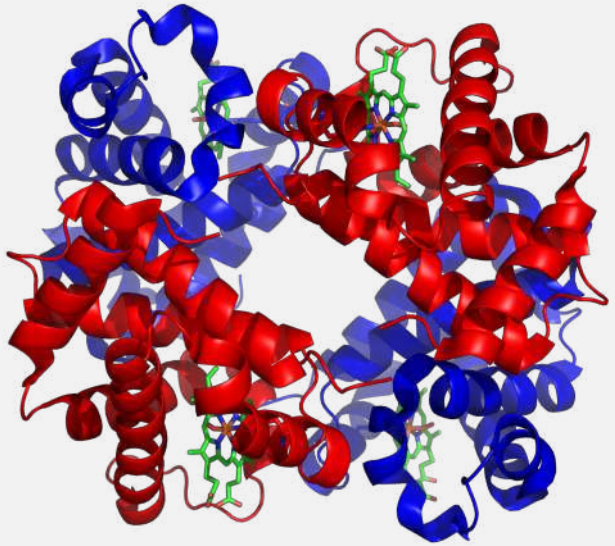
UNDERSTANDING PROTEIN FUNCTION THROUGH MULTIPLE MODELS OF STRUCTURE: BARRIERS TO INTEGRATION

Agnes Bolinska, Department of History and Philosophy of Science, University of Cambridge

Julie-Anne Gandier, Department of Bioproducts and Biosystems, Aalto University

UNDERSTANDING PROTEIN FUNCTION

- Structural – e.g. collagen
- Enzyme – catalysis of biochemical reactions
- Transport – e.g. hemoglobin
- Channels – control cell contents
- Receptors – sense stimuli, e.g. in neurons
- Immune Response – antibodies



UNDERSTANDING PROTEIN FUNCTION

- Studying protein function is difficult, since proteins are not directly accessible in their native environments
- So scientists must probe function indirectly, e.g. by determining protein structure
  - The "structure-function tenet" of protein science
- But determining protein structure is itself difficult
  - Protein structure is complex
  - It can only be accessed indirectly, through various experimental techniques

How can multiple models of protein structure best be integrated to inform our understanding of protein function?

"[T]he philosophical task is to understand what [integration] involves, how integrative practices operate, […] and what the challenges and limits to integration are"

-Brigandt (2013, pp. 461-62)

BARRIERS TO INTEGRATION IN NEUROSCIENCE

- Sullivan (2009) raises challenges for the possibility that integrative practices could establish the unity of neuroscience
- Because experimental protocols vary widely between labs, there is little hope for even the kind of non-reductive unity that Craver (2007) argues for

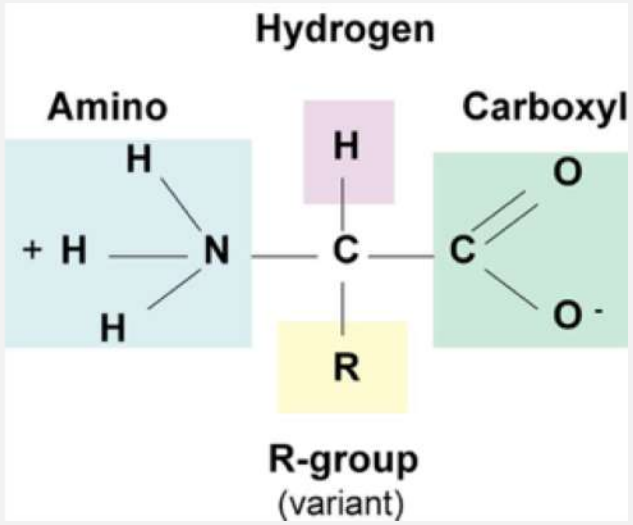
BARRIERS TO INTEGRATION IN PROTEIN SCIENCE

- Mitchell & Gronenborn argue that the relationship between multiple models of protein structure is "one of integration that maintains pluralism" (2017, p. 705)
- We show that the kind of integration they advocate is challenging:
  - Models are sometimes integrated in ways that...
    - afford certain experimental techniques more evidentiary weight
    - do not fully account for experimental context

OUTLINE

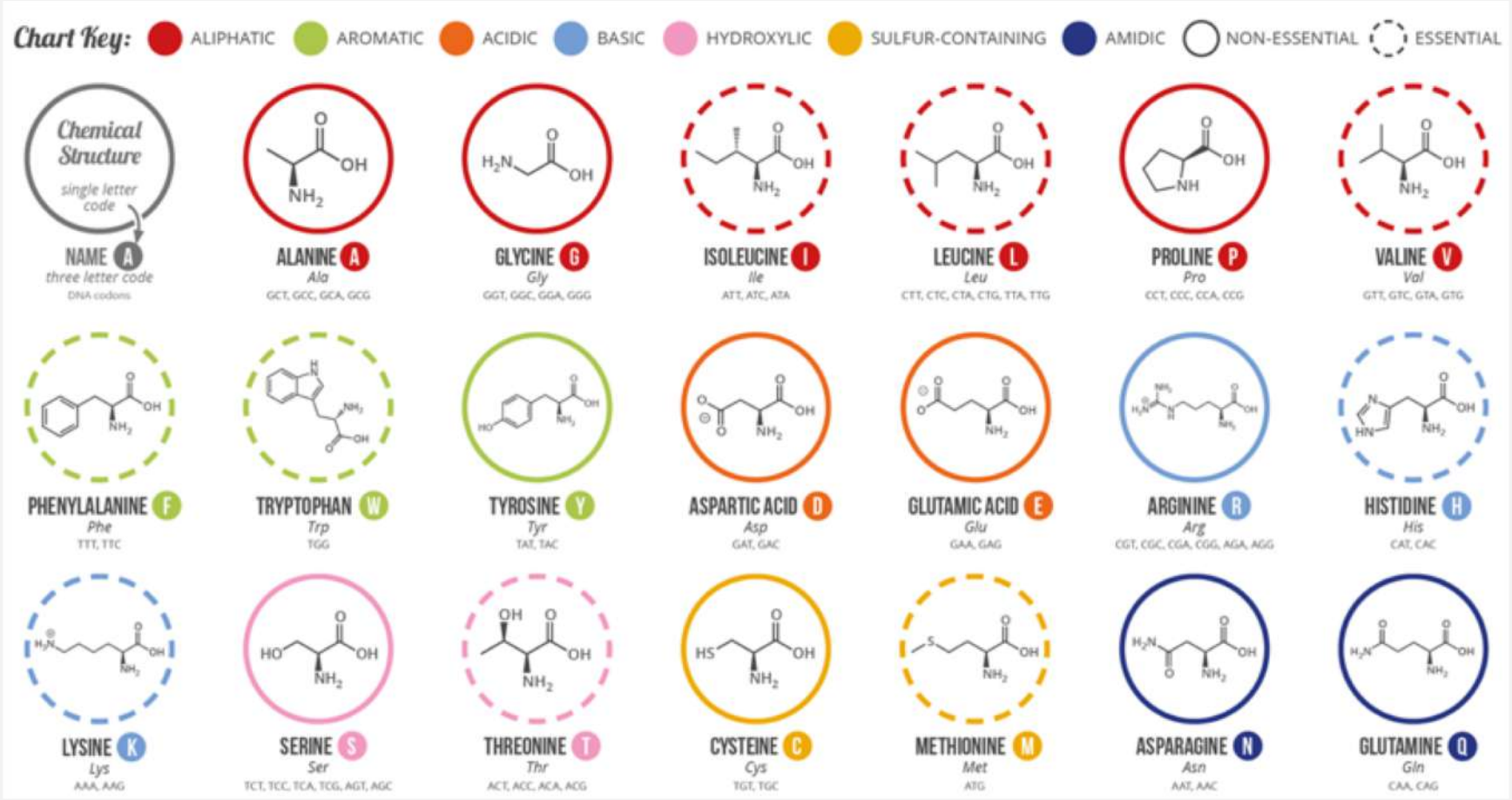
1. Protein structure and function
2. Techniques for protein structure determination
3. Barriers to integration of multiple models of protein structure
4. Case study: determining hydrophobin mechanism of action
5. Conclusion: on integrative pluralism

