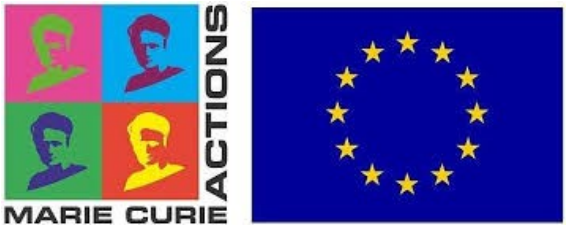





## 108° Congresso Nazionale SIF
## Università di Milano, Italy

# Proteins – a celebration of consilience

## *physics meets biology*

**Tatjana Škrbić**

**Jayanth Banavar, Amos Maritan,**
**Achille Giacometti, Trinh Hoang, George Rose**

# AlphaFold: Breakthrough of the Year 2021


Article

## Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2
Received: 11 May 2021
Accepted: 12 July 2021
Published online: 15 July 2021
Open access
Check for updates

John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4]✉

**Is protein folding problem solved?**

# AlphaFold: Breakthrough of the Year 2021


Article

## Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2
Received: 11 May 2021
Accepted: 12 July 2021
Published online: 15 July 2021
Open access
Check for updates

John Jumper[1,4✉], Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4✉]

## Is protein folding problem solved?

### The protein-folding problem: Not yet solved

We agree with H. H. Thorp ("Proteins, proteins everywhere," Editorial, 17 December 2021, p. 1415) and numerous others (*1*) that the advance in protein structure prediction achieved by the computer programs AlphaFold (*2*) and RoseTTAfold (*3*) is worthy of special notice. The accuracies of the predictions afforded by these new approaches, which use machine-learning methods that exploit the information about the relationship between sequence and structure contained in the databases of experimental protein structures and sequences, are much superior to previous approaches. However, we do not agree with Thorp that the protein-folding problem has been solved.

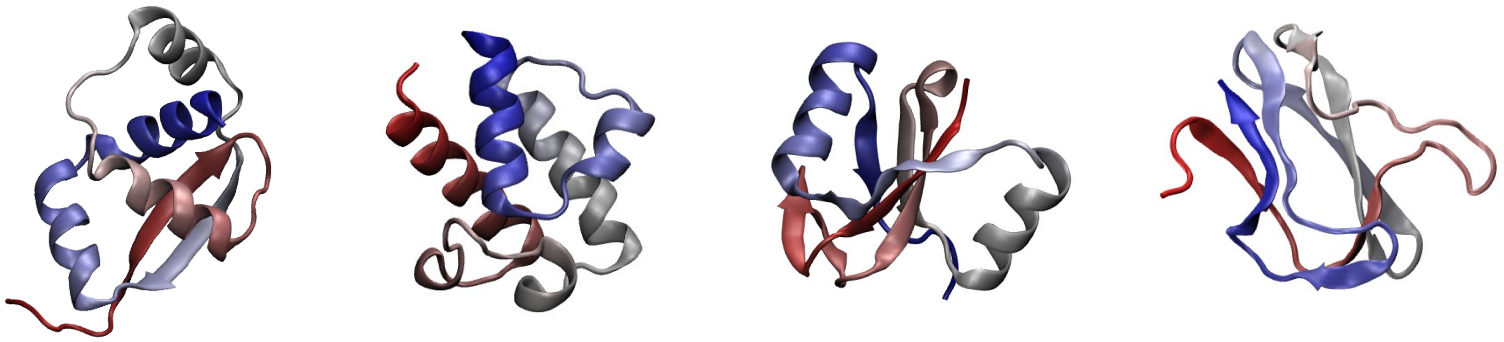
## Not yet, so let's keep doing it!

Peter B. Moore[1], Wayne A. Hendrickson[2], Richard Henderson[3], Axel T. Brunger[4]*
[1]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. [3]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [4]Department of Molecular and Cellular Physiology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.

4 FEBRUARY 2022 • VOL 375 ISSUE 6580

# A brief introduction to proteins

- Linear chains (of amino acids) with bond length ~3.8Å
- 20 amino acid types with distinct side chains
- Proteins fold rapidly and reproducibly (Anfinsen)
- Textbook paradigm: sequence determines structure
- Folding driven by hydrophobicity➡ *protein maximizes self-interaction & expels water from its interior*
- Space-filling building blocks: α – helices and almost planar β – sheets
- Modular structure, finite number of folds



# Linus Pauling

- **Geometries of protein building blocks predicted from quantum chemistry:** ***planarity of the peptide bond & need for a coherent placement of hydrogen bonds***

*THE STRUCTURE OF PROTEINS: TWO HYDROGEN-BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN*

By Linus Pauling, Robert B. Corey, and H. R. Branson*

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California†

Communicated February 28, 1951

**α-helix** **x**

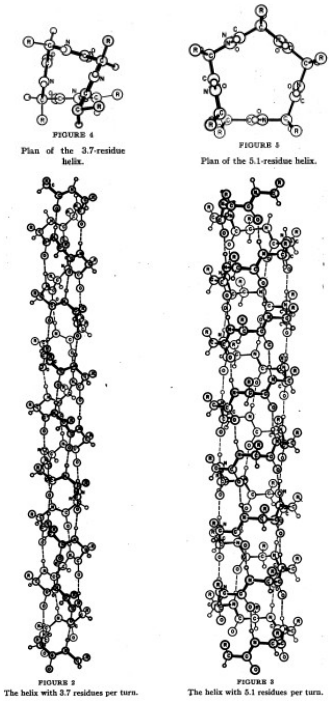


FIGURE 4
Plan of the 3.7-residue helix.

FIGURE 5
Plan of the 5.1-residue helix.

FIGURE 2
The helix with 3.7 residues per turn.

FIGURE 3
The helix with 5.1 residues per turn.

*CONFIGURATIONS OF POLYPEPTIDE CHAINS WITH FAVORED ORIENTATIONS AROUND SINGLE BONDS: TWO NEW PLEATED SHEETS*

By Linus Pauling and Robert B. Corey

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology, Pasadena, California

Communicated September 4, 1951

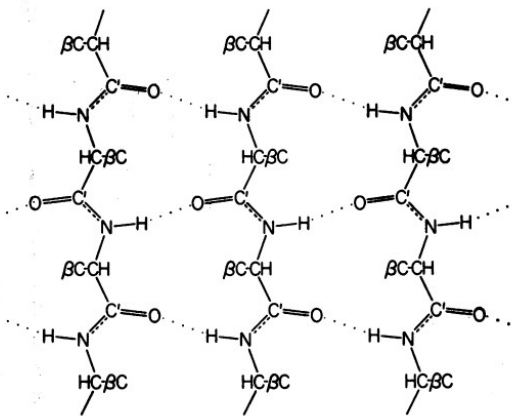


FIGURE 5

A diagrammatic representation of the parallel-chain pleated sheet structure.

