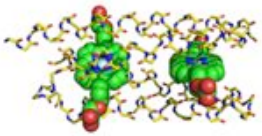
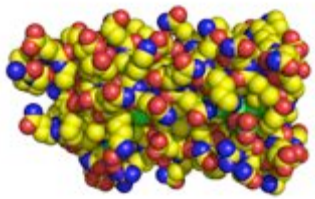
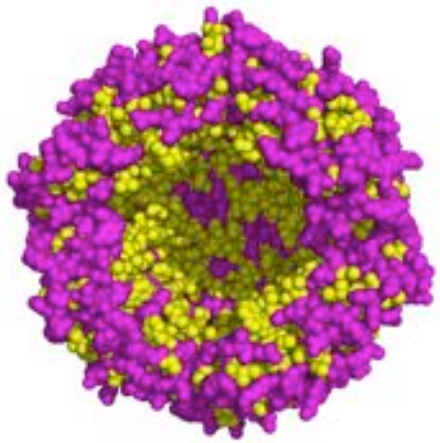
Theoretical approaches to designing and understanding proteins

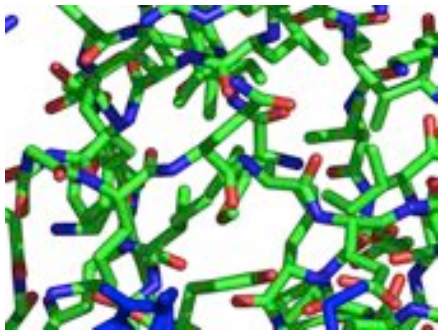
Jeffery G. Saven

University of Pennsylvania
Department of Chemistry

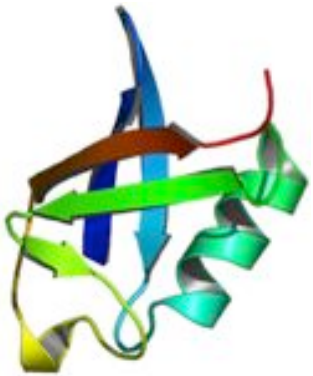





Proteins: many length scales and functions

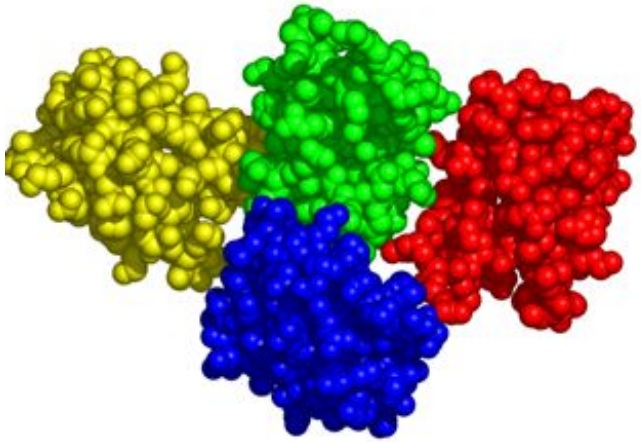
- Multiple environments: solution, membranes, surfaces…
- Many functionalities
  - Solution: catalysis, recognition, signals, pigments…
  - Membranes: channels, energy transduction, light harvesting, signaling…



Å



nm

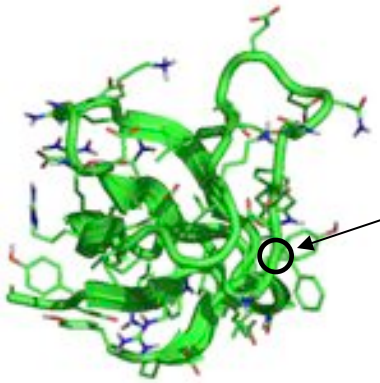


> 10 nm

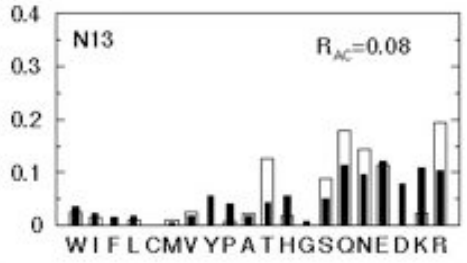
Theory and Design of Proteins (and Self-Organizing Macromolecules)

- Methods for probabilistic protein design
  - Input:
    - –Target tertiary and quaternary structure
    - –Features, e.g., well-packed, hydrophobic interior
    - –Atomistic energy functions
    - –Physical, synthetic and functional constraints on sequences
  - Output: Site-specific probabilities of the amino acids for a given structure
  - Advantages:
    - Large structures and diversity
    - Application to de novo protein design and combinatorial design
    - Transferable to nonbiological systems