PROTEIN STRUCTURE AND FUNCTION

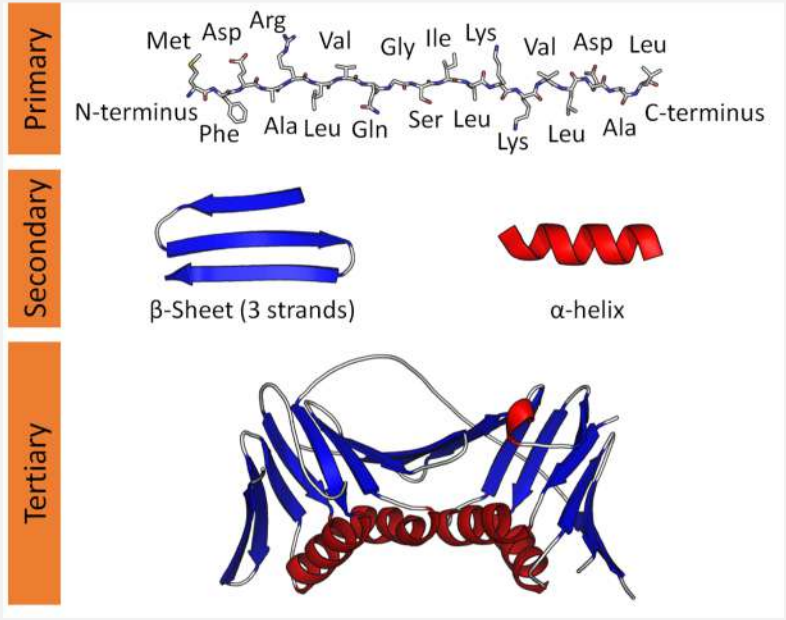


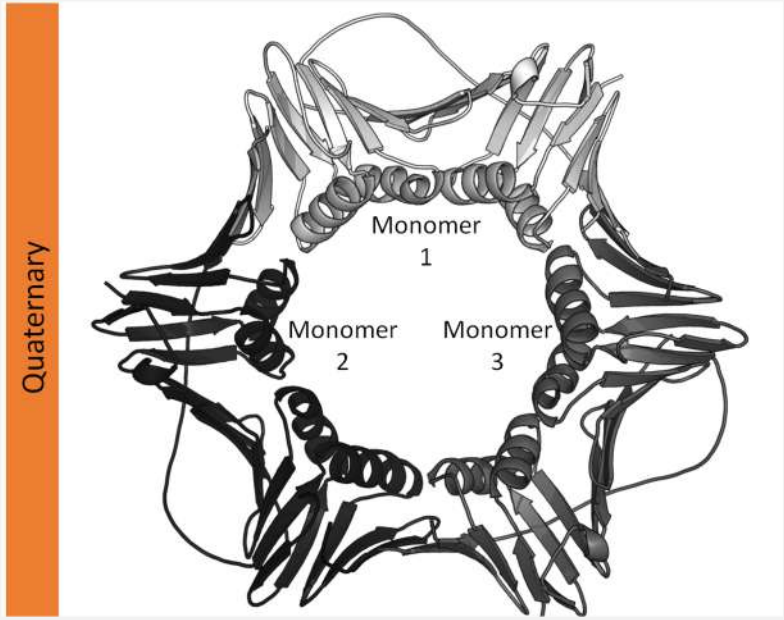
Proteins are chains of amino acids

PROTEIN STRUCTURE AND FUNCTION

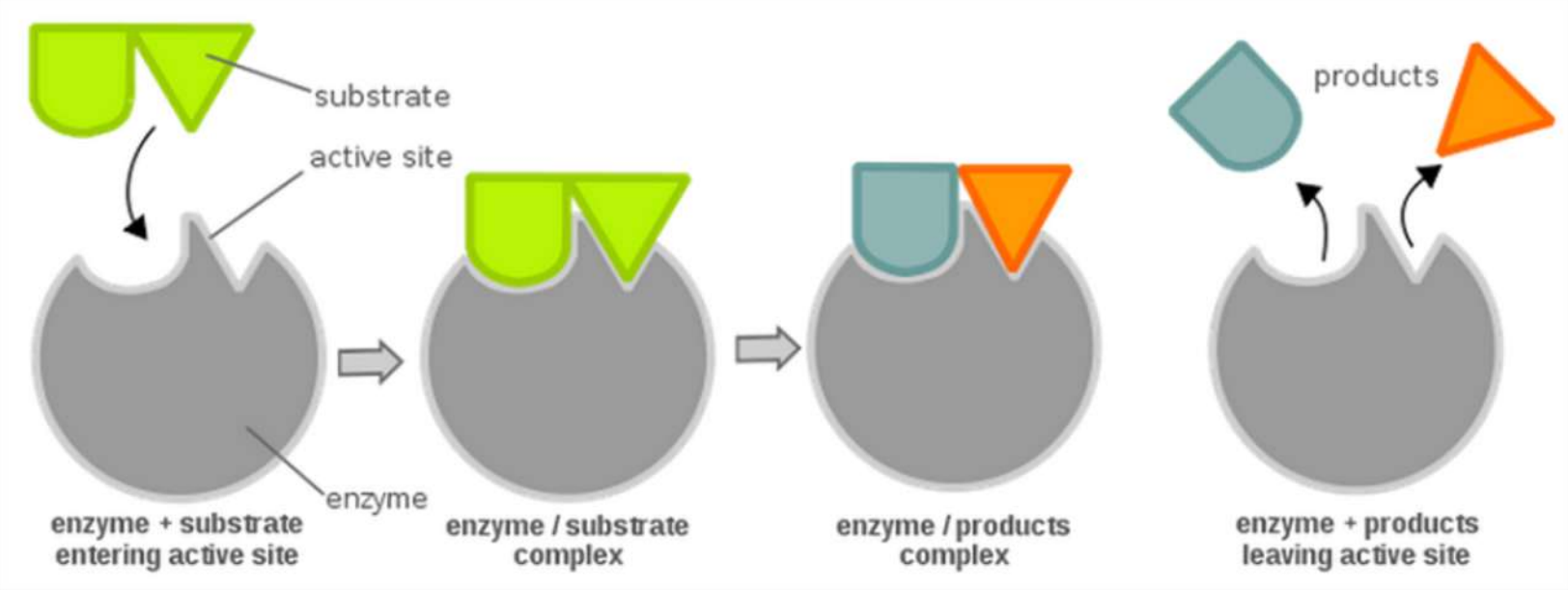


PROTEIN STRUCTURE AND FUNCTION





PROTEIN STRUCTURE AND FUNCTION



DETERMINING PROTEIN STRUCTURE

- Two types of experimental techniques:
  - **Coarse-grained** techniques provide data describing the protein at a larger scale (e.g. overall surface topology, approximate proportions of different secondary structures) and require less idealised conditions than atomistic techniques
  - **Atomistic** techniques provide the positions of atoms relative to each other in space; they require the protein to be isolated from its native environment and placed in a pure, highly concentrated state

ATOMISTIC TECHNIQUES

X-ray diffraction crystallography

- X-ray diffraction photographs are produced when a beam of X-rays is scattered by the electron clouds in a molecule
- The molecule is first crystallized

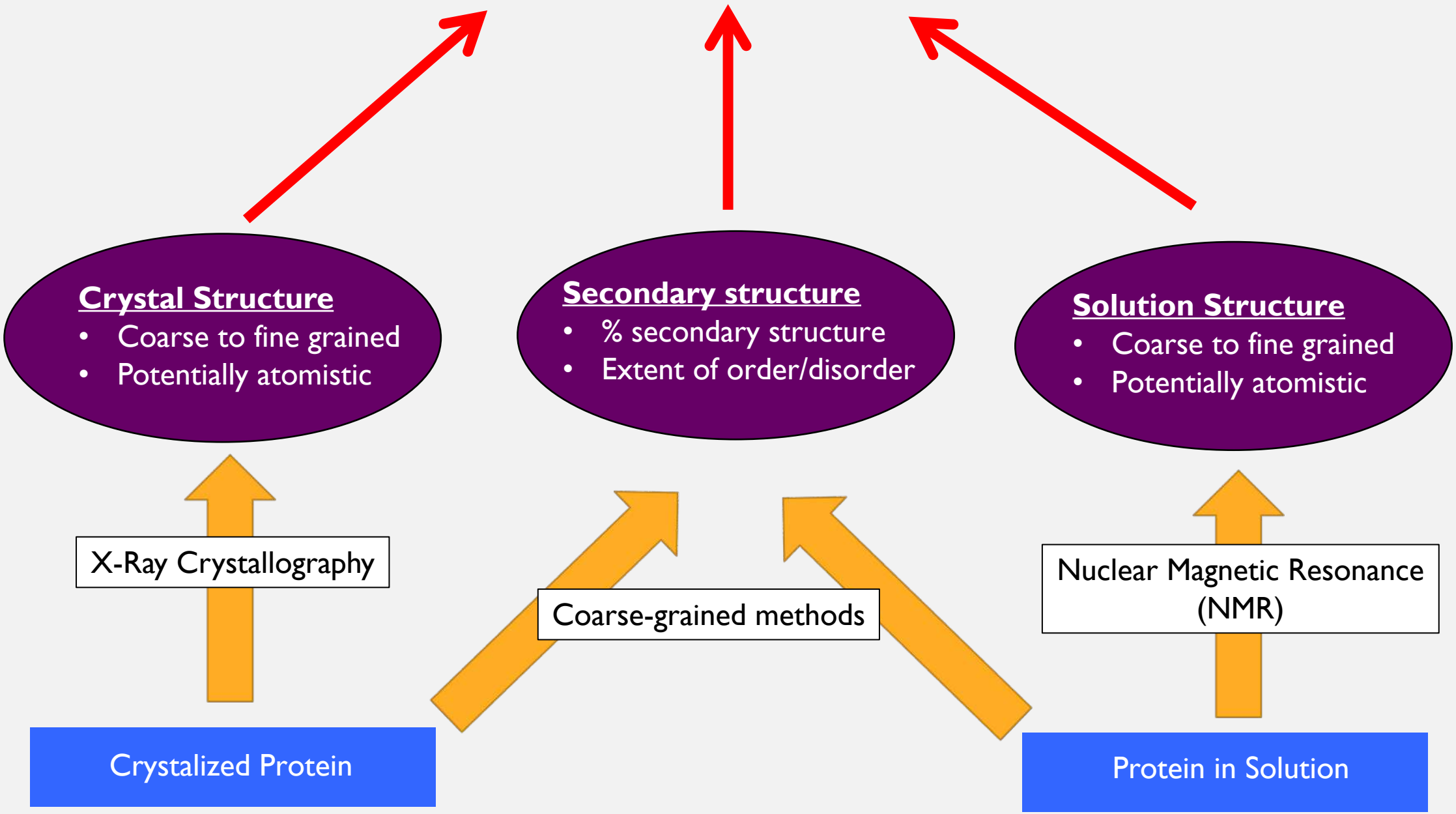
Solution NMR

- Depends upon the emission and absorption of electromagnetic radiation by the molecule's nuclei when exposed to a magnetic field
- The molecule is in aqueous solution

DETERMINING PROTEIN STRUCTURE

- Each technique produces a partial representation of protein structure, obscuring some features of structure while highlighting others
- The protein is removed from its native environment
- Solution NMR gives more direct access to protein dynamics than X-ray diffraction photography

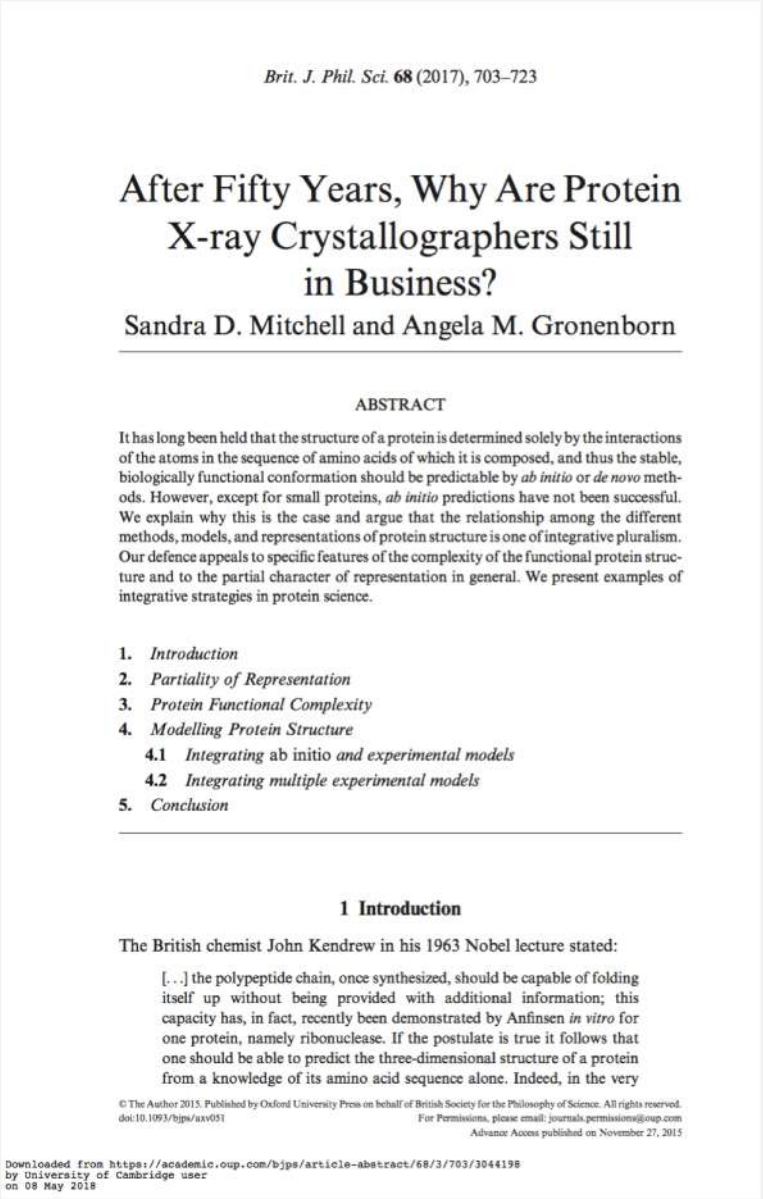
UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION



How can we best integrate multiple models to inform our understanding of protein structure?

