

3<sup>RD</sup> EDITION | ADVANCED  
TRAINING COURSE IN  
**MOLECULAR  
BIOENGINEERING**

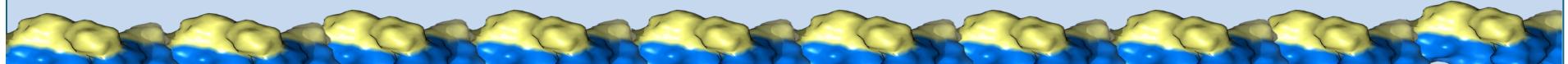


# Simulating Peptides: Insights into 3D Structure, Lipophilicity, and Aggregation

*Dr. William J. Zamora R*



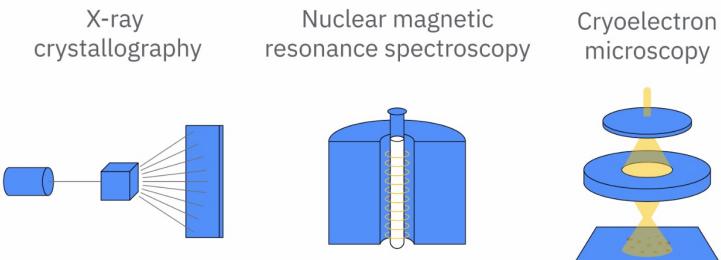
Laboratorio de Toxicología Computacional  
e Inteligencia Artificial  
LEBI Laboratorio de  
Ensayos Biológicos





# Introduction: Proteins and Structure

## Protein structures at atomic levels of detail (experimental methods)



These **experimental techniques** are:

*Inherently time and labour-consuming*

*Involve many troubleshooting steps*

SCIENCE VOL 343 28 MARCH 2014

PERSPECTIVES

### BIOCHEMISTRY

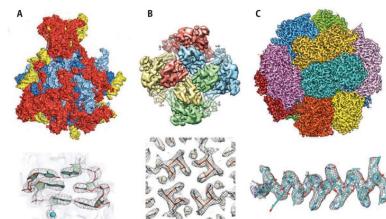
#### The Resolution Revolution

Werner Kühlbrandt

**P**recise knowledge of the structure of macromolecules in the cell is essential for understanding how they function. Structures of large macromolecules can now be obtained at near-atomic resolution by averaging thousands of electron microscope images recorded before radiation damage accumulates. This is what Amunts *et al.* have done in their article on page 1485 of this issue (1), reporting the structure of the large subunit of the mitochondrial ribosome at 3.2 Å resolution by electron cryo-microscopy (cryo-EM). Together with other recent high-resolution cryo-EM structures (2–4) (see the figure), this achievement heralds the beginning of a new era in molecular biology, where structures at near-atomic resolution are no longer the prerogative of x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

Ribosomes are ancient, massive protein-RNA complexes that translate the linear genetic code into three-dimensional proteins.

Advances in detector technology and image processing are yielding high-resolution electron cryo-microscopy structures of biomolecules.



Near-atomic resolution with cryo-EM. (A) The large ribosomal subunit of the yeast mitochondrial ribosome at 3.2 Å reported by Amunts *et al.* In the detailed view below, the base pair of an RNA double helix and a magnesium ion (blue) are clearly resolved. (B) TRPV1 ion channel at 3.4 Å (2), with a detailed view of residues lining the ion pore on the four-fold axis of the tetrameric channel. (C)  $F_{1\text{-}ATP}$ -reducing NifU(Fe) hydrogenase at 3.36 Å (3). The detail shows an α helix in the F1 subunit with resolved side chains. The maps are not drawn to scale.

#### Structure of the Yeast Mitochondrial Large Ribosomal Subunit

Alexey Amunts,<sup>1</sup> Alan Brown,<sup>2</sup> Xiao-chen Bai,<sup>2</sup> Jose L. Llúcar,<sup>2</sup> Tanweer Hussain, Paul Emsley,<sup>1</sup> Fei Long,<sup>1</sup> Garib Murshudov,<sup>1</sup> Sjors H. W. Scheres,<sup>1,†</sup> V. Ramakrishnan<sup>1</sup>

Mitochondria have specialized ribosomes that have diverged from their bacterial and cytoplasmic counterparts. We have solved the structure of the yeast mitochondrial large subunit using single-particle cryo-electron microscopy at a resolution of 3.2 Å. Our model is compatible with a previously proposed model to be built de novo and refined, including 39 proteins, 13 of which are unique to mitochondria, as well as expansion segments of mitochondrial RNA. The structure reveals a new exit tunnel path and architecture, unique elements of the E site, and a putative membrane docking site.

**M**itochondria are organelles in eukaryotic cells that play a major role in metabolism, especially the synthesis of adenosine triphosphate (ATP). During evolution, mitochondria have lost or transferred most of their genes to the nucleus, substantially reducing the size of their genome (1). In yeast, all but one of the few remaining protein-coding genes encode sub-

genome, whereas all but one of its ribosomal proteins are nuclear-encoded and imported from the cytoplasm. Mitochondria have diverged greatly from their counterparts in the cytosol of bacterial and eukaryotic cells, reflecting the unique variability depending on species (table S1) (2). Several genetic diseases map to mitochondria (3). In addition, the toxicity of many ribosomal antibiotics, in particular aminoglycosides, is thought to be due to their interaction with the ribosome (4).

Mitochondrial translation in the yeast *Saccharomyces cerevisiae* (6) has been used as a model to study human mitochondrial diseases (7). The 74S yeast mitochondrial ribosome has an overall molecular weight of 3 MD, some 30% greater than that of its bacterial counterpart. It consists of a 54S large subunit (1.9 MD) and a 37S small subunit (1.1 MD).

MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK.  
†These authors contributed equally to this work.

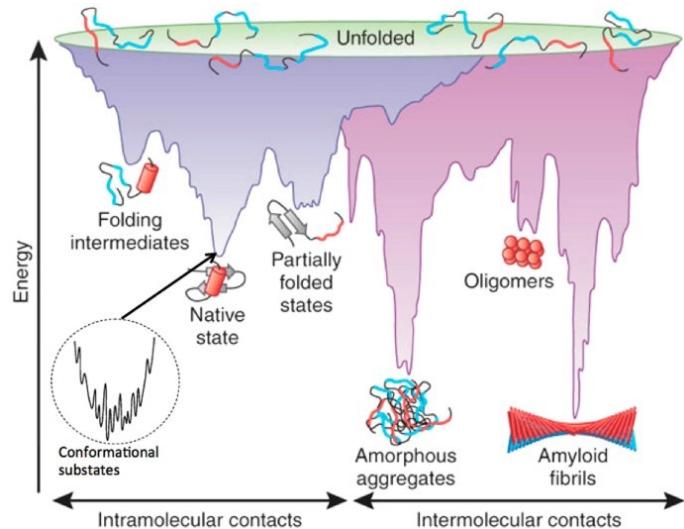
\*Corresponding author. E-mail: scheres@mrc-lmb.cam.ac.uk (S.H.W.S.); ramsak@mrc-lmb.cam.ac.uk (V.R.)



# Introduction: Proteins and Structure

## Levinthal's Paradox

Levinthal's paradox is that finding the **native folded state** of a protein by a **random search** among all possible configurations can take an enormously long time.



Raskatov JA, Teplow DB. Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins. Sci Rep. 2017 Oct 2;7(1):12433. doi: 10.1038/s41598-017-10525-5.

Extrait du Journal de Chimie Physique, 1968, 65 n° 1, p. 44.

ARE THERE PATHWAYS FOR PROTEIN FOLDING ?

by CYRUS LEVINTHAL.

[Massachusetts Institute of Technology, Department of Biology Cambridge, Massachusetts.]

Proc. Natl. Acad. Sci. USA  
Vol. 65, pp. 20-22, January 1968  
Biophysics

### Levinthal's paradox

ROBERT ZWANZIG, ATTILA SZABÓ, AND BIMAN BAGCHI\*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Building 2, National Institutes of Health, Bethesda, MD 20205

Contributed by Robert Zwanzig, October 7, 1967

**ABSTRACT** Levinthal's paradox is that finding the native folded state of a protein by a random search among all possible configurations can take an enormously long time. Yet proteins can fold in seconds or less. Mathematical analysis of a simple model shows that the probability of finding the native state against locally unfavorable configurations, of the order of a few  $10^{-7}$ , can reduce Levinthal's time to a biologically significant time to the fully correct state can be very much shorter. In fact, this time can become biologically significant.

### Model and Results

Since the goal is to try to understand the folding of any particular protein, but only represent an elementary result of Levinthal's paradox, precise details of the protein structure will be ignored. Consequently, the model to be treated is not expected to be directly useful in the theory of protein folding. It is also not intended to consider the kind of energetic effects that are known to be involved in folding a protein.

Lectures and articles dealing with protein folding dynamics often begin with a reference to the Levinthal "paradox" (1, 2). The main point of this paper is to show by mathematical analysis of a simple model that Levinthal's paradox becomes irrelevant when folding is considered in terms of the interactions between amino acids that take into account:

How long does it take for a protein to find its native structure? This is the central question of the Levinthal paradox: each bond connecting amino acids can have several (e.g., three possible) states, so the protein of, say,  $N = 10^3$  amino acids has  $3^{10^3} \approx 10^{300}$  states. Even if the protein is able to sample new configurations at the rate of  $10^6$  per second, or  $3 \times 10^6$  per day, it would take  $10^{24}$  years for them all. Levinthal concluded that random search is not an effective way of finding the correct state of a folded protein, because it is slow, and in a time scale of seconds or less, this is the paradox.

A clue to the resolution of the paradox is suggested by Dawkins (3). He considered the problem of the monkey hitting a sequence of small changes. He gave a more whimsical example of a similar paradox: how long will a random search take to produce a sonnet by William Shakespeare? The starting statement contains 28 characters, including 5 spaces; and there are 27 possible choices for each location, 26 letters and a space. The number of possible sequences is  $27^{28} \approx 10^{39}$ , which require about  $27^9 \approx 10^{27}$  key strokes. Dawkins observed that if the monkey cannot change those letters that are already correctly placed, it will never reach the final state, as predicted by a random search in only a few thousand key strokes.

In both examples, finding proteins or writing Hamlet, based on such a search is extremely slow. However, random searches, of course this is well known, in protein folding simulations, potential energy functions provide the necessary information to guide the search, as in molecular dynamics methods (5). However, these methods rely heavily on computer time and memory, and are not suitable for analysis. The goal of this paper is to provide the mathematical analysis of Levinthal's paradox for a highly simplified model of protein folding.

A first-passage time calculation shows that for an unbiased random search, Levinthal's protein folding estimate is essentially correct if  $k_0/k_1$  is not too small. The time  $\tau$  is essentially independent of the starting  $S$ ; even if the starting configuration is chosen to be a somewhat improbable protein, that it can wander further away before reaching  $S = 0$ . The mean first-passage time for a fully biased search, where the change  $c \rightarrow i$  is not allowed so that  $k_0 = 0$ , is

$$\tau(S) = (1/Nk_1)^{1/2} + A/\lambda k_1^2 \quad [1]$$

(The exact result is given later in Eq. 14.) This is approximately correct for large  $N$  if  $k_1$  is not too small. The time  $\tau$  is essentially independent of the starting  $S$ ; even if the starting configuration is chosen to be a somewhat improbable protein, that it can wander further away before reaching  $S = 0$ .

The mean first-passage time for a fully biased search, where the change  $c \rightarrow i$  is not allowed so that  $k_0 = 0$ , is

$$\tau(S) = (1/Nk_1)^{1/2} + A/\lambda k_1^2 \quad [1]$$

(The exact result is given later in Eq. 14.) This is approximately correct for large  $N$  if  $k_1$  is not too small. The time  $\tau$  is essentially independent of the starting  $S$ ; even if the starting configuration is chosen to be a somewhat improbable protein, that it can wander further away before reaching  $S = 0$ .

The mean first-passage time for a fully biased search, where the change  $c \rightarrow i$  is not allowed so that  $k_0 = 0$ , is

\*Permanent address: Indian Institute of Science, Bangalore, India.

†Present address: Department of Biochemistry, University of Oxford, United Kingdom.

Levinthal's paradox, in fact does not contain anything about it. Ref. 2, cited only a few times, contains Levinthal's estimate of folding times.



# Introduction: Proteins and Structure

## Anfinsen's Dogma

20 July 1973, Volume 181, Number 4096

SCIENCE

Folding problem encompasses two interrelated challenges:

- 1) understanding **the process** of protein chain folding
- 2) accurately predicting a protein's **final folded structure**

- In 1972 Christian Anfinsen shared the Nobel Prize in Chemistry for proposing that, in its standard physiological environment, **a protein's structure is determined by the sequence of amino acids that make it up**
- **Anfinsen's dogma implies that predicting the folded state of a protein does not necessarily require an understanding of the folding process**
- **Anfinsen's findings inspired a search for an efficient system based solely on its amino acid sequence. While challenging, this was at least theoretically possible.**

### Principles that Govern the Folding of Protein Chains

Christian B. Anfinsen

The telegram that I received from the Swedish Royal Academy of Sciences specifically cited "...studies on ribonuclease, in particular the relationship between the amino acid sequence and the biologically active conformation..." The work that my colleagues and I have carried out on the nature of the process that controls the folding of polypeptide chains into the unique three-dimensional structures of proteins was, indeed, strongly influenced by observations on the ribonuclease molecule. Many others, including Anson and Mirsky (1) in the 1930's and O'Farrell and Eysel (2) in the 1950's had observed and discussed the reversibility of denaturation of proteins. However, the true elegance of this consequence of natural selection was dramatized by the ribonuclease work, since the refolding of this molecule, after full denaturation by reductive cleavage of its four disulfide bonds (Fig. 1), required that only 1 of the 105 possible pairings of

eight sulfhydryl groups to form four disulfide linkages to form the native structure. The original observations that led to this conclusion were made together with my colleagues Michael Sela and Fred White in 1956-1957 (3). These were, in actuality, the beginnings of a long series of studies that rather vaguely aimed at the eventual total synthesis of the protein. As we all know, Gutfreund and Merrifield (4) at the Rockefeller Institute, and Ralph Hirschman and his colleagues at the Merck Research Institute (5), have now accomplished this monumental task.