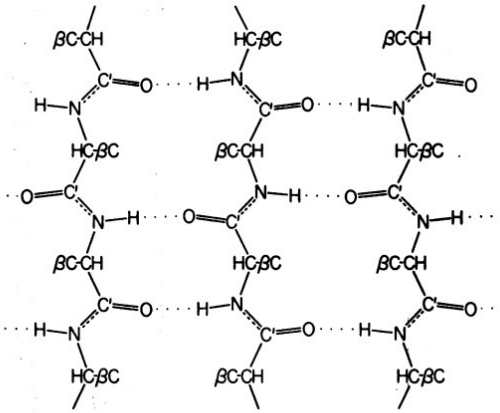


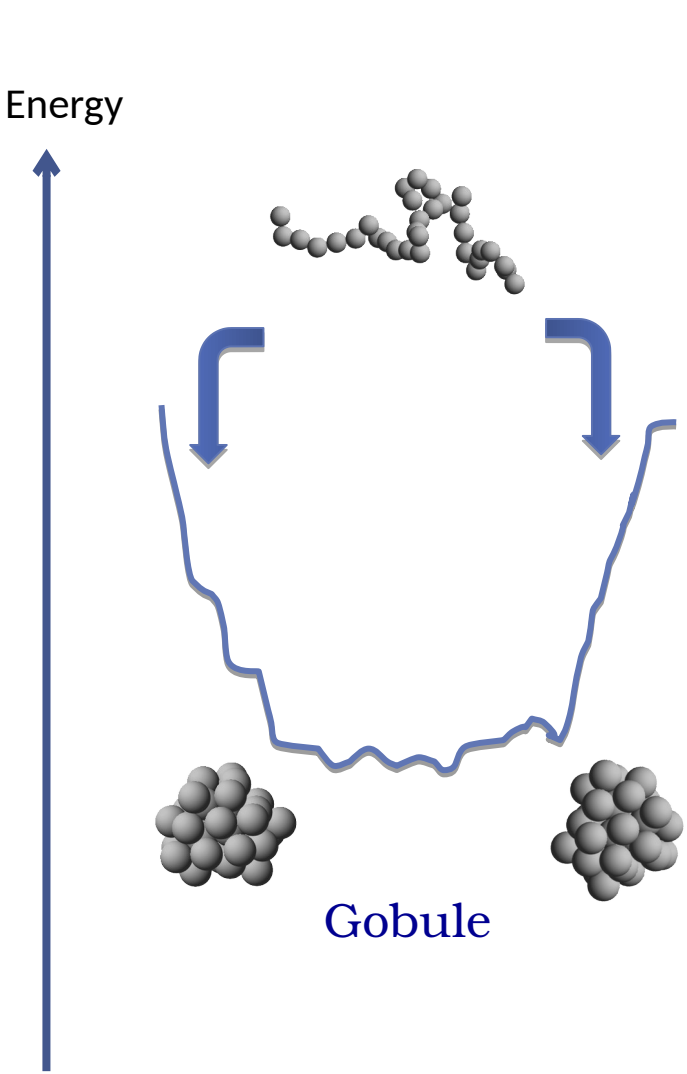
FIGURE 4

A diagrammatic representation of the antiparallel-chain pleated sheet structure.

