1. Proteins as drugs

2. Designing protein drugs *de novo*

Tanja Kortemme

Introduction

- About me:
  Background in Physical Chemistry & Biophysics
  Ph.D. in Biophysics from EMBL, Heidelberg, Germany

  Research in quantitative biology (computational & experimental)
  http://kortemmelab.ucsf.edu

- Development of computational protein design methods since 1999
  https://www.rosettacommons.org

- Earlier review on our work: Mandell et al, “Computer-aided design of functional protein interactions” *Nature Chemical Biology* 2009
- Recent highlight: Glasgow et al, “Computational design of a modular protein sense-response system” *Science* 2019

Agenda

Background

- Why protein drugs

State of the field

- Principles of engineering & optimizing antibody drugs

Current research

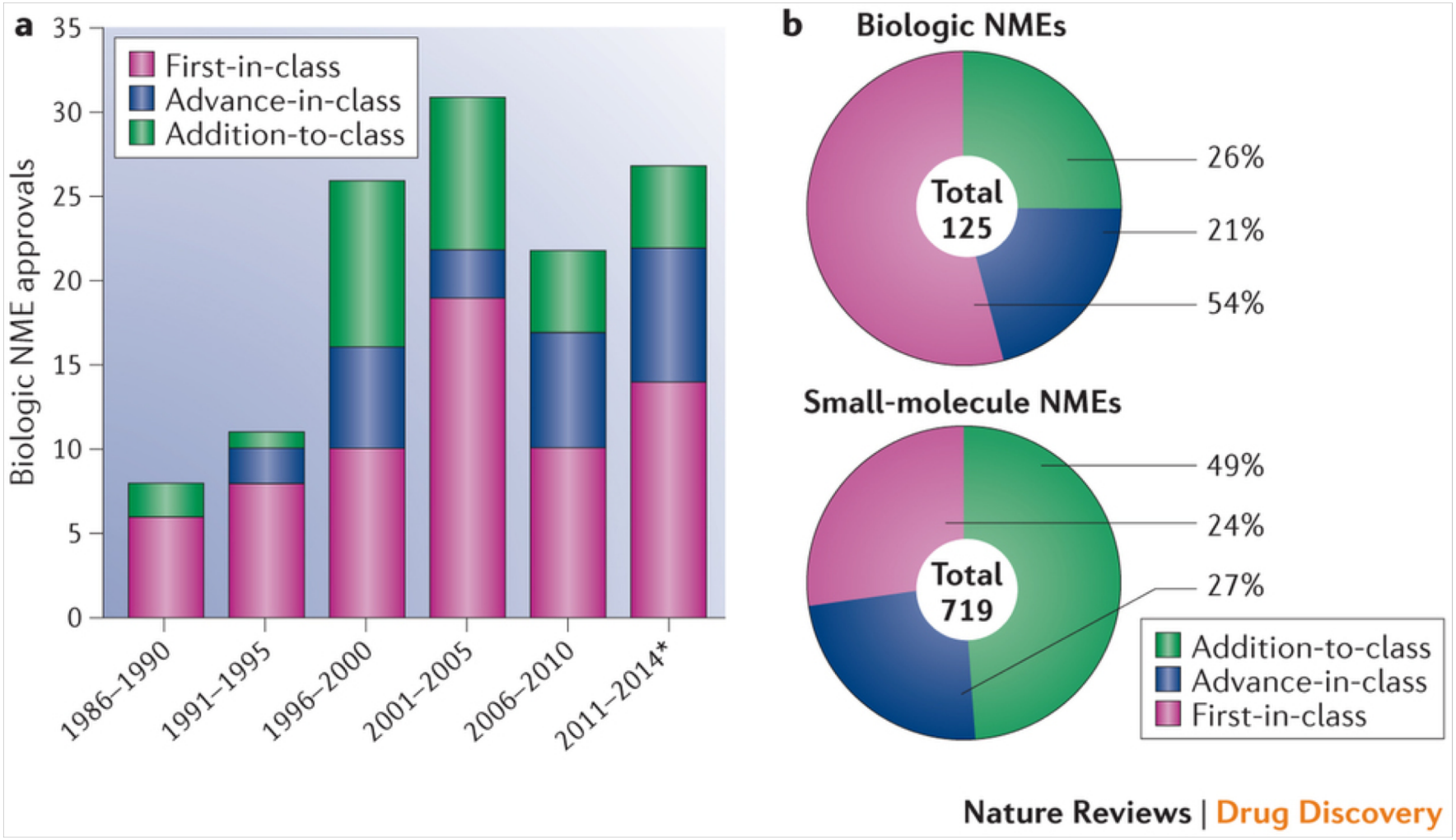
- Creating protein drugs from scratch
- Computational protein design (Rosetta)
- Successes and challenges

Proteins are important therapeutics

- > 200 protein therapeutics on the market

Name	Target disease	Market (2011)
Adalimumab/Humira Pen Etanercept/Enbrel	Rheumatoid arthritis	$8 billion
Infliximab/Remicade	arthritis	
Rituximab/Rituxan	Non-Hodgkin's B-cell lymphoma	$7 billion
Bevacizumab/Avastin Trastuzumab/Herceptin	Colorectal cancer Breast cancer	$6-5 billion
Insulin glargine/Lantus	Type I and II diabetes	$4.8 billion
Epogen (erythropoietin)	Renal anemia	$2.5 billion

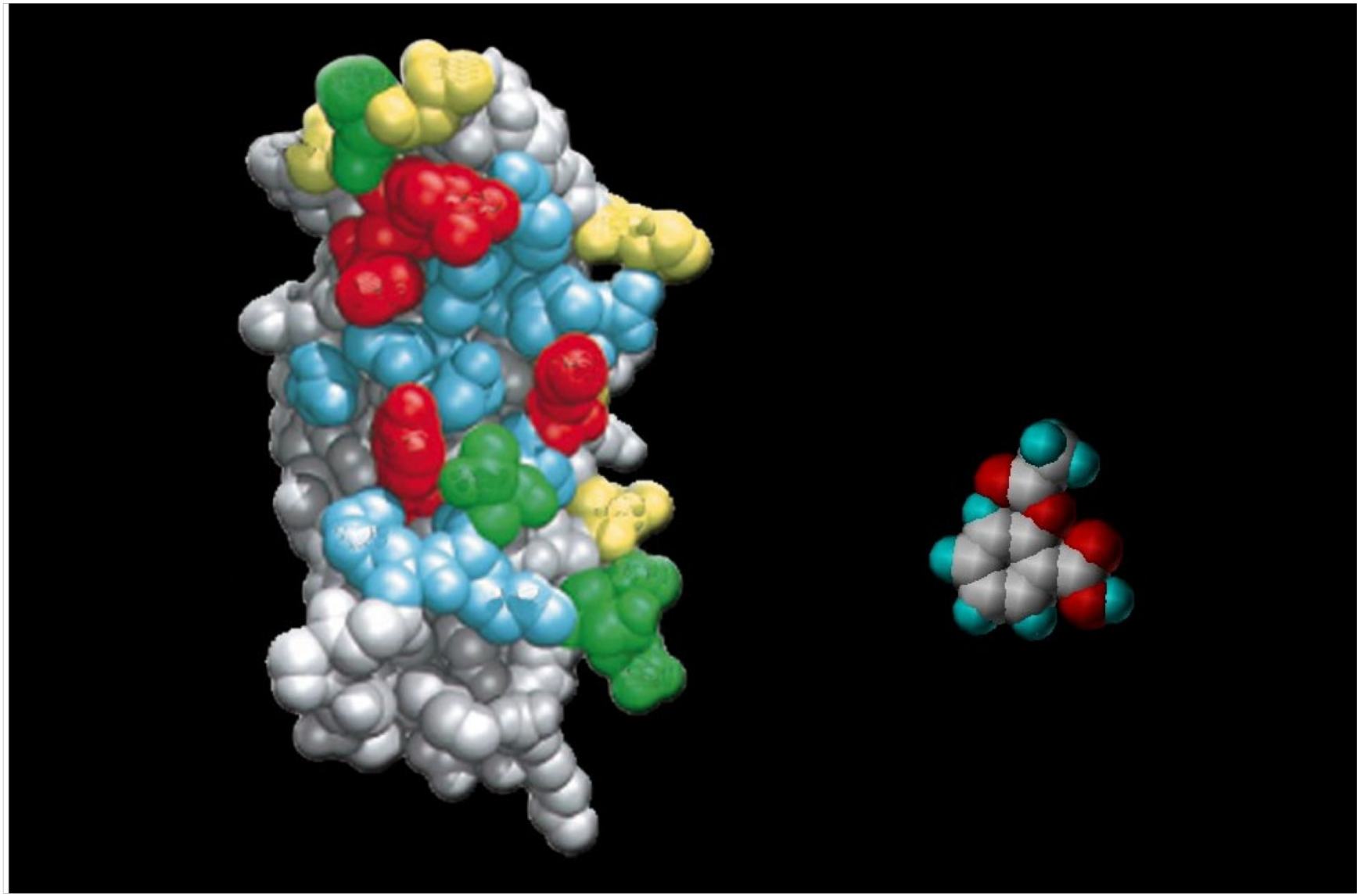
Biologics are innovative



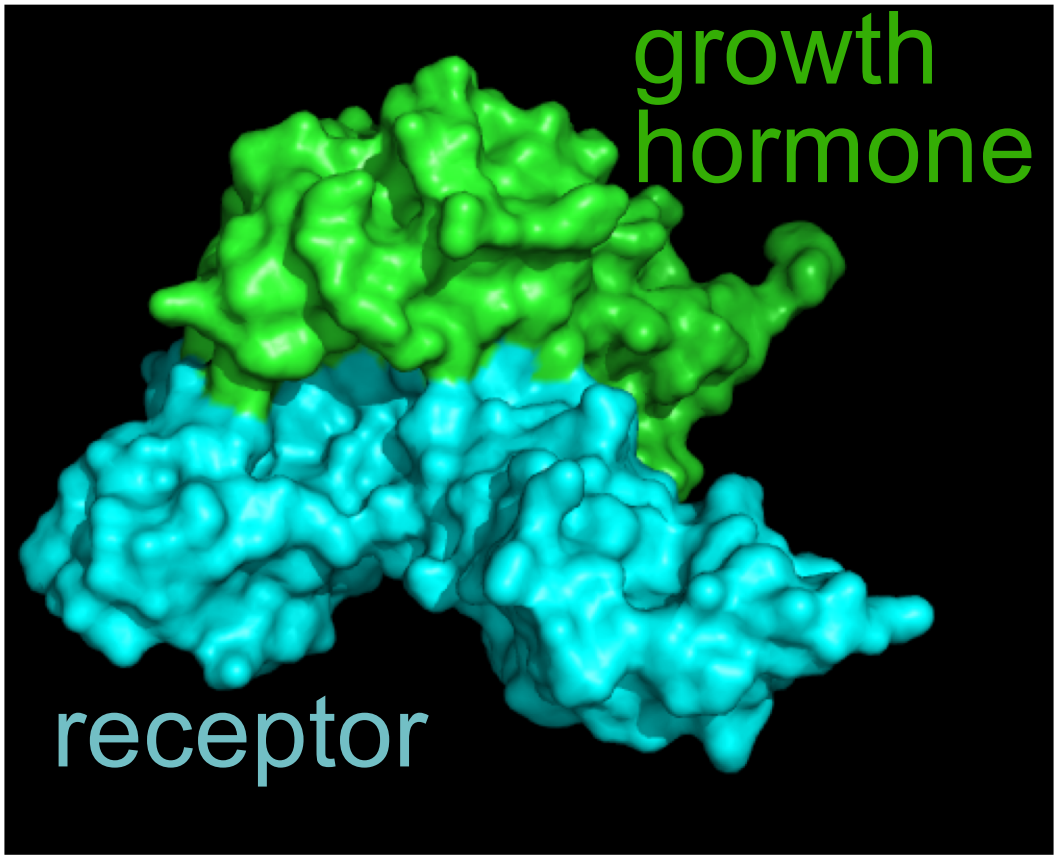
a | Biologic NME approvals by the US Food and Drug Administration in 1986–2014, split into innovation categories and 5-year time periods. **b** | Comparison of the innovativeness of biologic and small-molecule NMEs approved in 1986–2014. NME, new molecular entity. *The last bar is only a 4-year time period.

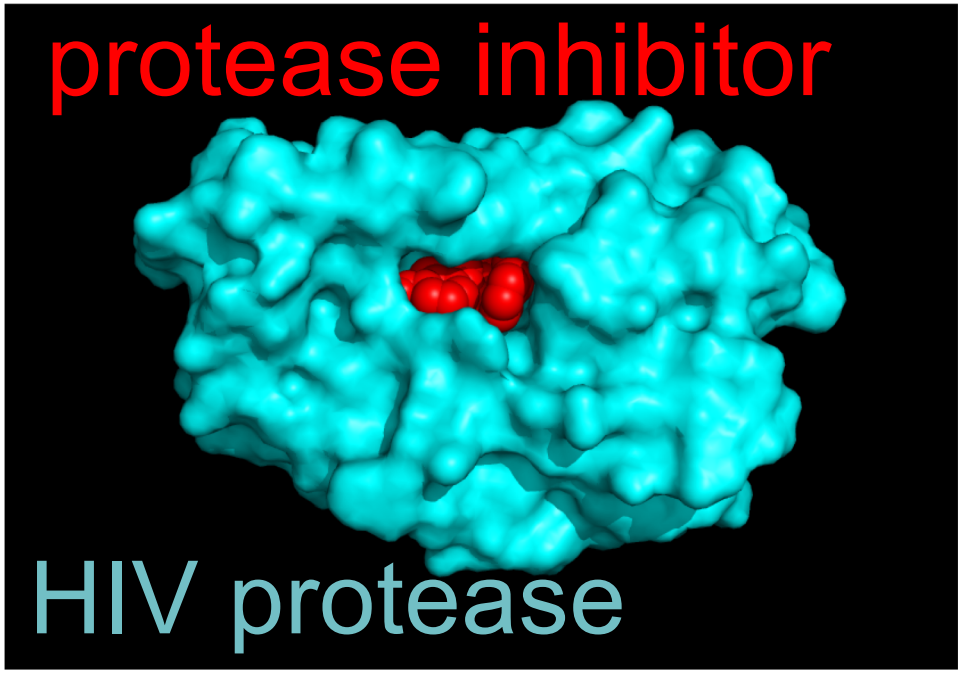
***Nature Reviews Drug Discovery* 14**, 83 (2015) doi:10.1038/nrd4535

