



Biophysical methods to guide protein crystallization and inhibitor binding studies

Paul Erbel, Frederic Villard, Allan d'Arcy

RAMC: September 2011



Introduction

overview

- Biophysical methods to characterize protein quality and guide crystallization
 - Limited proteolysis
 - Protein purification
 - Construct design
 - Special focus on NMR spectroscopy
- Biophysical methods to characterize compound binding
 - Support crystallization: ligand binding alters crystallization properties
 - Select compounds for cocrystallization



Biophysical methods: expensive toys or powerful tools?

Introduction: Classes of proteins

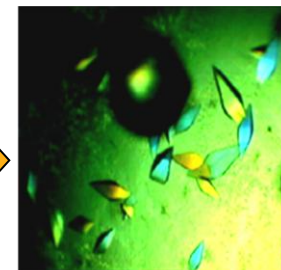
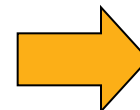
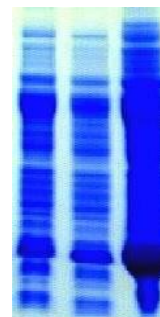
an oversimplified view

Proteins that cannot crystallize

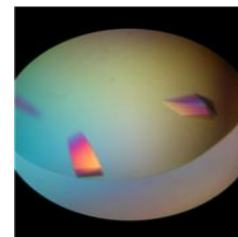
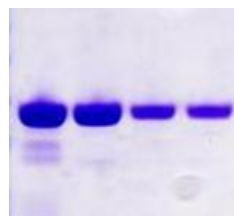
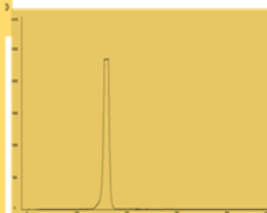
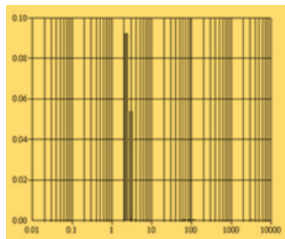


Aggregation/wrongly folded

Proteins that are easy to crystallize

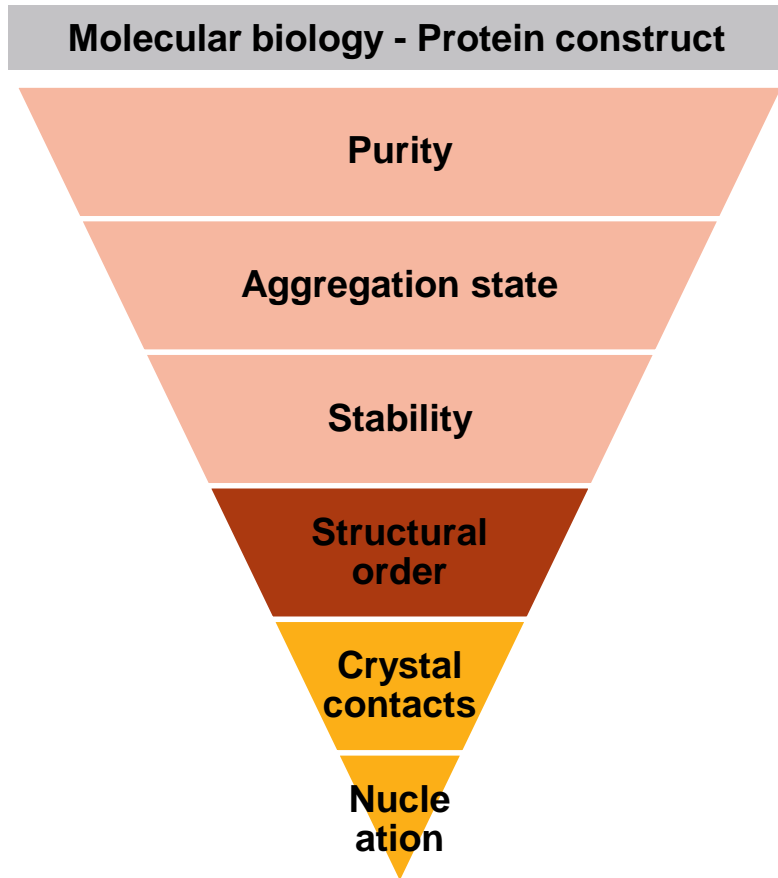


Proteins that are perfect but difficult to crystallize



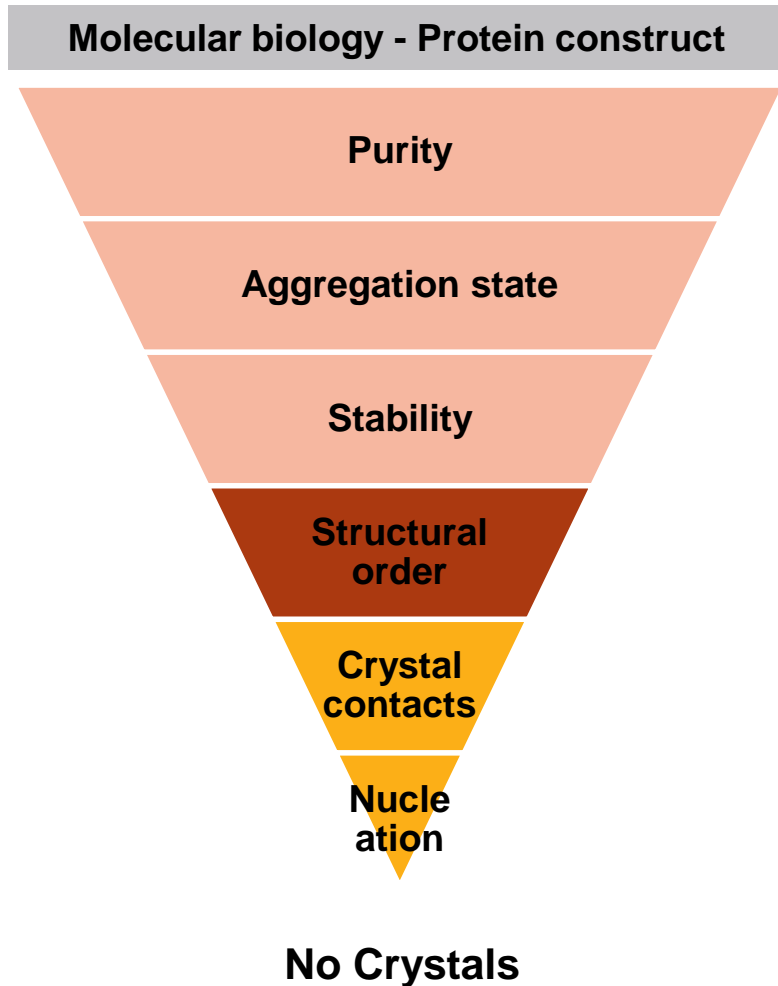
Introduction: critical parameters for protein crystallization?

Can we measure those parameters? Can we affect those parameters?



Introduction: critical parameters for protein crystallization?

Can we measure those parameters? Can we affect those parameters?



SDS Page
Size Exclusion Chromatography

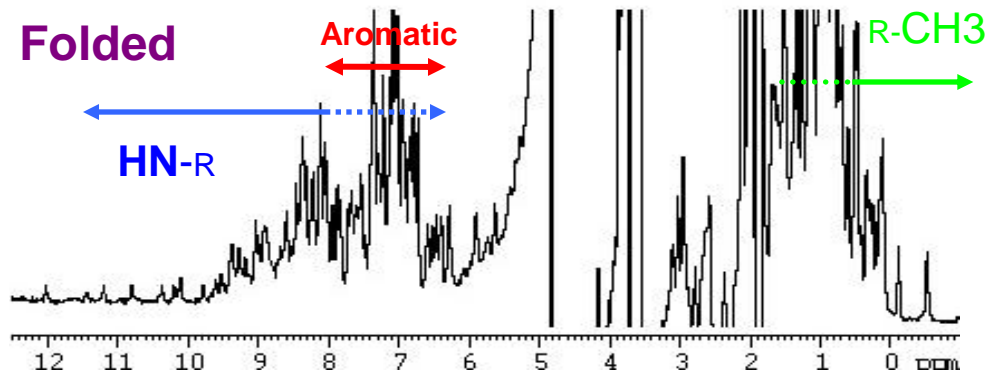
Dynamic Light Scattering

Does the protein concentrate?
Thermostability
Differential Scanning Fluorimetry (DSF)
-> Change buffer conditions

Specific Enzymatic activity
Limited proteolysis
Nuclear Magnetic Resonance
-> Ligand binding
-> Binding partners (additional domains)

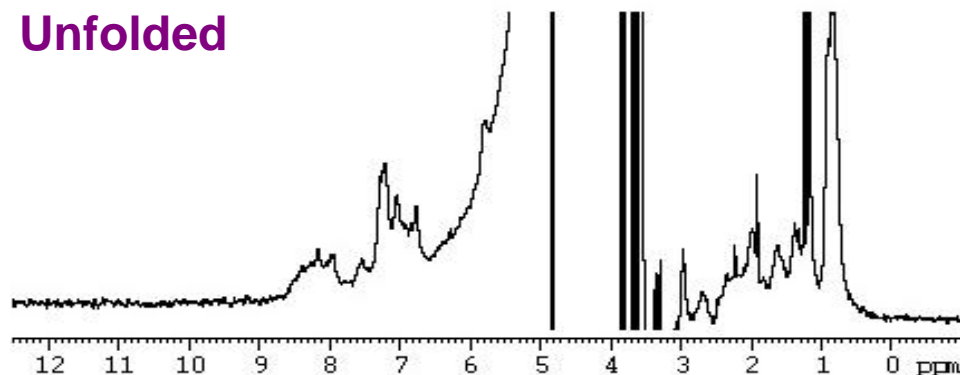
-> Natural variants
-> Engineering/surface mutations
-> Binding partners
-> Seeding
-> More screens / narrow screens