Target structure

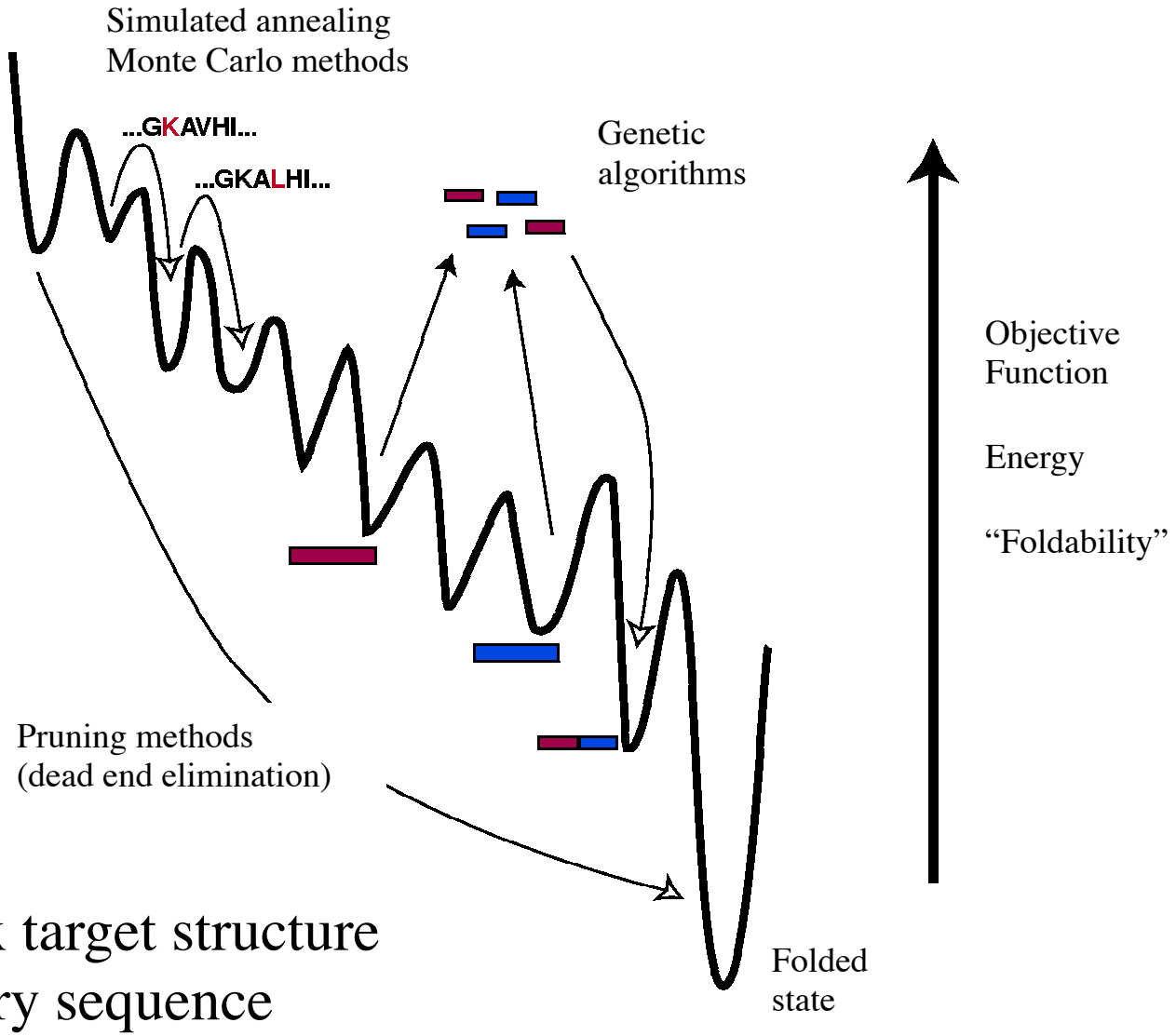


Atomistic model of structure and monomers



Site-specific monomer probabilities

Sequence design: search methods



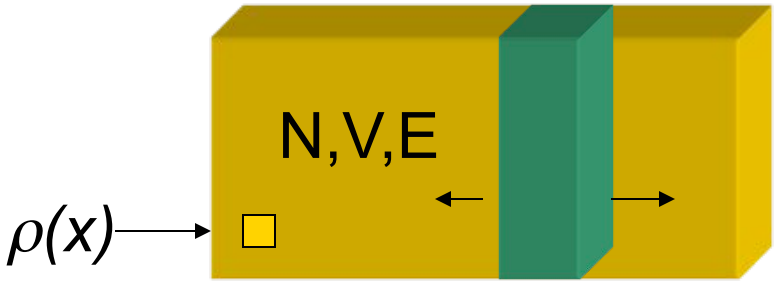
Fix target structure
Vary sequence

S. Mayo (Caltech), H. Hellinga (Duke); D. Baker (UW Seattle); T. Alber (UC Berkeley), P. Kim, B. Tidor, A. Keating (MIT); P. Harbury (Stanford); J. Desjarlais (Xencor); C. Floudas (Princeton); L. Lai (Beijing); S. Takada (Kyoto)…

Thermodynamic analogy

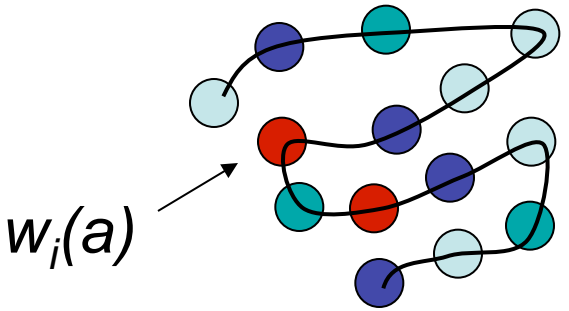
<u>Thermodynamic ensemble</u>

Ω=no. configurations



<u>Sequence ensemble</u>

Ω=no. sequences



Apply methods from statistical thermodynamics to estimate probabilities (effective thermodynamic quantities: T, E, S…)

- Solve for probabilities $w_i(a)$ subject to constraints on sequences
  - Self-consistent field methods based on entropy maximization

    H. Kono and J. G. Saven. *J Mol. Biol.*,. 306: 607-627 (2001).
    J. Zou and J. G. Saven, *J. Mol. Biol.,* 296: 281-294 (2000).
- Sample sequences and count frequencies of amino acids
  - Efficient (biased with replica exchange) Monte Carlo methods

    X. Yang and J. G. Saven, *Chem. Phys. Lett.*, 401: 205-210 (2005).
    J. Zou and J. G. Saven, *J. Chem. Phys*. 118: p. 3843–3854 (2003).

Self-consistent, entropy maximization

Sequences are not enumerated

Solve for probabilities $w_i(a)$: a = amino acid state, i = position in sequence

Maximize subject to physical and synthetic constraints on sequences: E_i, f_i

–Constrain effective energies E_i (low energy sequences for target structure)

–Other possible constraints:

- Pattern amino acids: hydrophobic inside, hydrophilic outside
- Specify identities and/or conformations of functionally important residues

$$V(\{w_i(a)\}) = S - \beta_1 E_1 - \beta_2 E_2 - \ldots - \lambda_1 f_1 - \lambda_2 f_2 - \ldots$$

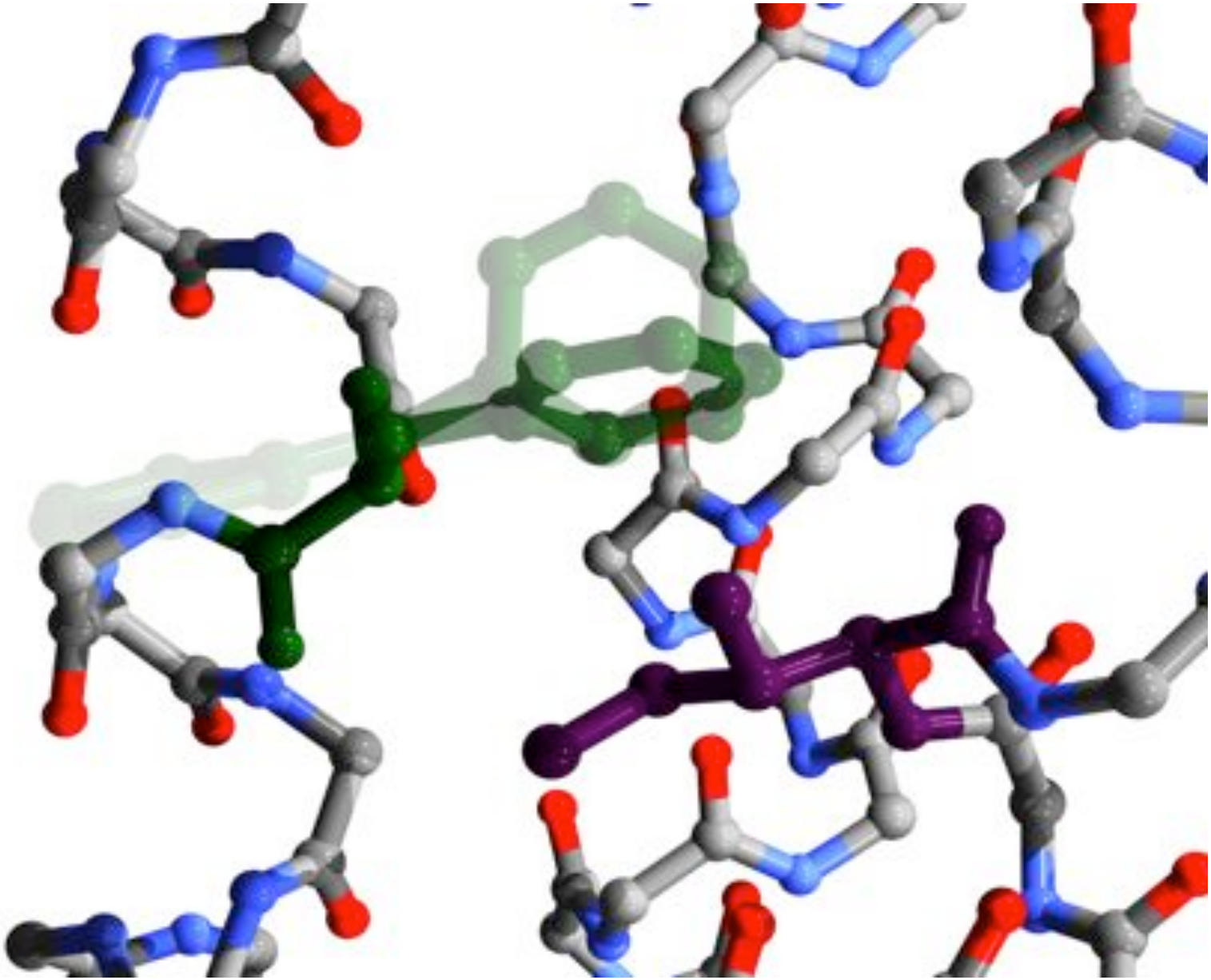
$$S = -\sum_i \sum_a w_i(a) \ln w_i(a)$$

$$E_1 = E_{folded}(\{a\}) - F_{unfolded}(\{a\})$$ Atomic potential energy

$$E_2 = E_{solvation}(\{a\})$$ Solvation energy

$$\frac{\partial V}{\partial w_i(a)} = 0, \text{ and } E_i = \langle E_i \rangle_{sequence}$$

H. Kono and J. G. Saven. *J Mol. Biol.*,. 306: 607-627 (2001).
J. Zou and J. G. Saven, *J. Mol. Biol.*, 296: 281-294 (2000).