The studies on the renaturation of fully denatured ribonuclease required many supporting investigations (6-8) to establish, finally, the generality which we have occasionally called (9) the "thermodynamic hypothesis." This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment. In terms of natural selection through the "design" of macromolecules during evolution, this idea emphasized the fact that a protein molecule only exists under conditions similar to those for which it was selected—the so-called physiological state.

After several years of study on the ribonuclease molecule it became clear to us, and to many others in the field of protein conformation, that proteins devoid of restrictive disulfide bonds or other covalent cross-linkages would make more convenient models for the study of the thermodynamic and kinetic aspects of the nucleation, and subsequent pathways, of polypeptide chain folding. Much of what will be reviewed with studies on the flexible and somewhat stippled-coated nucleic molecule, but I will first summarize some of the older background experiments on bovine pancreatic ribonuclease itself.

#### Support for the "Thermodynamic Hypothesis"

An experiment that gave us a particularly satisfaction in connection with the translation of information in the linear amino acid sequence into native conformation involved the rearrangement of so-called "scrambled" ribonuclease (8). When the fully reduced protein, with eight SH groups, is allowed to reoxidize under denaturing conditions such as exist in a solution of 8 molar urea, a mixture of products is obtained containing many or all of the possible 10<sup>5</sup> isomeric disulfide bonded forms (schematically shown in Fig. 2). A mixture having on the order of 1 percent of native 22-kilodalton native enzyme is essentially inactive.

Copyright © 1973 by the Nobel Foundation.

The author is chief of the Laboratory of Chemical Biology, National Institute of Arthritis,

Metabolism, and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20204. This article is the lecture he delivered in Stockholm, Sweden, on December 8, 1972, when he received the Nobel Prize for Chemistry, a prize he shared with Robert W. Williams and Stanford Moore.

This article is published here with the permission of the Nobel Foundation and will also be included in the proceedings of the meeting held in Stockholm on December 8, 1972, and published in 1973,

as well as in the series Nobel Lectures (in English) published by the Royal Publishing Company, Amsterdam and New York. Dr. Moore's and Dr. Stein's combined lecture appeared as a

single article in the 4 May issue of *Science*, page 458.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.

Editorial review by R. W. Williams.

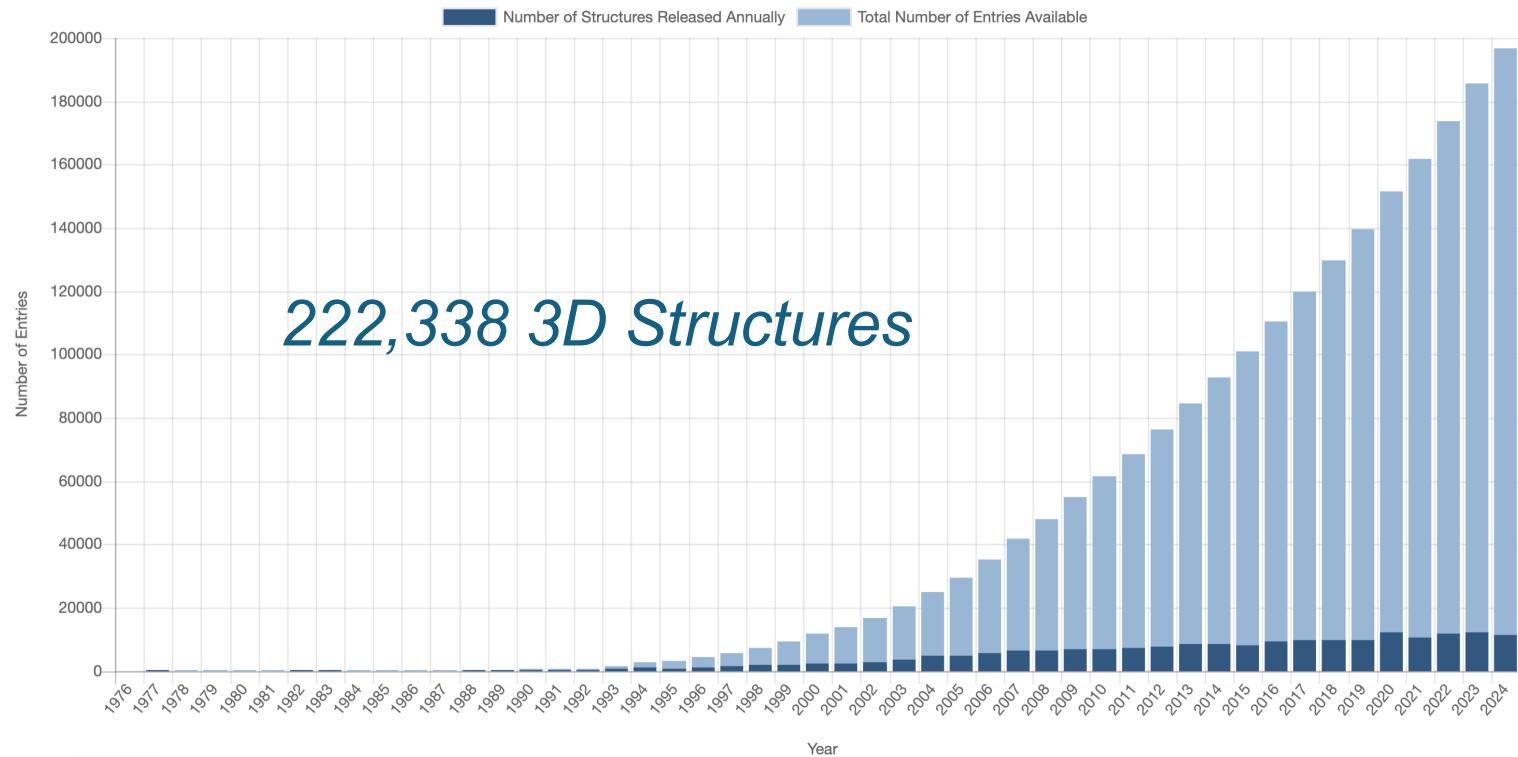
Published July 20, 1973.

Copyright © 1973 by the American Association for the Advancement of Science.

Received June 1, 1973; revised June 12, 1973.</p



## Predicting the Protein Structure: The Data



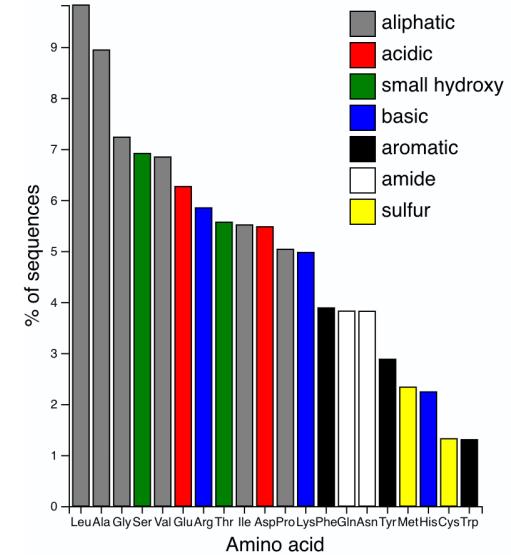
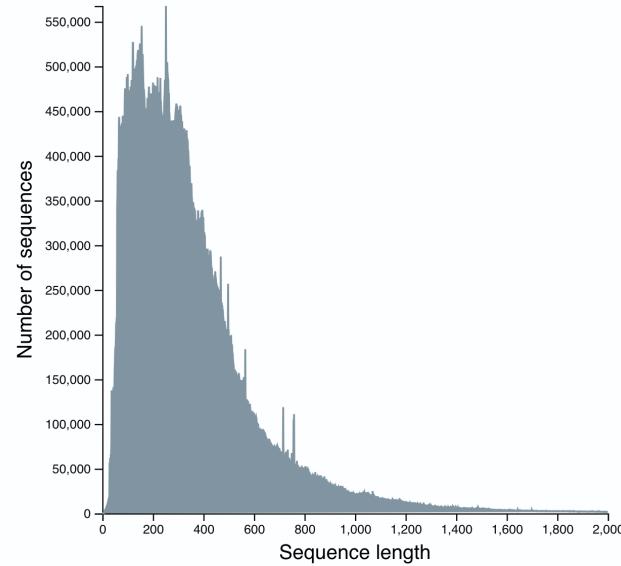
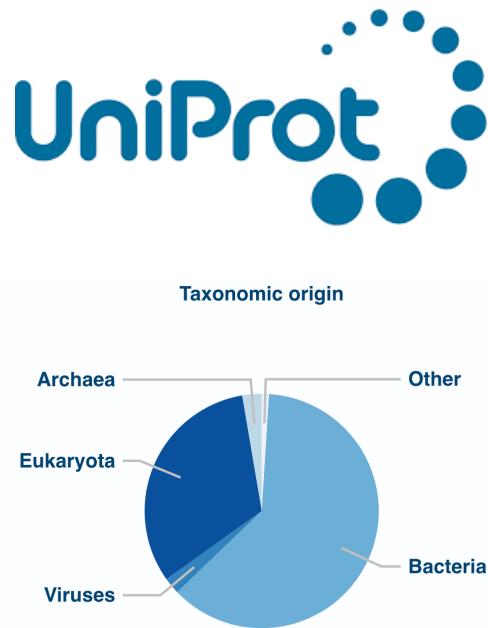
222,338 3D Structures

Data accessed on 26/11/24



## Predicting the Protein Structure: The Data

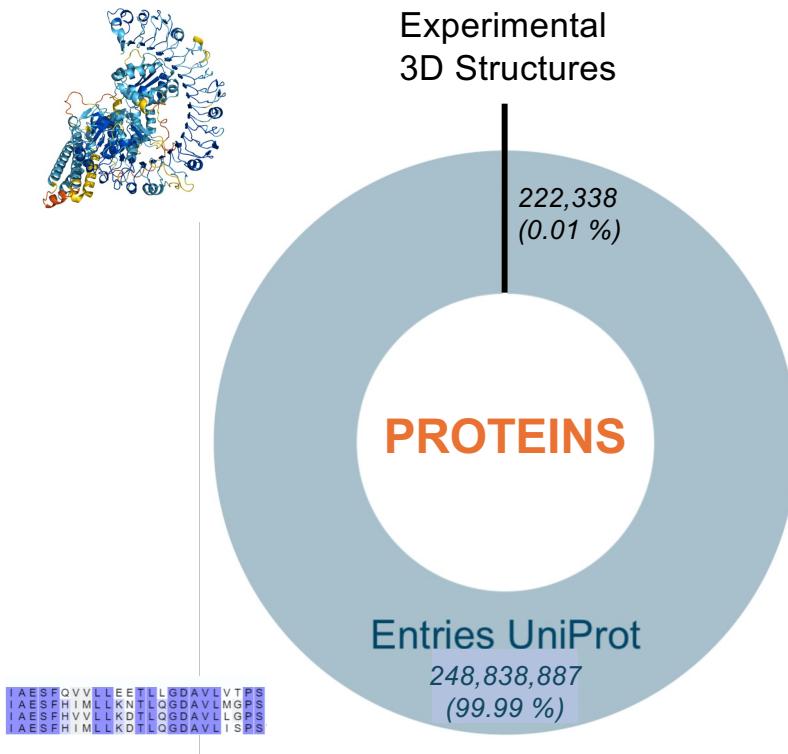
248,838,887 Sequences



Data accessed on 26/11/24



## Predicting the Protein Structure: The Data



The **gap** has widened considerably in recent years (advances in **DNA sequencing**)

**Experimental** structure determination can not keep up with the pace of sequencing

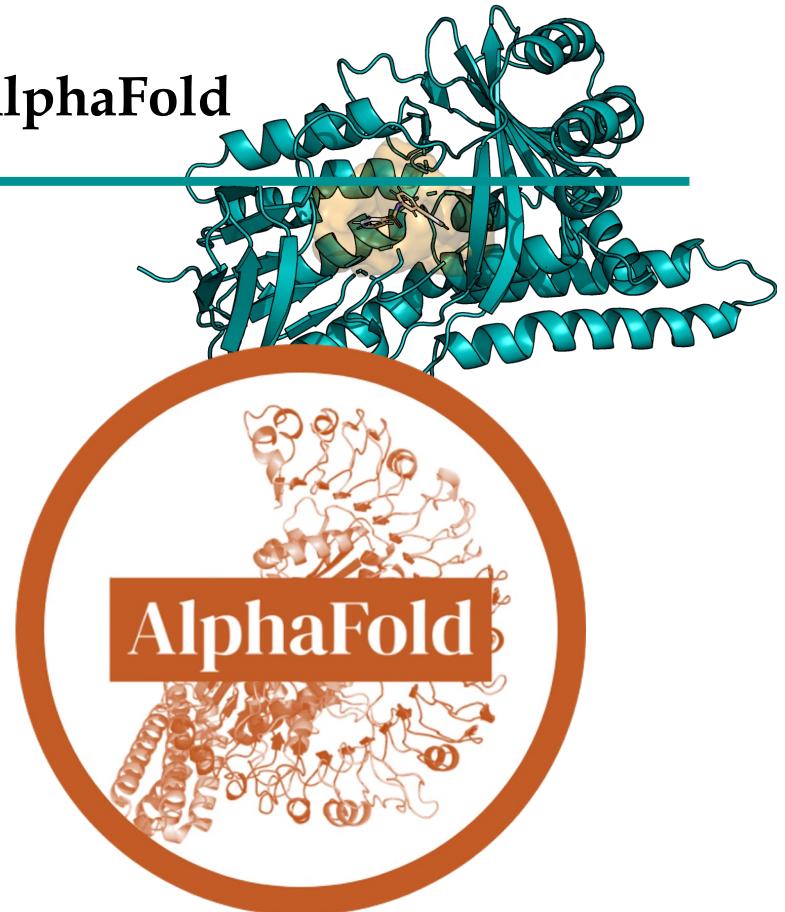
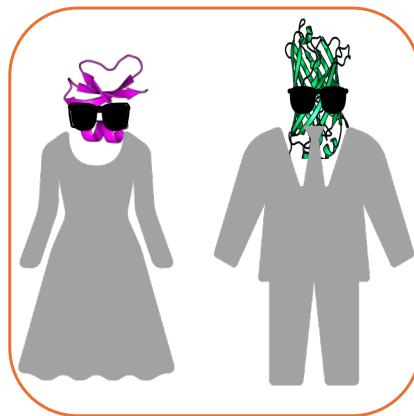
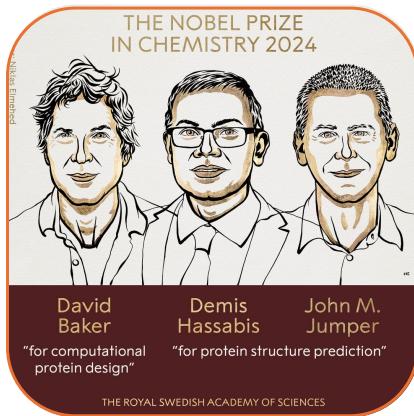
There have been **hundreds of millions** of known proteins with **unknown structures**.