Some background: Protein folding as seen by NMR



MMP12; Mw: 18 kDa

- + chemical dispersion in methyl region
- + chemical dispersion in amide region
- + sharp resonances (=unique conformation)



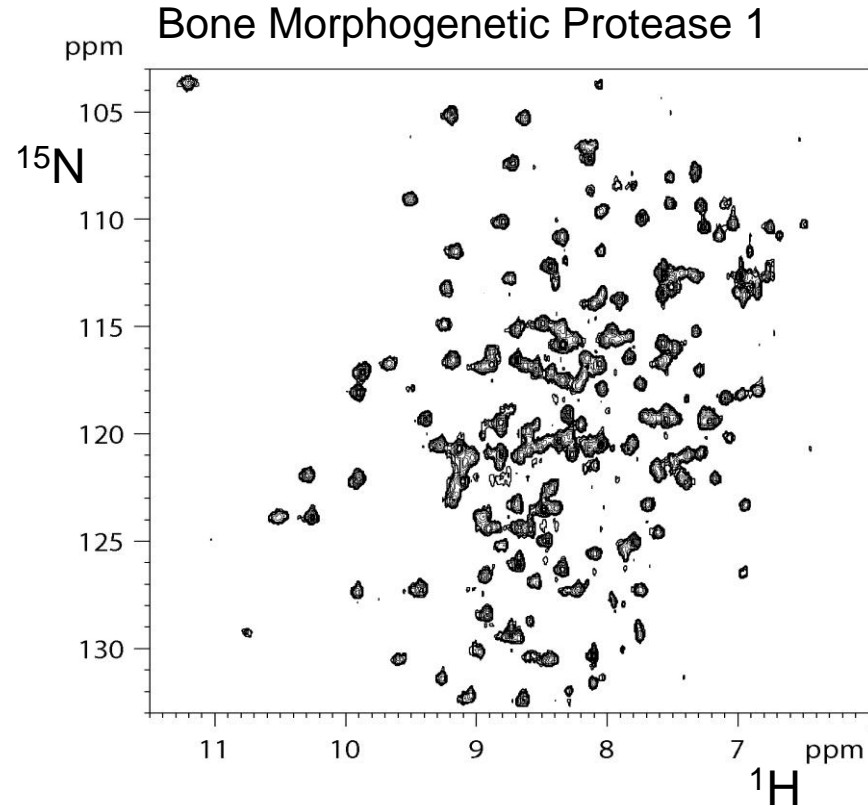
MMP9; Mw: 18 kDa (no fibronectin)

- poor chemical dispersion in methyl region
- poor chemical dispersion in amide region
- broad resonance (=multiply conformations)

- Simple experiment: requirements ~0.2mg of protein, Mw < 30kDa
 - Protein can be reused after NMR (concentrate for crystallization?)
- Folding assessment by NMR: situation often not black or white

Refolded protein: Tolloid like protease BMP1 and TLL1

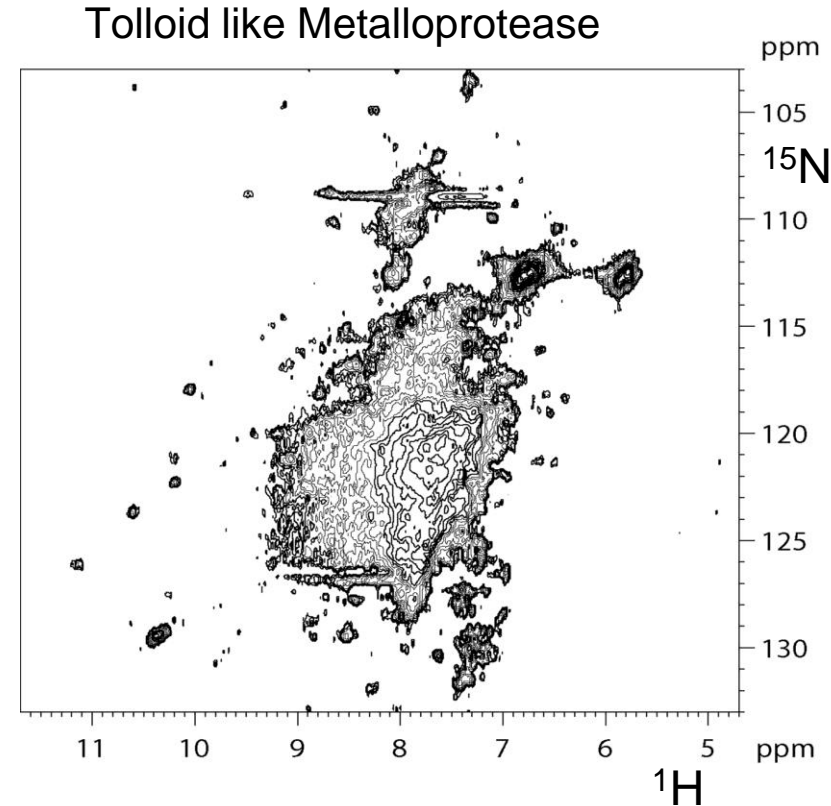
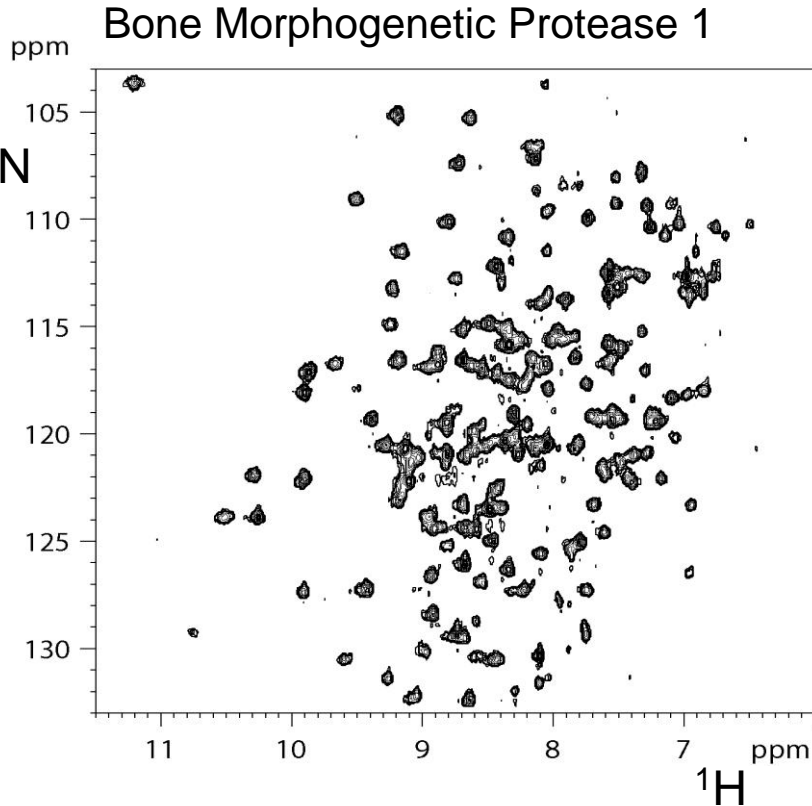
2D HSQC spectrum: folding seen by NMR



- Requirements: ~0.5 mg of ^{15}N labeled protein (*E. Coli* expression)
 - Mw < 40 kDa
 - Higher resolution (each peak corresponds to backbone amide)
 - BMP1 has 200 amino acids

Refolded protein: Tolloid like protease BMP1 and TLL1

Literature: Mac Sweeney et al., *J. Mol. Biol.* 384 (2008)



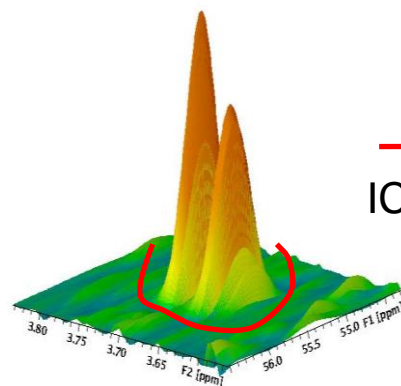
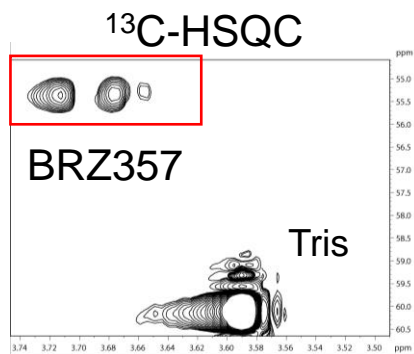
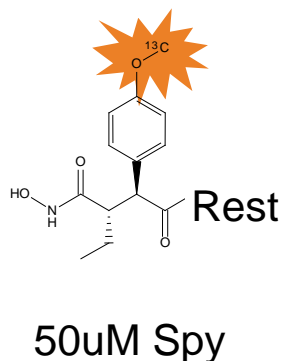
■ Refolded proteins:

- Sequence identity: 87%
- Interpretation of TLL spectrum: two species (only minor fraction well folded)

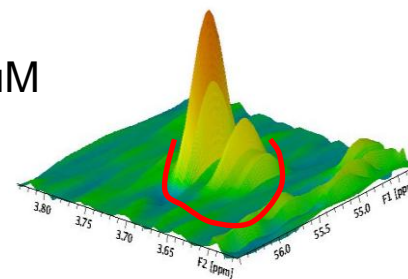
Refolded protein: Tolloid like protease BMP1 and TLL1

two species: quantification

- Quantification of the protease preparation by biochemical assay
 - Specific activity in U per mg (1 μmol substrate turned over per minute per mg of enzyme)
 - Active-site titration (molarity by active site titration / molarity as protein)
 - In our experience this is not straightforward: potent inhibitor required
- Active site titration by NMR is very simple
 - Spy molecule with $K_d < 1 \mu\text{M}$ (^{19}F or ^{13}C nice to have)