β ↑↑

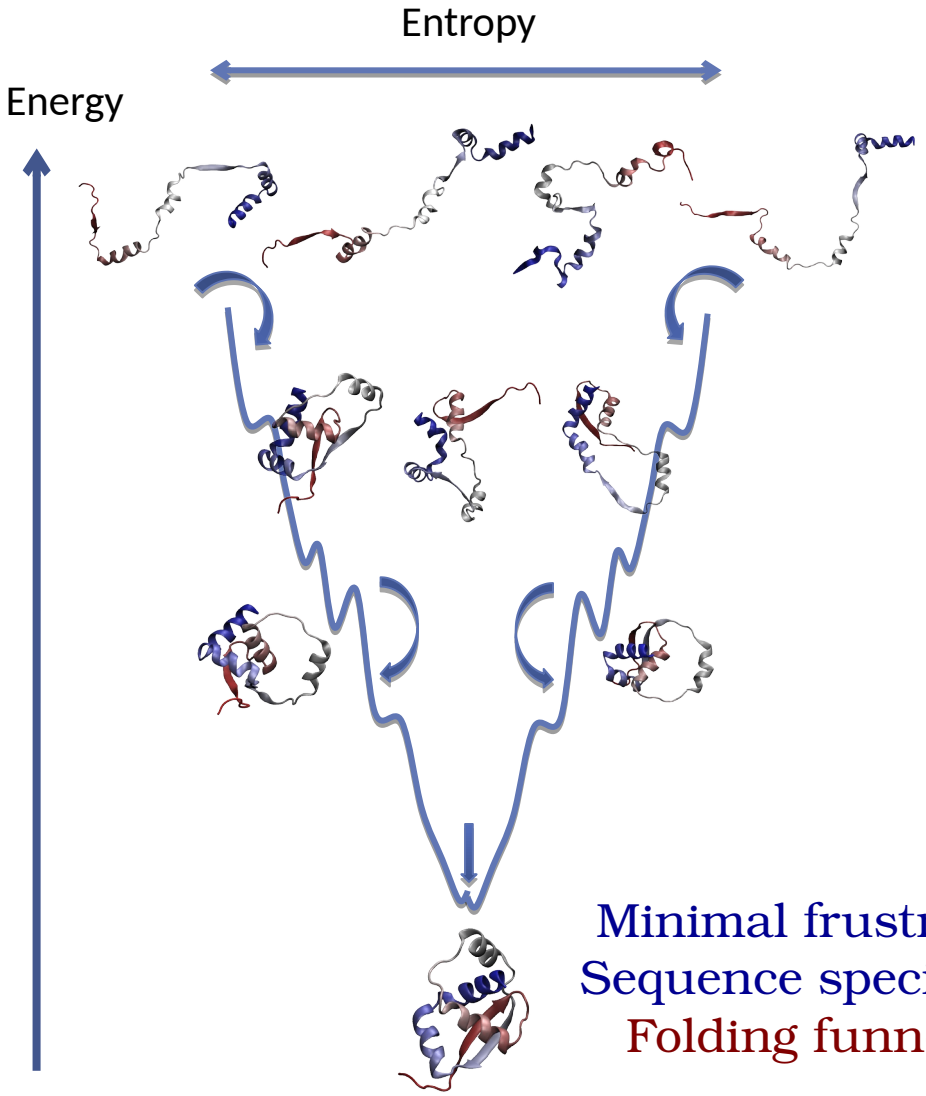
β ↑↓

# Polymers vs. Proteins




Gobule

Taylor, Paul, Binder, …




Minimal frustration
Sequence specificity
Folding funnel

Dill, Onuchic, Wolynes, …

**What is missing?**

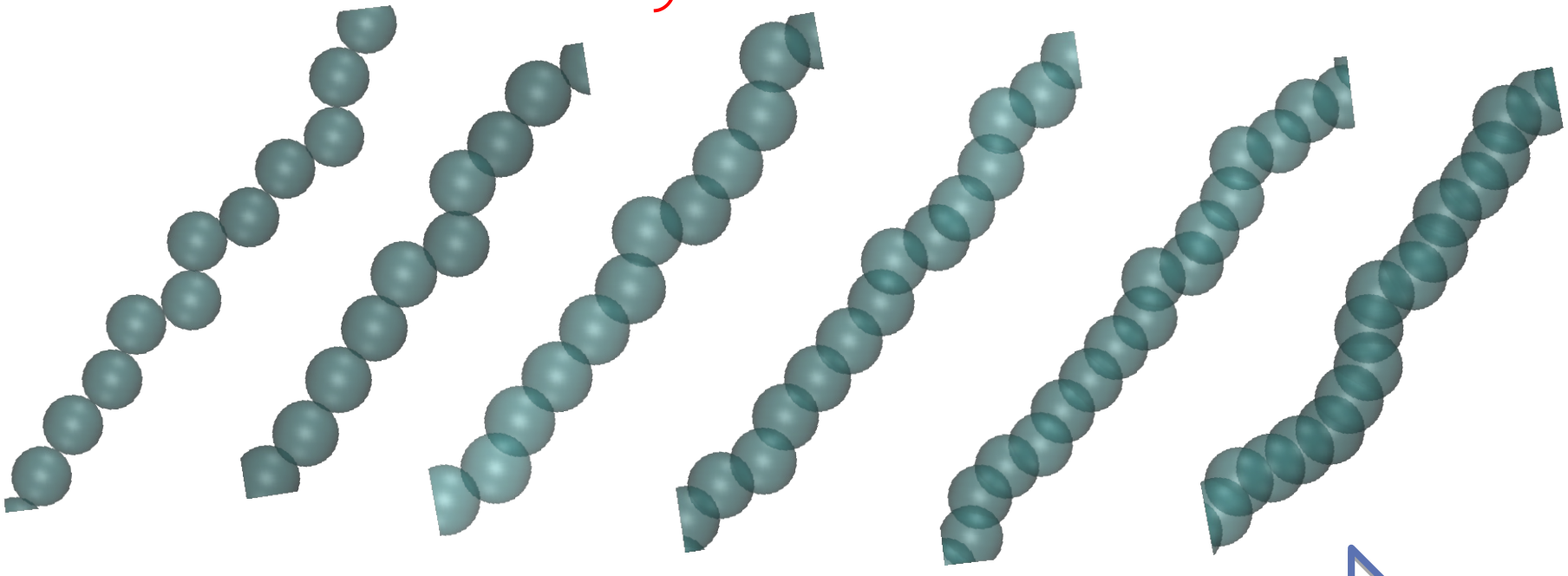
**Amino acid heterogeneity?  Chemistry?**

