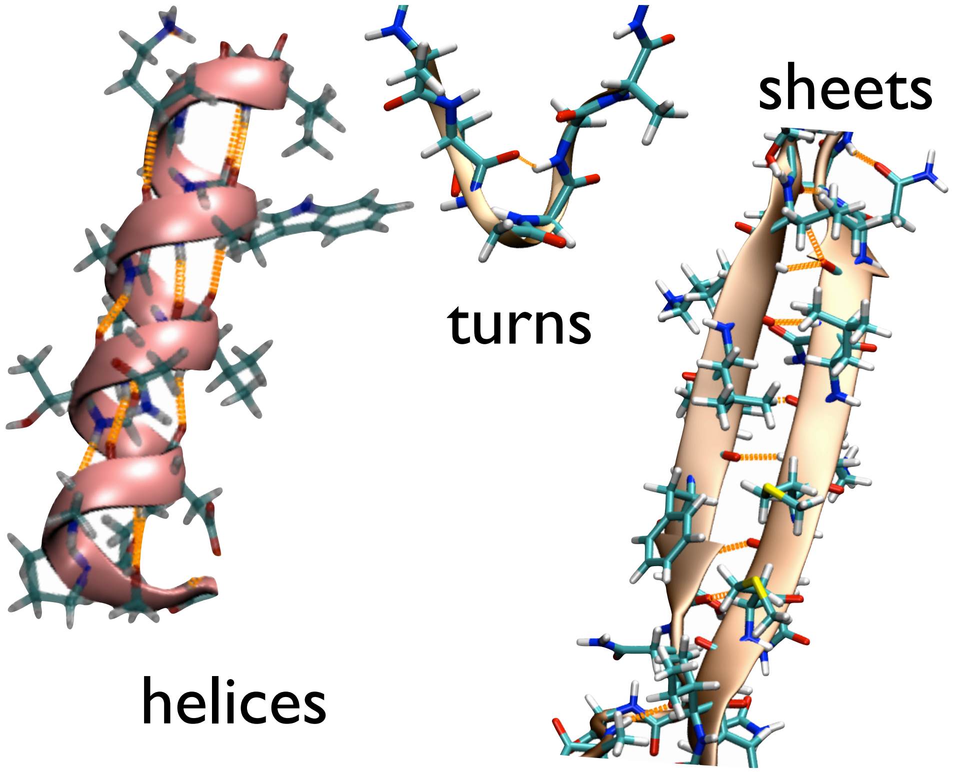
  This structure is not random, but tuned by the amino acid sequence.

  Typical "building blocks:"
  Secondary structure

Can we push "first principles" to predict secondary structure? (~100s of atoms)

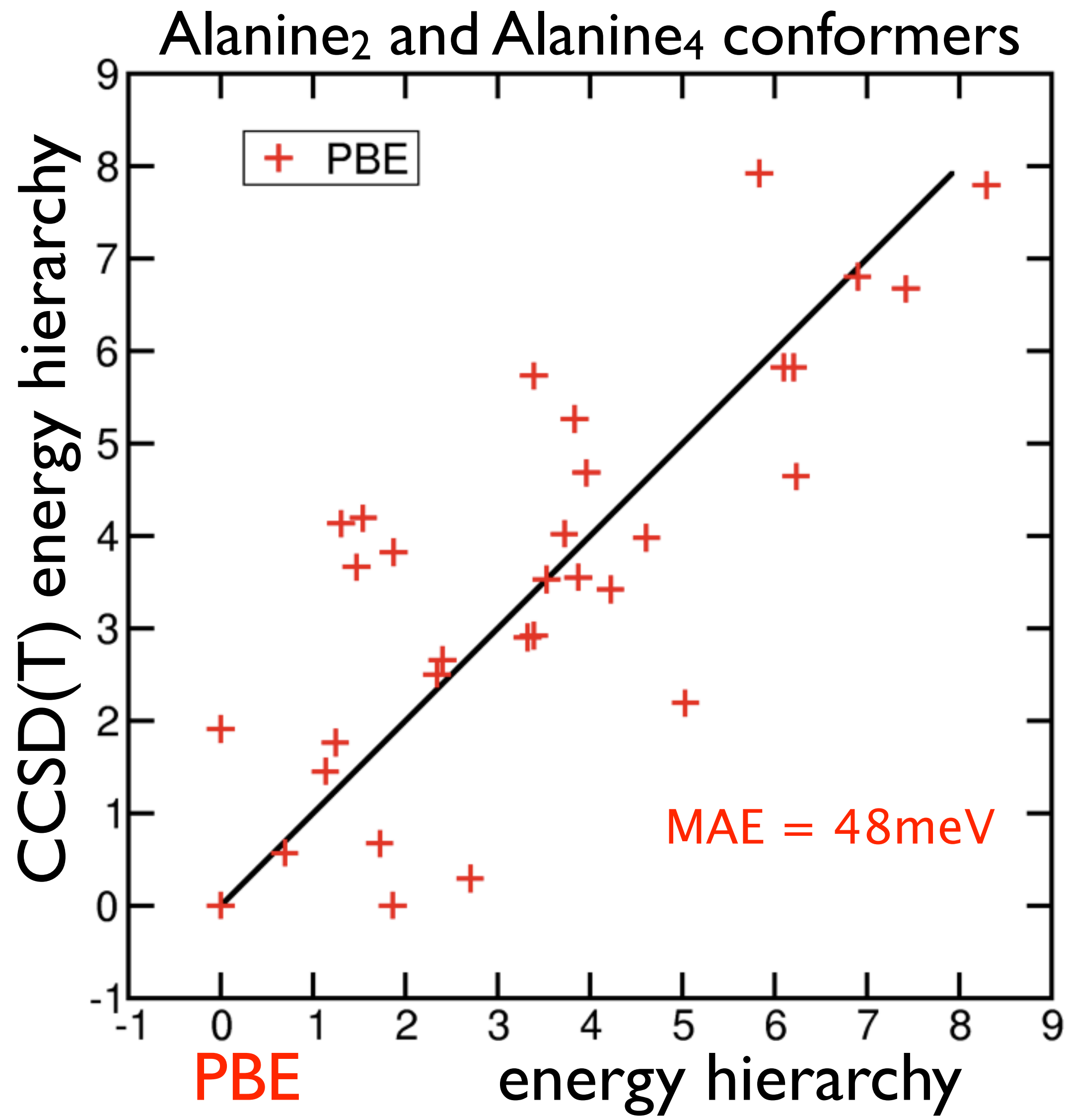
Hemoglobin





The first ingredient is weak: van der Waals

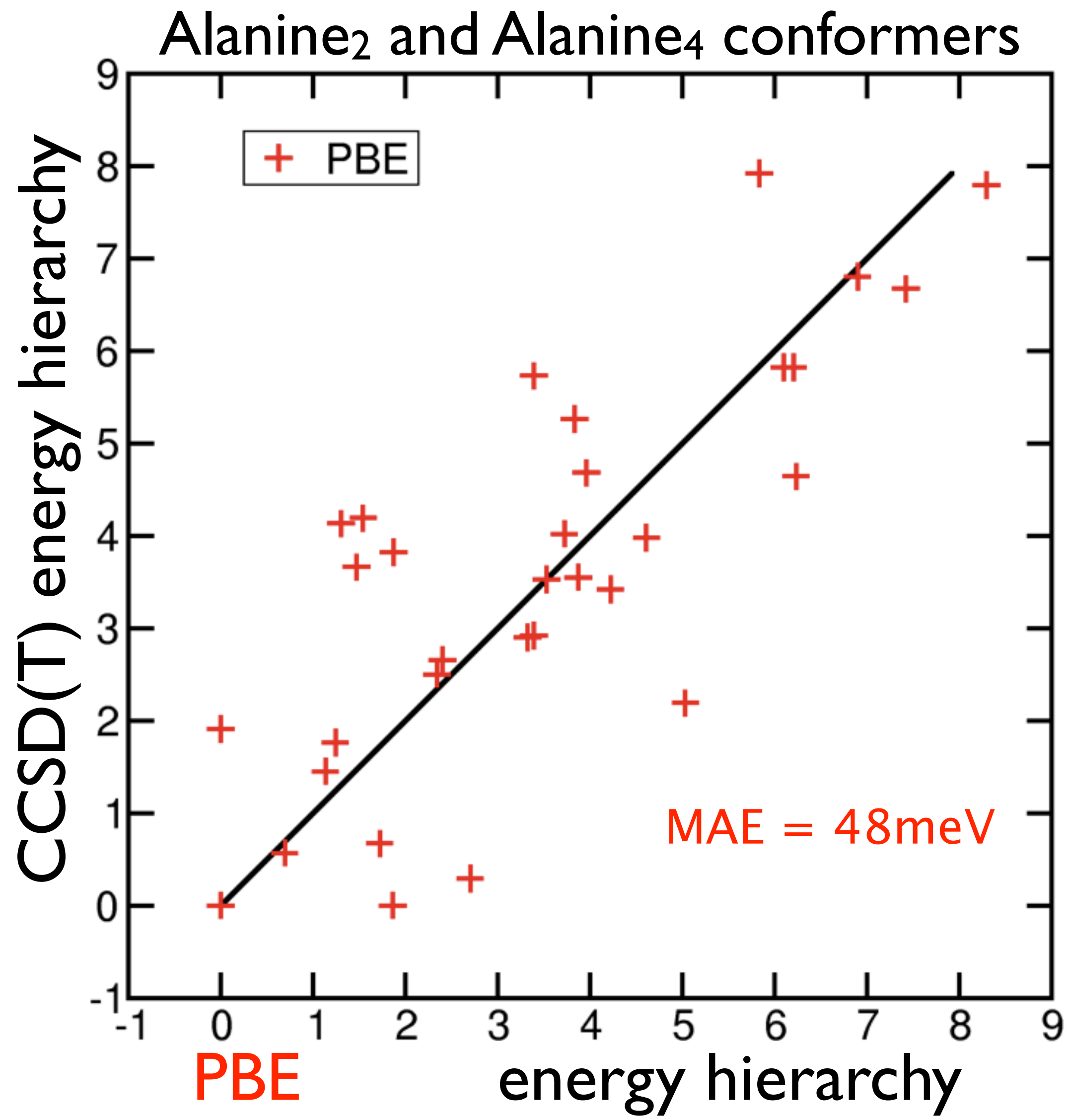
- DFT in the Perdew-Burke-Ernzerhof (1996) generalized gradient approximation
- Non-empirical, widely used, but *any* GGA lacks van der Waals tails



The first ingredient is weak: van der Waals

- DFT in the Perdew-Burke-Ernzerhof (1996) generalized gradient approximation
- Non-empirical, widely used, but *any* GGA lacks van der Waals tails
- Inclusion of van der Waals (vdW) dispersion via a $C_6[n]/R^6$ term: *(Yang, Grimme, Hobza, many others)*

$$E_{\mathrm{DFT}} + \sum_{\mathrm{atoms}\ ij} \frac{C_6^{ij}}{R_{ij}^6} \cdot f_{\mathrm{damp}}(R_{ij})$$



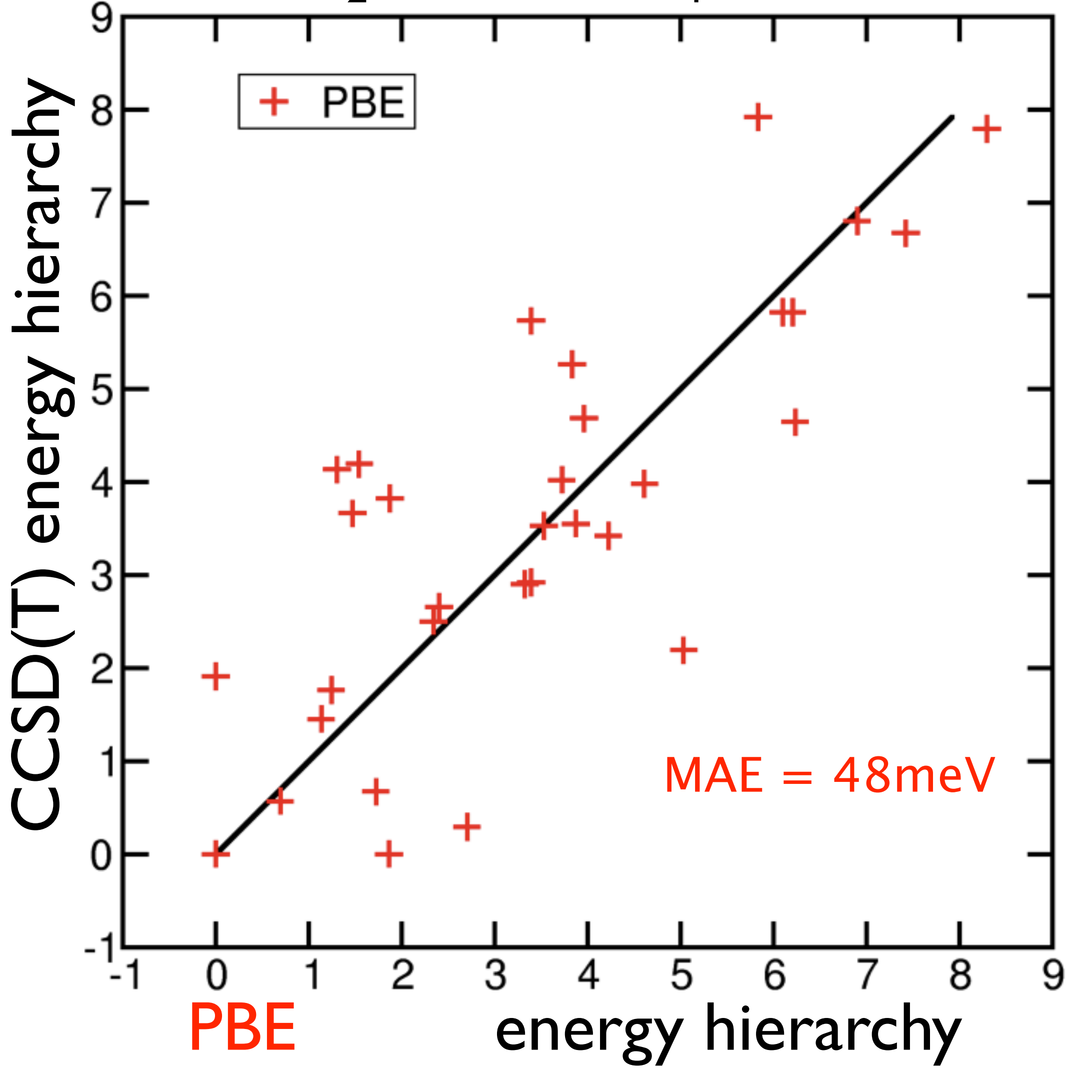
The first ingredient is weak: van der Waals

- DFT in the Perdew-Burke-Ernzerhof (1996) generalized gradient approximation
- Non-empirical, widely used, but *any* GGA lacks van der Waals tails
- Inclusion of van der Waals (vdW) dispersion via a $C_6[n]/R^6$ term:
  (Yang, Grimme, Hobza, many others)

$$E_{\mathrm{DFT}} + \sum_{\mathrm{atoms}\ ij} \frac{C_6^{ij}}{R_{ij}^6} \cdot f_{\mathrm{damp}}(R_{ij})$$

- Here: $C_6[n]$ term derived from self consistent electron density [1]

Alanine$_2$ and Alanine$_4$ conformers



[1] A. Tkatchenko and M. Scheffler, PRL ***102****, 073005 (2009)*

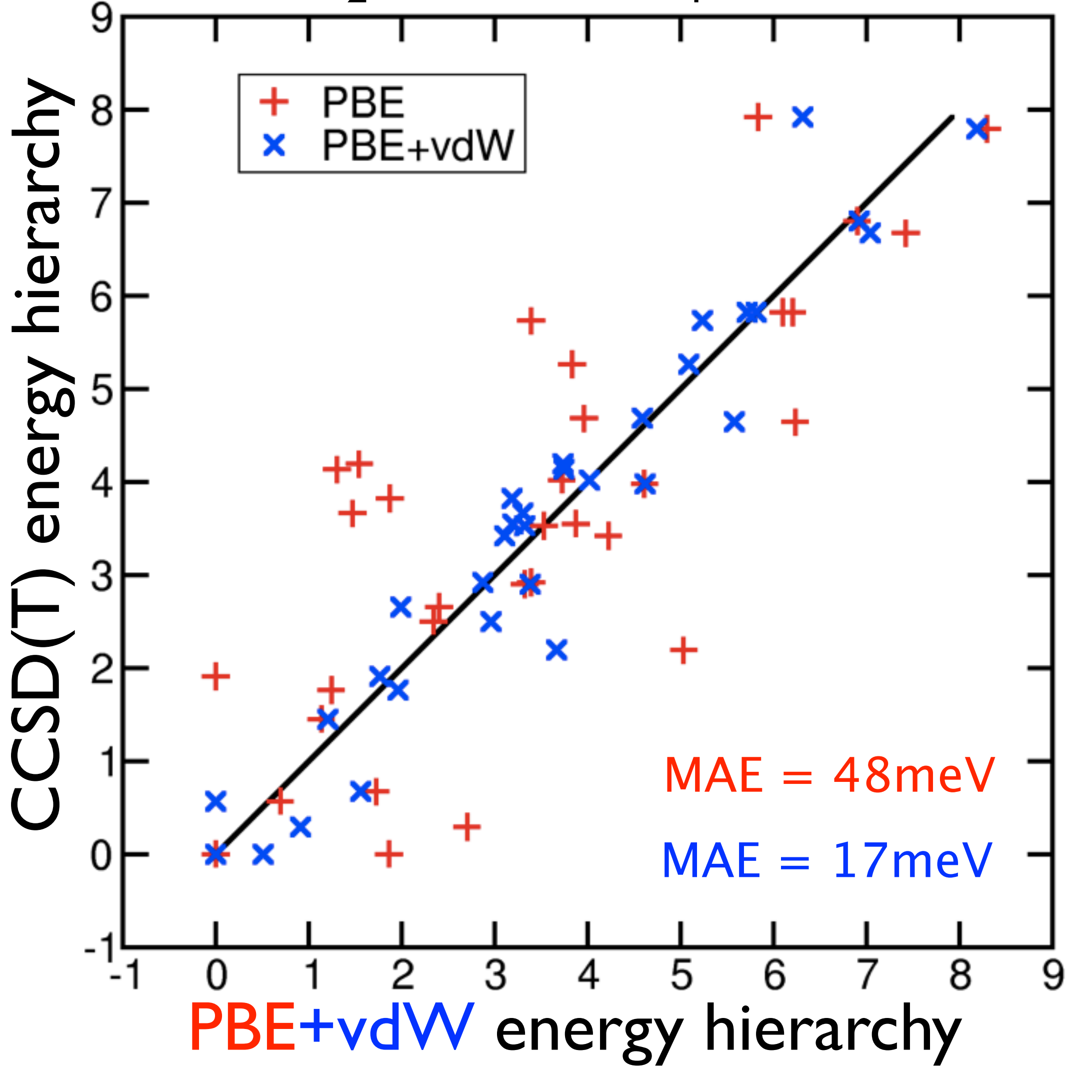
The first ingredient is weak: van der Waals

- DFT in the Perdew-Burke-Ernzerhof (1996) generalized gradient approximation
- Non-empirical, widely used, but *any* GGA lacks van der Waals tails
- Inclusion of van der Waals (vdW) dispersion via a $C_6[n]/R^6$ term:
  (Yang, Grimme, Hobza, many others)

$$E_{\rm DFT} + \sum_{\text{atoms } ij} \frac{C_6^{ij}}{R_{ij}^6} \cdot f_{\rm damp}(R_{ij})$$

- Here: $C_6[n]$ term derived from self consistent electron density [1]

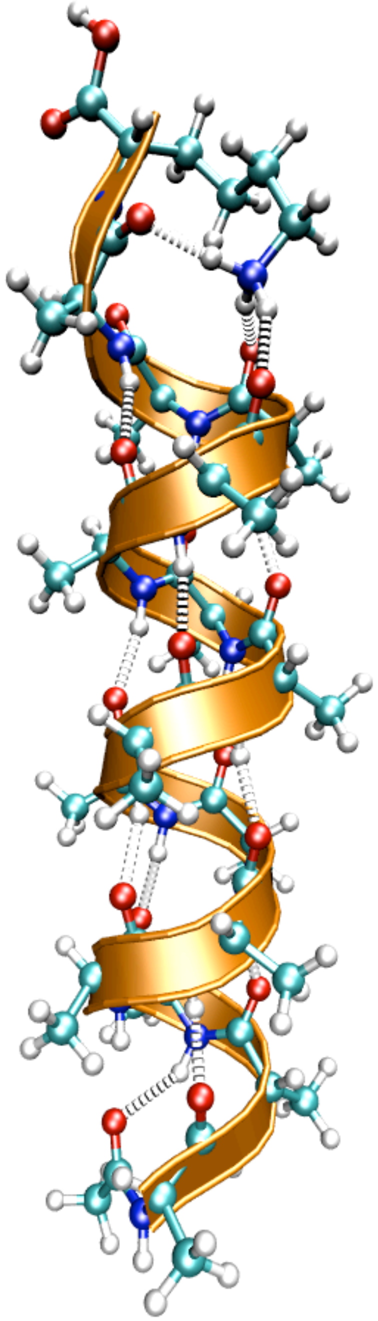
Alanine$_2$ and Alanine$_4$ conformers



[1] A. Tkatchenko and M. Scheffler, PRL ***102****, 073005 (2009)*

A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

Rossi, Blum, Kupser, von Helden, Bierau, Pagel, Meijer, Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*



α-helical Ac-Ala$_{15}$-LysH$^+$
(180 atoms): Helical?

