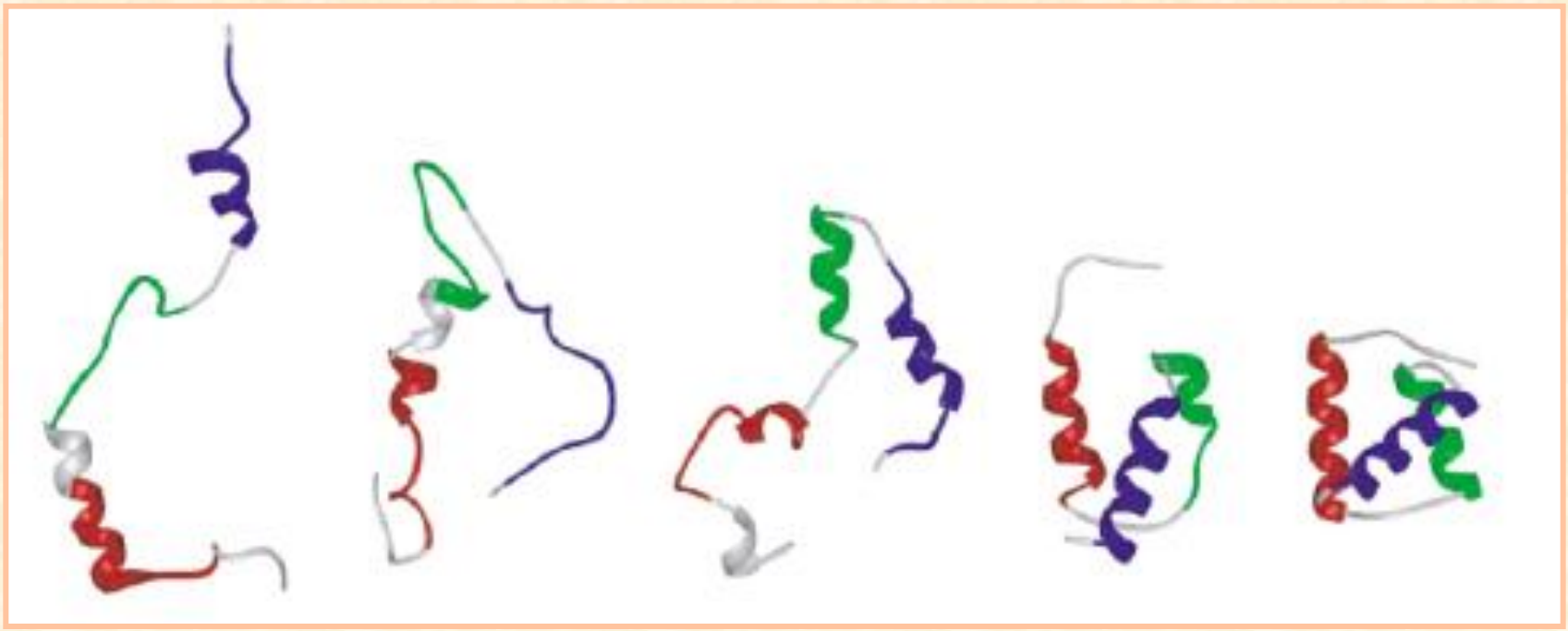
Sequence-Based Predictions of Protein Solubility

Michele Vendruscolo
Department of Chemistry
University of Cambridge

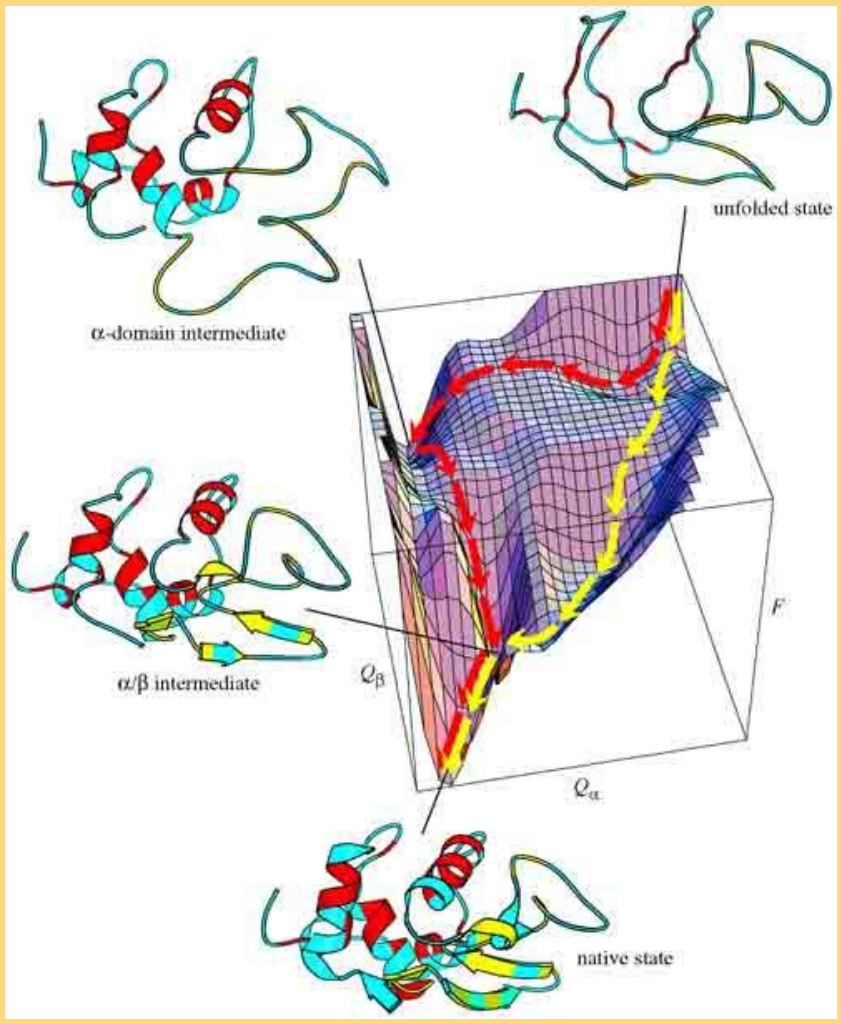
Molecular Interactions in Biopharmaceutical Formulations:
Can stability be rationalised and predicted?
Tuesday 30th October 2012 Trinity Centre, Cambridge, UK
MIBio 2012

