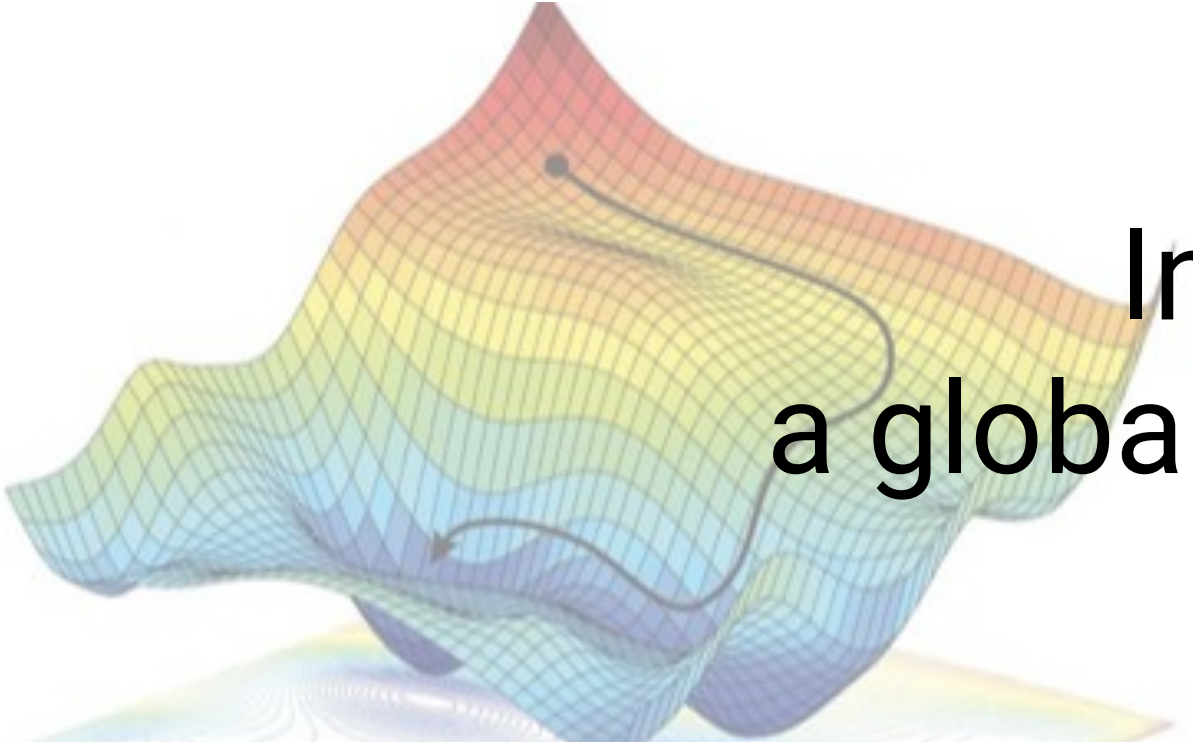
Class core values

1. Be **respect**ful to yourself and others
2. Be **confident** and believe in yourself
3. Always do your **best**
4. Be **cooperative**
5. Be **creative**
6. Have **fun**
7. Be **patient** with yourself while you learn
8. Don’t be shy to **ask “stupid” questions**



Week 6, Lecture 1

In search of a global minimum

Learning Objectives

1. Evaluate a problem to see if computational de novo design is a good fit for it
2. Understand the underlying concept of de novo design
3. Identify challenges in assessing energies of the systems
4. Describe a force field and its basic components
5. Identify proper minimization methods to use for design problems

De novo design

De novo design is the process of generating proteins from scratch

Creating new proteins often by physics-based and heuristic methods

de no·vo

/ˌdə ˈnōvō,dē ˈnōvō/

adverb

from the beginning; anew.
"in a pure meritocracy, everyone must begin de novo"

adjective

starting from the beginning.
"a general strategy for de novo protein design"