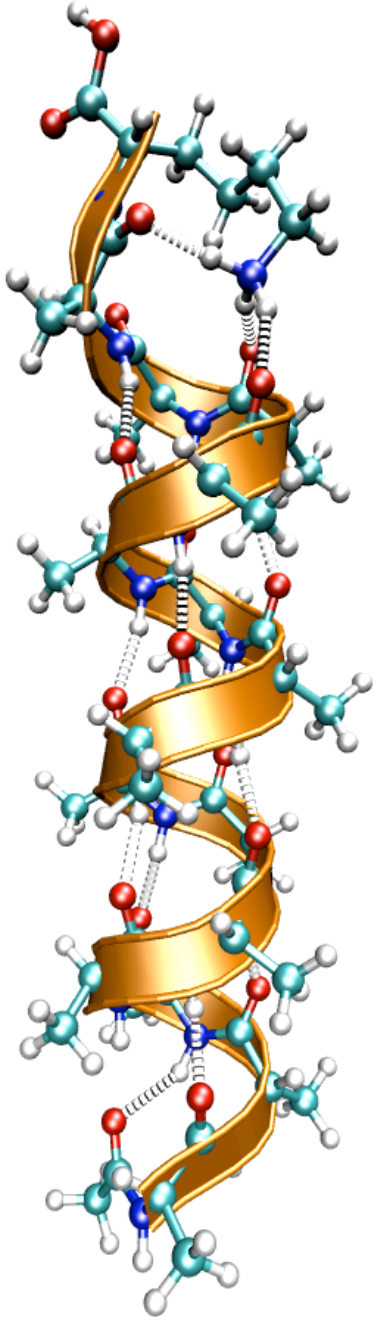
Experiment:
von Helden, Kupser, Bierau, Meijer,
Molecular Physics, FHI Berlin

Infrared multiphoton dissociation spectroscopy, FELIX free electron laser

Room temperature

A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

Rossi, Blum, Kupser, von Helden, Bierau, Pagel, Meijer, Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*



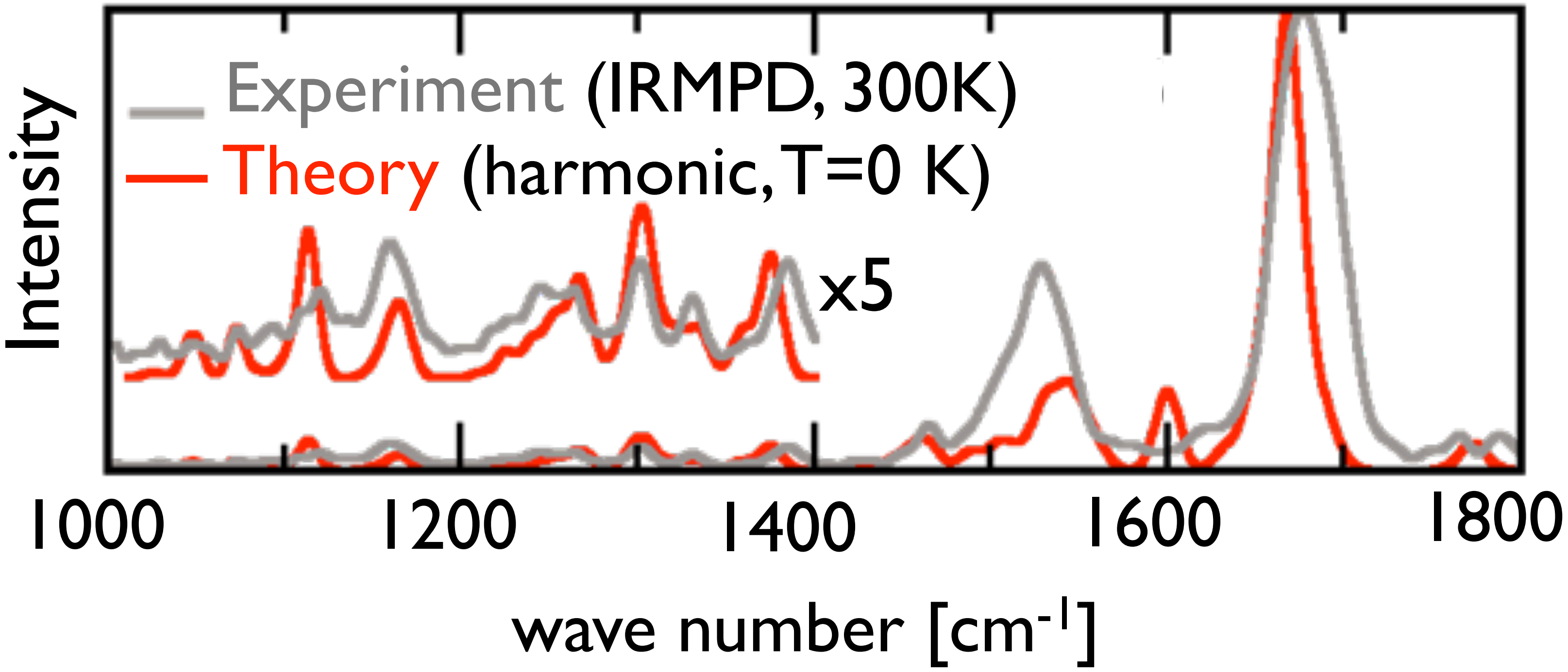
α-helical Ac-Ala$_{15}$-LysH$^+$
(180 atoms): Helical?

<u>Experiment</u>:
von Helden, Kupser, Bierau, Meijer,
Molecular Physics, FHI Berlin

Infrared multiphoton dissociation spectroscopy, FELIX free electron laser

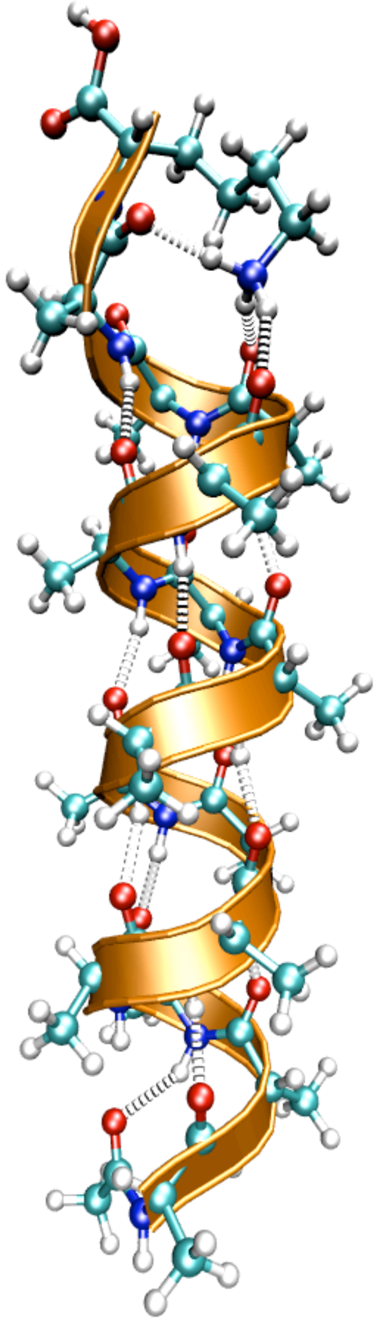
Room temperature

Theory: DFT-PBE+vdW; shifted, not scaled



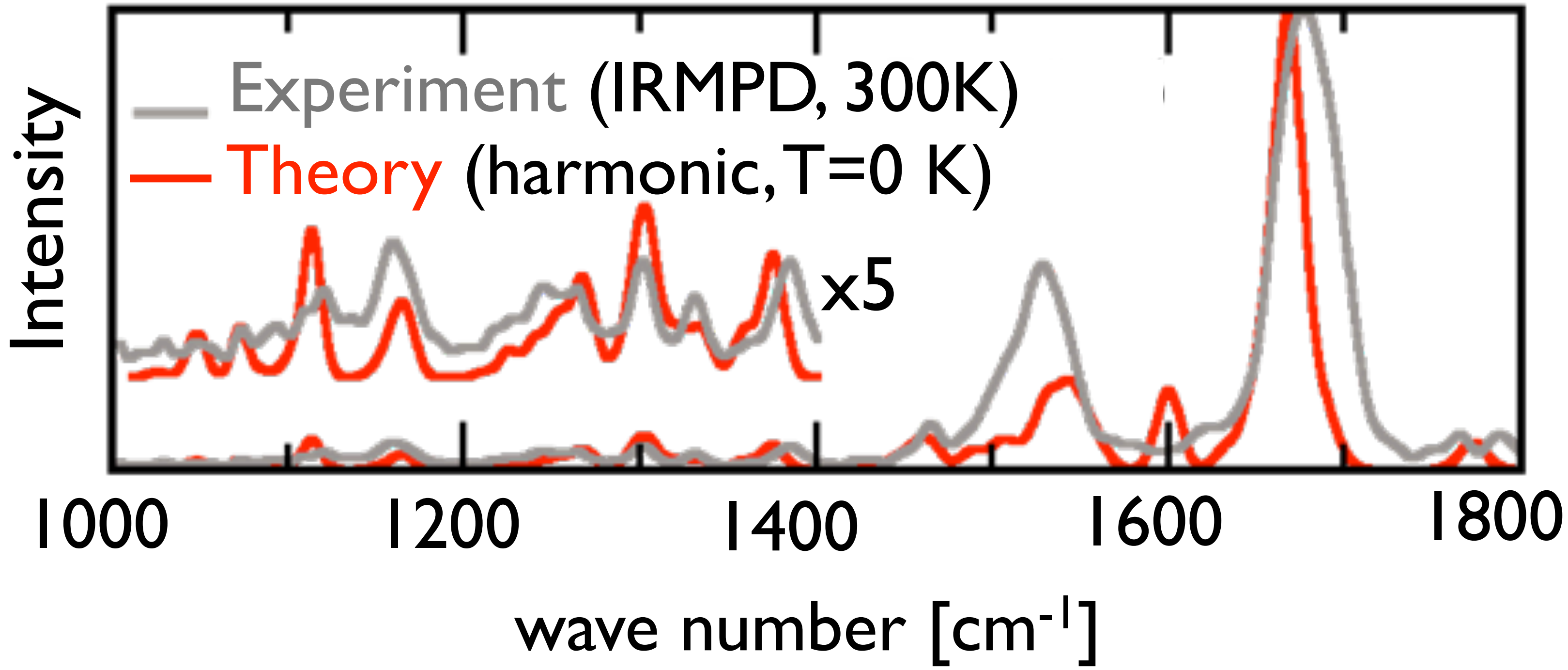
A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

Rossi, Blum, Kupser, von Helden, Bierau, Pagel, Meijer, Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*



α-helical Ac-Ala$_{15}$-LysH$^+$
(180 atoms): Helical?

Theory: DFT-PBE+vdW; shifted, not scaled

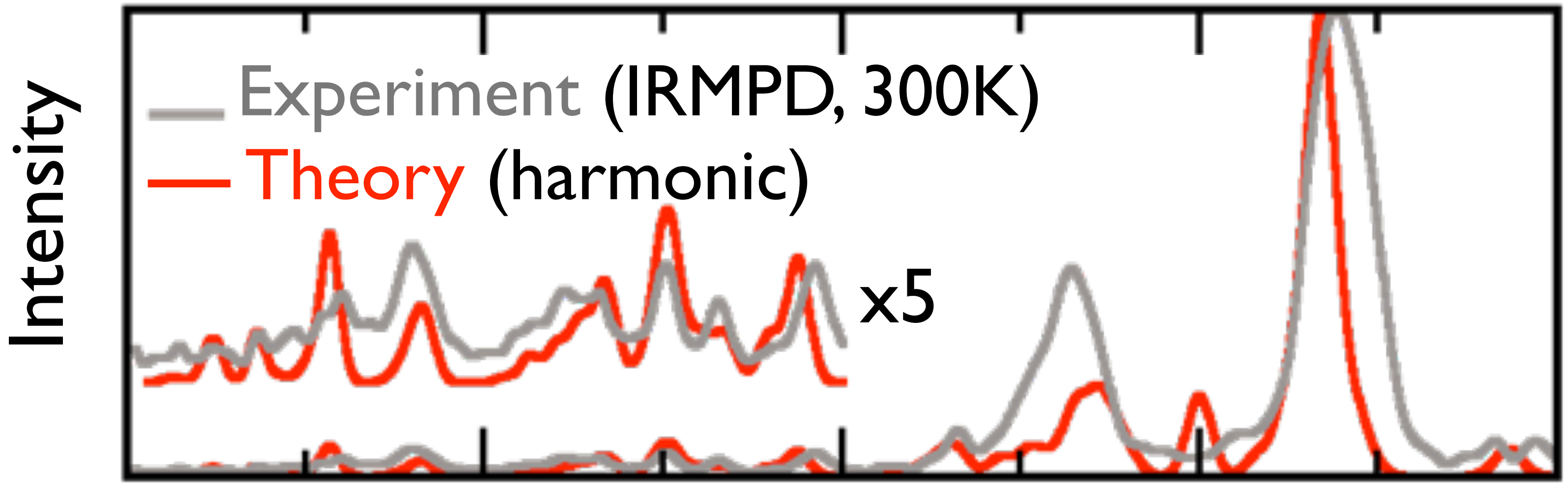


A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

25 ps Born-Oppenheimer molecular dynamics, "tight"(!), DFT-PBE+vdW

M. Rossi, V. Blum, P. Kupser, G. von Helden, F. Bierau, K. Pagel, G. Meijer, and M. Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*

Theory: PBE+vdW, shifted, not scaled



$$I(\omega) \propto \omega^2 \int_{-\infty}^{\infty} dt \, \underbrace{\langle \vec{M}(t) \cdot \vec{M}(0) \rangle}_{\text{dipole-dipole time correlation function}} e^{iwt}$$

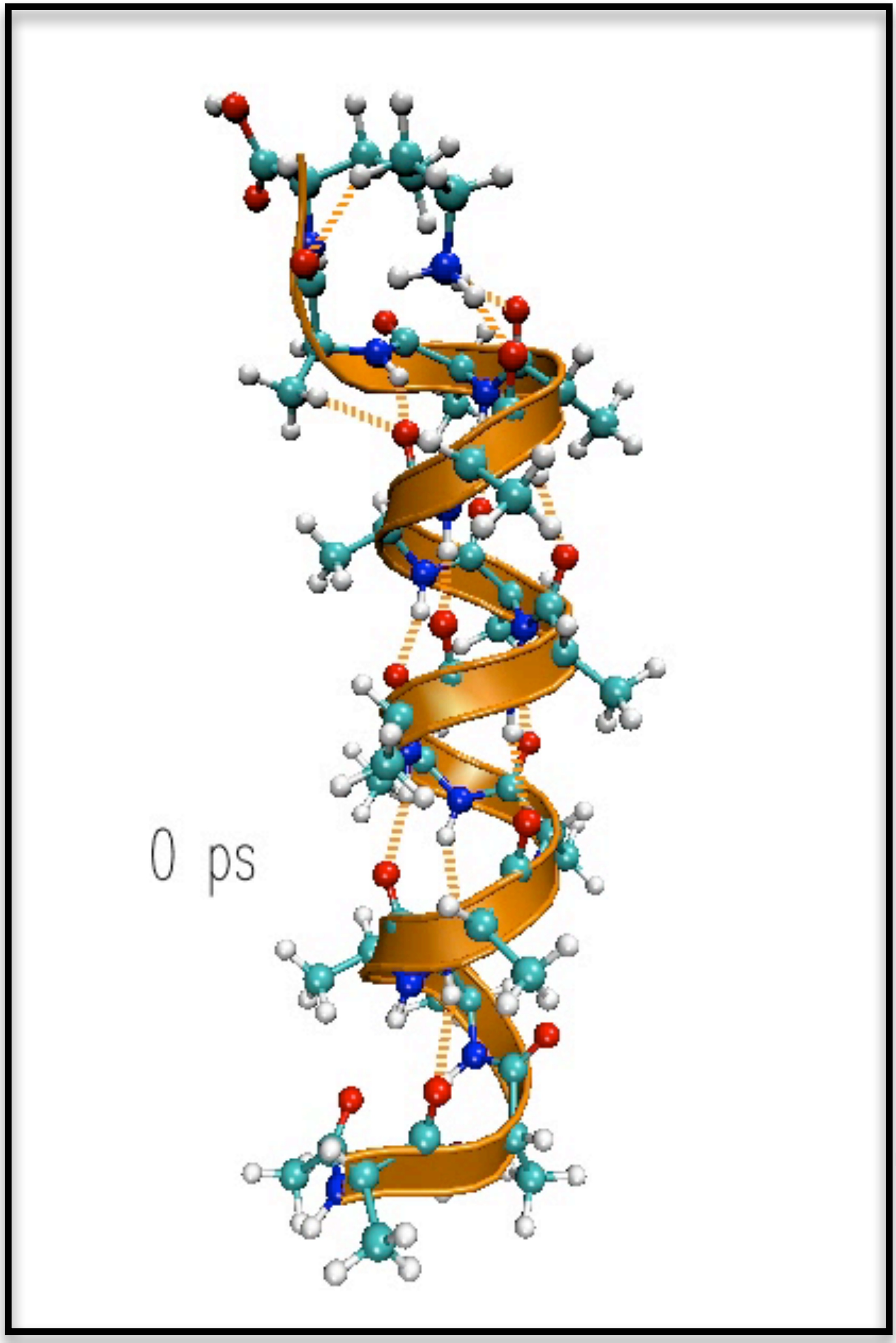
dipole-dipole time correlation function

see, e.g., *M.-P. Gaigeot, others*

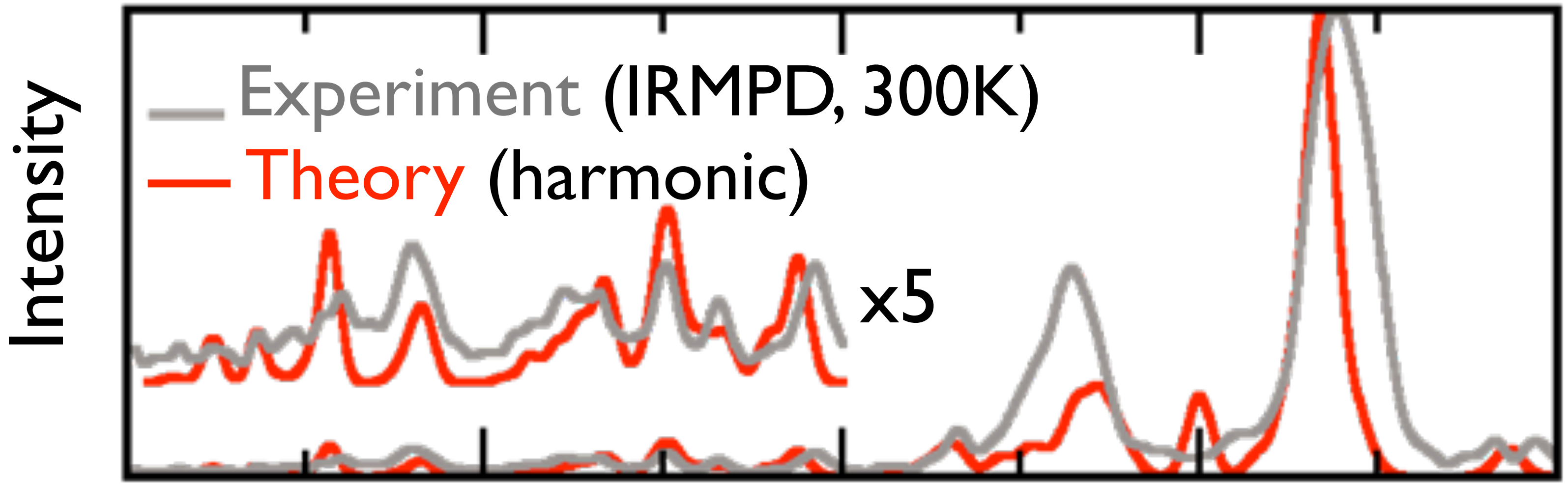
A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