Protein folding



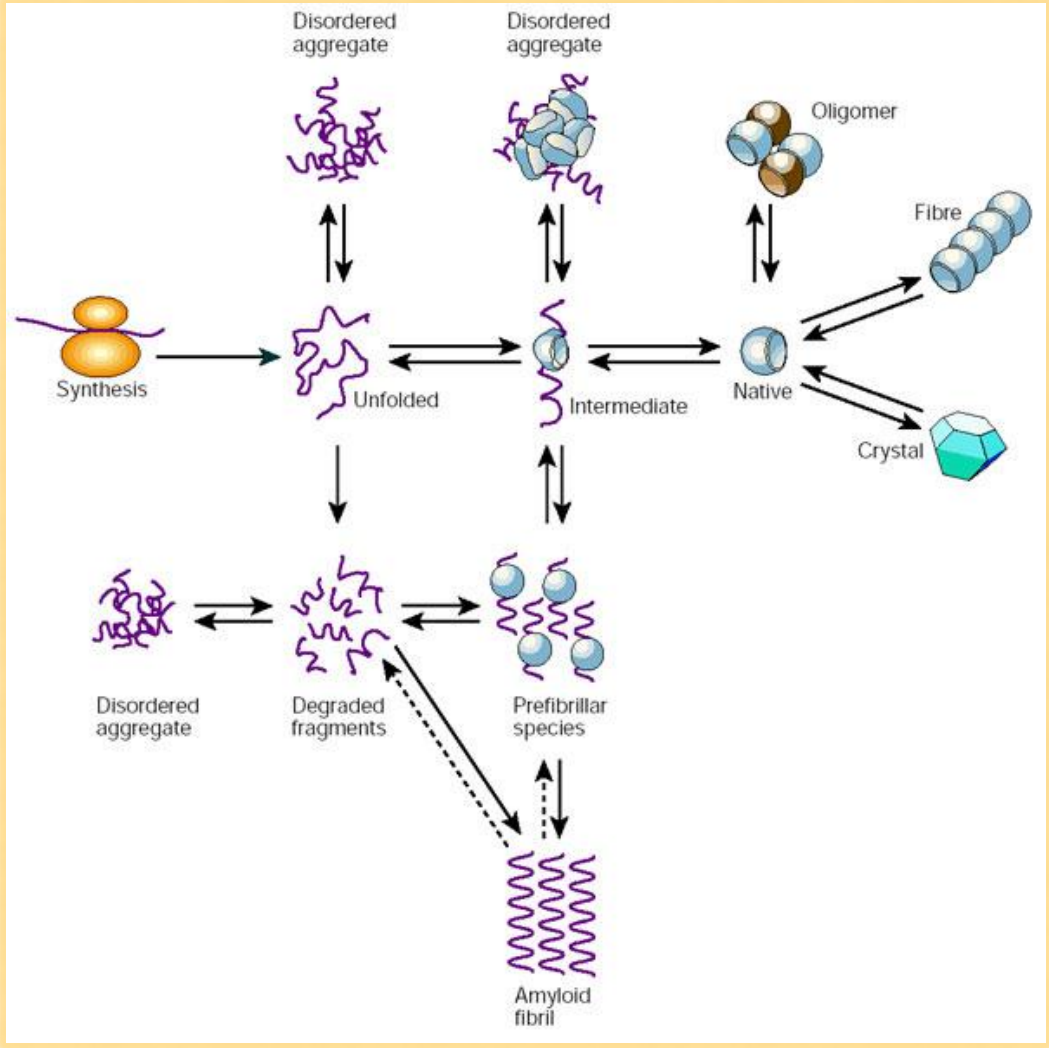
The fundamental code for protein folding is provided by the amino acid sequence.

Amino acid sequences encode the whole free energy landscape of proteins

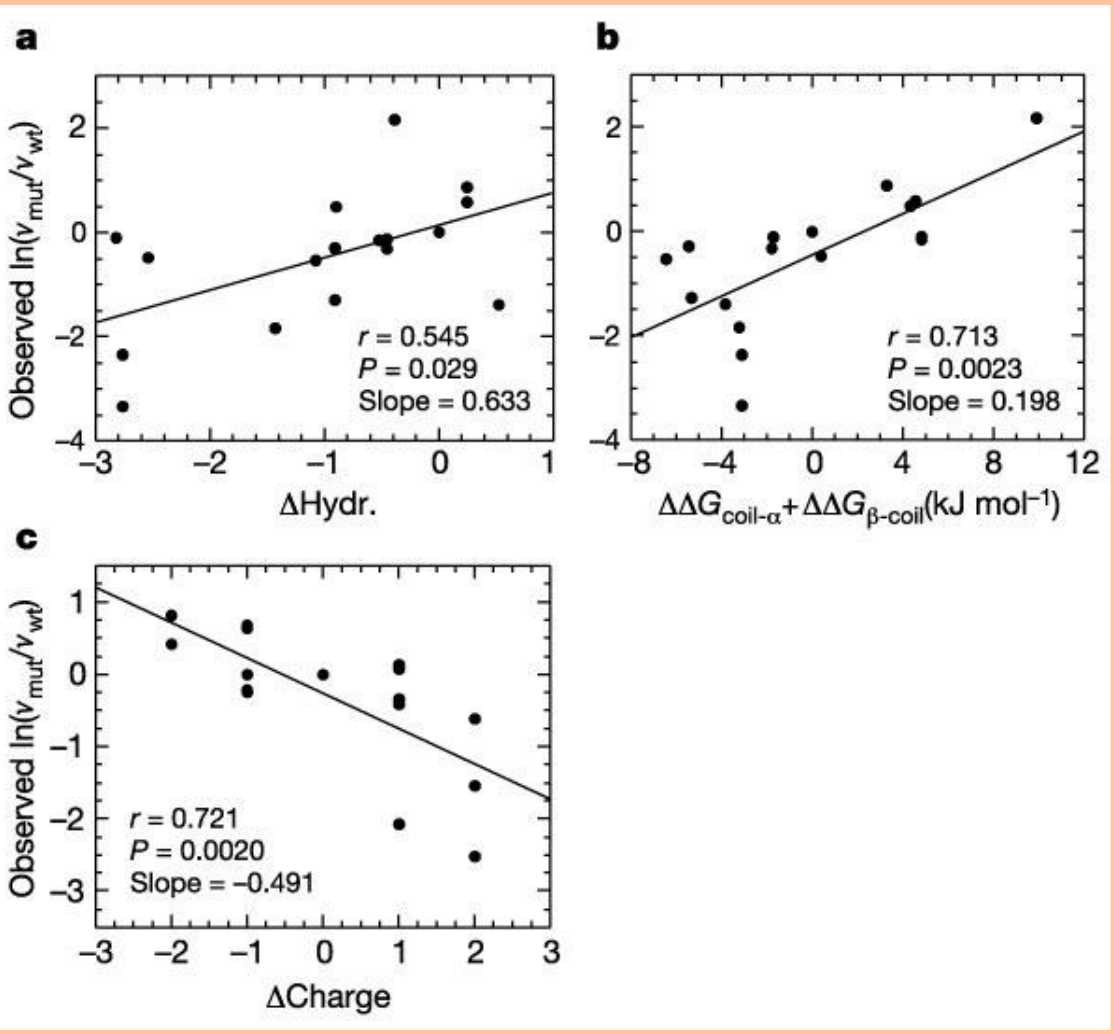


Not only the native structure but also all the other states and the corresponding pathways of interconversion are encoded in the amino acid sequence of a protein.

Does the amino acid sequence encode also for aggregation?



Physico-chemical principles of protein aggregation



Hydrophobicity, charge and secondary structure propensity are correlated with the changes in the aggregation rates upon mutation.