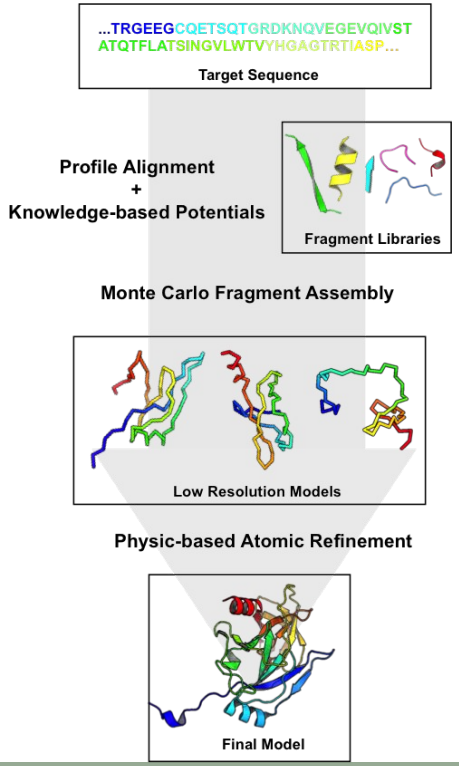
Definitions from Oxford Languages

De novo design is the process of generating proteins from scratch

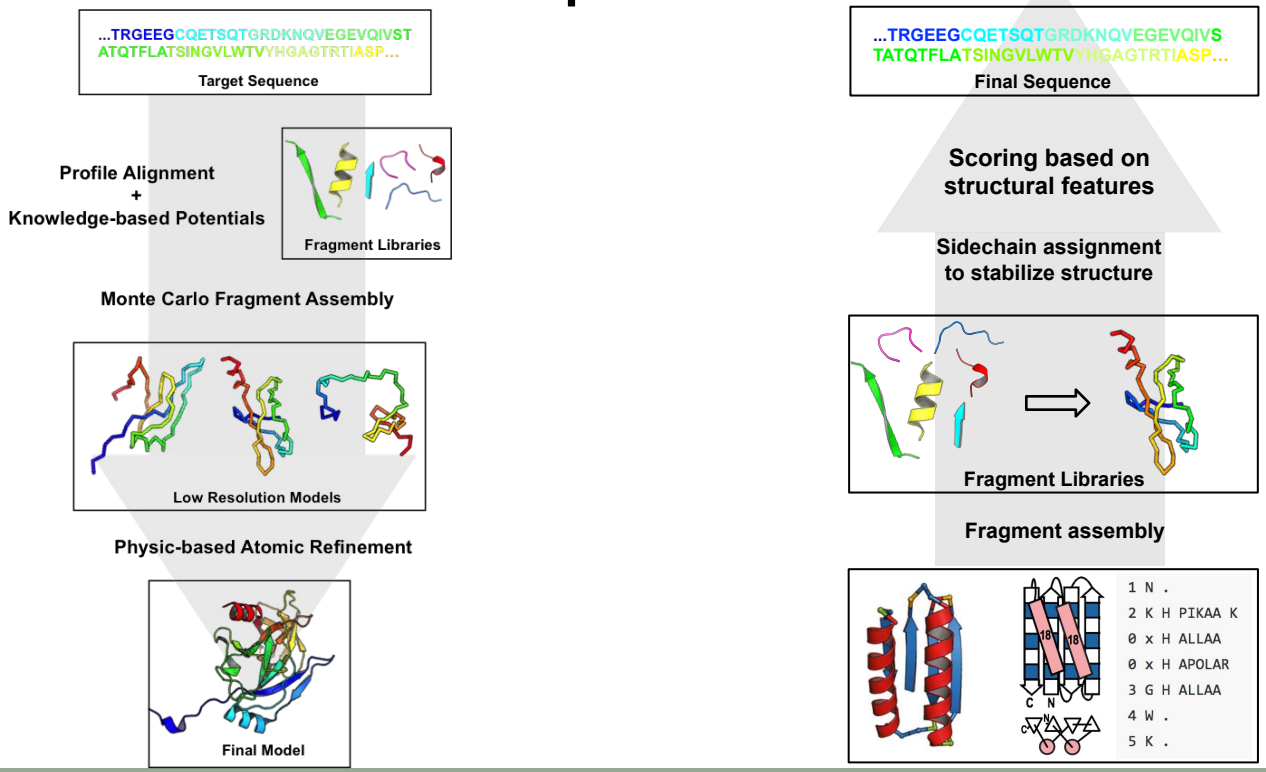
Creating new proteins often by physics-based and heuristic methods

Advantage	Disadvantage
If it works, it generates really good proteins because it considers all the rules	Low success rate
Novelty = less chance of resistance/defense	novel structures → may cause unknown problems downstream
Can create novel function/structure and is not limited to what we know	While not dependent on prior knowledge, still needs the rules
Great for specificity and full control	

In its most general form, de novo design is the reverse of structure prediction



In its most general form, de novo design is the reverse of structure prediction



The general process of de novo design

What is a good de novo design problem?

What is a good de novo design problem?

1. Is it **describable**?

What is a good de novo design problem?

1. Is it **describable**?
2. Is it **solvable**?

What is a good de novo design problem?

1. Is it **describable**?
2. Is it **solvable**?
3. Is it **testable**?

What is a good de novo design problem?

1. Is it **describable**?
2. Is it **solvable**?
3. Is it **testable**?
4. Is it **tractable**?

What is a good de novo design problem?

1. Is it **describable**?
2. Is it **solvable**?
3. Is it **testable**?
4. Is it **tractable**?
5. Is it **non-trivial**?

What is a good de novo design problem?

1. Is it **describable**?
2. Is it **solvable**?
3. Is it **testable**?
4. Is it **tractable**?
5. Is it **non-trivial**?

Designing proteins for function and predicting their structure are examples of good design problems.

Everything lies in the objective function

1. Pen-and-paper design

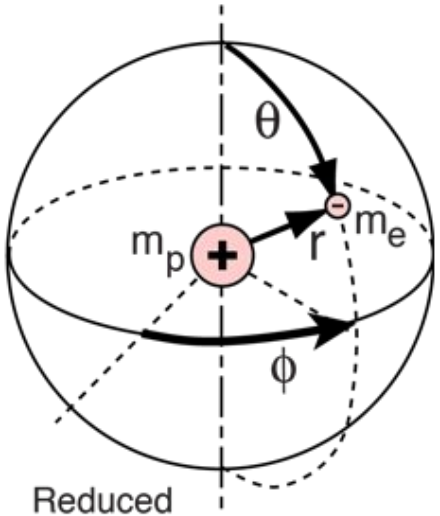
Everything lies in the objective function

1. Pen-and-paper design
    a. Physico-chemical knowledge about proteins and amino acids

Everything lies in the objective function

1. Pen-and-paper design
2. Automated design
    a. Gibbs free energy

Quantum mechanical calculations of free energy are intractable



Reduced mass $\mu = \frac{m_e m_p}{m_e + m_p}$

SCHRÖDINGER EQUATION

$$\hat{H}\Psi = E\Psi$$

$\hat{H}$ = Hamiltonian operator

Ψ = wavefunction (eigenfunction)

E = Energy (eigenvalue) → observable

For 1-D:

$$\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\Psi(x) + V(x)\Psi(x) = E\Psi(x)$$

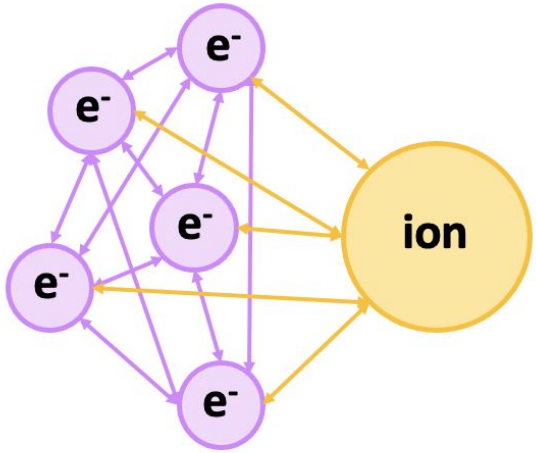
$\hat{K}$, Kinetic Energy $\hat{V}$, Potential Energy

↳ Classical counterpart:

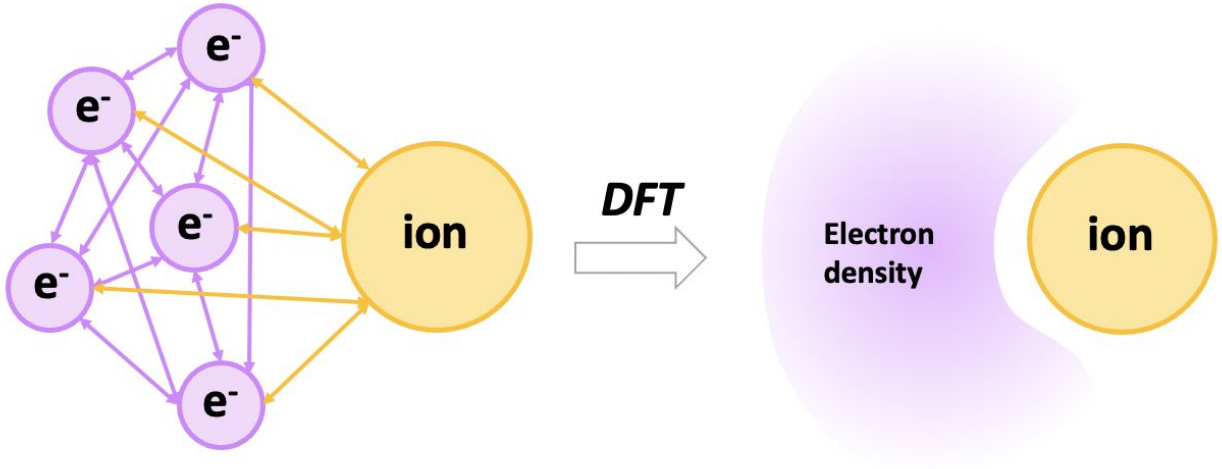
$K_E = \frac{1}{2}mv_x^2$ or $K_E = \frac{p_x^2}{2m}$

Image from hyperphysics: Dr. Debbie Gale Mitcher (@heydebigale)

Simplifications of the original Schrödinger's equation makes it tractable for small systems



Simplifications of the original Schrödinger's equation makes it tractable for small systems



Simplifications of the original Schrödinger's equation makes it tractable for small systems

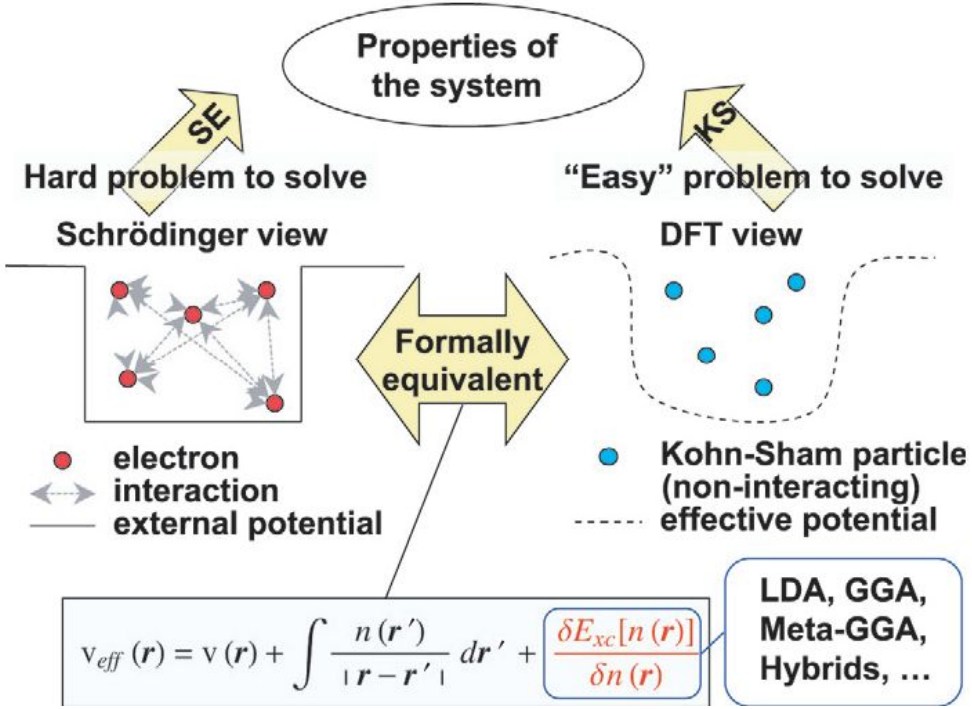
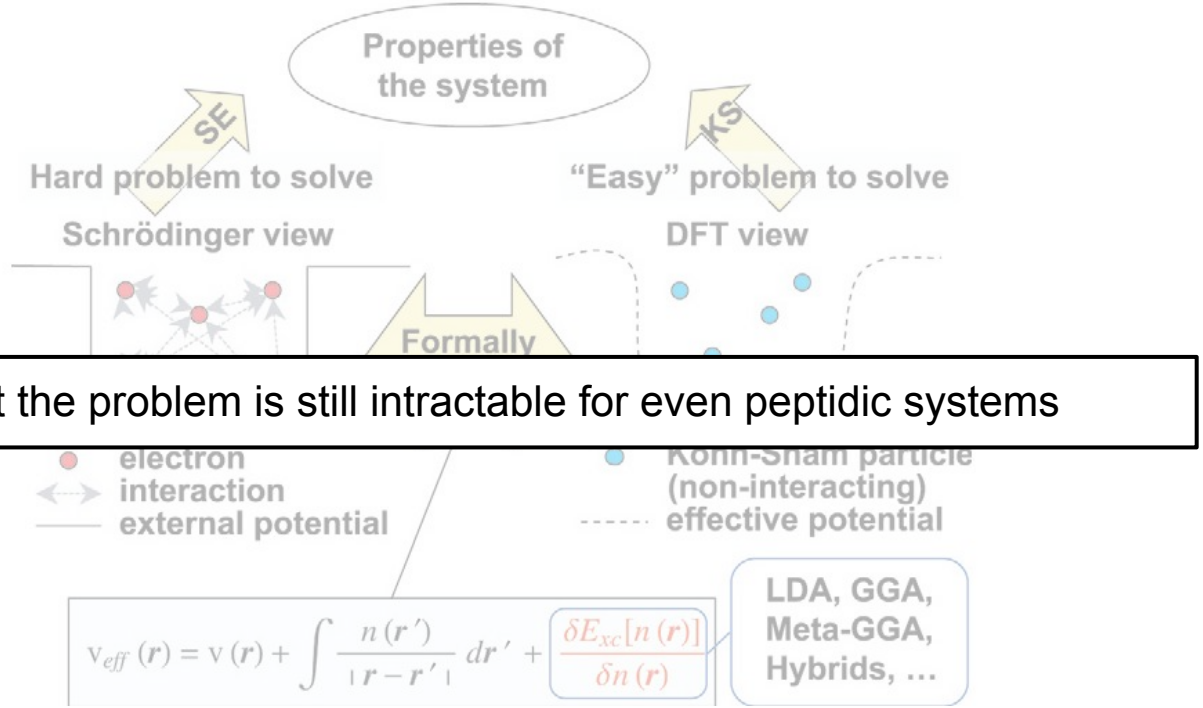


