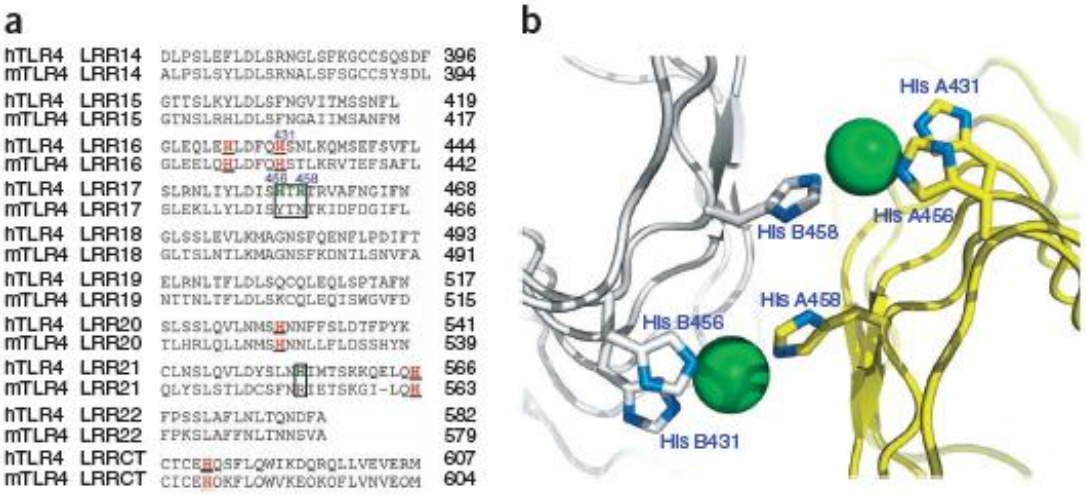
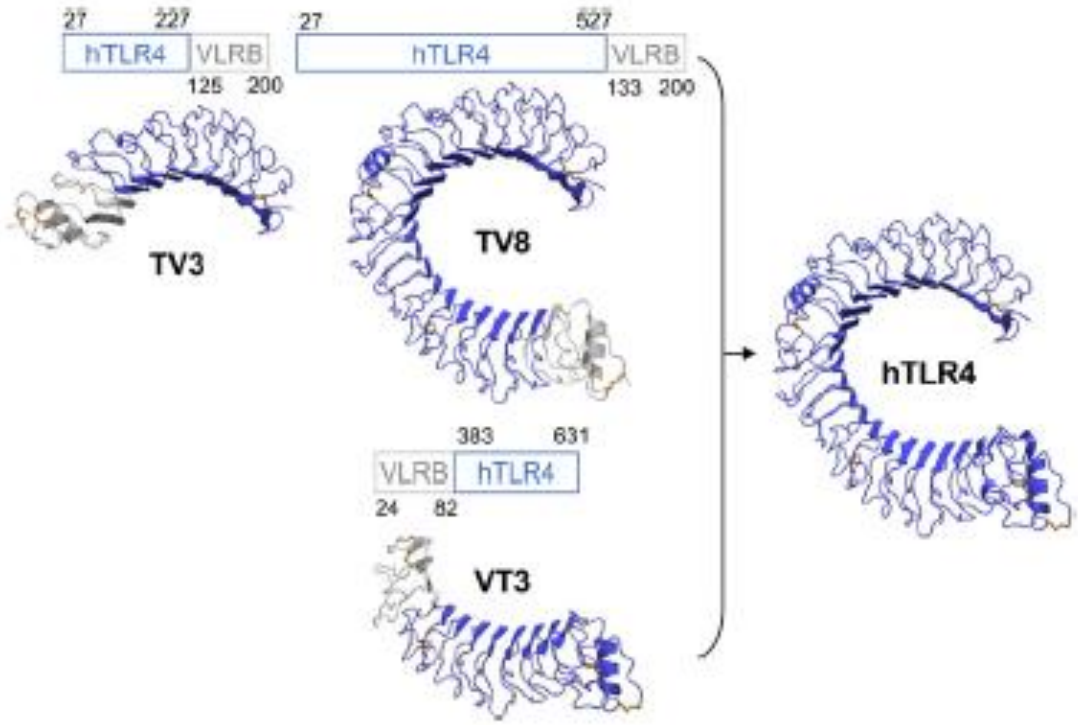
The Precipitator