Chiti et al. Nature 2003

Sequence-based prediction of aggregation rates

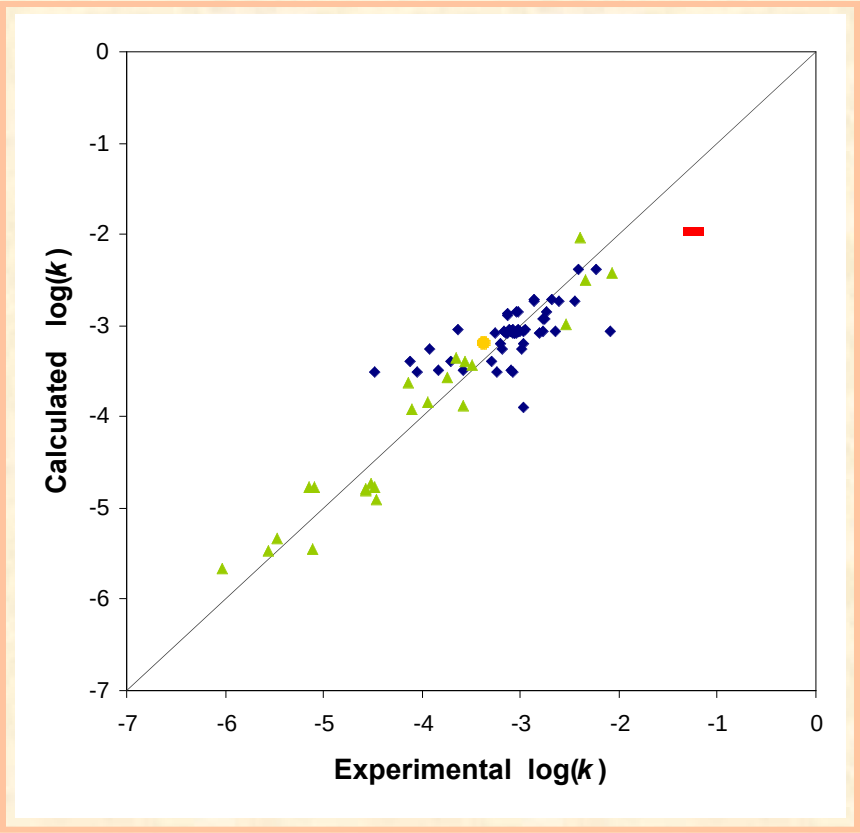
The combination of sequence-dependent factors and environmental factors enables the prediction of aggregation rates over a broad range of timescales (from seconds to weeks)

$$\ln(k) = \sum a_k I_k + \sum a_k E_k$$

ln(k): logarithm of the aggregation rate k

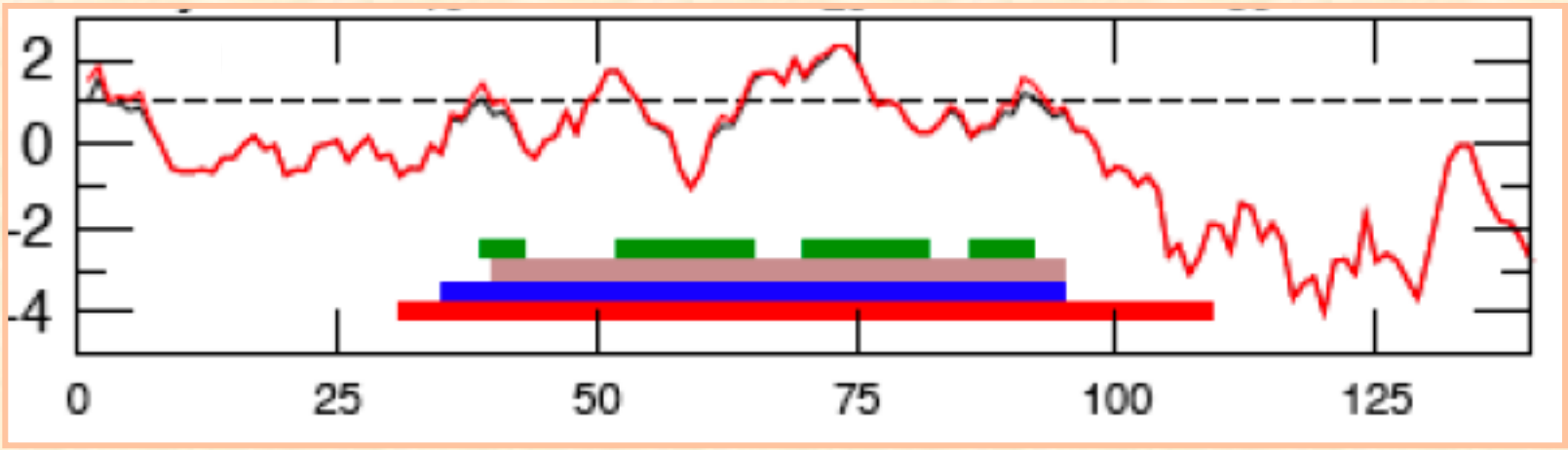
I^{hydr}: hydrophobicity
I^{pat}: hydrophobic patterns
I^{α} : α-helical propensity
I^{β} : β-sheet propensity
I^{ch}: charge contribution

E^{pH}: pH of the solution
E^{ionic}: ionic strength
E^{conc}: polypeptide concentration



DuBay et al. J. Mol. Biol. 2004

Prediction of aggregation-prone regions of α-synuclein

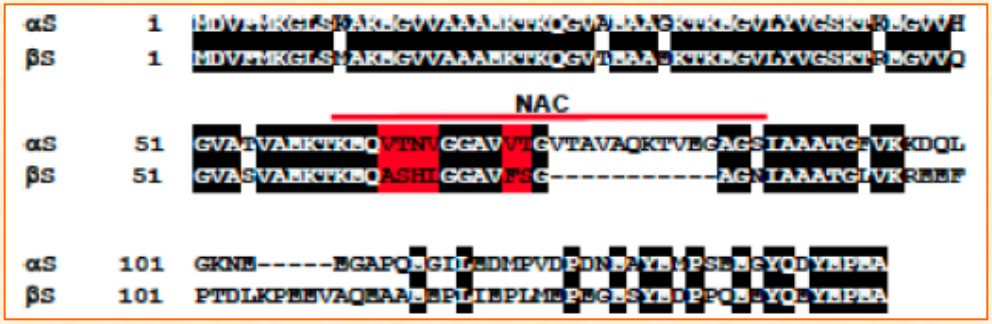


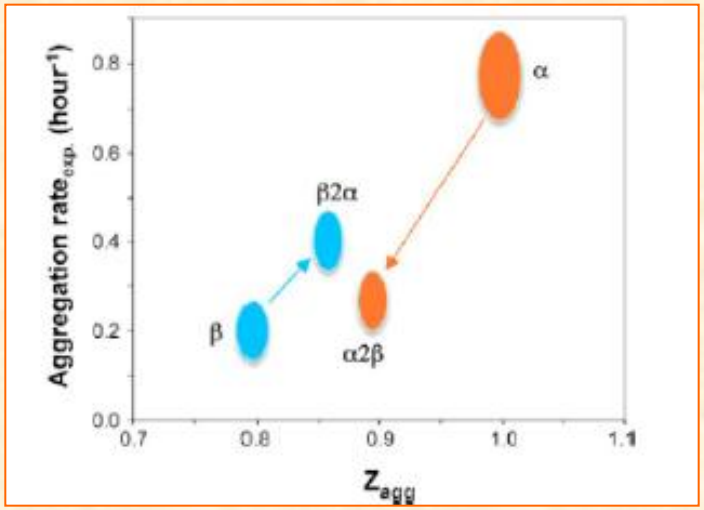
The aggregation propensity is a function of the physico-chemical properties of the amino acid sequence (hydrophobicity, charge, etc).