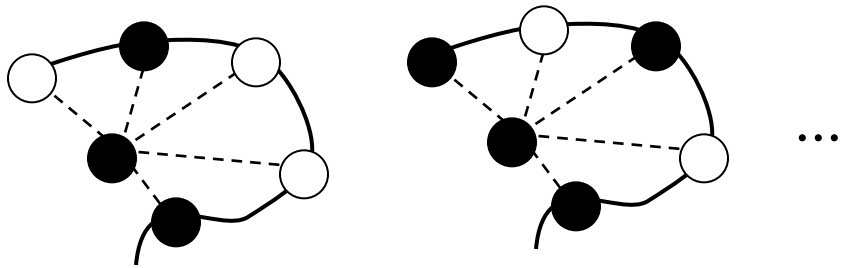
Local average energy

Energy of sequence $(a_1, \ldots, a_N)$ in structure

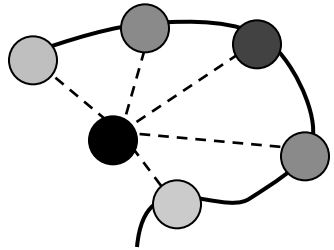
$$E(a_1,\ldots,a_N) = \sum_i \gamma^{(1)}(a_i) + \sum_{i<j} \gamma^{(2)}(a_i,a_j)$$

Local energy of a at site i

$$\varepsilon_i(a) = \gamma^{(1)}(a) + \sum_j \gamma^{(2)}(a,a_j)$$



$$\varepsilon_i(a) \approx \langle \varepsilon_i(a) \rangle = \gamma^{(1)}(a) + \sum_j \sum_{a_j} \gamma^{(2)}(a,a_j) w_j(a_j)$$



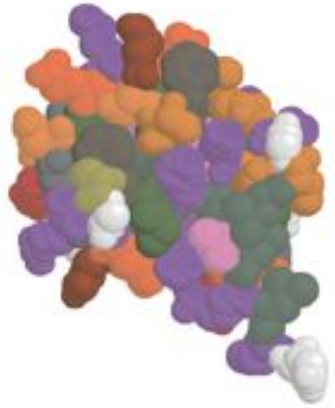
Average over sequences

Atomistic Models of Proteins

- Amino acid and side chain conformation
- "Energy"

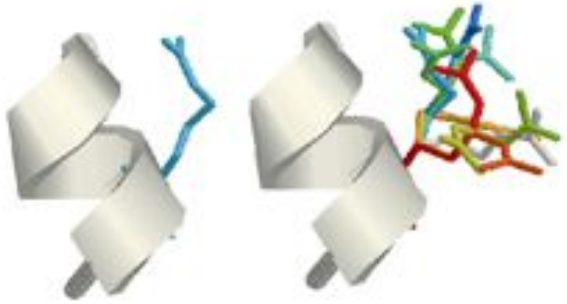
  Atomic interactions (AMBER)
  Solvation (Hydrophobic effect)

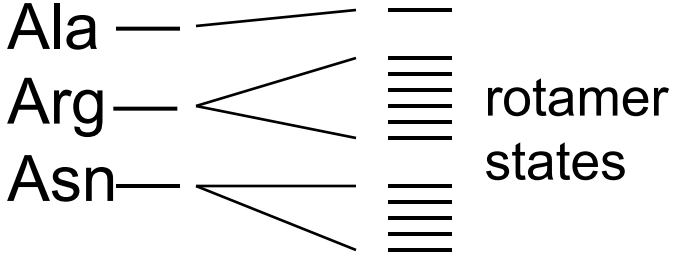
  [Kono & Saven, *J. Mol. Biol,* 306:607 (2001)]



Discrete conformational states for amino acids (rotamers)

Dunbrack & Cohen, *Protein Sci.,* 6:1661 (1997)

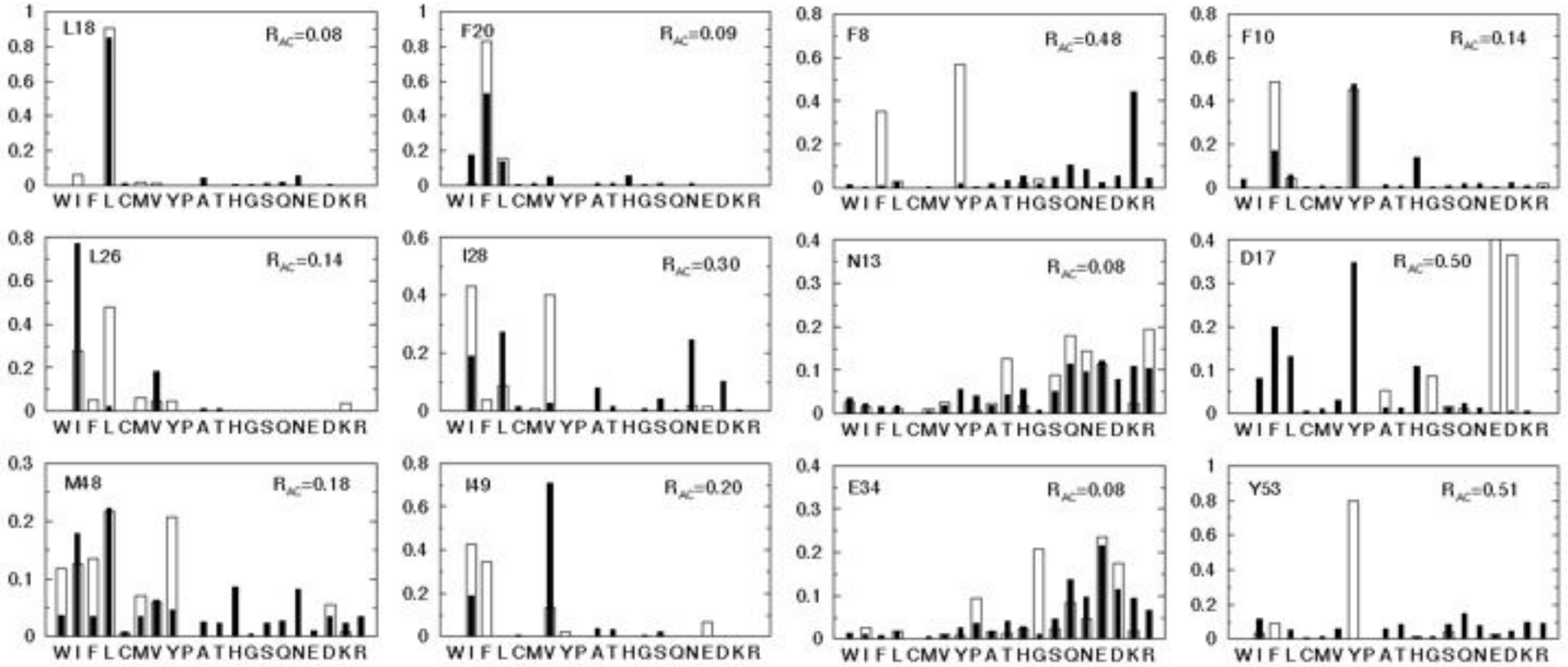




$$w_i(a) = \sum_{r^a} w_i(a, r^a)$$

Relative entropy: SH3 domain

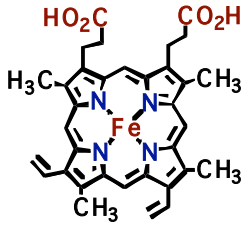
- 57 residue protein
-Allow all amino acids at each position
-Compare with multiple sequence alignment



