Hybrid quantum simulations of biomolecules: the role of copper in neurodegenerative diseases

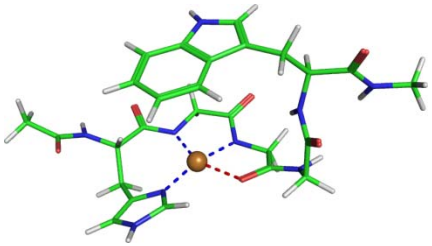
J. Bernholc, M. Hodak, W. Lu, F. Rose
Center for High Performance Simulation and Department of Physics
North Carolina State University, Raleigh,

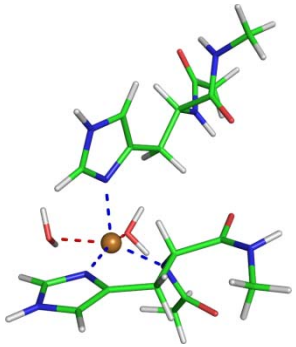
I. Introduction to molecular simulations in biology

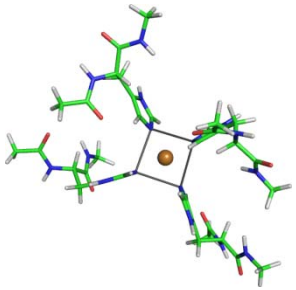
II. DFT/approximate-DFT hybrid method for biosimulations
- Orbital-free linear-scaling DFT for the solvent
- Full DFT for the biomolecule and its first solvation shells

III. Applications
- Prion protein and the role of copper
- Cu-induced misfolding in Parkinson disease

IV. Summary

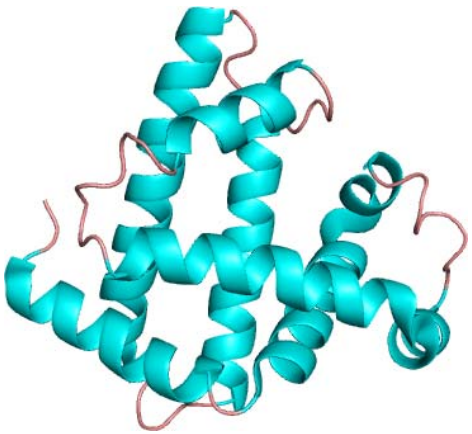


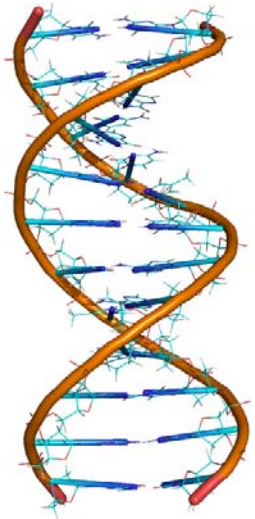




Biosimulations

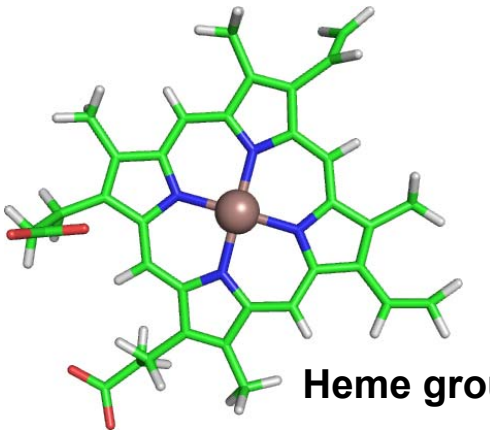
- Computer simulations involving organic molecules
  - Proteins, lipids, DNA, etc.
  - Biological macromolecules
    - amino acids, nucleotides, etc.
- Computational challenges
  - Size
    - Thousands, even millions of atoms
  - Time scale
    - Processes involving structural changes (folding etc.) take place over μs or more
  - Solvation
    - Increases size of the system
      - Solvent usually outnumbers solute



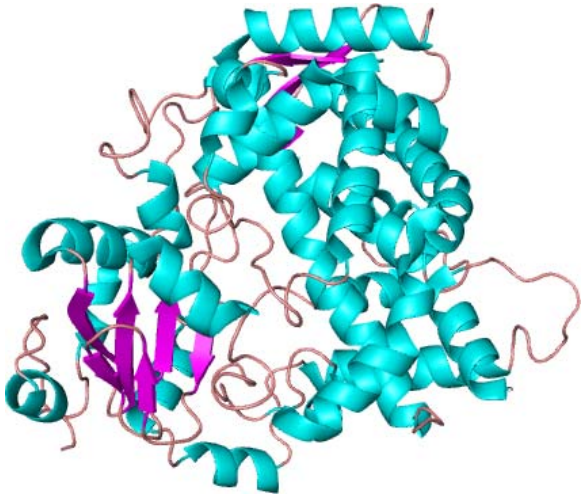


Biosimulations: methods

- Quantum molecular dynamics
  - can describe chemical processes
    - bond formation, bond breaking, charge transfer
  - ~10^2-10^3 atoms
  - timescale: ps
  - fragments and localized processes
- Classical molecular dynamics
  - force fields
  - cannot describe chemical processes
  - ~10^6 atoms
  - timescale: ns
- Coarse grained and lattice models
  - more simplified description
  - can describe protein folding



Heme group

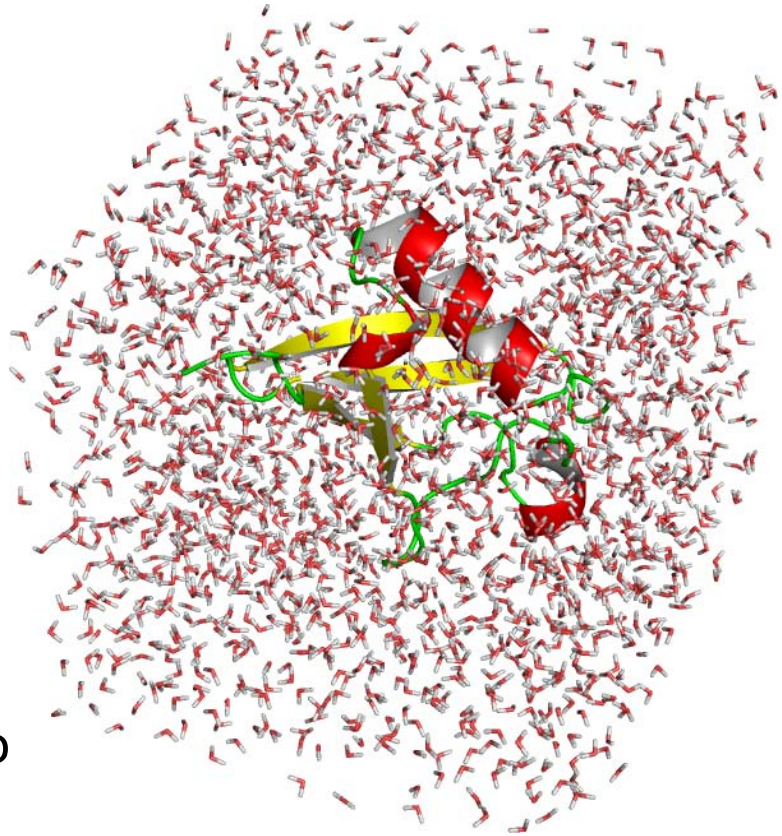


P4503A4 Oxidoreductase

Solvation in biosimulations

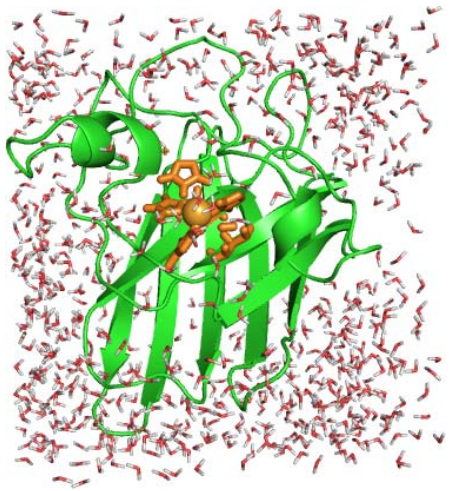
- Implicit
  - Surrounding solvent treated as continuum
  - Efficient
  - Incorrect results in certain cases
    - protein folding pathways etc.
- Explicit
  - More accurate than implicit
  - Solvent outnumbers solute
- DFT simulations
  - Implicit solvation is often used
    - Polarizable continuum model (PCM)
  - Simplified treatment of solvent has to be used for explicit solvation
    - Hybrid QM/MM approach

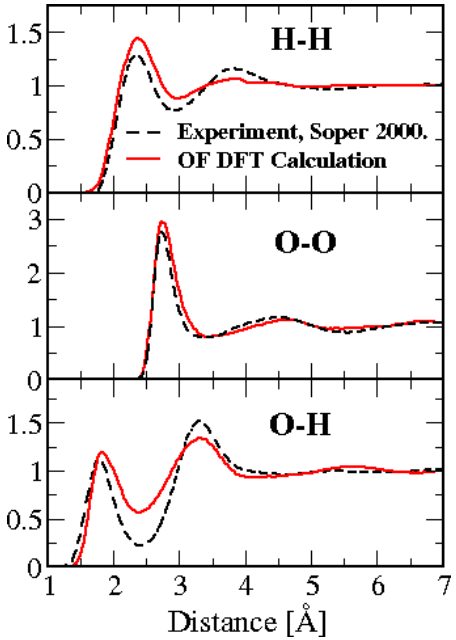
Fully solvated *ubiquitin*
Water: 83% of atoms