Data accessed on 26/11/24



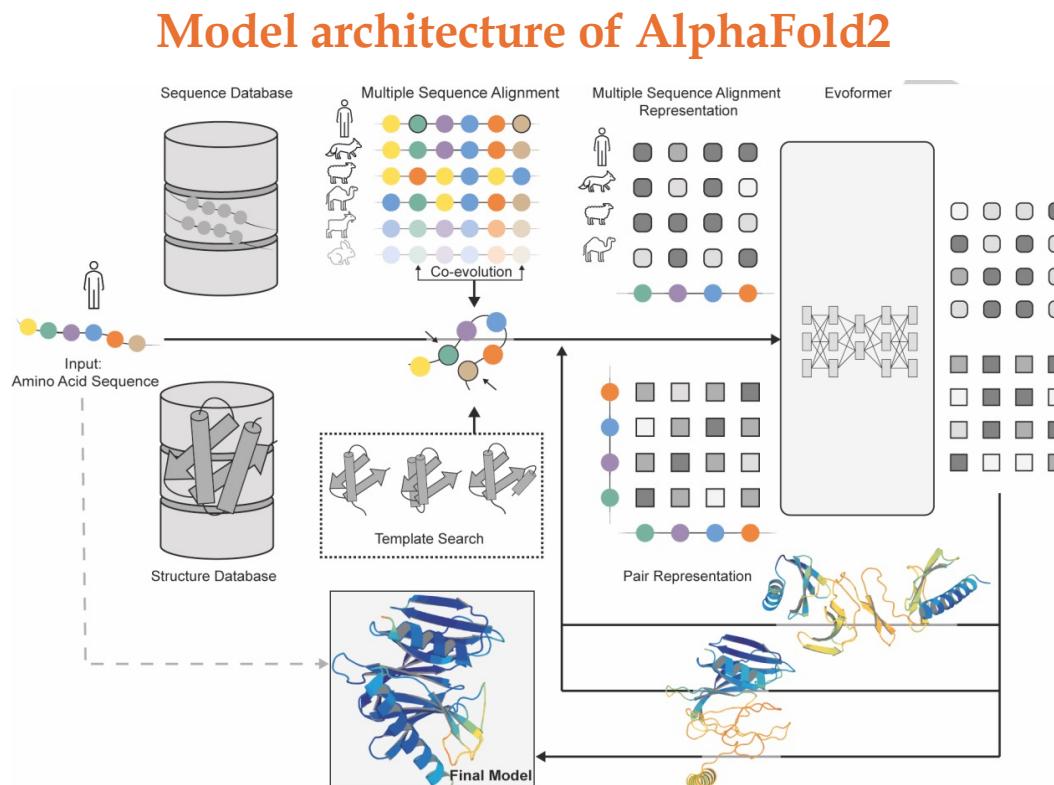
## Predicting the Protein Structure: AlphaFold

### Proteins in the Hall of Fame





# Predicting the Protein Structure: AlphaFold



pH Effects?



# Predicting the Protein Structure: pH

BBA - Biomembranes 1861 (2019) 182984  
 Contents lists available at ScienceDirect  
**BBA - Biomembranes**  
 journal homepage: [www.elsevier.com/locate/bbamem](http://www.elsevier.com/locate/bbamem)

## Activity and characterization of a pH-sensitive antimicrobial peptide

Morgan A. Hitchner<sup>a,1</sup>, Luis E. Santiago-Ortiz<sup>a,1</sup>, Matthew R. Necelis<sup>a</sup>, David J. Shirley<sup>a</sup>, Thaddeus J. Palmer<sup>a</sup>, Katharine E. Tarnawsky<sup>a</sup>, Timothy D. Vaden<sup>a</sup>, Gregory A. Caputo<sup>a,b,\*</sup>

<sup>a</sup>Department of Chemistry and Biochemistry, Rowan University, 201 Mullica Hill Road, Glassboro, NJ 08028, United States of America

<sup>b</sup>Department of Molecular and Cellular Biosciences, Rowan University, 201 Mullica Hill Road, Glassboro, NJ 08028, United States of America



Computational and Structural Biotechnology Journal 20 (2022) 1439–1455  
 **COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL**  
 journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

Investigation of the pH-dependent aggregation mechanisms of GCSF using low resolution protein characterization techniques and advanced molecular dynamics simulations

Suk Kyu Ko <sup>a,\*,1</sup>, Carolin Berner <sup>b</sup>, Alina Kulakova <sup>c</sup>, Markus Schneider <sup>d</sup>, Iris Antes <sup>d,1</sup>, Gerhard Winter <sup>b</sup>, Pernille Harris <sup>c</sup>, Günther H.J. Peters <sup>a,\*</sup>

<sup>a</sup>Technical University of Denmark, Department of Chemistry, 2800 Kongens Lyngby, Denmark

<sup>b</sup>Ludwig-Maximilian University of Munich, Department of Pharmacy, 81377 Munich, Germany

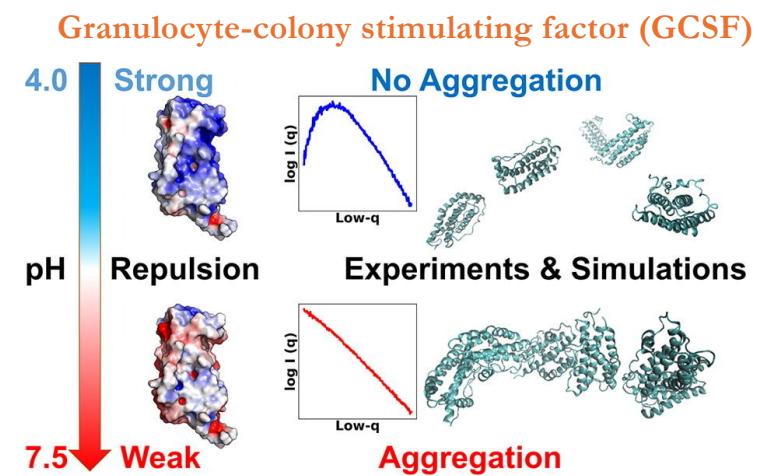
<sup>c</sup>University of Copenhagen, Department of Chemistry, 2100 Copenhagen, Denmark

<sup>d</sup>Technical University of Munich, TUM School of Life Sciences, 85354 Freising, Germany



**Table 1**  
 Antibacterial activity ( $\mu\text{M}$ ).

pH	<i>S. aureus</i>		<i>P. aeruginosa</i>		<i>A. baumannii</i>		<i>E. coli</i>	
	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC
5	0.47	0.94	3.75	3.75	0.23	0.94	0.47	0.94
6	15	15	7.5	7.5	1.88	3.75	1.88	1.88
7	15	> 15	> 15	> 15	7.5	7.5	3.75	7.5
8	15	> 15	> 15	> 15	> 15	> 15	15	15



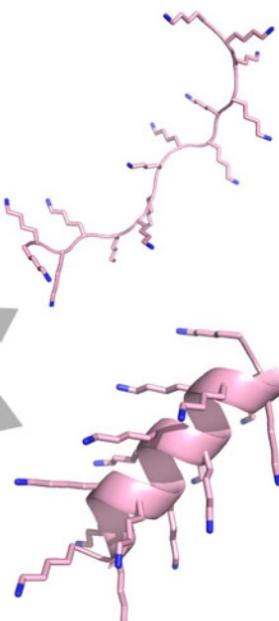


## Predicting the Protein Structure: PEP-FOLD4

**PEP-FOLD4**

**AMINO ACID  
SEQUENCE**

**pH  
VARIATION**



Prediction of peptides of less 40 amino acids in aqueous solution

Step beyond many machine-learning approaches, such as AlphaFold2, TrRosetta and RaptorX

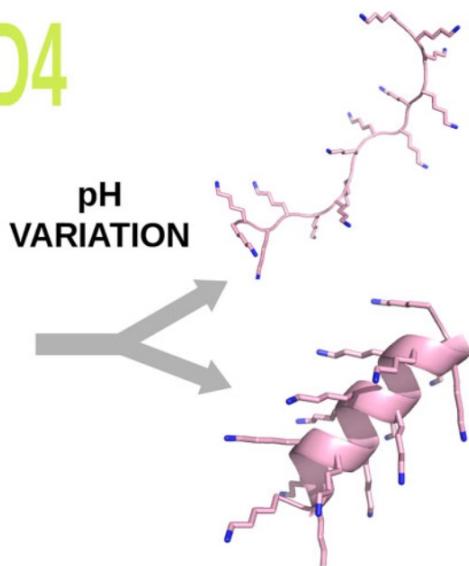
Debye-Hueckel formalism for charged-charged side chain interactions



## Predicting the Protein Structure: PEP-FOLD4

### PEP-FOLD4

AMINO ACID  
SEQUENCE



#### sOPEPforcefield

$$E = E_{local} + E_{nonbonded} + E_{H-bond} \quad (1)$$

$$E_{nonbonded} = E_{Mie} + E_{DH} \quad (2)$$

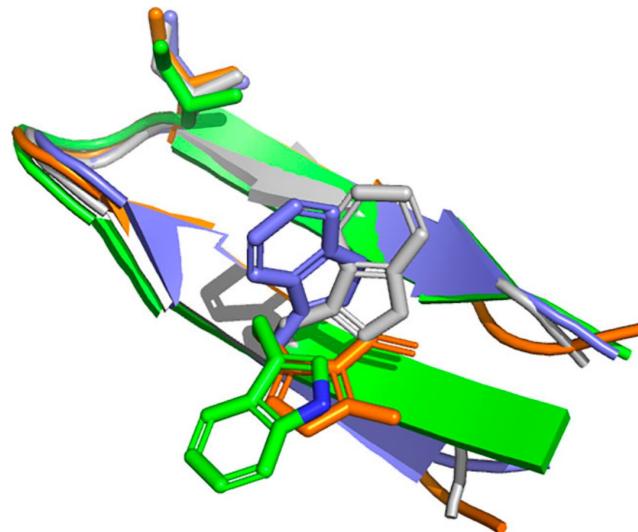
Mie formulation for non-bonded interactions.

$$E_{DH_{ij}} = (q_i * q_j * e^{-r_{ij}/l_{DH}}) / (\epsilon(r_{ij}) * r_{ij})$$

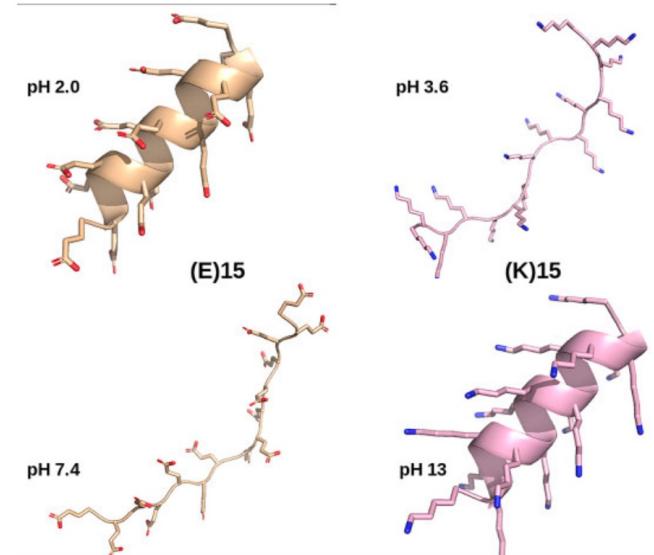
To take into account charge variation and pH dependence



## Predicting the Protein Structure: pH



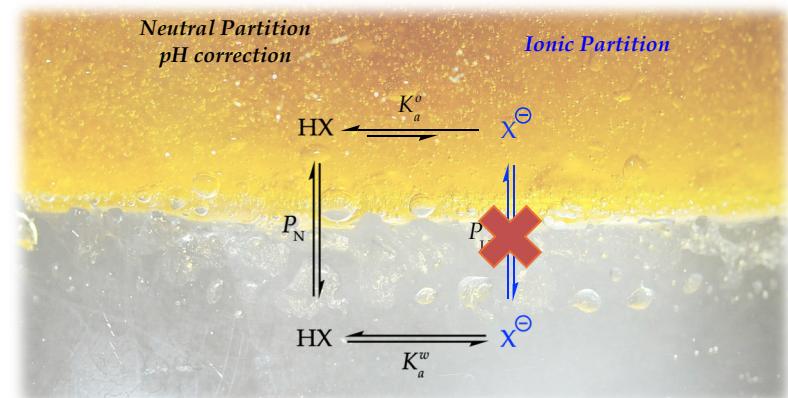
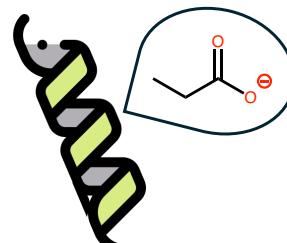
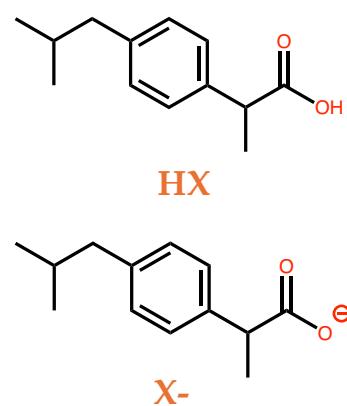
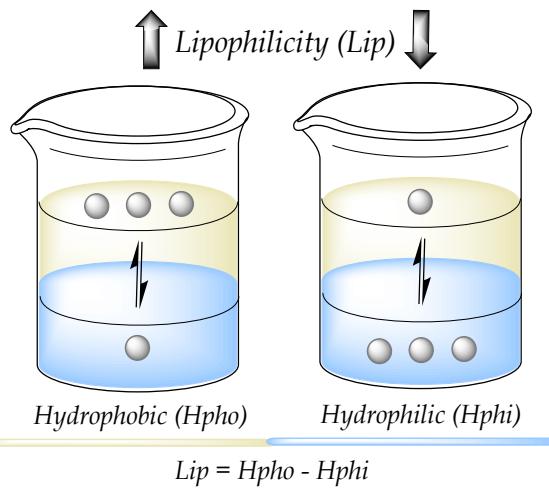
**Figure 2.** Structure predictions of the 14-residue peptide RGKWTYN-GITYEGR (PDB: 1j4m). Gray: NMR, green: PEP-FOLD4, orange: TrRosetta, blue: AlphaFold2. Side chains at two positions are depicted to show the structural agreement of the models.