We created a cellular, self-replicating purification device for His-tagged proteins. It is a completely artificially created fusion protein, which consists of a repeating LRRNT motif domain, coordinating Ni2+ Ions on its surface capped on N and C terminal end by hagfish sequence of a similar LRRNT motif. A second domain binds polystyrene surface. It is called the Precipitator.

The principal mechanism is comparable to Ni-NTA columns. Our Precipiator protein binds on the surface of the tube, presenting the chelated Nickel ions. Free binding sites of the Nickel ions are then exposed, so that a His-tagged protein can attach to them. Cells expressing a His-tagged protein can be dissolved by the light inducible lysis device. Subsequently, when the lysate is taken up with a serological pipette coated with the Precipitator protein, the His-tagged proteins bind to it. Cell debris is then washed off, while the His-tagged protein stays and is eluted afterwards, in the same fashion as done in Ni-NTA columns with imidazole solutions, increasing in concentration. The His-tagged protein is finally captured in a distinct fraction.



Prof. Martin Nickel Allergy – TLR4



- **TLR-4 (3FXI) structure paper**

- N-Terminal Hagfish 2z66
    - CPSRCSCSGTEIRCNSKGLTSVPTGIPSSATRLELESNKLQSLPHGVFDK
    - LTQLTK
    - Consensus: xLxxLxxLxL
- C-Terminal Hagfish 2z62
- LKELALDTNQLKSVPDGIFDR
- LTSLQKIWLHTNPWDCSCPRIDY
- LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP
- LxxLxLxxNxLx Consensus

Bacterial LRR Consensus:

LxxLxLxxNxLxxLPxxLPxx

TLR4

LxxLxLxxNxLxxLxxxxFxxLxx

3cvr bacterial ligase

LEVLDADNNQLTSLPE.LPASL

FSELQLNRLNLSSLPDNLPPQ

ITVLEITQNALISLPELPAS

LEYLDACDNRLSTLPELPAS

LKHLDVDNNQLTXLPELPAL

LEYINADNNQLTXLPELPTS

LEVLSVRNNQLTFLPELPES

EIFFRCRENRITHIPENILSLDP

Extracted ideal consensus, by consensus sequence and 3d rational analysis:

LEVLDVSNNQLTSLPDNLPAS

Rational1:

LxALHCSxNxLxSLPxxLPxx LxHLACSxNxLxSLPxxLPxx

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPDGVFDK

LTALHKSNNQLTSLPDNLPAS

LEHLAVSNNQLTSLPDNLPAS

LEALHVSNNQLTSLPDNLPAS

LEHLAVSNNQLTSLPDNLPAS

LEALHVSNNQLTSLPDNLPAS

LEHLAVSNNQLTSLPDNLPAS

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLQTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

Rational2:

LxxLxCSxNHLHSLPxxLPxx

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPHGVFDK

LTQLTKSNNHLHSLPDNLPAS

LEVLDVSNNHLHSLPDNLPAS

LEVLDVSNNHLHSLPDNLPAS

LEVLDVSNNHLHSLPDNLPAS

LEVLDVSNNHLHSLPDNLPAS

LEVLDVSNNHLHSLPDNLPAS

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLHTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

Rational3:

LxxLxCSxNxLxSLPHHLPxx

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPHGVFDK

LTQLTKSNNQLTSLPHHLPAS

LEVLDVSNNQLTSLPHHLPAS

LEVLDVSNNQLTSLPHHLPAS

LEVLDVSNNQLTSLPHHLPAS

LEVLDVSNNQLTSLPHHLPAS

LEVLDVSNNQLTSLPHHLPAS

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLHTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