Hybrid DFT/DFT calculations

- QM/MM
  see e.g. **Warshel, Levitt, J. Mol. Biol. 1976.**
  - QM: chemically active part
    - small part of the system
  - MM: rest of the system
  - QM-MM interface
    - special interaction terms have to be defined
      - electrostatic, repulsion, etc.
      - MM parameters are used
      - problem with exchange of atoms
- Our approach: hybrid DFT/DFT method
  Hodak, Lu, Bernholc, JCP, 2008.
  - Kohn-Sham DFT for the chemically active part of biomolecule
  - Orbital-free DFT for the solvent in the system





Real-space Multi-Grid method (RMG)

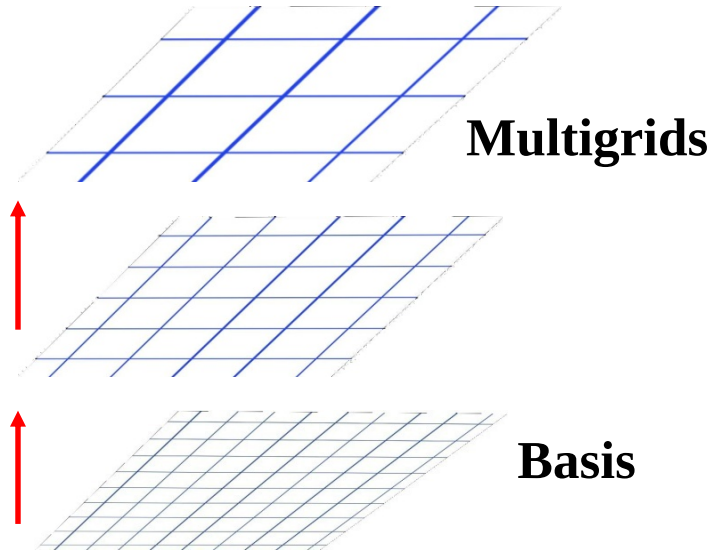
- Density functional equations solved directly on the grid
- Multigrid techniques remove instabilities by working on one length scale at a time
- Convergence acceleration and automatic preconditioning on all length scales
- Non-periodic boundary conditions are as easy as periodic
- Compact "Mehrstellen" discretization

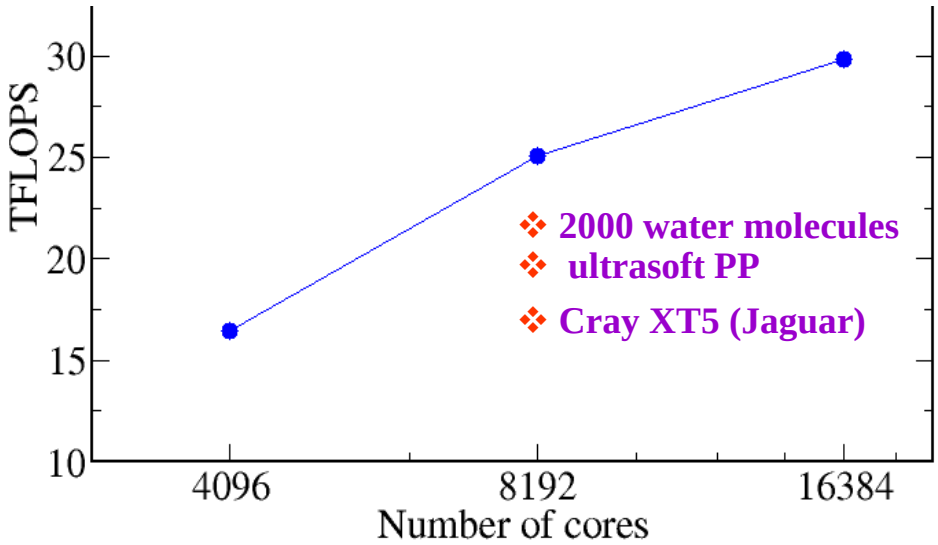
$$A[\phi_i] + B[(V_{eff} + V_{NL})\phi_i] = \varepsilon_i B[S\phi_i]$$

- Allows for efficient massively parallel implementation

❖ *See E. L. Briggs, D. J. Sullivan and J. Bernholc Phys. Rev. B 54, 14362 (96).*

❖ Ultrasoft pseudopotentials: *M. Hodak, S. Wang, W. Lu and J. Bernholc, PRB 76, 085108 (07)*

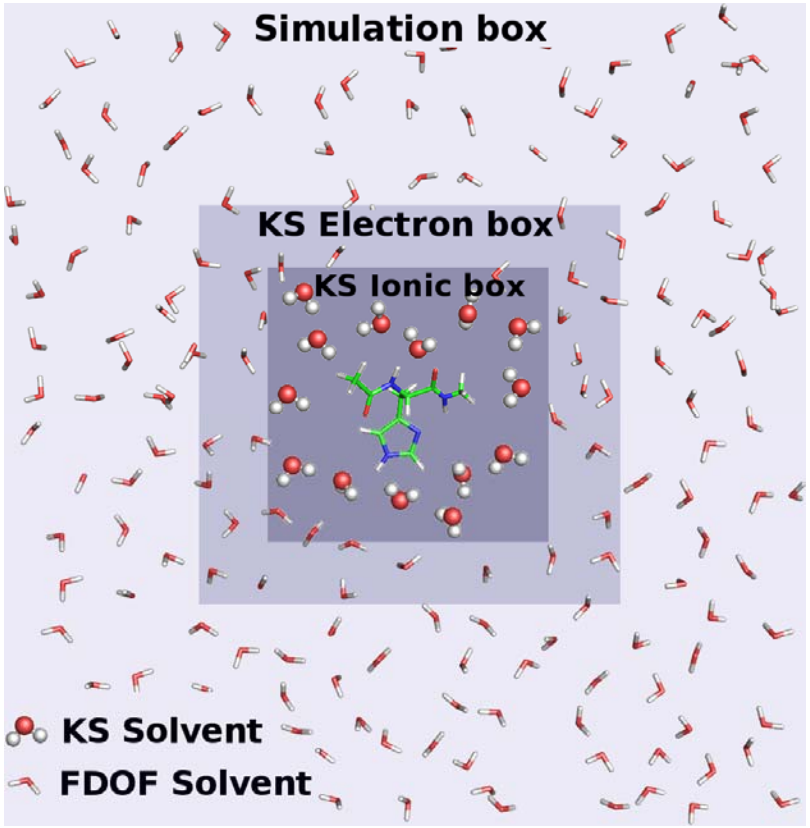




Kohn-Sham DFT/Orbital-free DFT calculation

- Kohn-Sham DFT for chemically active part
  - KS Ions must be located inside KS Ionic Box
    - biomolecules must be trimmed at the boundary
  - KS equations solved in KS electron box with zero boundary conditions (non-periodic)
    - extra space needed for wavefunctions and density to decay to zero
- Orbital-free DFT for the rest of the system (Lee, Lee, Parr functional, 1991)
  - takes 5% or less even when treating 100× more atoms
  - Generalization of Gordon-Kim approach for closed-shell molecules

    Gordon and Kim, JCP, 1972, see also Barker and Sprik, Mol. Phys. 2003



Orbital-free DFT

- **Minimize energy functional without using orbitals**
- **Efficient, linear scaling**
- **Mixed success in reproducing K-S electron densities**
- **Kinetic energy functional has to be explicit in *ρ* (Thomas-Fermi)**
- **Frozen-density orbital-free model for liquid water**

  LLP KE functional (TF + GGA) **Lee, Lee, Parr, PRA, 1991.**

$$T_{LLP} = C_{TF} \int_{\Omega} \varrho_{TF}^{5/3} \left(1 + \frac{0.0045x^2}{1 + 0.0253x \sinh^{-1} x} \right) d\mathbf{r}, \; x = 2^{1/3} \frac{|\nabla \varrho_{TF}(\mathbf{r})|}{\varrho_{TF}^{4/3}(\mathbf{r})}$$

  - **Individual molecules have fixed geometries and frozen densities**
  - **Ultrasoft and norm-conserving pseudopotentials (only local parts needed)**
  - **Charge density given by gaussians**
    - **Constants determined from dipole moment and RDF**

      see also Barker and Sprik, Mol. Phys. 2003

- **Electronic energy functional**

$$E_{FDOF}[\varrho] = T_{LLP}[\varrho] + E_H[\varrho] + E_{XC}[\varrho] + E_{PPlocal}[\varrho] + E_{ii}$$

Hybrid KS DFT/OF DFT energy

- System consisting of KS and OF ions
  - Total density: $\varrho_{tot}(\mathbf{r}) = \varrho_{OF}(\mathbf{r}) + \varrho_{KS}(\mathbf{r})$
  - Total kinetic energy

$$T_{tot}[\varrho_{OF}, \varrho_{KS}] = T_{OF}[\varrho_{OF}] + T_{KS}[\varrho_{KS}] + T_{nadd}[\varrho_{OF}, \varrho_{KS}]$$