**Figure 3.** pH-dependent conformations of (E)15 and (K)15 peptides by PEP-FOLD4.

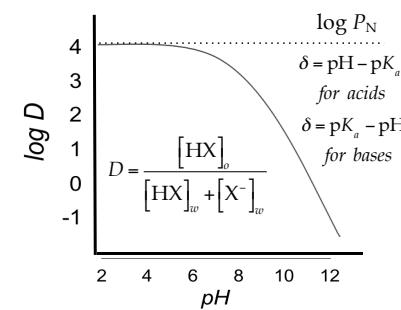


## Predicting Protein Properties: Lipophilicity



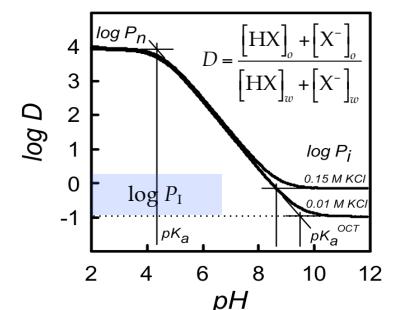
Model 1. pH Correction

$$\log D = \log \underline{P_N} - \log(1 + 10^\delta)$$



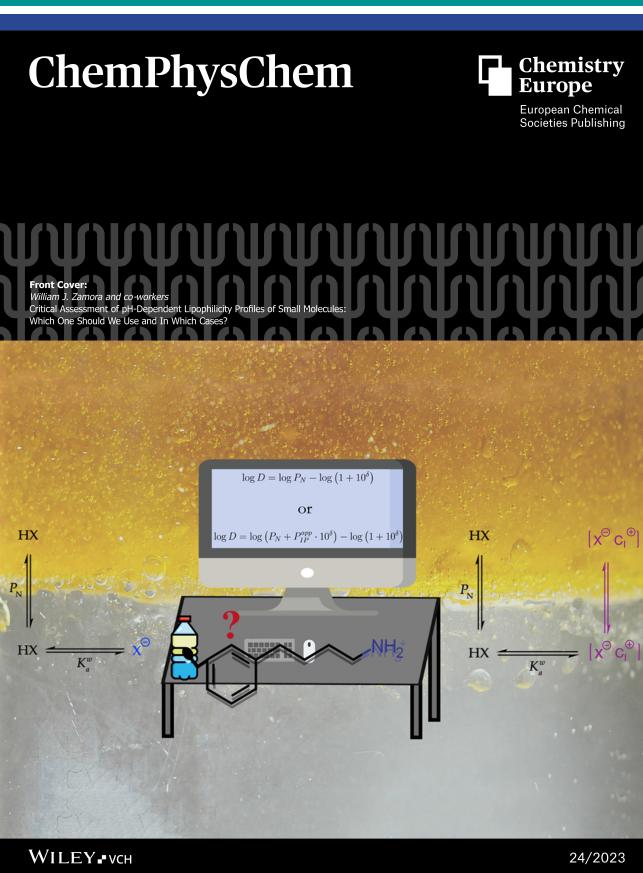
Model 2. Ionic Partition

$$\log D = \log \underline{(P_N + P_I \cdot 10^\delta)} - \log(1 + 10^\delta)$$





# Predicting Protein Properties: Lipophilicity

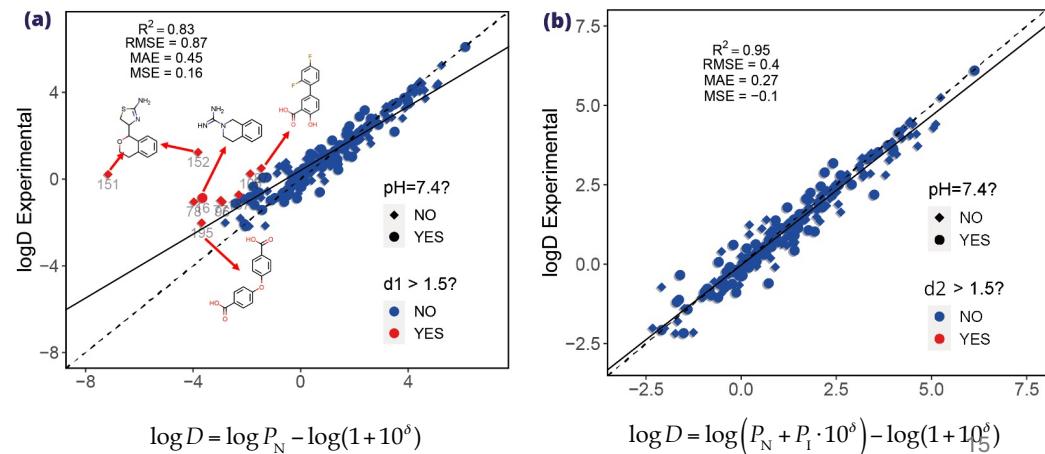


Research Article

## Critical Assessment of pH-Dependent Lipophilicity Profiles of Small Molecules: Which One Should We Use and In Which Cases?

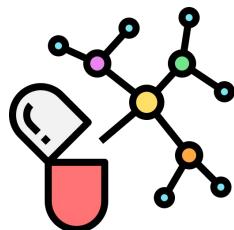
Esteban Bertsch, Sebastián Suñer, Dr. Silvana Pinheiro, Prof. Dr. William J. Zamora

First published: 03 October 2023 | <https://doi.org/10.1002/cphc.202300548>





## Predicting Protein Properties: Lipophilicity



**Small molecules**  
MW < 500 Da

### ○ Drug Development

Acceptable level of in vivo clearance for a drug is logD of 0–3

$$\log(\text{Sol}_{\text{aq}}) = 0.5 - \log P - 0.01(\text{MP}-25)$$

### *The Rule of 5*

1. 5 H-bond donors
2. 10 H-bond acceptors
3. Molecular weight < 500 Da
4.  $\text{Log } P < 5$ .

Acceptable permeability log D of 1–3.5

LipE should be prioritized based on the importance of **enthalpic optimization**

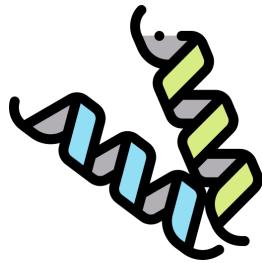
$$\text{LipE} = \text{LLE} = -\log(\text{potency}) - \log D$$

to achieve **low dose** and **dosing frequency**



## Predicting Protein Properties: Lipophilicity

### Beyond the Rule of 5 (bRo5) Chemical Space

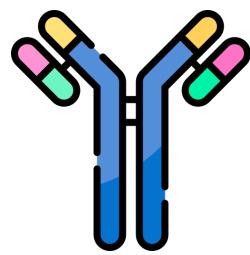


Peptides

MW ~ 500–5000 Da

Rapid clearance, short half-life

Low membrane permeability  
(chemically unmodified peptides)



Therapeutic proteins

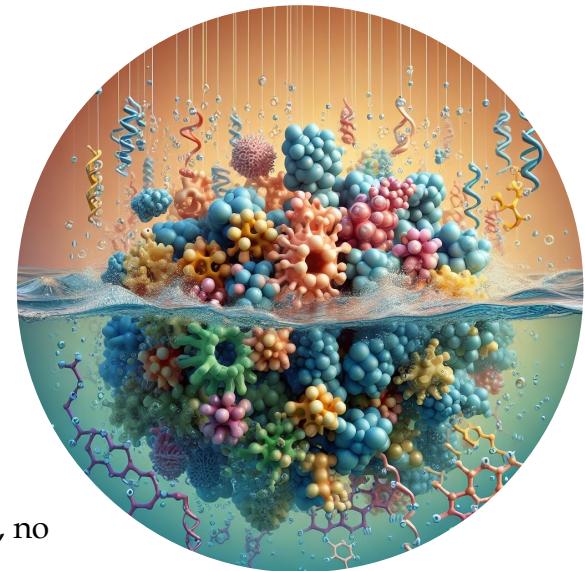
MW > 5000 Da

Complex (often recombinant) production, no  
easy chemical modification

Immunogenicity

Low membrane permeability due to size

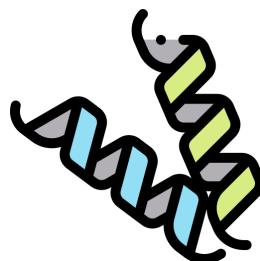
Only extracellular or surface-exposed targets



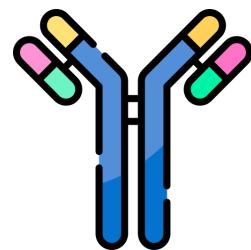
**No clear guidelines for  
rational drug design!!!**



## Predicting Protein Properties: Lipophilicity

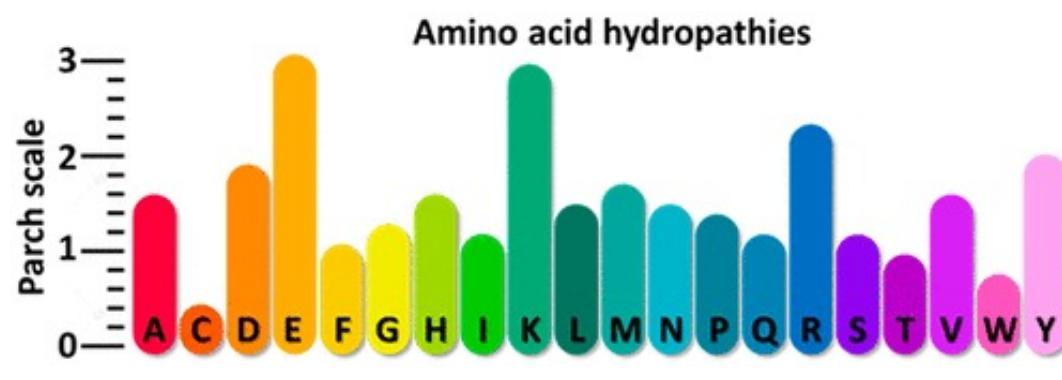


Peptides  
MW ~ 500-5000 Da



Therapeutic proteins  
MW > 5000 Da

Computational



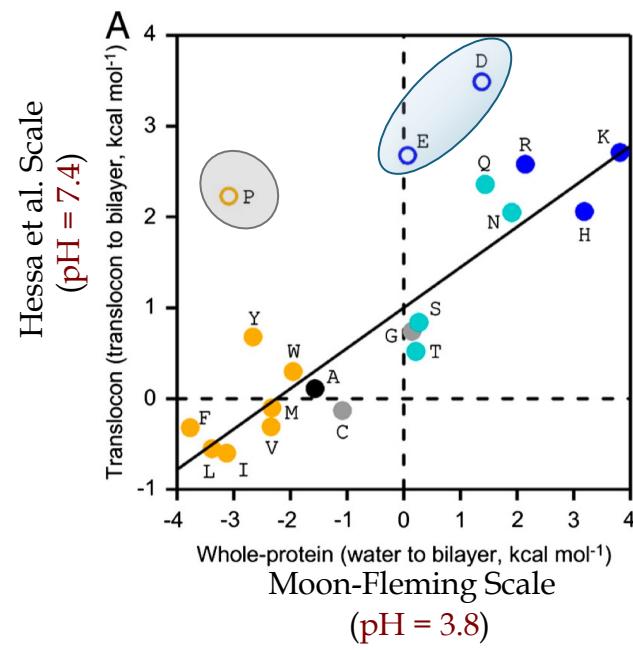
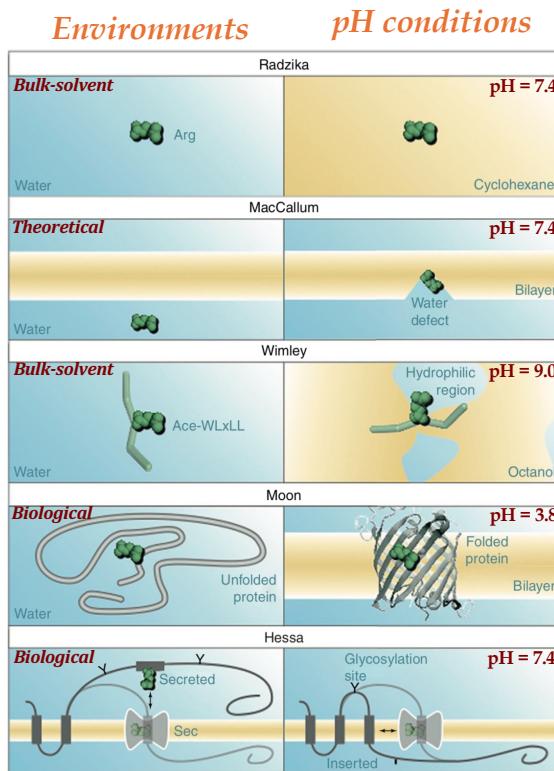
AKLDR  
RDLKA

LDRKA

Same logP or logD values!!!