25 ps Born-Oppenheimer molecular dynamics, “tight”(!), DFT-PBE+vdW

M. Rossi, V. Blum, P. Kupser, G. von Helden, F. Bierau, K. Pagel, G. Meijer, and M. Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*



Theory: PBE+vdW, shifted, not scaled



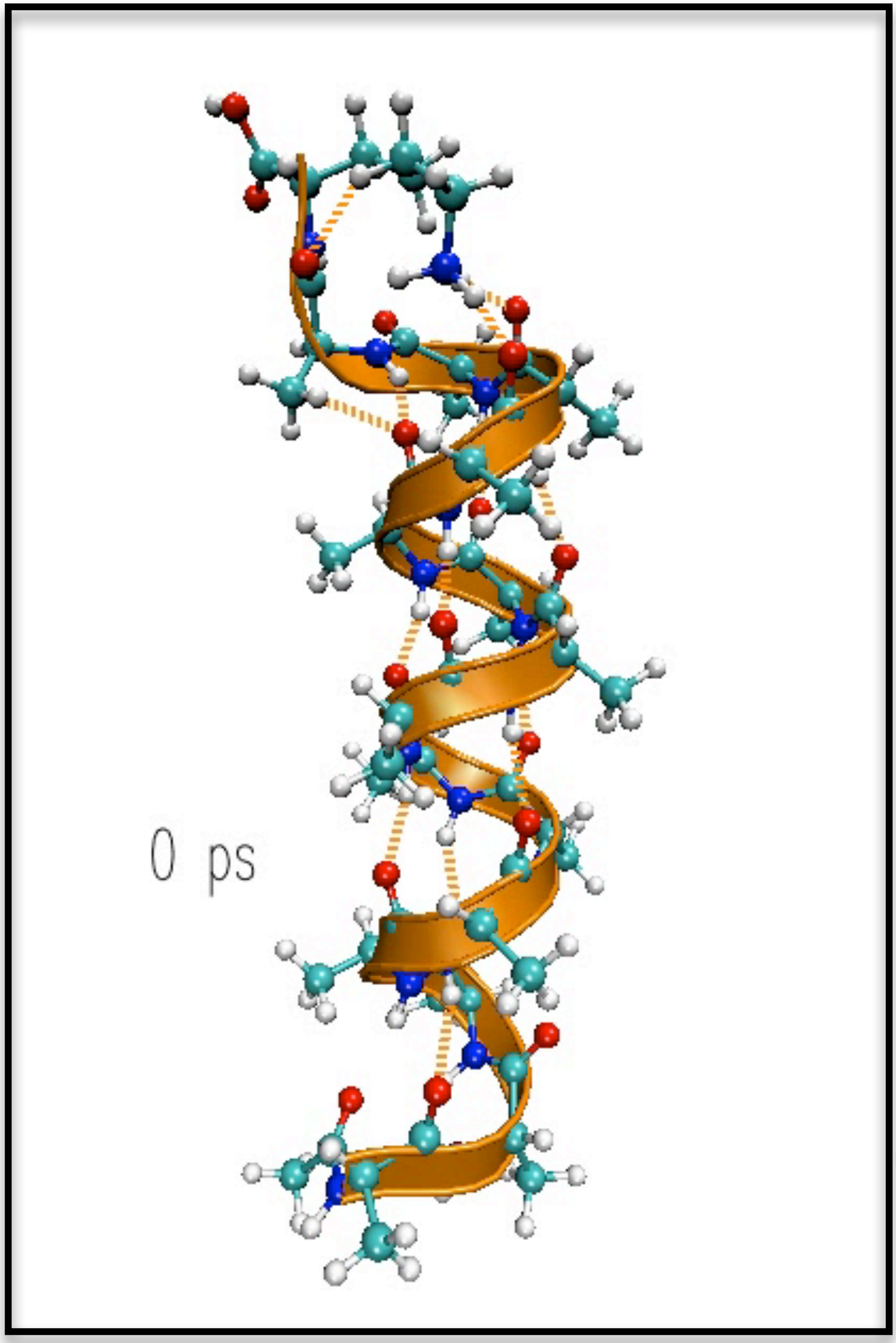
$$I(\omega) \propto \omega^2 \int_{-\infty}^{\infty} dt\, \underbrace{\langle \vec{M}(t) \cdot \vec{M}(0) \rangle}_{\text{dipole-dipole time correlation function}} e^{iwt}$$

see, e.g., *M.-P. Gaigeot, others*

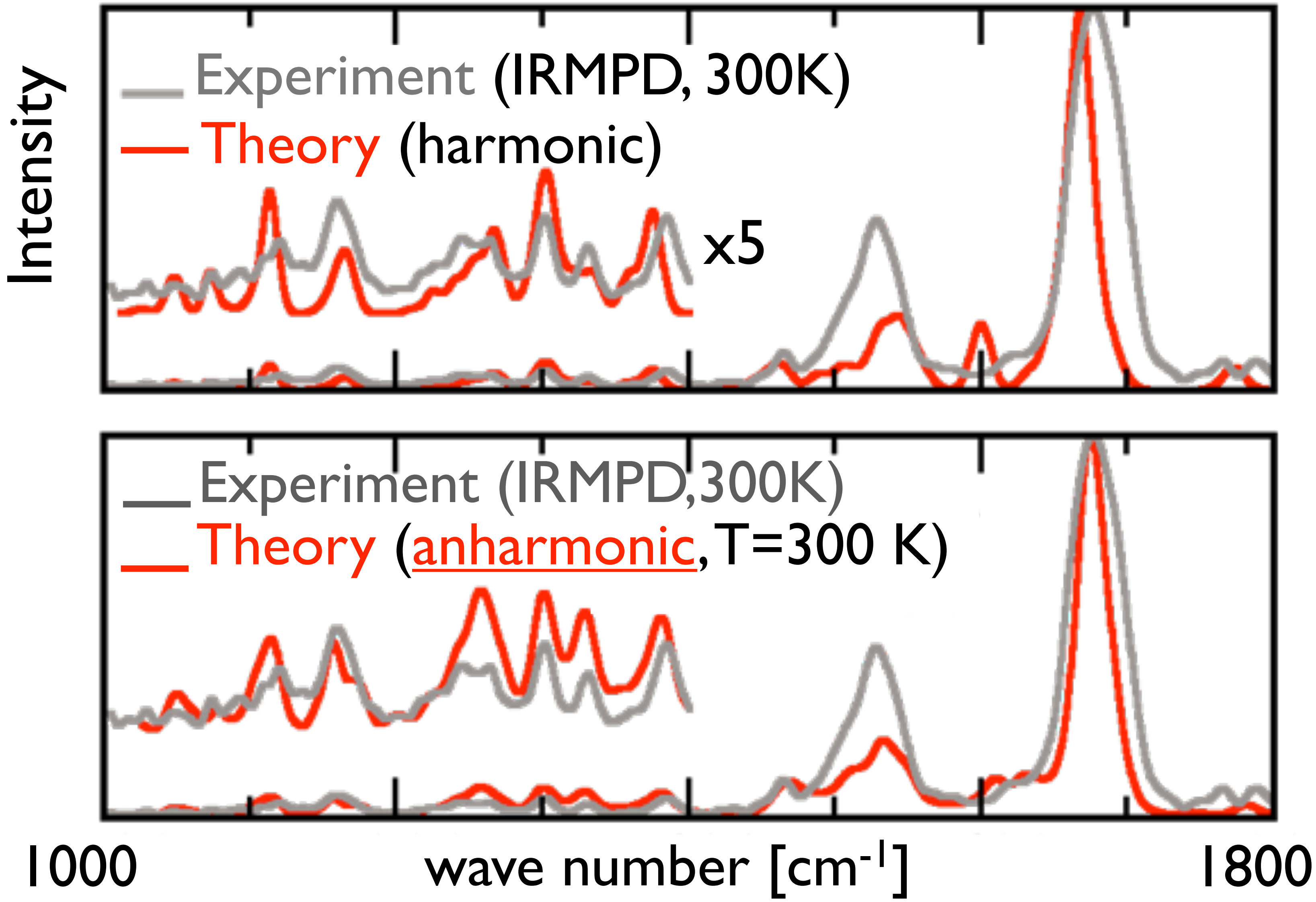
A known helix in the gas phase: Ac-Ala$_{15}$-LysH$^+$

25 ps Born-Oppenheimer molecular dynamics, "tight"(!), DFT-PBE+vdW

M. Rossi, V. Blum, P. Kupser, G. von Helden, F. Bierau, K. Pagel, G. Meijer, and M. Scheffler, J. Phys. Chem. Lett. ***1****, 3465 (2010)*



Theory: PBE+vdW, shifted, not scaled



$$I(\omega) \propto \omega^2 \int_{-\infty}^{\infty} dt \, \underbrace{\langle \vec{M}(t) \cdot \vec{M}(0) \rangle} e^{iwt}$$

dipole-dipole time correlation function

see, e.g., *M.-P. Gaigeot, others*

Now what happens without van der Waals?

Kohtani et al., JACS ***126****, 7420 (2004)*: "Extreme stability of an unsolvated helix"
Ac-Ala$_{15}$-LysH$^+$ **α-helix** is stable up to ≈ 650 K

Now what happens without van der Waals?

Kohtani et al., JACS ***126****, 7420 (2004)*: "Extreme stability of an unsolvated helix"
Ac-Ala_{15}-$LysH^+$ **α-helix** is stable up to ≈ 650 K

PBE

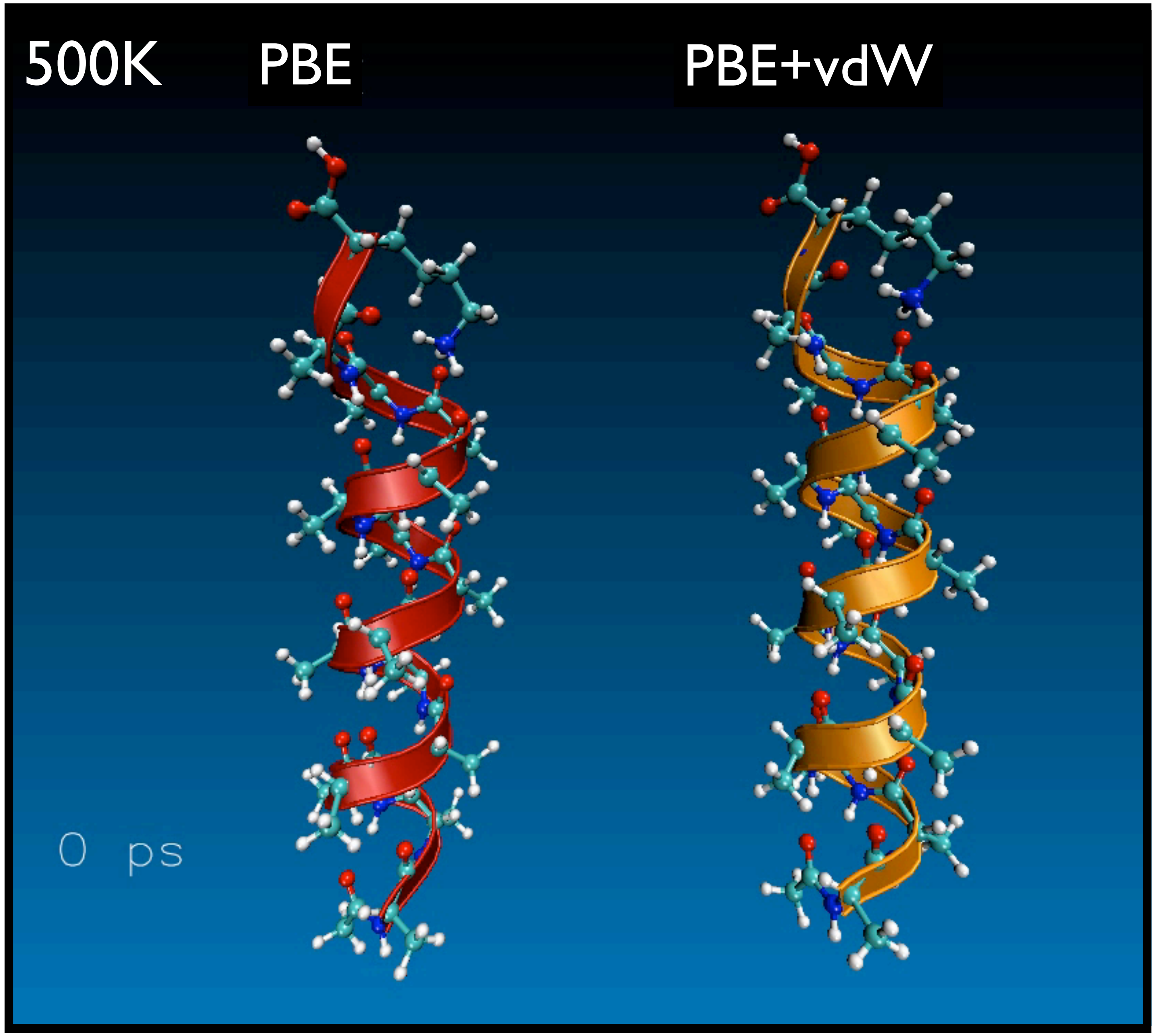
PBE+vdW

A. Tkatchenko, M. Rossi, V. Blum, J. Ireta, and M. Scheffler, Phys. Rev. Lett. ***106****, 118102 (2011)*

Now what happens without van der Waals?

Kohtani et al., JACS ***126****, 7420 (2004)*: "Extreme stability of an unsolvated helix"

$\text{Ac-Ala}_{15}\text{-LysH}^+$ **α-helix** is stable up to ≈ 650 K

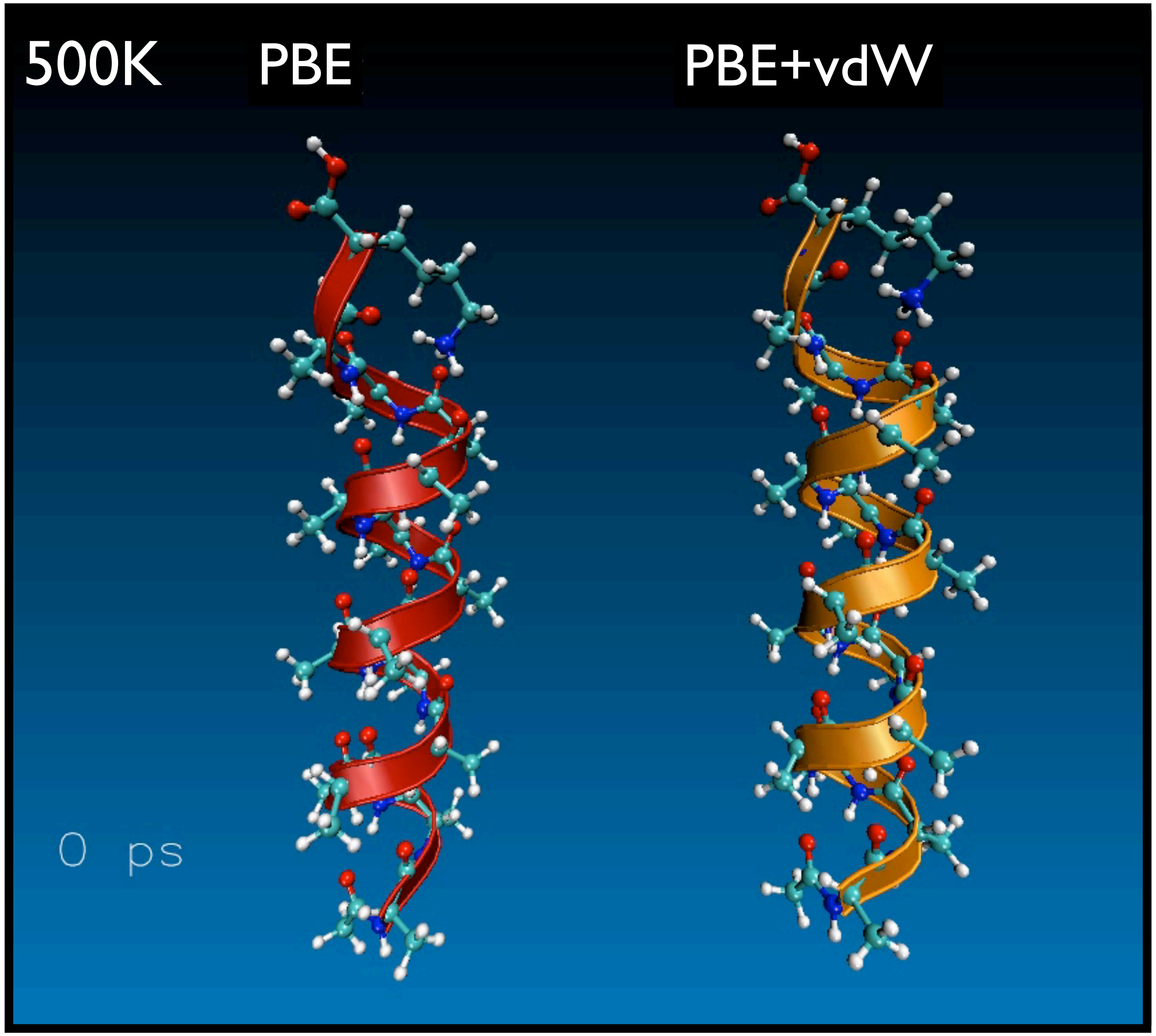


A. Tkatchenko, M. Rossi, V. Blum, J. Ireta, and M. Scheffler, Phys. Rev. Lett. ***106****, 118102 (2011)*

Now what happens without van der Waals?