- Mitchell and Gronenborn (2017, p. 711): "Our thesis is that these multiple models are integrated in the service of [a] shared [scientific] goal."

- They highlight several techniques for the integration of crystal and NMR models: using an NMR structure to solve a crystal structure via molecular replacement, using a crystal structure as an input for an NMR model, joint refinement approach



Brit. J. Phil. Sci. **68** (2017), 703–723

After Fifty Years, Why Are Protein X-ray Crystallographers Still in Business?

Sandra D. Mitchell and Angela M. Gronenborn

ABSTRACT

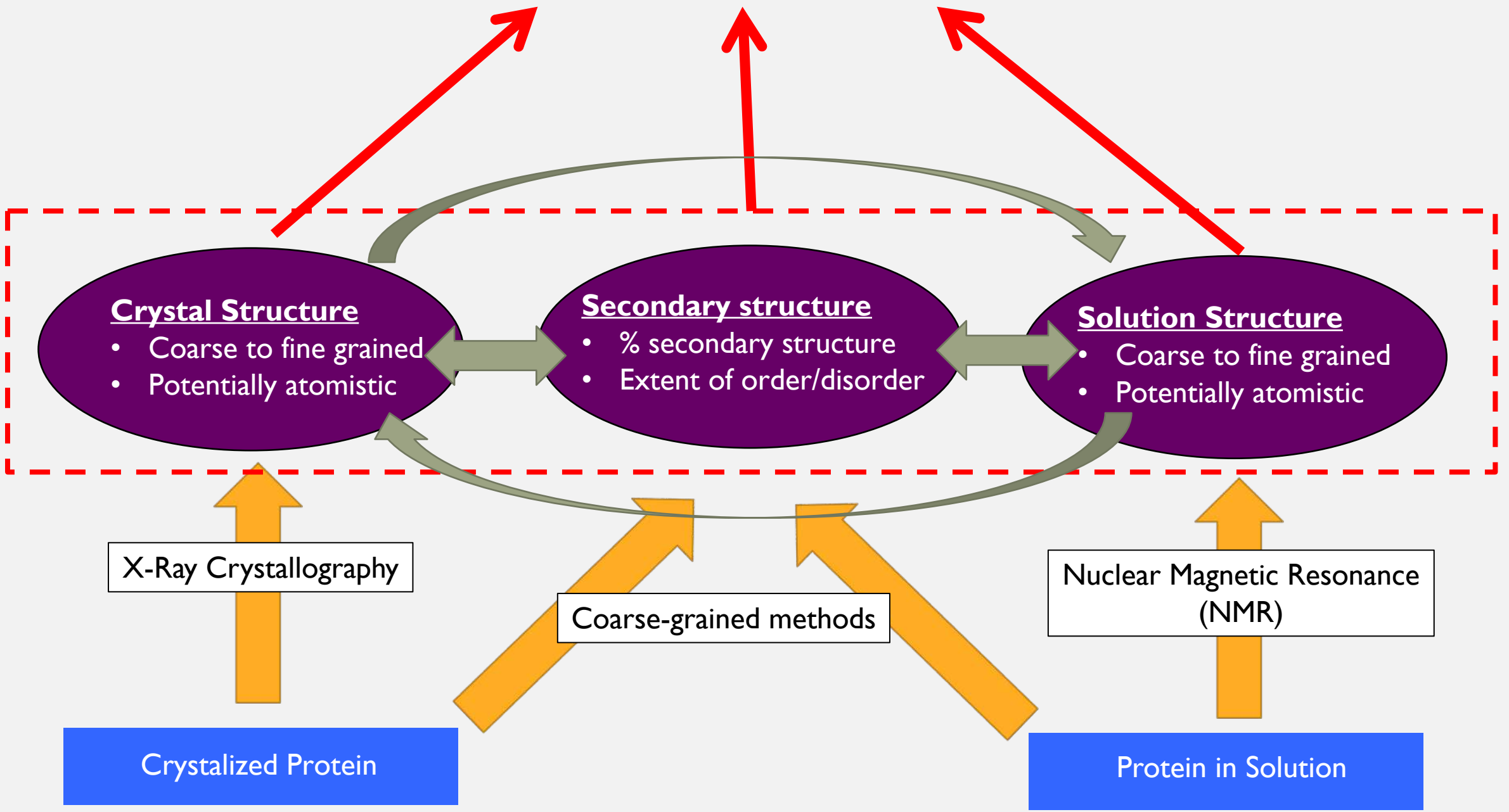
It has long been held that the structure of a protein is determined solely by the interactions of the atoms in the sequence of amino acids of which it is composed, and thus the stable, biologically functional conformation should be predictable by *ab initio* or *de novo* methods. However, except for small proteins, *ab initio* predictions have not been successful. We explain why this is the case and argue that the relationship among the different methods, models, and representations of protein structure is one of integrative pluralism. Our defence appeals to specific features of the complexity of the functional protein structure and to the partial character of representation in general. We present examples of integrative strategies in protein science.

1 Introduction

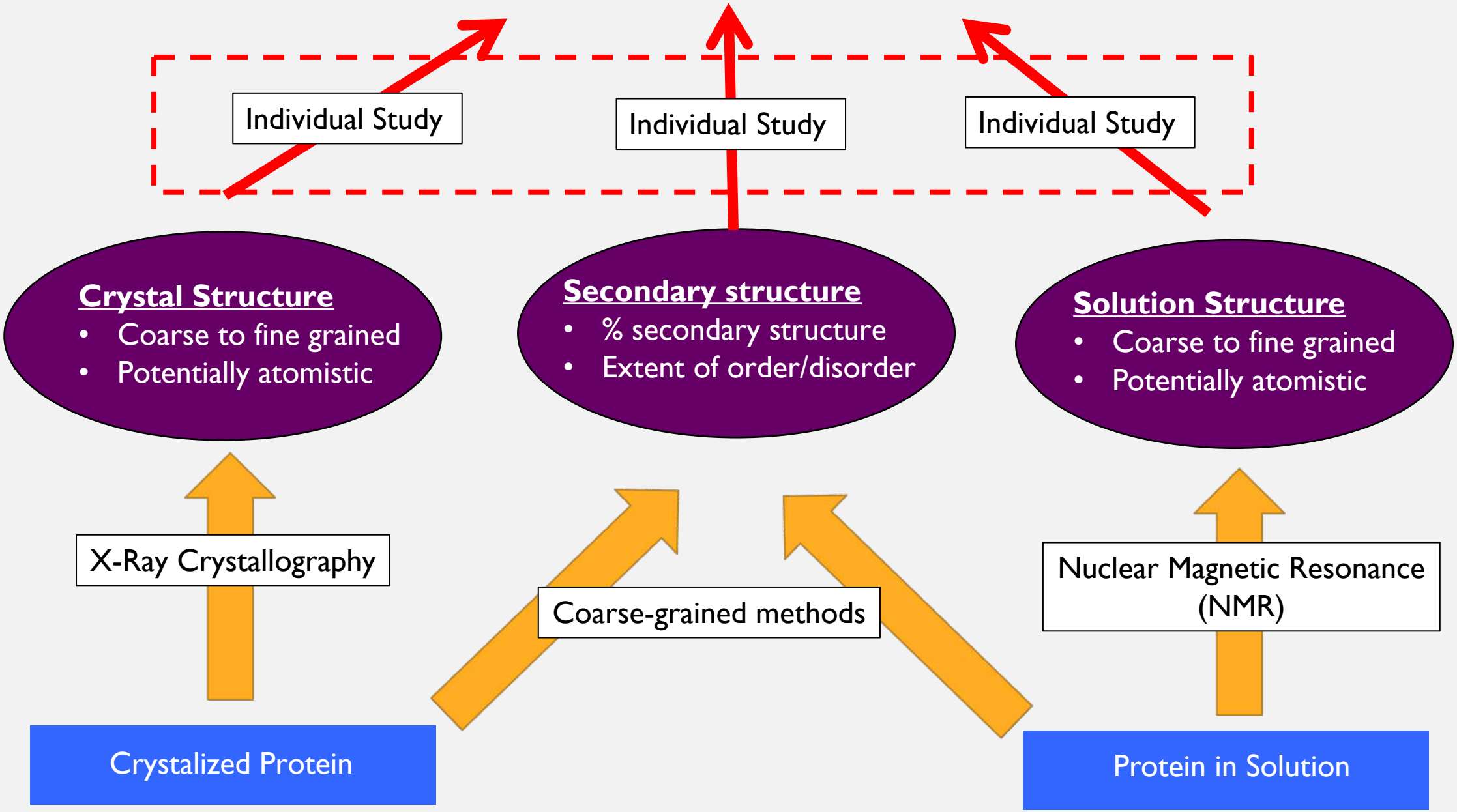
The British chemist John Kendrew in his 1963 Nobel lecture stated:

> [. . .] the polypeptide chain, once synthesized, should be capable of folding itself up without being provided with additional information; this capacity has, in fact, recently been demonstrated by Anfinsen *in vitro* for one protein, namely ribonuclease. If the postulate is true it follows that one should be able to predict the three-dimensional structure of a protein from a knowledge of its amino acid sequence alone. Indeed, in the very

doi:10.1093/bjps/axv051
Advance Access published on November 27, 2015


UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION
Crystal Structure
• Coarse to fine grained
• Potentially atomistic
Secondary structure
• % secondary structure
• Extent of order/disorder
Solution Structure
• Coarse to fine grained
• Potentially atomistic
X-Ray Crystallography
Coarse-grained methods
Nuclear Magnetic Resonance (NMR)
Crystalized Protein
Protein in Solution

UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION

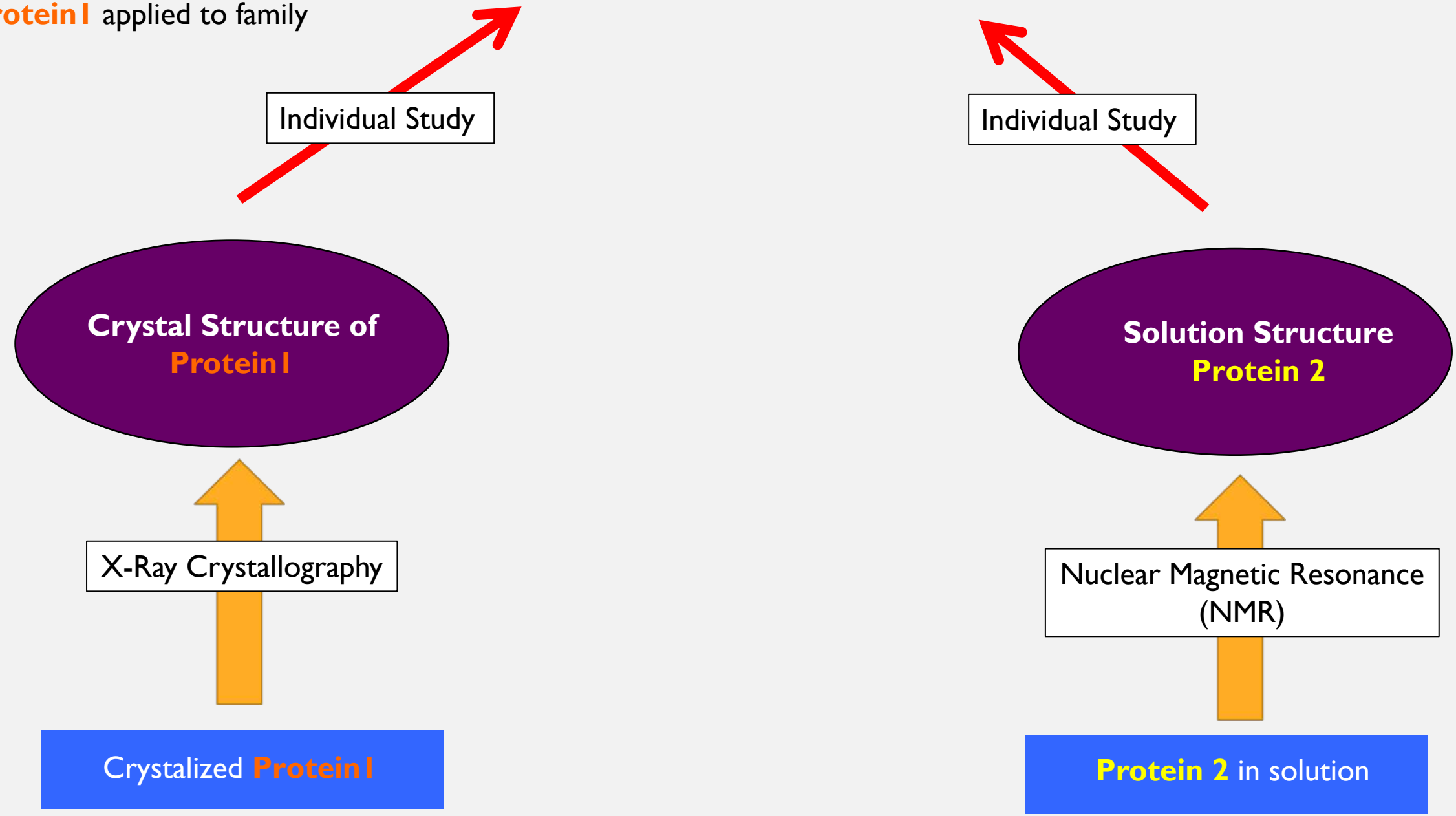


BARRIERS TO INTEGRATION

- It's often impossible to apply several techniques to the study of a single protein
- So instead, different techniques are applied to different proteins, each of which produces a model that serves as a partial representation
- Moreover, models produced using certain techniques are afforded more evidentiary weight than others, such that subsequent models are understood as standing in a confirmatory relationship to them
  - We call this the *model-ladenness of interpretation*
- And structures produced using them are taken to be representative of the family as a whole

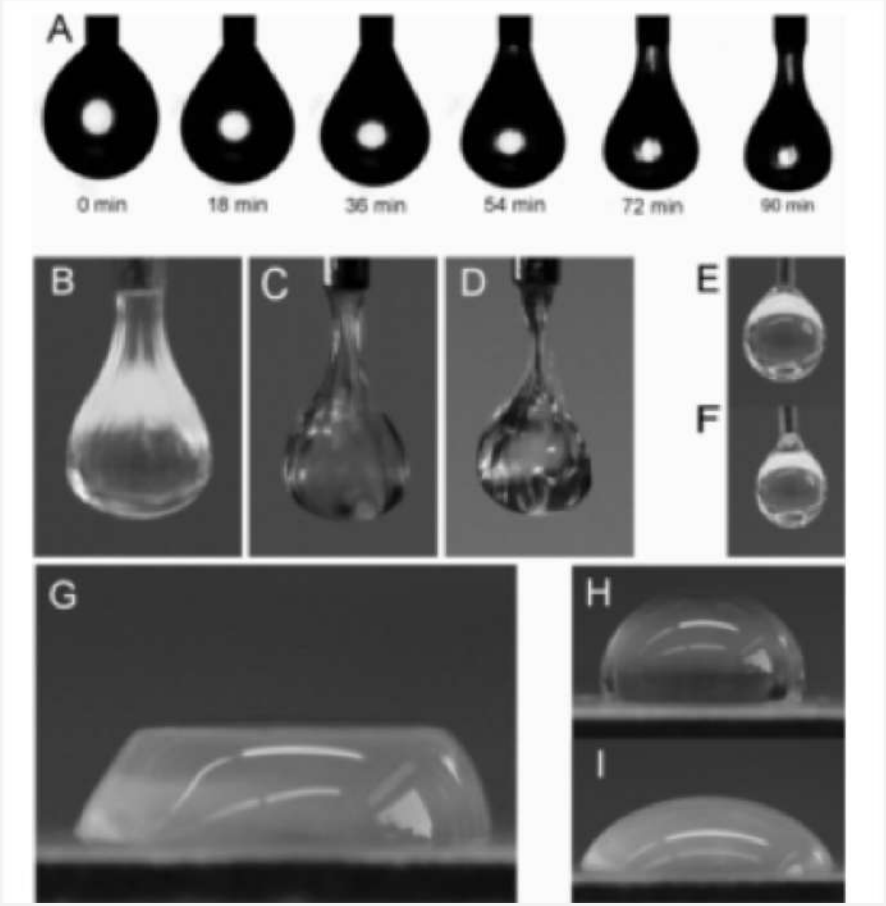
UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION

Protein I applied to family



DETERMINING HYDROPHOBIN STRUCTURE

- Hydrophobins are produced and secreted ubiquitously by filamentous fungi
- They self-assemble to form highly stable films at any interface (i.e. solid-liquid, liquid-liquid, or liquid-air)
- To determine the mechanism underlying this function, scientists look to determine the structure of the protein

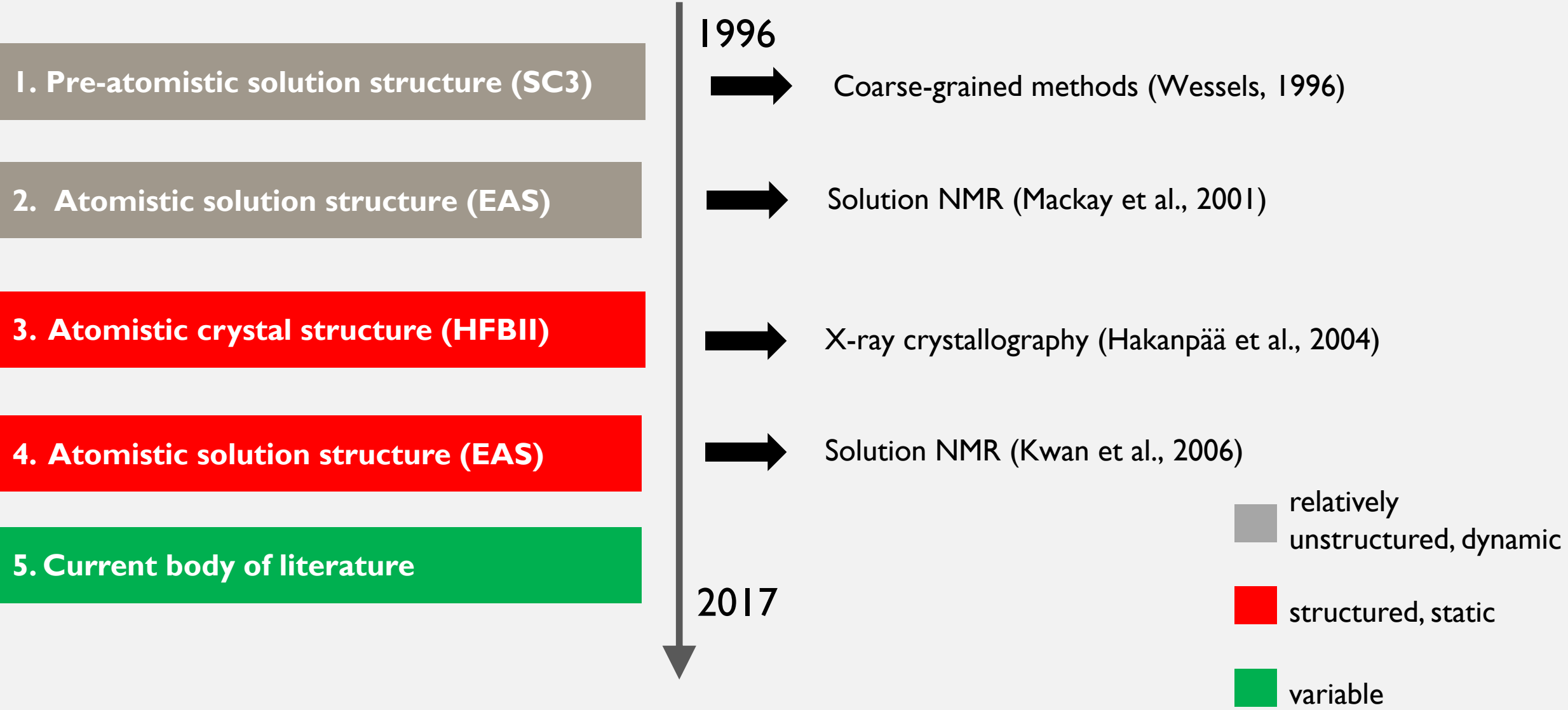


(Szilvay et al. 2007, p. 2348)