## Predicting Protein Properties: Lipophilicity





# Predicting Protein Properties: Lipophilicity

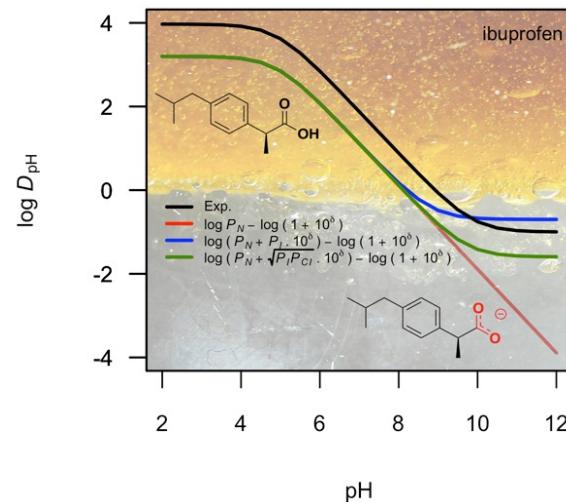
THE JOURNAL OF  
PHYSICAL CHEMISTRY B

Article  
pubs.acs.org/JPCB

## Prediction of pH-Dependent Hydrophobic Profiles of Small Molecules from Miertsu–Scrocco–Tomasi Continuum Solvation Calculations

Published as part of The Journal of Physical Chemistry virtual special issue "Manuel Yáñez and Otilia Mó Festschrift".

William J. Zamora,<sup>†,‡</sup> Carles Curutchet,<sup>‡</sup> Josep M. Campanera,<sup>\*§,¶</sup> and F. Javier Luque<sup>\*†</sup>



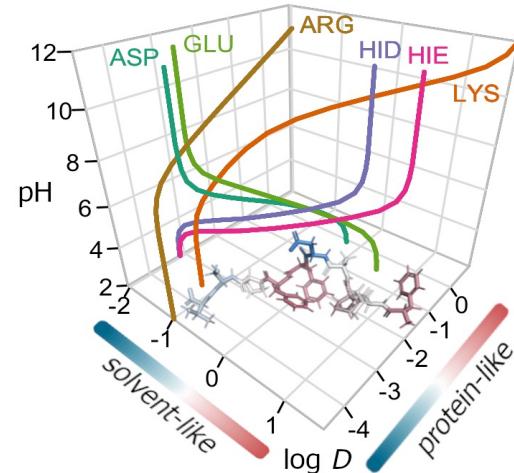
THE JOURNAL OF  
PHYSICAL CHEMISTRY  
Letters

Cite This: *J. Phys. Chem. Lett.* 2019, 10, 883–889

Letter  
pubs.acs.org/JPCL

## Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations

William J. Zamora,<sup>¶</sup> Josep M. Campanera,<sup>\*§</sup> and F. Javier Luque<sup>\*¶</sup>



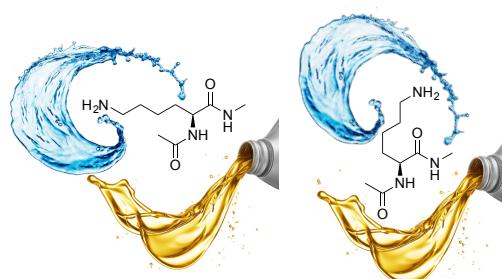


# Predicting Protein Properties: Lipophilicity

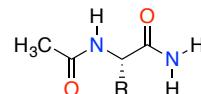
**Table 5.** Number of Conformers Considered for each Residue with a Population  $\geq 5\%$ .

Residue	Total no. Conformers	% Conformational Space Covered
ALA	8	99.9
ARG	26	42.0
ARN	26	42.0
ASN	21	93.9
ASP	17	97.2
ASH	17	97.2
CYS	8	99.9
CYX	8	99.9
GLN	23	72.3
GLU	28	81.7
GLH	28	81.7
GLY	8	99.9
HIP	18	89.6
HID	18	89.6
HIE	18	89.6
ILE	9	96.0
LEU	10	94.3
LYS	27	52.5
LYN	27	52.5
MET	28	80.3
PHE	12	91.5
PRO	8	97.9
SER	9	99.0
THR	7	97.5
TRP	24	84.1
TYR	14	96.2
TYN	14	96.2
VAL	7	98.9
<b>Total</b>	<b>572</b>	

## Development of the Scale



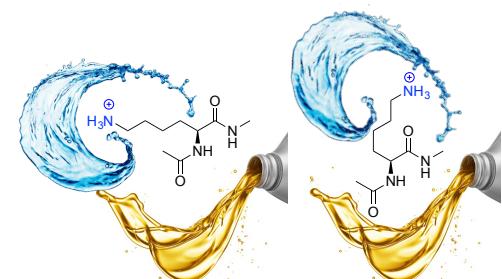
IEFPCM-MST  
SMD



$$\log D = \log(P_N + P_I \cdot 10^\delta) - \log(1 + 10^\delta) \quad (9)$$

## Solvent-like Scale (SolvL)

Boltzmann's weighting scheme



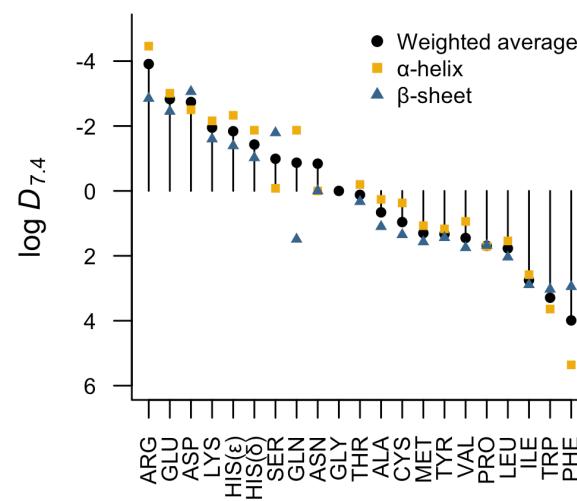
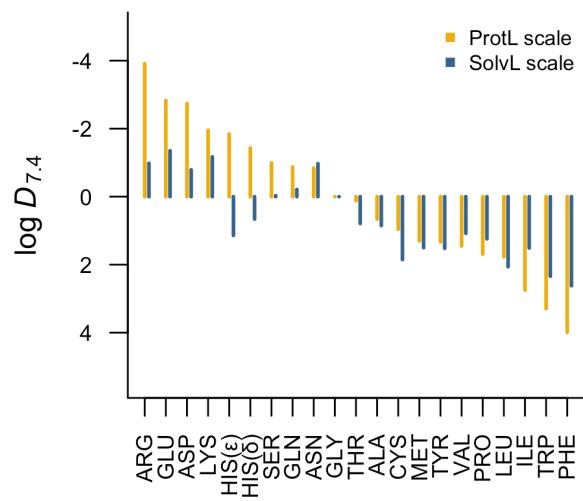
## Protein-like Scale (ProtL)

Backbone-dependent conformational library of Dunbrack et al. weights



## Predicting Protein Properties: Lipophilicity

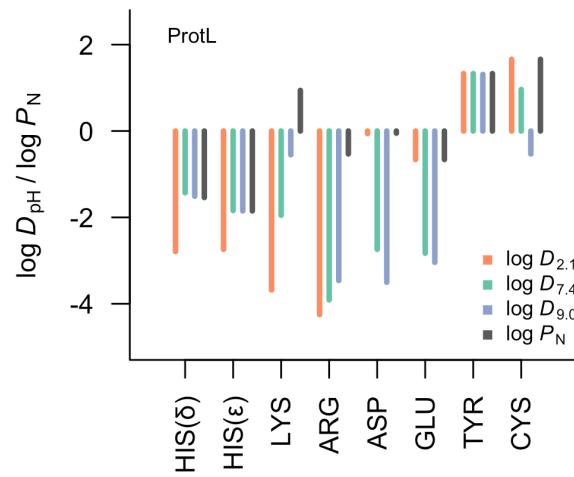
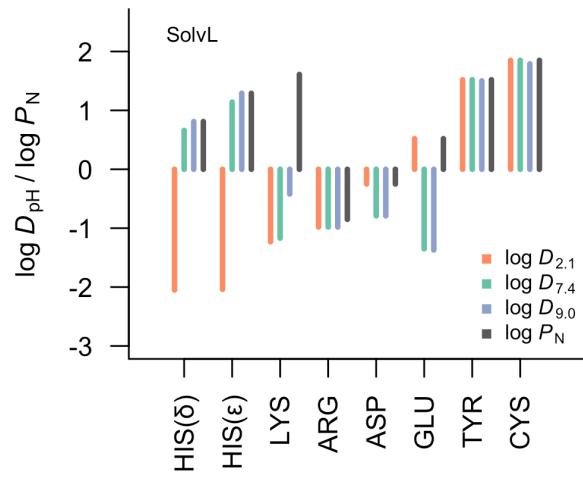
*SolvL & ProtL Schemes: Environment Dependence*





## Predicting Protein Properties: Lipophilicity

*SolvL & ProtL Schemes: pH Dependence*



$$\log D = \log(P_N + P_I \cdot 10^\delta) - \log(1 + 10^\delta)$$



# Predicting Protein Properties: Lipophilicity

Solv

EUR. J. MED. CHEM. — CHIM. THER., 1983-18, N° 4, pp. 369-375

## Hydrophobic parameters $\pi$ of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides

Jean-Luc FAUCHÈRE (\*) and Vladimir PLIŠKA

Departments of Molecular Biology and Biophysics and of Animal Production, Swiss Federal Institute of Technology, ETH, CH 8093 Zürich, Switzerland.

(Received January 12, 1983, accepted March 28, 1983)

**Summary.** — A hydrophobic parameter for the side chain of each of the twenty naturally occurring amino-acids was estimated quantitatively by partitioning of their N-acetyl-amino-acid amides ( $\text{CH}_3\text{CO}-\text{NH}-\text{CH}(\text{R})-\text{CONH}_2$ , R = variable side chain) in octanol/water. Suitable forms of the derivatives were synthesized chemically where necessary. The new series affords reliable distribution coefficients especially for polar amino-acids. A set of Hansch side chain parameters  $\pi$  is obtained for the twenty common amino-acids, for half-cystine as well as for the extremely lipophilic amino-acids carboranylalanine and adamantlylalanine. It is found that these hydrophobicity indices, obtained at neutral pH, are tightly correlated with the degree of solvent exposure of the side chains in globular proteins and it is proposed that they be used in future biophysical and pharmacological studies.

**Résumé.** — Un paramètre d'hydrophobie des chaînes latérales des 20 acides aminés naturels a été déterminé par mesure de la distribution entre l'octanol et l'eau de leurs dérivés amidés, N-acetylés. Plusieurs de ces composés ont été synthétisés et caractérisés. Cette nouvelle série a produit des coefficients de distribution fiables, notamment pour les acides aminés polaires. La constante d'hydrophobie  $\pi$  de Hansch a été ainsi obtenue pour les chaînes latérales des 20 acides aminés naturels, de l'hémicystine ainsi que des deux composés extrêmement lipophiles, l'adamantlylalanine et la carboranylalanine. Ces indices d'hydrophobie sont étroitement corrélés avec le degré d'exposition au solvant des chaînes latérales dans les protéines globulaires et de nature à être utilisées avec profit dans de futures études biophysiques et pharmacologiques.

**Zusammenfassung.** — Ein hydrophober Parameter für die Seitenkette der zwanzig natürlich vorkommenden Aminosäuren wurde bestimmt durch Messung des Verteilungskoeffizienten in Octanol/Wasser der N-acetylierten Aminosäuren Amide. Die nicht erhältlichen Derivate wurden synthetisiert und charakterisiert. Die neue Serie lieferte zuverlässige Verteilungskoeffizienten, auch für polare Aminosäuren. Die Hansch Hydrophobie Konstante  $\pi$  wurde auf diese Weise für die Seitenketten der 20 natürlichen Aminosäuren sowie für diejenigen von Hemicystein und von den extrem lipophilen Adamantlylalanin und Carboranylalanin erhalten. Es konnte gezeigt werden, daß diese Parameter mit dem Ausmass der Lösungsmittelkontakte mit den Seitenketten in globulären Proteinen eng korrelieren uns es wird vorgeschlagen, daß sie in zukünftigen biophysikalischen und pharmakologischen Studien herangezogen werden.