Image from Leung et al, *Modelling Simul. Matter Sci.* 2005 **13** R1

Simplifications of the original Schrödinger's equation makes it tractable for small systems



But the problem is still intractable for even peptidic systems

Image from Leung et al, *Modelling Simul. Matter Sci.* 2005 **13** R1

Force fields are generated to approximate the energies of the system

Newtonian systems that simplify interactions among molecules

Force fields are generated to approximate the energies of the system

Newtonian systems that simplify interactions among molecules

Reactions Involving Hydrogen Molecules and Atoms

J. Chem. Phys. 4, 170 (1936); https://doi.org/10.1063/1.1749815

J. Hirschfelder, H. Eyring, *and* B. Topley

In class activity

Writing up a simple forcefield

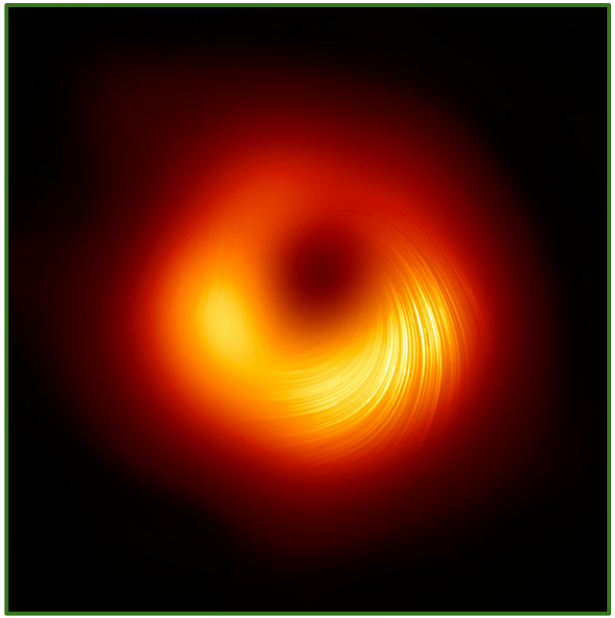


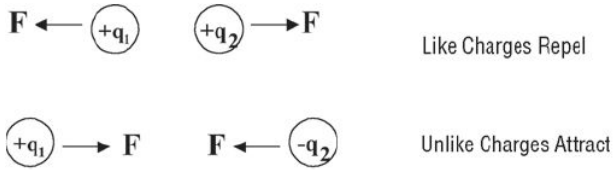
Image from Scientific American: Magnetic force field around a black hole

A simple force field

E = Intermolecular interactions + Intramolecular interactions

A simple force field

E = **Intermolecular interactions** + Intramolecular interactions



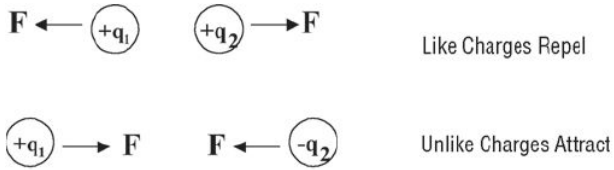
$$F = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r^2}\right) = k\frac{q_1 q_2}{r^2}$$

$$k \approx 9\times10^9\ N\,m^2/C^2$$

Coulomb's law

A simple force field

E = **Intermolecular interactions** + Intramolecular interactions

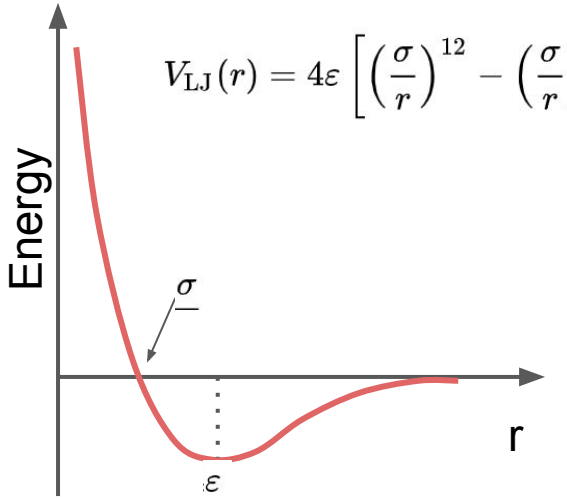


$$F = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r^2}\right) = k\frac{q_1 q_2}{r^2}$$

$$k \approx 9\times10^9\ N\,m^2/C^2$$

Coulomb's law

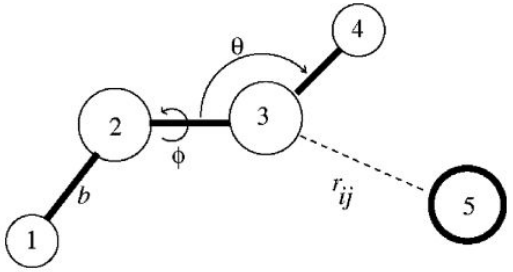
$$V_{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