Designing protein complexes with nonbiological cofactors

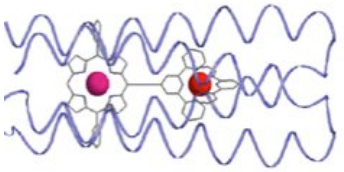
Groups of
Michael Therien, William DeGrado,
& J. Saven

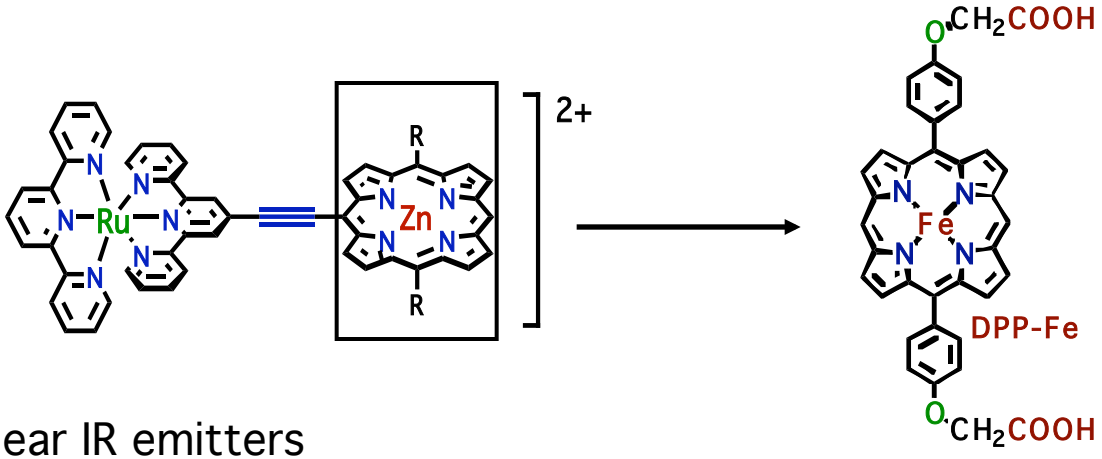
Protein complexes with nonbiological cofactors

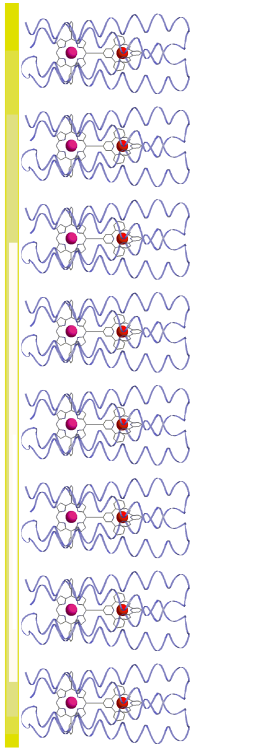


PPIX-Fe

- Cofactors confer function to proteins
  - e.g., Heme (oxygen binding, catalysis)
- New function and materials: proteins containing nonbiological cofactors
  - Controlled cofactor environment
  - Controlled protein assembly







- near IR emitters
- large molecular hyperpolarizability (NLO)
- long lived charge separated states
  (M. J. Therien)

The equation of the helical coiled-coil

Acta Cryst. (1953). **6**, 685

The Fourier Transform of a Coiled-Coil

BY F. H. C. CRICK

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, The Cavendish Laboratory, Cambridge, England

(*Received* 14 *March* 1953)

The Fourier transforms are given for a continuous coiled-coil, and for a set of atoms spaced at regular intervals along a coiled-coil. The nature of the solution is briefly discussed.

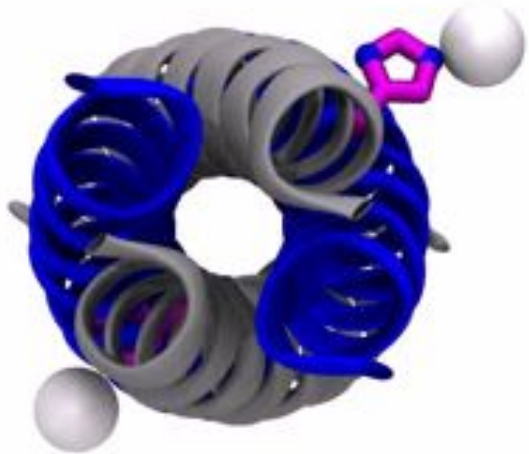
$$x = r_2 \cos \omega_2 t + r_1 \cos \omega_2 t . \cos \omega_1 t - r_1 \cos \alpha . \sin \omega_2 t . \sin \omega_1 t ,$$

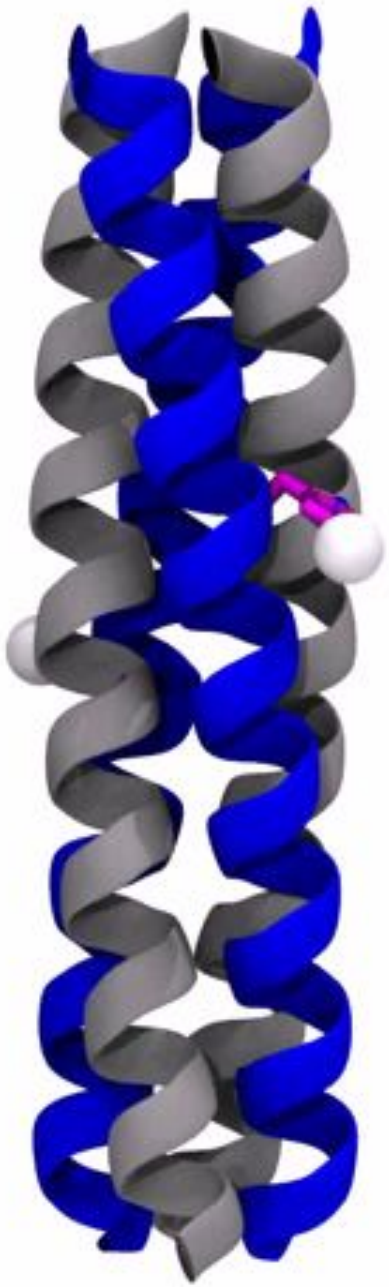
$$y = r_2 \sin \omega_2 t + r_1 \sin \omega_2 t . \cos \omega_1 t + r_1 \cos \alpha . \cos \omega_2 t . \sin \omega_1 t ,$$

$$z = P(\omega_2 t/2\pi) - r_1 \sin \alpha . \sin \omega_2 t ,$$

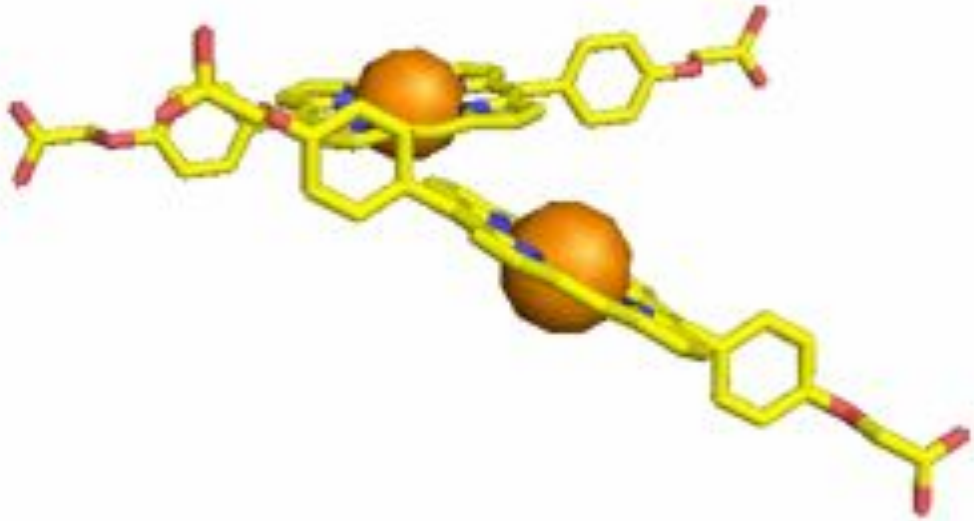
Crick predicted the structural topology of the coiled-coil