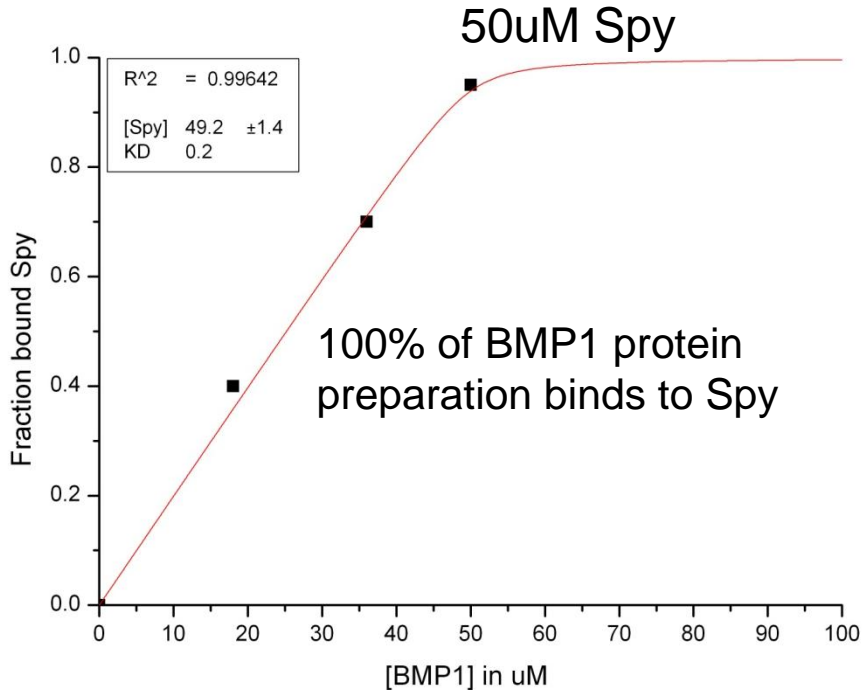


$\text{IC}_{50} = 0.2 \mu\text{M}$

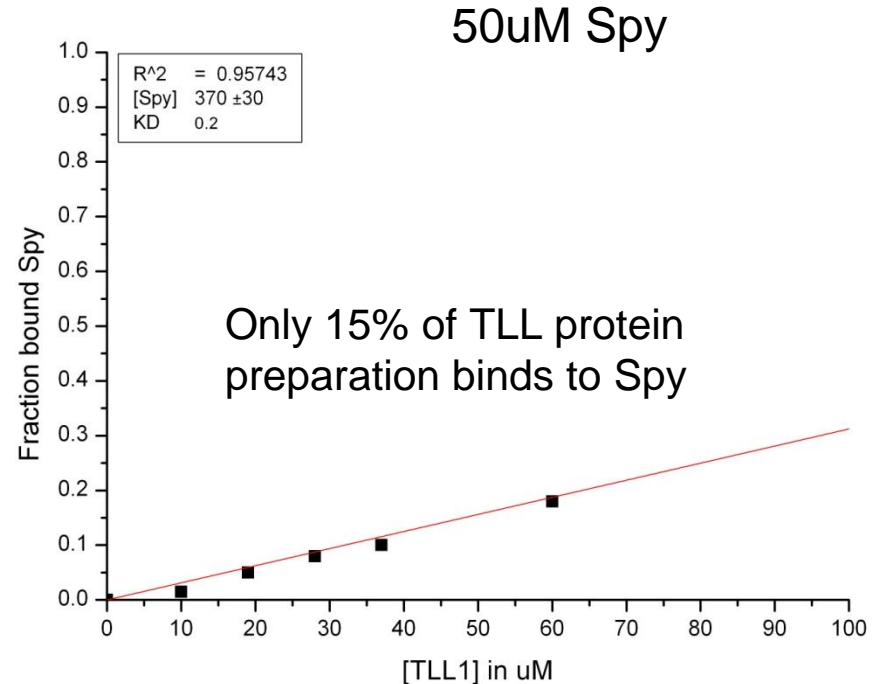


Refolded protein: Tolloid like protease BMP1 and TLL1

ONLY 15% OF TLL1 IS FOLDED



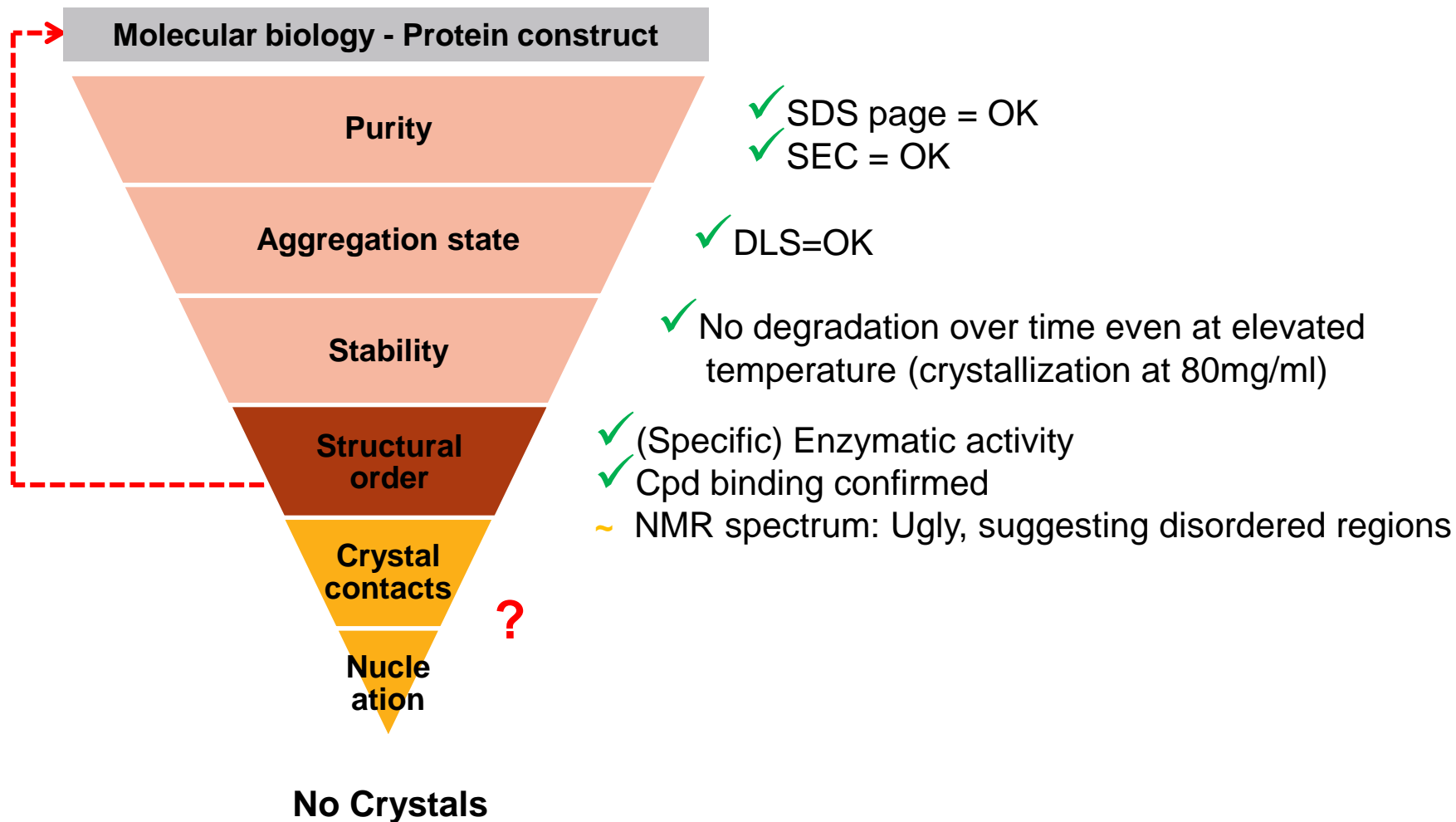
- Xtal grow spontaneous
 - In Tris buffer at 4 °C
 - Resolution 1.3 Å



- Xtal obtained in PEG screen
 - Screening conditions 11mg/ml at RT (1.4 Å)
 - ⇒ Folded protein crystallizes out?
 - ⇒ Seen this more often for refolded proteins
 - presence of detergent, glycerol, arginine?
 - mix of disulphide bonds?

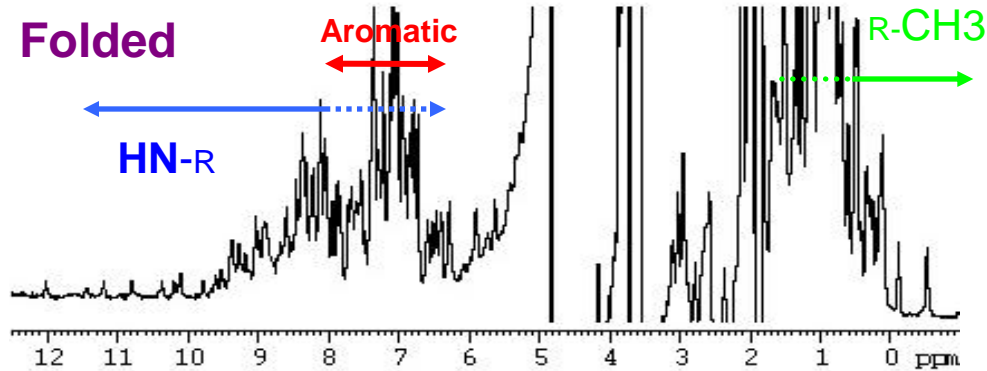
Rather difficult case: Dengue Virus Protease

D'Arcy et al., Acta Cryst. F62 (2006) and Erbel et al., Nat. Struct. Biol. 13 (2006)