  - Non-additive kinetic energy can be estimated

Wesolowski and Warshel, JPC 1993

$$T_{nadd}[\varrho_{OF}, \varrho_{KS}] = T_{OF}[\varrho_{tot}] - T_{OF}[\varrho_{OF}] - T_{OF}[\varrho_{KS}]$$

  - Hybrid electronic energy functional

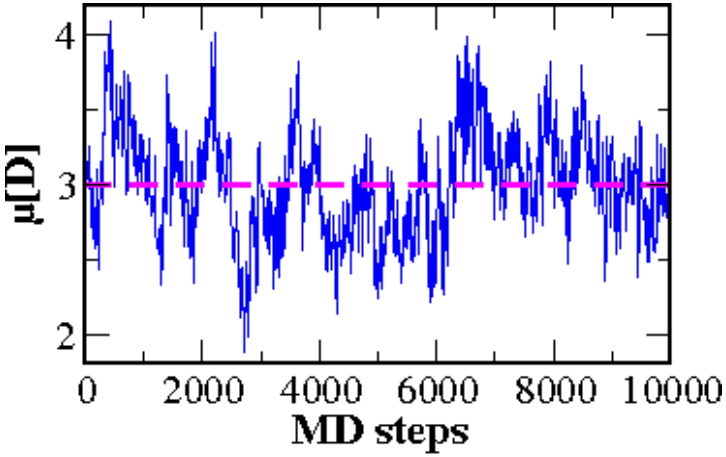
$$E_{Hybrid}[\varrho_{OF}, \varrho_{KS}, \{\mathbf{R}_i\}_{OF}, \{\mathbf{R}_i\}_{KS}] = T_{tot}[\varrho_{OF}, \varrho_{KS}] + E_H[\varrho_{tot}]$$

$$+E_{XC}[\varrho_{tot}] + E_{PP_{nonloc}}[\varrho_{KS}, \{\mathbf{R}\}_{KS}] + E_{PPloc}[\varrho_{tot}, \{\mathbf{R}\}] + E_{ii}[\{\mathbf{R}\}]$$

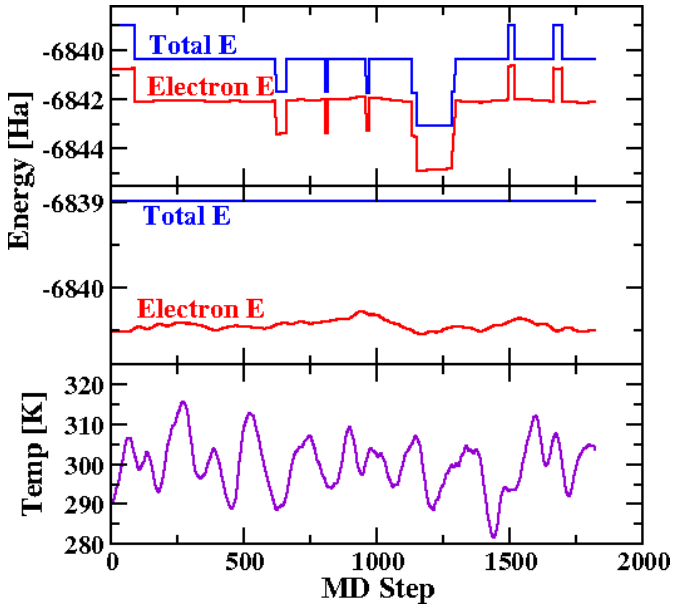
  - Leads to modified KS equations for ρ_{KS}
  - Interactions between subsystems on quantum (DFT) level

Solvent exchange and solvation effects

- Water molecule polarization:
  - Gas phase: 1.85 D
  - Liquid phase:
    - Experiments: 2.9-3.15 D
    - ab initio calculations: ~3.0 D
      - Silvestrelli, Parrinello, JCP 3572, 1999; PRL 3311, 1999.
    - **Hybrid: 3.0 D**



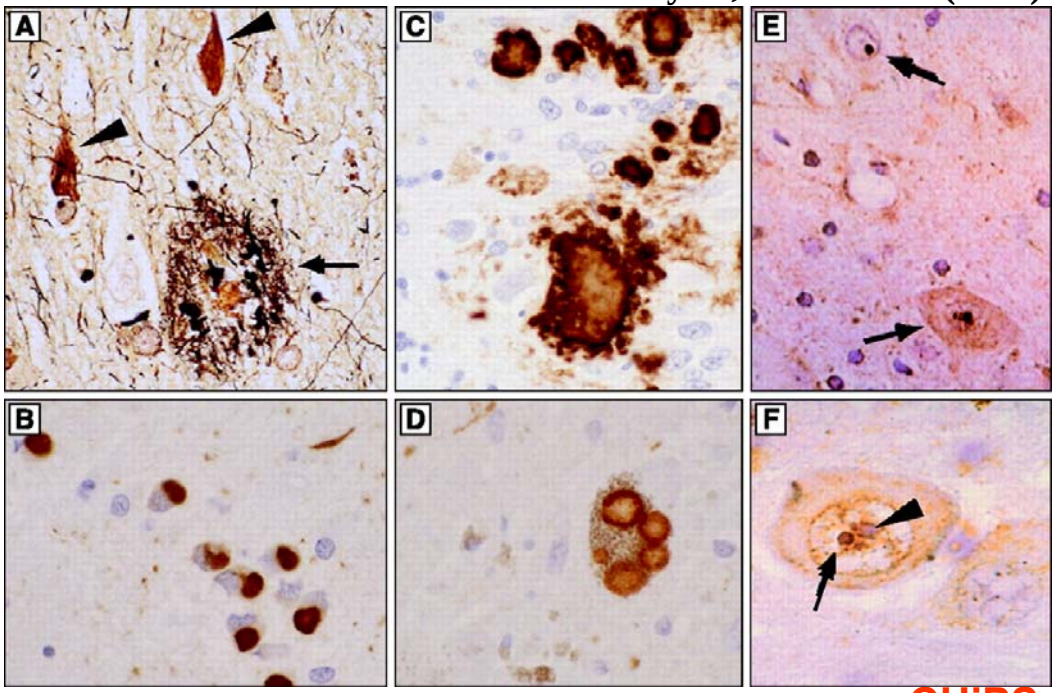
- Exchange of solvent between OF and KS regions
  - smooth energy curves can be obtained
  - energy conserving MD



Applications: Neurodegenerative diseases

- **Diseases related to neuronal damage**
  - **characterized by insoluble deposits (plaques, amyloids) composed of proteins**
  - **Protein-only hypothesis: Diseases are caused by improperly folded proteins unable to perform their functions**
  - **New class of infectious diseases, no DNA or RNA involved**
  - **Metals ions may play a role**
  - **Alzheimer's disease (A)**
  - **Pick's disease (B)**
  - **Prion disease(C)**
  - **Parkinson's disease (D)**
  - **Machado-Joseph's disease (E,F)**
  - **Huntington disease**

❖Taylor, Science 296 (2002).



Pathogenesis in Neurodegenerative diseases

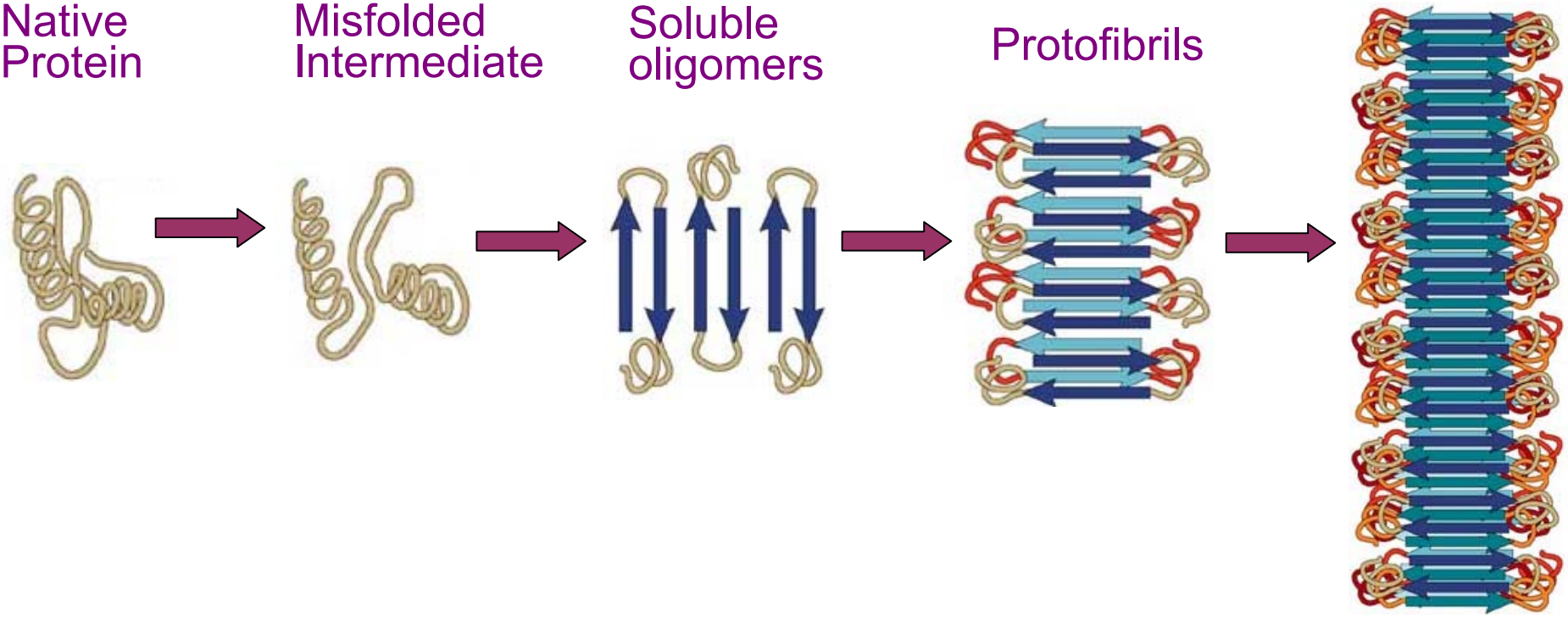
Amyloid Fibrils
- Cross-β sheet
- β-strands perpendicular to fiber axis