Kohtani et al., JACS ***126****, 7420 (2004)*: "Extreme stability of an unsolvated helix"
Ac-Ala$_{15}$-LysH$^+$ **α-helix** is stable up to ≈ 650 K

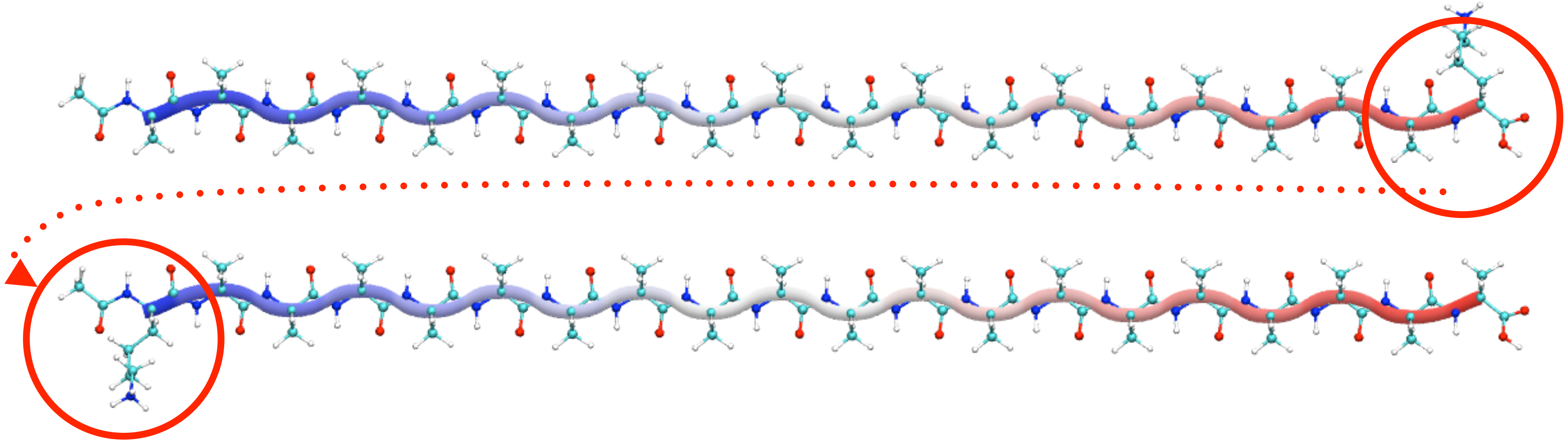
Mostly 3$_{10}$ helical!



α helical!

A. Tkatchenko, M. Rossi, V. Blum, J. Ireta, and M. Scheffler, Phys. Rev. Lett. ***106****, 118102 (2011)*

... and what if we "structurally destroy" the helix?



Ac-Ala$_{19}$-LysH$^+$

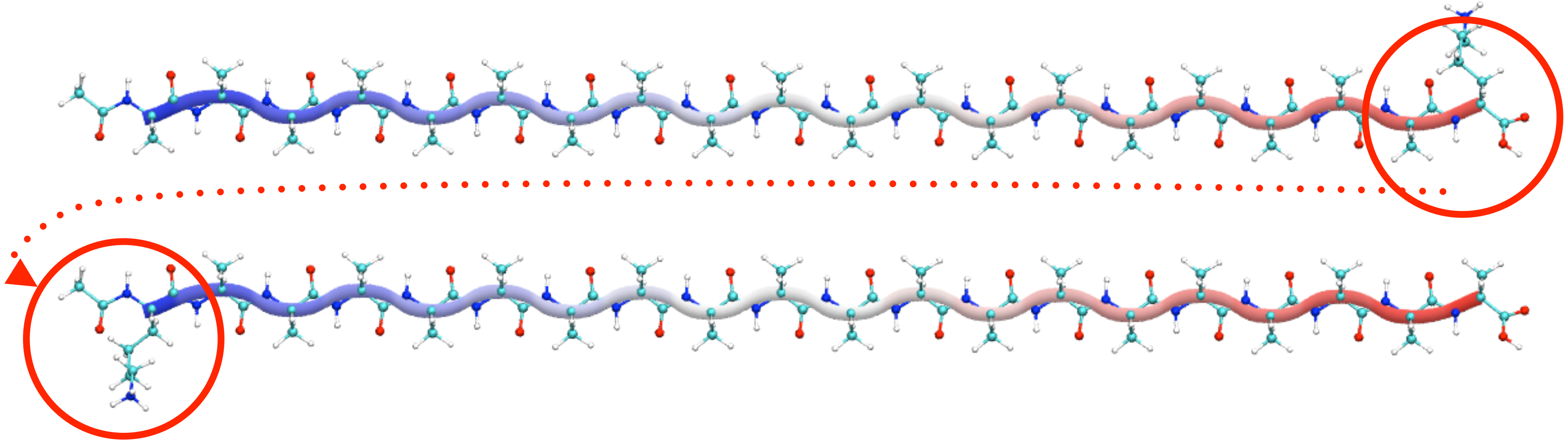
Ac-LysH$^+$-Ala$_{19}$

Can we tackle the conformational space of a 220-atom system?
Combinatorial explosion!

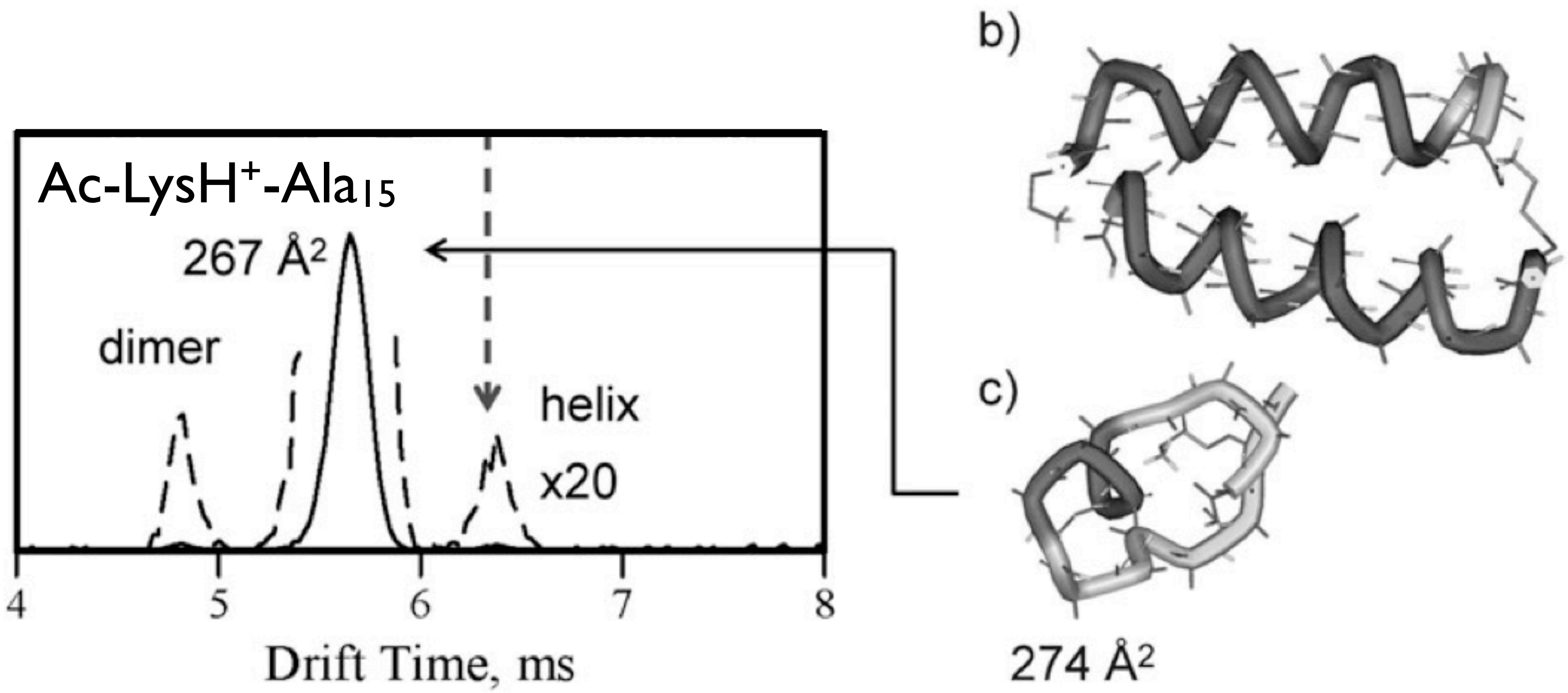
... and what if we “structurally destroy” the helix?

Ac-Ala$_{19}$-LysH$^+$

Ac-LysH$^+$-Ala$_{19}$



Can we tackle the conformational space of a 220-atom system?
Combinatorial explosion!

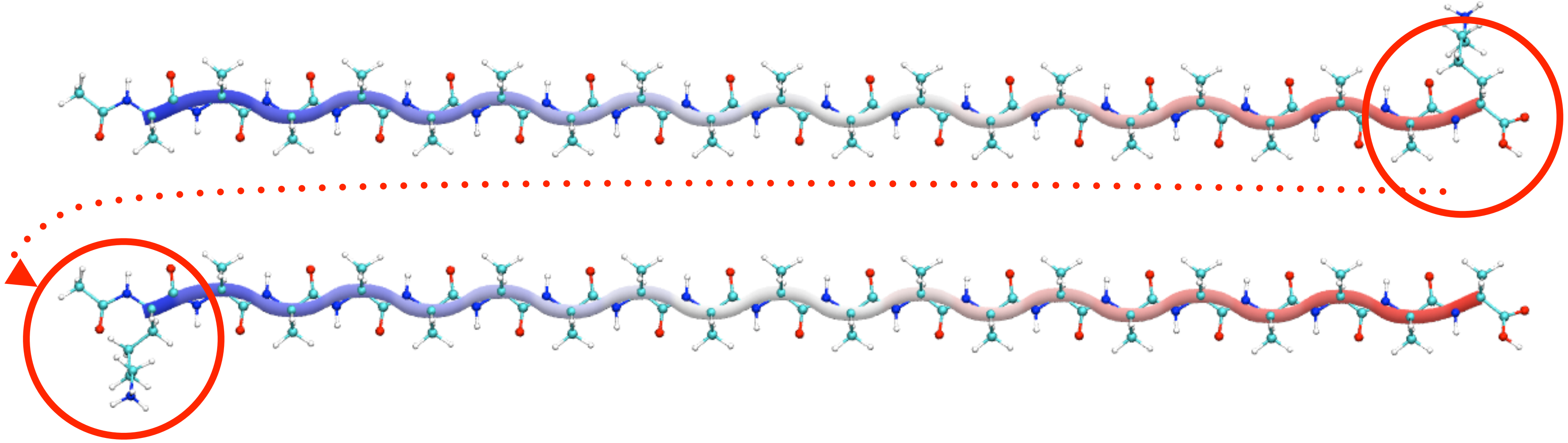


[2] Jarrold, PCCP 9, 1659, 2007

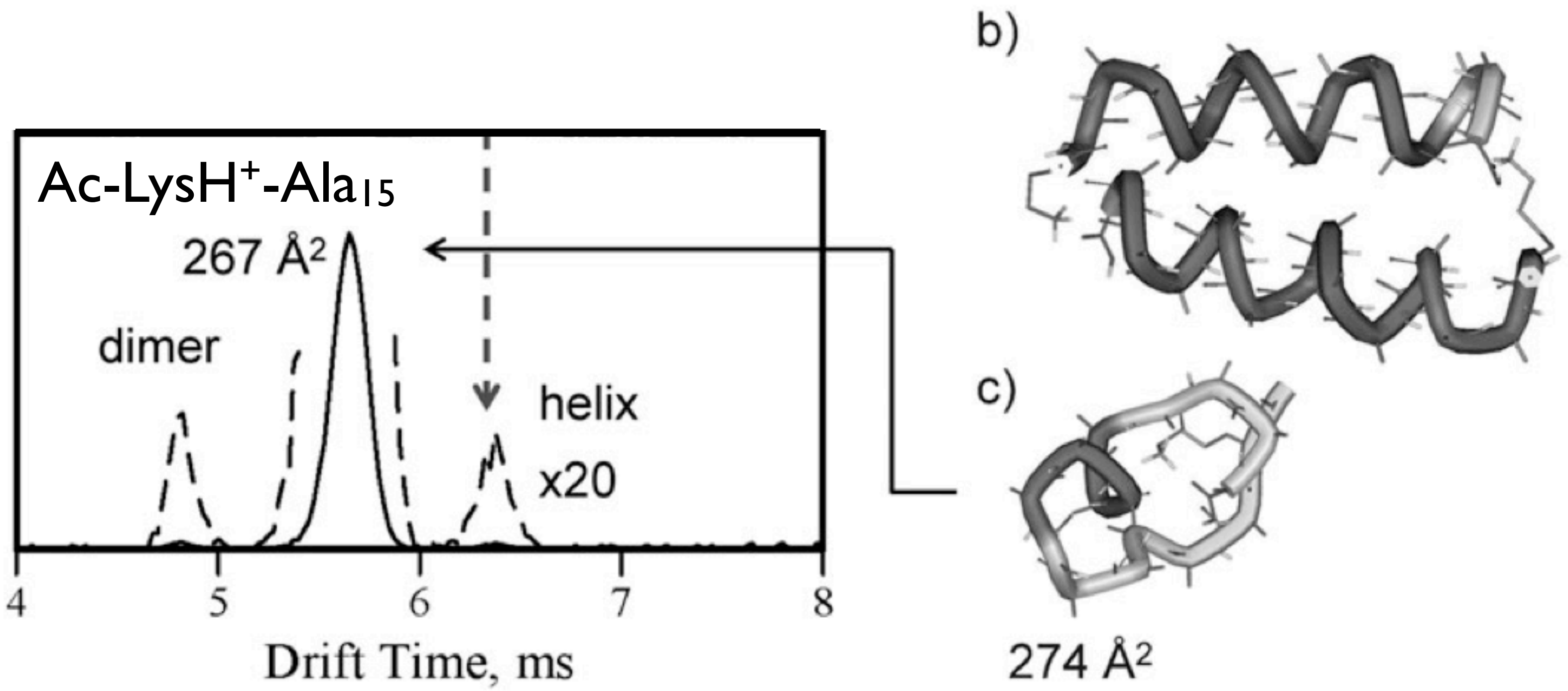
... and what if we “structurally destroy” the helix?

Ac-Ala$_{19}$-LysH$^+$

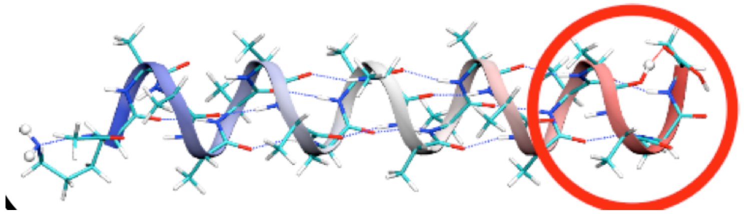
Ac-LysH$^+$-Ala$_{19}$



Can we tackle the conformational space of a 220-atom system?
Combinatorial explosion!



[2] Jarrold, PCCP 9, 1659, 2007

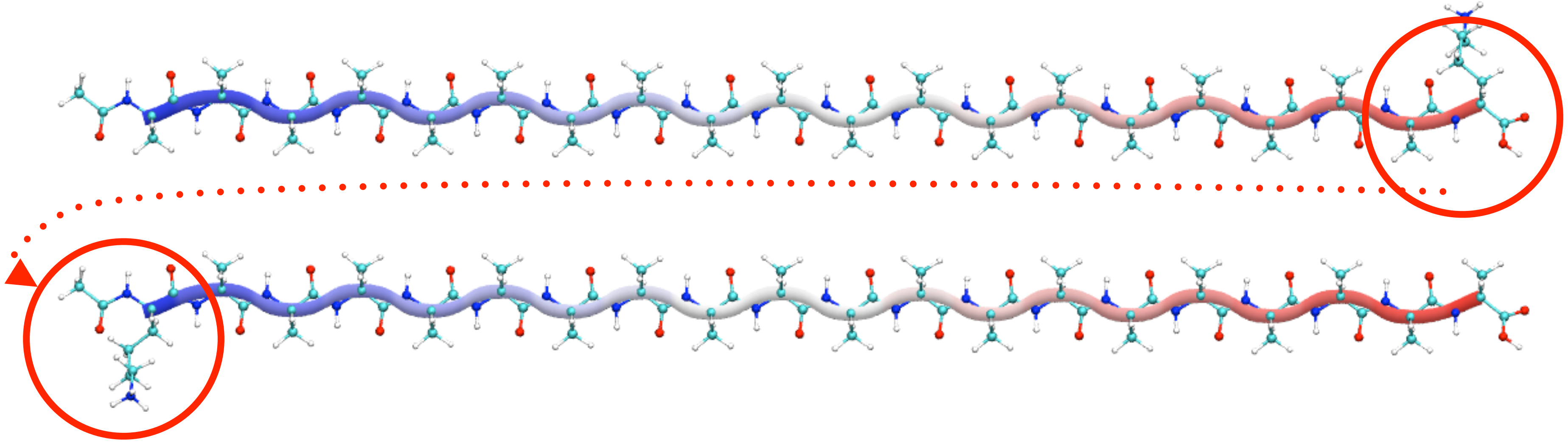


Helix, protonated C-terminus?

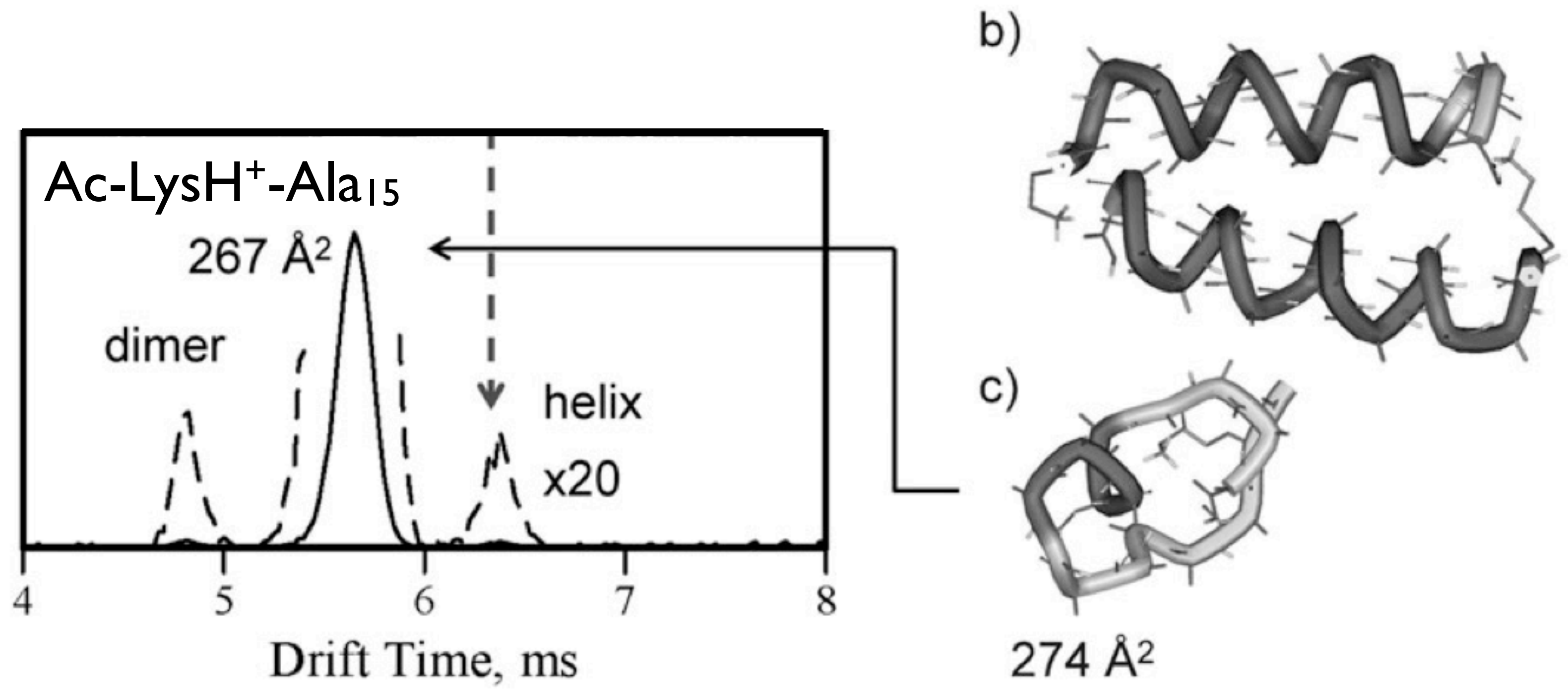
... and what if we “structurally destroy” the helix?