We have developed the Zyggregator method to predict aggregation rates and aggregation-prone regions (www-vendruscolo.ch.cam.ac.uk)

Tartaglia et al. J. Mol. Biol. 2008

Conversion of α-synuclein into β-synuclein by a six-residue swap

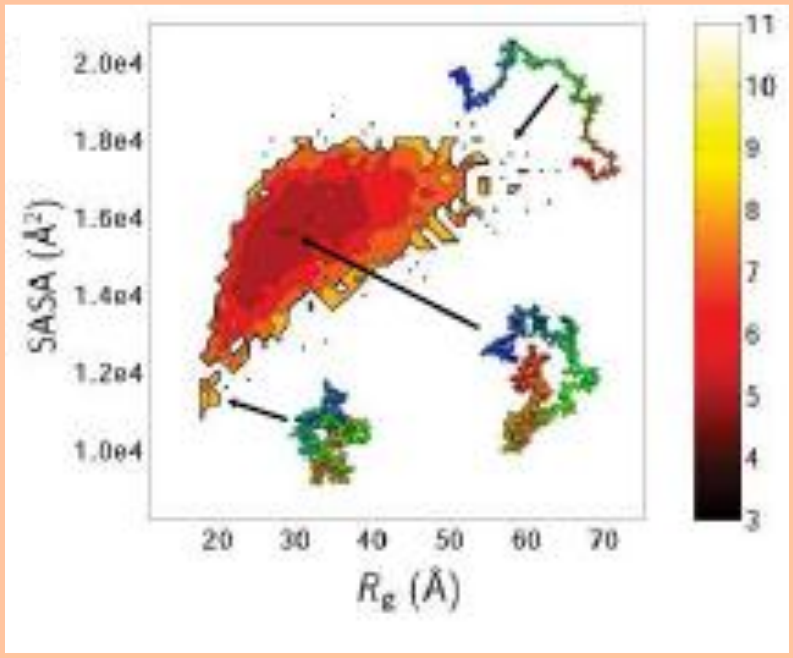




By using the Zyggregator method we have rationally designed a mutant form of α-synuclein with the same aggregation behaviour of β-synuclein

Roodveldt et al. Biochemistry, in press

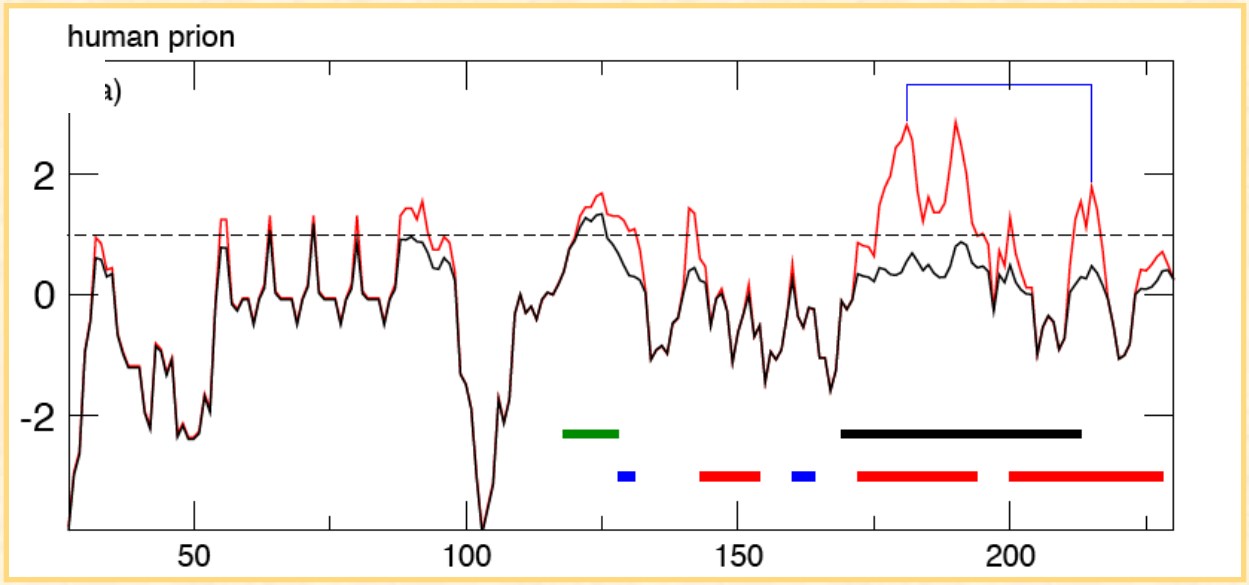
NMR determination of the natively unfolded state of α-synuclein

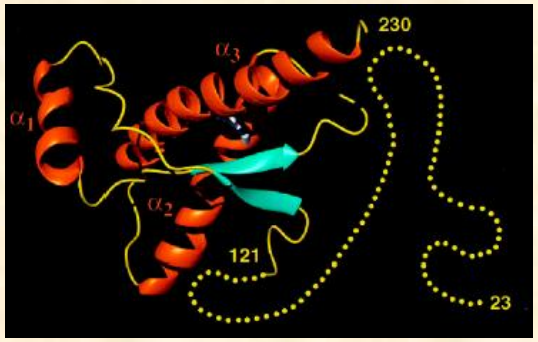


We used NMR spectroscopy in combination with molecular dynamics simulations to determine an ensemble of conformations representing the natively unfolded state of α-synuclein.

Dedmon et al. J. Am. Chem. Soc. 2005
Allison et al. J. Am. Chem. Soc. 2009

Folding against aggregation: Aggregation-prone regions of the human prion protein