PITCH: 132.87
RADIUS: 4.65
RPT: 3.64
HEL. ROT: -175.00
HEL. OFF: 0.00

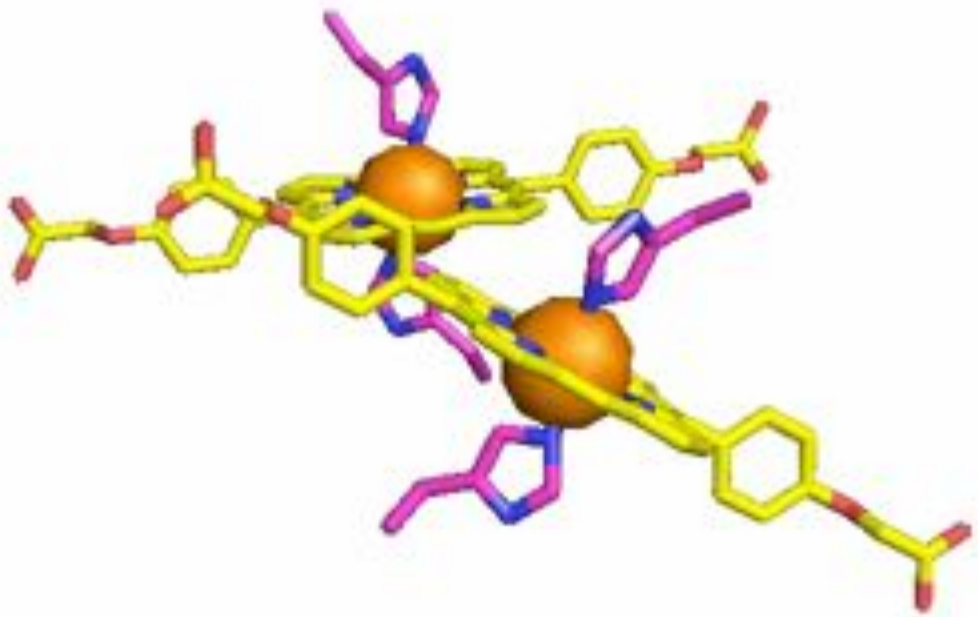




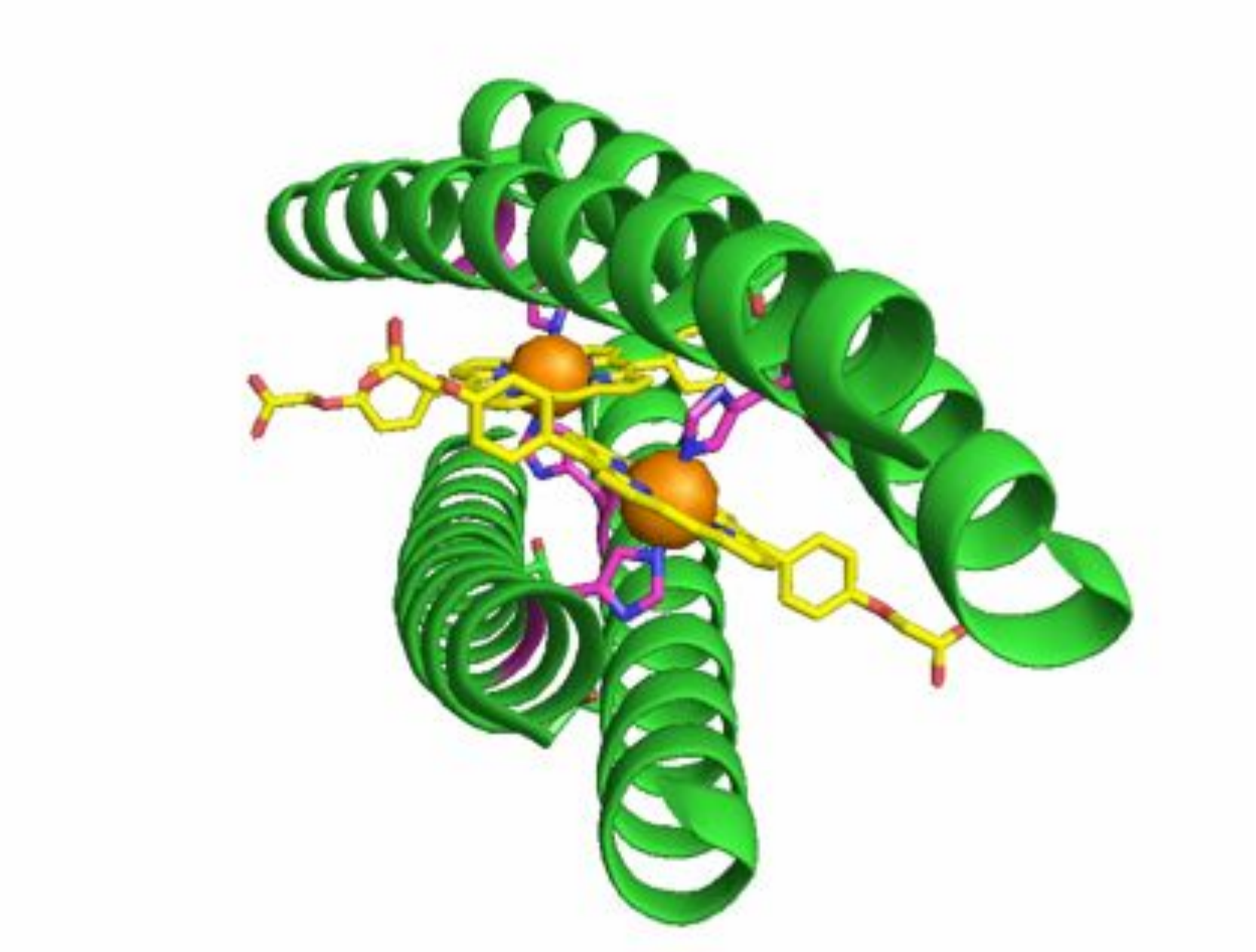
Porphyrin



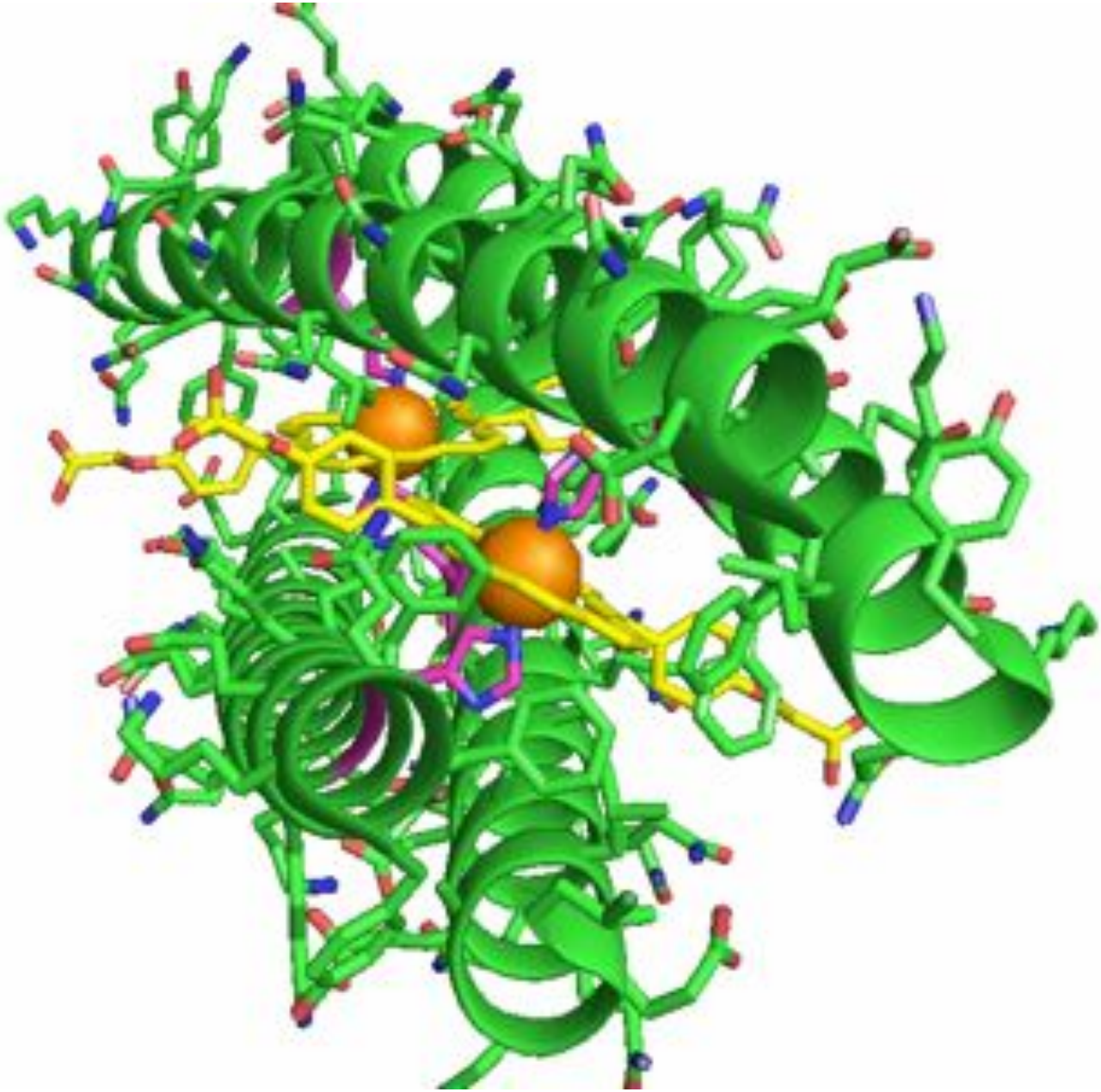
Porphyrin + His



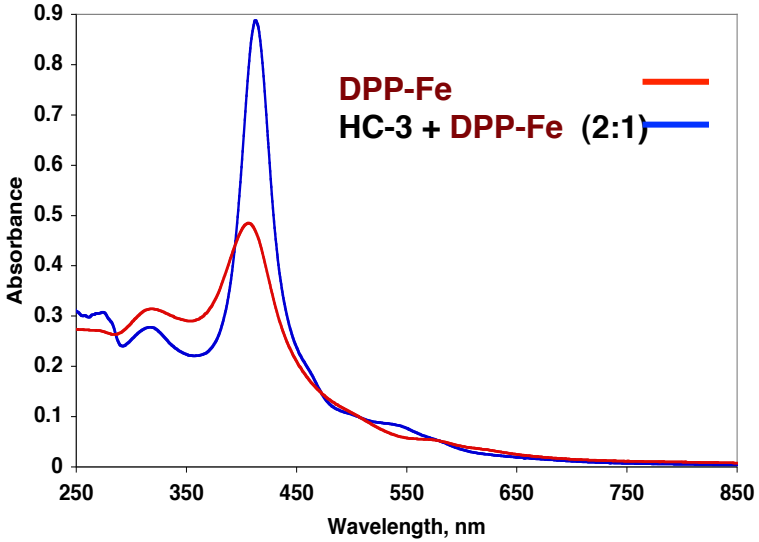
Porphyrin + His + Thr + tetrameric backbone

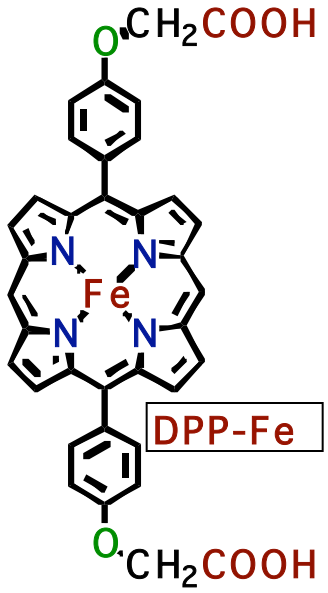


Porphyrin + His + Thr + tetrameric backbone + sequence