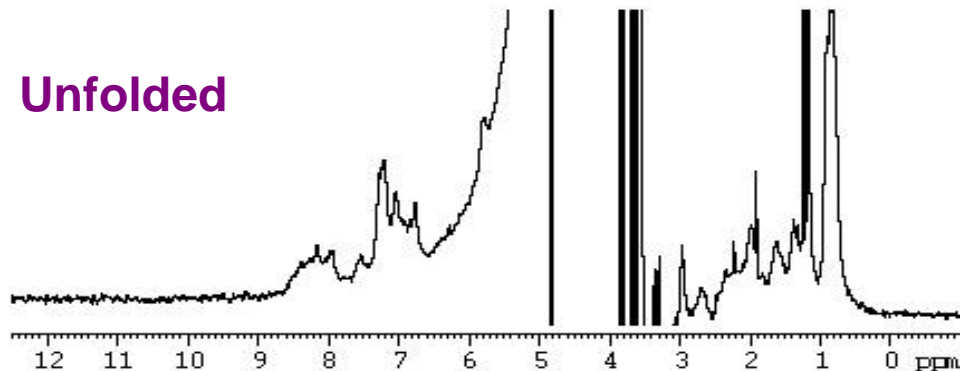
Dengue virus protease

First assessment of structural order



MMP12; Mw: 18 kDa

- + chemical dispersion in methyl region
- + chemical dispersion in amide region
- + sharp resonances (=unique conformation)



MMP9; Mw: 18 kDa (no fibronectin)

- poor chemical dispersion in methyl region
- poor chemical dispersion in amide region
- broad resonance (=multiply conformations)

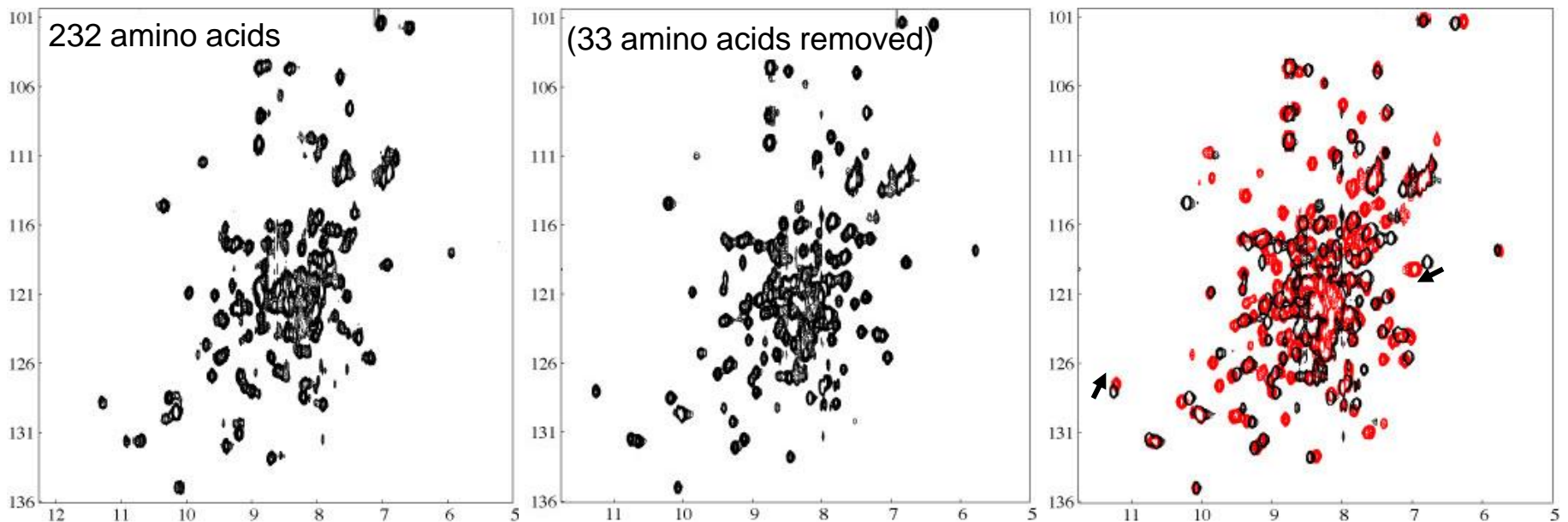
Reducing structural disorder of Dengue protease

improve construct, ligand binding, ...

49 NS2B 95 GGGGSGGGG 1 NS3pro 185

49 NS2B 95 EVKKQR↓AG 17 NS3pro 170

Upon addition of Bz-Nleu-Lys-Arg-Arg-H

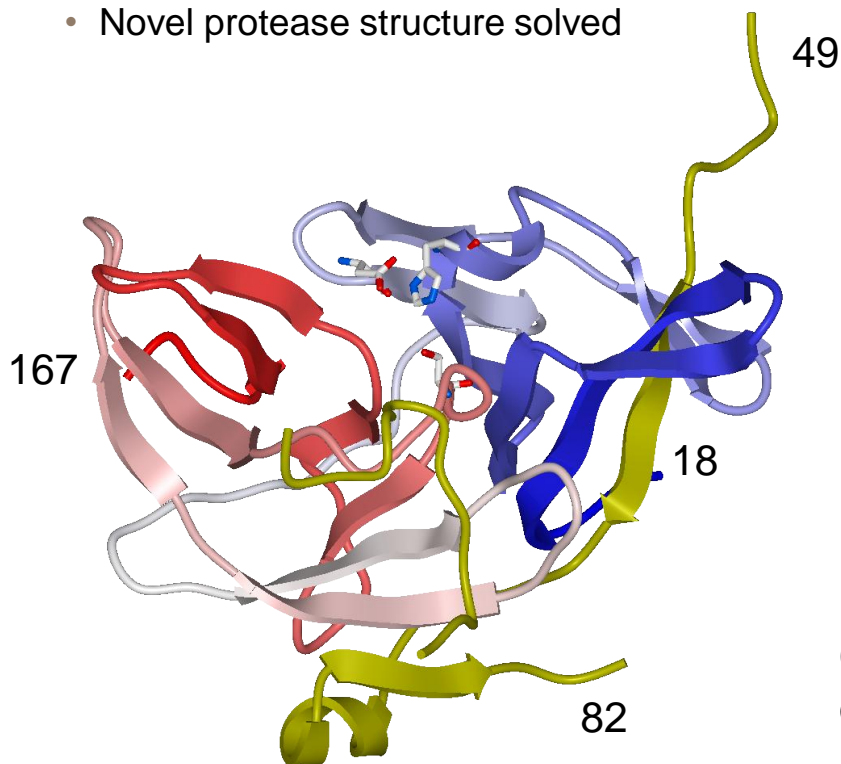
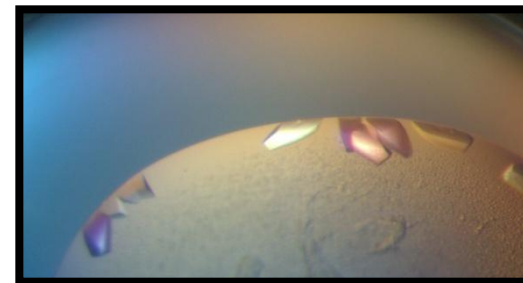


- Construct optimized (further reduction makes construct unstable / inactive)
 - Still NMR spectrum not great
 - Ligand binds and affects protein dynamics / structure
- => No crystallization conditions found with the short construct**

Rather difficult case: Dengue Virus Protease

Well diffracting crystals obtained with Lysine mutants

- One crystallization condition found out of screens
 - Diffraction is good (~1.7 Å)
 - Novel protease structure solved



Visible in structure

49 NS2B 80

18 NS3pro 167

Construct Lys-> Arg

49 NS2B 95 GGGGSGGGG 1 NS3pro 185

Construct optimization

49 NS2B 95 EVKKQR↓AG 17 NS3pro 170

Blue-red: NS3 protease
 Yellow: NS2B cofactor
 Cofactor contributes β-strand to N-terminal β-barrel

Rather difficult case: Dengue Virus Protease

we feel really smart ...

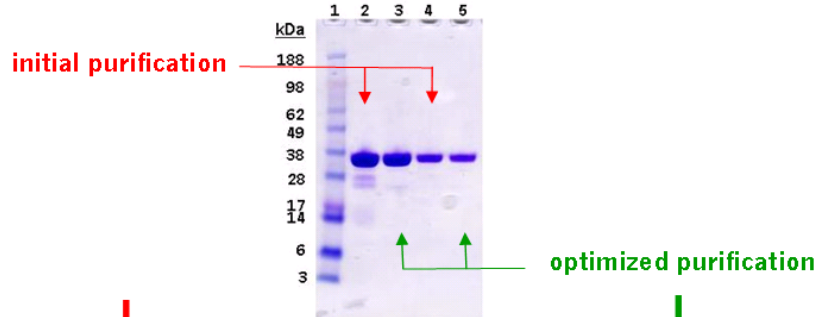
- Protein construct: indeed very dynamic
 - Quickly observed by NMR
 - Optimal truncated construct made (see crystal structure)
 - However, not the critical factor for crystallization of Dengue Protease
- Introducing crystal contacts (7x Lys->Arg) did the trick !?
 - Repetitive purification of Dengue protease Lysine mutants allow to optimize purification protocol
affinity – thrombine digestion at 4 °C - anion exchange - size exclusion chromatography
- Better stop the story now, but for once we took the time to look back
 - Can we obtain crystals with original construct using the improved purification protocol?