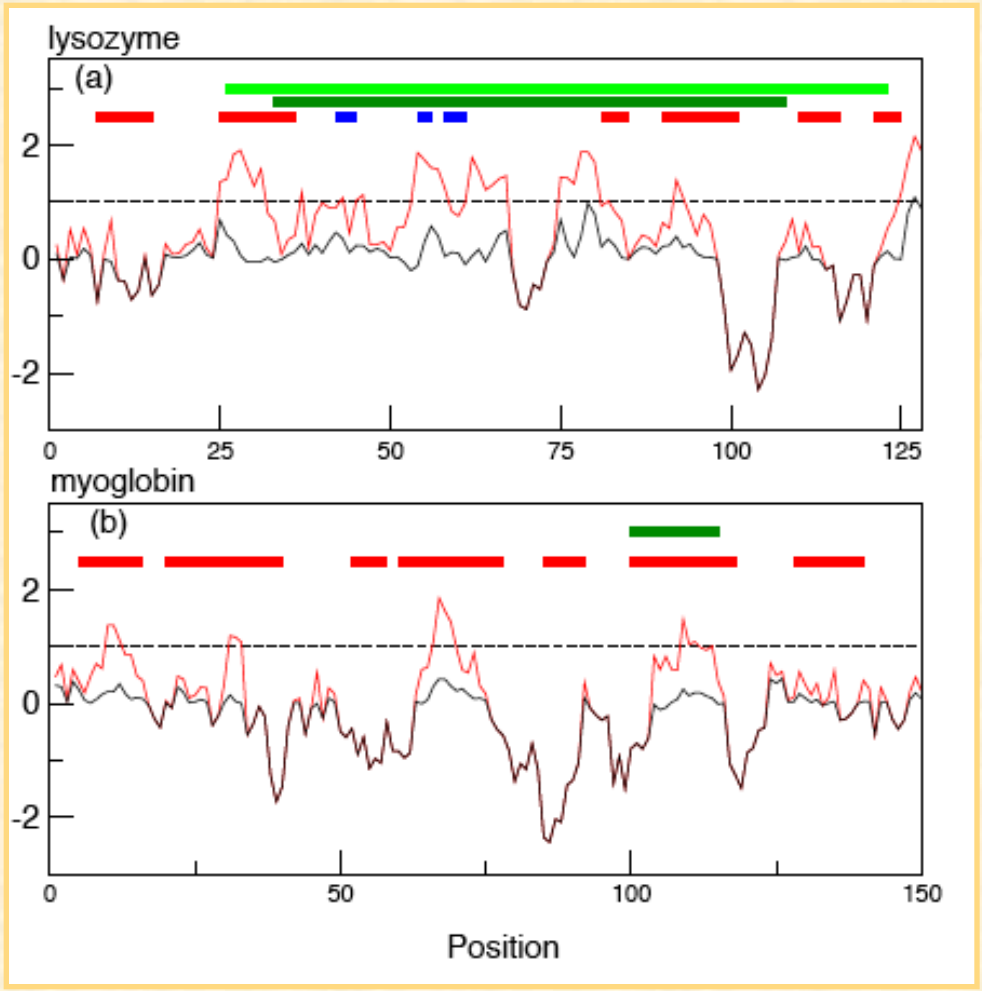
Tartaglia et al. J. Mol. Biol. 2008

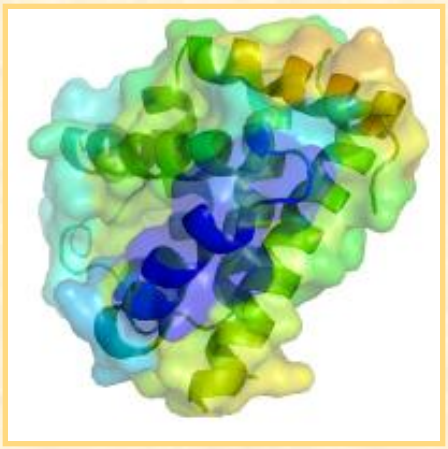
The region 181-186 (helix 2) as a high intrinsic propensity to aggregate.
However, the folded structure of the PrP^{C} state prevents this tendency to initiate aggregation.

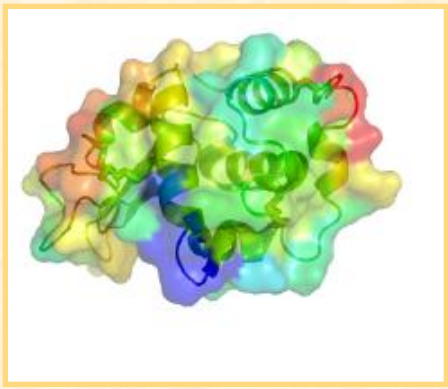
After the PrP^{C} state is destabilised into the PrP^{Sc} state , the C-terminal regions of high aggregation propensities become available to form the structural core of the amyloid fibrils.

This is negative design principle against aggregation.

Protection against aggregation in globular proteins



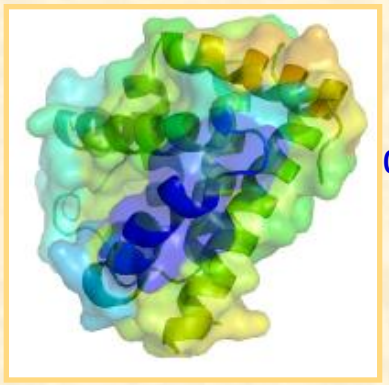




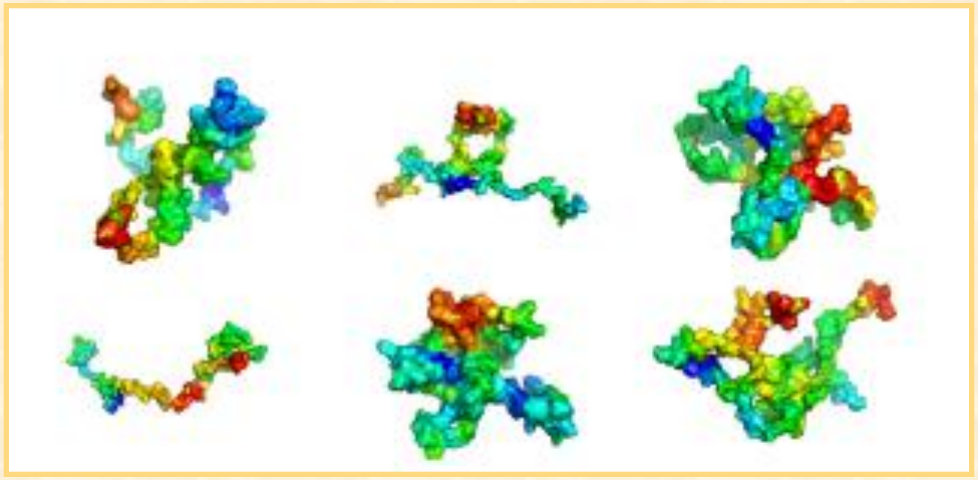
Regions of high intrinsic aggregation propensity are not exposed in native states.
Globular proteins usually do not aggregate unless they are destabilised.

Tartaglia et al. J. Mol. Biol. 2008

Protection against aggregation in native states: Aggregation-prone surfaces

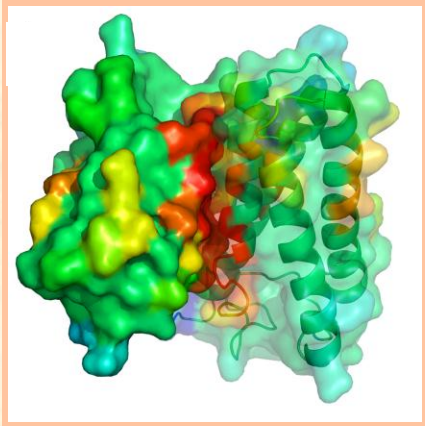


destabilisation

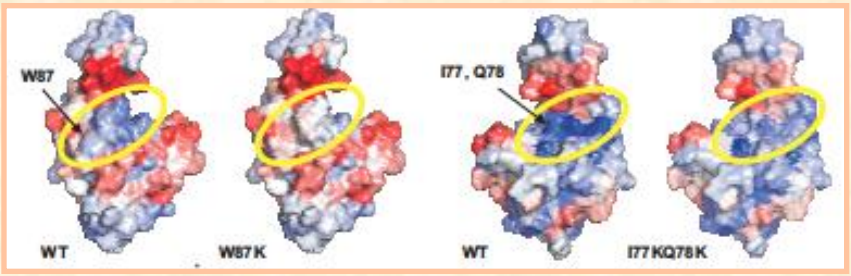


Regions of high intrinsic aggregation propensity are solvent exposed in non-native states, but they are protected in native states.