Rational4:

LxxLxCSxNxLxSLPxHLPHx

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPHGVFDK

LTQLTKSNNQLTSLPDHLPHS

LEVLDVSNNQLTSLPDHLPHS

LEVLDVSNNQLTSLPDHLPHS

LEVLDVSNNQLTSLPDHLPHS

LEVLDVSNNQLTSLPDHLPHS

LEVLDVSNNQLTSLPDHLPHS

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLHTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

Rational5:

LHxLxCSxNxLxSLPxxLPxH

LHVLDVSNNQLTSLPDNLPAH

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPHGVFDK

LHQLTKSNNQLTSLPDNLPAH

LHVLDVSNNQLTSLPDNLPAH

LHVLDVSNNQLTSLPDNLPAH

LHVLDVSNNQLTSLPDNLPAH

LHVLDVSNNQLTSLPDNLPAH

LHVLDVSNNQLTSLPDNLPAH

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLHTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

Rational6:

LHHLxCSxNxLxSLPxxLPxx

LHHLDVSNNQLTSLPDNLPAS

CPSRCSCSGTEIRCNSKGLTSVPTGIPSS

ATRLELESNKLQSLPHGVFDK

LHHLTKSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LHHLDVSNNQLTSLPDNLPAS

LKELALDTNQLKSVPDGIFDR
LTSLQKIWLHTNPWDCSCPRIDY
LSRWLNKNSQKEQGSAKCSGSGKPVRSIICP

After evaluating the C-Scores of I-TASSER we chose the three best results and reverse translated them, using http://www.bioinformatics.org/sms2/rev_trans.html

Rational1

MW=27164.440000000006

MCPSRCSCSGTEIRCNSKGLT
SVPTGIPSSATRLELESNKLQ
SLPDGVFDKLTALHKSNNQLT
SLPDNLPASLEHLAVSNNQLT
SLPDNLPASLEALHVSNNQLT
SLPDNLPASLEHLAVSNNQLT
SLPDNLPASLEALHVSNNQLT
SLPDNLPASLEHLAVSNNQLT
SLPDNLPASLKALHLDTNQLK
SVPDGIFDRLTSLQKIWLQTN
PWDCSCPRIDYLSRWLNKNSQ
KEQGSAKCSGSGKPVRSIICP

Results for 251 residue sequence "Untitled" starting "CPSRCSCSGT"

>reverse translation of Untitled to a 753 base sequence of most likely codons.

Gaattcgcggccgcttctag

ATGtgcccgagccgttgcagctgtagcggcaccgaaattcgctgcaacagcaaaggcctgaccagcgtgccgaccggcattccgagcagcgcgacccgcctggaactggaaagca
acaaactgcaaagcctgccggatggcgtgtttgataaactgaccgcgctgcataaaagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaacatctggcggt
gagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaagcgctgcatgtgagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaa
catctggcggtgagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaagcgctgcatgtgagcaacaaccagctgaccagcctgccggataacctgccggc
gagcctggaacatctggcggtgagcaacaaccagctgaccagcctgccggataacctgccggcgagcctgaaagcgctgcatctggataccaaccagctgaaaagcgtgccggatg
gcatttttgatcgcctgaccagcctgcaaaaaatttggctgcaaaccaacccgtgggattgcagctgcccgcgcattgattatctgagccgctggctgaacaaaaacagccagaaagaac
agggcagcgcgaaatgcagcggcagcggcaaaccggtgcgcagcattatttgcccgTAGTAa