**Key words :** Amino-acid hydrophobicity. — Partition coefficients. — N-acetyl-amino-acid amides.

Calculated log  $D_{7.4}$

Calculated log  $D_{7.4}$

Pro

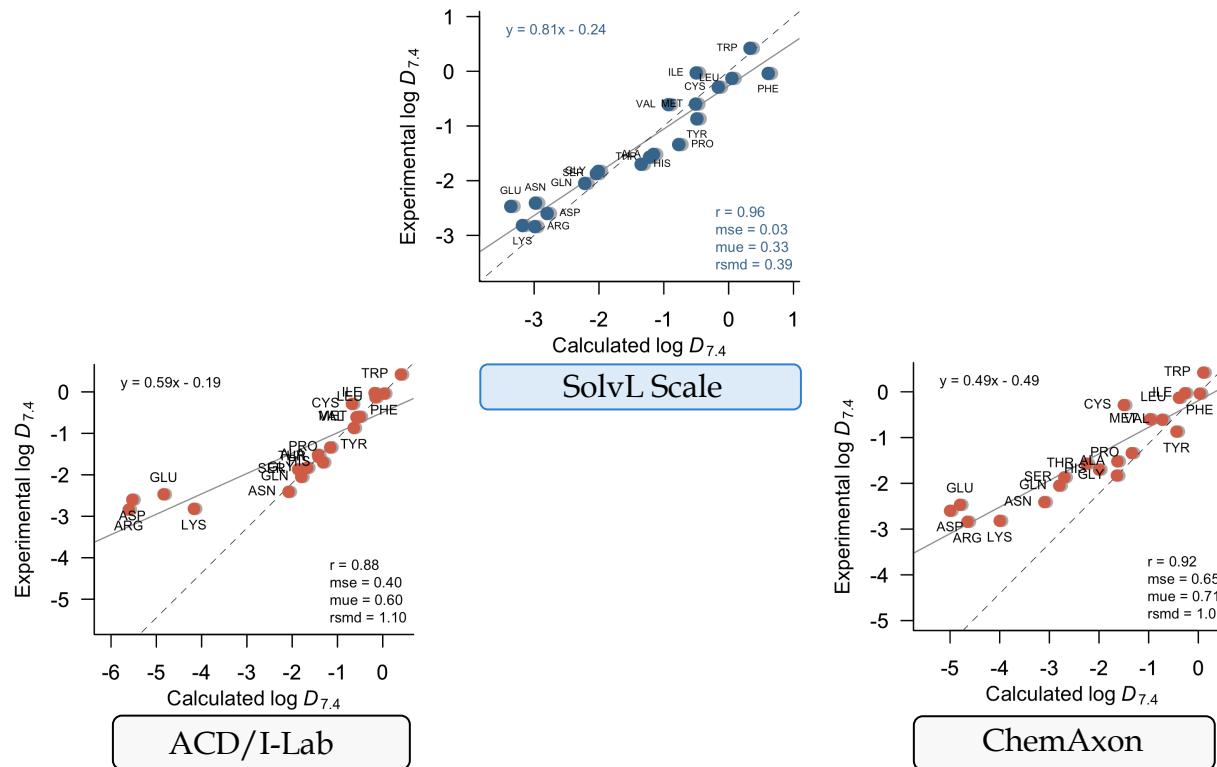
1  
0.42  
.83  
1.07

5



## Predicting Protein Properties: Lipophilicity

*Validation: Reproduction of Fauchère-Pliska Experimental Scale*

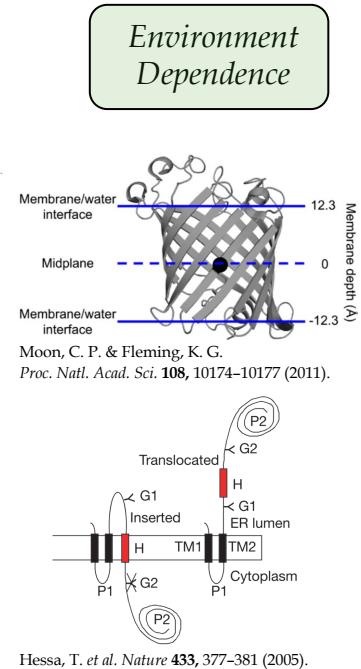




# Predicting Protein Properties: Lipophilicity

## Validation: Comparison with Other Scales

**Table 6.** Statistical Parameters of the Comparison of the SolvL and ProtL Scales with other Lipophilicity Scales. Comparison Was Made Using the Values Adapted to the Specific pH of each Scale and Relative to Gly.



Scale <sup>a</sup>	SolvL				ProtL			
	mse	mue	rsmd	<i>r</i> p-value	mse	mue	rsmd	<i>r</i> p-value
<i>Bulk-Solvent Adapted Scale</i>								
Fauchère-Pliska	-0.20	0.36	0.46	<b>0.94</b> $2 \times 10^{-10}$	0.36	0.98	1.28	<b>0.92</b> $6 \times 10^{-9}$
Eisenberg-McLachlan	-0.20	0.44	0.57	<b>0.90</b> $3 \times 10^{-8}$	0.36	1.08	1.35	<b>0.91</b> $2 \times 10^{-8}$
Hopp-Woods	-0.49	0.60	0.74	<b>0.91</b> $2 \times 10^{-8}$	0.07	0.84	1.08	<b>0.89</b> $9 \times 10^{-8}$
Wimley et al.	-0.60	1.02	1.16	<b>0.59</b> 0.006	0.04	1.24	1.64	<b>0.61</b> 0.004
	-0.87 <sup>b</sup>	0.92	1.03	<b>0.87</b> $2 \times 10^{-6}$	-0.30	1.03	1.25	<b>0.87</b> $2 \times 10^{-6}$
<i>Biological-Based Scale</i>								
Moon-Fleming	-0.12	0.57	0.67	<b>0.94</b> $4 \times 10^{-10}$	0.24	0.72	0.93	<b>0.91</b> $7 \times 10^{-9}$
Hessa et al.	-0.92	0.93	1.18	<b>0.79</b> $3 \times 10^{-5}$	-0.36	1.08	1.46	<b>0.82</b> $6 \times 10^{-6}$
<i>Knowledge-Based Scale</i>								
Koehler et al.	-0.91	1.10	1.33	<b>0.78</b> $4 \times 10^{-5}$	-0.35	1.55	1.87	<b>0.80</b> $2 \times 10^{-5}$
Janin et al.	-1.06	1.11	1.32	<b>0.78</b> $3 \times 10^{-5}$	-0.51	1.36	1.71	<b>0.74</b> $2 \times 10^{-4}$
<i>Consensus Scale</i>								
Kyte-Doolittle	-0.81	1.43	1.71	<b>0.72</b> $3 \times 10^{-4}$	-0.25	1.13	1.41	<b>0.78</b> $3 \times 10^{-5}$

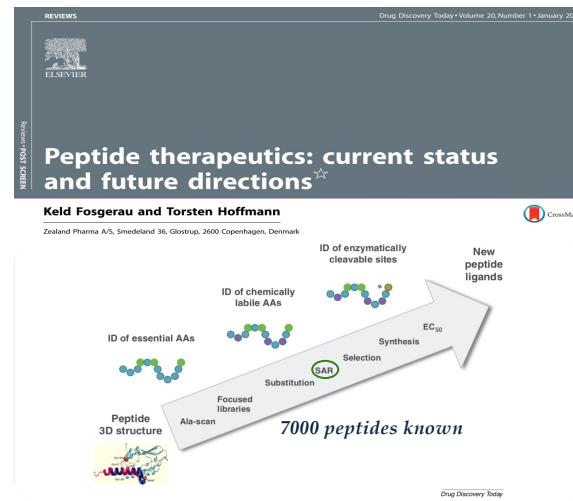
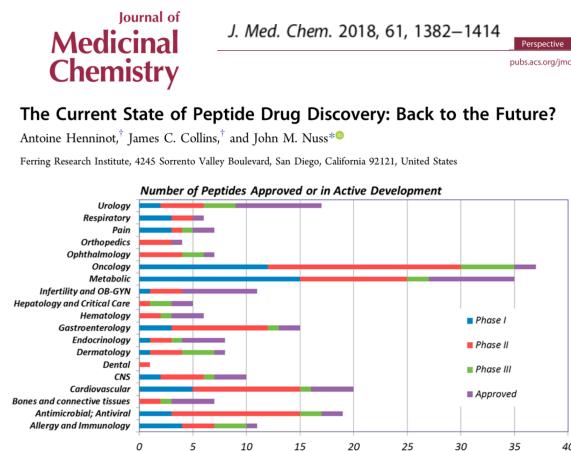
pH 9.0

pH 3.8



# Predicting Protein Properties: Lipophilicity

## Application: Lipophilicity of Peptides Using a Cumulative Lipophilicity



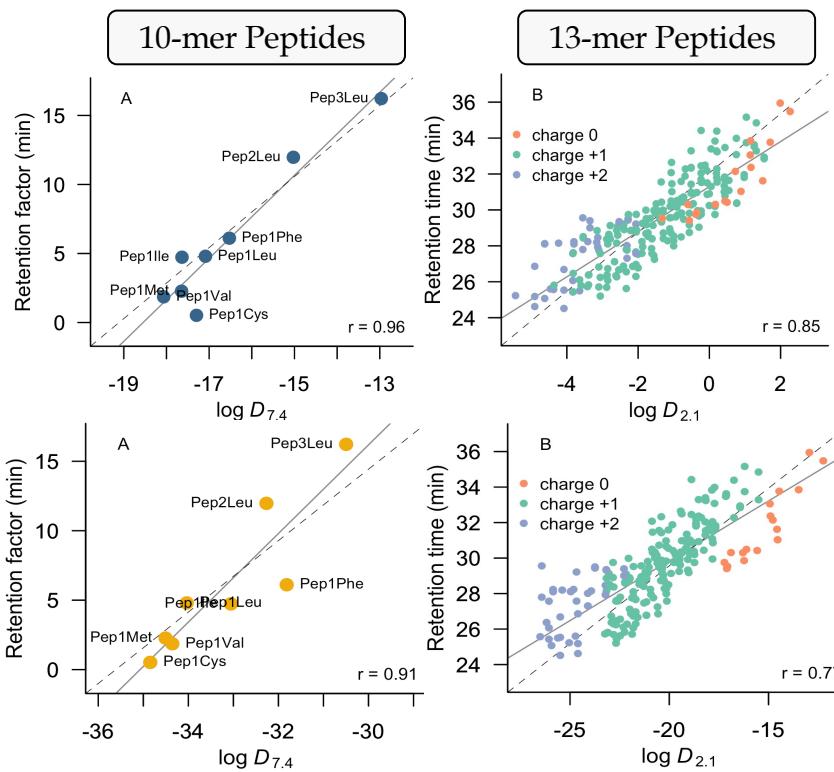
$$\log(P_N / D_{pH})^{peptide} = \sum_{i=1}^{N_{res}} \log(P_N^i / D_{pH}^i)^{bb+sc} + \sum_{i=1}^{N_{cg}} \log(P_N^i / D_{pH}^i)^{cg} \quad (29)$$



## Predicting Protein Properties: Lipophilicity

*Application: Resolution of peptides in RP-HPLC*

SolvL Scale

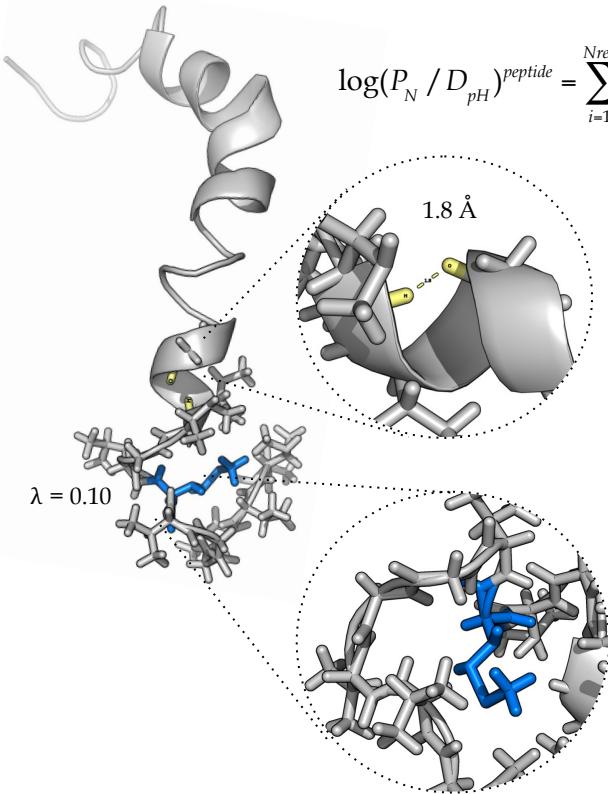


ProtL Scale



## Predicting Protein Properties: Lipophilicity

*Application: Lipophilicity of Peptides Using a Local-Context Dependent Lipophilicity*



$$\log(P_N / D_{pH})^{peptide} = \sum_{i=1}^{N_{res}} (\lambda^i \cdot \log(P_N^i / D_{pH}^i)^{bb+sc} + \lambda^i \cdot \log(P_N^i / D_{pH}^i)^{cg} + \alpha^i + \beta^i) \quad (30)$$

### 1. Parameter $\lambda$

$$\lambda^i = \frac{S_{peptide}^i}{\langle S_{AA\ model}^i \rangle} \quad (31)$$

### 2. Parameter $\alpha$

0.73 (log P units) per HB

### 3. Parameter $\beta$

$$\beta^i = H_{res}^i \cdot (1 - \lambda^i)^{sc} \quad (32)$$

$$H_{res}^i = \frac{0.023 \cdot SASA_{res}^{sc}}{2.303RT} \quad (33)$$

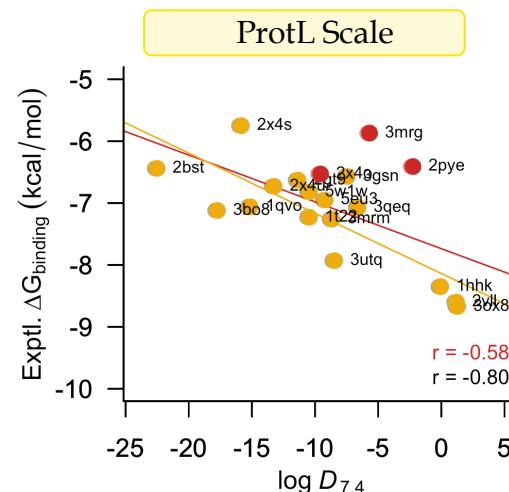
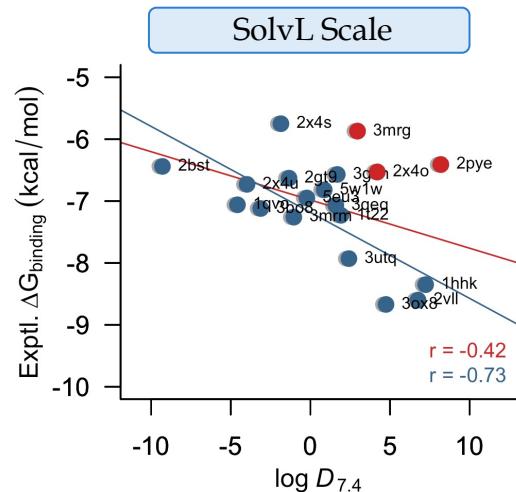
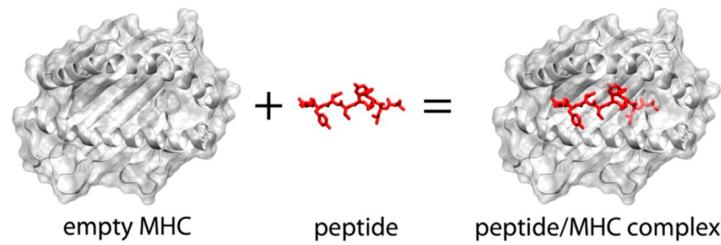
**Table 7.** Average Solvent Accessible Surface Area (SASA) for the Side Chain of the Hydrophobic Residues and the Hydrophobic Effect Contribution Value when the Side Chain is Fully Buried.