DETERMINING HYDROPHOBIN STRUCTURE

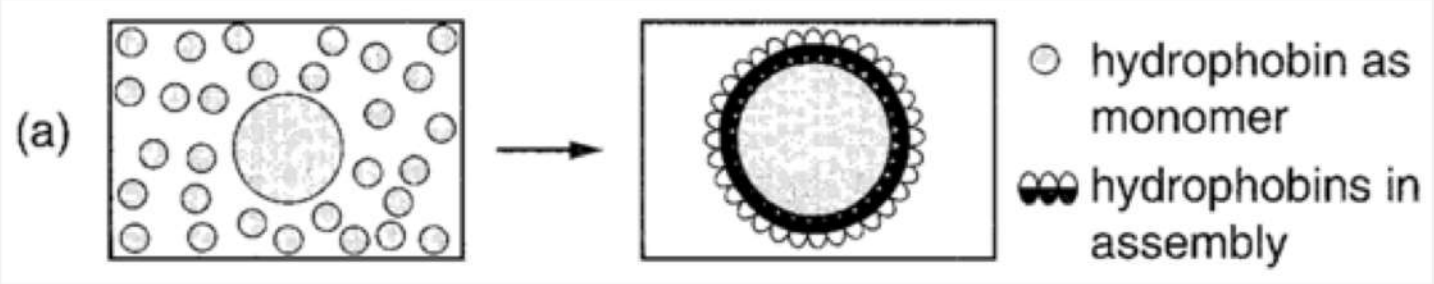
- We should exercise caution when integrating multiple partial representations of structure in the hydrophobin family
- Normally, proteins in a family share high sequence similarity (35% or higher)
- However, hydrophobins are defined by the position and chemical nature of only eight amino acids (out of approximately one hundred)
- Normally, proteins in a family share a well-defined specific function
- However, the unifying function in the hydrophobin family is relatively vague ("self-assembly" at interfaces)

DETERMINING HYDROPHOBIN STRUCTURE



1. Pre-atomistic solution structure (SC3)

Early view of the hydrophobin structure: dynamic in solution and structured at the interface



"Compared to other proteins, SC3 is very surfactive, the lowering of the surface tension mainly resulting from a conformational change during assembly of the monomers into an amphipathic film" (Wessels 1996, p. 12).

DETERMINING HYDROPHOBIN STRUCTURE

1. Pre-atomistic solution structure (SC3)

2. Atomistic solution structure (EAS)

3. Atomistic crystal structure (HFBII)

4. Atomistic solution structure (EAS)

5. Current body of literature

1996

→ Coarse-grained models: Hydrophobins are dynamic in solution and structured at the interface

→ Solution NMR (Mackay et al., 2001)

→ X-ray crystallography (Hakanpää et al., 2004)

→ Solution NMR (Kwan et al., 2006)

2017

relatively unstructured, dynamic

structured, static

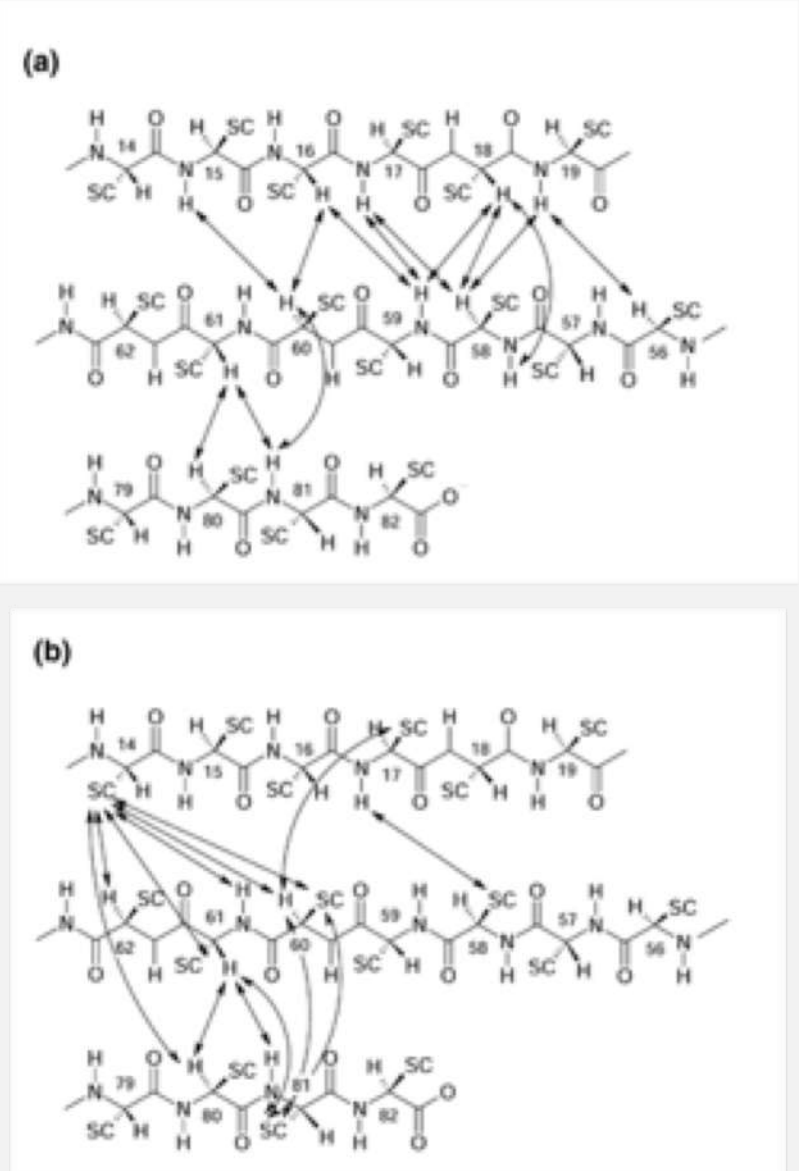
variable

2. Atomistic solution structure (EAS)

Solution NMR produces high resolution "structural information" on a member of the hydrophobin family

" We have found that EAS is monomeric, but mostly unstructured in solution, except for a small region of antiparallel β sheet …" (Mackay et al. 2001, p. 83)

"EAS joins an increasing number of proteins that undergo a disorder to order transition in carrying out their normal function..." (Mackay et al. 2001, p. 83)



β-sheet topology of EAS (a) between backbone atoms and (b) between sidechains

DETERMINING HYDROPHOBIN STRUCTURE

1. Pre-atomistic solution structure (SC3)
2. Atomistic solution structure (EAS)
3. Atomistic crystal structure (HFBII)
4. Atomistic solution structure (EAS)
5. Current body of literature

1996

→ Coarse-grained models: Hydrophobins are dynamic in solution and structured at the interface

→ NMR structure (Mackay et al., 2001): Hydrophobins undergo a disorder to order transition at the interface

→ X-ray crystallography (Hakanpää et al., 2004)

→ Solution NMR (Kwan et al., 2006)

2017

relatively unstructured, dynamic

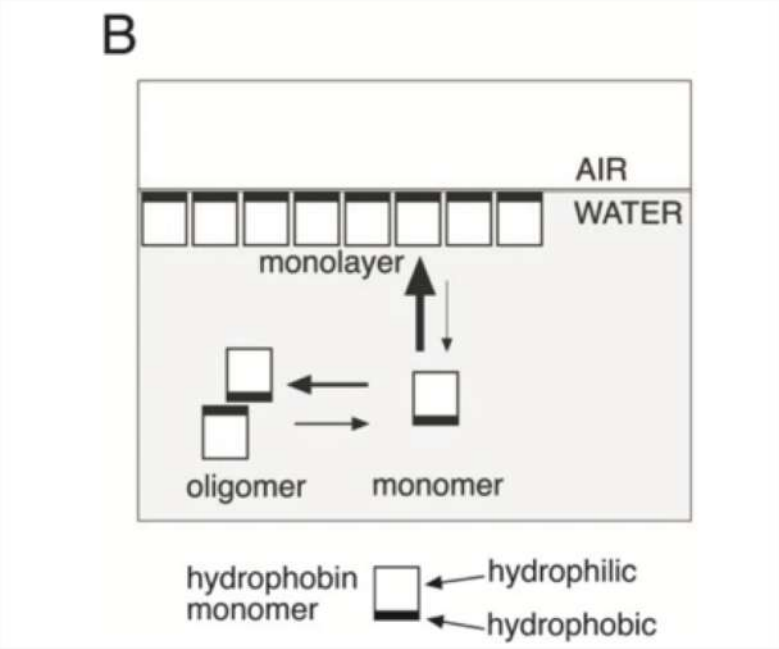
structured, static

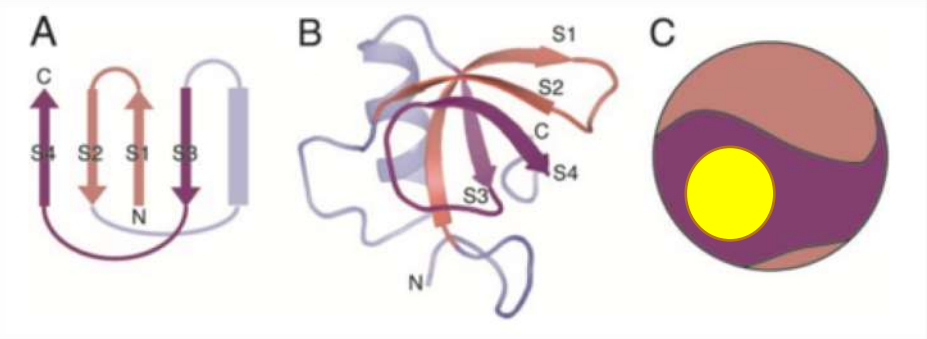
variable

3. Atomistic crystal structure (HFBII)

A seemingly incompatible account …

The hydrophobin protein that could be confined to a crystal displayed a highly rigid structure with little disorder





(Hakanpää et al. 2004, p.536)

3. Atomistic crystal structure (HFBII)

A seemingly incompatible account …

The hydrophobin protein that could be confined to a crystal displayed a highly rigid structure with little disorder

- Hakanpää et al. concluded that "the data presented show that much of the current views on structure-function relations in hydrophobins must be re-evaluated …" (2004, p. 538)
- But this was unwarranted, especially considering how dissimilar the primary sequences of proteins in this family are
- The crystal structure's application to the whole family reflects the greater evidentiary weight placed on this than on models from NMR and coarse-grained techniques

DETERMINING HYDROPHOBIN STRUCTURE

1. Pre-atomistic solution structure (SC3)
2. Atomistic solution structure (EAS)
3. Atomistic crystal structure (HFBII)
4. Atomistic solution structure (EAS)
5. Current body of literature

1996

→ Coarse-grained models: Hydrophobins are dynamic in solution and structured at the interface

→ NMR structure (Mackay et al., 2001): Hydrophobins undergo a disorder to order transition at the interface

→ Crystal structure (Hakanpää et al., 2004): Hydrophobins are highly rigid and structured. Previous data must be reevaluated.