Tactagtagcggccgctgcag

Alignment of our reverse translated sequence and the one optimized by ATG:biosynthetics:

```
                *         *         *         *         *         *         *         *         *         *
  1>ATGtgcccgagccgttgcagctgtagcggcaccgaaattcgctgcaacagcaaaggcctgaccagcgtgccgaccggcattccgagcagcgcgacccgcc>100
  1>ATGTGCCCGAGTCGTTGCAGTTGTTCAGGCACCGAAATTCGCTGCAATAGCAAAGGCTTGACCTCTGTGCCGACCGGTATTCCAAGTTCAGCAACCCGTC>100

                *         *         *         *         *         *         *         *         *         *
101>tggaactggaaagcaacaaactgcaaagcctgccggatggcgtgtttgataaactgaccgcgctgcataaaagcaacaaccagctgaccagcctgccgga>200
101>TTGAGTTAGAATCCAATAAACTTCAGAGCTTGCCCGATGGTGTCTTCGACAAACTGACCGCTCTGCACAAAAGCAACAACCAGCTCACCTCCCTCCTGA>200

                *         *         *         *         *         *         *         *         *         *
201>taacctgccggcgagcctggaacatctggcggtgagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaagcgctgcatgtgagc>300
201>CAATTTGCCGGCGTCCTCGAACAGCTCGCGGTGAGTAACAATCAACTGACCTCGTTGCCTGATAACCTGCCGGCTAGCTTAGAAGCACTGCATGTGTCC>300

                *         *         *         *         *         *         *         *         *         *
301>aacaaccagctgaccagcctgccggataacctgccggcgagcctggaacatctggcggtgagcaacaaccagctgaccagcctgccggataacctgccgg>400
301>AATAACCAACTTACGAGCTTACCAGACAACTTACCTGCCAGCTTAGACCATCTGGCCGTAGCAACAACCAATTAACAAGCCTTCCGGACAACCTGCCGG>400

                *         *         *         *         *         *         *         *         *         *
401>cgagcctggaagcgctgcatgtgagcaacaaccagctgaccagcctgccggataacctgccggcgagcctggaacatctggcggtgagcaacaaccagct>500
401>CGAGCCTTGAGCGCCTTCATGTCTCTAACAACCAGTTGACTTCTTTACCGGATAATCTGCCTGCCTCACTGGAACAGCTTGCGGTTTCAAACAATCAGTT>500

                *         *         *         *         *         *         *         *         *         *
501>gaccagcctgccggataacctgccggcgagcctgaaagcgctgcatctggataccaaccagctgaaaagcgtgccggatggcatttttgatcgcctgacc>600
501>AACGTCGCTGCCGGATAACTTACCGGCGTCACTTAAGCGCTTACATTTGGATACGAATCAGCTCAAAAGCGTGCCTTGATGGCATCTTTGATCGCCTGACA>600

                *         *         *         *         *         *         *         *         *         *
601>agcctgcaaaaaatttggctgcaaaccaacccgtgggattgcagctgcccgcgcattgattatctgagccgctggctgaacaaaaacagccagaaagaac>700
601>TCATTACAGAAAATTTGGTTGCAGACTAATCCATGGGATTGTAGTTGTCCACGCATCGACTATCTCTCCGTTGGTTGAATAAAAATTCCCAAAACGAAC>700

                *         *         *         *         *         *
701>agggcagcgcgaaatgcagcggcagcggcaaaccggtgcgcagcattatttgcccgTAGTAa>763
701>AGGGCAGCGCAAATGCTCTGGCTCCGGCAAACCGGTACGCTCGATCATTTGTCCGTAATAAT>763
```

Rational2:

MW=27495.59

MCPSRCSCSGTEIRCNSKGLT
SVPTGIPSSATRLELESNKLQ
SLPDGVFDKLTQLTKSNNHLH
SLPDNLPASLEVLDVSNNHLH
SLPDNLPASLEVLDVSNNHLH
SLPDNLPASLEVLDVSNNHLH
SLPDNLPASLEVLDVSNNHLH
SLPDNLPASLEVLDVSNNHLH
SLPDNLPASLKELALDTNHLH
SVPDGIFDRLTSLQKIWLQTN
PWDCSCPRIDYLSRWLNKNSQ
KEQGSAKCSGSGKPVRSIICP