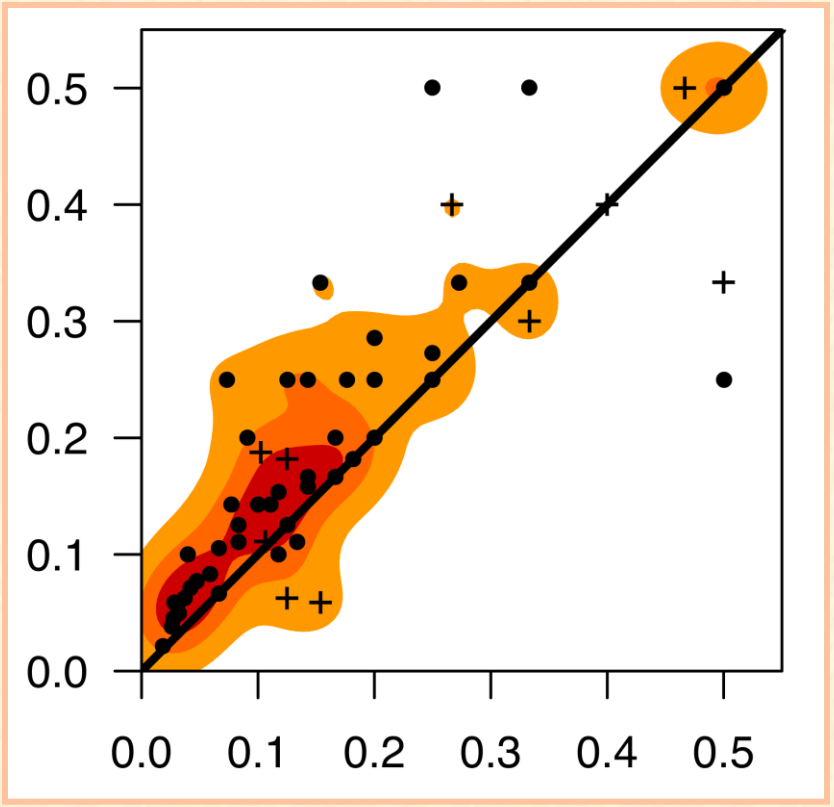
Competition between functional and dysfunctional association



Protein-protein complex interfaces are highly aggregation-prone



Aggregation propensity is a very good predictor of protein-protein interfaces



Disulfide bonds are found preferentially near aggregation-prone interfaces

Pechmann et al. PNAS 2009

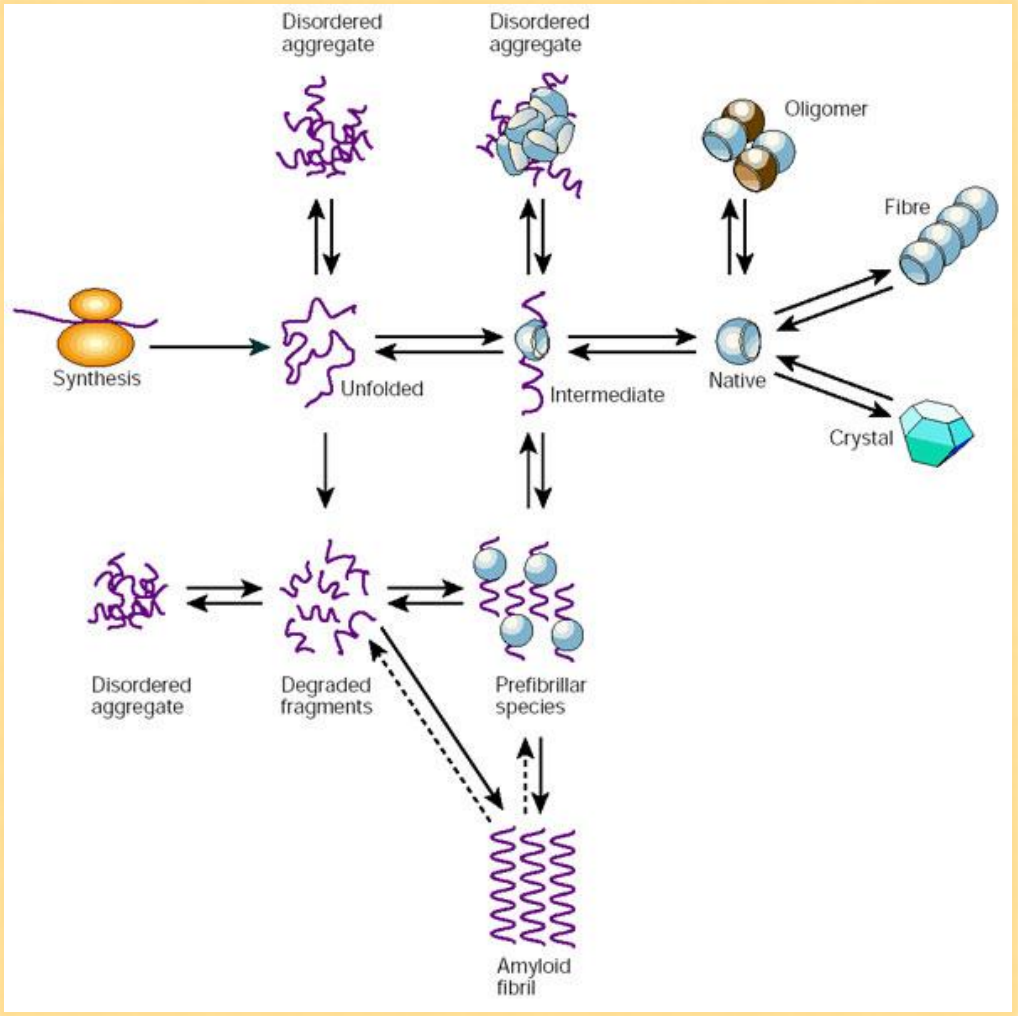
Folding, misfolding and aggregation are driven by the interplay of the same basic interactions