Designed protein binds nonbiological Fe-porphyrin cofactor



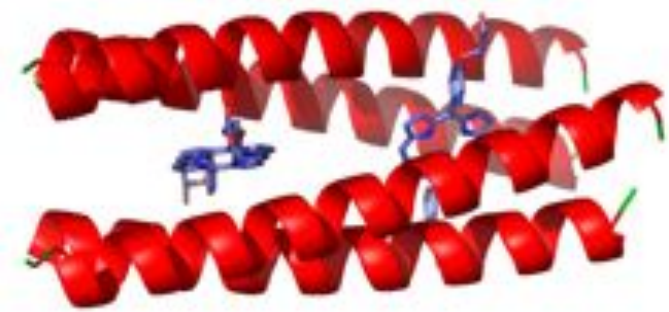


Computationally designed tetra-α-helical has target structure and protein selectively binds DPP-Fe.

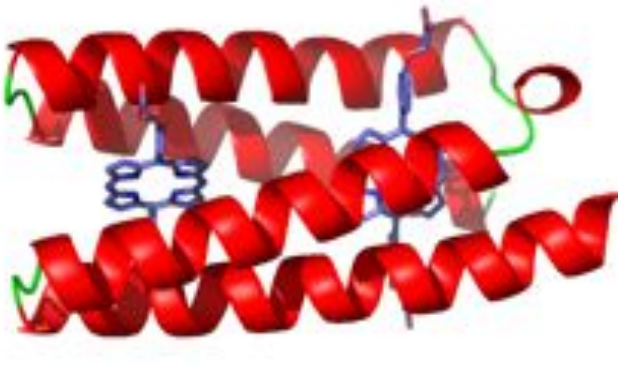
Helical peptide (CD)
Helicity increases upon addtion of cofactor (CD)
Asymmetry about cofactor (CD in Soret band)
Correct MW (Gel filtration; HPLC)
Selective binding to nonbiological cofactor