Reverse Translate results

Results for 251 residue sequence "Untitled" starting "CPSRCSCSGT"

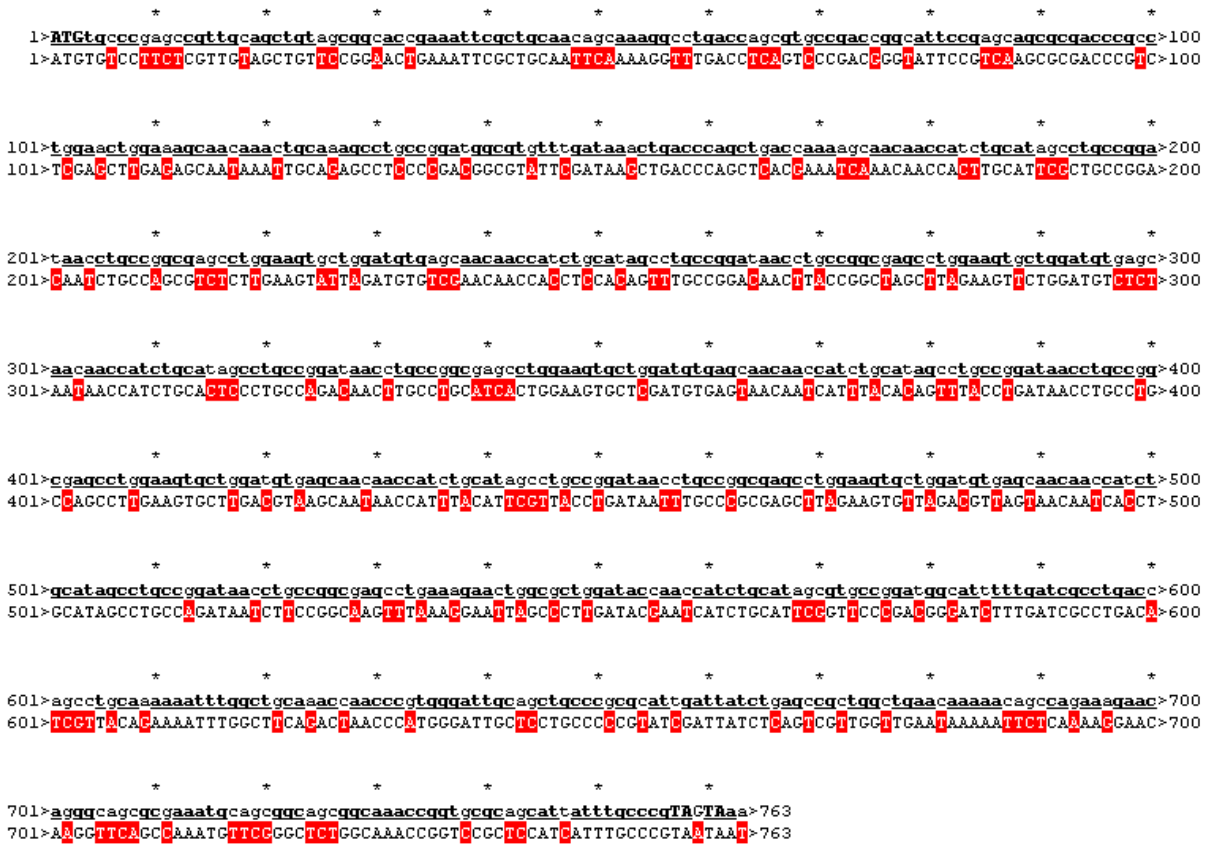
>reverse translation of Untitled to a 753 base sequence of most likely codons.