Hydrophobic interactions

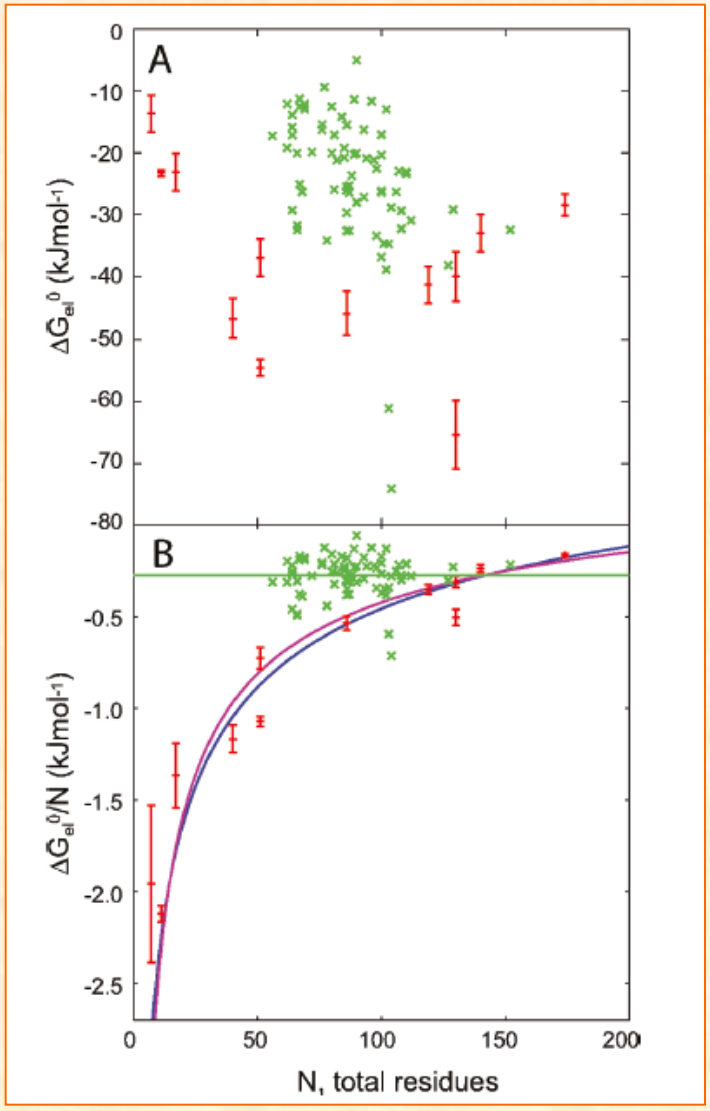
Hydrogen bonding

Electrostatic interactions

van der Waals interactions



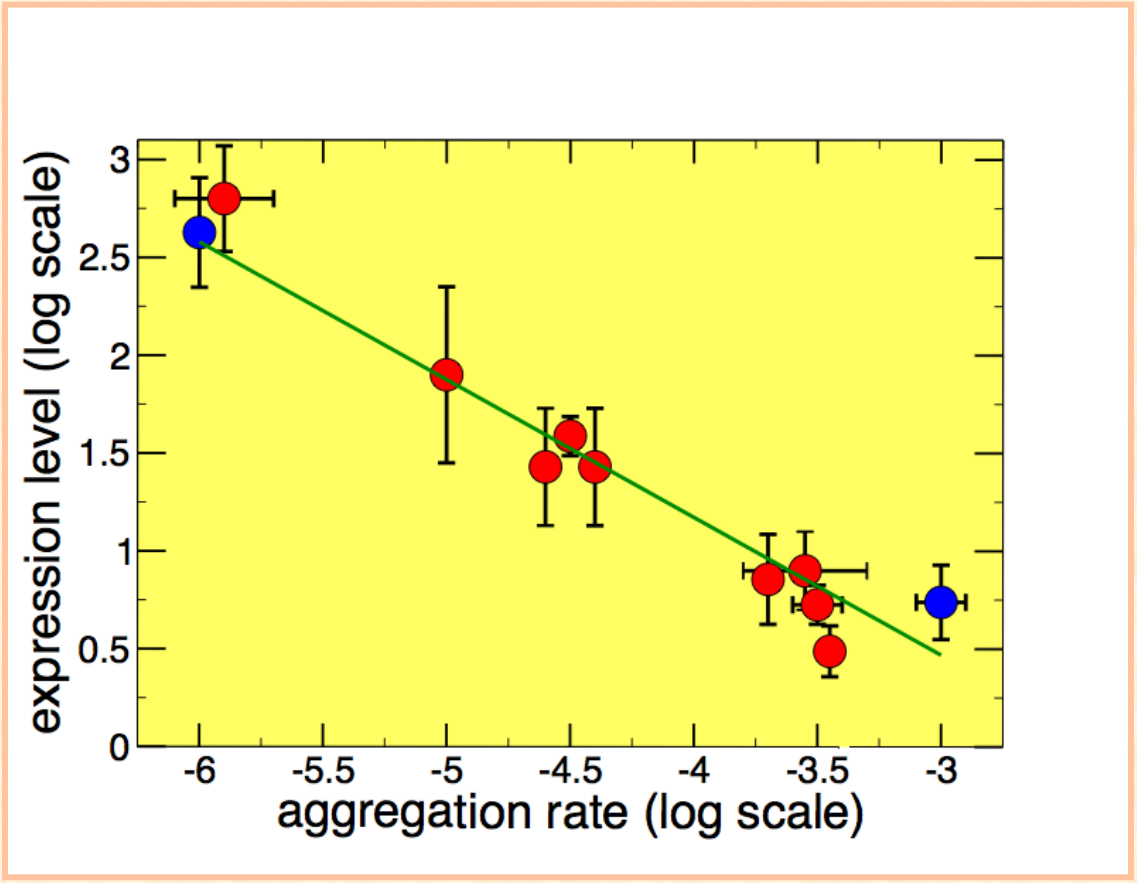
Native states of proteins are metastable against aggregation



Aggregated forms are more stable than native states but they are separated by high kinetic barriers from them.

Baldwin et al. JACS 2011

Proteins are expressed at their critical concentrations



The amino acid sequence of proteins have co-evolved with the cellular environment to resist aggregation

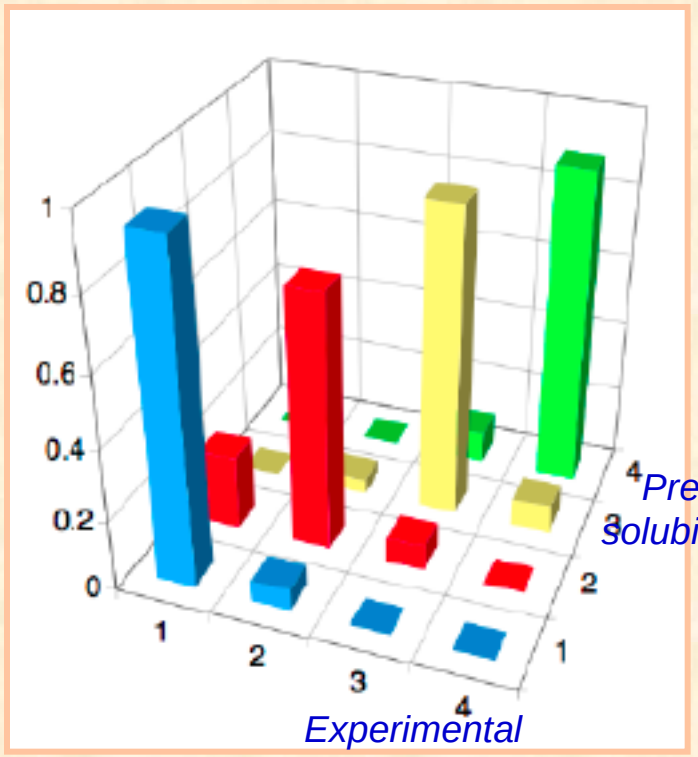
…but only up to the concentrations required for their optimal function.