Protein and small molecule drugs have different targets



Protein and small molecule drugs have different targets

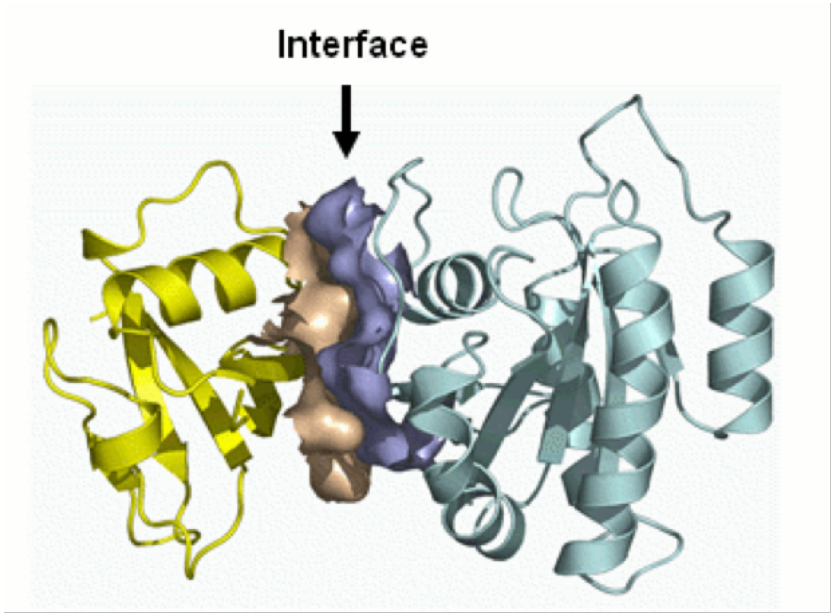




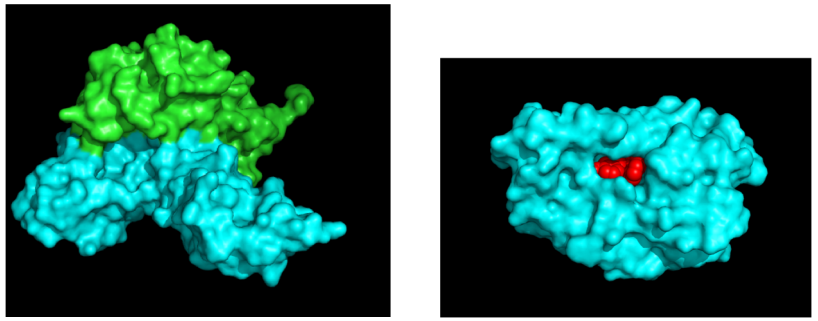
- flat interface
- many interactions

- binding pocket (“druggable”)
- fewer interactions

Proteins can be very potent in blocking protein-protein interactions

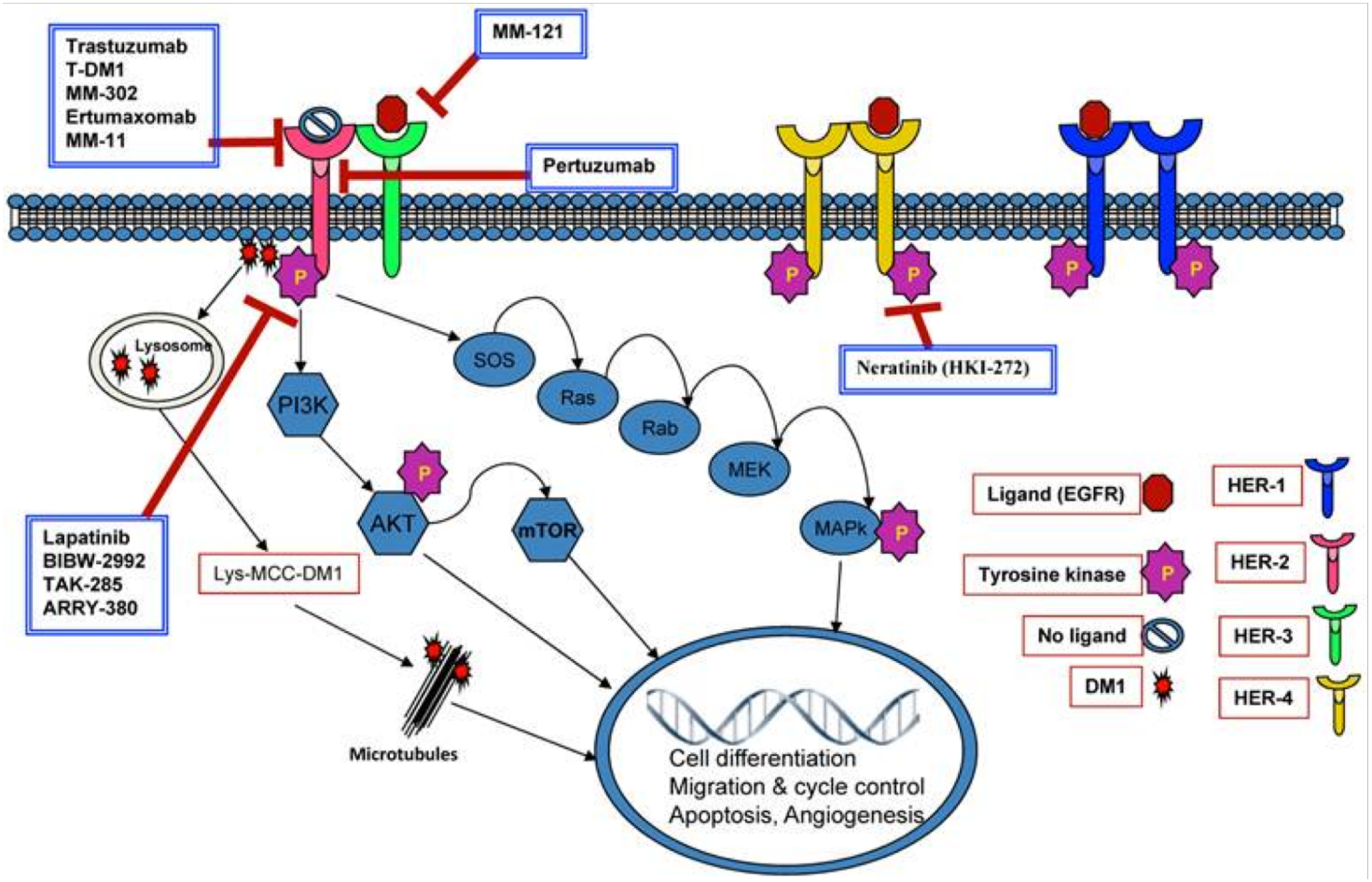


- a small molecule drug may not have enough surface/binding energy to efficiently block a protein-protein interaction *



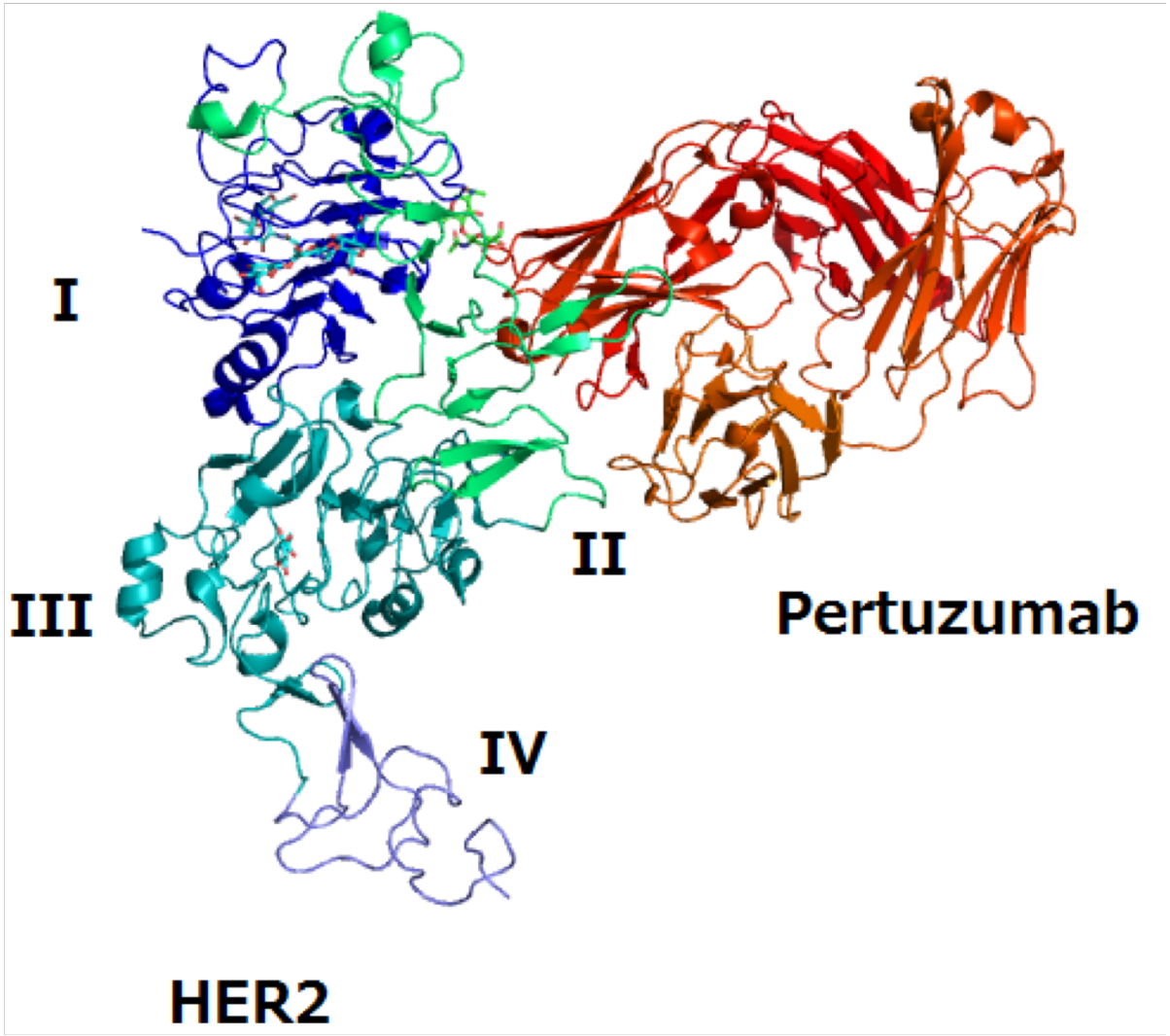
** large protein-protein interfaces are difficult to disrupt with a small molecule, although sometimes possible: see Jim Wells & Chris McClendon, Nature 2007*

Protein-protein interactions important in cancer are key targets for inhibition

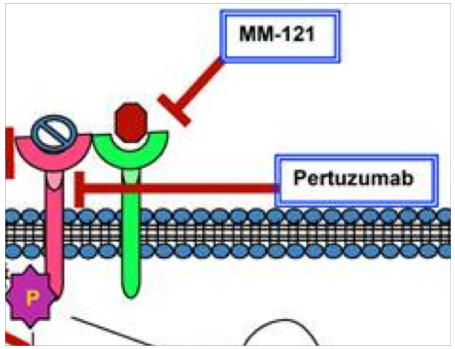


Journal of Cancer 04: 0117

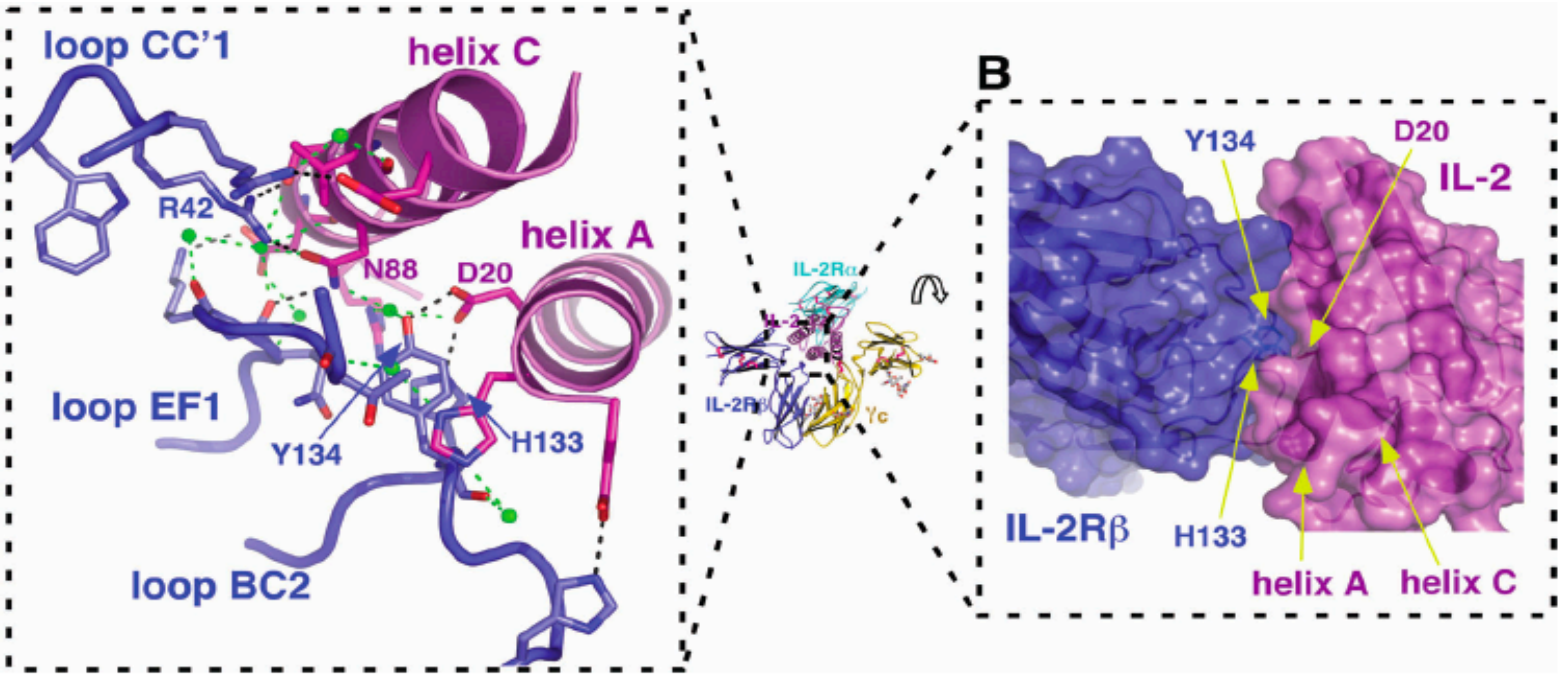
Proteins drugs can be very potent in blocking protein-protein interactions



- Pertuzumab inhibits dimerization of Her2 with other Her receptors



Proteins can be very specific, because they can form many defined interactions with their targets



Example IL-2/ IL-2 receptor, Wang et al., Science 310: 1159-1163, 2005

A protein drug may be able to form similar specific interactions to distinguish a target, such as the IL-2 receptor, from a paralog that has slightly different amino acids in the interface.