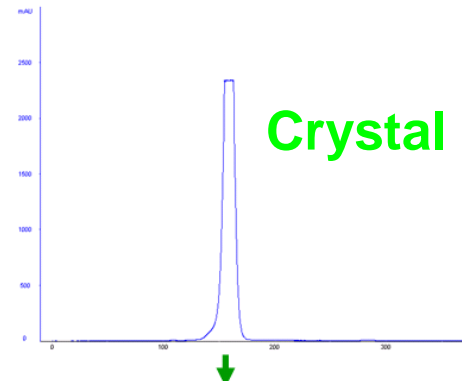
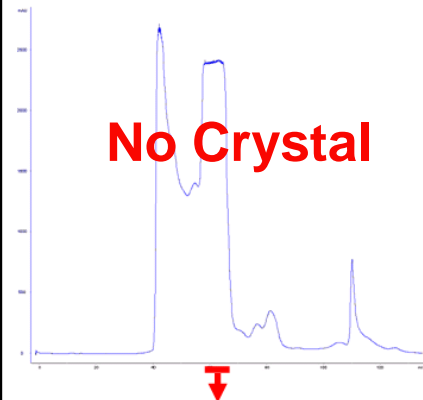
Dengue protease purification

original construct - two different purifications

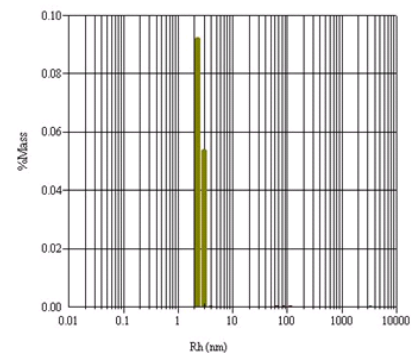
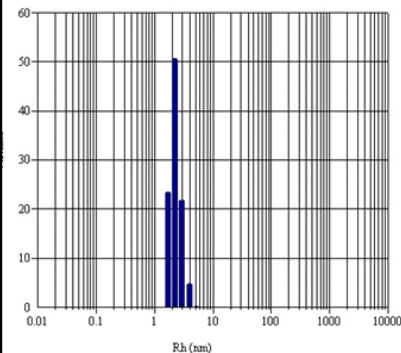
SDS gel



SEC profile



DLS



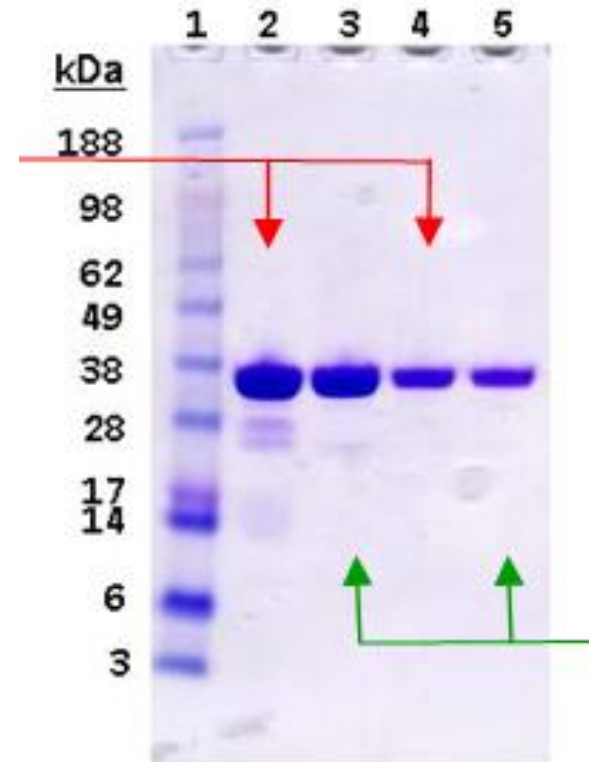
Rather difficult case: Dengue Virus Protease

our learning, take home message

- Critical for crystallization of Dengue protease: protein purification
 - How clean needs a protein preparation be?
 - SDS page does not show DNA / RNA / carbohydrates

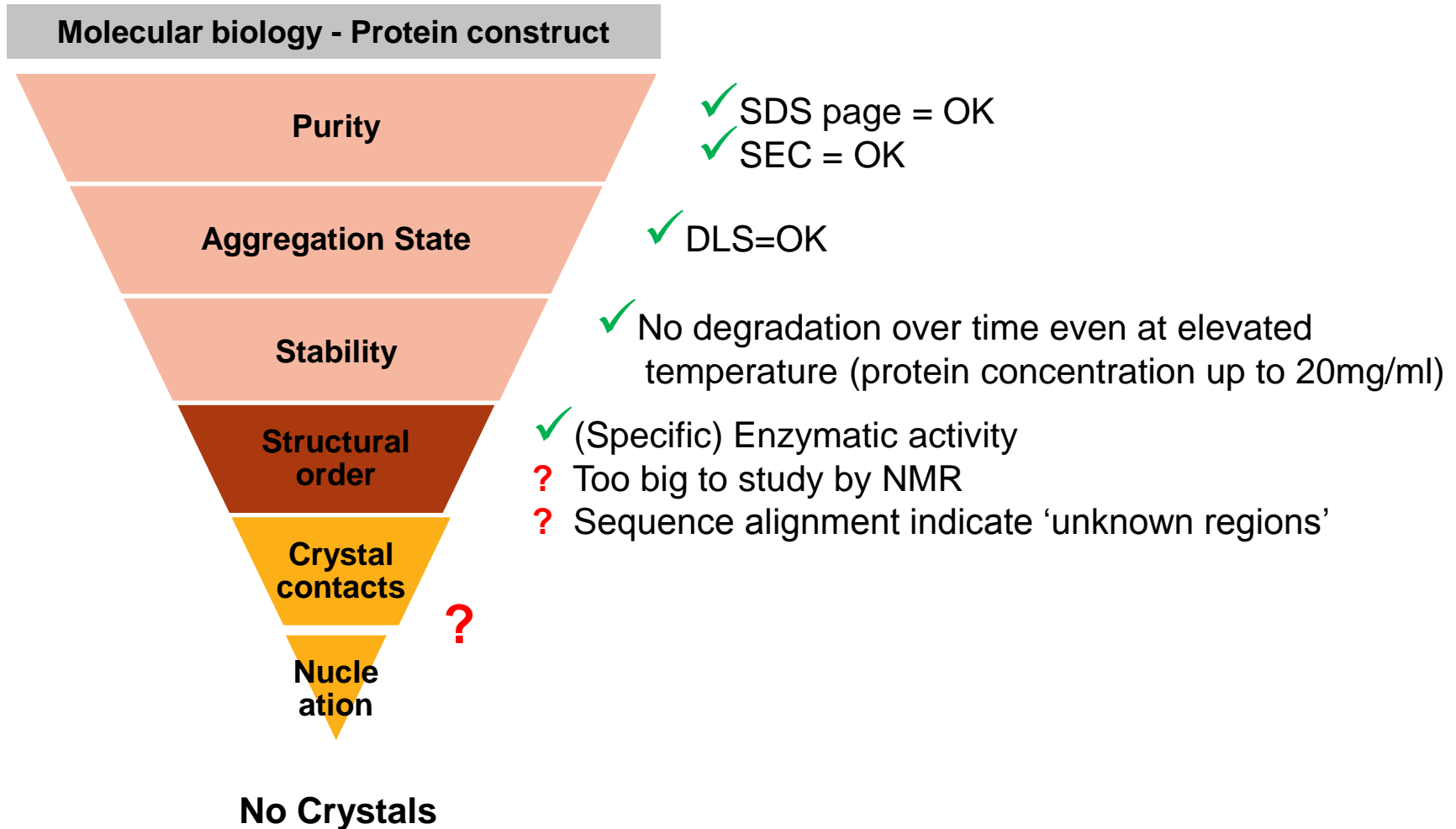


- Standard Operating Procedure: 3 purification steps
 - Affinity purification (typically Ni-NTA)
 - Ionic exchange (or HIC) (=removal of DNA/RNA)
 - Size exclusion: only polishing step (=analytics, buffer exchange)



Rather difficult case: Cysteine protease

no structural information: 56 kDa multidomain protein required for biochemical activity



Rather difficult case: Cysteine protease

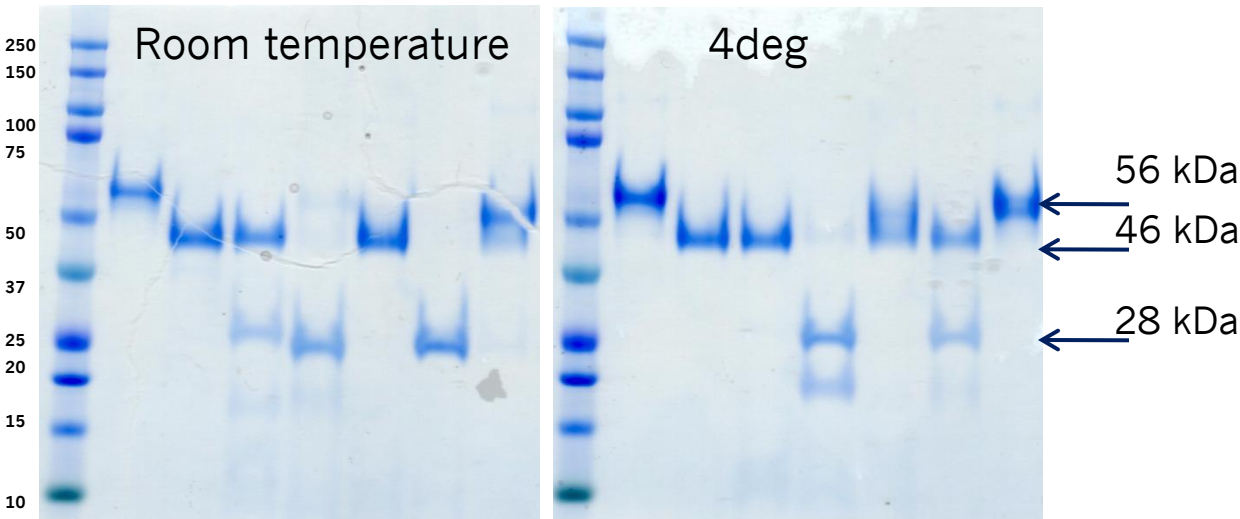
limited proteolysis: an efficient way to probe the structural order of protein

■ Limited proteolysis:

- Requirement ~ 1mg of protein
- Proteolysis profile of up to 18 different proteases (Hampton kit, see F. Villard for details)
- Digestion O/N at room temperature and 4 °C, typically at 1:1000 dilution
- Analysis cleavage product by SDS page and LC-MS

kDa

1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8



1. Marker
2. Control
3. O/N with α -chymotrypsin
4. O/N with trypsin
5. O/N with elastase
6. O/N with papain
7. O/N with subtilisin
8. O/N with EndoGlu-C

Rather difficult case: Cysteine protease

truncated construct: only the first step.