Residue	Average SASA (Å <sup>2</sup> )	$H_{res}^i$ (log P units)
Ala	69	1.2
Val	130	2.2
Leu	158	2.7
Ile	157	2.6
Met	166	2.8
Pro	115	1.9
Phe	188	3.2
Trp	232	3.9
Tyr	201	3.4



## Predicting Protein Properties: Lipophilicity

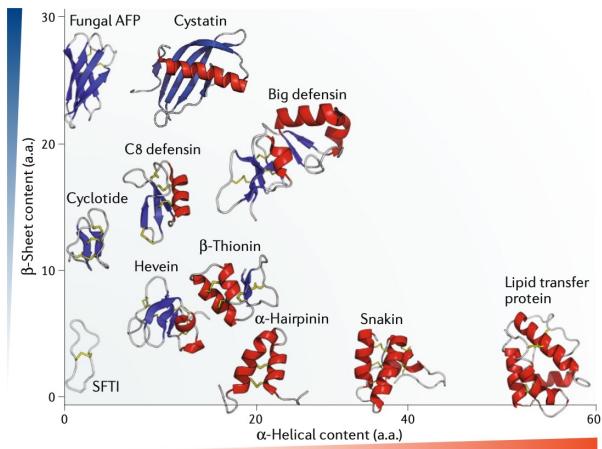
*Application: Relationship Lip. & Exptl. Binding Free Energies*



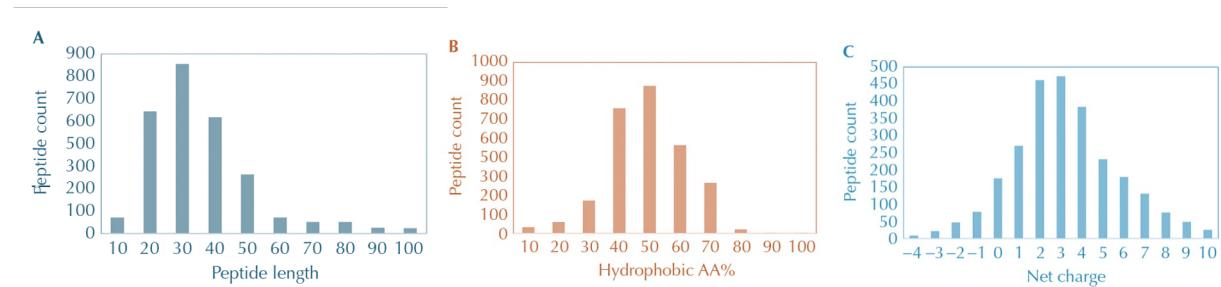


## Predicting Protein Properties: Lipophilicity

### ○ Cationic Host Defense Peptides (HDPs)



Mookherjee, N., Anderson, M. A., Haagsman, H. P., & Davidson, D. J. (2020). Antimicrobial host defence peptides: functions and clinical potential. *Nature reviews. Drug discovery*, 19(5), 311–332.

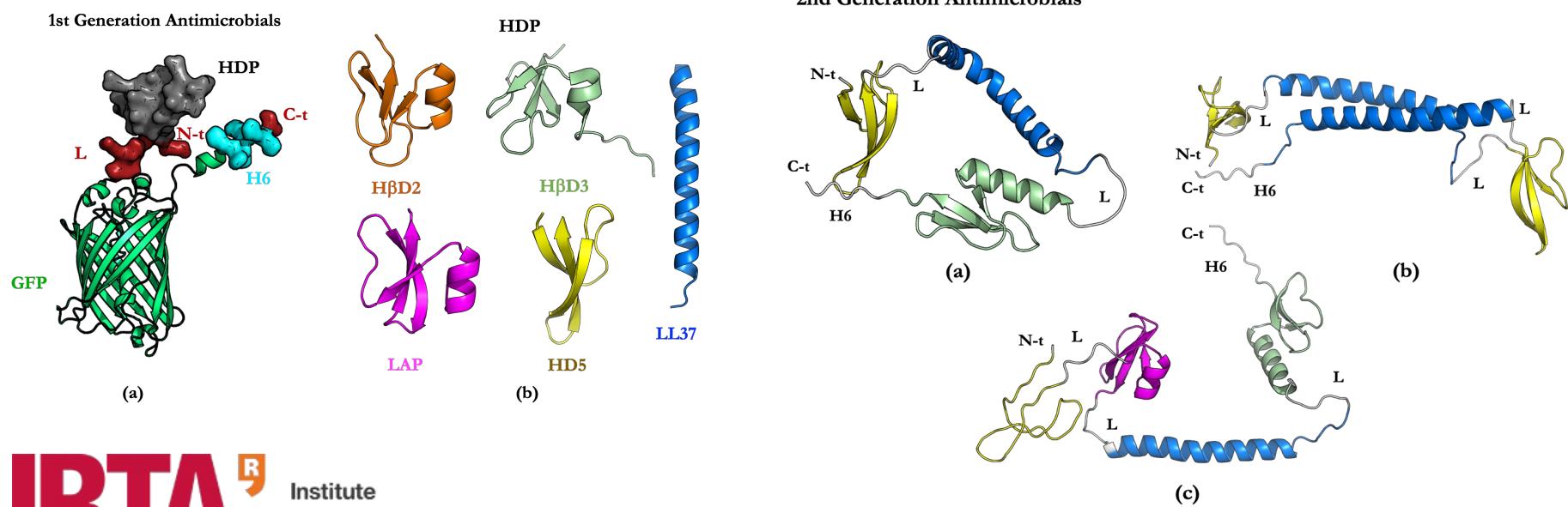


Hydrophobicity/Lipophilicity  
Amphipathicity



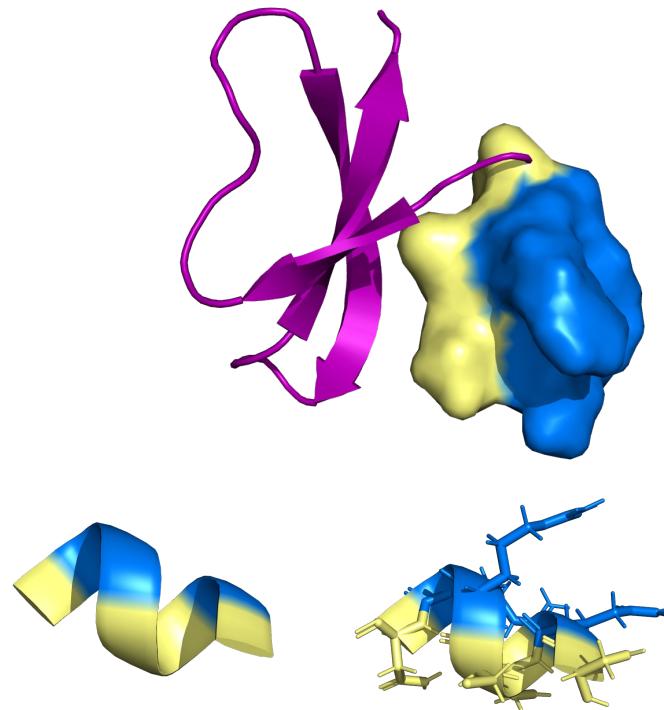
## Predicting Protein Properties: Lipophilicity

*Structure-Antimicrobial Activity Relationships of Recombinant Host Defense Peptides Against Drug-Resistant Bacteria*

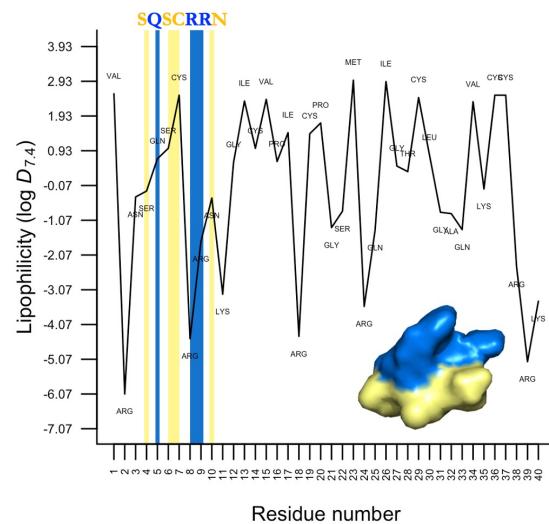




## Predicting Protein Properties: Lipophilicity



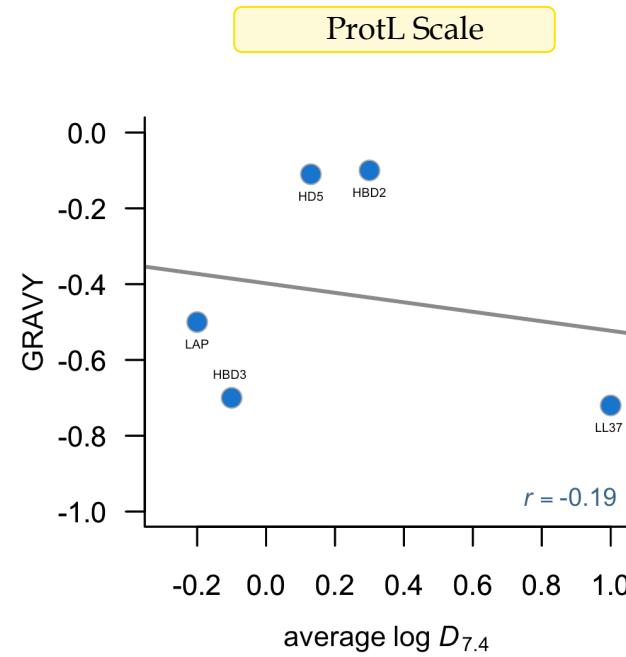
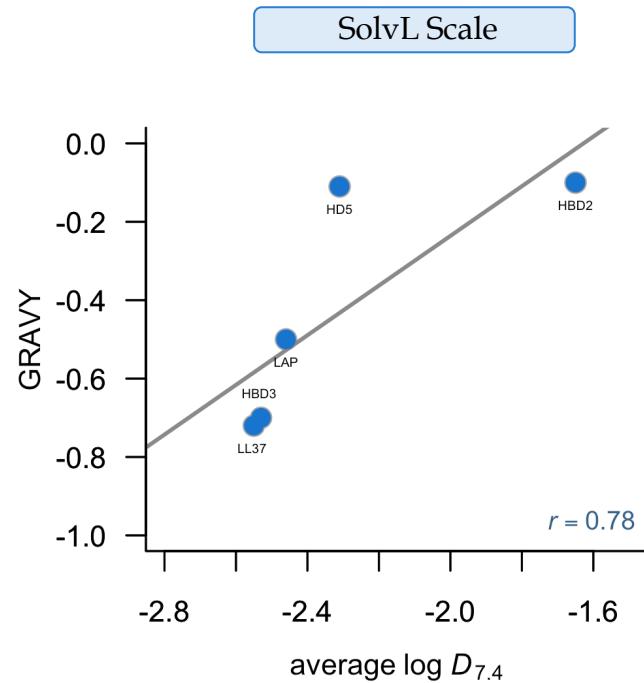
$$Amp = \sum_{i=1}^n \log D_i - \sum_{j=1}^m \log D_j$$



$$\log(D_{\text{pH}}/P_N)^{\text{peptide}} = \sum_{i=1}^N (\lambda^i \cdot \log(D_{\text{pH}}/P_N)^{\text{bb+sc}} + \lambda^i \cdot \log(D_{\text{pH}}/P_N)^{\text{cg}} + \alpha^i + \beta^i + \gamma^i)$$



## Predicting Protein Properties: Lipophilicity

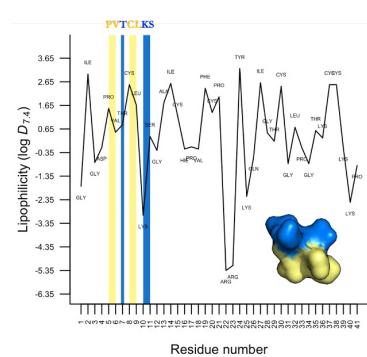




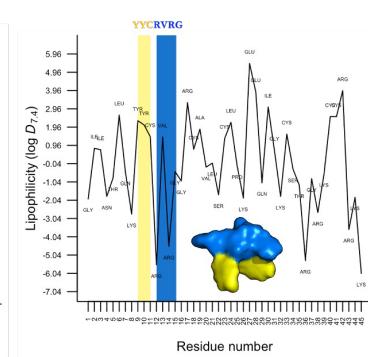
# Predicting Protein Properties: Lipophilicity