$\text{Ac-Ala}_{19}\text{-LysH}^+$

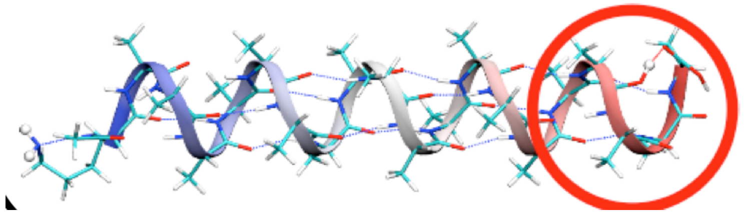
$\text{Ac-LysH}^+\text{-Ala}_{19}$



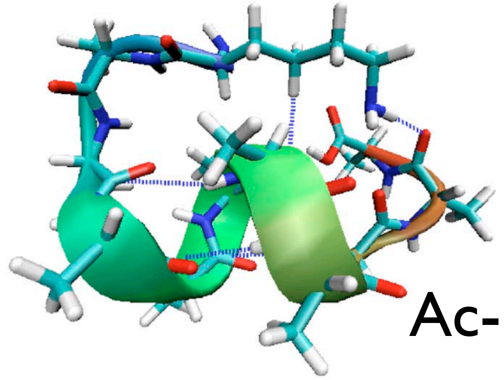
Can we tackle the conformational space of a 220-atom system?
Combinatorial explosion!



[2] Jarrold, PCCP 9, 1659, 2007



Helix, protonated C-terminus?



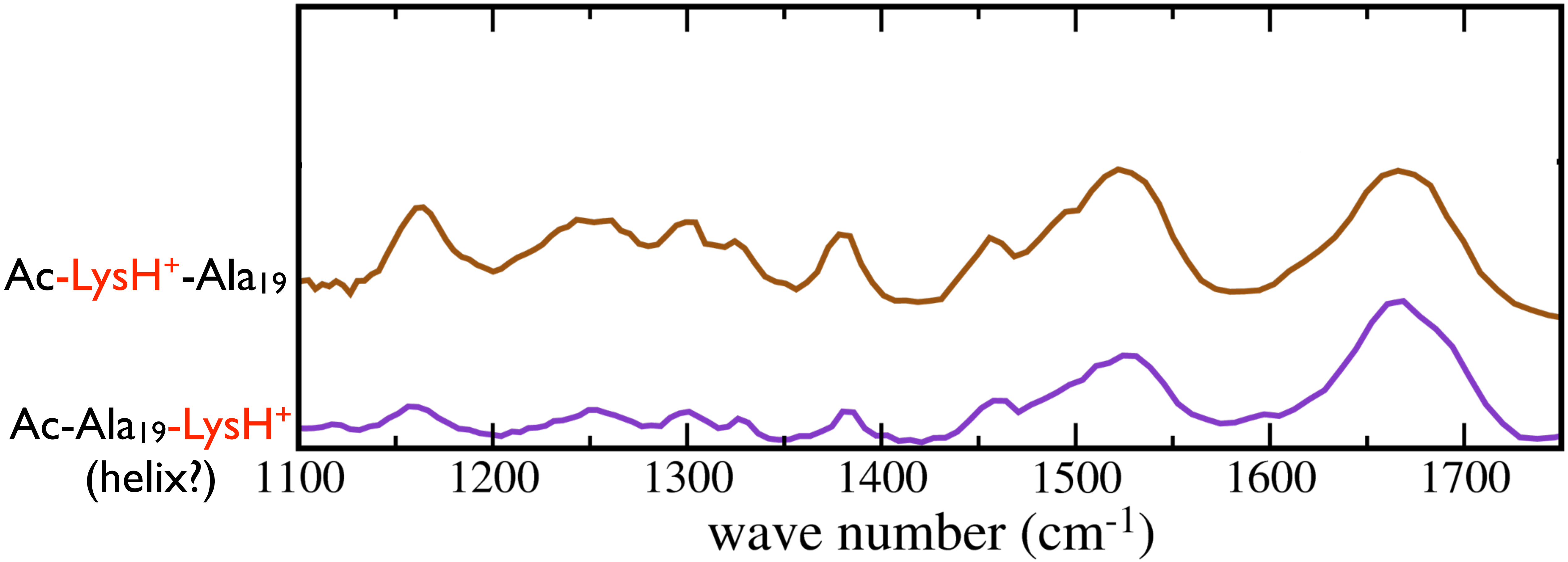
$\text{Ac-Lys}^+\text{-Ala}_{10}$

Wei et al., JCP 126, 204307 (2007)
Replica Exchange / Force Field

Turn again to vibrational spectroscopy: Helical?

Peter Kupser, Gert von Helden, Frank Filsinger, Kevin Pagel, Gerard Meijer,
Molecular Physics, FHI Berlin

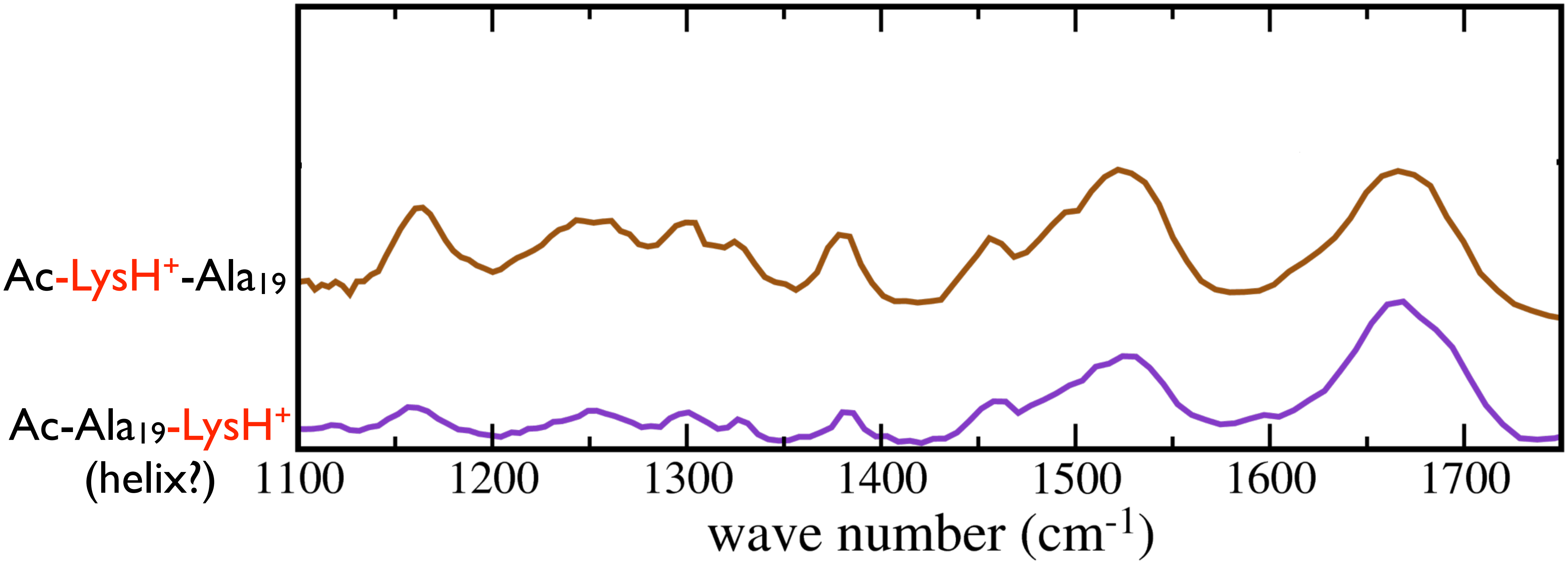
Infrared multiphoton dissociation spectroscopy, FELIX, room temperature

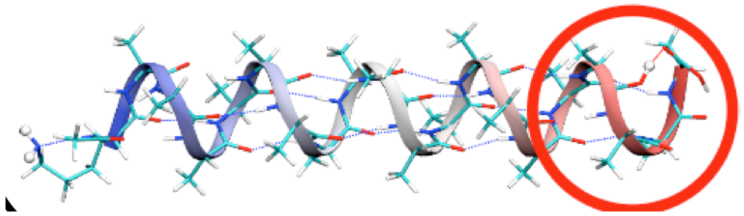


Turn again to vibrational spectroscopy: Helical?

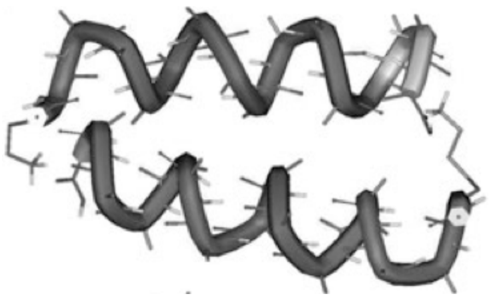
Peter Kupser, Gert von Helden, Frank Filsinger, Kevin Pagel, Gerard Meijer,
Molecular Physics, FHI Berlin

Infrared multiphoton dissociation spectroscopy, FELIX, room temperature





helix after all?

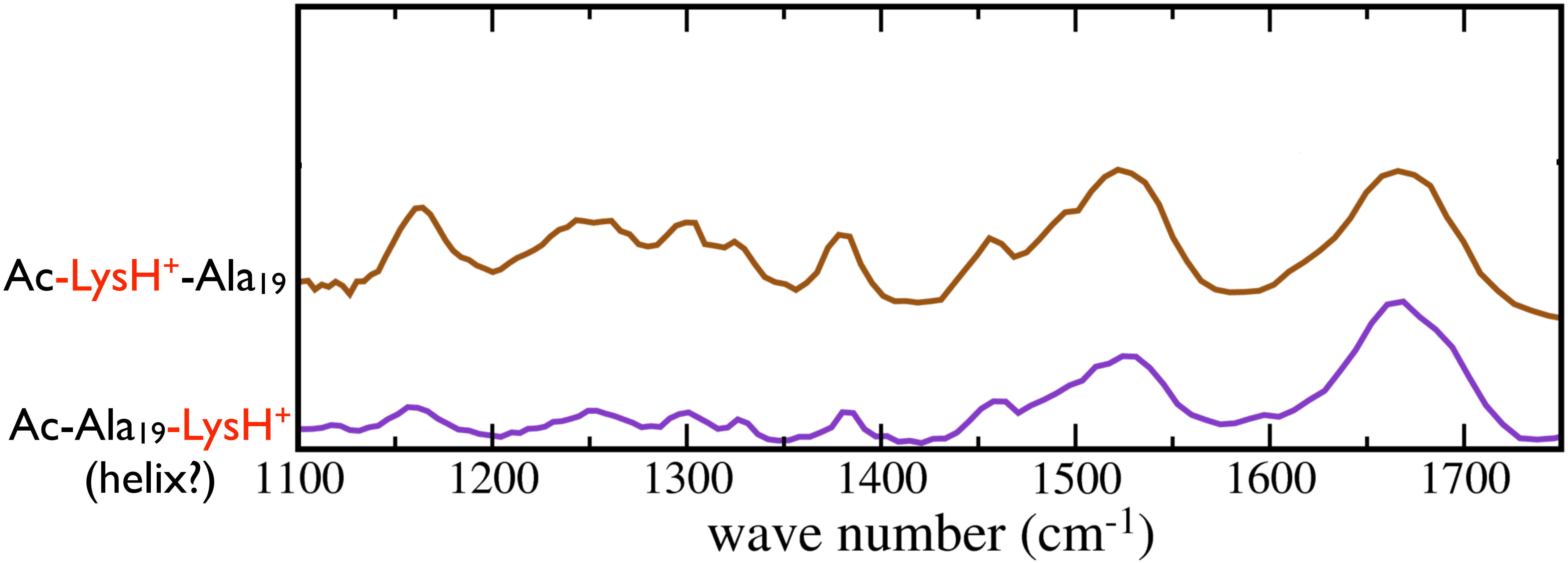


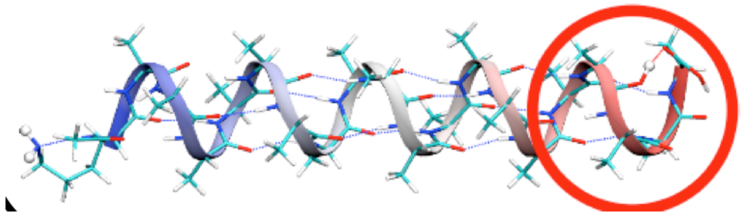
helix dimer?

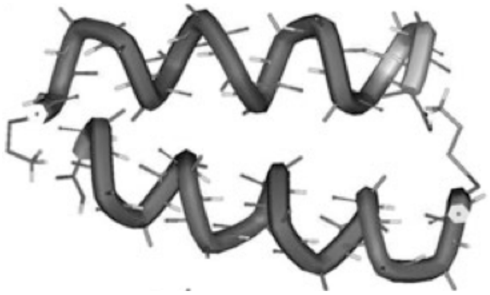
Turn again to vibrational spectroscopy: Helical?

Peter Kupser, Gert von Helden, Frank Filsinger, Kevin Pagel, Gerard Meijer,
Molecular Physics, FHI Berlin

Infrared multiphoton dissociation spectroscopy, FELIX, room temperature




helix after all?

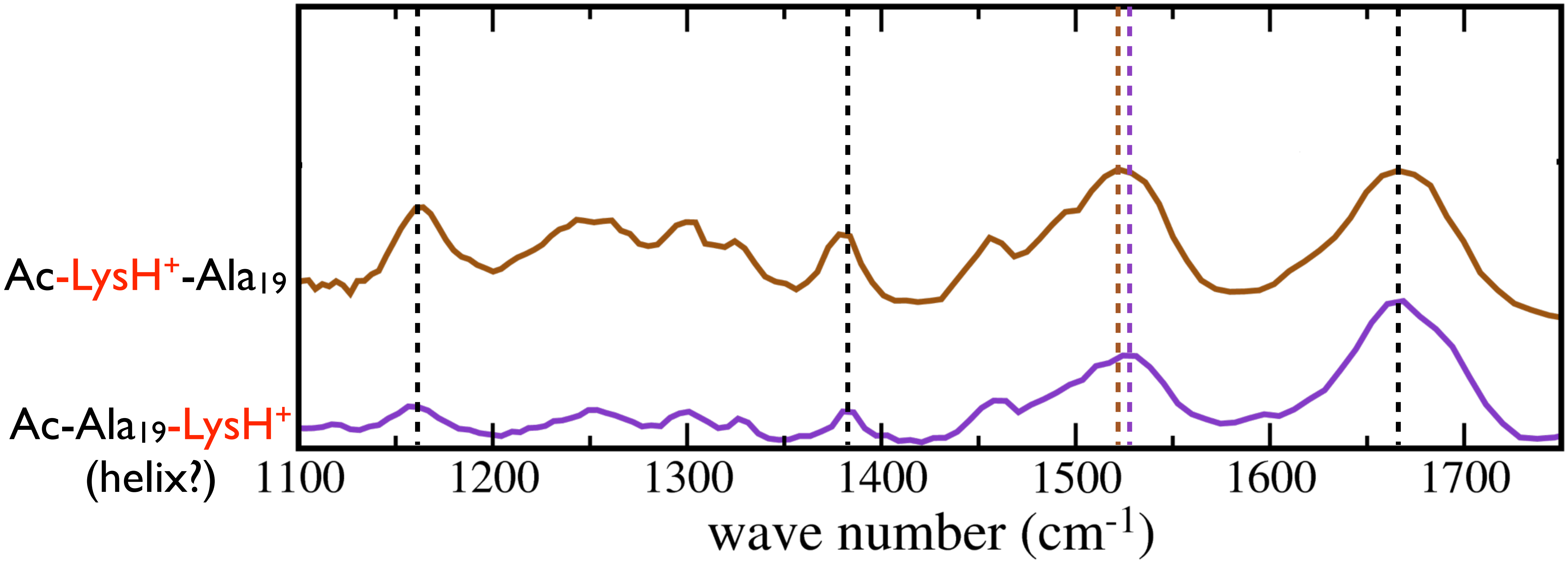

helix dimer?

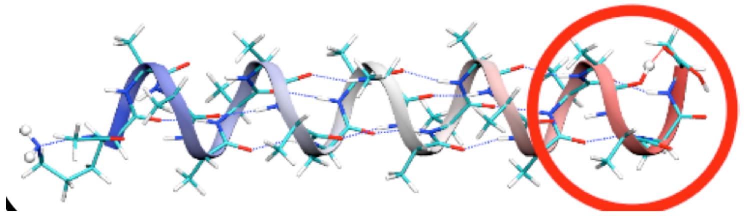

not
structure-
sensitive?

Turn again to vibrational spectroscopy: Helical?

Peter Kupser, Gert von Helden, Frank Filsinger, Kevin Pagel, Gerard Meijer,
Molecular Physics, FHI Berlin

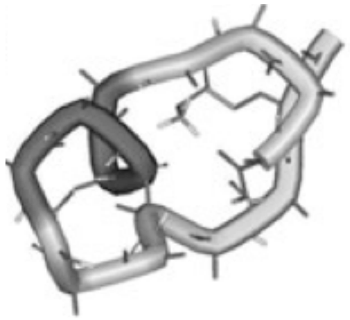
Infrared multiphoton dissociation spectroscopy, FELIX, room temperature





helix after all?

helix dimer?

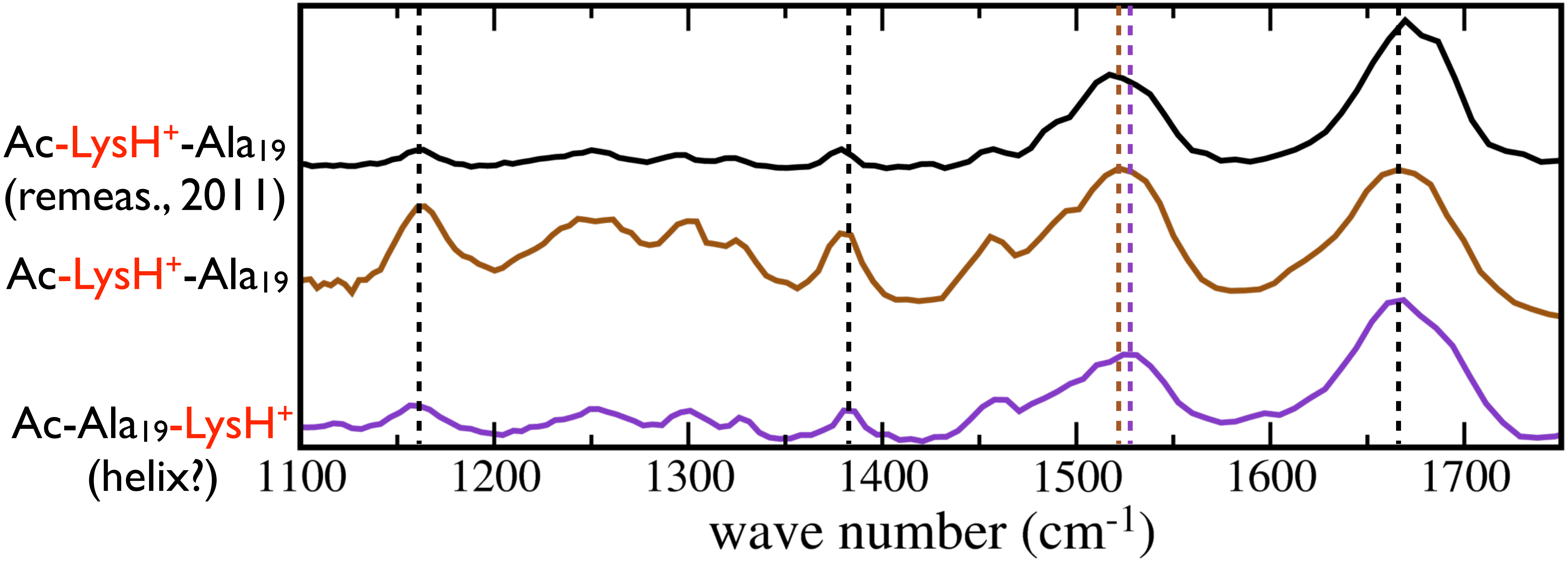


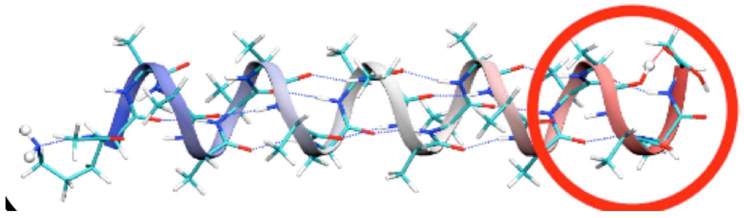
not structure-sensitive?