Life on the edge:

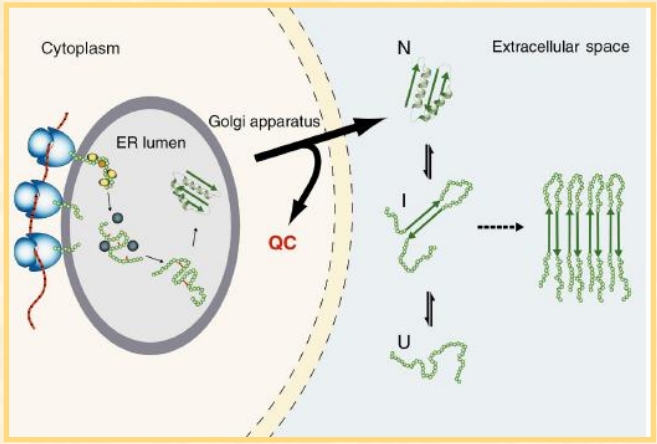
A small increase in expression or a small decrease in solubility lead to aggregation.

Tartaglia et al. TiBS 2007

Sequence determinants of protein solubility (i.e. of the balance between folding and aggregation)



The solubility of recombinant human proteins in *E.coli* (hEx1 database) can be predicted from their amino acid sequences.

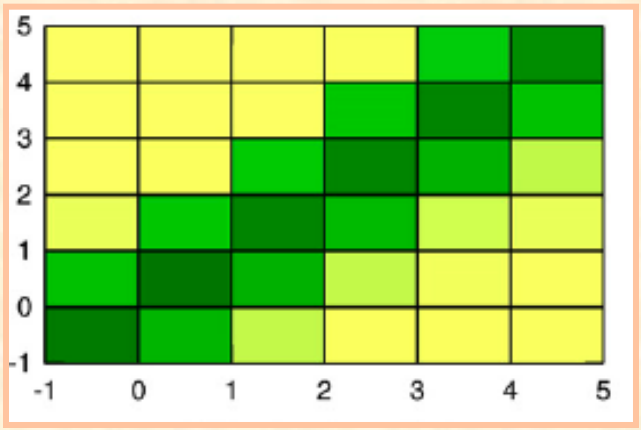


Tartaglia et al. JMB 2009

CamEL method: http://www-vendruscolo.ch.cam.ac.uk/camel.php

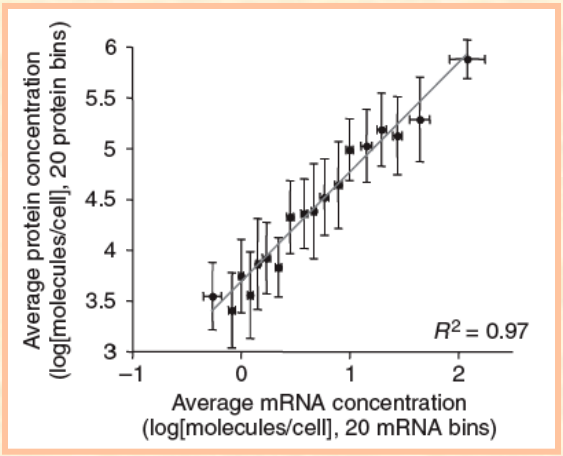
Sequence determinants of the maximal levels of protein abundance (i.e. as allowed by the solubility)

Experimental mRNA max expression levels



Predicted mRNA max expression levels

Tartaglia et al. JMB 2009



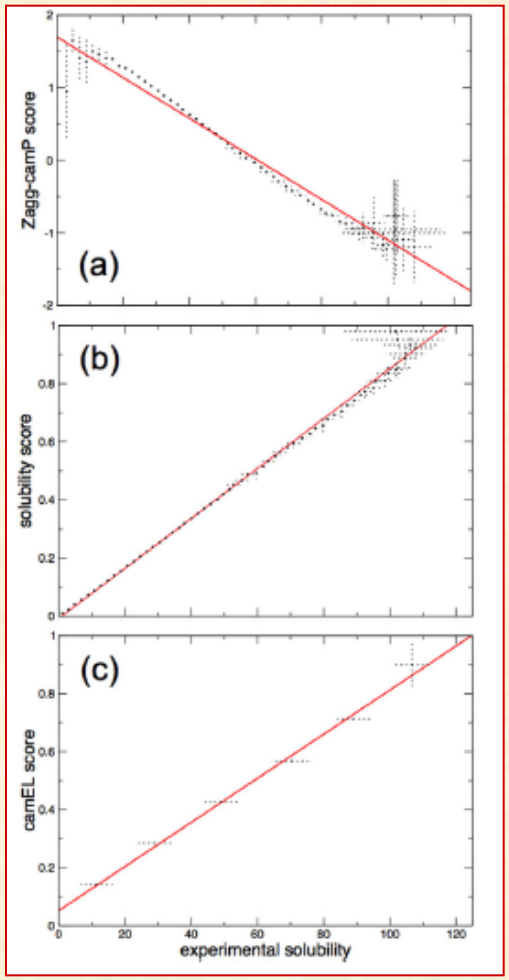
The maximal levels of mRNA expression in *E.coli* (CCDB database) can be predicted from the amino acid sequences of the corresponding proteins.

CamEL method: http://www-vendruscolo.ch.cam.ac.uk/camel.php

Lu et al. Nat. Biotech 2007

Sequence determinants of protein solubility

Correlation between experimental solubility and:



Predicted aggregation propensities (Zyggregator)

Predicted solubility (CCSOL)

Predicted abundance (CamEL)

www-vendruscolo.ch.cam.ac.uk/software.html

Agostini et al. J. Mol. Biol. 2012

The amino acid sequence encodes a series of propensities of a protein

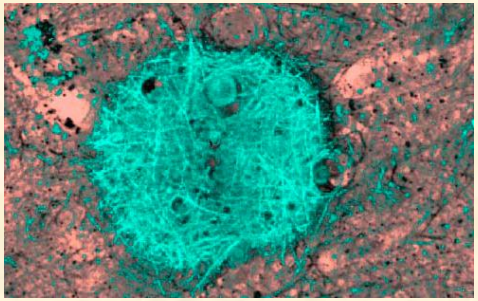
…although its actual behaviour will be eventually controlled by the environment.

Acknowledgements

Amol Pawar
Kateri DuBay
Prajwal Ciryam
Sebastian Pechmann
Pietro Sormanni

Gian Gaetano Tartaglia (CRG, Barcelona)
Chris Dobson (Cambridge)
Fabrizio Chiti (Florence)
Ulrich Hartl (MPI, Martinsried)

élan Elan Drug Technologies



CAMBRIDGE BRISTOL TORONTO HAMBURG
NEURODEGENERATIVE DISEASE CONSORTIUM