**Or does one need to remove spurious symmetries?**

# Breaking spherical symmetry

- **Chain is uniaxial**
- **Sphere is isotropic**

This inconsistency can be removed by overlap

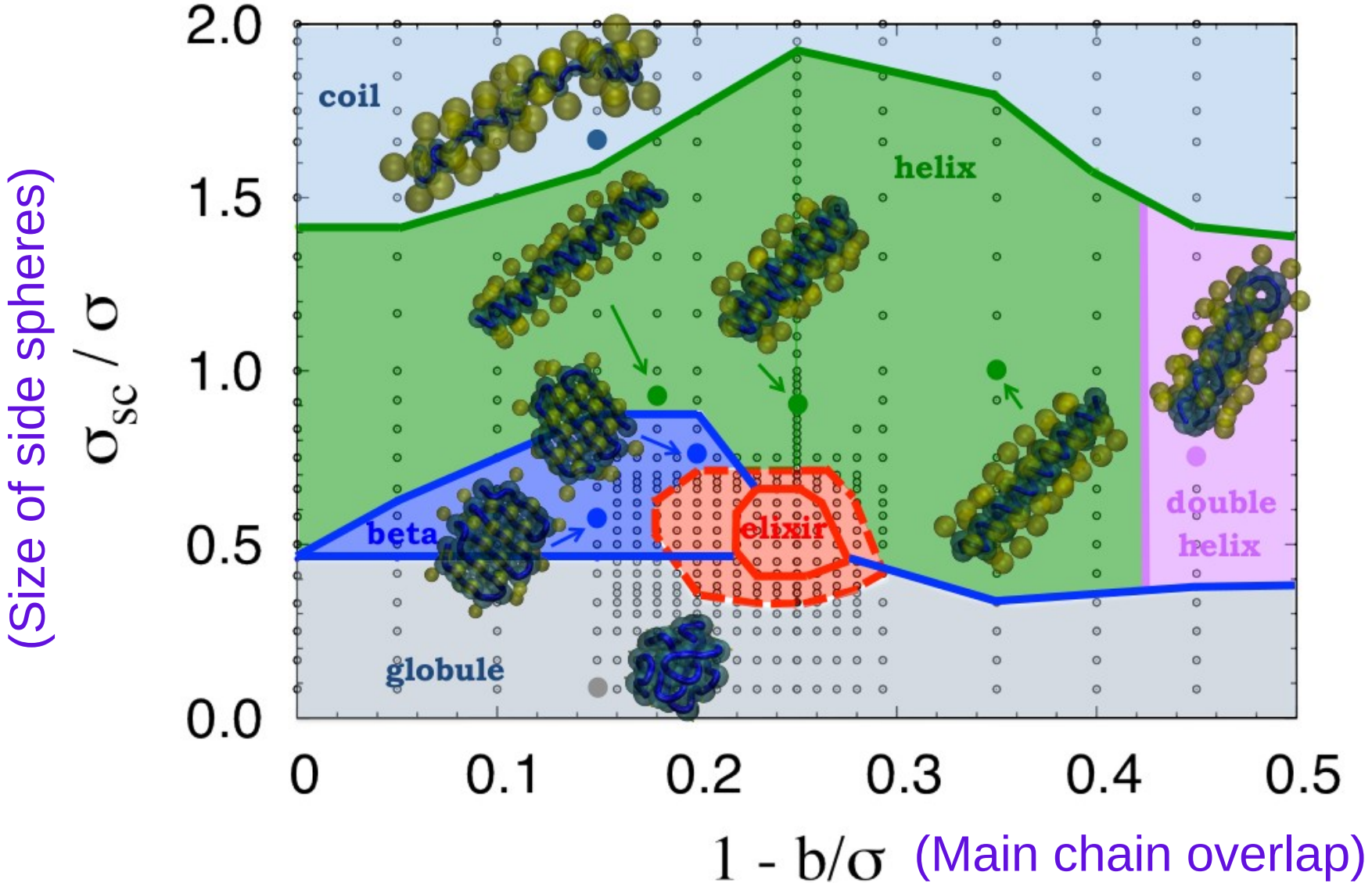


Sphere overlap introduces cylindrical symmetry

**+ *effective attraction* mimicking hydrophobic collapse**

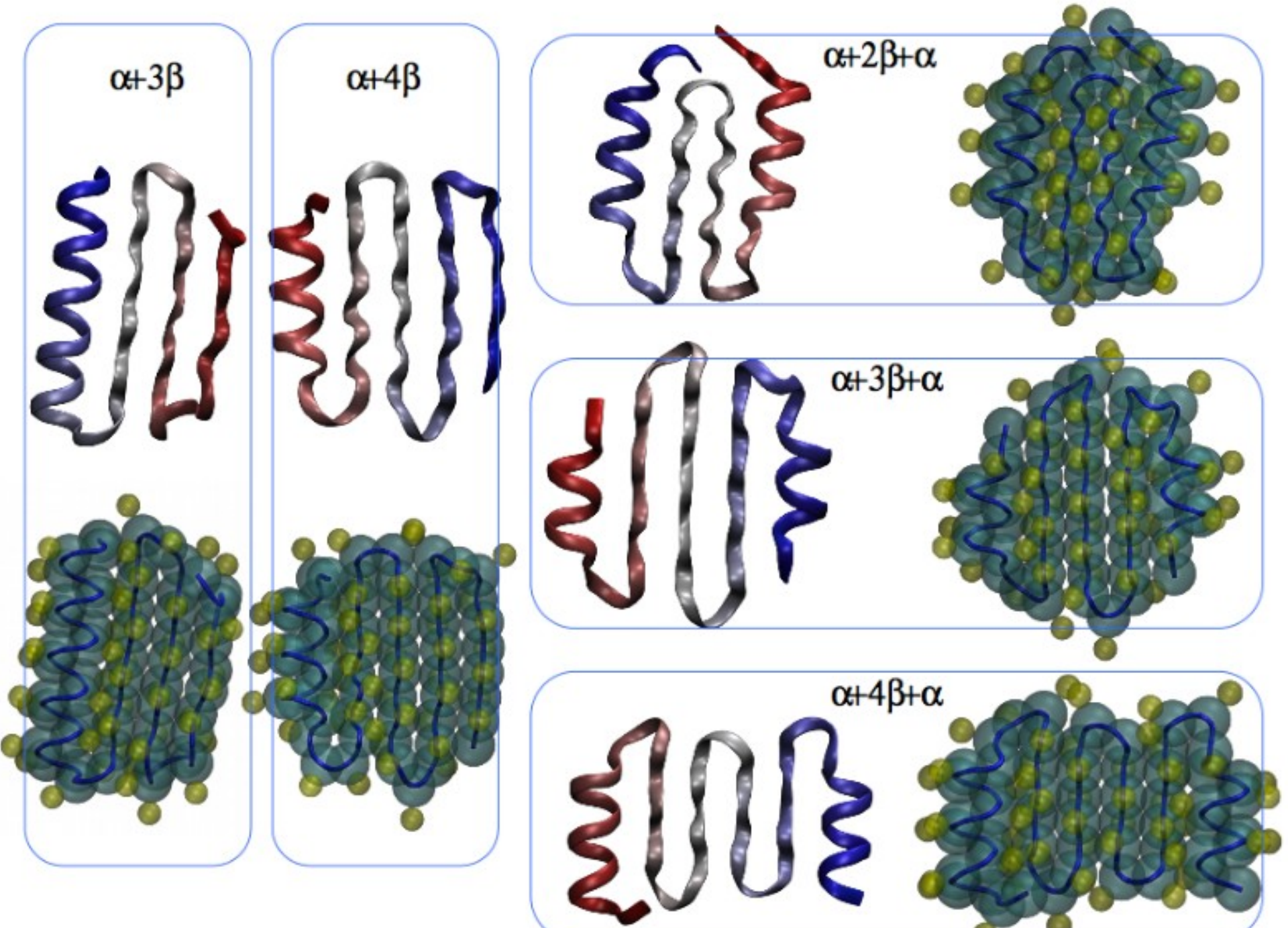
**+ *side spheres* in anti-normal direction (steric hindrance)**

# Elixir phase





Škrbić, Hoang, Maritan, Banavar & Giacometti, *Proteins* (2019)
Škrbić, Hoang, Maritan, Banavar & Giacometti, *Soft Matter* (2019)

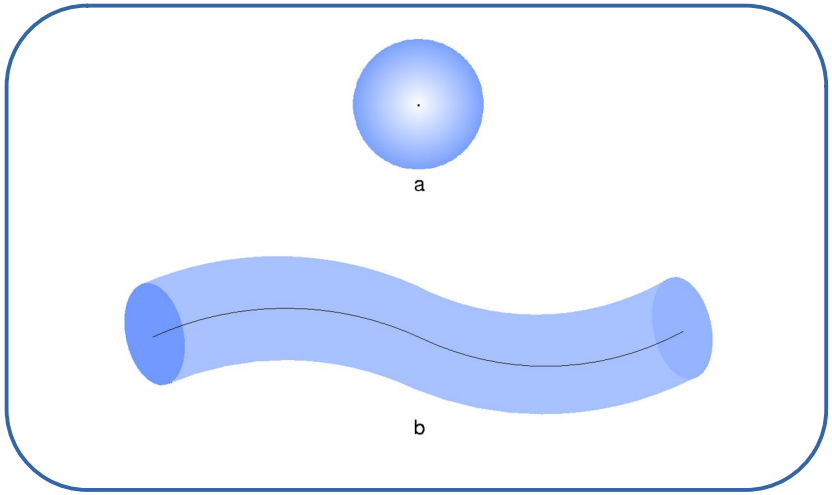

α+3β
α+4β
α+2β+α
α+3β+α
α+4β+α

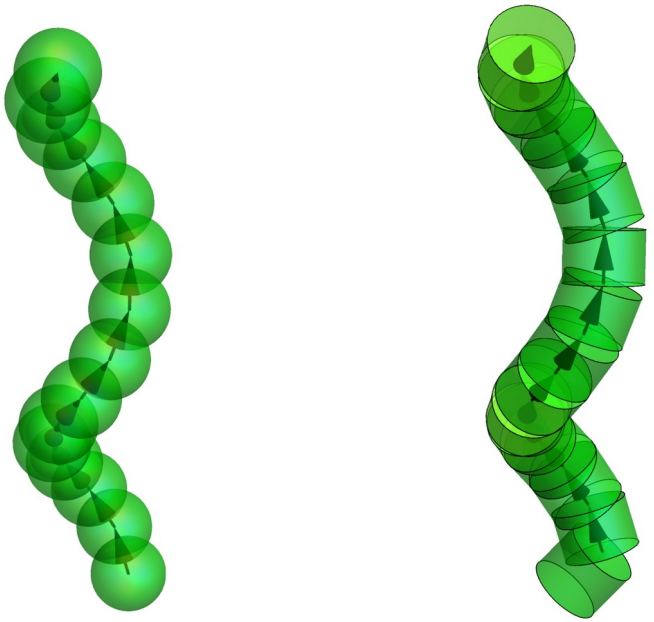
- When inspected carefully, helices & sheets do not have the right geometries
- Breaking spherical symmetry helps