## *Physicochemical properties of 1st generation antimicrobials*

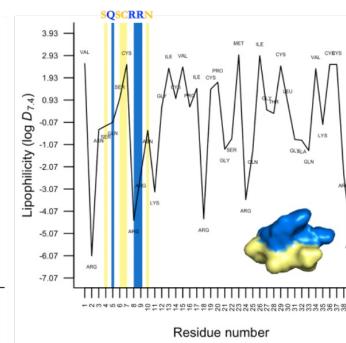
H $\beta$ D2



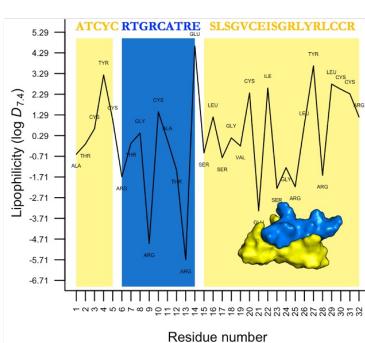
H $\beta$ D3



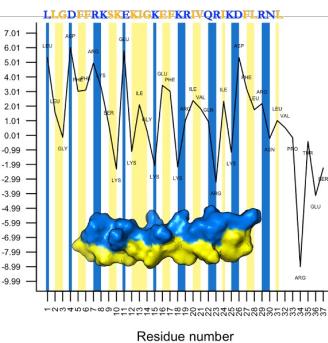
LAP



HD5



LL37



### Host Defense Peptide (HDP)

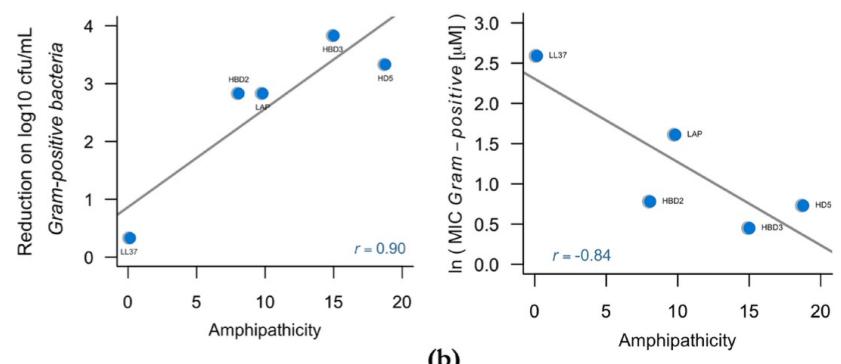
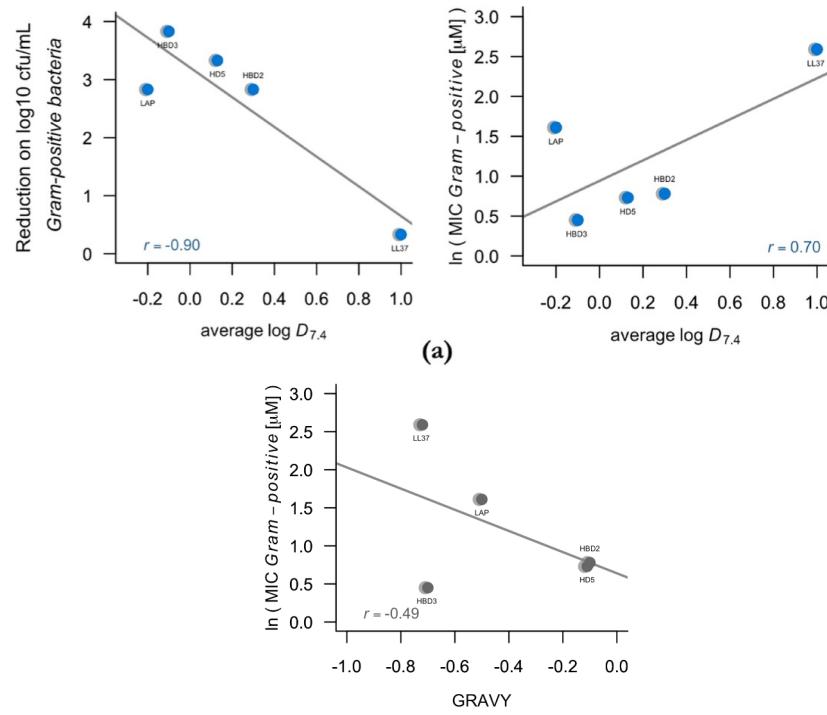
### Amphipathicity (Amp)

H $\beta$ D2	8.08
H $\beta$ D3	15.02
LAP	9.83
HD5	18.78
LL37	0.16



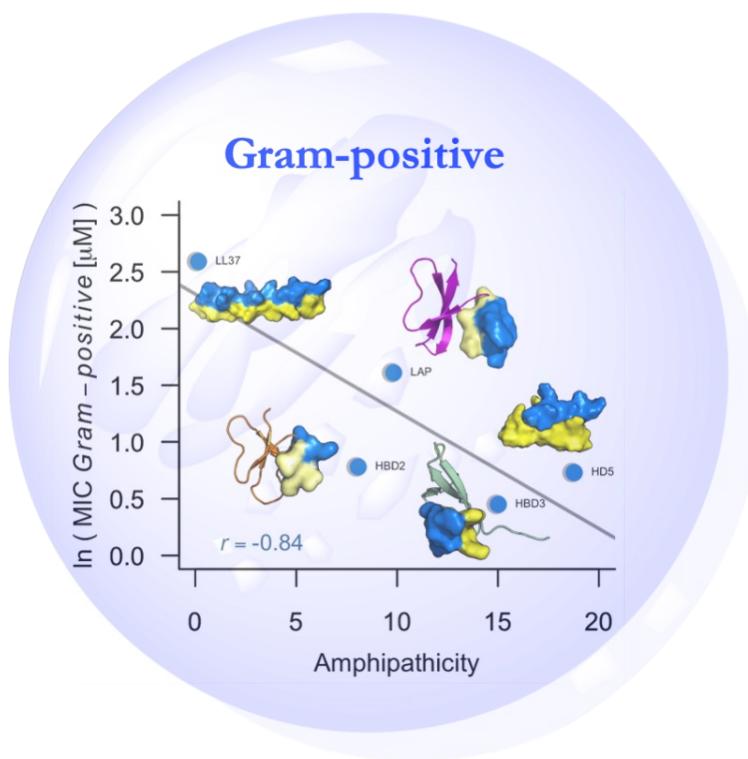
## Predicting Protein Properties: Lipophilicity

*Structure-antimicrobial activity relationships of  
1st generation host defense peptides against Gram-positive*





## Predicting Protein Properties: Lipophilicity



$$Amp = \sum_i^n \log D_i - \sum_j^n \log D_j$$



# Predicting Protein Properties: Protein Aggregation

Received: 1 August 2024 | Revised: 3 September 2024 | Accepted: 6 September 2024

DOI: 10.1002/pro.5180

**TOOLS FOR PROTEIN SCIENCE**



## Aggrescan4D: A comprehensive tool for pH-dependent analysis and engineering of protein aggregation propensity

Mateusz Zalewski<sup>1</sup> | Valentin Iglesias<sup>2,3</sup> | Oriol Bárcenas<sup>2,4</sup> |  
Salvador Ventura<sup>2,5</sup> | Sebastian Kmiecik<sup>1</sup>

<sup>1</sup>Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Warsaw, Poland

<sup>2</sup>Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>3</sup>Clinical Research Centre, Medical University of Białystok, Białystok, Poland

<sup>4</sup>Institute of Advanced Chemistry of Catalonia (IQAC), CSIC, Barcelona, Spain

<sup>5</sup>Hospital Universitari Parc Taulí, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Universitat Autònoma de Barcelona, Sabadell, Spain

Parkinson

Alzheimer

Type II diabetes

Various cancers



## Predicting Protein Properties: Protein Aggregation

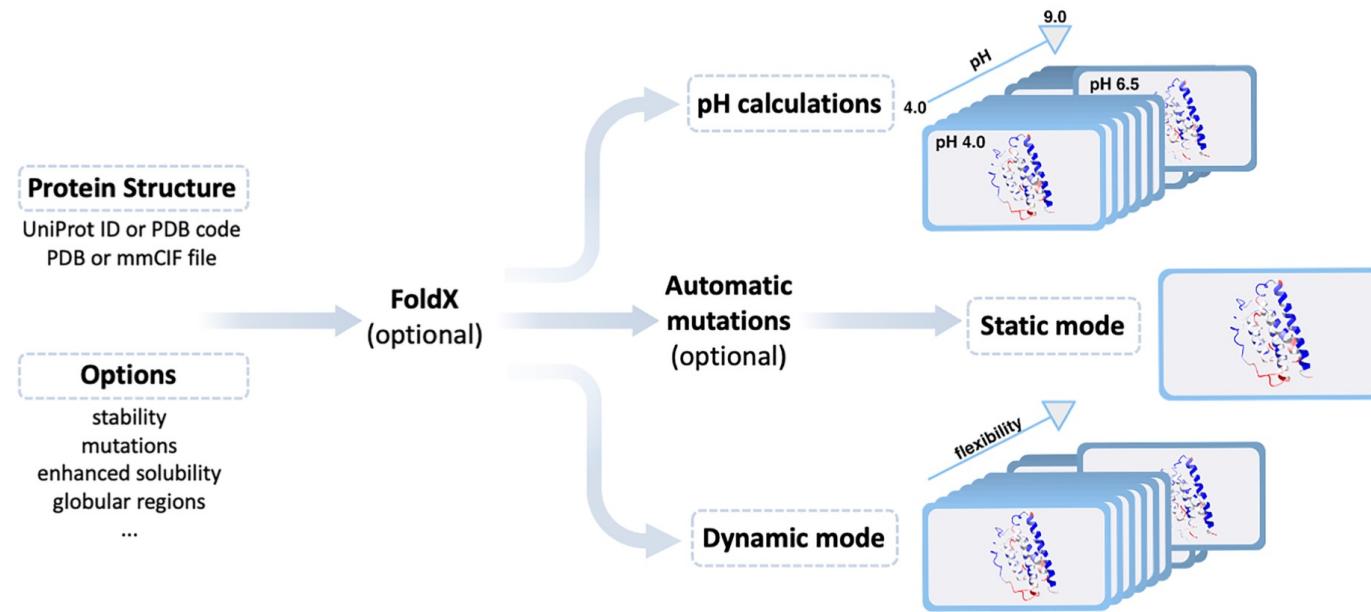
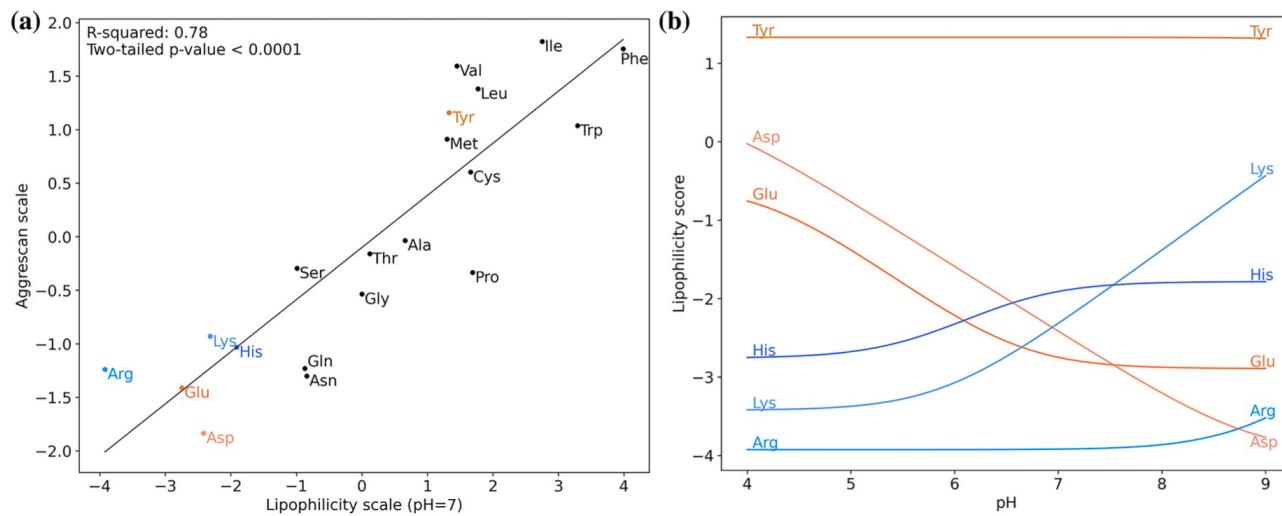


FIGURE 1 A4D pipeline.



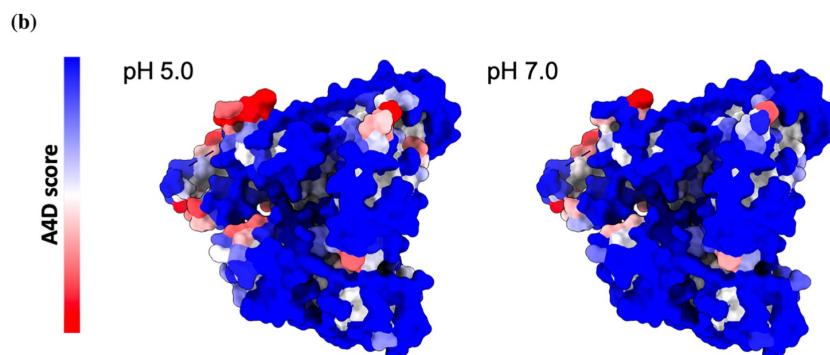
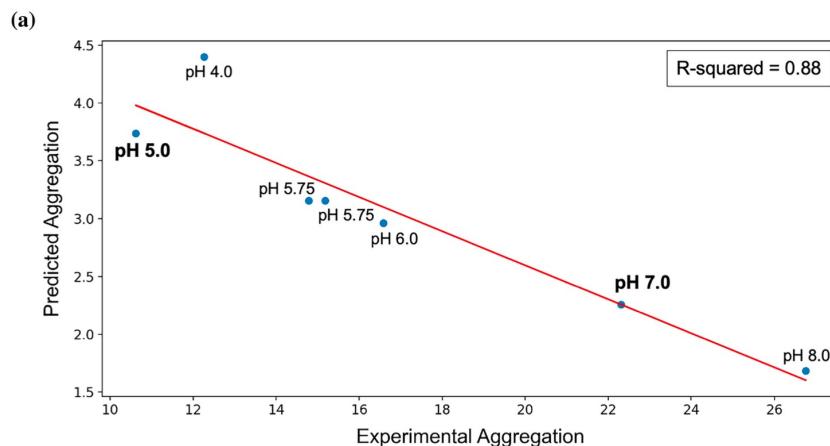
## Predicting Protein Properties: Protein Aggregation



**FIGURE 2** Incorporation of pH modulated changes in aggregation propensity in A4D. (a) Linear correlation between the pH-dependent lipophilicity scale used for pH calculations at neutral pH and the *in vivo* derived Aggrescan scale used for non-pH calculations. (b) pH-driven changes in the lipophilicity score. Due to limitations on satisfactory 3D pKa estimation for Arginine, its score is considered constant in the A4D assayed pH-range.



## Predicting Protein Properties: Protein Aggregation



$$A4D \text{ score} = \text{Agg}_i(\alpha e^{\beta RSA_i}) + \sum_e [\text{Agg}_e(\alpha e^{\beta RSA_e})(\gamma e^{-\delta \text{dist}})]$$

*Human Serum Albumin (PDB ID: 6YG9) aggregation propensity predictions using A4D pH calculations*

Blue for solubility

Red for aggregation

White for neutral region



# Thank you for your attention

*Graciès!*

*Merci!*

*Gracias!*

*Thanks!*

*Obrigado!*



[william.zamoraramirez@ucr.ac.cr](mailto:william.zamoraramirez@ucr.ac.cr)