Agenda

Background

- Why protein drugs:
  - potent as protein-protein interaction inhibitors
  - can perform "endogenous" activity: insulin, human growth hormone, IL-2
  - often very specific

Agenda

Background

- Why protein drugs

State of the field

- Principles of engineering & optimizing antibody drugs

Protein drug FDA approval 2011-2016

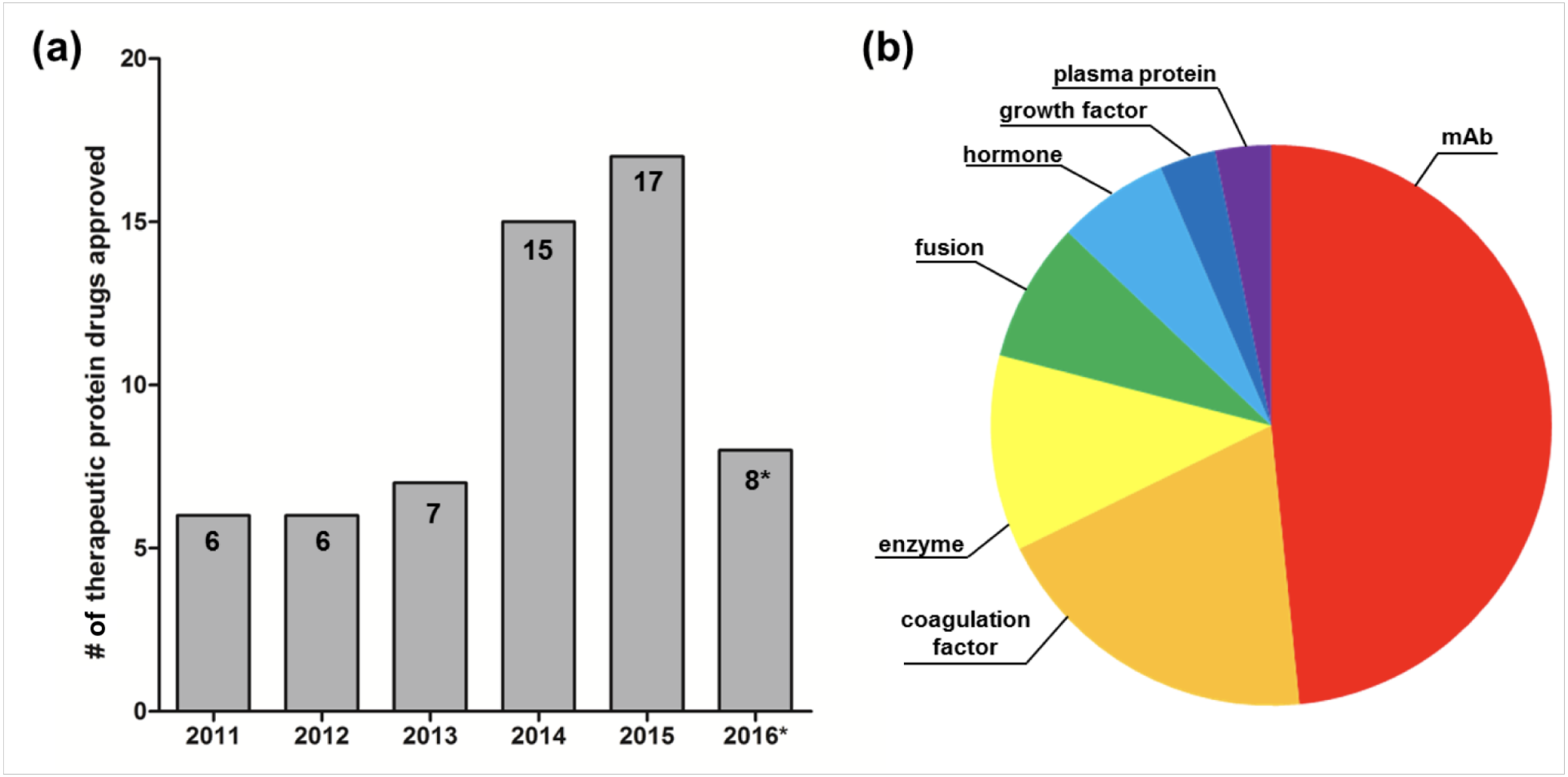
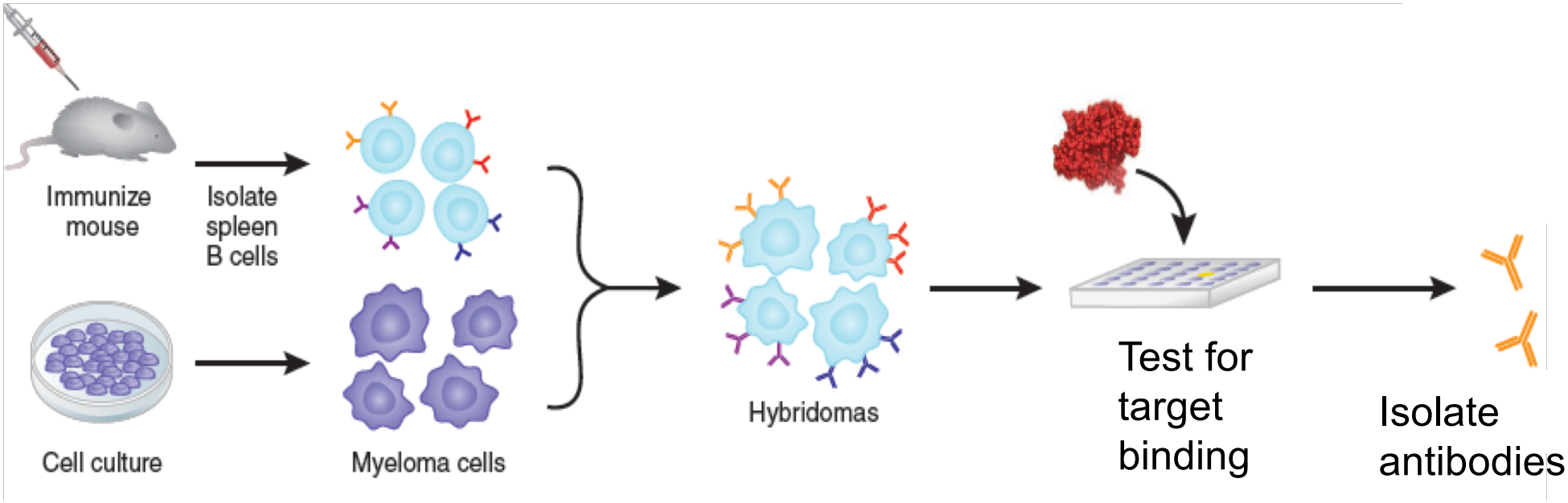


Figure 2. U.S. Food and Drug Administration (FDA)-approved therapeutic proteins (2011–2016*). (**a**) Bar graph showing the number of therapeutic protein FDA approvals by year (2011–2016*). (**b**) Pie chart showing the distribution of FDA-approved therapeutic proteins (2011–2016*) by drug class. *January 1, 2011, through August 31, 2016.

Lagassé HAD, Alexaki A, Simhadri VL *et al.* **Recent advances in (therapeutic protein) drug development [version 1; referees: 2 approved]** *F1000Research* 2017, **6**(F1000 Faculty Rev):113

Monoclonal antibodies can be generated *via* immunizing mice



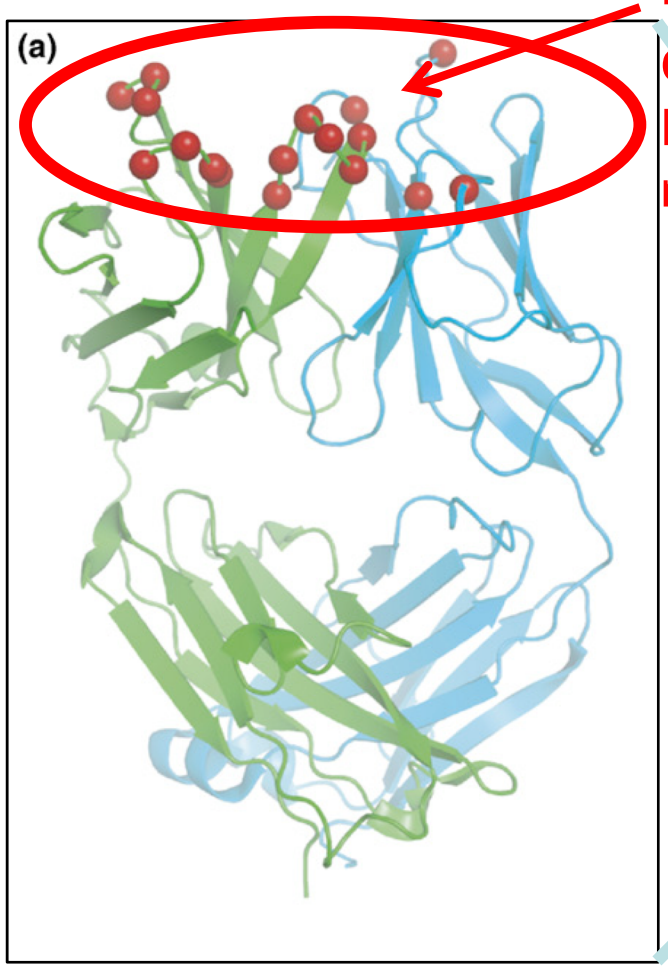
But the problem with this approach: antibodies are from mouse and can lead to immune reactions in humans!

(but even completely human antibodies can cause immune reactions)

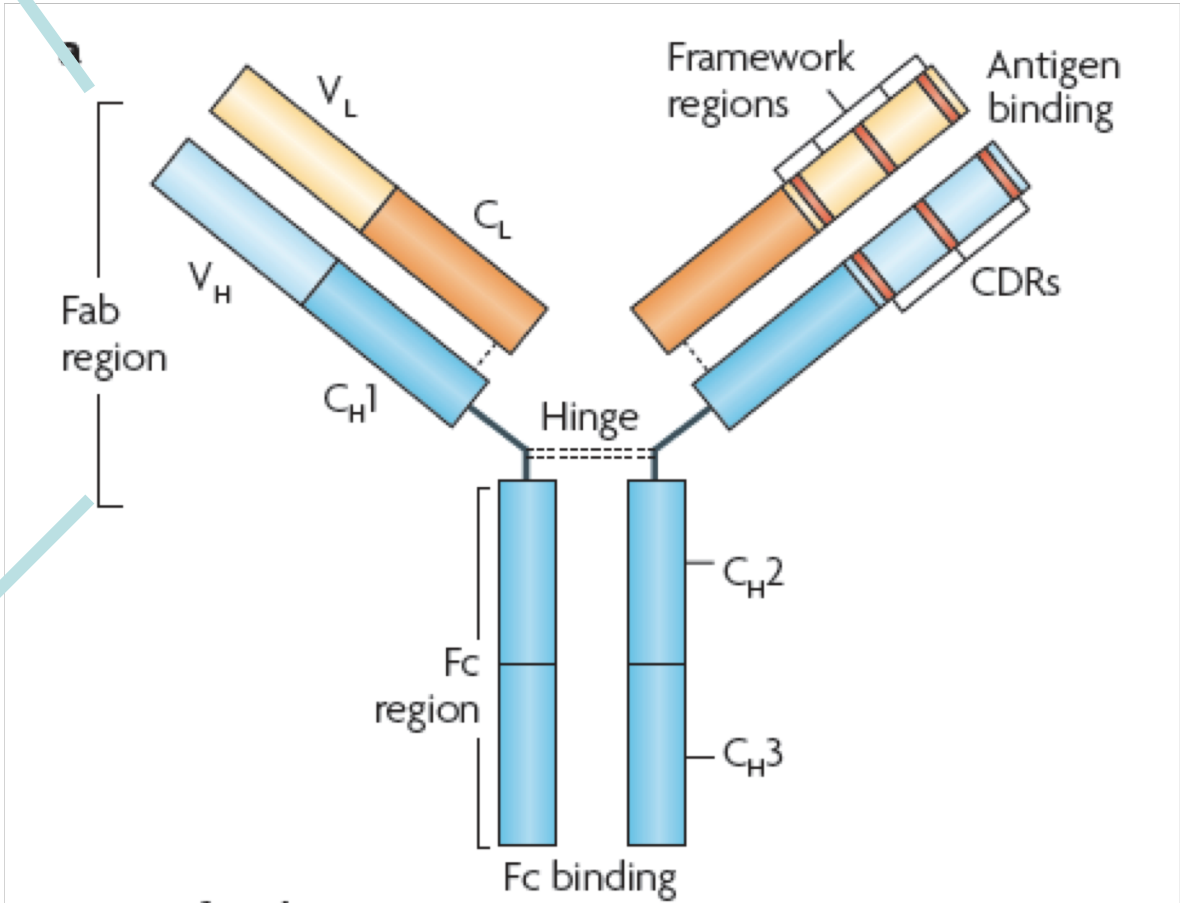
Antibody Structure

CDRs (complementarity-determining regions) that bind the antigen are formed by a small region: loops in the heavy (H1-H3) and light chains (L1-L3)

Red spheres indicate amino acid residues mainly responsible for target recognition