**But what is missing?**

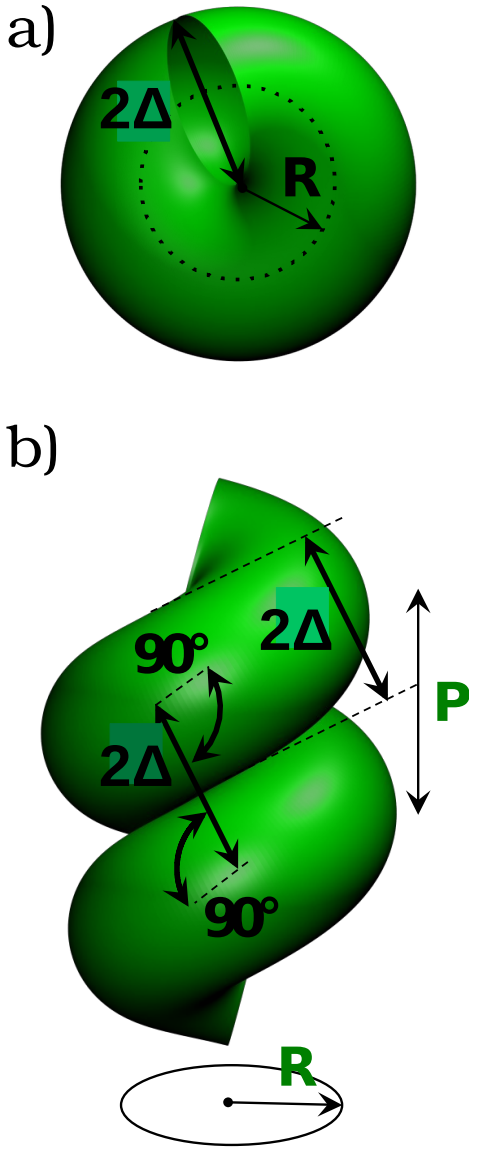
**Space-filling is not correctly captured!**

# Sphere vs Tube



# Continuum case




Maritan, Micheletti,
Trovato & Banavar, *Nature* (2000)

# The discrete space-filling helix

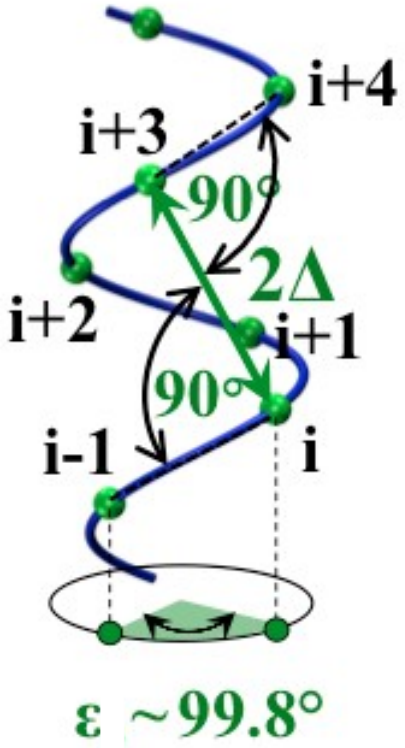
- Bond length equal to 3.81Å
- Space-filling discrete helix determined from elementary mathematics & physics

Tube radius $\Delta \approx 2.63$Å

Rotation angle $\varepsilon \approx 99.8°$

Pitch to radius ratio

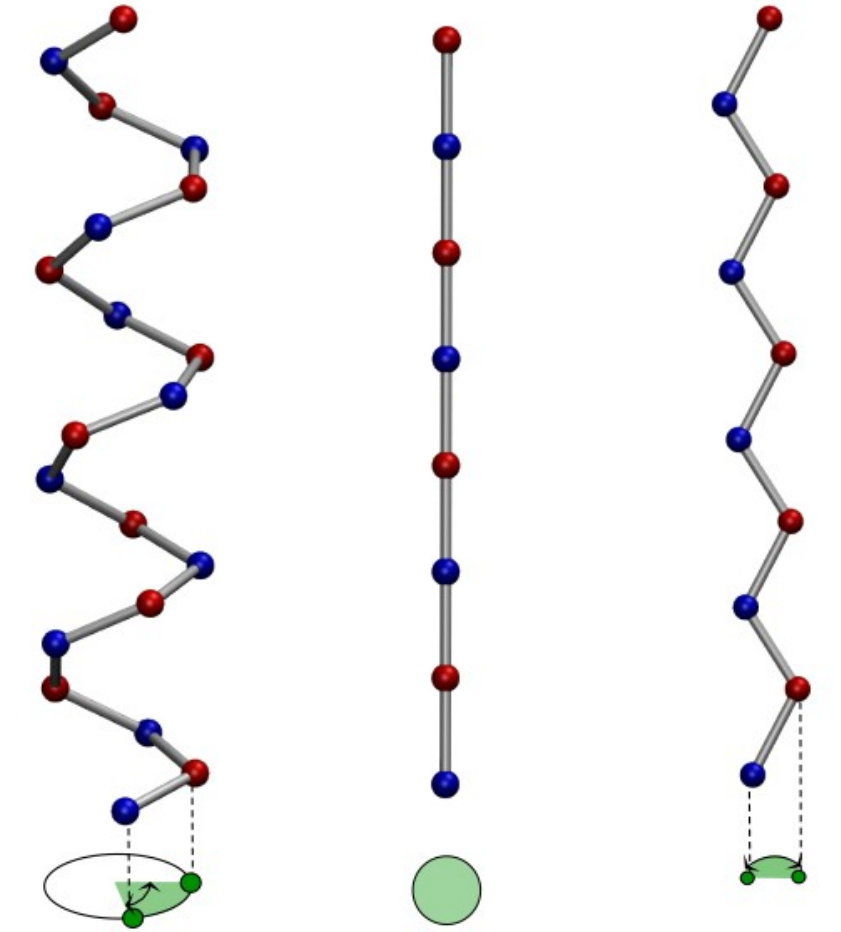
$\eta = P/(2\pi R) \approx 0.400$



# Discreteness: β-strands

$r(t) = R(\sin t, \cos t, \eta t)$
$\varepsilon$ is rotation angle
$t = n\,\varepsilon$, $n = 1, 2, 3, \ldots$
R = helix radius
P = helix pitch
$\eta = P/(2\pi R)$