F. Cochran, W. Wang, V. Nanda, S. Wu, W. F. DeGrado, J. G. Saven, M. J. Therien, *JACS*, 2005, 127, 1346-1347

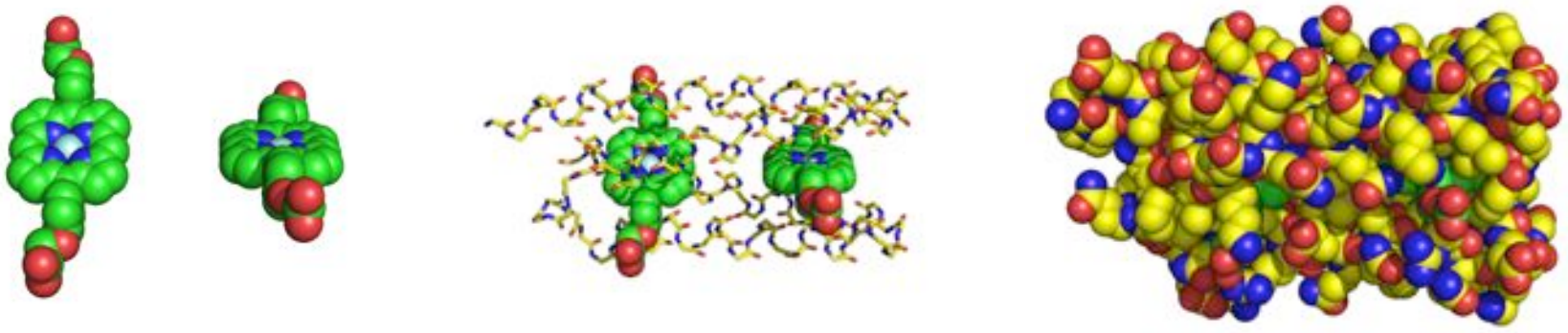
Tetramer to single chain protein



Tetramer

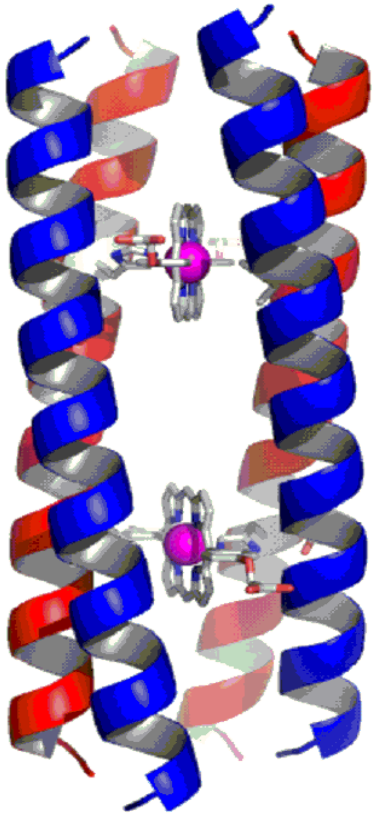


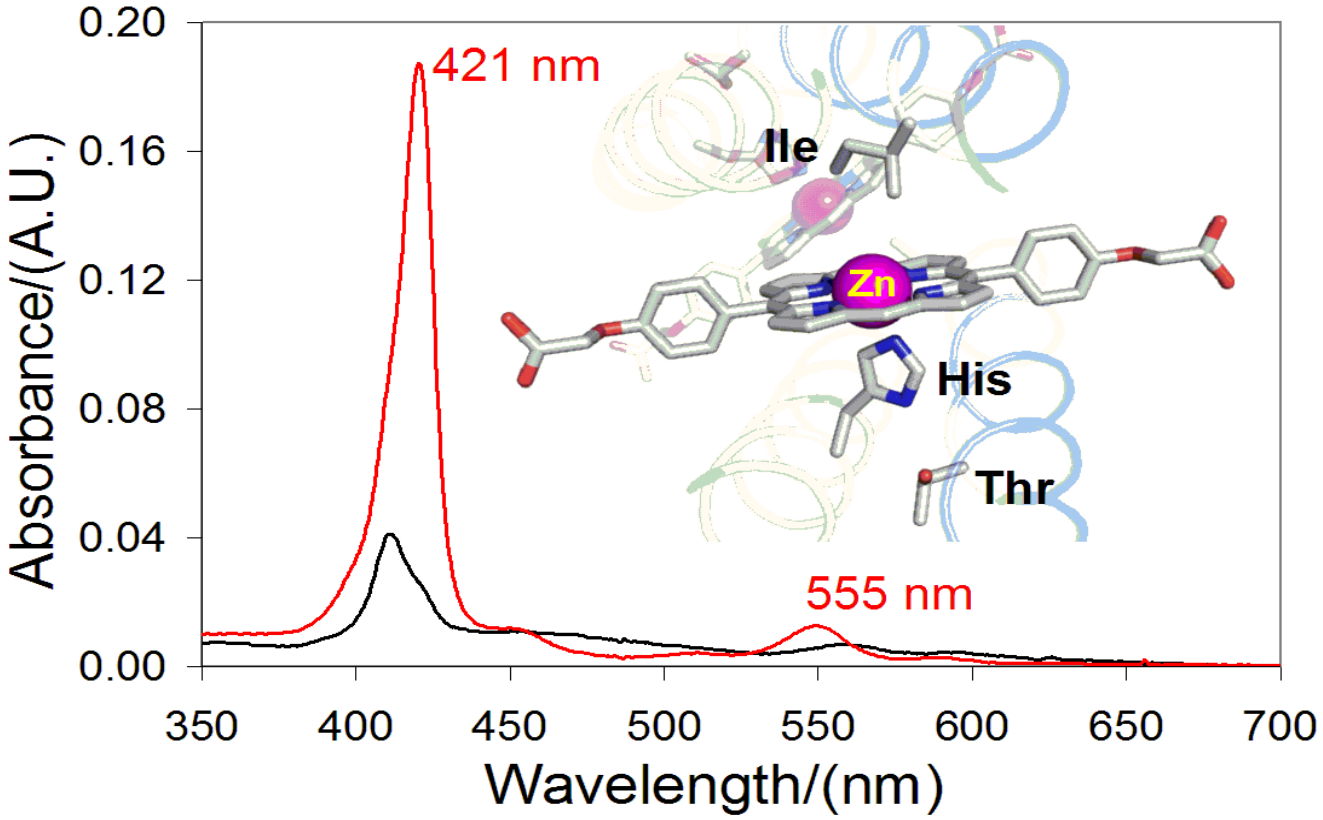
Single-Chain



Gretchen M. Bender, Andreas Lehmann, Hongling Zou, Hong Cheng, H. Christopher Fry, Don Engel, Michael J. Therien, J. Kent Blasie, Heinrich Roder, Jeffrey G. Saven, and William F. DeGrado, *J. Am. Chem. Soc.*, 2007. 129: 10732-10740.

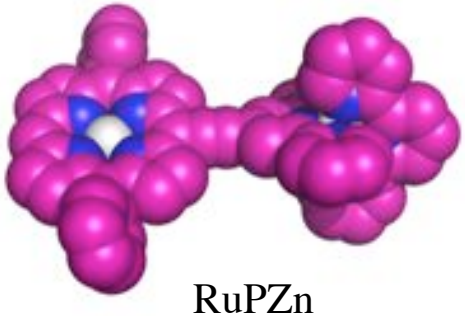
Selective binding of nonbiological Zn-porphyrin to designed α-helical A_2B_2 hetero-tetramer



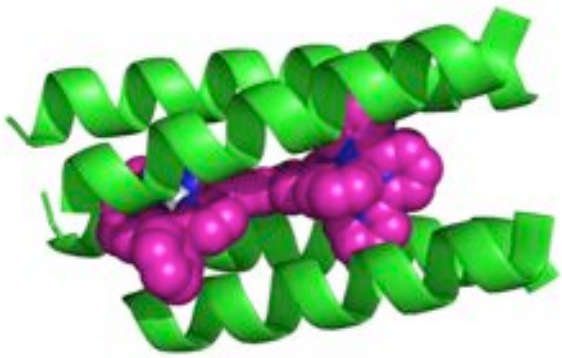


C. Fry, A. Lehmann, J. Saven, W. DeGrado, M. Therien. *J. Am. Chem. Soc.* (2010)

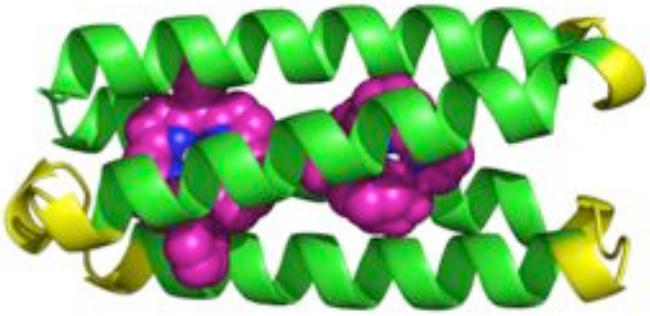
Tailoring protein to NLO cofactor: RuPZn