Native Protein

Misfolded Intermediate

Soluble oligomers

Protofibrils

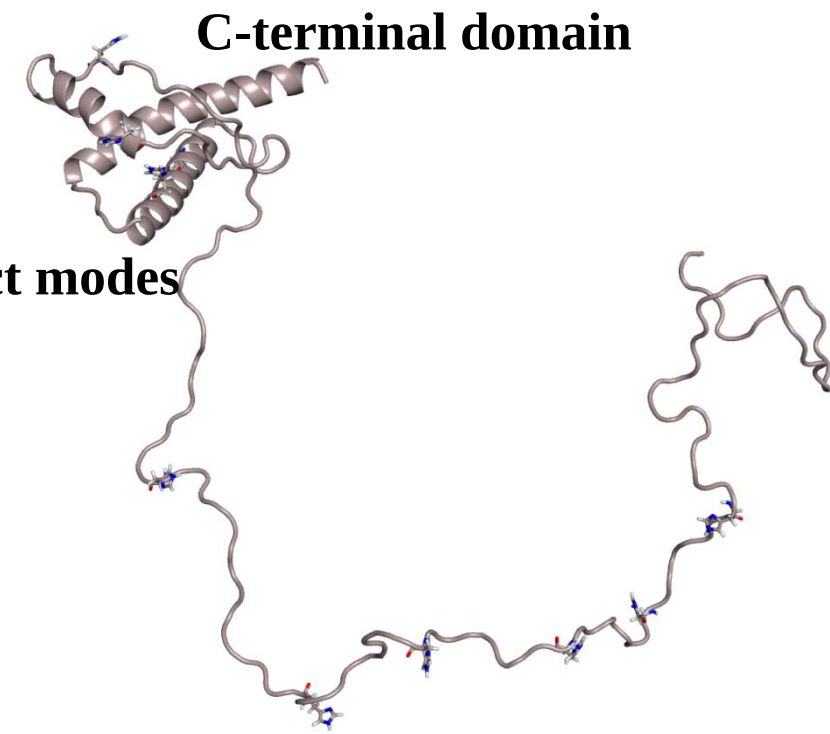


Exact nature of neurotoxic species is unknown

C. Soto,
Nat. Rev. Neurosci, 2003.

Prion protein

- **Involved in prion diseases**
  - **mad cow disease (cattle), scrapie (sheep, goat)**
  - **Creutzfeldt-Jakob disease, kuru (human)**
- **Prion protein (PrP)**
  - **Folded and unfolded regions**
  - **Unknown normal function**
  - **Efficiently binds Cu(II)**
    - **Cu may be bound in three distinct modes**
    - **May be related to PrP's function**
      - **Copper buffer?**
  - **Role of copper in prion diseases?**



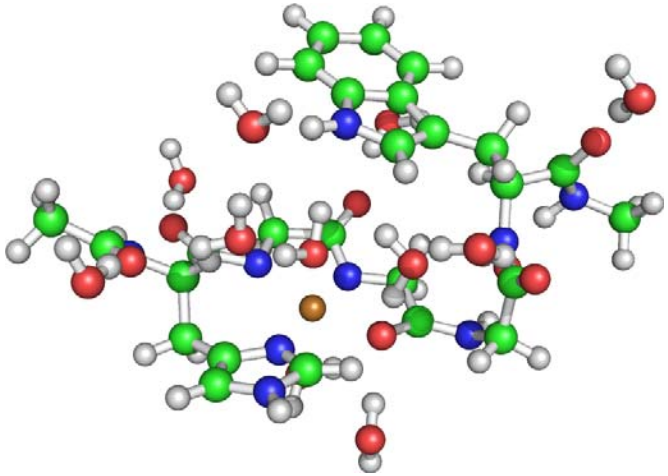
Hybrid calculation setup

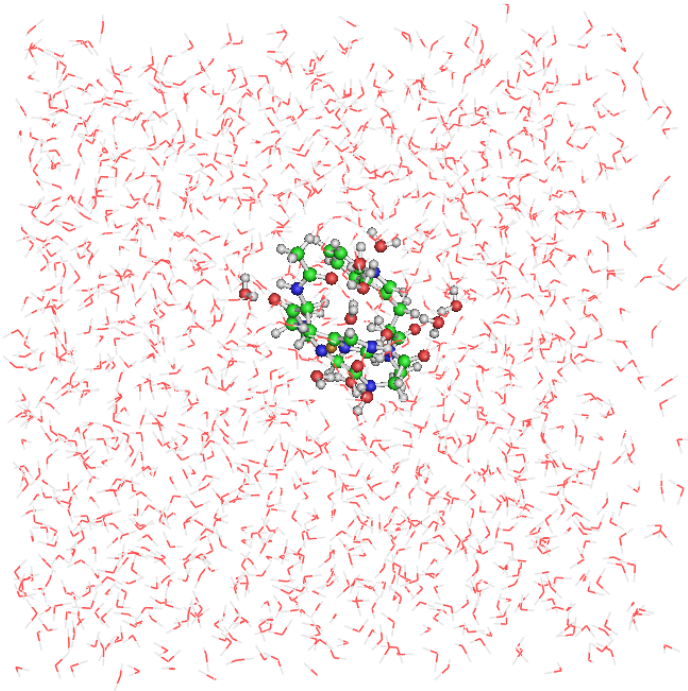
Quantum system:

- PrP-Cu binding fragment
- 1st solvation shell around Cu

Hybrid system:

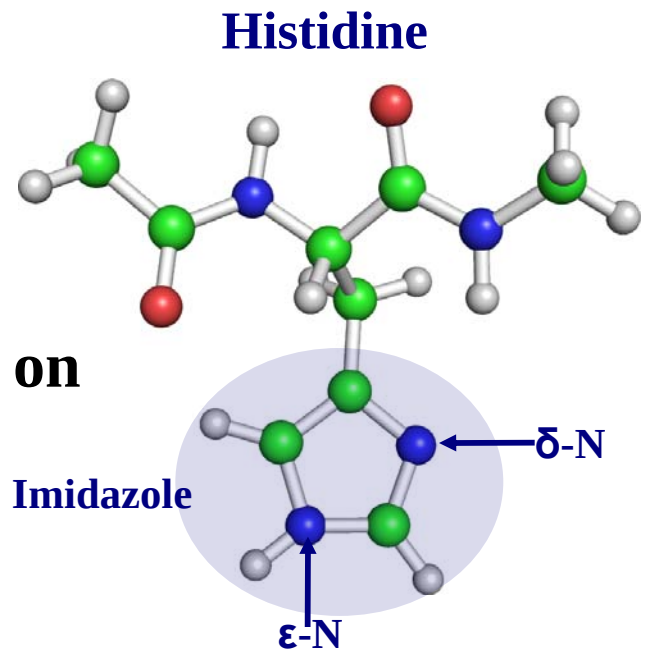
- Full solvation
- 2000-3000 water molecules treated by OF DFT



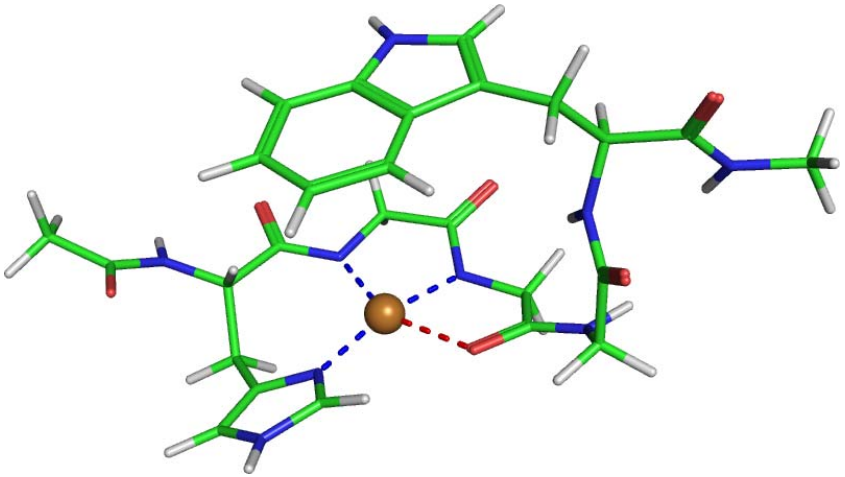


Copper binding to PrP

- **High affinity binding sites located in octarepeat region**
  - **copper anchored by histidine residues**
  - **octarepeat contains four histidines**
- **Three distinct binding modes depending on copper concentration**