- Three distinct geometries
- Discreteness enables a second building block



generic ε
3d

ε = 2π
1d

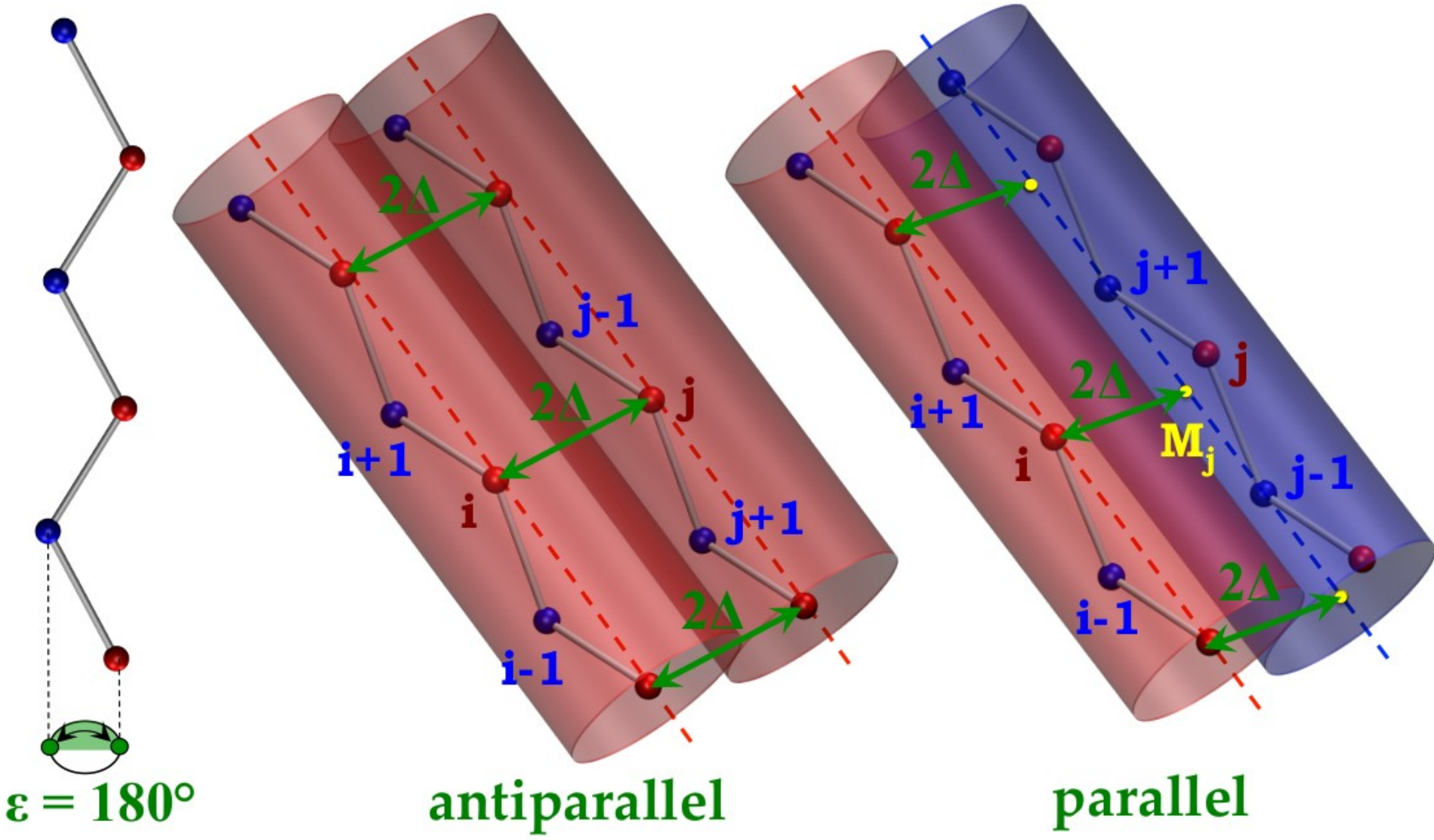
ε = π
2d

STANDARD HELIX

FORBIDDEN BY STERICS

RED AND BLUE SITES

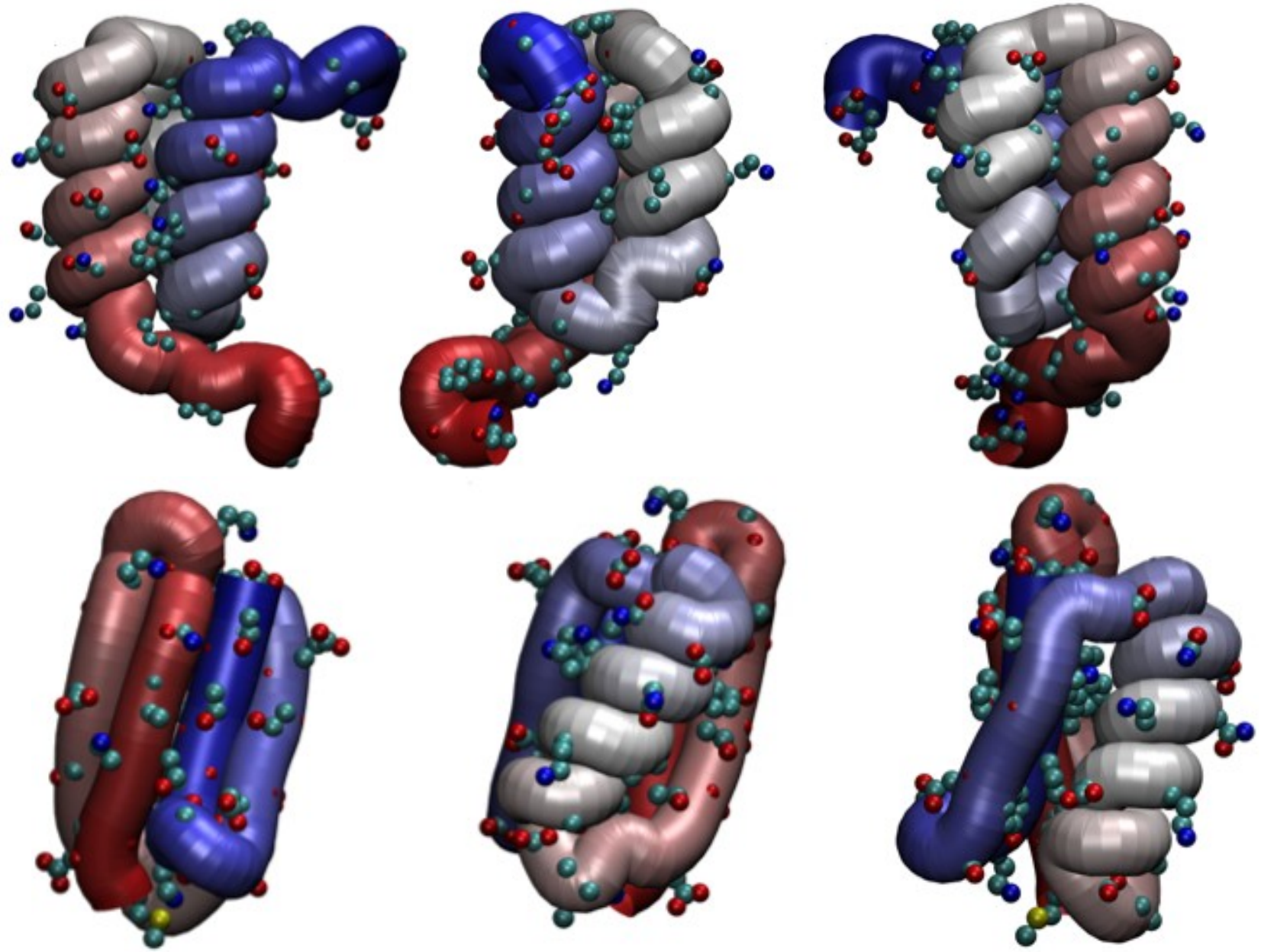
# Two types of assembly of β-strands



# Tube radius of 2.63Å

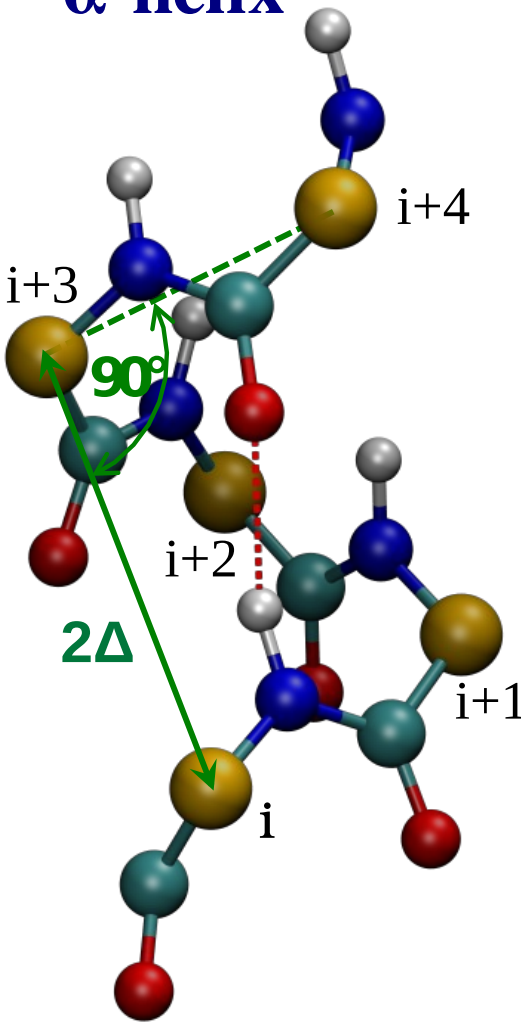
3α

α+4β



# Experimental data on more than 4000 protein structures from PDB are in accord with zero-parameter theory

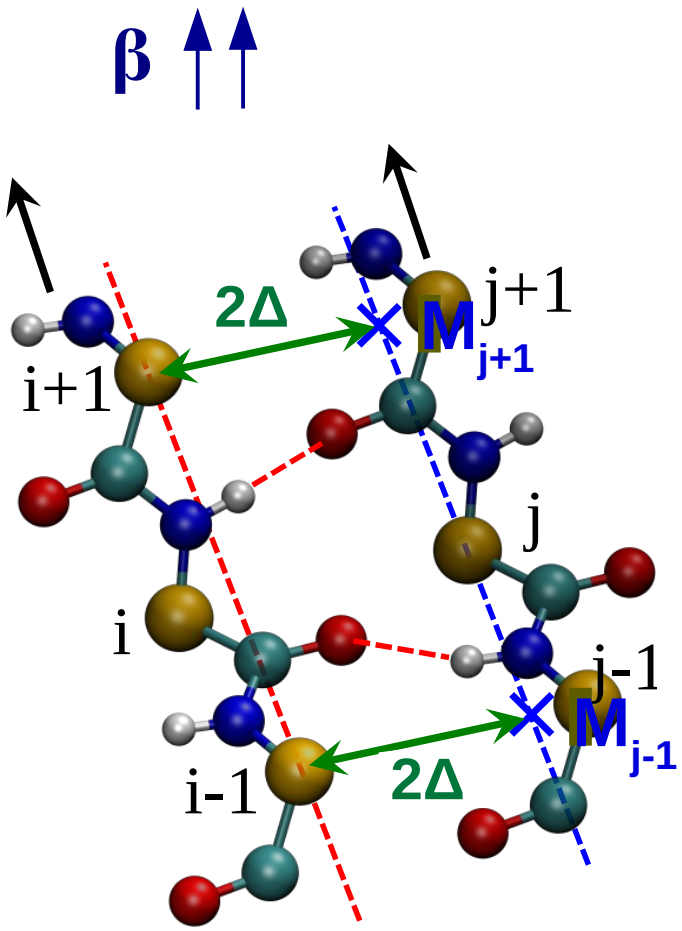
## α-helix




| HELIX | Theory | PDB data |
|---|---|---|
| Rotation angle $\varepsilon$ [°] | 99.8 | 99.1 ± 3.4 |
| # of residues per turn | 3.61 | 3.63 ± 0.13 |
| Helix radius R [Å] | 2.27 | 2.30 ± 0.07 |
| Helix pitch P [Å] | 5.69 | 5.47 ± 0.49 |
| Ratio $\eta = \mathrm{P}/(2\pi\mathrm{R})$ | 0.400 | 0.377 ± 0.046 |
| $\angle(i, i+3, i+4)$ [°] | 90.0 | 86.5 ± 3.9 |
| $d(i, i+3)$ [Å] | 2Δ=5.26 | 5.12 ± 0.16 |
| $\theta$ [°] | 91.8 | 91.3 ± 2.2 |
| $\mu$ [°] | 52.4 | 49.7 ± 3.9 |


Škrbić, Maritan, Giacometti, Rose & Banavar, *Phys. Rev. E* (2021)
Škrbić, Hoang, Giacometti, Maritan & Banavar, *Int. J. Mod. Phys. B* (2022)