→ Solution NMR (Kwan et al., 2006)

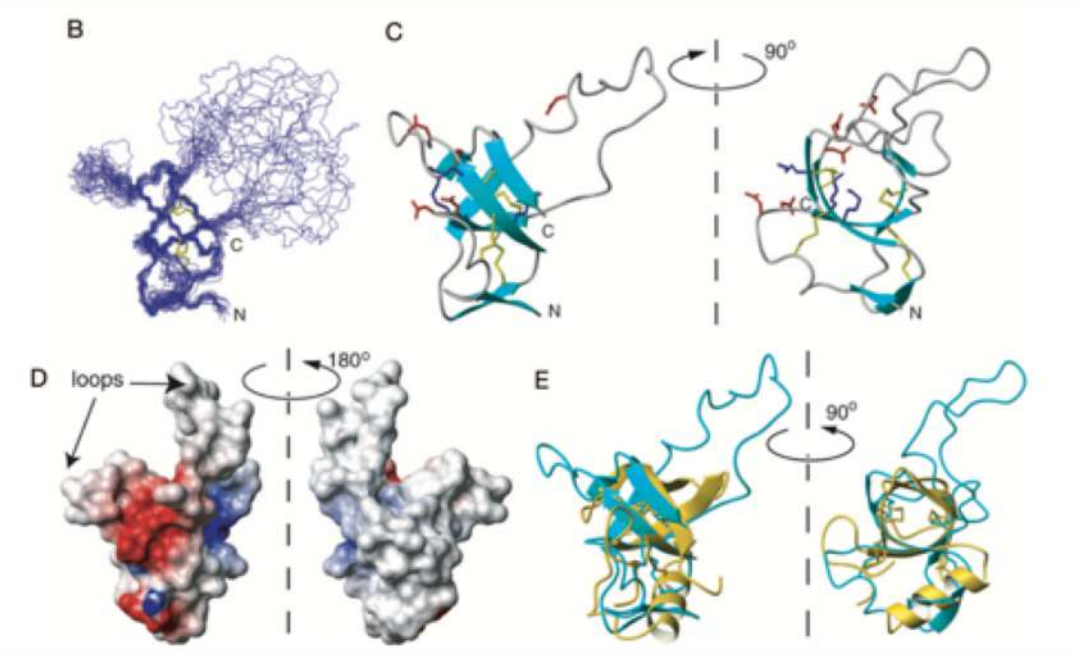
2017

relatively unstructured, dynamic

structured, static

variable

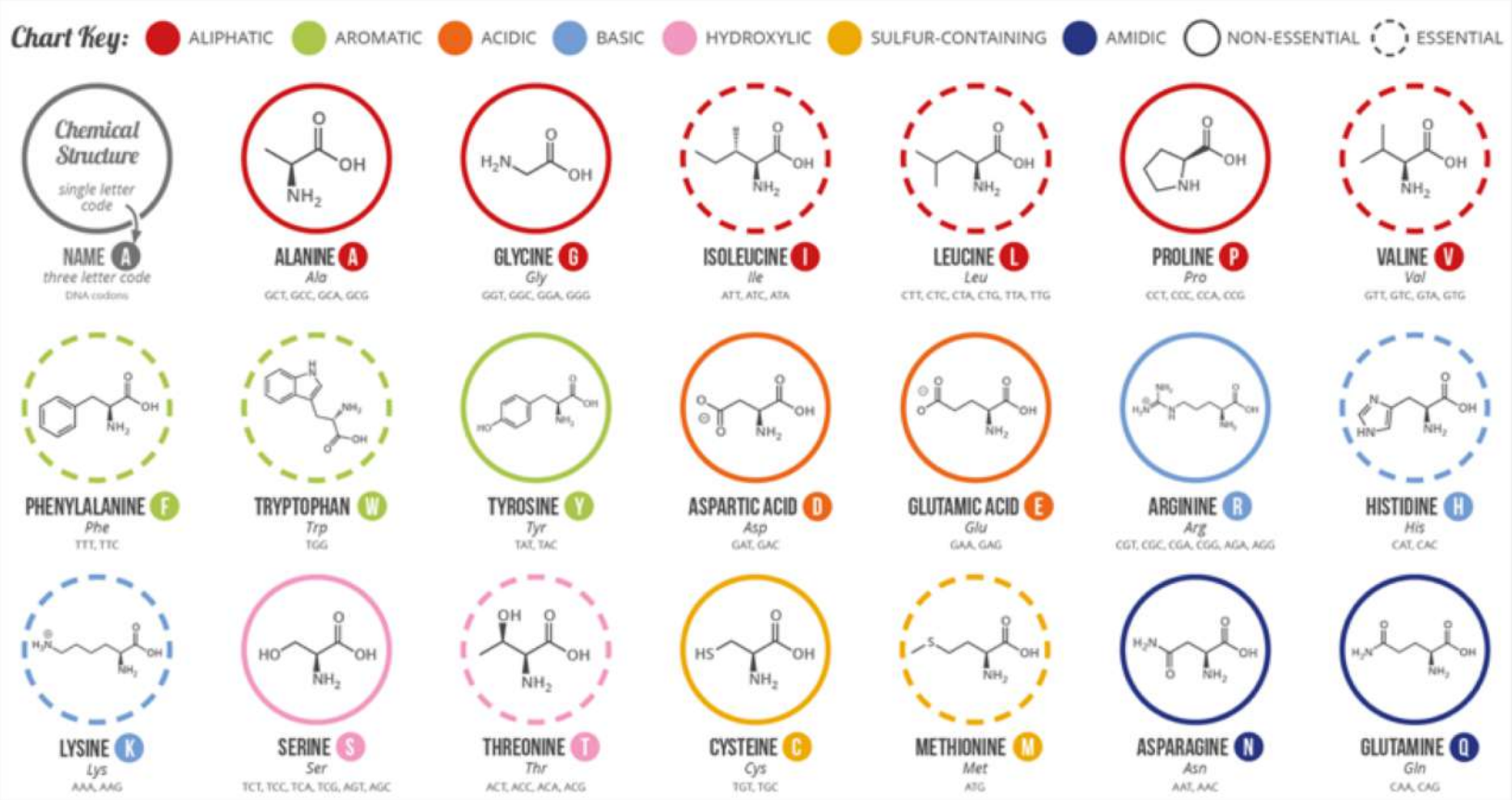
4. Atomistic solution structure (revisiting EAS)



(Kwan et al. 2006, p. 3622)

"EAS forms a β-barrel structure punctuated by several disordered regions and displays complete segregation of charged and hydrophobic residues on its surface…"

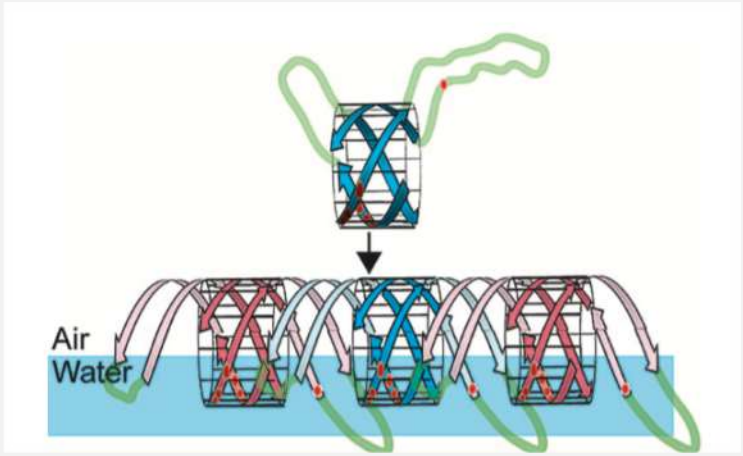
4. Atomistic solution structure (revisiting EAS)



- The hydrophobic patch of HFBII contains mostly amino acids with aliphatic chains, the most hydrophobic residues you can have
- In contrast, the "hydrophobic patch" of EAS contains merely the *absence of* the most hydrophilic residues, *viz*. charged residues