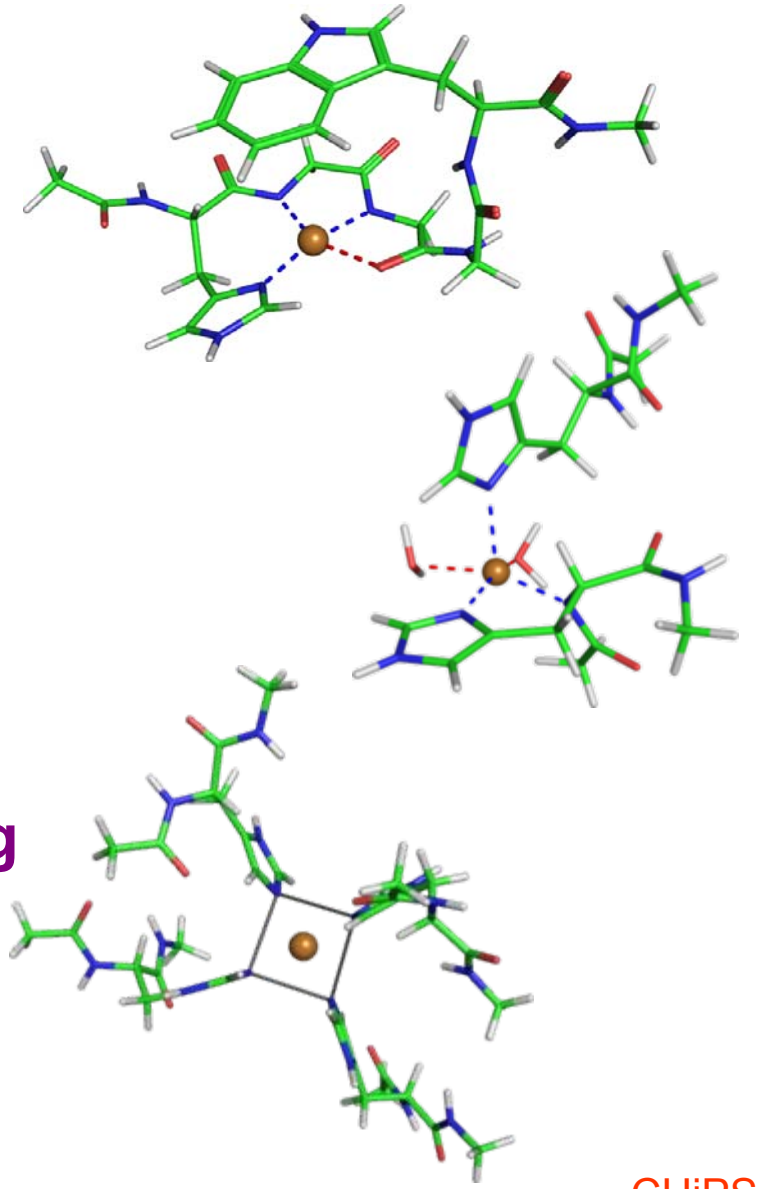


- **High concentration binding**
  - **Sequential residues HGGGW**
  - **One histidine involved**
  - **Four Cu ions can be bound**
  - **Three N and 1 O**
  - **Planar binding**



Concentration dependent Cu^{2+} binding to PrP

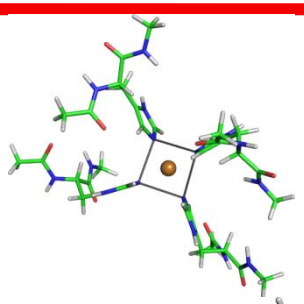
- **High Cu^{2+}concentration**
  - Sequential residues HGGGW
  - One histidine involved
  - Four Cu ions can be bound
  - ~ 150 KS atoms, 3000 OF water molecules
- **Medium Cu^{2+}concentration**
  - Two histidines are involved
    - One in axial position
  - Two in-plane waters
  - Two Cu ions can be bound
- **Low Cu^{2+}concentration binding**
  - Four histidine are involved
    - All histidines in octarepeat domain
  - Coordinating atoms arranged on square
  - One Cu ion can be bound



Energetics of PrP's function as a copper buffer

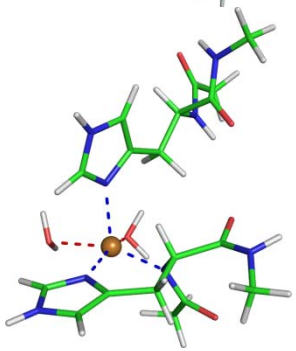
Cu concentration

Low: 1 Cu bound @ 5.22 eV



Intermediate: 2 Cu bound, each @ 2.75 eV

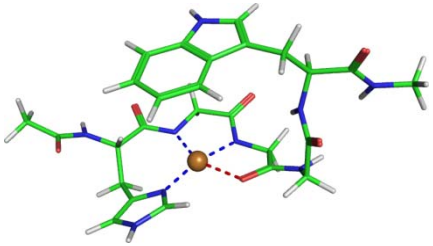
Energy balance: 2*2.75 eV – 5.22 eV = 0.28 eV
Entropy*T increase (at 300 K): 0.16 eV
calculated using Schlitter's formula, CPL 1993.
Addition of 2nd Cu is **exothermic**



High: 4 Cu bound, each at 1.76 eV

Energy balance: 4*1.76 eV – 2*2.75 eV = 1.54 eV
Entropy*T increase relative to previous case: 0.23 eV

Addition of more Cu is **exothermic**

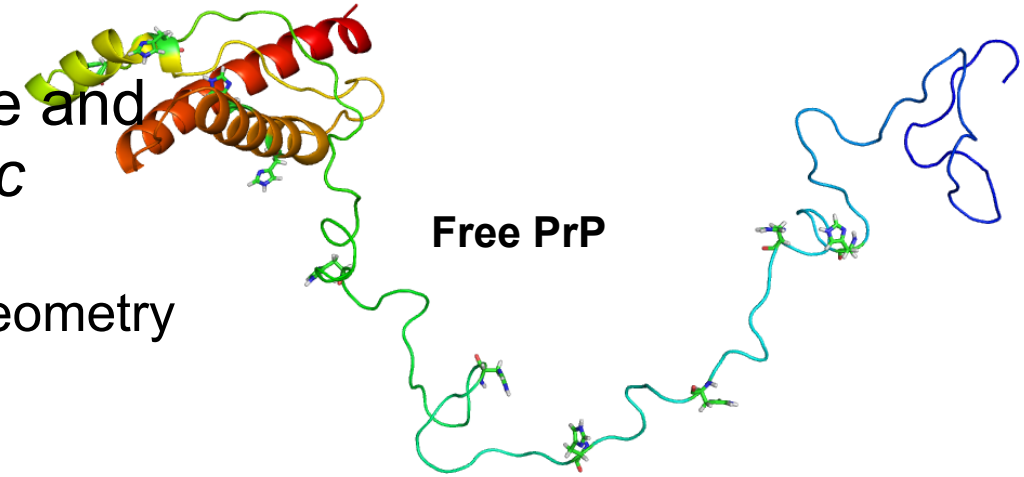


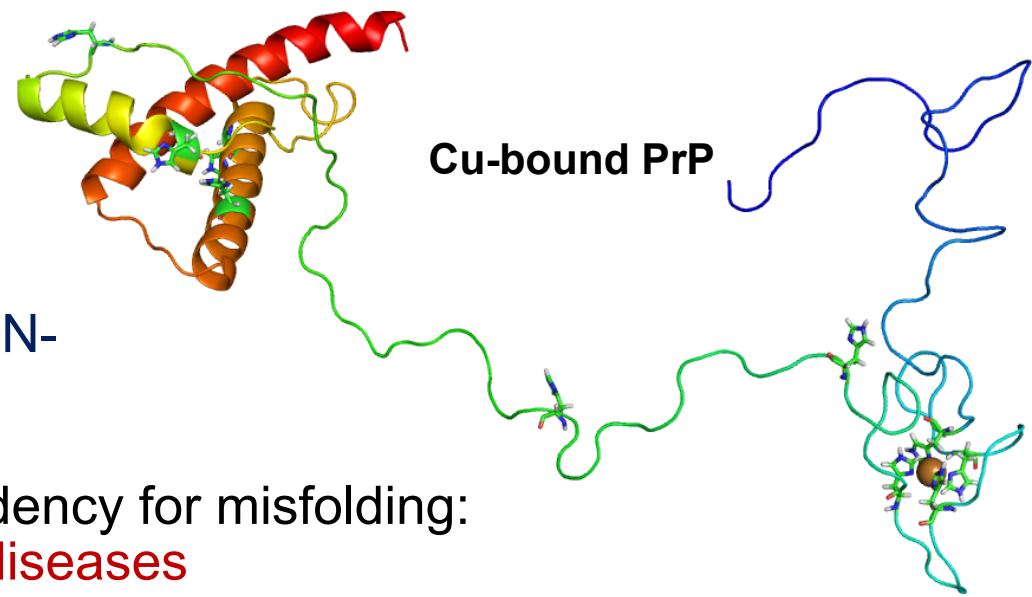
Conclusion: It is favorable to bind more copper ions when they are available

PrP can function as a copper buffer !