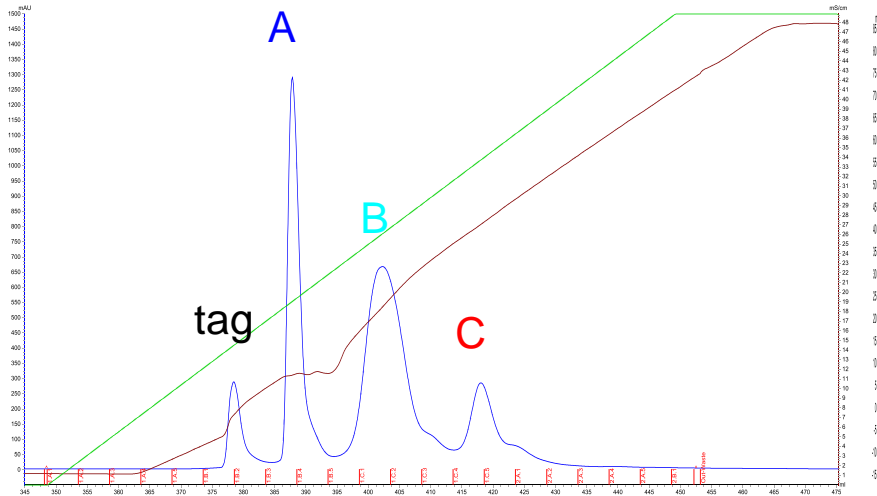
- Limited proteolysis – option 1:
 - Size exclusion chromatography to polish digested protein
 - Crystallization trials with digested protein
 - ⇒ fastest way to change protein construct
 - Crystals only obtained in presence of substrate analogue
 - substrate induces dimers
 - ⇒ no use for drug discovery

- Limited proteolysis – option 2:
 - Generate new constructs based on results of limited proteolysis
 - Issue: protein only expresses as monomer (expression system Baculo virus expression)

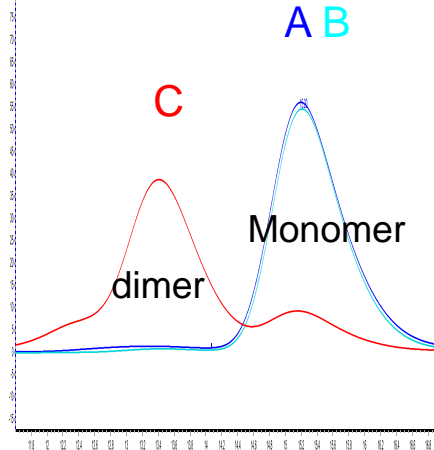
- Truncated construct (46kDa) revived interest in biophysical studies
 - Protein expression in *E.coli* explored
 - Soluble expressed protein in high yield obtained for truncated construct

Rather difficult case: Cysteine protease purification of truncated construct

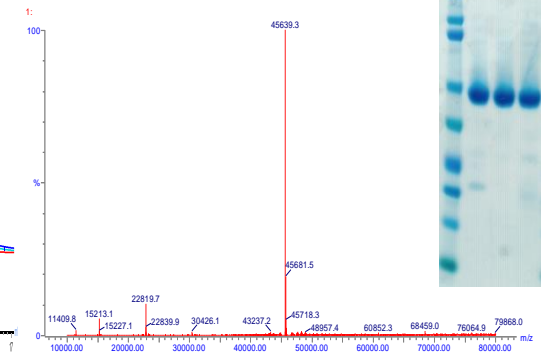
IEX on Source 15Q



(Analytical) SEC



45639 Da



- SOP purification: interesting IEX profile

- Observations:

- No crystals obtained with monomeric protein (used for binding studies by NMR/SPR)
- Intensity of dimeric peak C depends on protein yield
 - *E.coli* > *Baculovirus* expression system (dimeric signal overlooked)

Dimeric form: crystals obtained in absence of substrate analogue

Rather difficult case: Cysteine protease

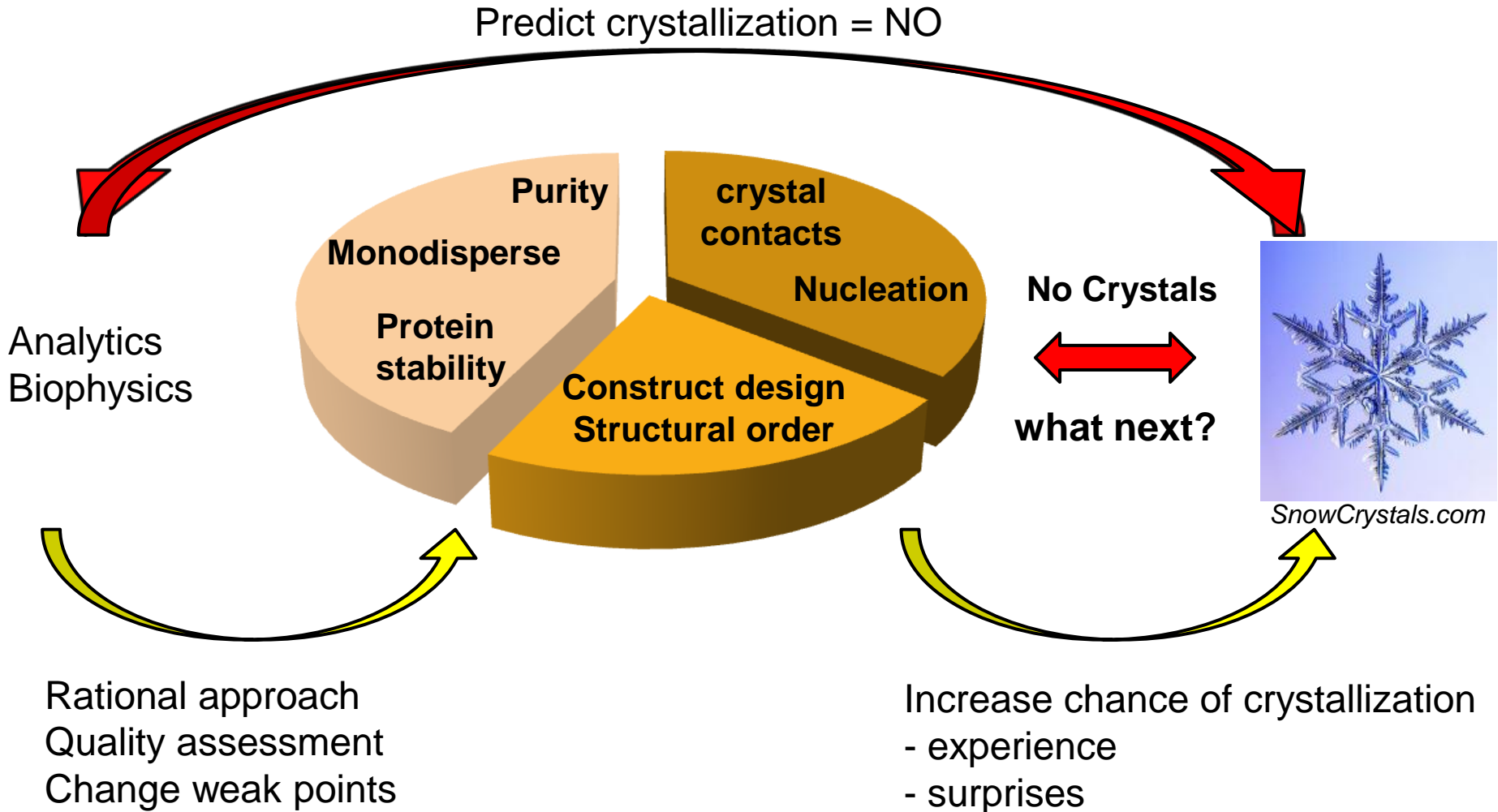
structural of novel protease solved

Key steps:

- Improving structural order:
 - Limited proteolysis: C-terminal truncation of ~100 amino acids
 - Focus on dimeric form
- Expression and Purification:
 - Changing the expression system
 - Ion Exchange chromatography

Which parameters are critical for protein crystallization?

Can we measure those parameters? Can we affect those parameters?



Biophysical characterization for compound binding

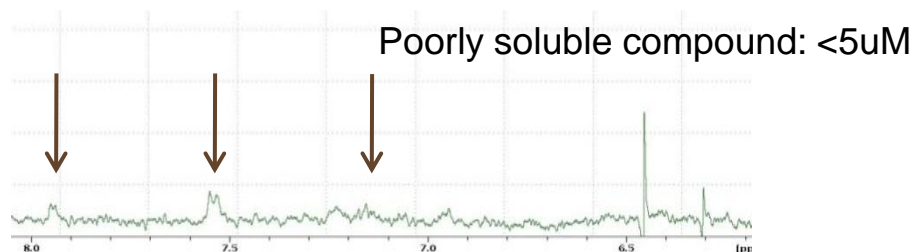
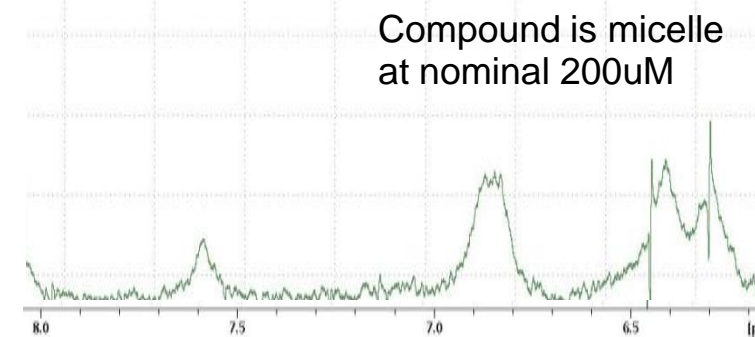
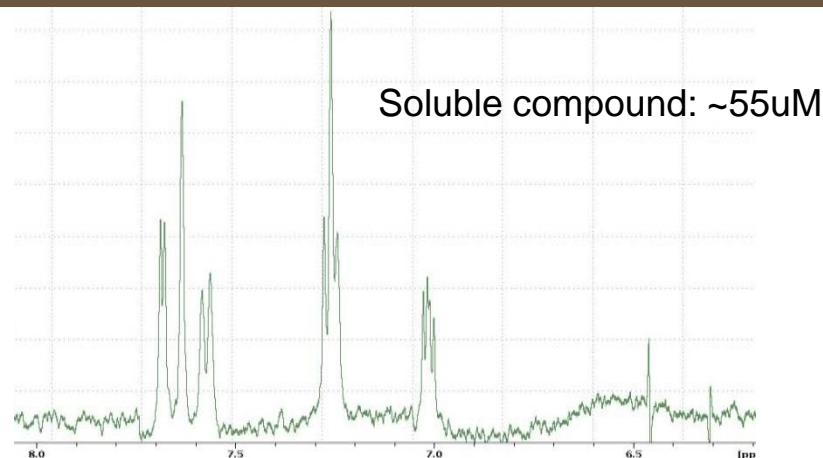
support crystallization and drug discovery