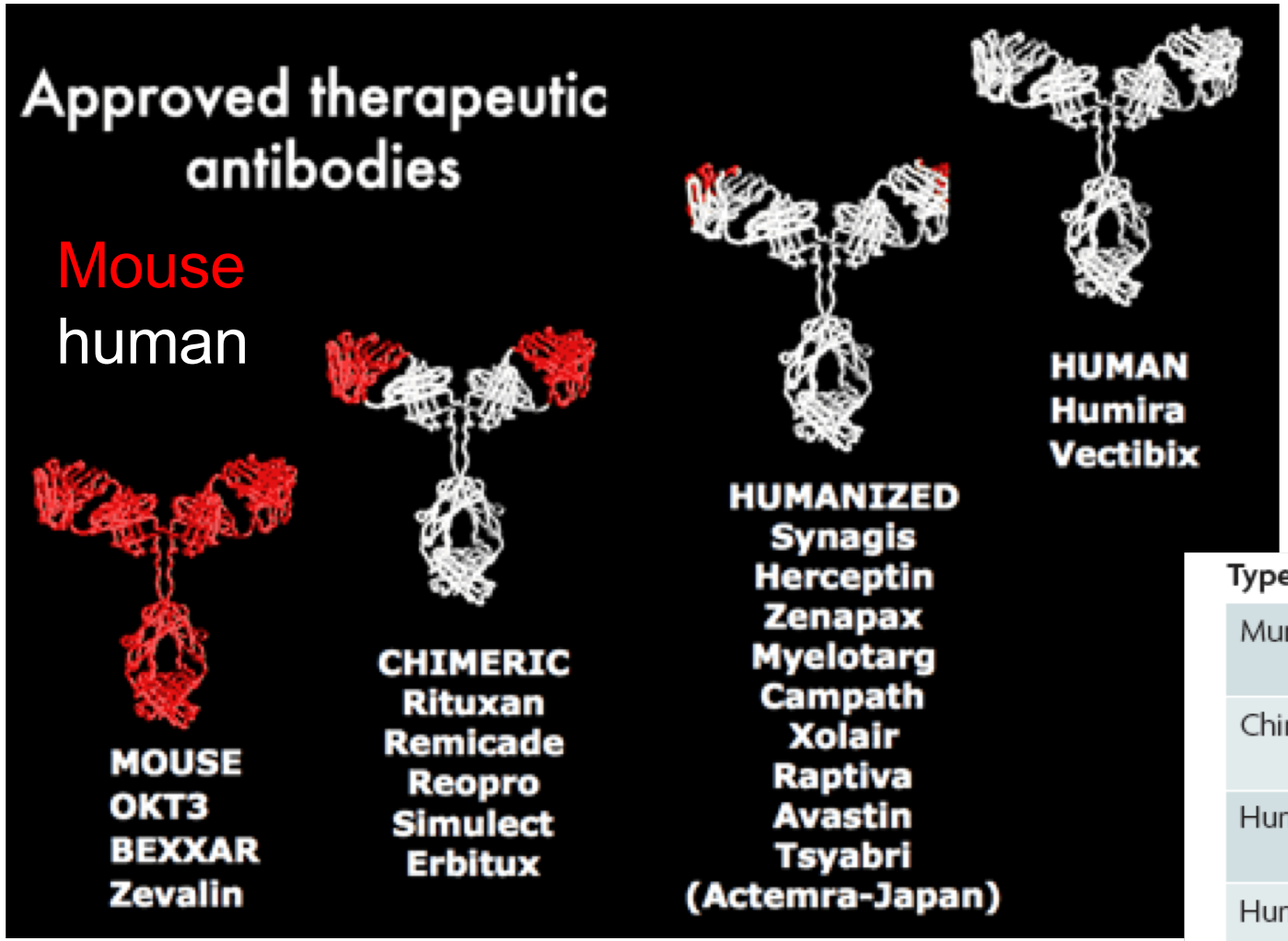


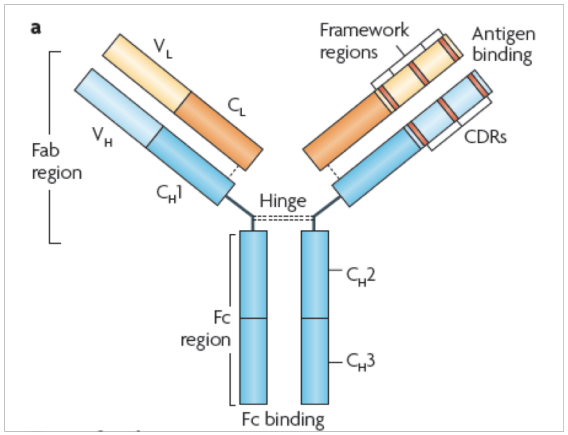
Antigen-binding Fragment (Fab)



Different classes of therapeutic antibodies have a different fraction of mouse amino acid sequence

These are most desired



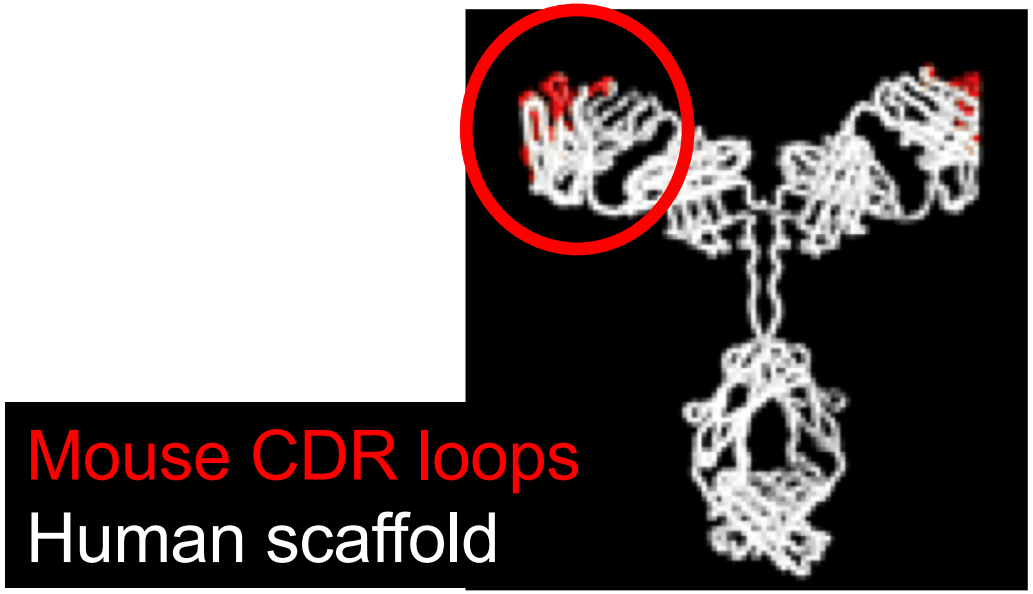


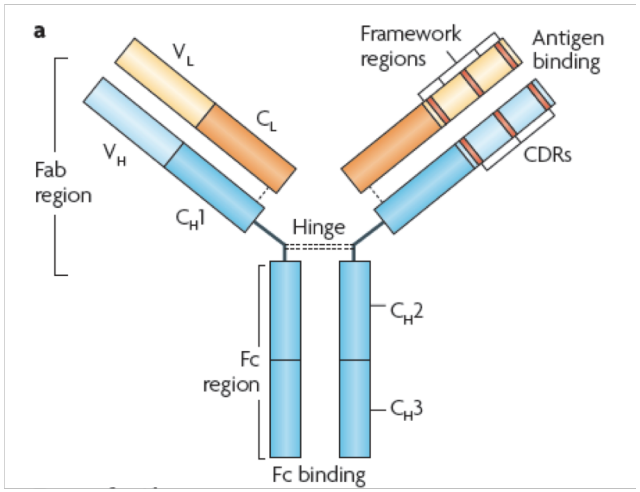
Types of mAbs

Murine	Entirely murine amino acids
Chimeric	Human constant (C) + murine variable (V) regions
Humanized	Murine complementarity determining regions (CDRs)
Human	Entirely human amino acids

“Humanized” antibodies are a common solution to decease the immunogenicity of mouse antibodies

- “transplant” the CDR loops from the mouse antibody onto a human constant “scaffold”
- How? Insert into the gene sequence



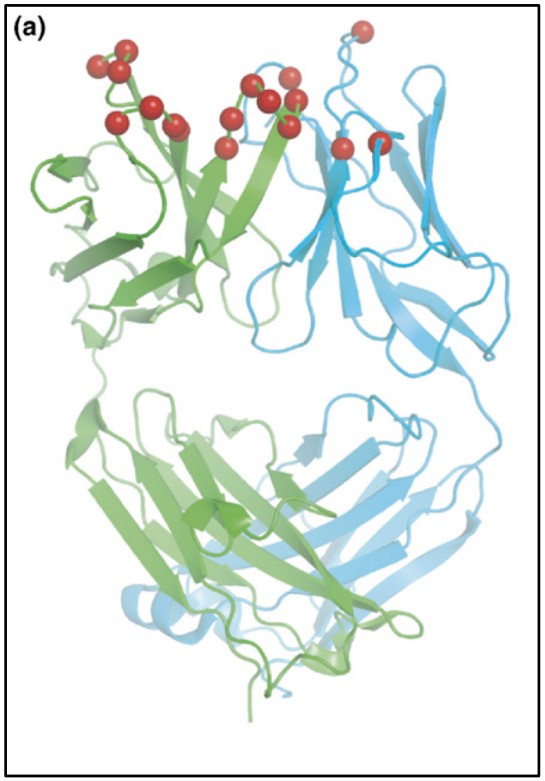


The problem with this approach: “transplanting” generally decreases potency (small sequence differences in the scaffold can change the precise structure / positioning of the CDR loops)

Can we use a human antibody scaffold in the first place?

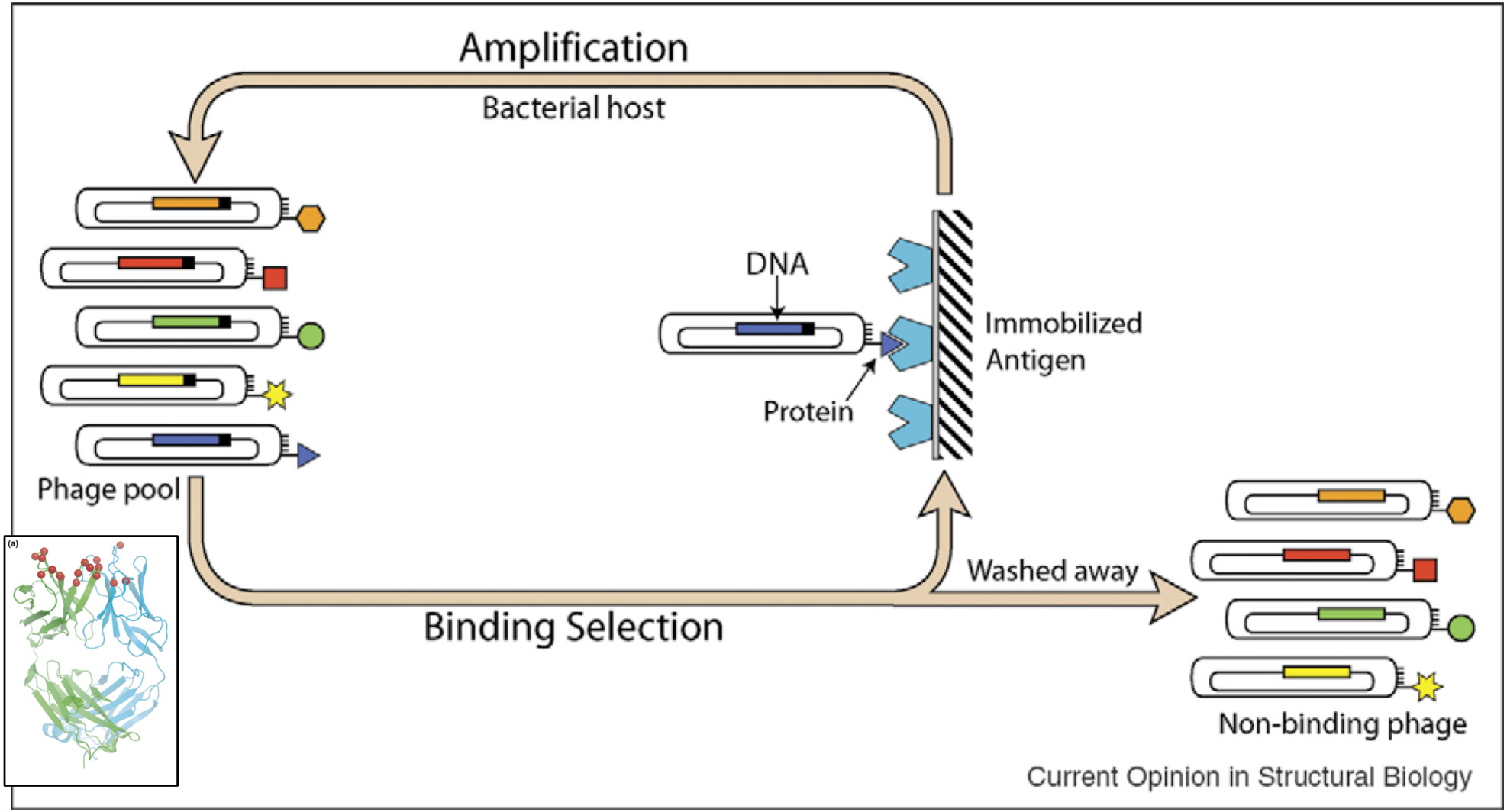
- Can be done using a technique called "phage display":

  select from "libraries" of antibody variants where the scaffold is human and the red residues are varied



- Phage display can also be used to improve potency: antibody "affinity maturation" by selection in the laboratory

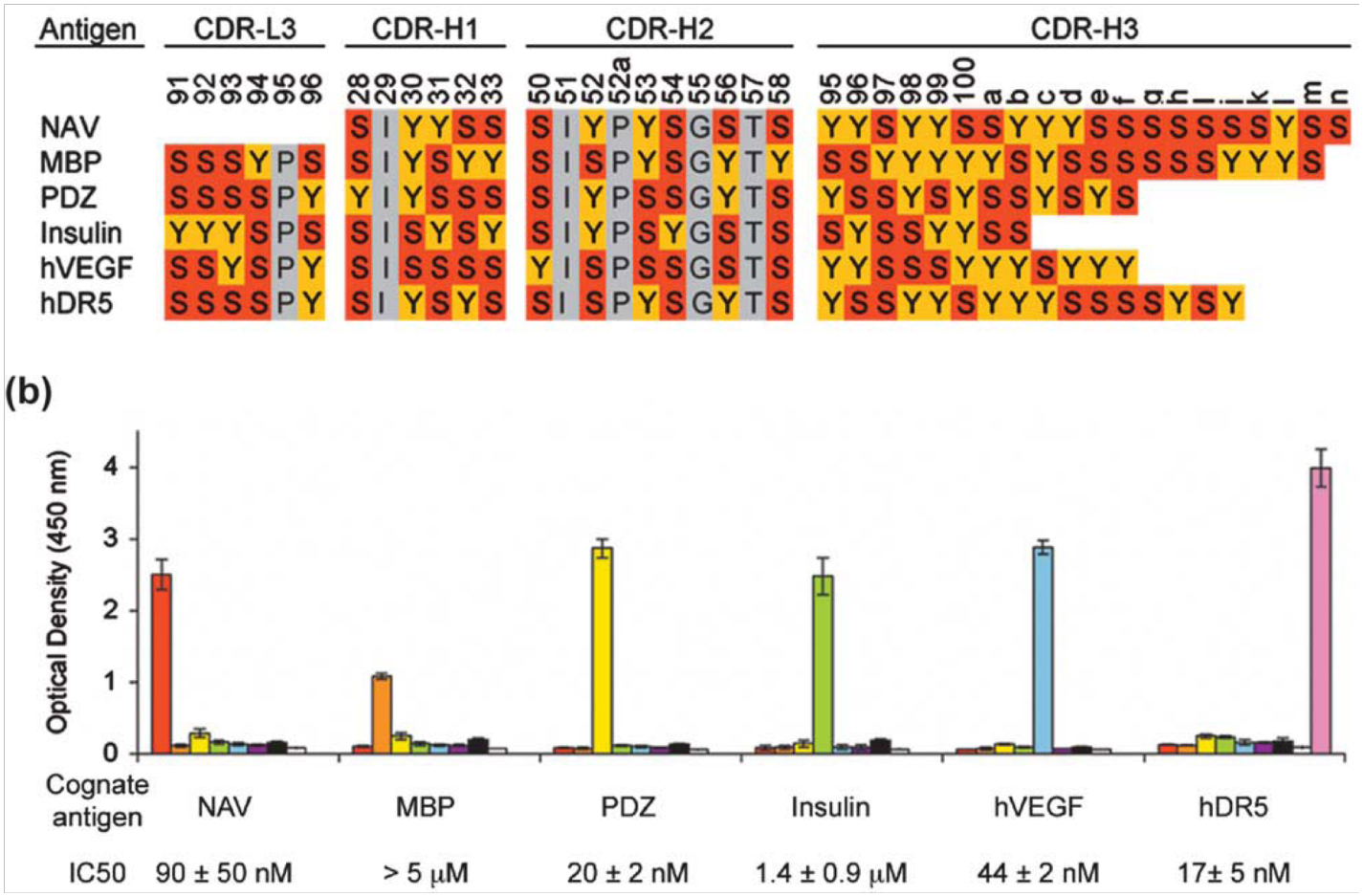
“Phage display” can efficiently generate specific antibodies (and also optimize them)