Lennard-Jones potential

A simple force field

E = Intermolecular interactions + **Intramolecular interactions**



$$E = \underset{\text{all bonds}}{\Sigma \tfrac{1}{2} K_b (b - b_0)^2} + \underset{\text{all bond angles}}{\Sigma \tfrac{1}{2} K_\tau (\tau - \tau_0)^2} + \underset{\text{all dihedral angles}}{\Sigma \tfrac{1}{2} K_\theta \{1 + \cos(n\theta - \delta)\}}$$

$$+ \underset{\substack{\text{all non-bonded} \\ \text{pairs}}}{\Sigma \epsilon_{ij} \{(r^0_{ij}/r_{ij})^{12} - 2 (r^0_{ij}/r_{ij})^6\}} + \underset{\substack{\text{all atomic} \\ \text{co-ordinates}}}{\Sigma \tfrac{1}{2} w (x_i - x^0_i)^2}$$

Force field generation has been a long quest

CHARMM: A program for macromolecular energy, minimization, and dynamics calculations†

Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, Martin Karplus

First published: Summer 1983 | https://doi.org/10.1002/jcc.540040211 | Citations: 10,422

Force field generation has been a long quest

Intramolecular (internal, bonded terms)

$$\sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{\substack{improper \\ dihedrals}} K\phi(\phi - \phi_0)^2 + \sum_{Urey-Bradley} K_{UB}\left(r_{1,3} - r_{1,3;0}\right)^2 \quad (1)$$

Intermolecular (external, nonbonded terms)

$$\sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

J Comput Chem. 2010 Mar; 31(4): 671–690.

Force field generation has been a long quest

Ariel Warshel

Martin Karplus

Michael Levitt

Force fields generated for molecular dynamics are not well suited for design

Design force fields simplify physics-based terms and combine them with heuristic terms

$$\Delta E_{\text{total}} = \sum_i w_i E_i(\Theta_i, \text{aa}_i)$$

Table 1. Summary of Terms in REF15 for Proteins

term	description	weight	units	ref(s)
fa_atr	attractive energy between two atoms on different residues separated by a distance d	1.0	kcal/mol	5, 6
fa_rep	repulsive energy between two atoms on different residues separated by a distance d	0.55	kcal/mol	5, 6
fa_intra_rep	repulsive energy between two atoms on the same residue separated by a distance d	0.005	kcal/mol	5, 6
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol	36
lk_ball_wtd	orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol	50, 71
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol	36
fa_elec	energy of interaction between two nonbonded charged atoms separated by a distance d	1.0	kcal/mol	50
hbond_lr_bb	energy of short-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sr_bb	energy of long-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_bb_sc	energy of backbone−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sc	energy of side-chain−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
dslf_fa13	energy of disulfide bridges	1.25	kcal/mol	49
rama_prepro	probability of backbone ϕ, ψ angles given the amino acid type	(0.45 kcal/mol)/kT	kT	50, 51
p_aa_pp	probability of amino acid identity given backbone ϕ, ψ angles	(0.4 kcal/mol)/kT	kT	51
fa_dun	probability that a chosen rotamer is native-like given backbone ϕ, ψ angles	(0.7 kcal/mol)/kT	kT	52
omega	backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	(0.6 kcal/mol)/AU	AU[a]	72
pro_close	penalty for an open proline ring and proline ω bonding energy	(1.25 kcal/mol)/AU	AU	51
yhh_planarity	sinusoidal penalty for nonplanar tyrosine χ_3 dihedral angle	(0.625 kcal/mol)/AU	AU	49
ref	reference energies for amino acid types	(1.0 kcal/mol)/AU	AU	1, 51

[a]AU = arbitrary units.

Table 1. Summary of Terms in REF15 for Proteins

term	description	weight	units	ref(s)
fa_atr	attractive energy between two atoms on different residues separated by a distance d	1.0	kcal/mol	5, 6
fa_rep	repulsive energy between two atoms on different residues separated by a distance d	0.55	kcal/mol	5, 6
fa_intra_rep	repulsive energy between two atoms on the same residue separated by a distance d	0.005	kcal/mol	5, 6
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol	36
lk_ball_wtd	orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol	50, 71
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol	36
fa_elec	energy of interaction between two nonbonded charged atoms separated by a distance d	1.0	kcal/mol	50
hbond_lr_bb	energy of short-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sr_bb	energy of long-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_bb_sc	energy of backbone−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sc	energy of side-chain−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
dslf_fa13	energy of disulfide bridges	1.25	kcal/mol	49
rama_prepro	probability of backbone ϕ, ψ angles given the amino acid type	(0.45 kcal/mol)/kT	kT	50, 51
p_aa_pp	probability of amino acid identity given backbone ϕ, ψ angles	(0.4 kcal/mol)/kT	kT	51
fa_dun	probability that a chosen rotamer is native-like given backbone ϕ, ψ angles	(0.7 kcal/mol)/kT	kT	52
omega	backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	(0.6 kcal/mol)/AU	AU[a]	72
pro_close	penalty for an open proline ring and proline ω bonding energy	(1.25 kcal/mol)/AU	AU	51
yhh_planarity	sinusoidal penalty for nonplanar tyrosine χ_3 dihedral angle	(0.625 kcal/mol)/AU	AU	49
ref	reference energies for amino acid types	(1.0 kcal/mol)/AU	AU	1, 51

[a]AU = arbitrary units.