Effect of Cu on full-length PrP

- Classical MD (NAMD) on free and Cu-bound PrP at low Cu *conc*
  - Cu modeled by either fixing the binding site to the calculated geometry or using Cu-bond "springs"
- Flexible N-terminal part
  - Free PrP
    - only temporary turns
  - Cu-bound PrP
    - Stable turns around Cu
    - Metastable short β-hairpins
  - Cu enhances stability of whole N-terminal domain

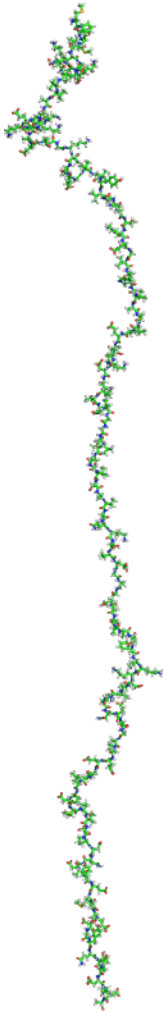




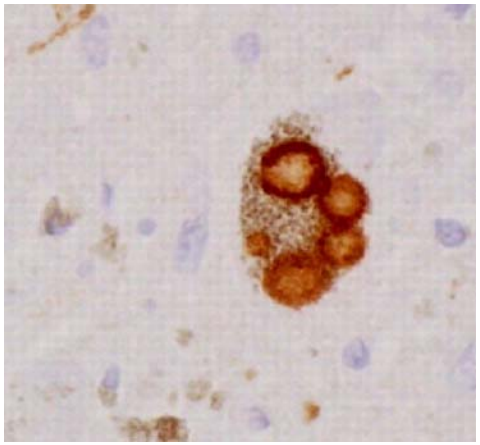
Higher stability → lesser tendency for misfolding:
Protective role of copper in prion diseases

Hodak, Chisnell, Lu, Bernholc, Proc. Nat. Acad. Sci. 2009

Parkinson's disease



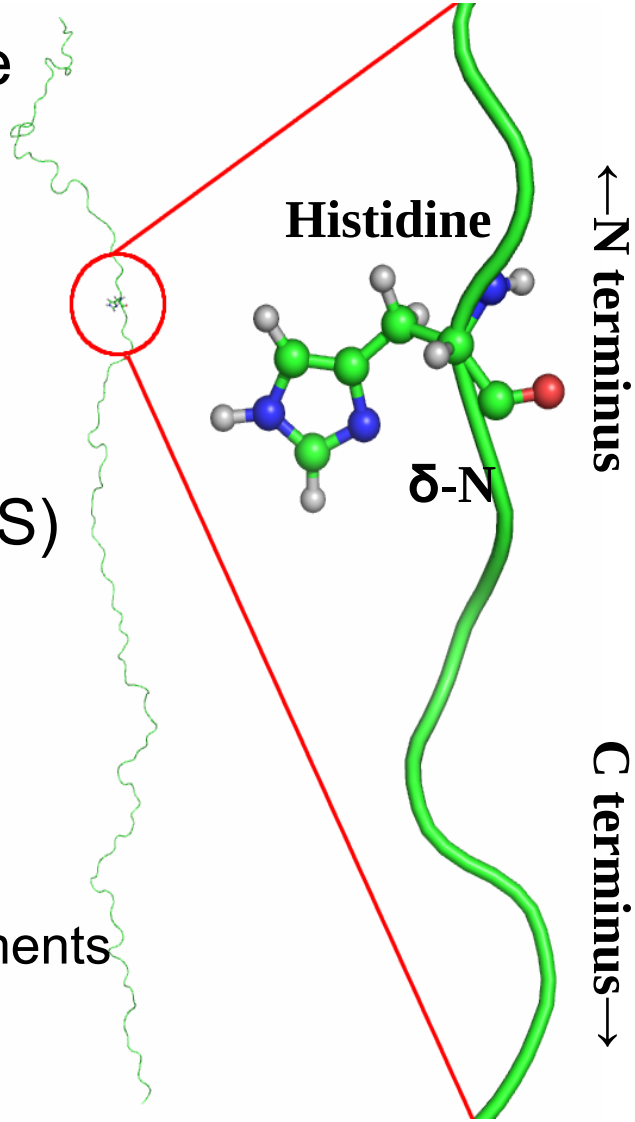
- Involves α-synuclein
  - 140 aminoacids
  - misfolded proteins form Lewy bodies
  - present in neurons of the central nervous system
  - natively unstructured
  - Unknown function
    - May affect neuron plasticity and neurotransmitter release
  - binding of metal ions causes misfolding and subsequent oligomerization
    - Cu(II), Al(III), Fe(III), Co(III), Mn(II)
    - Cu(II) most effective



Taylor, Science 296 (2002).

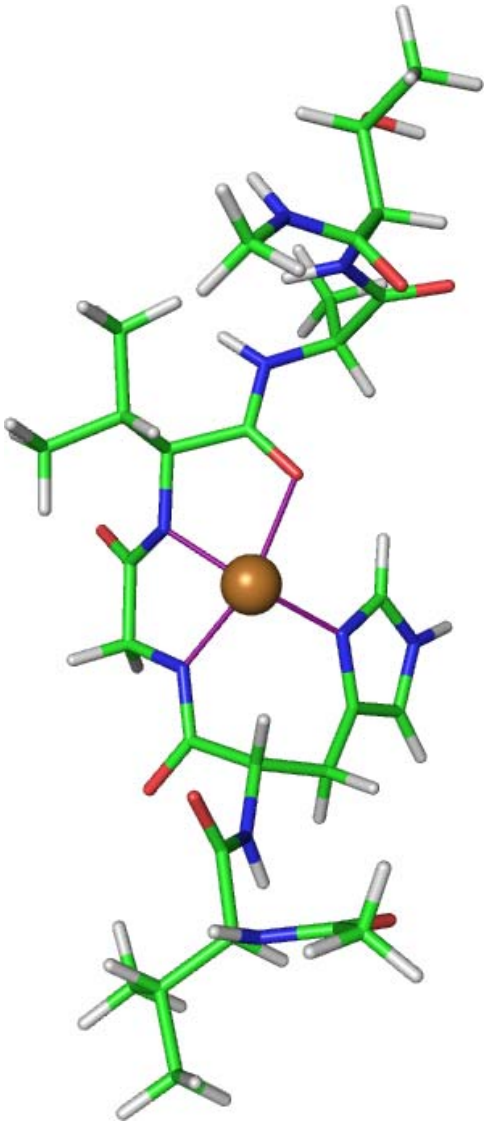
Cu^{2+} and α-synuclein

- Cu^{2+} is the most effective metal ion in the formation of α-synuclein plaques
- Cu^{2+} is found in α-synuclein plaques of Parkinson's disease patients and is speculated to be causal
- NMR Structural characterization of Cu binding to α-synuclein (Rasia et al., PNAS)
  - Preferentially bound to δ-N of Histidine 50
  - Secondary binding occurs at N terminus
- Our Objective:
  - Validate Cu binding as a viable progenitor of stable β-sheet formation
  - Explore the extent of available β-sheet alignments