- Compound binding to support crystallization
 - Compound binding can alter protein conformation and dynamics.
 - This can affect crystallization behavior.
- Drug discovery: protein crystallization in industry requires complex structures
 - IC_{50} sufficient for compounds selection?
 - Early HTS hits, Fragment based screening hits
- Critical factors to select compounds for crystallization:
 - Solubility - NMR application
 - Binding (yes/no) and Binding pocket - NMR application
 - Affinity (K_d), Specific or Unspecific binding - SPR application

Biophysical characterization for compound binding

solubility determination in buffer

- Compound solubility by NMR
 - reliable method in our experience
 - Acceptable throughput (30min / sample)
 - Solubility can be determined >5uM
 - Good range of buffers (including detergents)
- Impact of solubility determination
 - ratio IC_{50} vs solubility
 - problematic if solubility < IC_{50}
 - IC_{50} valid? What is the assay measuring?
 - 'absolute' solubility
 - Prioritization of compounds
 - Design co-crystallization experiments
- Compound solubility: low tech, high impact



Biophysical characterization for compound binding

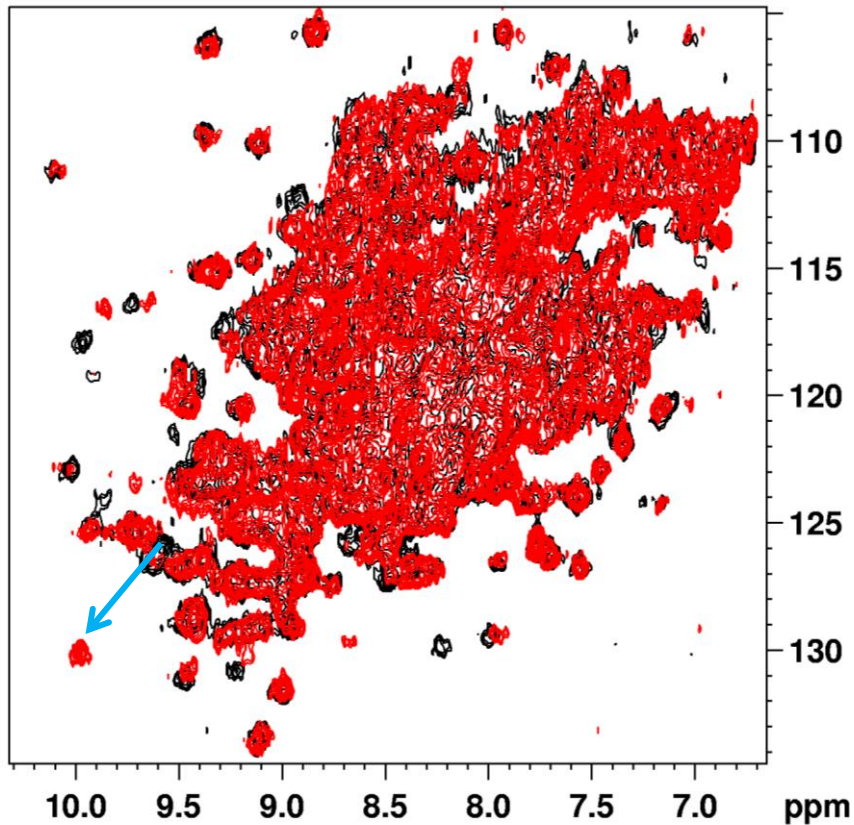
binding (yes / no) and where?

- Bias toward NMR spectroscopy for determining compound binding
- NMR ligand based methods like STD, Waterlogsy, relaxation filtered exp. (also as ^{19}F NMR)
 - Fine screening technologies, but not good enough to (de)validate hits for crystallization
 - Neat applications as reporter set-up (provides K_d information)
- Only way: NMR protein observed
 - Very low false positives (experimental mistakes like changes in pH and DMSO concentration)
 - Very low false negatives (low solubility in combination with weak affinity)
 - Protein Mw <50 kDa
- Big trick box to simplify NMR spectra – amino acid selective labeling
 - Provides binding site information

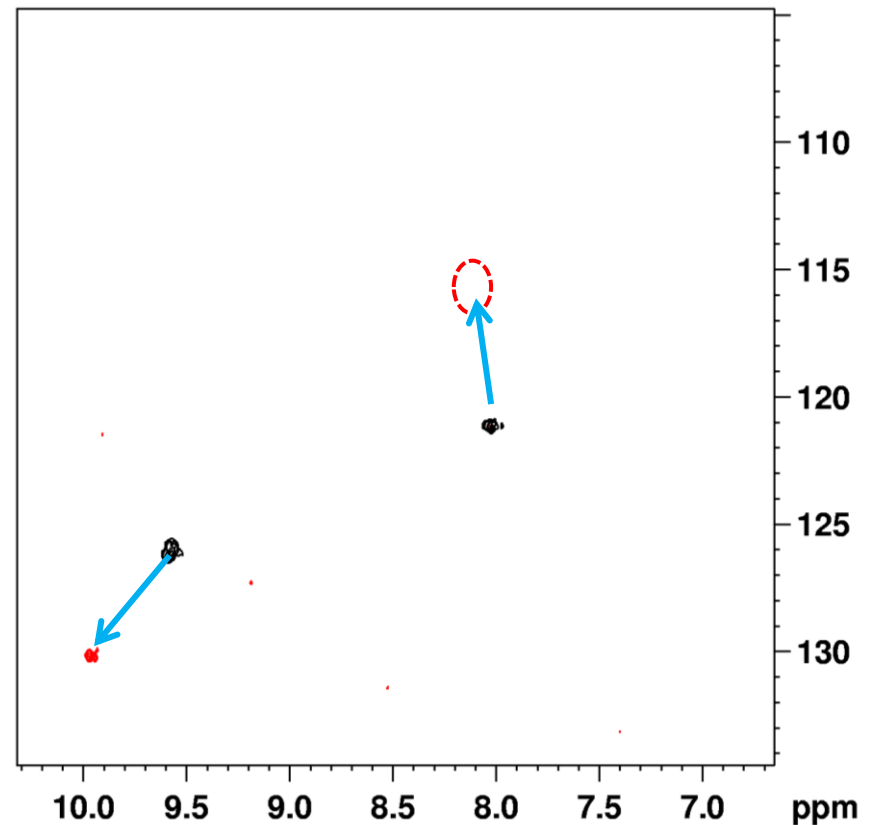
Biophysical characterization for compound binding

binding (yes / no) and where?

All backbone ^{15}N -labeled



Only Trp ^{15}N labeled



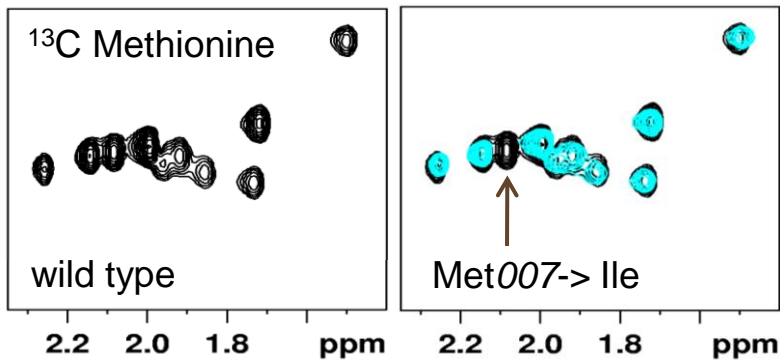
■ Amino acid selective labeling

- BEV and *E.coli* (in minimal growth medium supplemented with amino acids)
- Reduces signal overlap, increased resolution

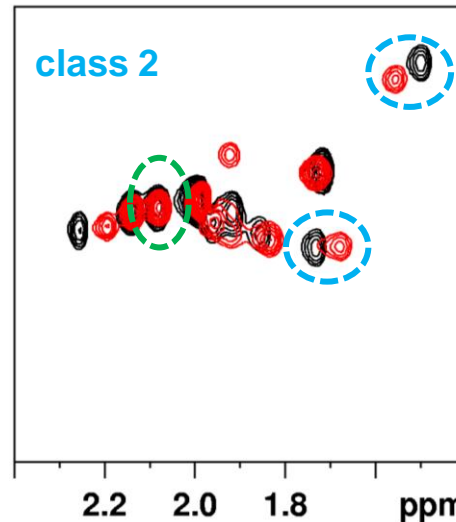
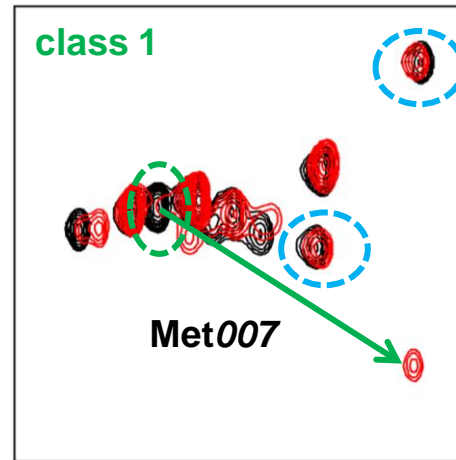
Biophysical characterization for compound binding

binding (yes / no) and where?