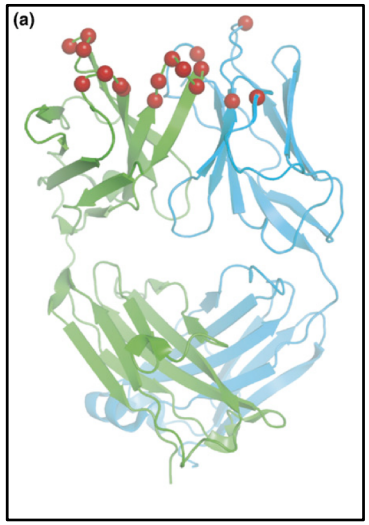


Sidhu & Koide, 2007

By phage display, antibodies can be selected to be specific for diverse antigens

The same “library” of antibody variants (that had the sequences of their CDR loops randomized) was screened against 6 targets to yield 6 different specific antibodies





Fellouse et al., JMB 2005

Agenda

Background

- Why protein drugs

State of the field

- Principles of engineering & optimizing antibody drugs

Current research

- Creating protein drugs from scratch
- Computational protein design (Rosetta)
- Successes and challenges

Can we design (non-antibody) protein therapeutics *de novo*?

- let's first think about why we would want to…

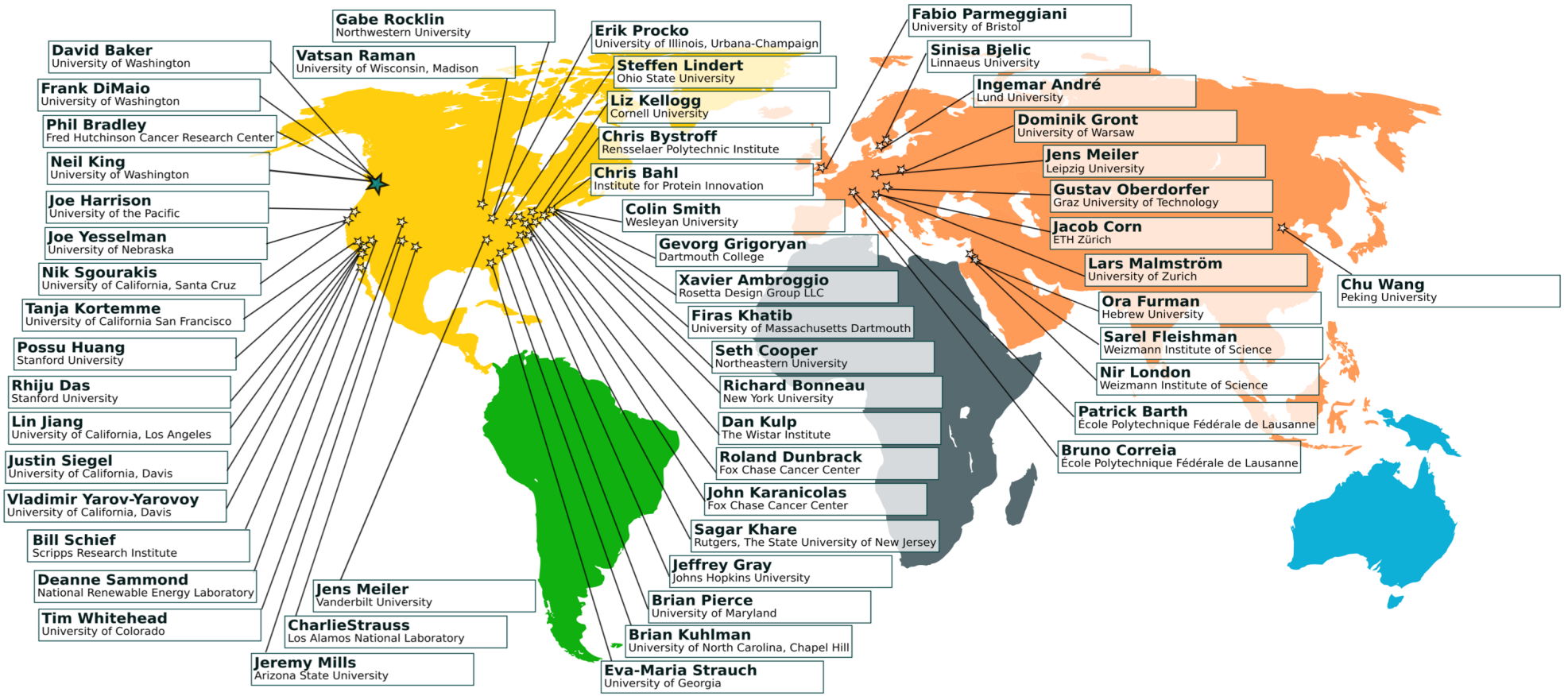
Can we design (non-antibody) protein therapeutics *de novo*?

- Two case-studies:
  - Vaccines
  - Potent and selective mimics of endogenous proteins

Before we get to applications: How does *de novo* protein design work? -> in Rosetta

A large community of research labs develop Rosetta to model & design proteins

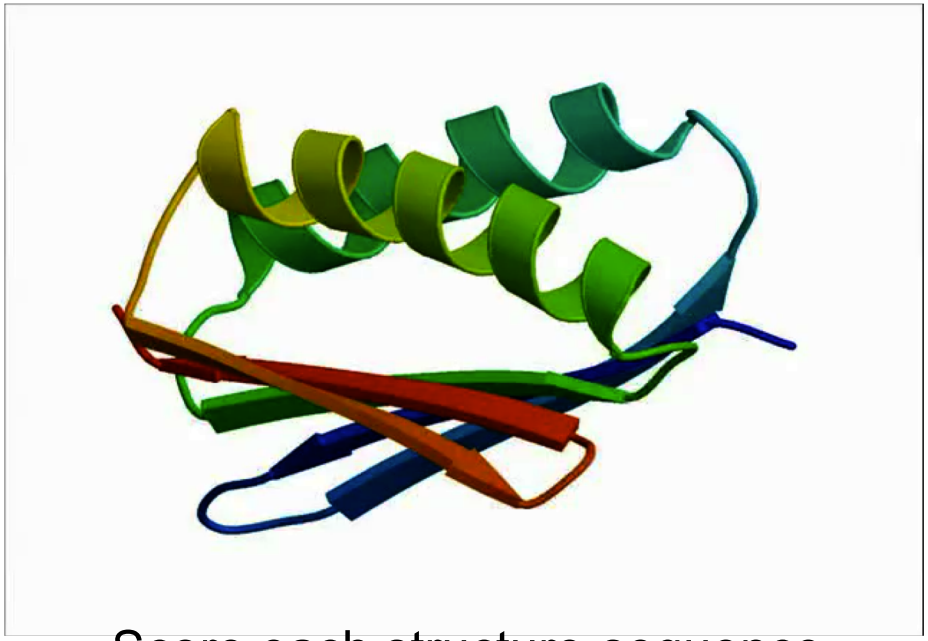
Source code free for academia: rosettacommons.org
Licensed for a fee by >70 companies



Computational protein design (Rosetta) is an optimization problem

INPUT

Design Objective: Structure & Function (represented at all-atom level)



Score each structure-sequence combination, find “best”

Monte-Carlo simulated annealing
Genetic Algorithms
SAT solvers
Dead-end Elimination
…

OUTPUT

Amino acid Sequences Optimized for Design Objective & Structure

The key challenge is that the possible space is absolutely enormous

---

possible sequences for 100 residue protein: $20^{100} \sim 10^{130}$
(most proteins are larger; only a small fraction will be functional)

number of atoms in the universe: $\sim 10^{80}$

The key challenge is that the possible space is absolutely enormous

---

possible sequences for 100 residue protein: $20^{100} \sim 10^{130}$
(most proteins are larger; only a small fraction will be functional)

number of atoms in the universe: ~ 10^{80}

number of different proteins on earth today: ~ 10^{12}

number of proteins sampled in evolution: 10^{21} - 10^{50}

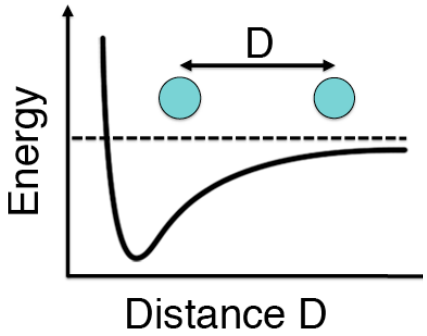
As a consequence, need to make simplifications

- How we ”sample” space (reduced degrees of freedom)

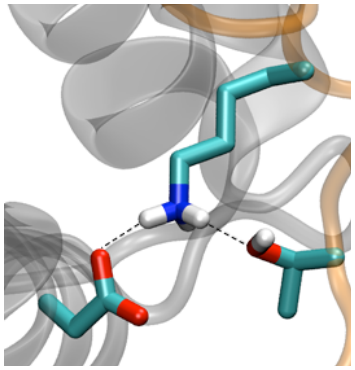
- How we ”score” solutions (approximate energy function)

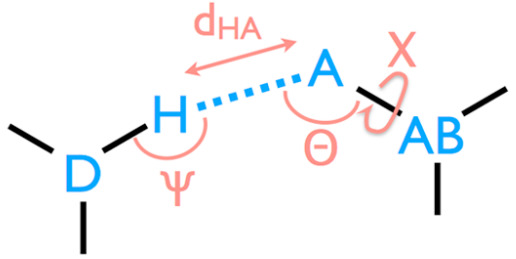
Rosetta all-atom energy function

1. van der Waals packing



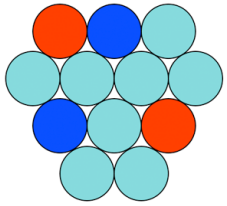
2. hydrogen bonding

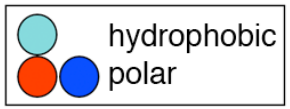




Kortemme & Baker, PNAS 2002
Kortemme et al., JMB 2003

3. implicit solvation



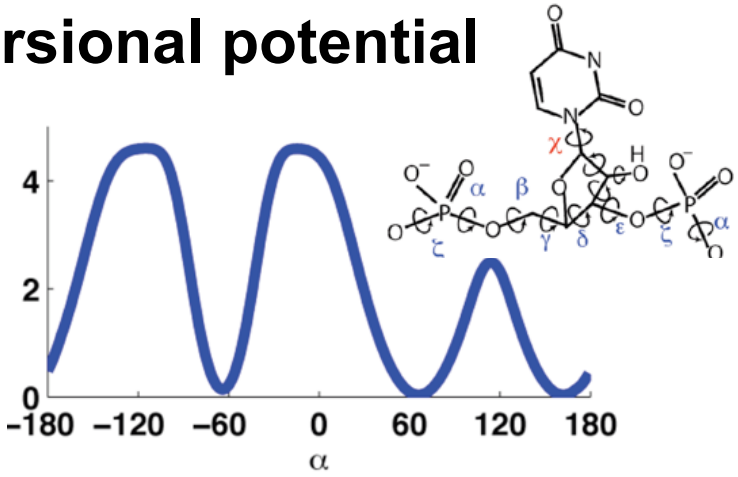


$$\Delta G_i^{slv} = \Delta G_i^{ref} - \sum_{j \neq i} f_i(r_{ij}) V_j$$

ΔG^{ref} : reference solvation free energy
V_j : Volume of group j
f_i : solvation free energy density of i
r_{ij} : distance between i and j

Lazaridis & Karplus, Proteins 1999

4. torsional potential



5. electrostatic repulsion (screened)