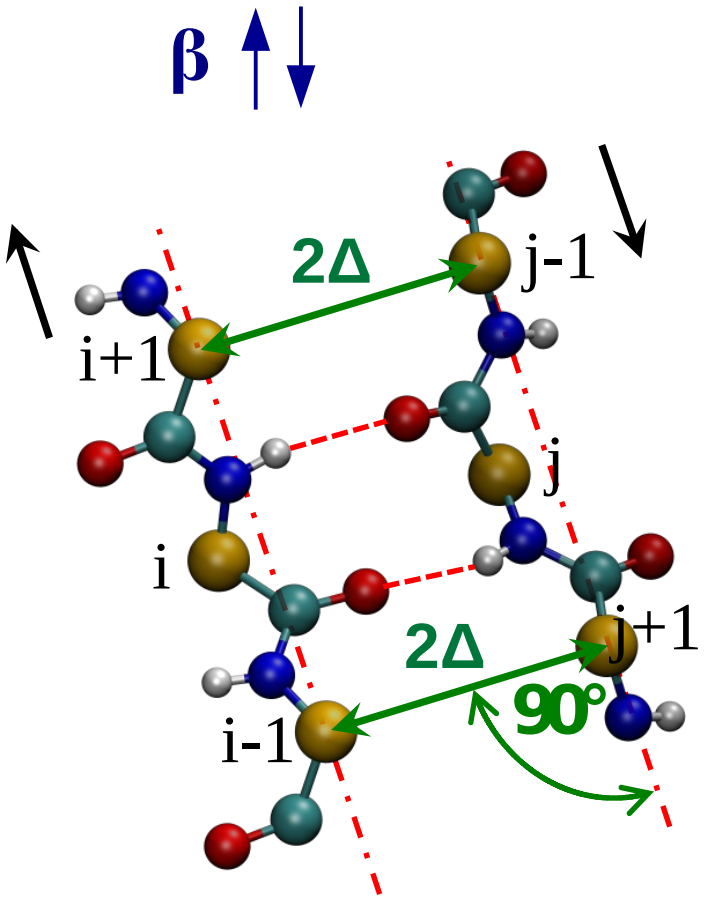
# Experimental data on more than 4000 protein structures from PDB are in accord with zero-parameter theory




| SHEET | Theory | PDB data |
|---|---|---|
| Parallel $\beta$-sheet | | |
| $\theta$ [°] | flexible | $121 \pm 10$ |
| $\mu$ [°] | $\sim 180$ | $191 \pm 17$ |
| $d(i+1, M_{j+1})$ [Å] | $2\Delta = 5.26$ | $5.26 \pm 0.16$ |
| $\angle(i-1, M_{j-1}, M_{j+1})$ [°] | 90.0 | $87.2 \pm 4.5$ |
| $\angle(M_{j-1}, i-1, i+1)$ [°] | 90.0 | $87.6 \pm 3.6$ |


Škrbić, Maritan, Giacometti, Rose & Banavar, *Phys. Rev. E* (2021)
Škrbić, Hoang, Giacometti, Maritan & Banavar, *Int. J. Mod. Phys. B* (2022)

# Experimental data on more than 4000 protein structures from PDB are in accord with zero-parameter theory




| SHEET | Theory | PDB data |
|---|---|---|
| Antiparallel $\beta$-sheet | | |
| $\theta$ [°] | flexible | $127 \pm 10$ |
| $\mu$ [°] | $\sim 180$ | $186 \pm 20$ |
| $d(i,j)$ [Å] | $2\Delta = 5.26$ | $5.26 \pm 0.20$ |
| $\angle(i-1, j+1, j+3)$ [°] | 90.0 | $87.0 \pm 4.2$ |
| $\angle(j+1, i-1, i+1)$ [°] | 90.0 | $86.4 \pm 4.1$ |


Škrbić, Maritan, Giacometti, Rose & Banavar, *Phys. Rev. E* (2021)
Škrbić, Hoang, Giacometti, Maritan & Banavar, *Int. J. Mod. Phys. B* (2022)

**Textbook paradigm: *Sequence determines structure***

⇩

***Sequence selects structure***

**from the menu of pre-determined folds:**
**Sequence independent free energy landscape**

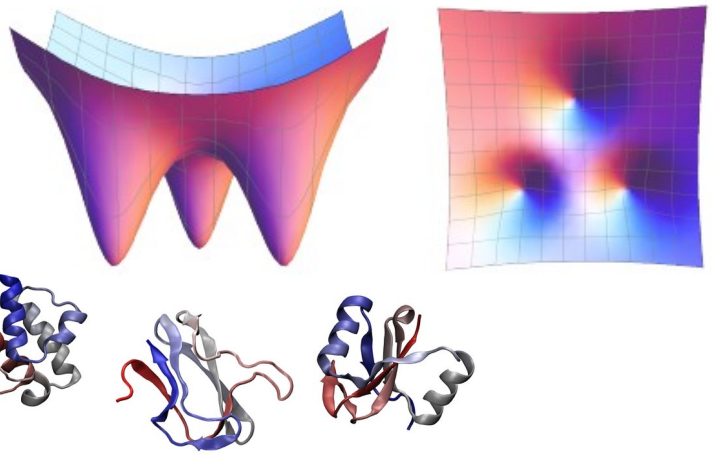
**– Banavar, Maritan et al. (2000 onwards)**

**Finite # of distinct folds (~10,000 structures in PDB)**

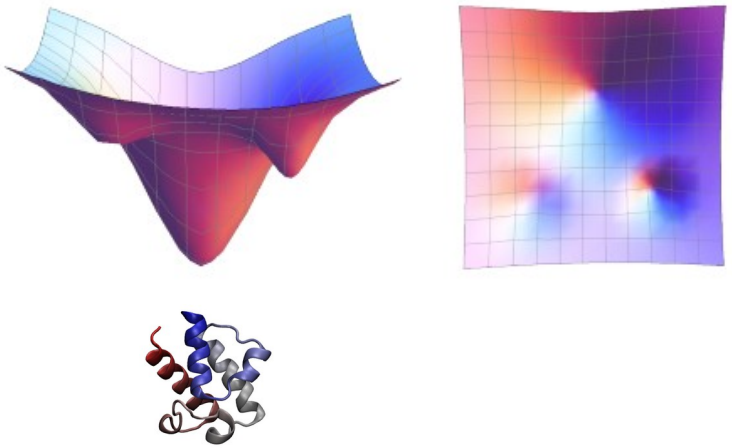
**vs.**

**Astronomical # of putative sequences (~100 $^{20}$)**

**Homopolymer :**
***all minima are equivalent***



**Specific sequence:**
***one minimum is selected***



# Summary

* First principles, zero parameter predictions of protein building blocks geometries validated by experimental structures

* Discreteness is essential for life

* Possibility of new classes of protein-like nano-machines not necessarily based on carbon chemistry

* While sequences and functionalities undergo *Darwinian evolution*, protein structures are *Platonic and immutable* – this helps maintain functionality during evolution

* Future investigation: understanding the role of amino acid specificity in dictating the choice of the native state fold

* Consilience: Chemistry & Biology provide a perfect fit to the dictates of Mathematics & Physics