4. Atomistic solution structure (revisiting EAS)

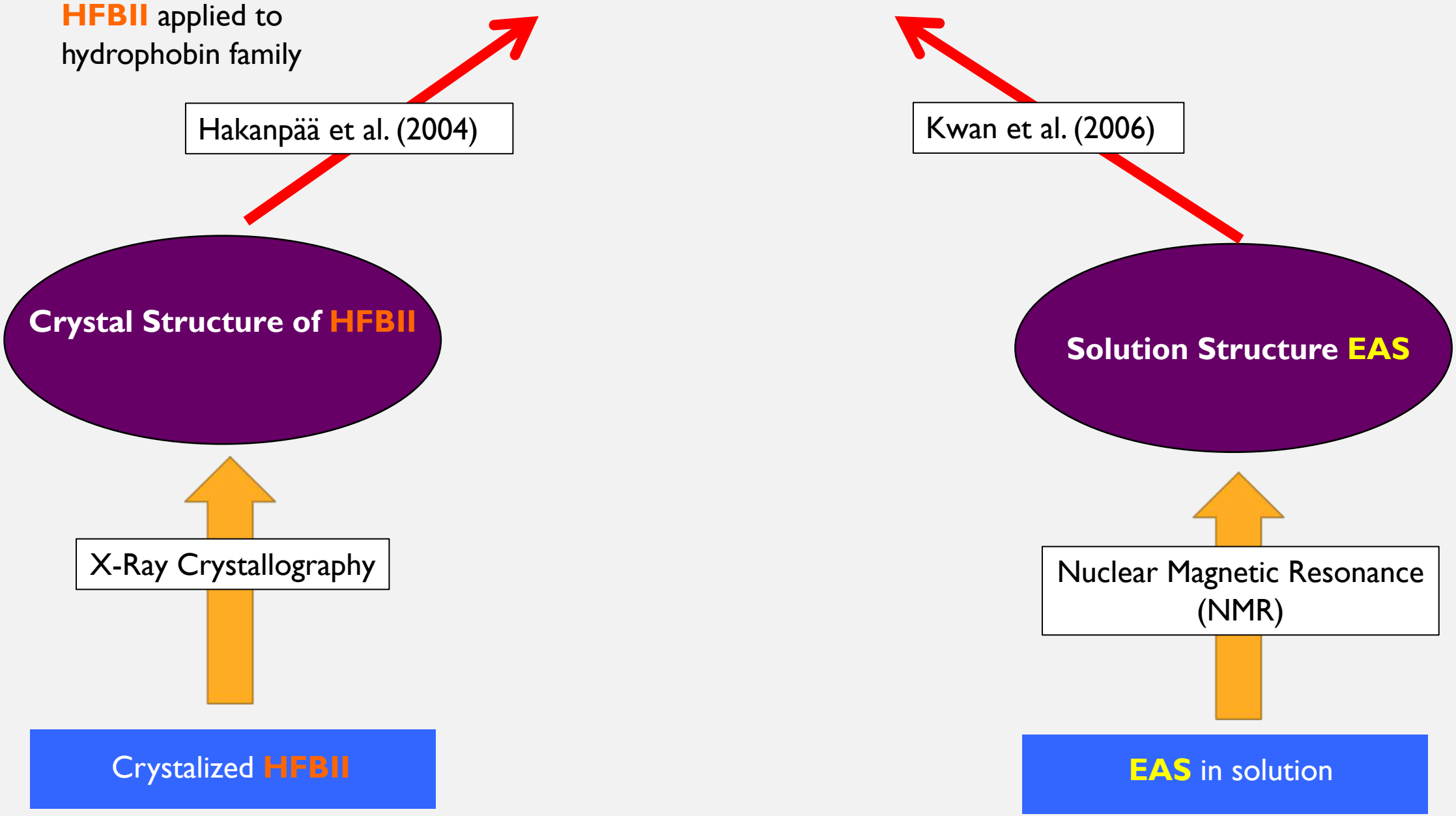
What does this imply in the explanation of function i.e. self-assembly?



"Given that there is a single continuous charged patch on the surface of the EAS core and that the diametrically opposite face is completely hydrophobic, the simplest way of arranging monomers in the monolayer is for the charged side to face the water." (Kwan et al. 2006, p. 3623)

Significant changes in structure at the interface are not considered in this explanation of self-assembly

UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION



DETERMINING HYDROPHOBIN STRUCTURE

1. Pre-atomistic solution structure (SC3)
2. Atomistic solution structure (EAS)
3. Atomistic crystal structure (HFBII)
4. Atomistic solution structure (EAS)
5. Current body of literature

1996

→ Coarse-grained models: Hydrophobins are dynamic in solution and structured at the interface

→ NMR structure (Mackay et al., 2001): Hydrophobins undergo a disorder to order transition at the interface

→ Crystal structure (Hakanpää et al., 2004): Hydrophobins are highly rigid and structured. Previous data must be reevaluated.

→ Solution NMR (Kwan et al., 2006): relatively structured, no conformational change at the interface

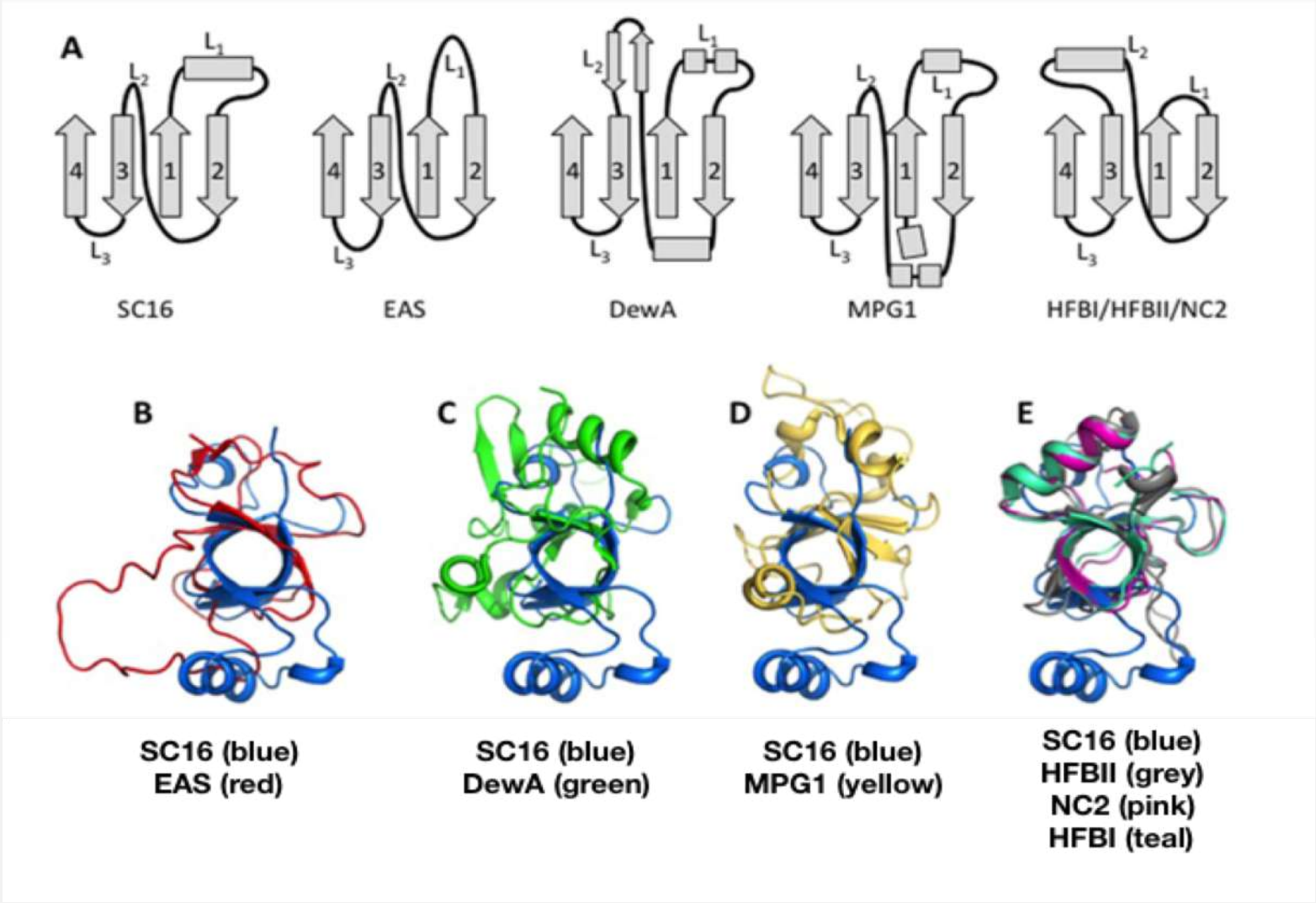
2017

relatively unstructured, dynamic

structured, static

variable

5. Current body of literature



(Gandier et al., 2017)

5. Current body of literature

- If integration is so challenging, and faces the barriers we describe, then how were scientists able to determine these structures?
- Protein science is an iterative process: over the course of multiple studies, self-correction can take place
- But had due attention been paid to different experimental contexts, this process could have been more efficient

SUMMARY

- We highlighted some barriers to the integration of multiple models of protein structure:
  - The structures of different proteins, resolved using different experimental techniques, are compared to one another
  - And some models are afforded more evidentiary weight, thereby influencing how the results of subsequent investigations into structure and function are interpreted

SUMMARY

- We demonstrated this using the hydrophobin case study
  - Early work on hydrophobin structure and function held that hydrophobins are mostly unstructured in solution and undergo a conformational change during self-assembly
  - The crystal model of HFBII was ordered, with no conformational change driving function
  - This influenced a second solution NMR model of EAS and the understanding of hydrophobin function more generally

How should models of structure best integrated with one another with the aim of understanding protein function?

HOW SHOULD MODELS BE INTEGRATED?

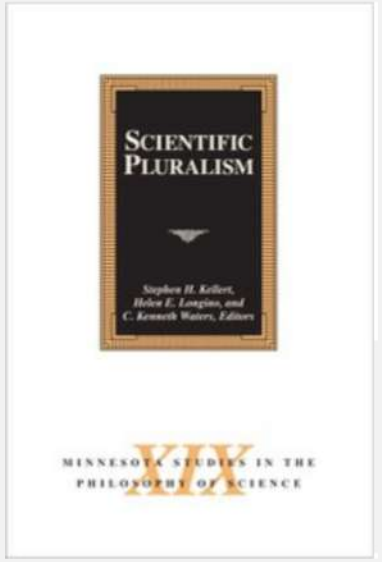
- Models produced using different experimental techniques should be integrated in a way that allows each to highlight those features that it is designed to exhibit
- Thus, when models produced using different techniques are integrated, careful attention should be paid to how this is done
  - Certain features can be integrated, while others are retained
  - But how do we know which ought to be integrated and which ought to be retained?

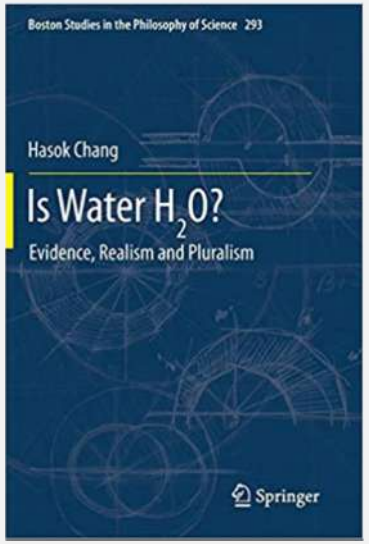
ON INTEGRATIVE PLURALISM

Mitchell and Gronenborn (2017, p. 705): the relationship between models of protein structure produced by different techniques is “one of integration that maintains pluralism,” rather than of unification into a single, comprehensive model

ON INTEGRATIVE PLURALISM

"Pluralism is here to stay: science lives in a world of multiple models, and they cannot always, or perhaps even often, be reduced or unified into one 'complete' model" (Mitchell & Gronenborn 2017, p. 711).