Turn again to vibrational spectroscopy: Helical?

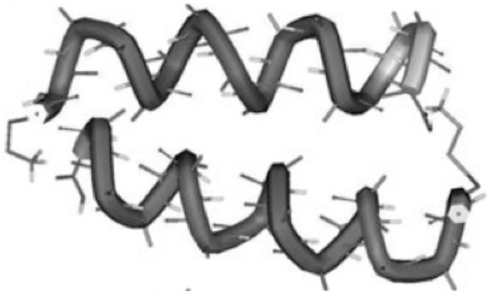
Peter Kupser, Gert von Helden, Frank Filsinger, Kevin Pagel, Gerard Meijer,
Molecular Physics, FHI Berlin

Infrared multiphoton dissociation spectroscopy, FELIX, room temperature

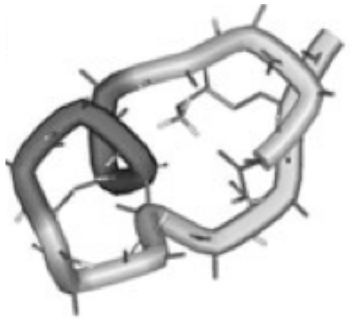




helix after all?



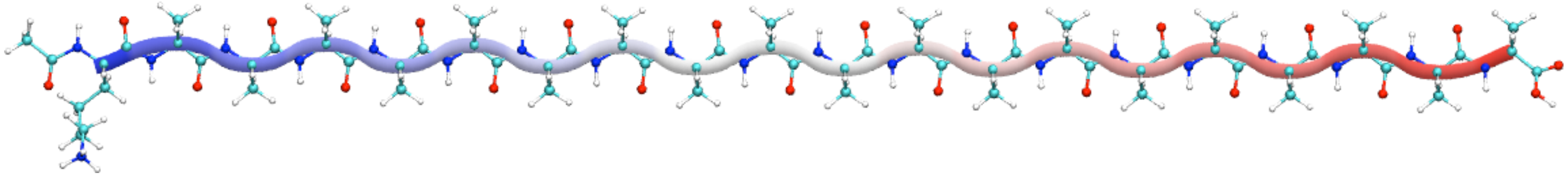
helix dimer?



not
structure-
sensitive?

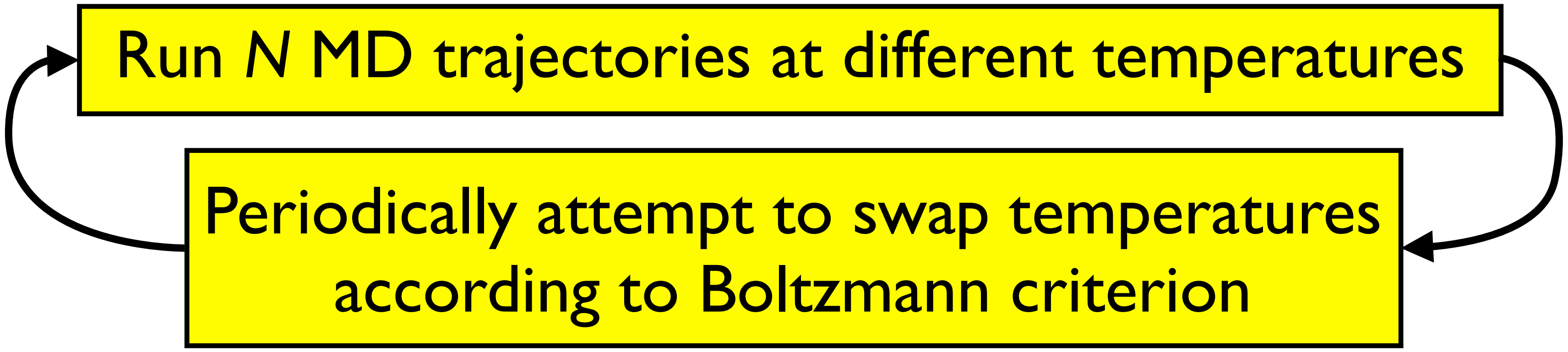
Ac-LysH$^+$-Ala$_{19}$: Searching a huge conformational space

Franziska Schubert, Carsten Baldauf, Mariana Rossi, VB - FHI



220 atoms - well beyond brute force “enumeration”.

One alternative:
Replica Exchange Molecular Dynamics (REMD)



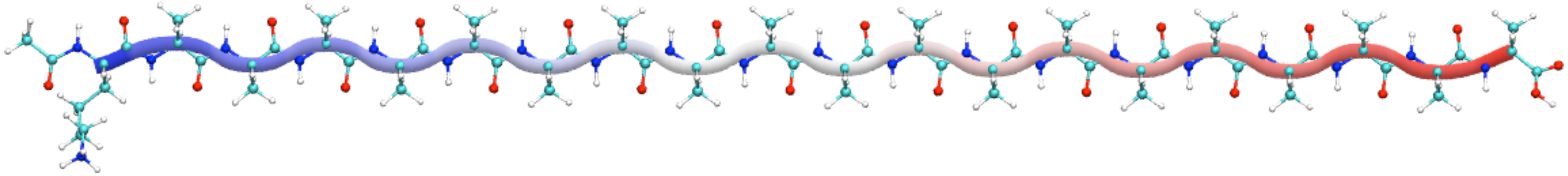
Run *N* MD trajectories at different temperatures

Periodically attempt to swap temperatures according to Boltzmann criterion

Marinari, Parisi, Europhys. Lett ***19****, 451 (1992); U.H.E. Hansmann, Chem. Phys. Lett.* ***281****, 140 (1997); Y. Sugita, Y. Okamoto, Chem. Phys. Lett.* ***314****, 14 (1999); many others*

Ac-LysH$^+$-Ala$_{19}$: Searching a huge conformational space

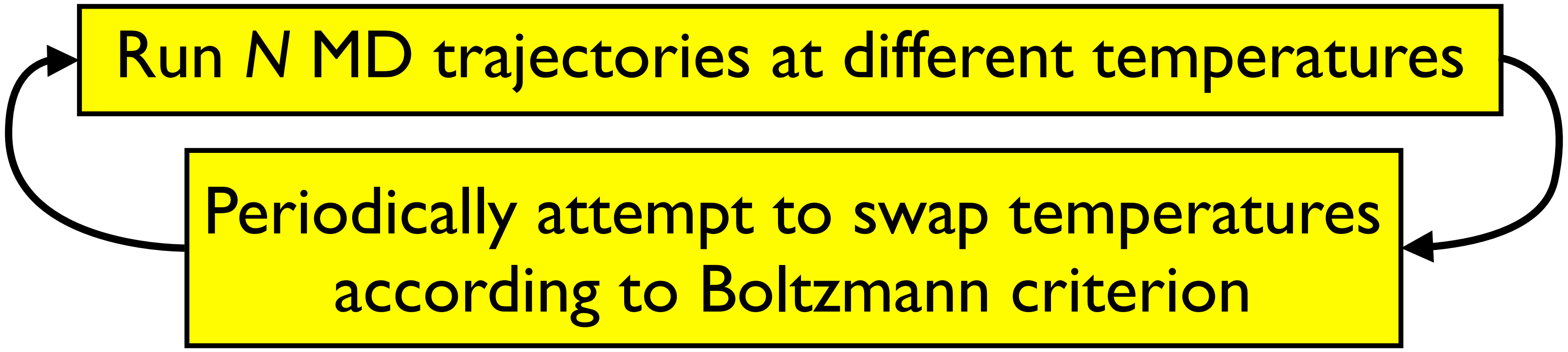
Franziska Schubert, Carsten Baldauf, Mariana Rossi, VB - FHI



220 atoms - well beyond brute force "enumeration".

One alternative:

Replica Exchange Molecular Dynamics (REMD)



Run *N* MD trajectories at different temperatures

Periodically attempt to swap temperatures according to Boltzmann criterion

But typical (even just to scan for structures):
100s of nanoseconds, O(10) trajectories
→ first principles??

Marinari, Parisi, Europhys. Lett ***19****, 451 (1992); U.H.E. Hansmann, Chem. Phys. Lett.* ***281****, 140 (1997); Y. Sugita, Y. Okamoto, Chem. Phys. Lett.* ***314****, 14 (1999); many others*

Coupling REMD & first principles:

<u>Strategy:</u>

"Global" FF REMD (e.g., 500 ns × 16 trajectories)

(Gromacs OPLS-AA)

Postrelax, reweight $O(10^3)$ midpoint structures by DFT-PBE+vdW

"Local" *ab initio* REMD for few "best" conformers DFT-PBE+vdW, 10-20 ps × 16 trajectories

ai-REMD in FHI-aims: Luca Ghiringhelli, FHI

Coupling REMD & first principles:

Strategy:

(Gromacs OPLS-AA)

"Global" FF REMD (e.g., 500 ns × 16 trajectories)

Postrelax, reweight O(10^3) midpoint structures
by DFT-PBE+vdW

"Local" *ab initio* REMD for few "best" conformers
DFT-PBE+vdW, 10-20 ps × 16 trajectories

ai-REMD in FHI-aims: Luca Ghiringhelli, FHI

For:

Monomers
Dimers
Helices with "mobile proton"
"Unlikely α-helix" with proton at LysH^+

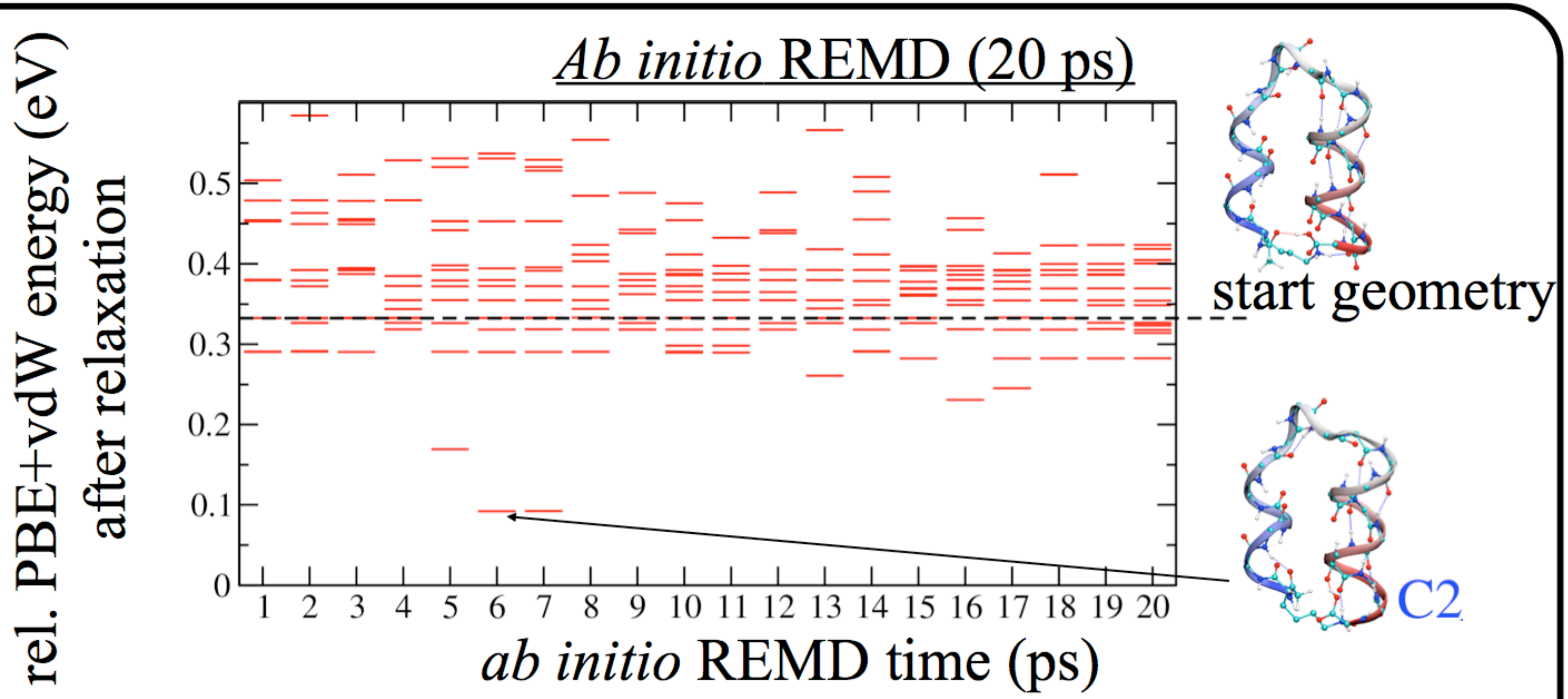
Coupling REMD & first principles:

Strategy:

"Global" FF REMD (e.g., 500 ns × 16 trajectories)

(Gromacs OPLS-AA)

Postrelax, reweight $O(10^3)$ midpoint structures



Ab initio REMD (20 ps)

→ rearrange local bond networks (termination!) successfully

Conformation prototypes vs. IRMPD: (1) Helical monomers

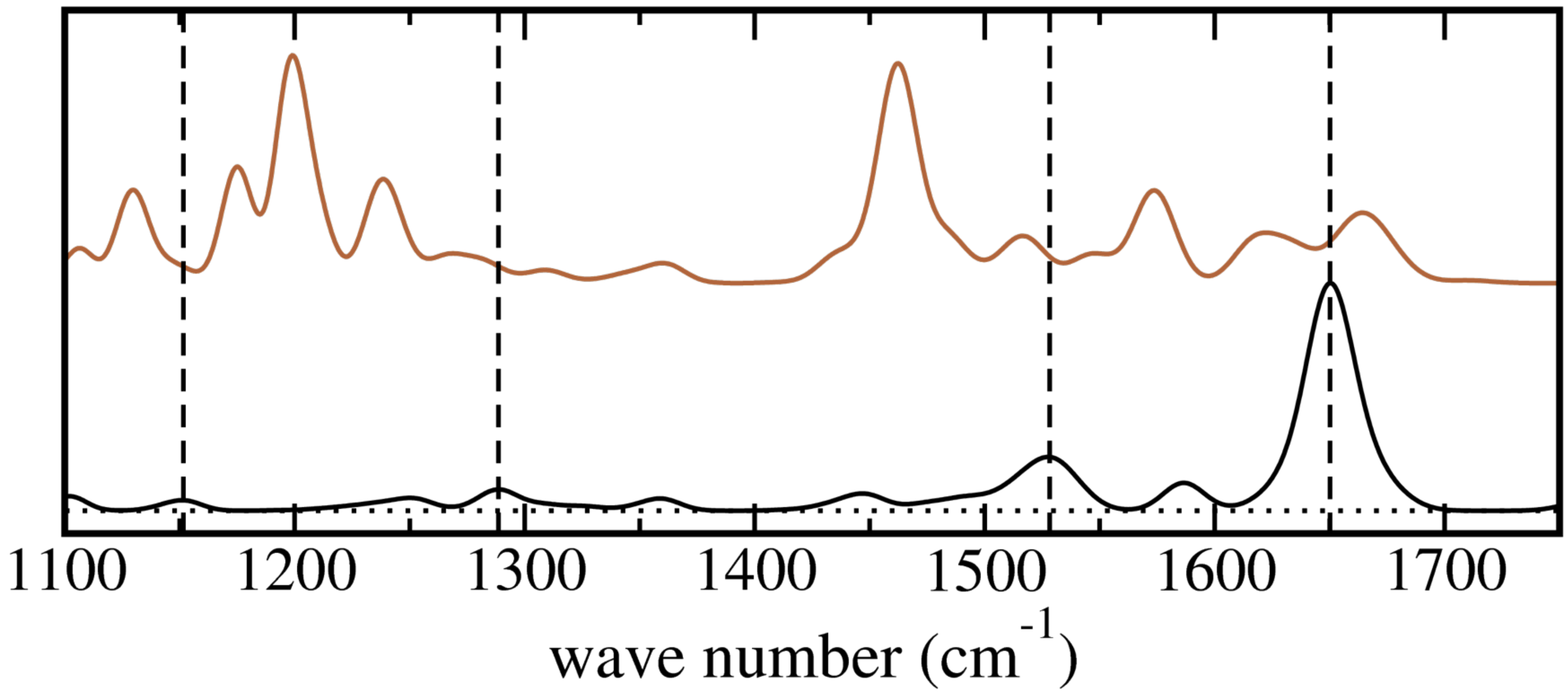
Theory-theory comparison, harmonic:

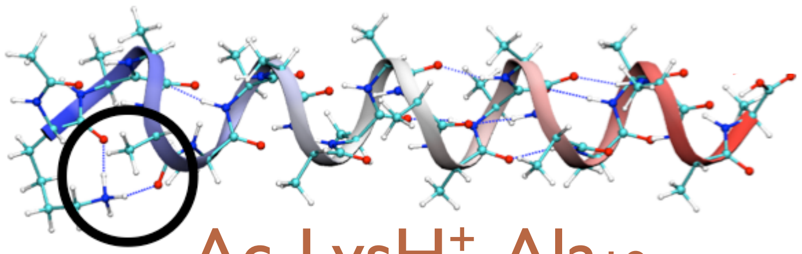
$\underline{\text{Ac-LysH}^+\text{-Ala}_{19} \leftrightarrow \text{Ac-Ala}_{19}\text{-LysH}^+}$

Conformation prototypes vs. IRMPD: (1) Helical monomers

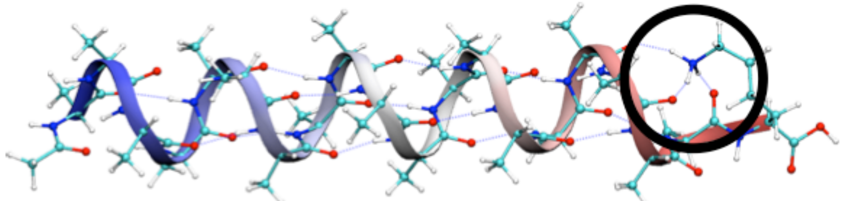
Theory-theory comparison, harmonic:

Ac-LysH$^+$-Ala$_{19}$ ↔ Ac-Ala$_{19}$-LysH$^+$





Ac-LysH$^+$-Ala$_{19}$
proton at N terminus

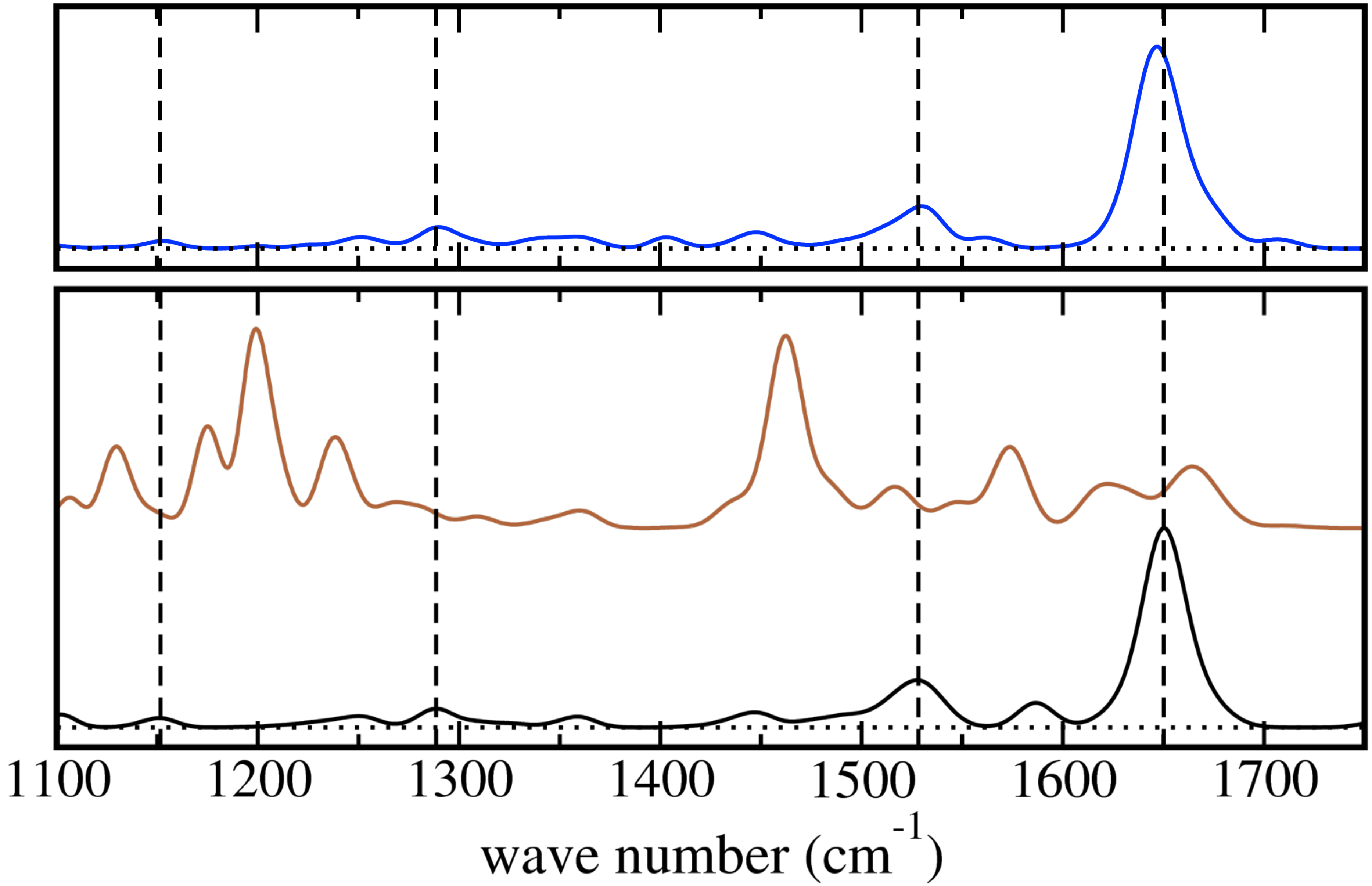


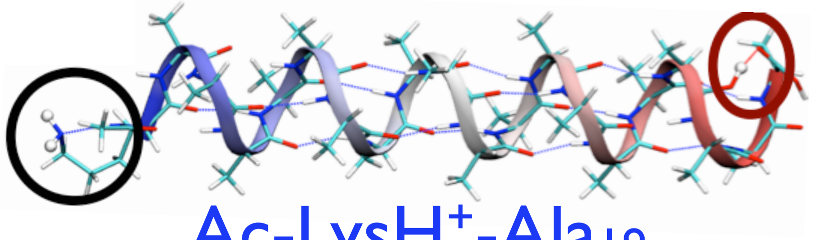
Ac-Ala$_{19}$-LysH$^+$
helix

Conformation prototypes vs. IRMPD: (1) Helical monomers

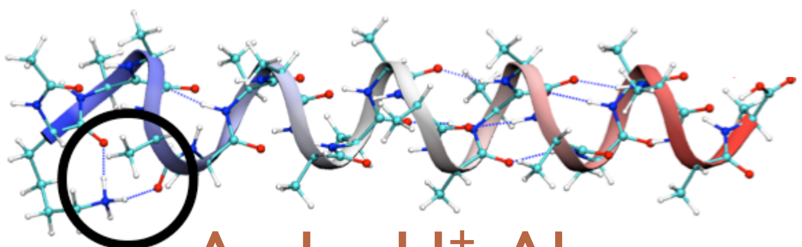
Theory-theory comparison, harmonic:

Ac-LysH$^+$-Ala$_{19}$ ↔ Ac-Ala$_{19}$-LysH$^+$

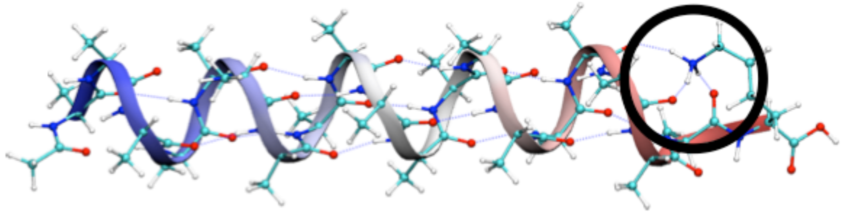




Ac-LysH$^+$-Ala$_{19}$
proton at C terminus



Ac-LysH$^+$-Ala$_{19}$
proton at N terminus



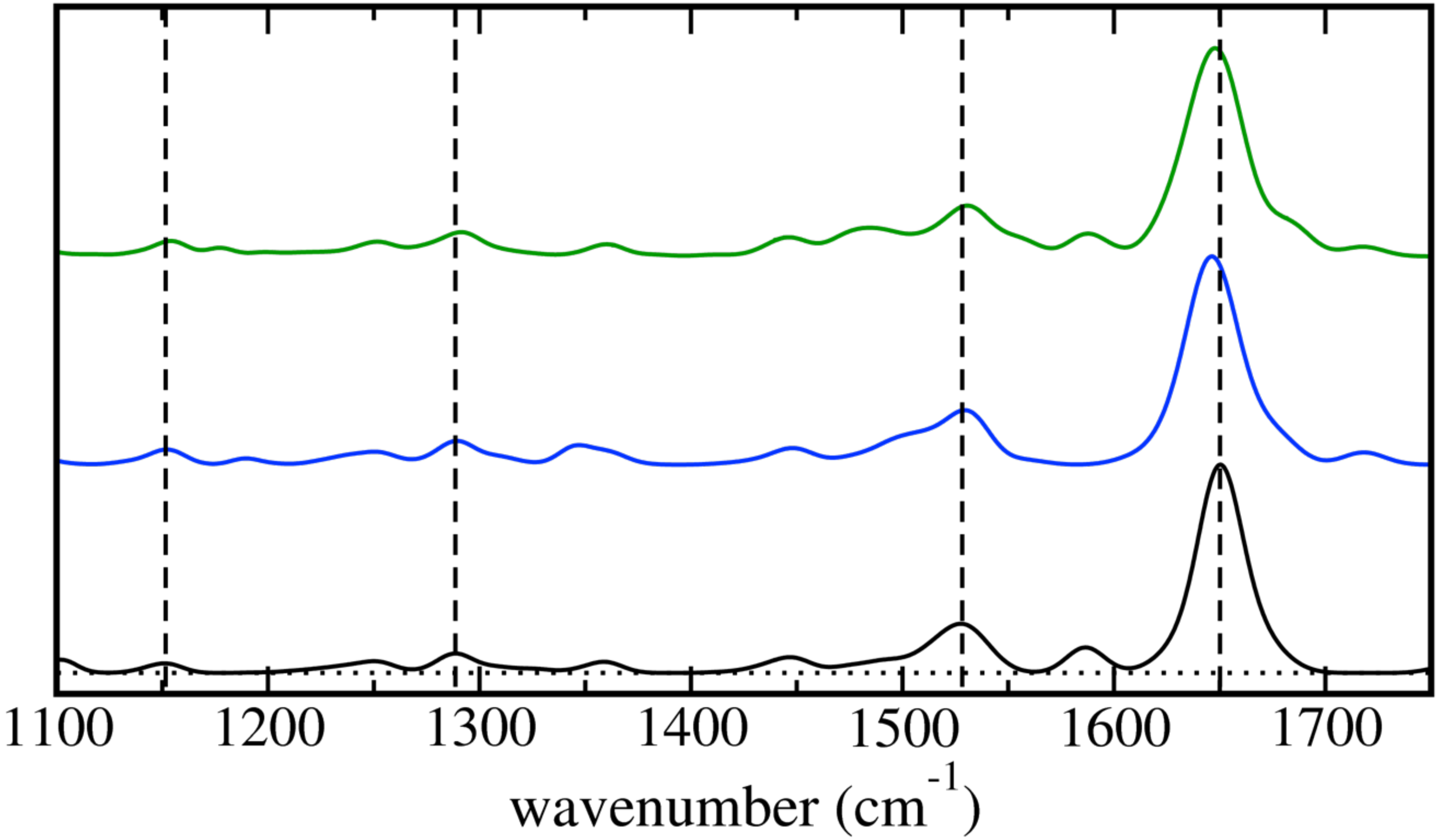
Ac-Ala$_{19}$-LysH$^+$
helix

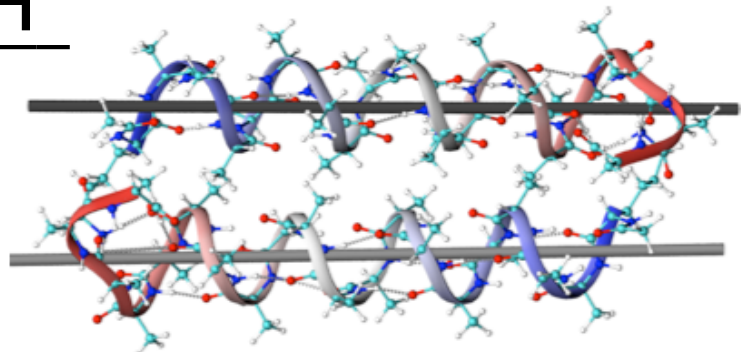
1) "Just helical" models: Need C-terminus proton. No Amide-II shift.

Conformation prototypes vs. IRMPD: (2) Helical dimers

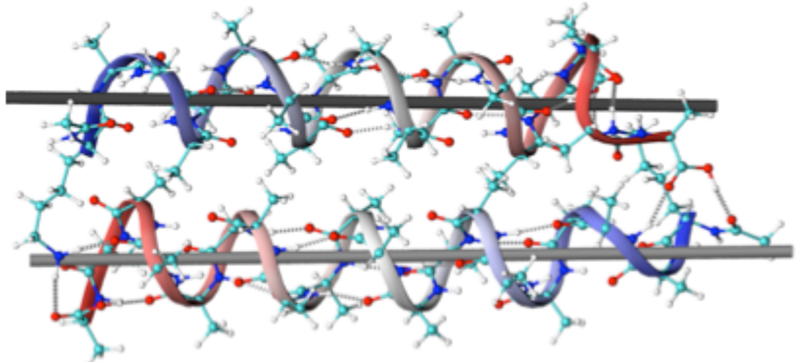
Theory-theory comparison, harmonic:

Ac-LysH$^+$-Ala$_{19}$ ↔ Ac-Ala$_{19}$-LysH$^+$

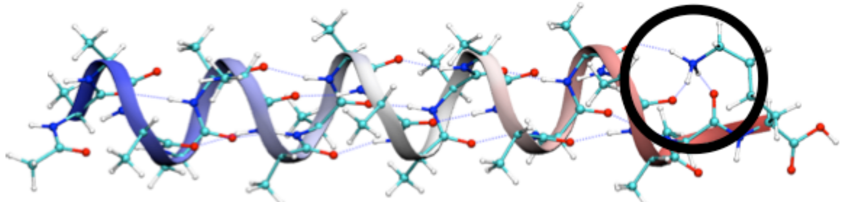




Ac-Ala$_{19}$-LysH$^+$, dimer 2



Ac-Ala$_{19}$-LysH$^+$, dimer 1

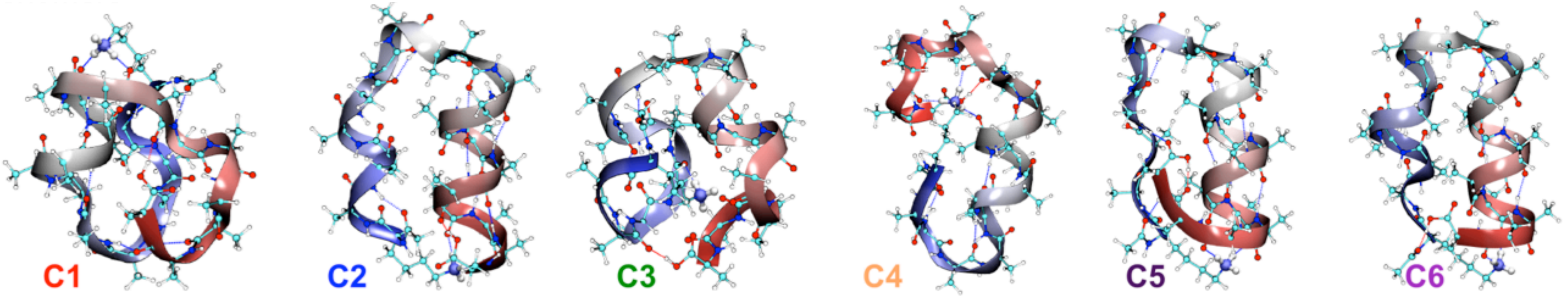


Ac-Ala$_{19}$-LysH$^+$
helix

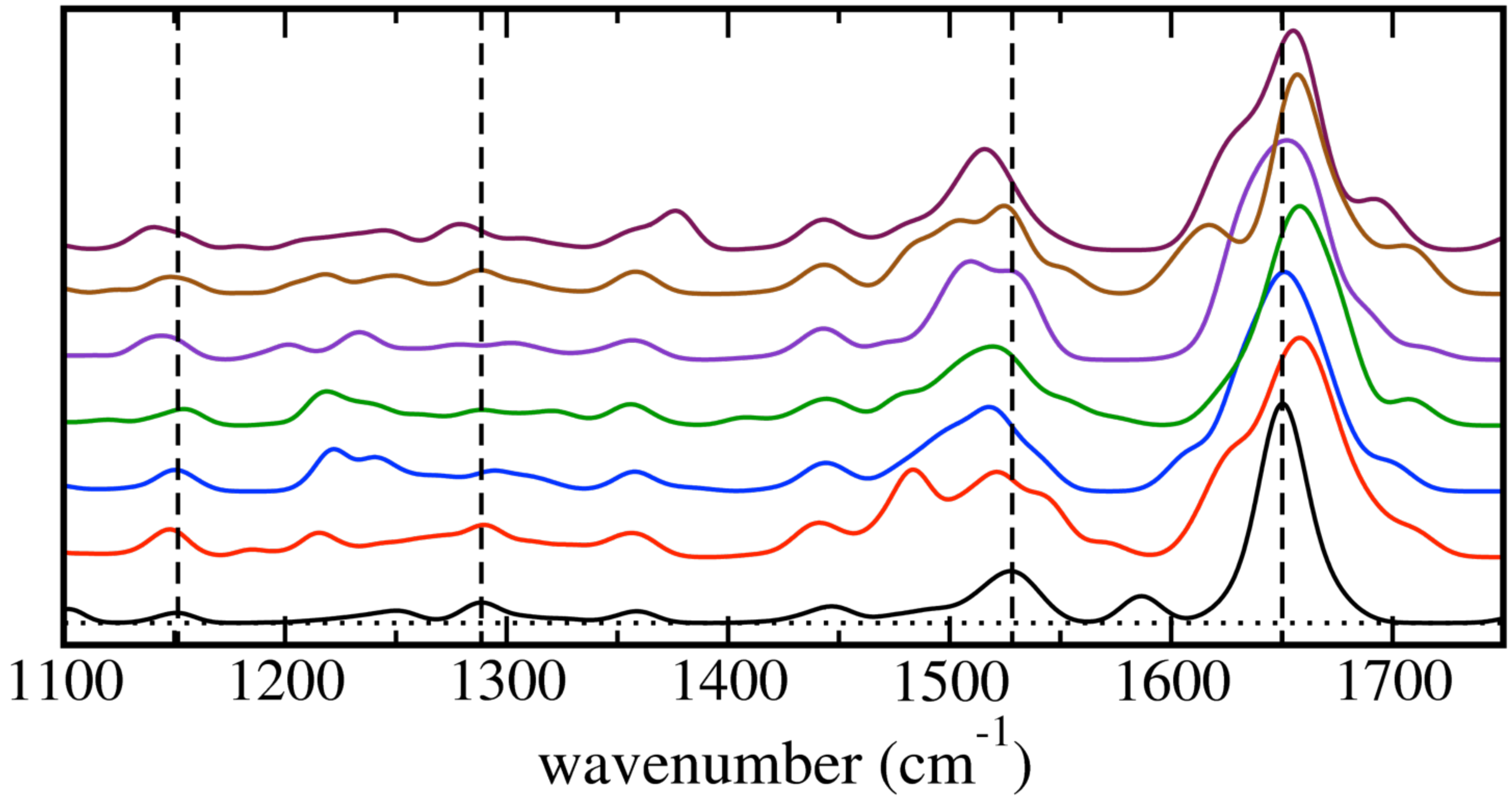
2) “Best dimer” models: Plausible, but no Amide-II shift.

Conformation prototypes vs. IRMPD: (3) *Any* monomers

Plenty of options close in (harmonic) free energy:

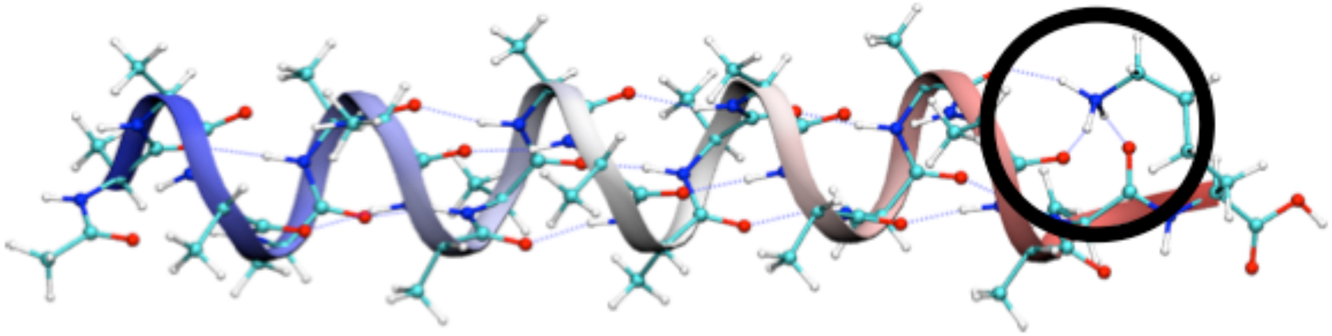


ΔF (300K) 0.0 0.06 eV 0.09 eV 0.15 eV 0.09 eV 0.12 eV



3) *Many* possible monomers.
Amide-II shift.