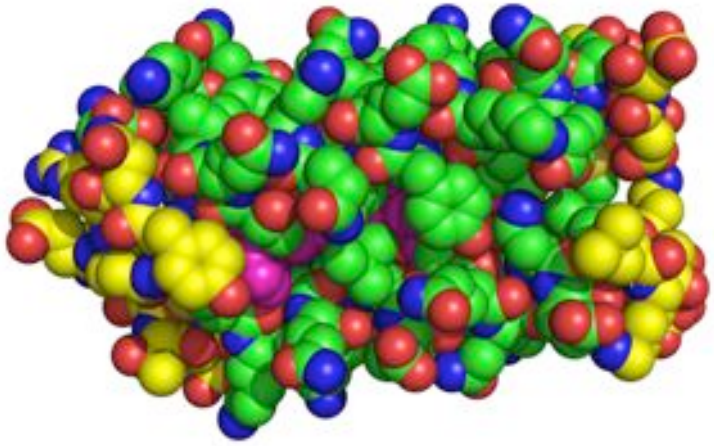
RuPZn

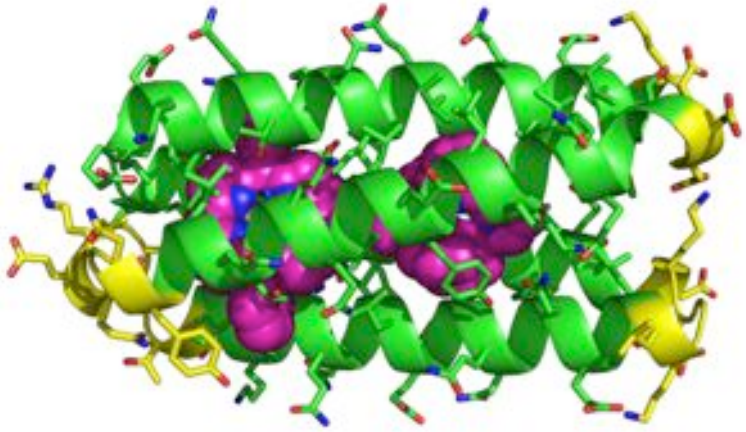


Build helical bundle



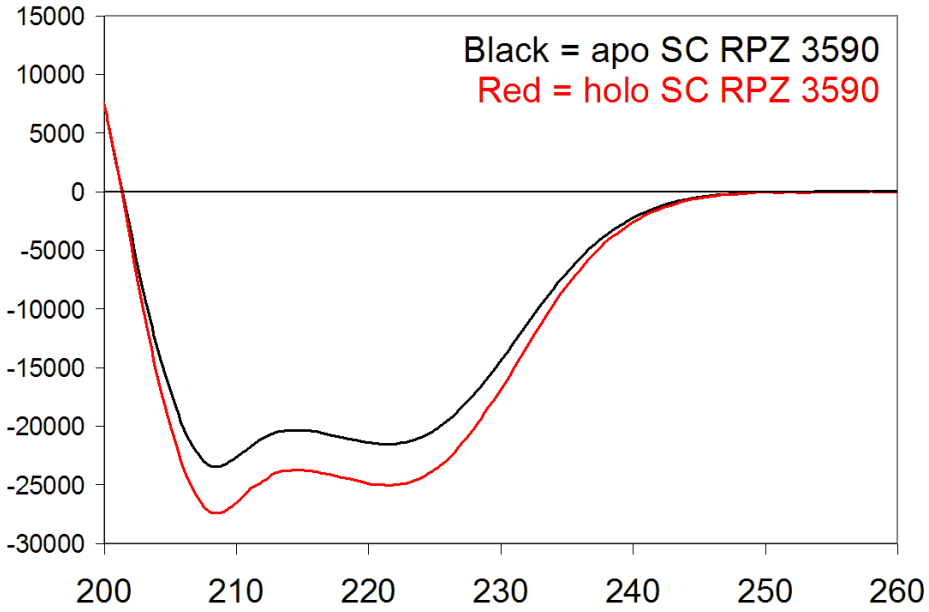
Build loops to arrive at Single chain

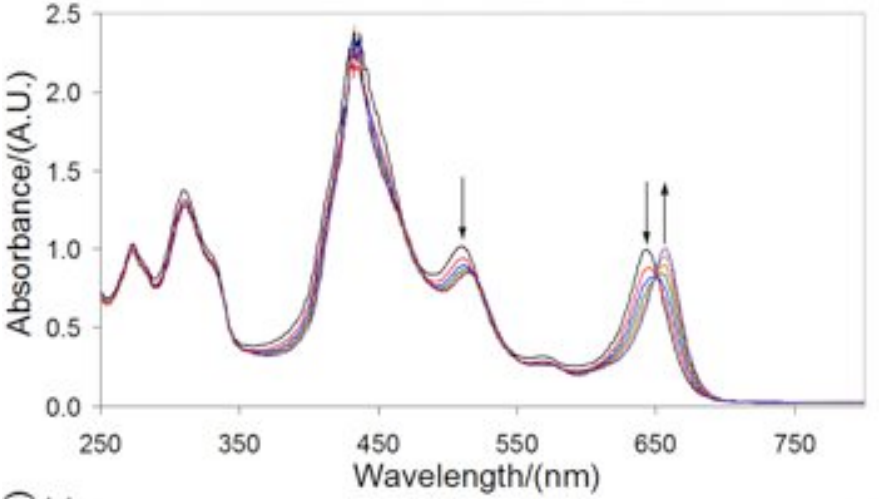


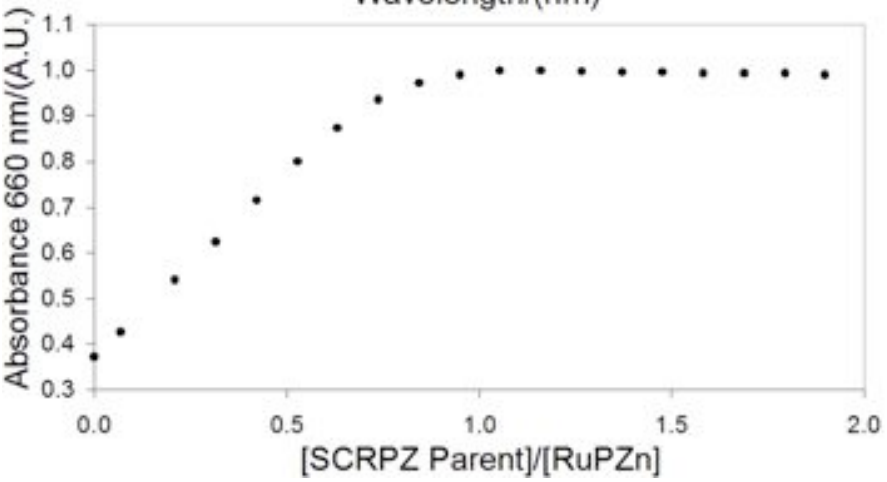


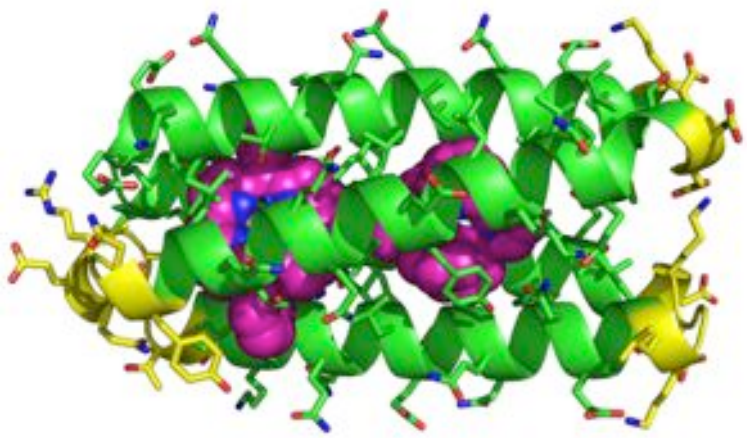
Computational design of sequence

Designed proteins are helical and bind cofactor









Shift in Q band as cofactor binds
Titration has correct stoichiometry
K_d ~ 30 nM, 1 uM

Acknowledgments

J. G. Saven group (Penn)

Present
Chris Lanci
Chris MacDermaid
Jose Manuel Perez Aguilar

Former
Hidetoshi Kono
(Japan Atomic Energy Res. Inst.)
Andreas Lehmann
(Fox Chase Cancer Center)
Wei Wang (Colgate-Palmolive, Inc.)
Xi Yang (Standard & Poors)

William F. DeGrado group (Penn)

Michael J. Therien group (Penn & Duke)
Christopher Fry (Argonne Nat'l Lab)

Feng Gai group (Penn)

Ivan Dmochowski group (Penn)

Support
National Science Foundation
(MRSEC)
(NSEC)
Department of Energy
National Institutes of Health
University of Pennsylvania