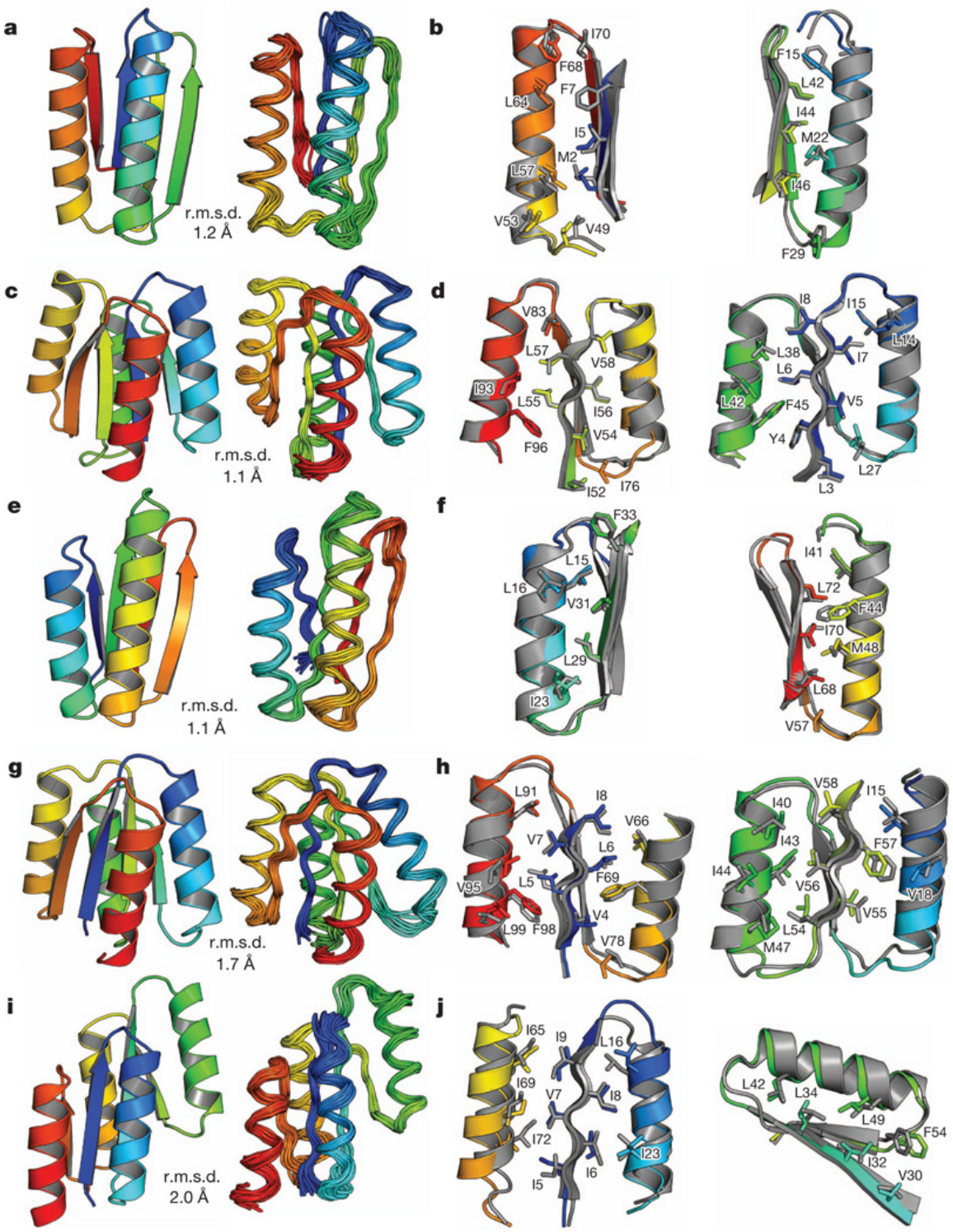
free energy - configurational entropy

As a consequence, need to make simplifications

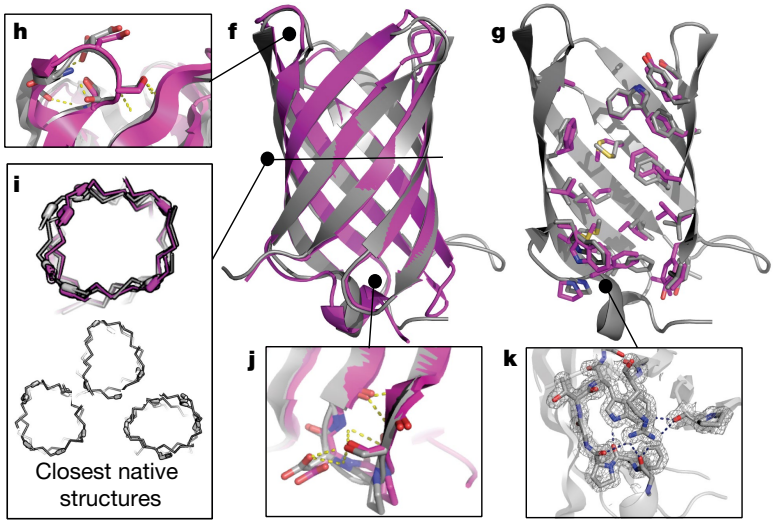
- How we ”sample” space (reduced degrees of freedom)

- How we ”score” solutions (approximate energy function)

Both of these are main causes of errors and determine current state of the field.

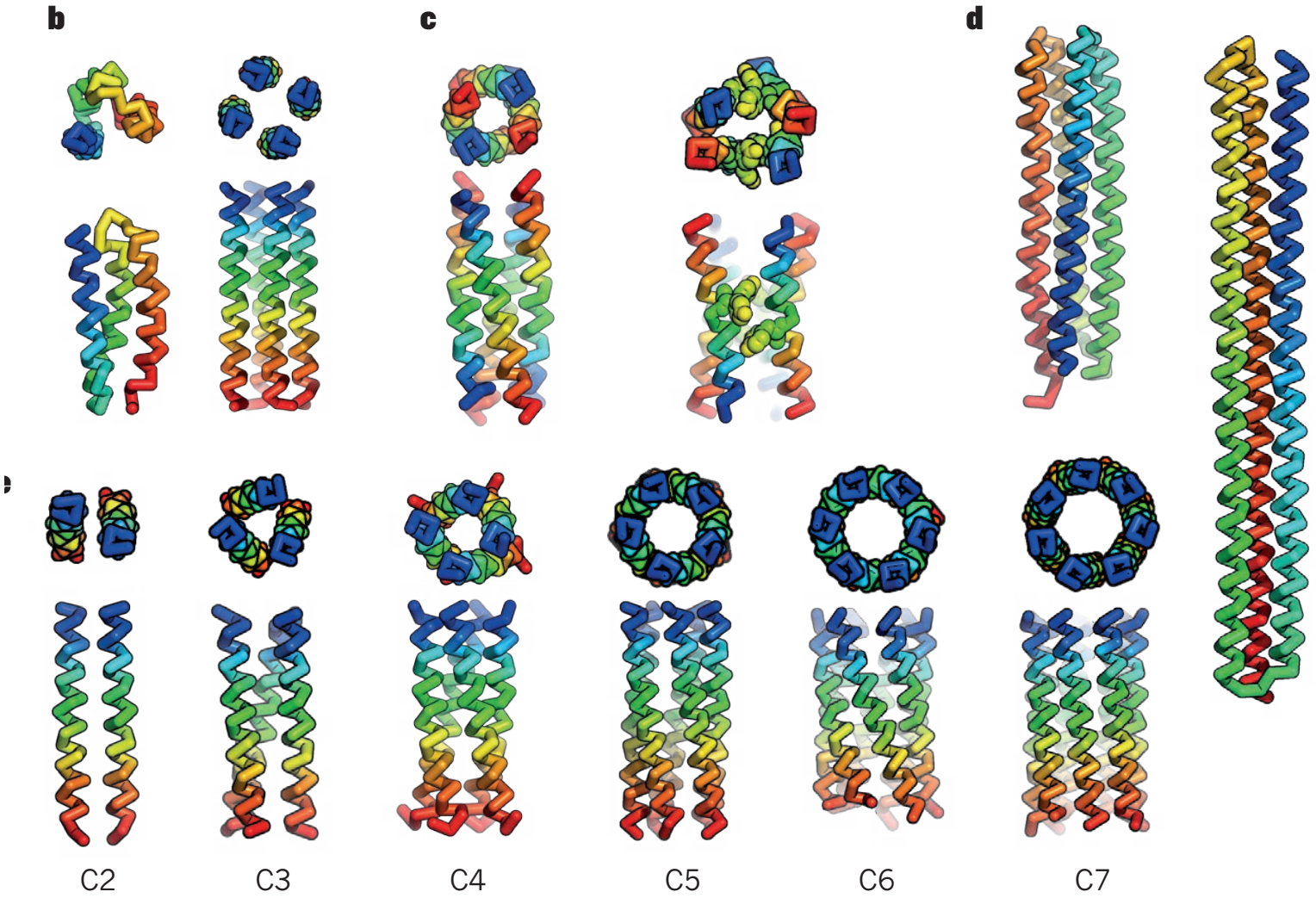


Despite these simplifications, many design successes: small, ”idealized” folds



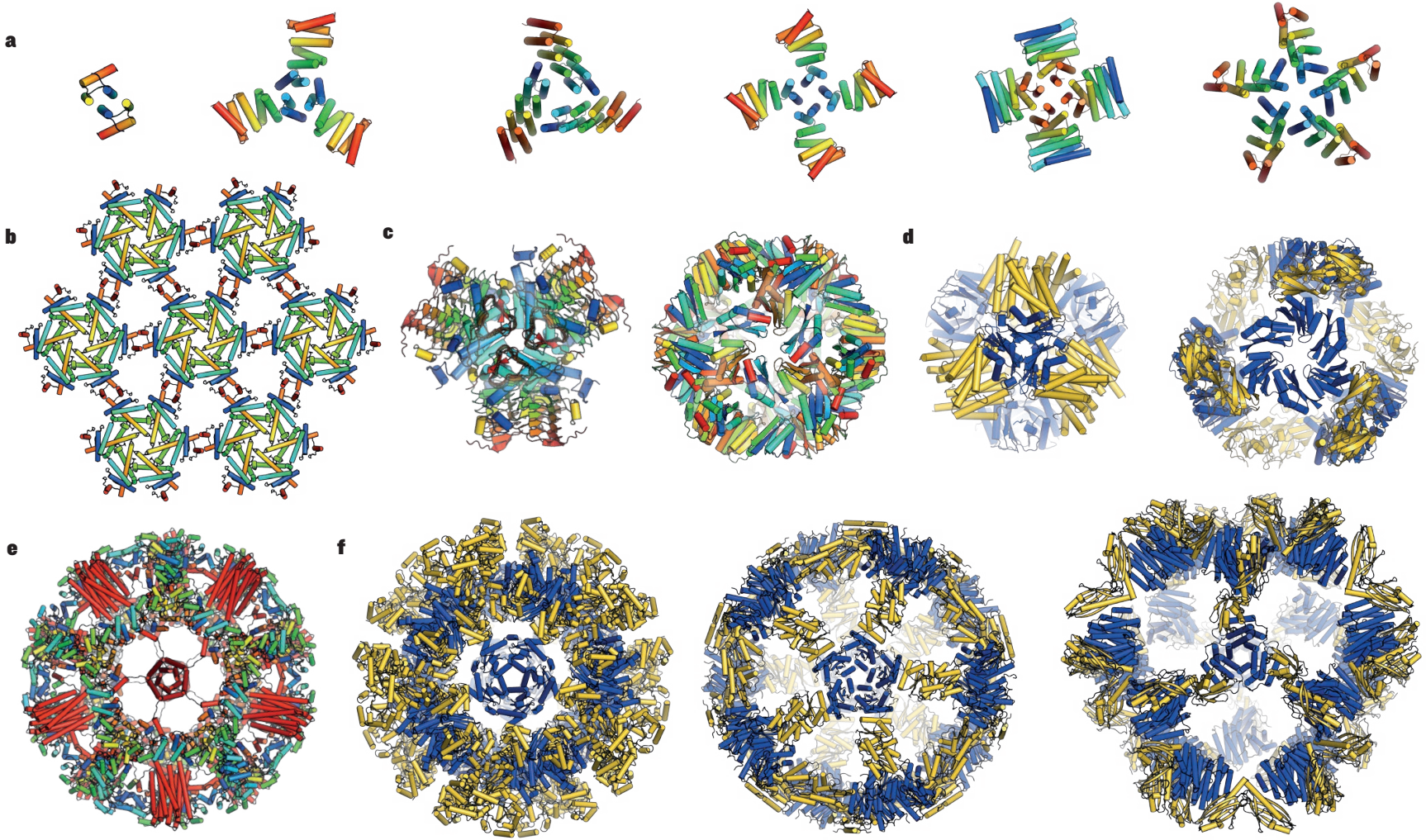
Koga et al, Nature 2012 DOI: 10.1038/nature11600
Dou, Vorobieva et al, Nature 2018 DOI: 10.1038/s41586-018-0509-0

In particular, a wealth of architectures from helical bundles (that can be “functionalized”, more later)



Huang, Boyken & Baker, Nature 2016 doi:10.1038/nature19946

Helical structures can be assembled into a range of higher-order architectures



Huang, Boyken & Baker, Nature 2016 doi:10.1038/nature19946

Computational protein design: state of the field

(>3 decades of fundamental work, enormous progress in applications in the last 15 years!)

- *de novo* folds built from rules: a, a/b and all-beta proteins
- new architectures, symmetrical assemblies & materials
- helical bundles, can be functionalized

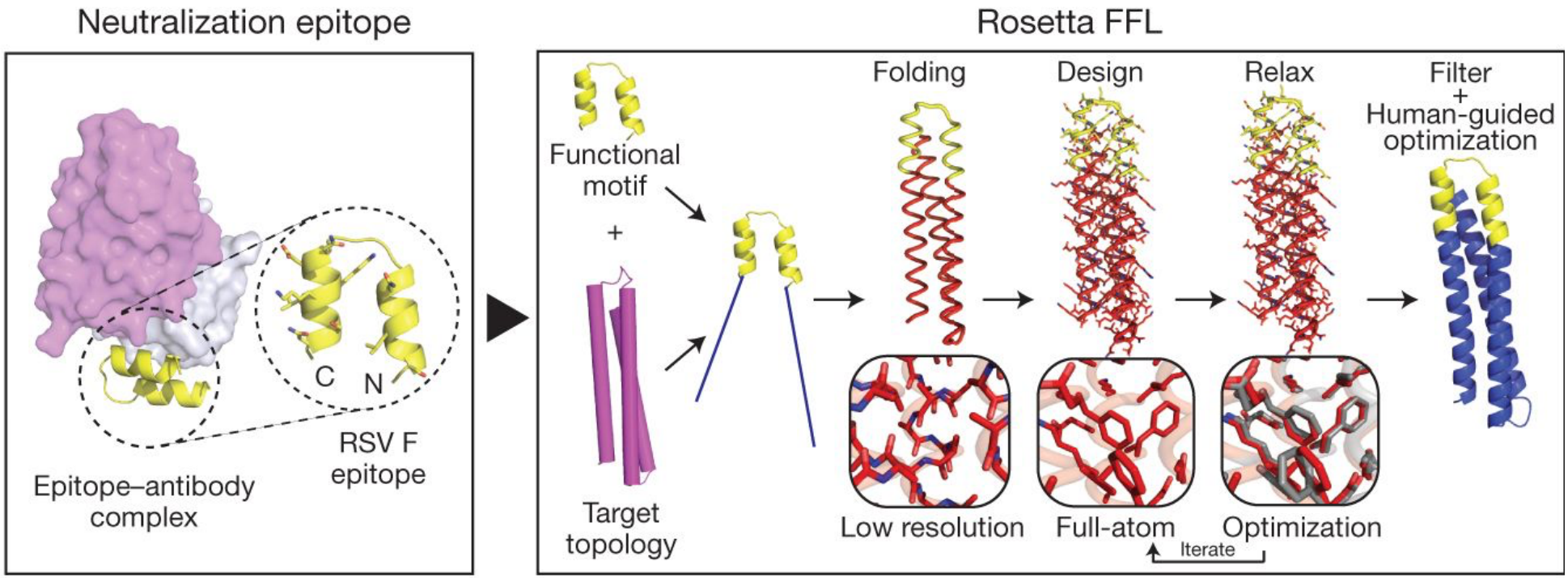
Current challenges are in designing function

- sensors
- switches
- efficient enzymes
- machines
- … many complex and composite functions

key difficulties:

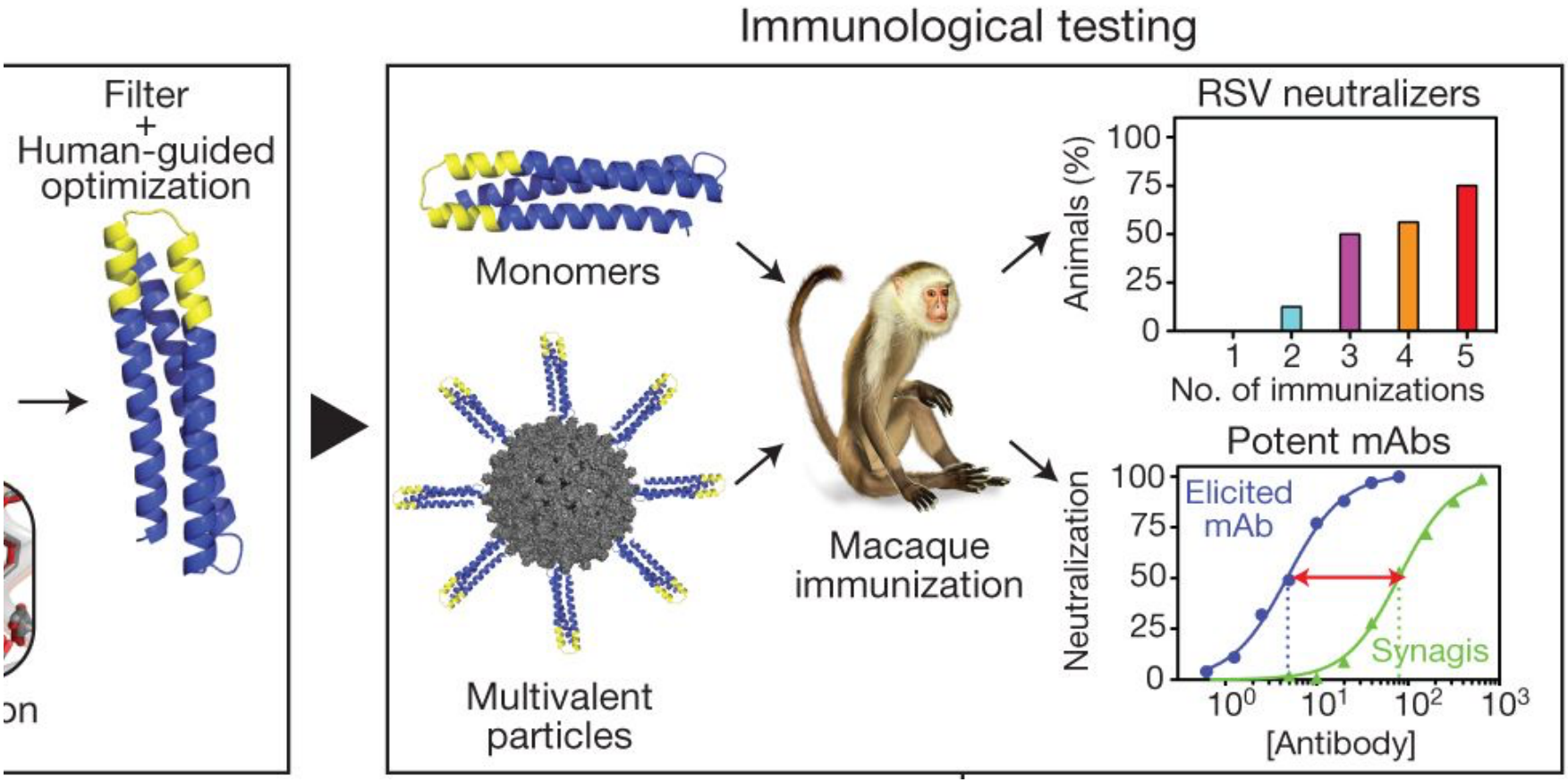
- precise control over irregular functional geometries
- often polar recognition
- switchable states (not deep minima)

An approach to circumvent this problem: “transplant” functional region to new stable protein



Correia et al, Nature 2014
DOI: 10.1038/nature12966

Proof of principle for epitope-focused vaccine design



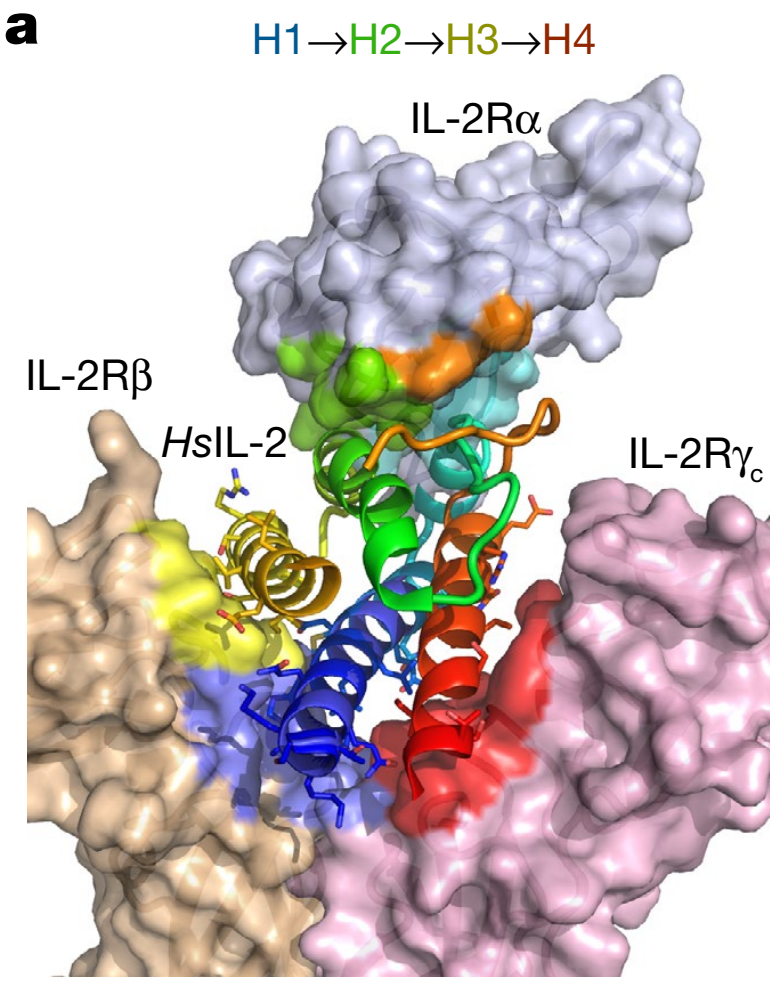
Correia et al, Nature 2014
DOI: 10.1038/nature12966

De novo design of potent and selective and protein mimics as therapeutics (case study IL-2)

Challenges with using IL-2 as drug:

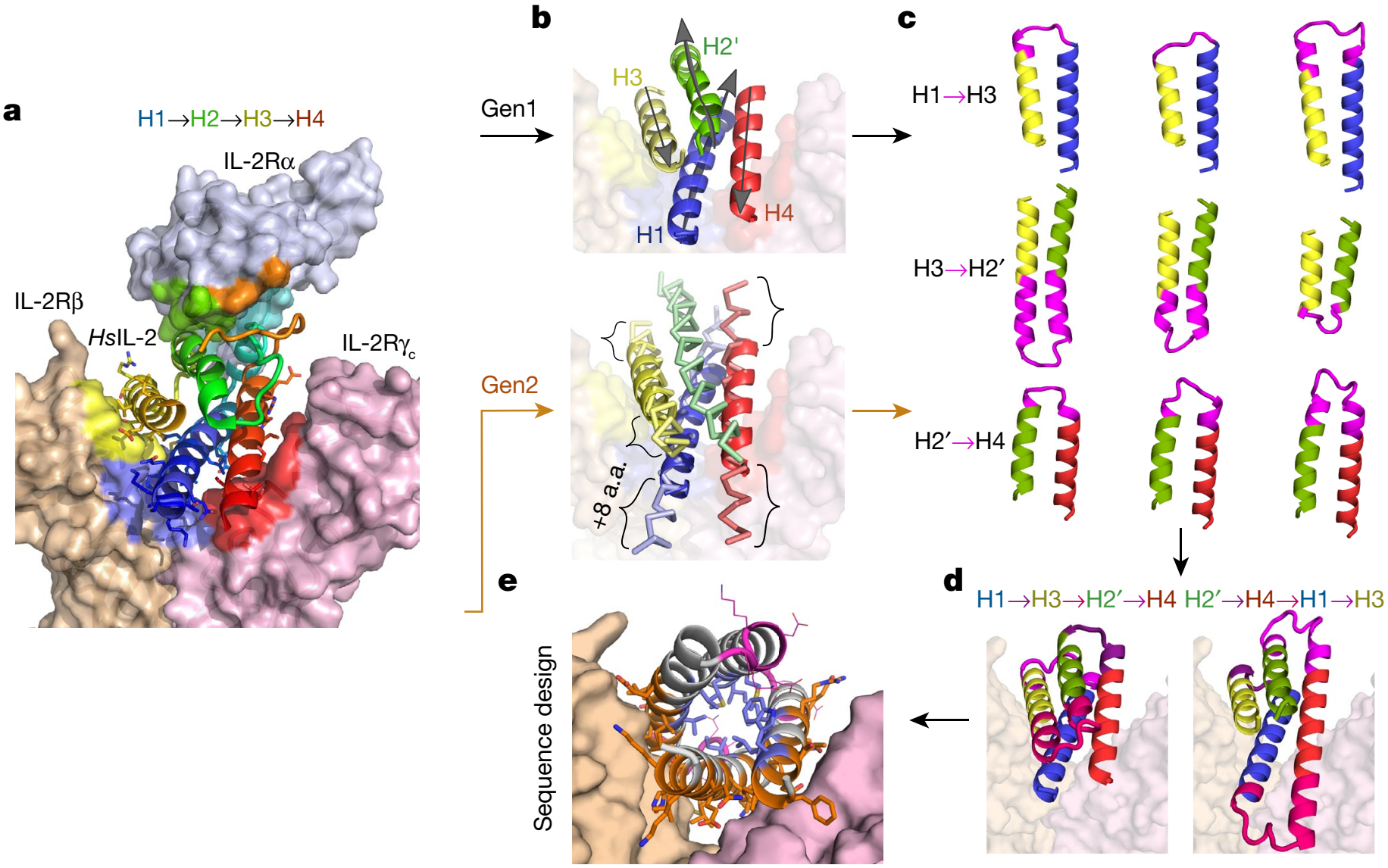
- marginal stability, aggregation
- toxicity (perhaps by interaction with IL-2Ralpha)
- immune response will also target endogenous IL-2

- previous engineering efforts compromised activity and / or stability



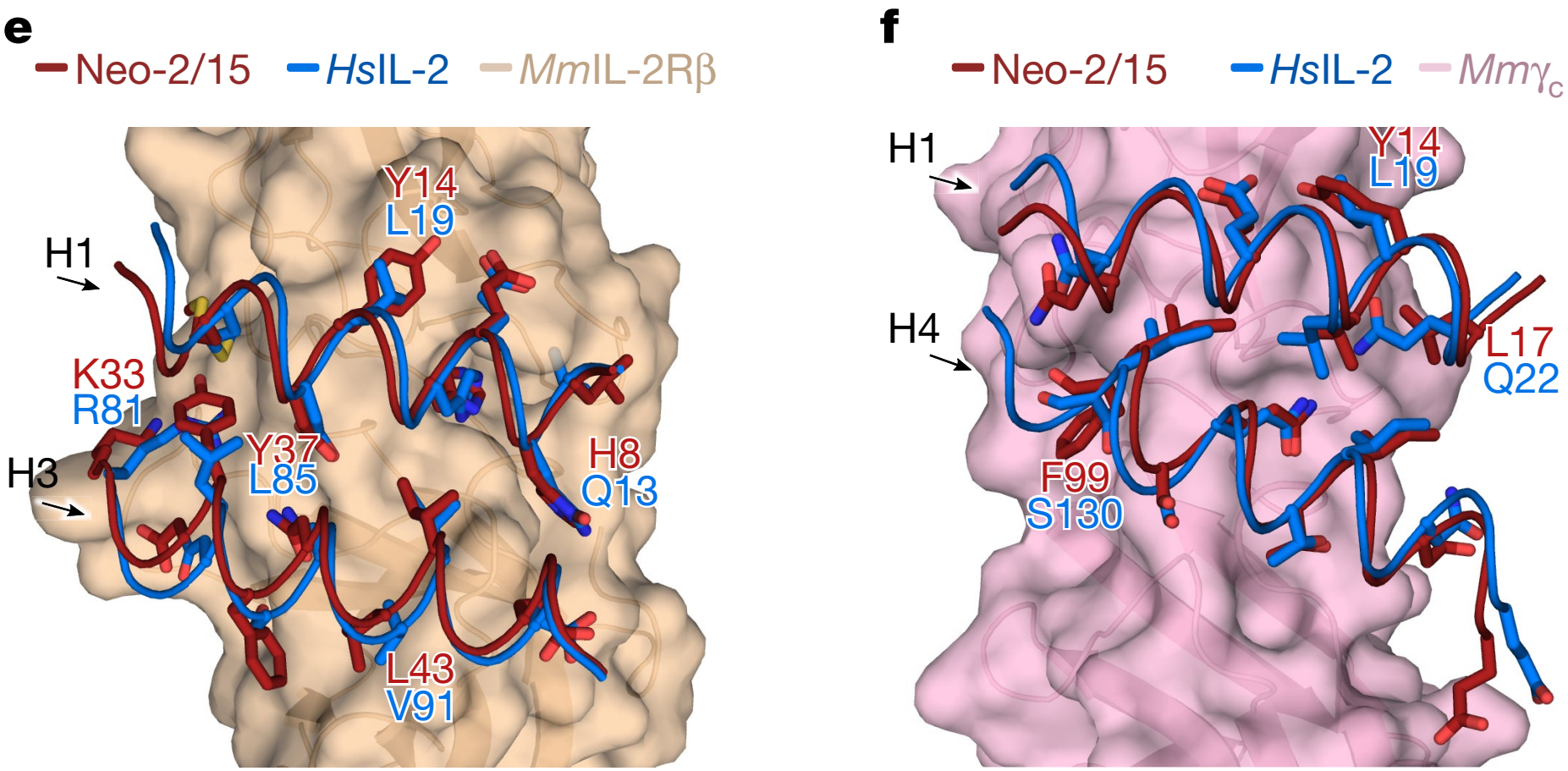
Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