Table 1. Summary of Terms in REF15 for Proteins

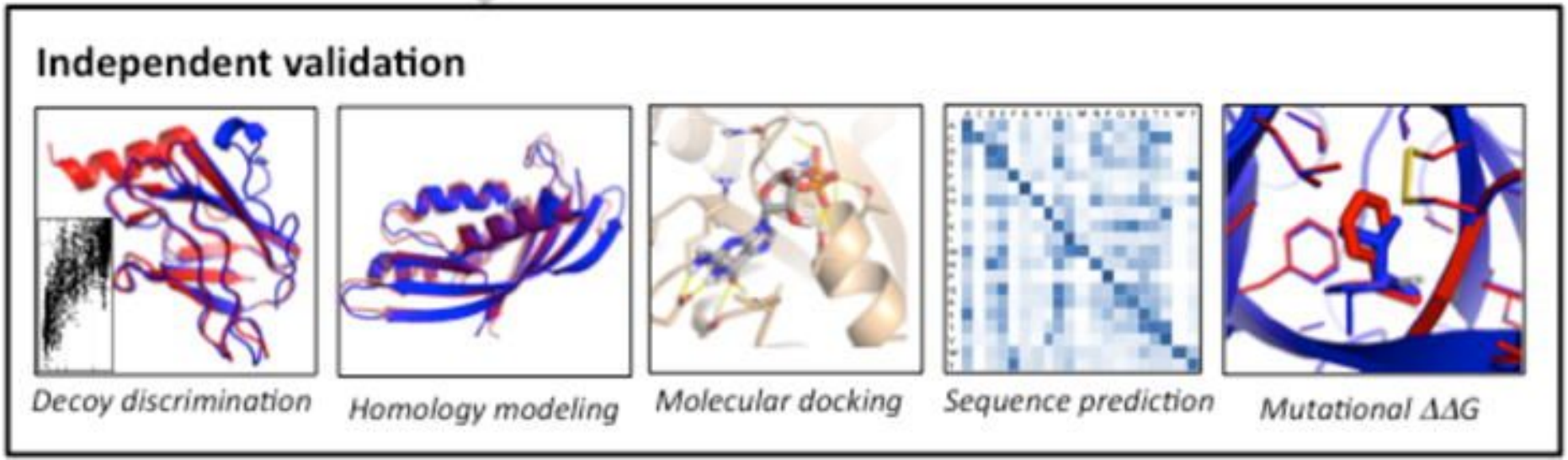
term	description	weight	units	ref(s)
fa_atr	attractive energy between two atoms on different residues separated by a distance d	1.0	kcal/mol	5, 6
fa_rep	repulsive energy between two atoms on different residues separated by a distance d	0.55	kcal/mol	5, 6
fa_intra_rep	repulsive energy between two atoms on the same residue separated by a distance d	0.005	kcal/mol	5, 6
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol	36
lk_ball_wtd	orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol	50, 71
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol	36
fa_elec	energy of interaction between two nonbonded charged atoms separated by a distance d	1.0	kcal/mol	50
hbond_lr_bb	energy of short-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sr_bb	energy of long-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_bb_sc	energy of backbone−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sc	energy of side-chain−side-chain hydrogen bonds	1.0	kcal/mol	38, 49
dslf_fa13	energy of disulfide bridges	1.25	kcal/mol	49
rama_prepro	probability of backbone ϕ, ψ angles given the amino acid type	(0.45 kcal/mol)/kT	kT	50, 51
p_aa_pp	probability of amino acid identity given backbone ϕ, ψ angles	(0.4 kcal/mol)/kT	kT	51
fa_dun	probability that a chosen rotamer is native-like given backbone ϕ, ψ angles	(0.7 kcal/mol)/kT	kT	52
omega	backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	(0.6 kcal/mol)/AU	AU[a]	72
pro_close	penalty for an open proline ring and proline ω bonding energy	(1.25 kcal/mol)/AU	AU	51
yhh_planarity	sinusoidal penalty for nonplanar tyrosine χ_3 dihedral angle	(0.625 kcal/mol)/AU	AU	49
ref	reference energies for amino acid types	(1.0 kcal/mol)/AU	AU	1, 51

[a]AU = arbitrary units.

Force fields makes assumptions/simplifications

Force fields makes assumptions/simplifications
→ they must be parameterized/trained

Force fields makes assumptions/simplifications → they must be parameterized/trained



In class activity

Score optimization

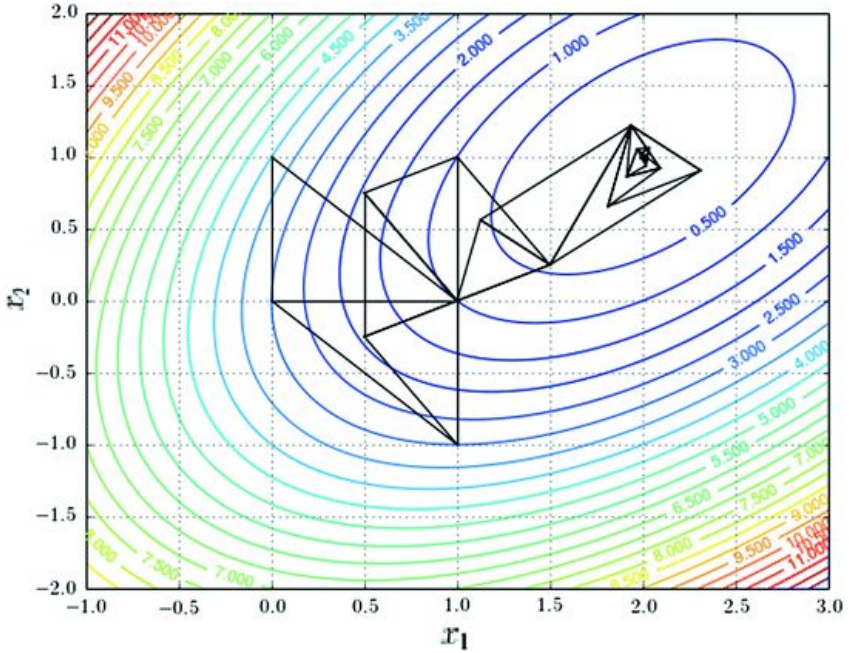


Image from: *Advances in hydroinformatics* **2018** pp997-1012

Methods for optimizing the objective function (value of force field) depend on our problem