ON INTEGRATIVE PLURALISM

"Using a joint refinement approach, an overall better model of a protein structure can be derived by combining X-ray and NMR data (Shaanan et al. 1992). This type of integration reduces the under-determination in the models inherent to each methodology. Using data from both methods in refinement reduces the total range of possible models by mutually correcting individual model bias" (Mitchell & Gronenborn 2017, p. 17)

ON INTEGRATIVE PLURALISM

"While it is clear that a given protein will have a given structure under specific conditions, neither experimental nor inferential protocol is expected to perfectly or completely detect it, and different environments can further modulate the molecular behaviour that influences the targeted atomic properties. Integrating multiple models from different experimental protocols provides a means to reach more accurate results than relying on any single method."
(Mitchell & Gronenborn 2017, p. 17)

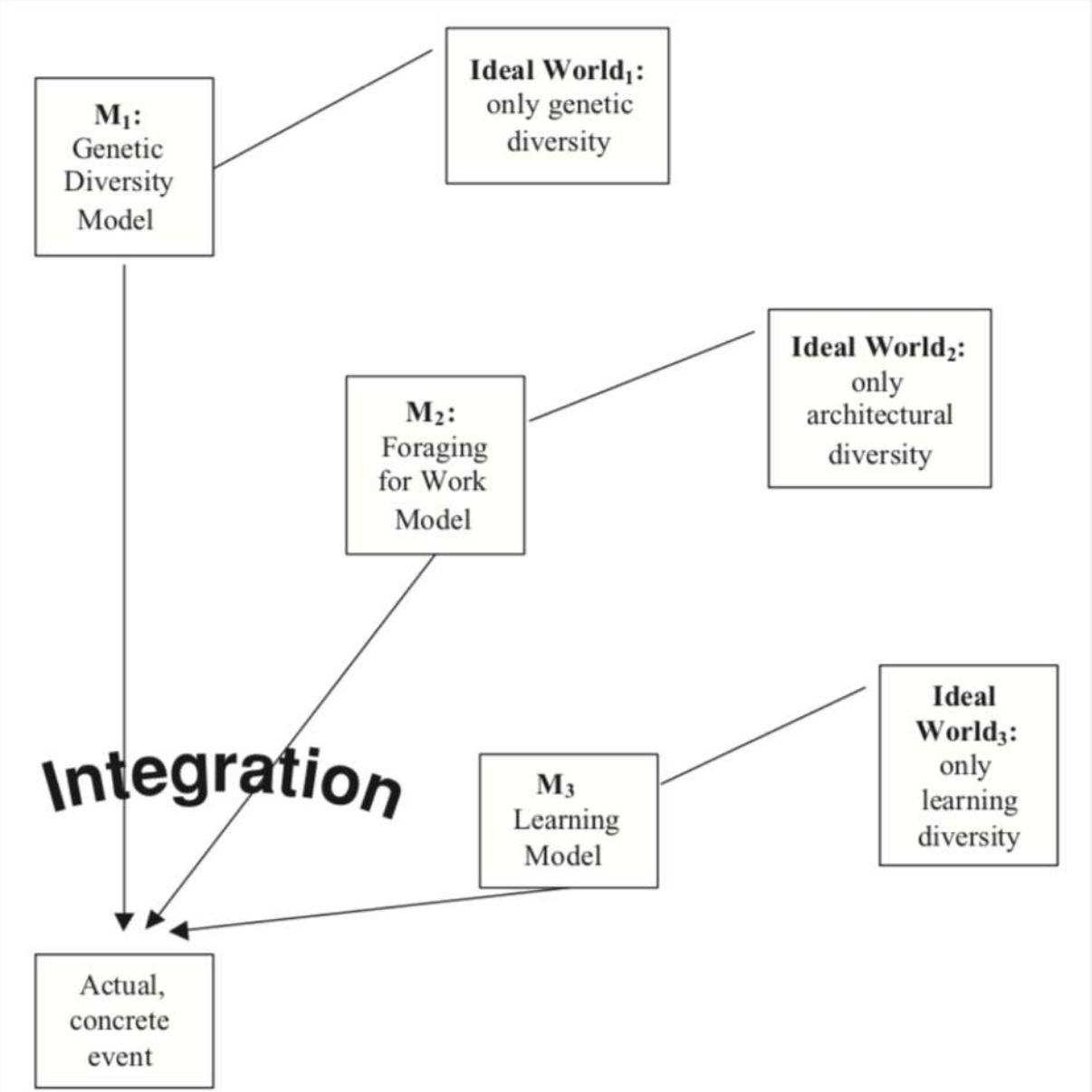
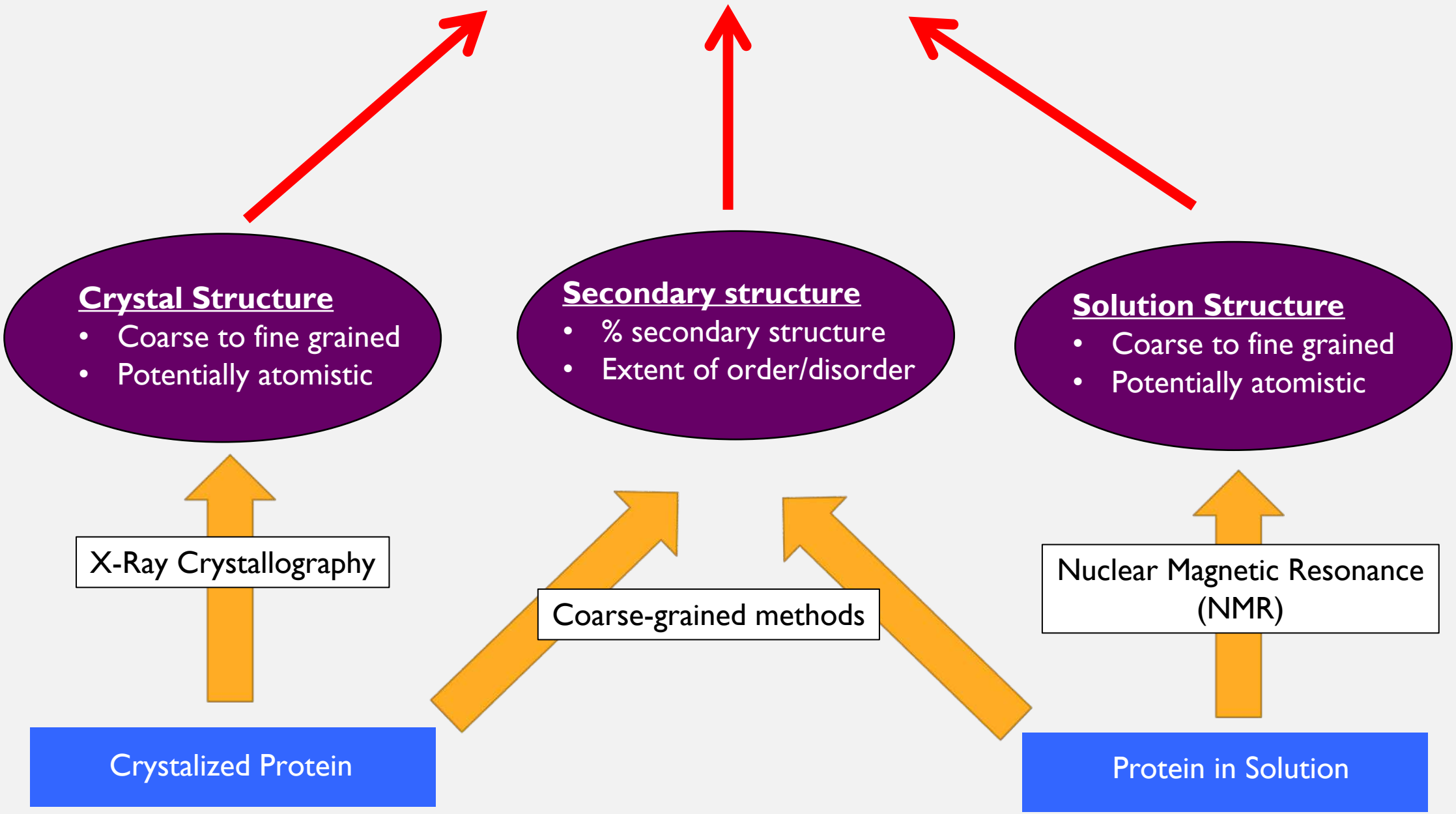


Figure 6.1. Causal models are idealizations.

Mitchell (2003, p. 215)

UNDERLYING MECHANISM OF BIOLOGICAL FUNCTION



Thank you!

amb273@cam.ac.uk

julie-anne.gandier@aalto.fi

References:

- Brigandt, I. (2013). Integration in biology: Philosophical perspectives on the dynamics of interdisciplinarity. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44: 461-65.
- Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
- Gandier et al. (2017) Characterization of a Basidiomycota hydrophobin reveals the structural basis for a high-similarity ClassI subdivision. *Scientific Reports*. **7**:45863.
- Hakanpää, J. *et al.* (2004). Atomic Resolution Structure of the HFBII Hydrophobin, a Self-assembling Amphiphile. *J. Biol. Chem.* **279,** 534–539.
- Kwan, A. H. Y. *et al.* (2006). Structural basis for rodlet assembly in fungal hydrophobins. *Proc. Natl. Acad. Sci. USA* **103,** 3621–3626.
- Mackay, J. P. *et al.* (2001) *The* hydrophobin EAS is largely unstructured in solution and functions by forming amyloid-like structures. *Structure*. **9,** 83–91.
- Mitchell, S. D. (2003). *Biological Complexity and Integrative Pluralism.* Cambridge: Cambridge University Press.
- Mitchell, S. D., & Gronenborn, A. M. (2017). After Fifty Years, Why Are X-Ray Crystallographers Still in Business? *The British Journal for the Philosophy of Science*, 68 (3):703-723.
- Sullivan, J. A. (2009). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese* 167: 511–539.
- Szilvay S.R. et al. (2007). Self-Assembled Hydrophobin Protein Films at the Air-Water Interface: Structural Analysis and Molecular Engineering. Biochemistry, 46:.2345-2354
- Wessels, J. G. H. (1996). Fungal hydrophobins: proteins that function at an interface. *Trends in Plant Science Reviews* **1**, 9-15.