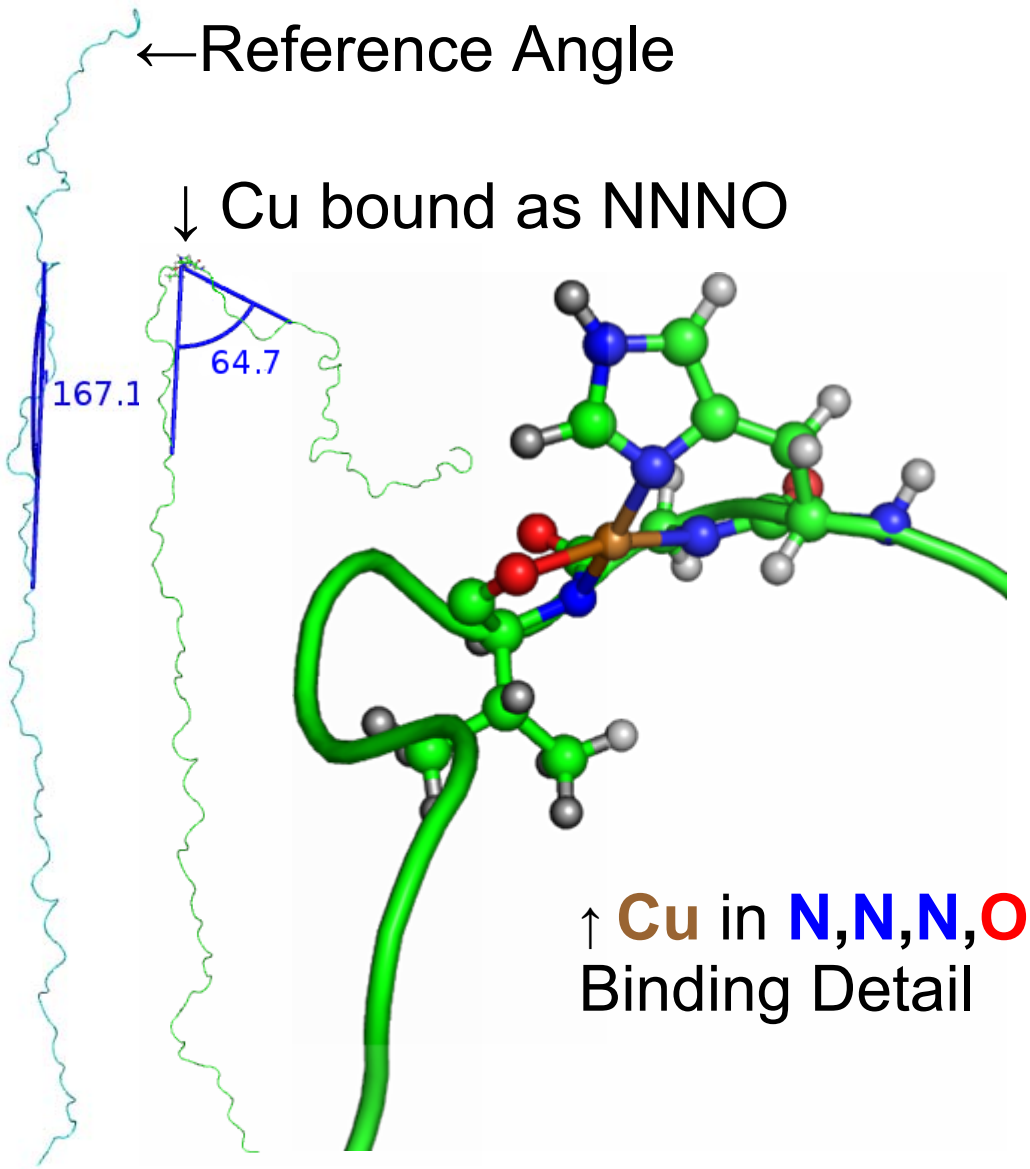


Cu(II) binding to α-synuclein fragment

- **Experiment: NMR**
  - **primary binding site: His50**
  - **residues 49-54 most affected**
- **Our model:**
  - **residues 49-54: VHGVAT**
  - **Cu(II) anchored by δ-N on His50 imidazole**
  - **deprotonated N and O in protein backbone**



NNNO Binding Geometry



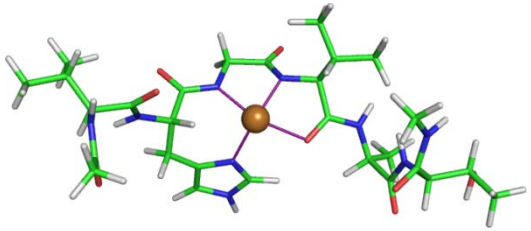
- *ab-initio* fragment consists of 91 atoms
- Backbone deforms from reference 167° angle to 65° angle
- Neutrally charged system
- NNNO is energetically preferred

NNNO	**0.0 eV (ref)**
NNOO	**+ 0.82 eV**
NONO	**+ 1.56 eV**

Cu(II)-bound α-synuclein

Free

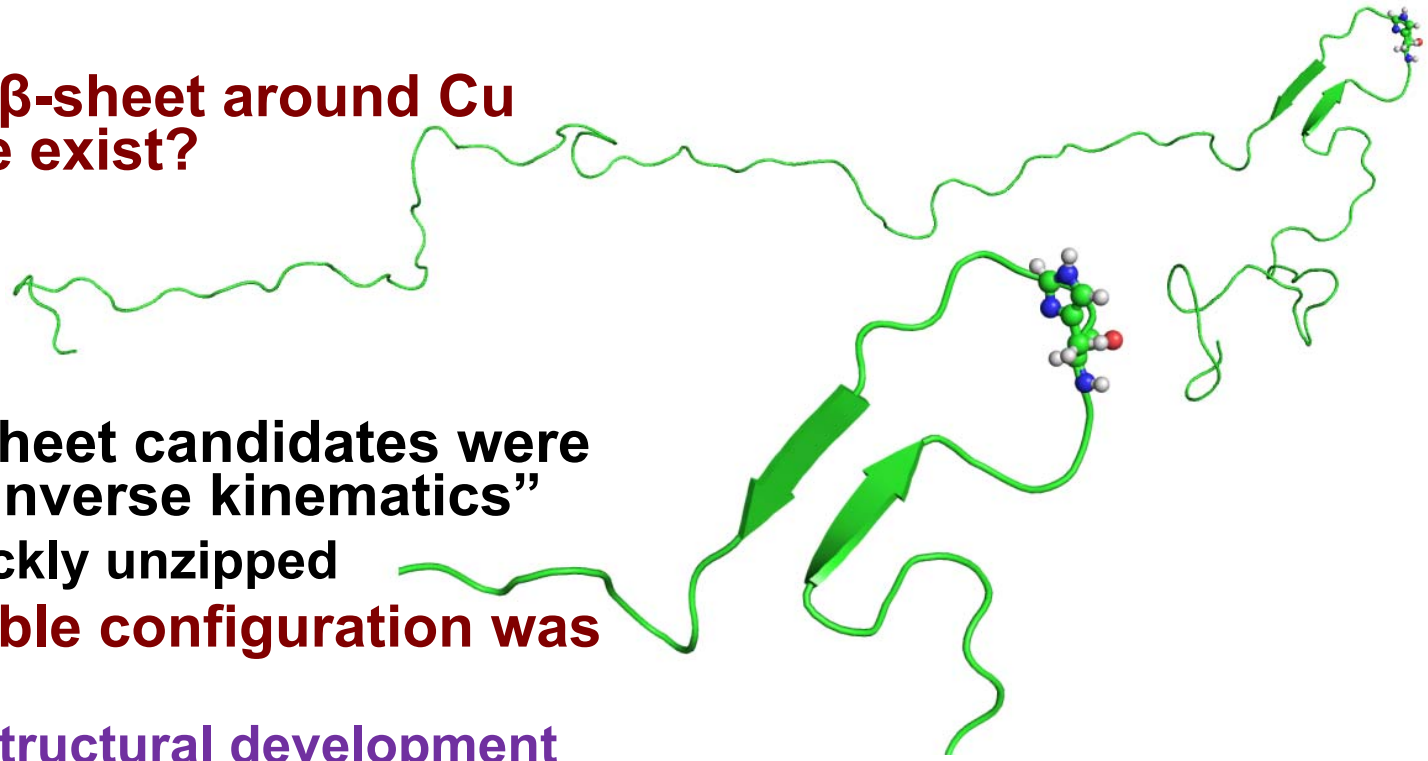
- Free α-synuclein
  - mostly linear
  - angle around His50: 167°
- Cu-bound α-synuclein
  - imposing binding site geometry introduces turn to the protein structure
  - angle around His50: 65°
  - continued development using classical MD (NAMD) shows hydrogen contacts around binding site and alignment of the backbone
    - Possible β-sheet development
    - Misfolded α-synuclein contains β-sheets !



Cu-bound

β-sheet in Cu-bound α-synuclein

- Can stable β-sheet around Cu binding site exist?



- Several β-sheet candidates were tested by "inverse kinematics"
  - Most quickly unzipped
- But... a stable configuration was found
  - Further structural development observed in MD
    - Backbone starts aligning next to existing β-sheet

Suggests mechanism for Cu-initiated misfolding of α-synuclein!

Mechanism of Cu-initiated misfolding of α-synuclein (α-syn)

Free α-syn

- Mostly linear
- Little tendency for structural development

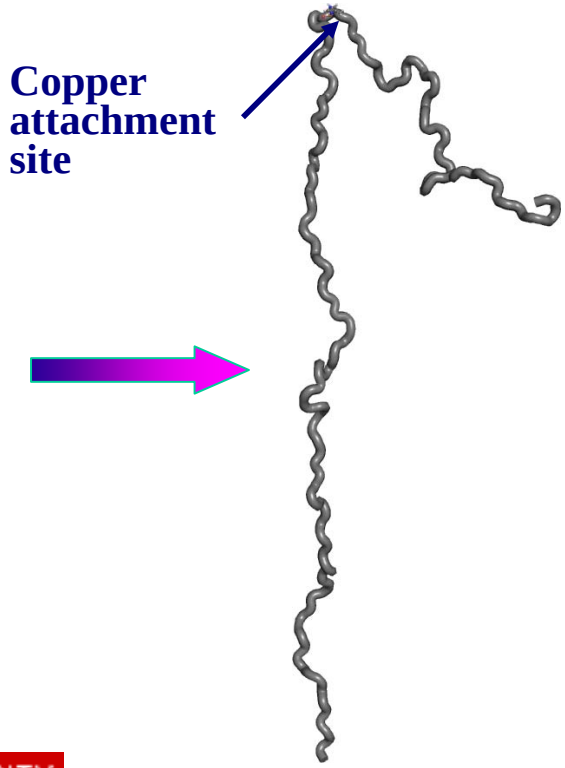
Cu-bound α-syn

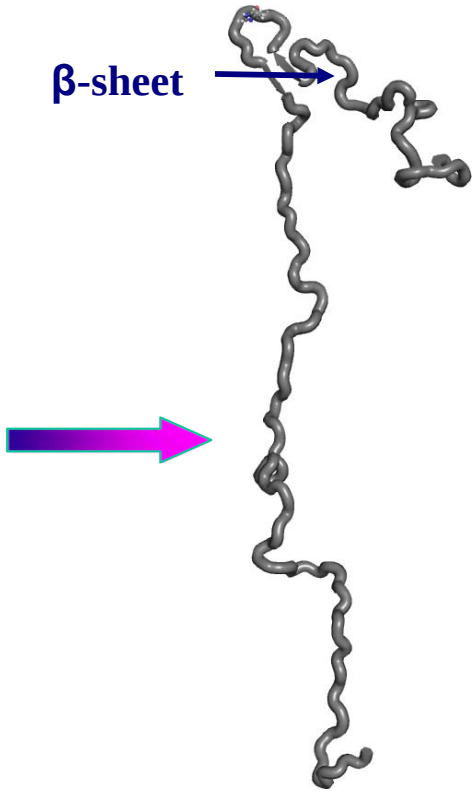
- Cu attachment causes a turn-like feature at the binding site
- Turn creates favourable conditions for β-sheet development