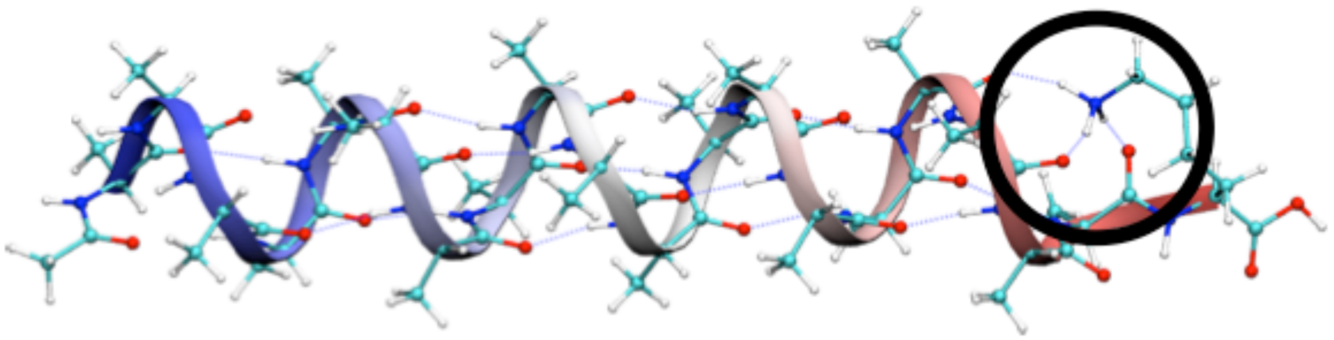
Upshot: Ac-Ala$_{19}$-LysH$^+$ vs. Ac-LysH$^+$-Ala$_{19}$



Ac-Ala$_{19}$-LysH$^+$:

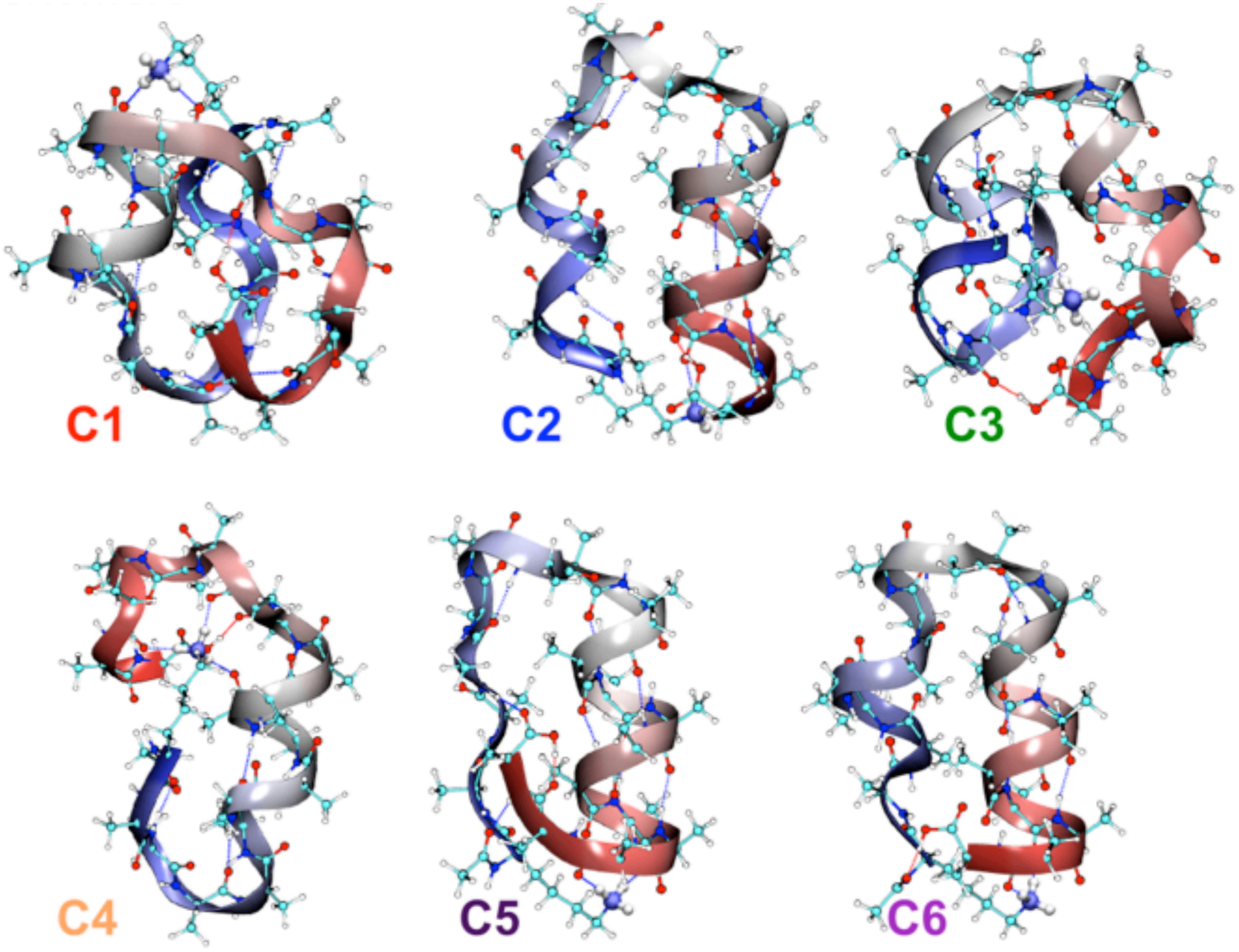
“Helix seeker” - Alanine likes helices, and proton at electrostatically favorable end

Upshot: Ac-Ala$_{19}$-LysH$^+$ vs. Ac-LysH$^+$-Ala$_{19}$



Ac-Ala$_{19}$-LysH$^+$:

“Helix seeker” - Alanine likes helices, and proton at electrostatically favorable end



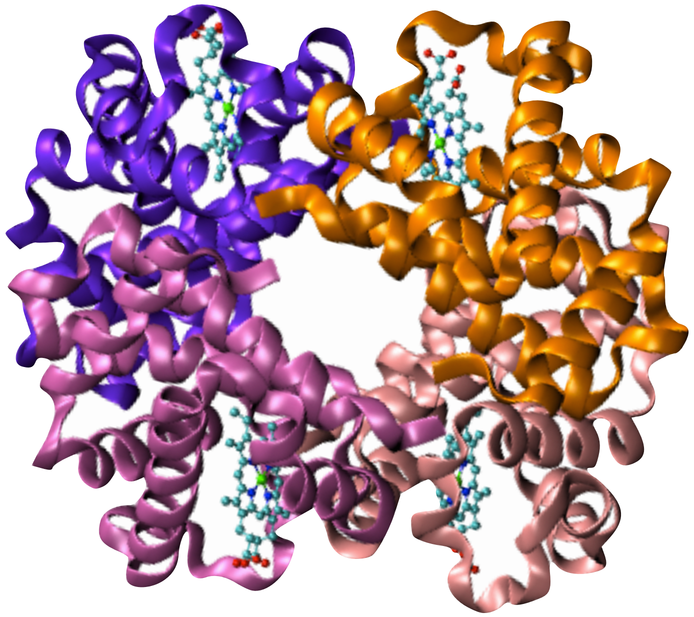
Ac-LysH$^+$-Ala$_{19}$

“Frustrated helix seeker” - Alanine likes helices, but proton at “wrong” end

→Mix of energetically similar “bent” helix segments that twist proton to “right” end of helix explains spectra!

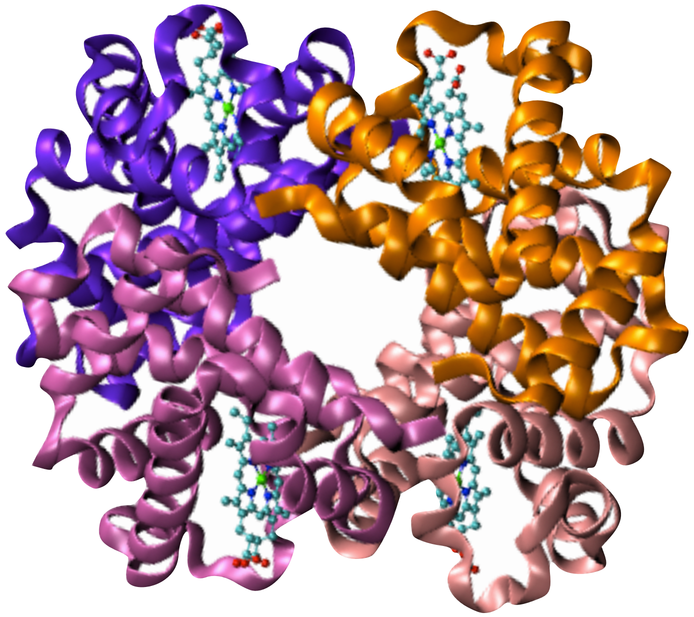
Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)



Cold, not so dense world: Challenges

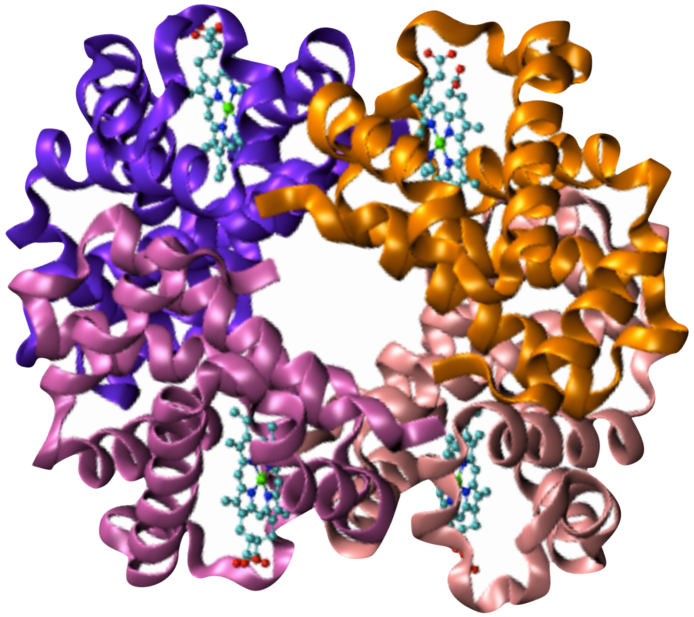
Hemoglobin
("Real" protein)



"Trivial": System sizes, simulation times -
~100 picoseconds, ~1000 atoms is still low end

Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)

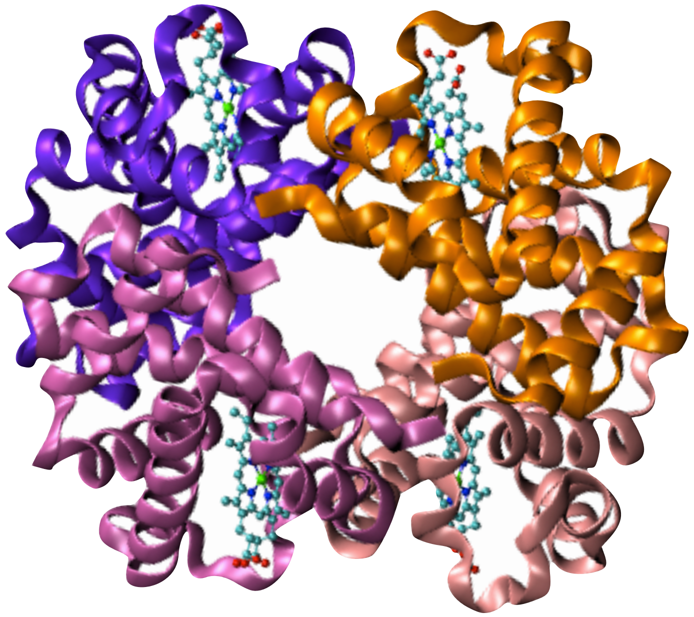


"Trivial": System sizes, simulation times - ~100 picoseconds, ~1000 atoms is still low end

"Less trivial": Meaningful free energy based predictions beyond harmonic approximation

Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)



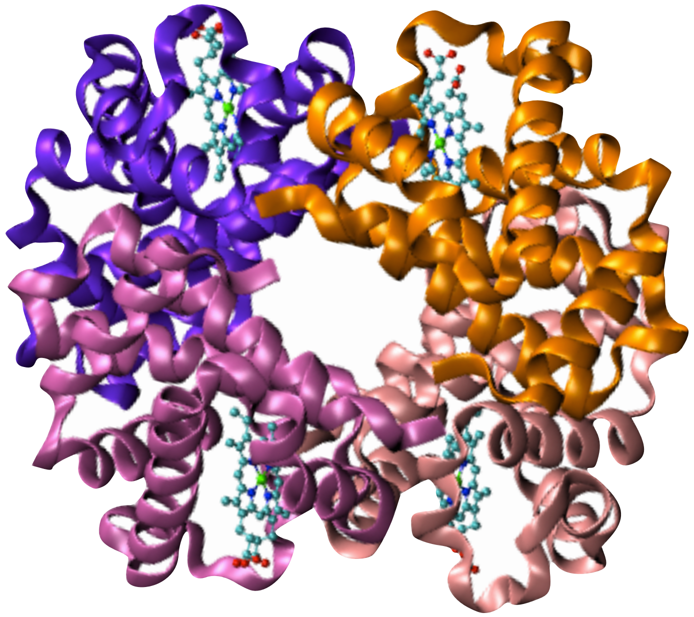
"Trivial": System sizes, simulation times - ~100 picoseconds, ~1000 atoms is still low end

"Less trivial": Meaningful free energy based predictions beyond harmonic approximation

"Less trivial": "Coarse-grained" techniques to sample "rare events" without wasteful simulations

Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)



"Trivial": System sizes, simulation times - ~100 picoseconds, ~1000 atoms is still low end

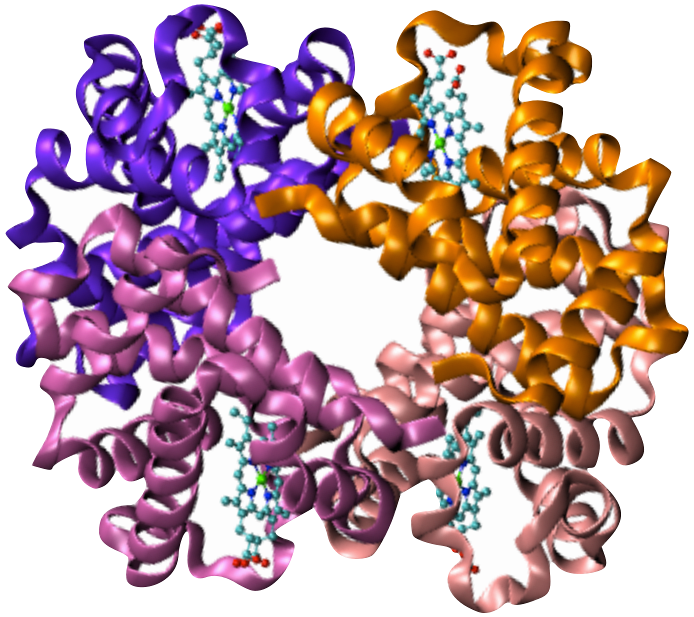
"Less trivial": Meaningful free energy based predictions beyond harmonic approximation

"Less trivial": "Coarse-grained" techniques to sample "rare events" without wasteful simulations

"Less trivial": Yet more accurate energies, but not (much) more expensive. GGA still barely good enough for ~few meV

Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)



"Trivial": System sizes, simulation times - ~100 picoseconds, ~1000 atoms is still low end

"Less trivial": Meaningful free energy based predictions beyond harmonic approximation

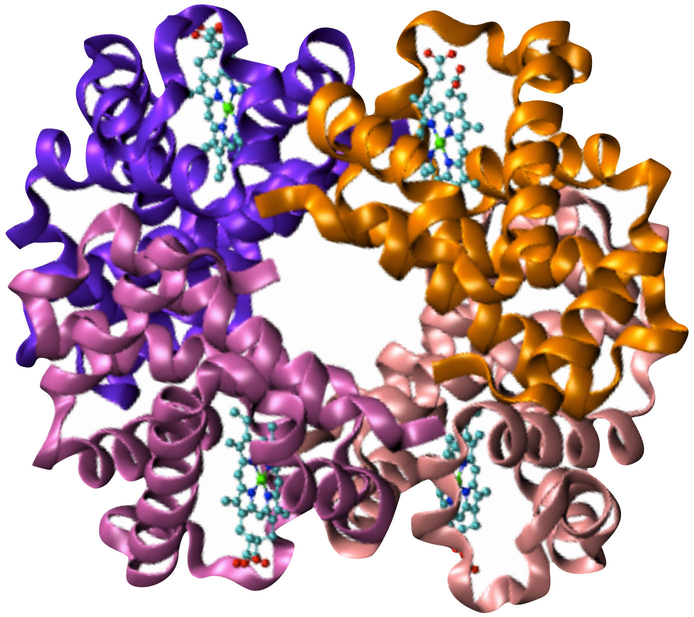
"Less trivial": "Coarse-grained" techniques to sample "rare events" without wasteful simulations

"Less trivial": Yet more accurate energies, but not (much) more expensive. GGA still barely good enough for ~few meV

"Not trivial": Nuclei (esp. hydrogen) are not classical particles. Electronic excitations?

Cold, not so dense world: Challenges

Hemoglobin
("Real" protein)



"Trivial": System sizes, simulation times - ~100 picoseconds, ~1000 atoms is still low end

"Less trivial": Meaningful free energy based predictions beyond harmonic approximation

"Less trivial": "Coarse-grained" techniques to sample "rare events" without wasteful simulations

"Less trivial": Yet more accurate energies, but not (much) more expensive. GGA still barely good enough for ~few meV

"Not trivial": Nuclei (esp. hydrogen) are not classical particles. Electronic excitations?

Most of these sound suspiciously like "warm, (somewhat) dense" as well!

Warm, dense: “Hopes” and challenges

“Hopes:”

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.

Warm, dense: “Hopes” and challenges

“Hopes:”

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.
- Still “pretty compact” basis sets compared to possible alternatives

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.
- Still "pretty compact" basis sets compared to possible alternatives

Challenges:

S.B. Carroll, Science 316, 1427 (2007)

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.
- Still "pretty compact" basis sets compared to possible alternatives

Challenges:

S.B. Carroll, *Science* 316, 1427 (2007)

- Many atoms per volume = high density of basis functions per volume. We suffer like all others.

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.
- Still "pretty compact" basis sets compared to possible alternatives

Challenges:

S.B. Carroll, Science 316, 1427 (2007)

- Many atoms per volume = high density of basis functions per volume. We suffer like all others.
- For *very dense* conditions, add basis functions for low-distance limit?

Warm, dense: "Hopes" and challenges

"Hopes:"

- Sound, accurate basis sets, elements 1-102
- Simple basic recipe, no implicit limit to all-electron accuracy by cutoff radii, pseudoization radii, etc.
- Still "pretty compact" basis sets compared to possible alternatives

Challenges:

S.B. Carroll, Science 316, 1427 (2007)

- Many atoms per volume = high density of basis functions per volume. We suffer like all others.
- For *very dense* conditions, add basis functions for low-distance limit?
- Are sums over high-lying states really the way?

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- "Properties": Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.
- LDA, GGA, van der Waals corrections, hybrid functionals, Hartree-Fock+MP2, RPA, *GW*, ...
- Massively parallel - scalable eigensolver ELPA

So where are we at?

- Sound, accurate basis sets, elements 1-102
- Basic numerical techniques for DFT
  Real-space integrals, Poisson equation, scalar relativity etc. *(Becke, Delley, Baerends, many others)*
- Non-periodic, periodic boundary conditions on exactly equal footing
- "Properties": Structure optimization, *ab initio* molecular dynamics, vibrations/phonons, spectroscopy, etc.
- LDA, GGA, van der Waals corrections, hybrid functionals, Hartree-Fock+MP2, RPA, *GW*, ...
- Massively parallel - scalable eigensolver ELPA

Thank you!