1. Brute-force trial of all possible cases
2. Branching and dead-end elimination
3. Heuristic approaches

In class activity

Finding the minimum

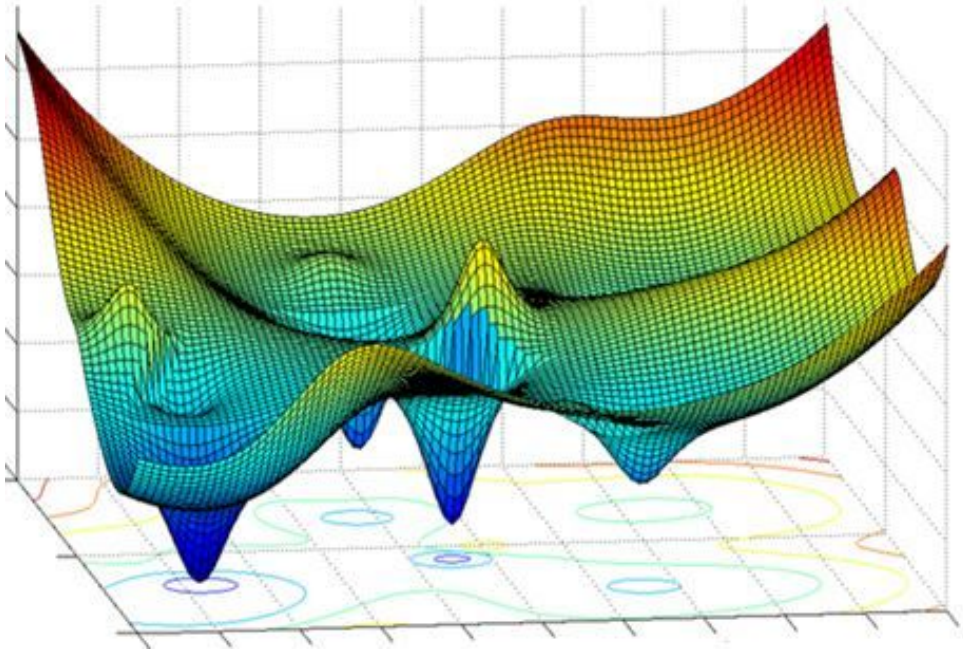
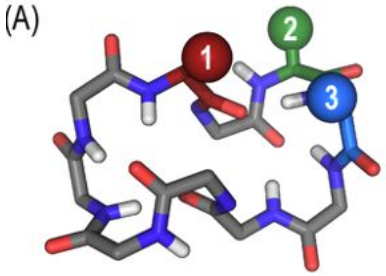


Image from: *From the Genesis*, Gradient descent part I

Methods for optimizing the objective function (value of force field) depend on our problem



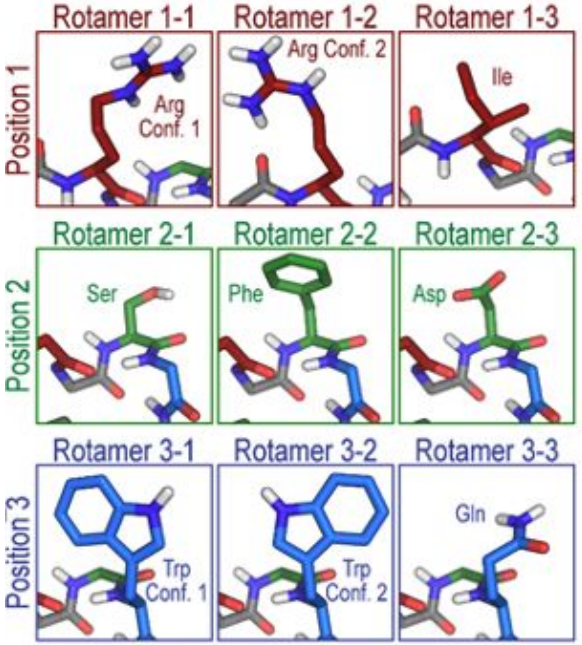


Image from Mulligan and Hosseinzadeh, ACS Symposium Series, In press

Methods for optimizing the objective function (value of force field) depend on our problem

1. Brute-force trial of all possible cases

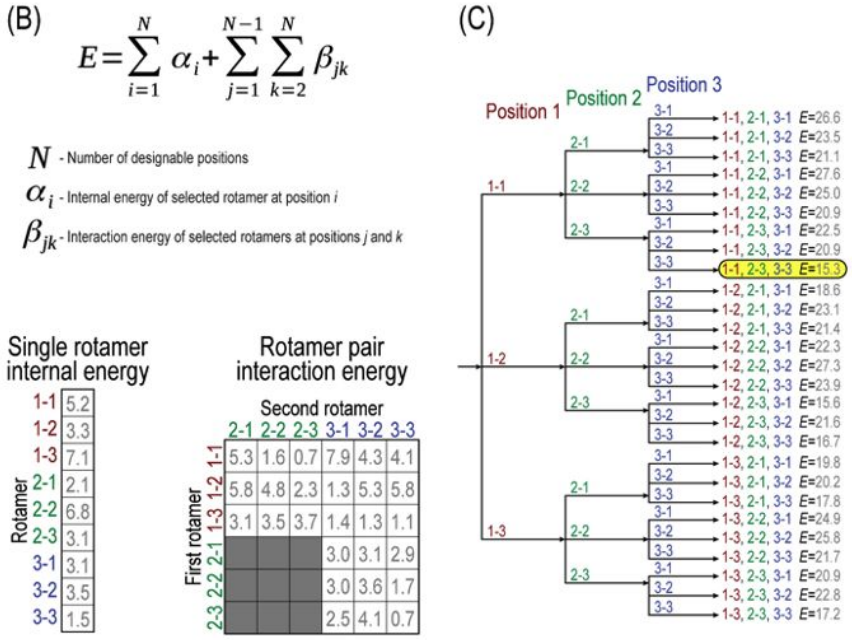


Image from Mulligan and Hosseinzadeh, ACS Symposium Series, In press

Methods for optimizing the objective function (value of force field) depend on our problem

1. Brute-force trial of all possible cases
2. Branching and dead-end elimination

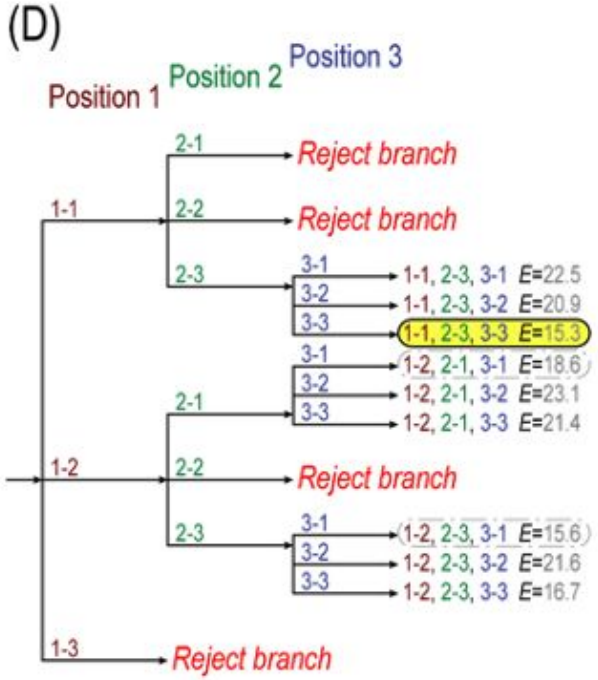


Image from Mulligan and Hosseinzadeh, ACS Symposium Series, In press

Methods for optimizing the objective function (value of force field) depend on our problem