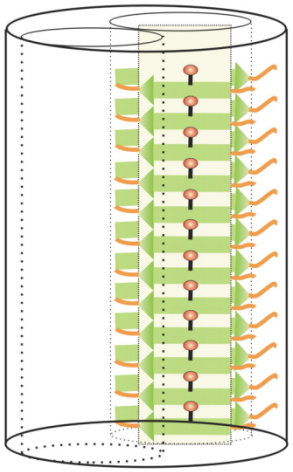
β-sheet development

- β-sheet is formed around the binding site
- Structural development continues

Structural development continues until the protein is completely misfolded







Oligomerization

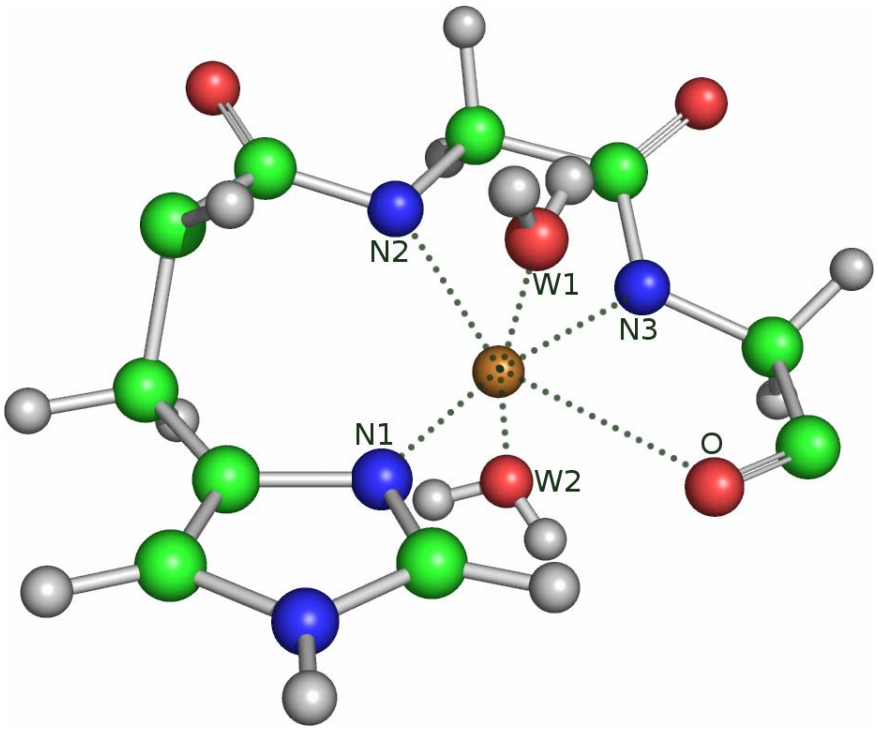
Summary

- New orbital-free DFT method for including solvent effects at low cost
  - Large-scale hybrid DFT/approximate-DFT quantum molecular dynamics
  - Biomolecule and its first solvation shells treated by full DFT
  - Solvent treated by orbital-free DFT: cheap & linear-scaling
  - Exchange of solvent molecules between full and approximate DFT regions is allowed; total energy is conserved
- Suitable for Exascale
  - Linear-scaling full DFT & "beyond DFT": multi-level hierarchical *ab initio*
  - Replica dynamics and multiple time-steps acceleration
- Applications to transition metal centers in proteins
  - Prion protein serves as a copper storage medium: multistage binding
  - Copper plays a protective role in the onset of prion diseases
  - Copper in Parkinson's disease protein
  - Accelerates misfolding and the onset of Parkinson's disease

Cu-PrP binding site geometry

- Copper bond distances (Å):

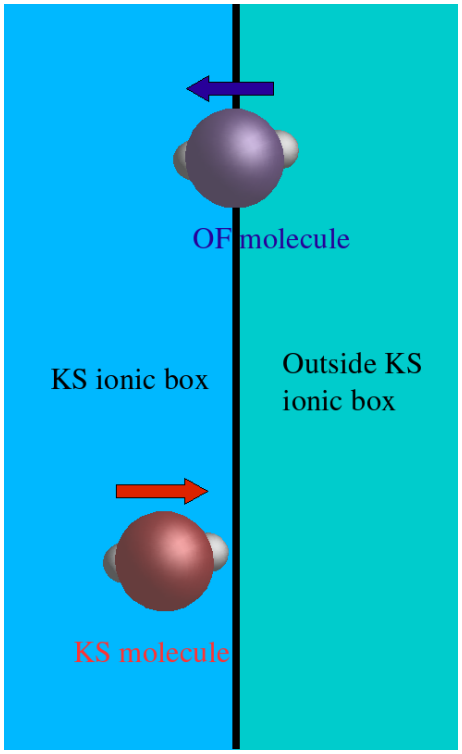
	Dry protein	partial solvation	full solvation
Cu-N1	1.99	1.98	1.98
Cu-N2	2.01	2.00	2.00
Cu-N3	1.91	1.91	1.92
Cu-O	2.19	2.38	2.4
Cu-W1	-	2.83	2.88
Cu-W2	-	2.71	2.64



- Solvation is important for proper description of binding
- Gap between occupied and unoccupied states
  - No solvation: **2.11** eV
  - Partial solvation: **2.25** eV
  - Full solvation: **2.34** eV
- Solvation is important for electronic properties and cannot be neglected

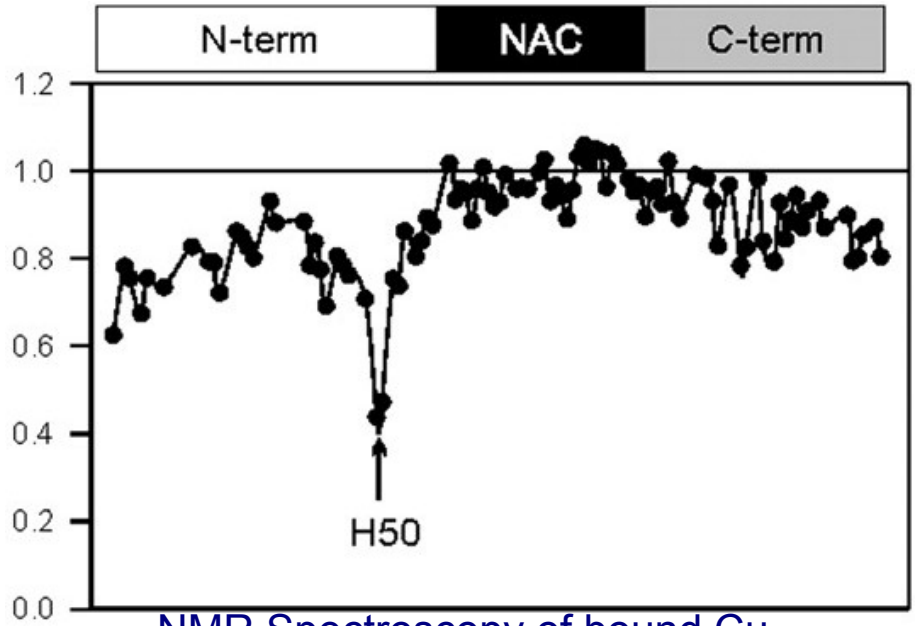
Exchange of solvent II.

- Comparing hybrid simulations with enabled and disabled switching
  - 1 exchange event: OF -> KS
  - Difference at the end of simulation: 200 MD steps after the exchange
    - Displacement: 5.6 % (0.05 Bohr)
      - timestep = 20 a.u (0.5 fs), T= 300K



Protein based Onset Mechanisms

Copper (II) binding at residue 50 of α-synuclein has been experimentally verified to accelerate oligomerization



NMR Spectroscopy of bound Cu
(Rasia et al., PNAS 2005)

Inherited mutation of α-synuclein at residue 53 (Alanine to Threonine - A53T) invariably leads to early onset PD

α-synuclein sub-sequence

40 50 60
... GSKTKEGVVH GVATVAEKTK ...

40 50 60
... GSKTKEGVVH GVTTVAEKTK ...