A design strategy for IL-2 mimics



Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

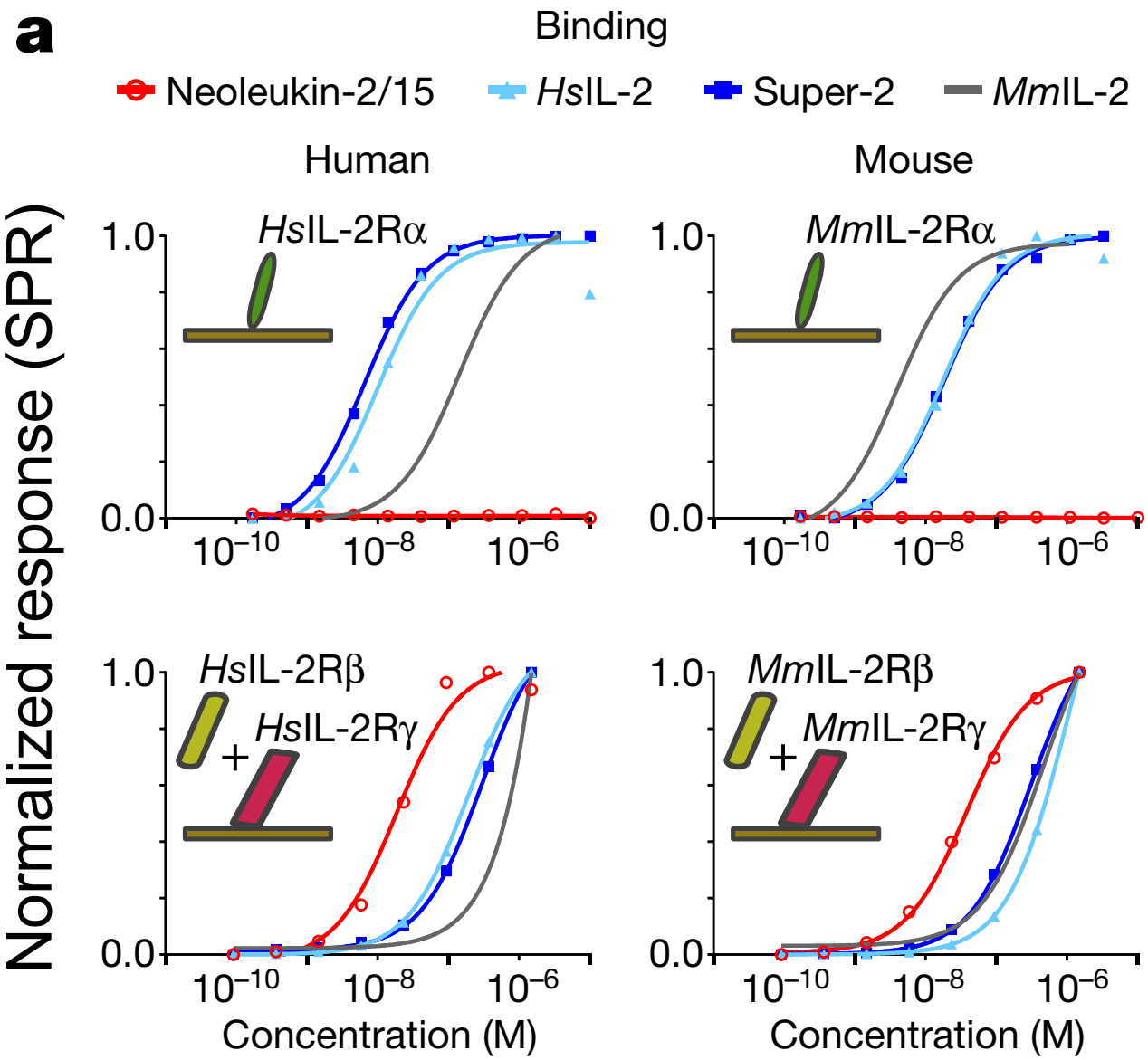
The design structure confirms the preserved (and improved) interactions with beta and gamma



Outside these regions, Neo-2/15 is quite different from IL-2:
Sequence identity to human 14%, mouse 24%

Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

Neo-2/15 has the desired binding properties (*in vitro*)

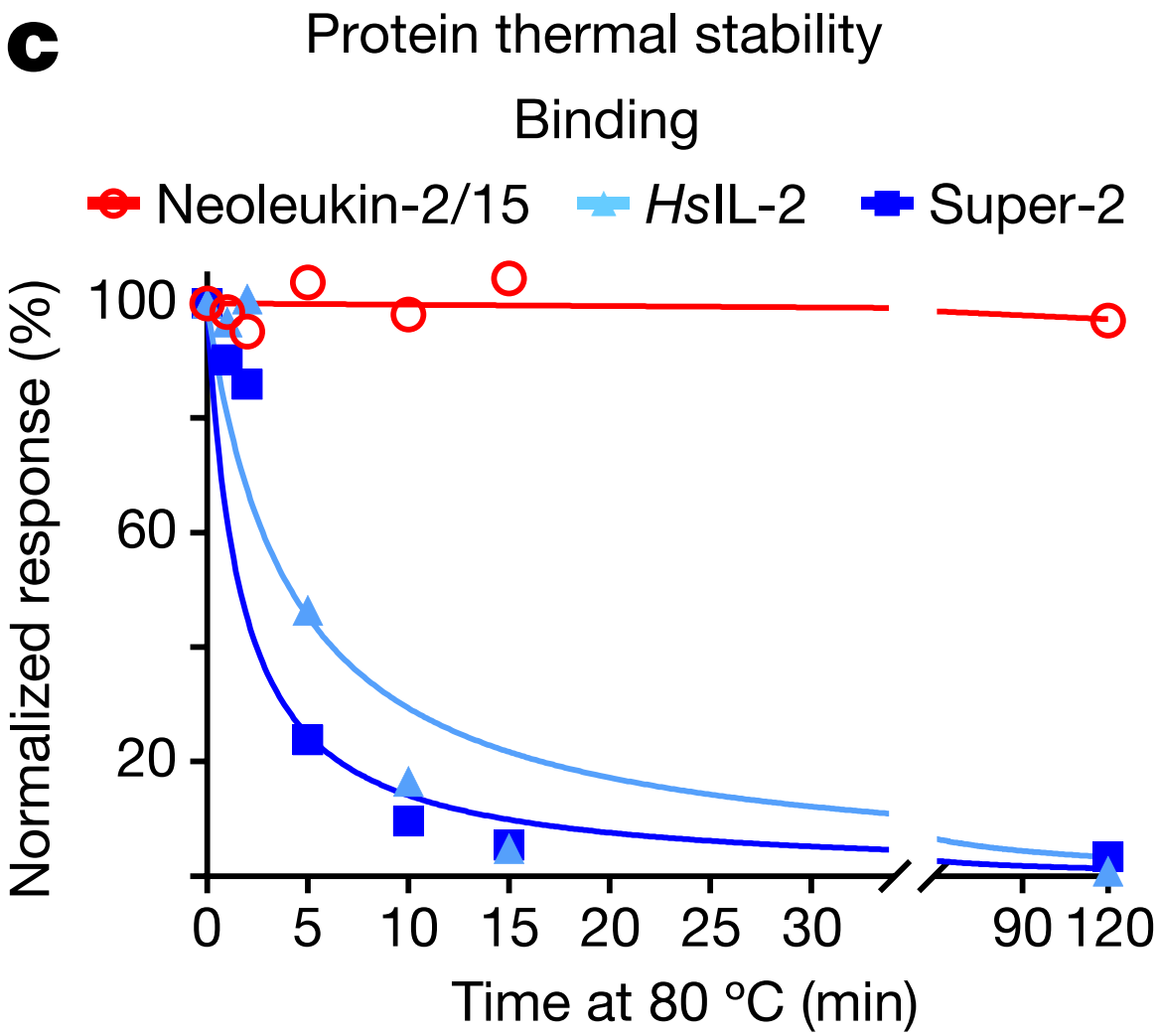


no detectable binding to alpha

strong binding to beta-gamma

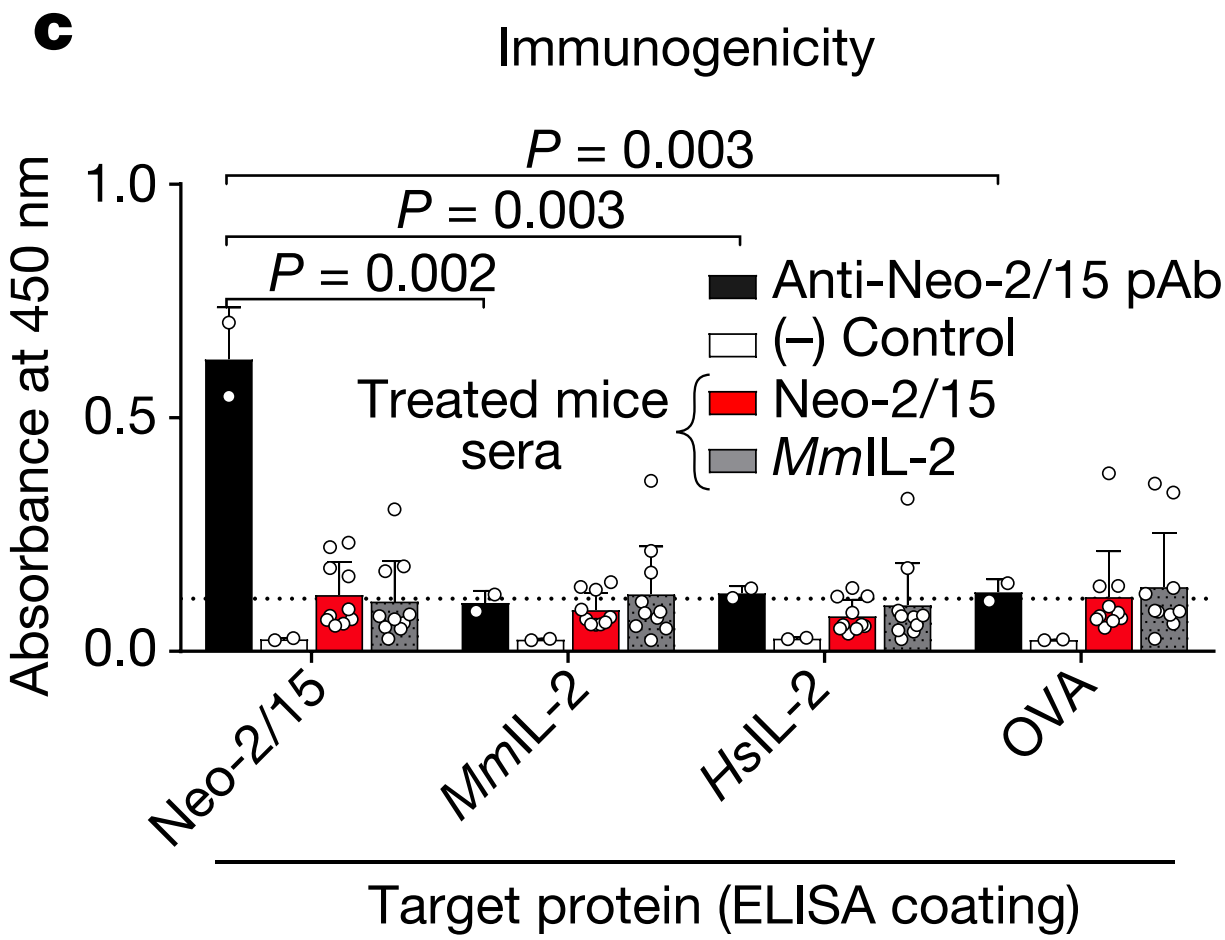
Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

Neo-2/15 has increased thermal stability



Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

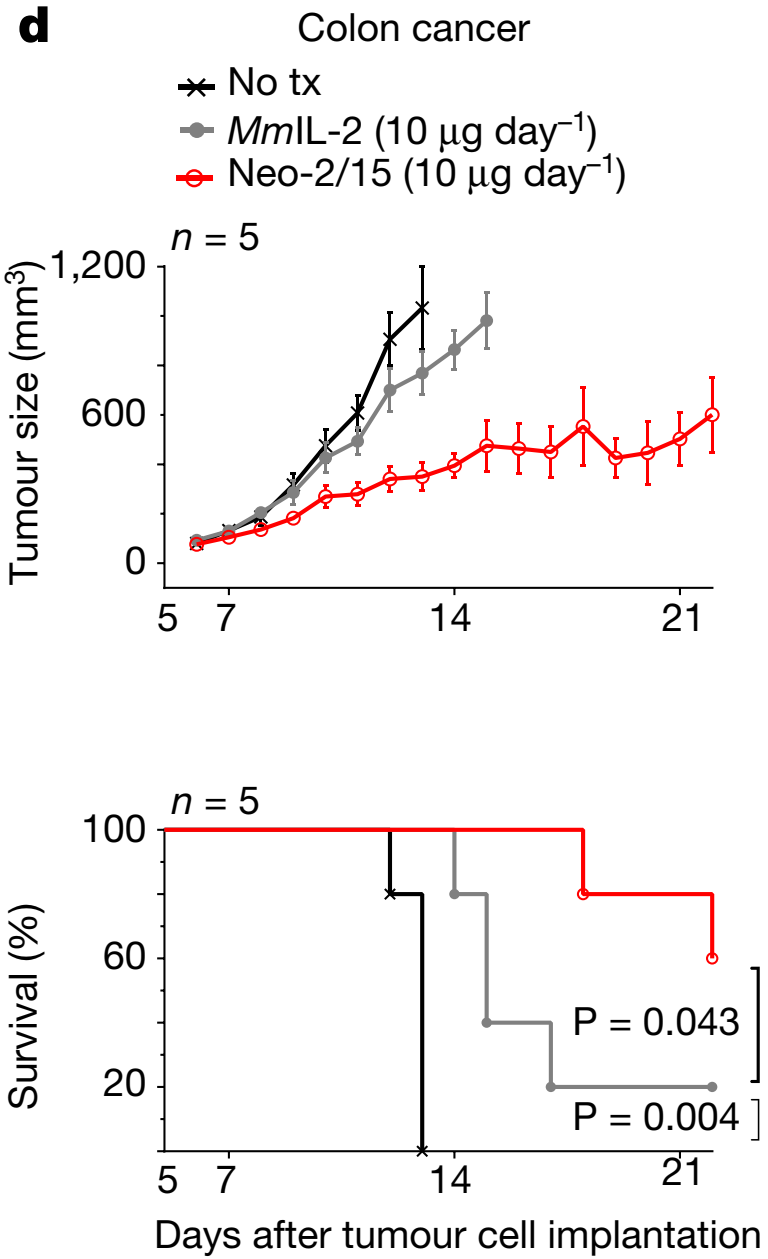
Neo-2/15 shows limited immunogenicity



Anti-Neo-2/15 polyclonal antibody does not cross-react with human or mouse IL-2

Silva et al, Nature 2019 doi.org/10.1038/s41586-018-0830-7

Neo-2/15 is more effective than mouse IL-2 in a colon cancer model



Silva et al, Nature 2019
doi.org/10.1038/s41586-018-0830-7

In summary

- enormous progress in *de novo* design of protein structures
- promise to create fine-tuned new architectures for many new functions

- design of function is more challenging
- some of these challenges can already be overcome by building known functional elements into *de novo* architectures with improved properties