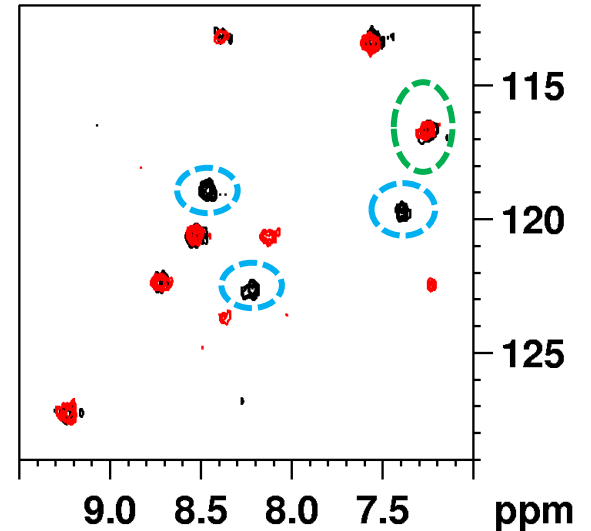
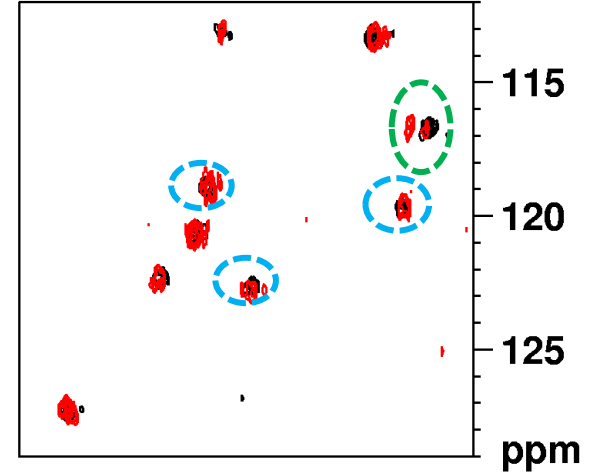
- Chemical shift perturbation:
 - Binding confirmed for both compounds
- Different chemical shift pattern:
 - Class 1 and class 2 compounds bind in different pocket
- Assignment of amino acid / pocket:
 - Mutation of labeled amino acid



Methionine ¹³C labeled



Cysteine ¹⁵N labeled



Biophysical characterization for compound binding

Affinity (K_d), Specific or Unspecific binding

SPR = Slightly Plausible Results

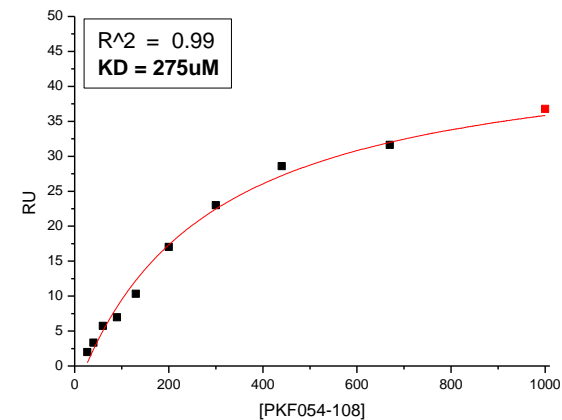
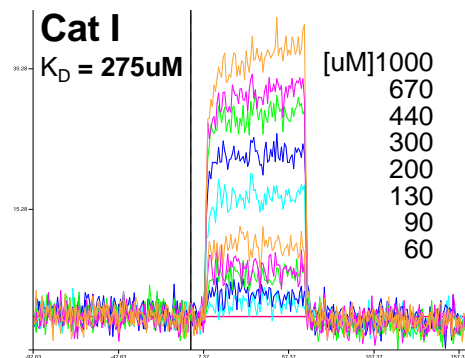
Category 1:

- NMR binding (yes)
- SPR: 1-1 binding with K_D

>60%

⇒ NMR and SPR aligned

⇒ First priority for crystallization

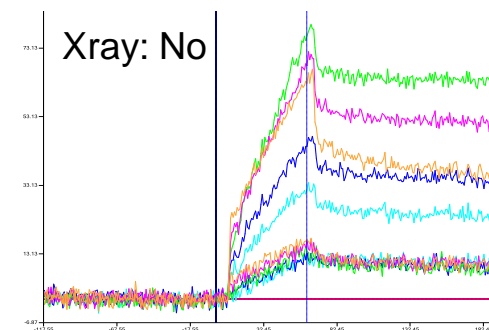
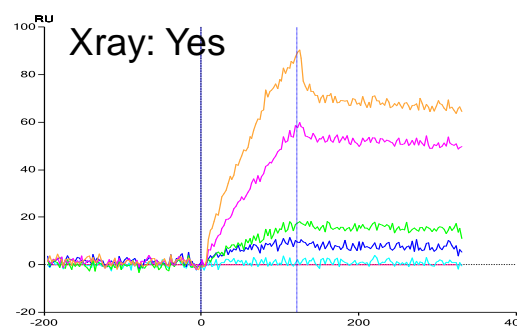


Category 2:

- NMR binding (yes)
- SPR: response
 - slow kinetics, super stoichiometric
 - Binding to surface?

~20%

⇒ Low priority for crystallization



Lessons learnt:

- SPR: strict filter for crystallization (solubility, cpd aggregation, non specific binding, affinity)?
- Combination of NMR and SPR links robustness, sensitivity and hit characterization (reversed strategy?)

Structural Science Unit of Protease Platform

Allan d'Arcy

Frederic Villard

Martin Renatus

Arnaud Decock

Aengus MacSweeney

Nicola Hughes

Daniela Vinzenz

Simon Ruedisser

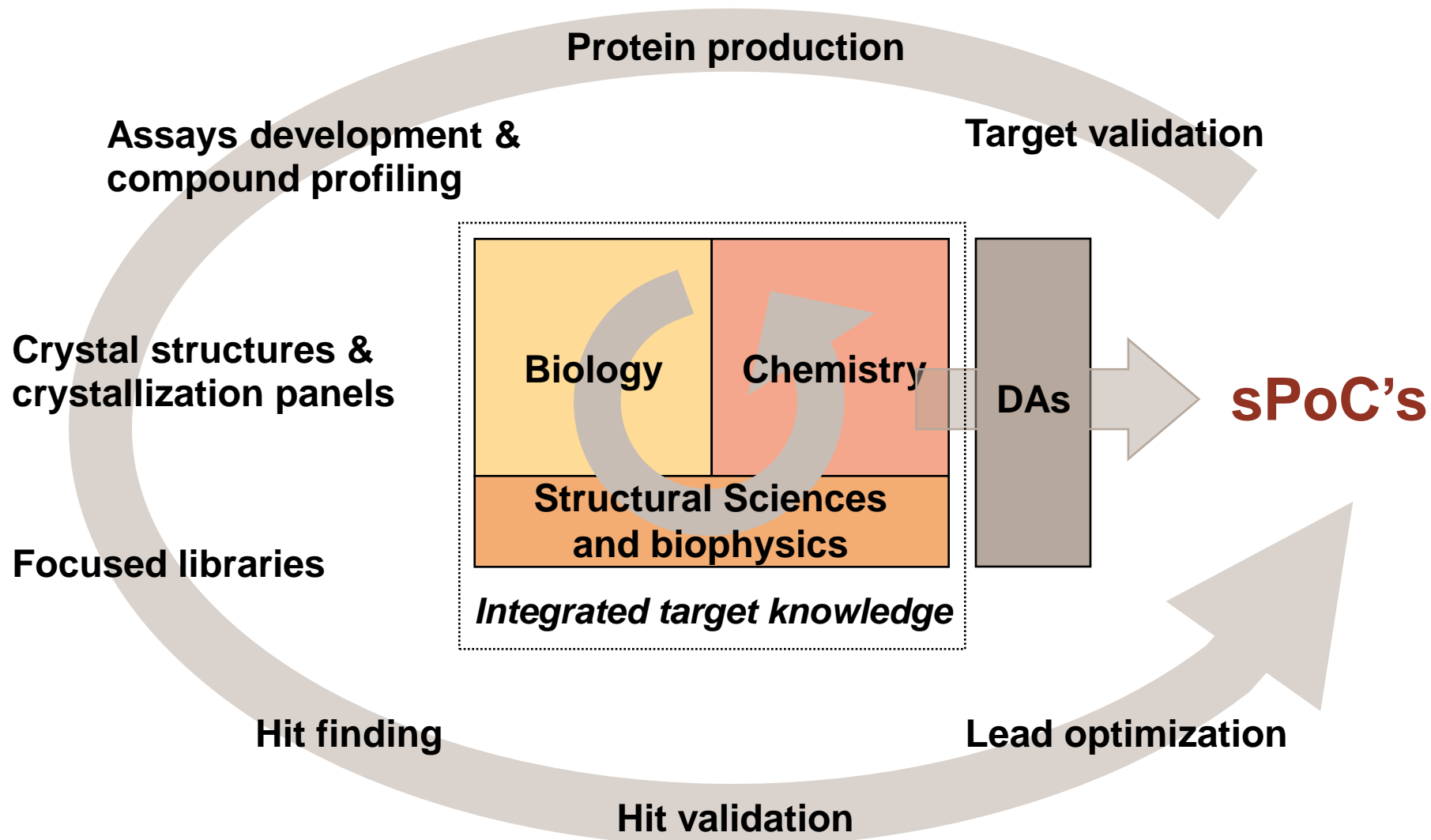
Nikolaus Schiering

Christian Wiesmann



Expertise Platform Proteases (EPP)

an integrated approach



Some theory

solubility vs IC50 – specific vs unspecific

Compound soluble / insoluble