1. Brute-force trial of all possible cases
2. Branching and dead-end elimination
3. Heuristic approaches

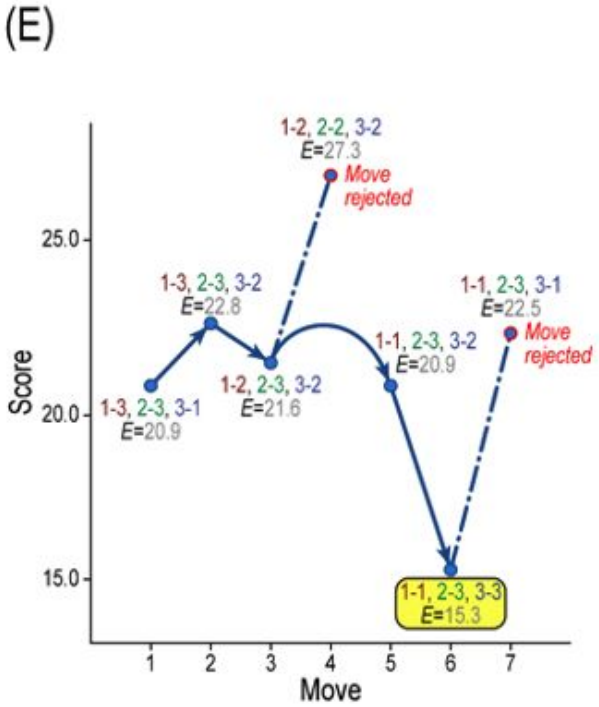


Image from Mulligan and Hosseinzadeh, ACS Symposium Series, In press

Methods for optimizing the objective function (value of force field) depend on our problem

1. Brute-force trial of all possible cases
2. Branching and dead-end elimination
3. Heuristic approaches

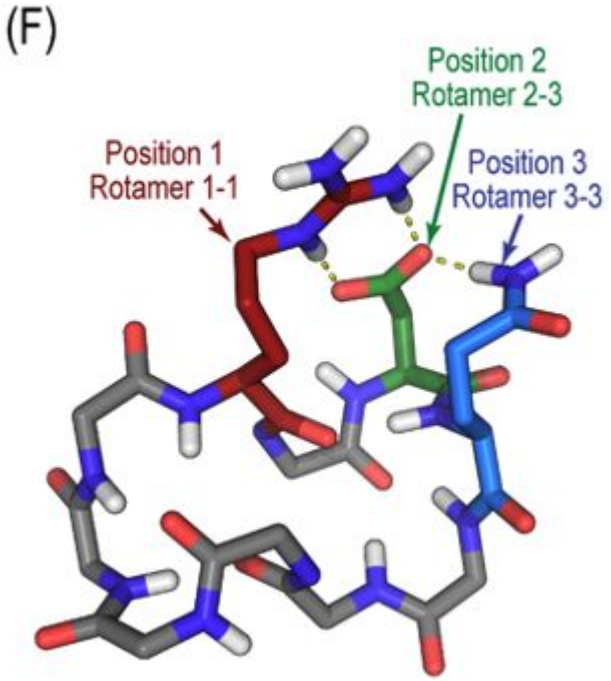


Image from Mulligan and Hosseinzadeh, ACS Symposium Series, In press

For the next lecture:

1. Read JC paper moderated by **group III**
2. Keep working on your specific aims!

Next lecture:

To control function is to control structure

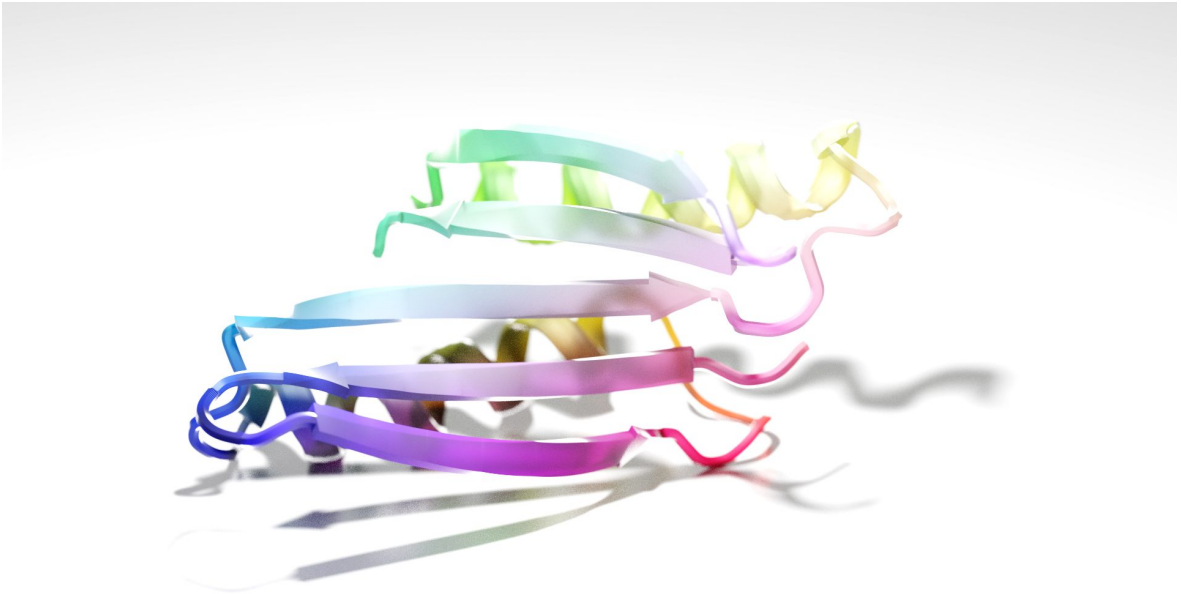


Image from IPD website: Deep learning dream up new protein structures