Gaattcgcggccgcttctag

ATGtgcccgagccgttgcagctgtagcggcaccgaaattcgctgcaacagcaaaggcctgaccagcgtgccgaccggcattccgagcagcgcgacccgcctggaactggaaagca
acaaactgcaaagcctgccggatggcgtgtttgataaactgacccagctgaccaaaagcaacaaccatctgcatagcctgccggataacctgccggcgagcctggaagtgctggatgtg
agcaacaaccatctgcatagcctgccggataacctgccggcgagcctggaagtgctggatgtgagcaacaaccatctgcatagcctgccggataacctgccggcgagcctggaagtgc
tggatgtgagcaacaaccatctgcatagcctgccggataacctgccggcgagcctggaagtgctggatgtgagcaacaaccatctgcatagcctgccggataacctgccggcgagcctg
gaagtgctggatgtgagcaacaaccatctgcatagcctgccggataacctgccggcgagcctgaaagaactggcgctggataccaaccatctgcatagcgtgccggatggcatttttgat
cgcctgaccagcctgcaaaaaatttggctgcaaaccaacccgtgggattgcagctgcccgcgcattgattatctgagccgctggctgaacaaaaacagccagaaagaacagggcagc
gcgaaatgcagcggcagcggcaaaccggtgcgcagcattatttgcccgTAGTAa

tactagtagcggccgctgcag

Alignment of our reverse translated sequence and the one optimized by ATG:biosynthetics:



Rational4:

consensus
SLPxHLPHxLxxLxCSxNxLx

MW=27844.059999999994

MCPSRCSCSGTEIRCNSKGLT
SVPTGIPSSATRLELESNKLQ
SLPDGVFDKLTQLTKSNNQLT
SLPDHLPHSLEVLDVSNNQLT
SLPDHLPHSLEVLDVSNNQLT
SLPDHLPHSLEVLDVSNNQLT
SLPDHLPHSLEVLDVSNNQLT
SLPDHLPHSLEVLDVSNNQLT
SLPDHLPHSLKELALDTNQLK
SVPDHIFHRLTSLQKIWLQTN
PWDCSCPRIDYLSRWLNKNSQ

KEQGSAKCSGSGKPVRSIICP

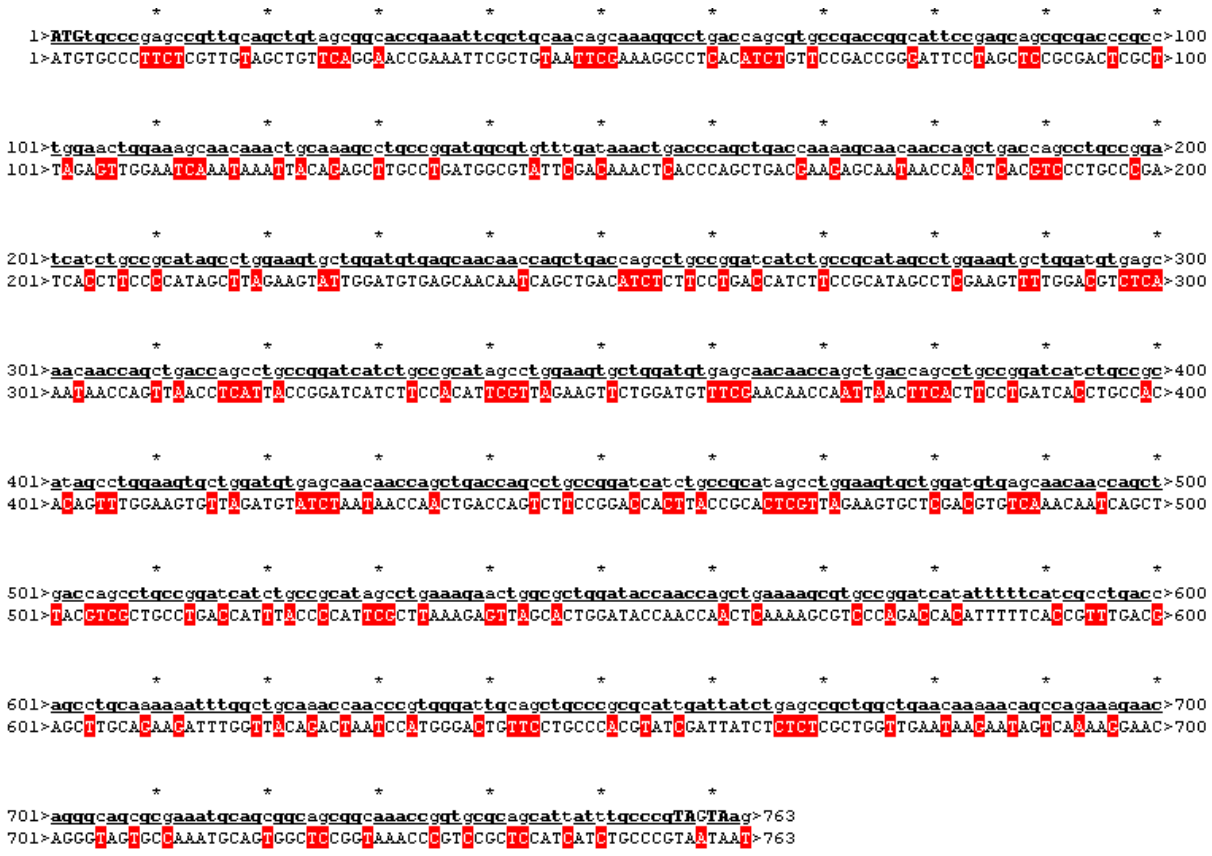
Reverse translation, optimized codon usage for e.coli:

ATGtgcccgagccgttgcagctgtagcggcaccgaaattcgctgcaacagcaaaggcctgaccagcgtgccgaccggcattccgagcagcgcgacccgcctggaactggaaagca
acaaactgcaaagcctgccggatggcgtgtttgataaactgacccagctgaccaaaagcaacaaccagctgaccagcctgccggatcatctgccgcatagcctggaagtgctggatgtg
agcaacaaccagctgaccagcctgccggatcatctgccgcatagcctggaagtgctggatgtgagcaacaaccagctgaccagcctgccggatcatctgccgcatagcctggaagtgct
ggatgtgagcaacaaccagctgaccagcctgccggatcatctgccgcatagcctggaagtgctggatgtgagcaacaaccagctgaccagcctgccggatcatctgccgcatagcctg
gaagtgctggatgtgagcaacaaccagctgaccagcctgccggatcatctgccgcatagcctgaaagaactggcgctggataccaaccagctgaaaagcgtgccggatcatatttttcat
cgcctgaccagcctgcaaaaaatttggctgcaaaccaacccgtgggattgcagctgcccgcgcattgattatctgagccgctggctgaacaaaaacagccagaaagaacagggcagc
gcgaaatgcagcggcagcggcaaaccggtgcgcagcattatttgcccgTAGTAa

ATG:biosynthetics optimization:

We submitted the DNA sequence to our sponsor. They kindly optimized us the sequence for secondary sturctures and RNA trafficking.

Alignment of our reverse translated sequence and the one optimized by ATG:biosynthetics:



Sequenzen von ATG:

Alignment machen.

I-Tasser: Structure prediction

I-Tasser is a 3-dimensional protein prediction method which can be used via a downloadable program or a online server. It is provided by the Yang-Zhang lab of the University of Kansas. The community-wide Critical Assessment of Structure Prediction (CASP) ranked the I-Tasser as the best protein prediction method in the server section.
The method of the I-Tasser can be divided into 4 stages:

Stage 1 is called threading. In this stage the sequence or parts of the sequence or motifs are compared with already known protein structures to identify evolutionary relatives and predict secondary structures. This gives a sequence profile and together with the query sequence it is threaded by LOMETS, a server combining seven threading programs which evaluate the sequence respectively the templates with an individual score. The top templates are selected for further consideration.

Stage 2 is responsible for structural assembly. The structure is modeled *ab initio* by aligning different templates from the threading. This is a very complex process and different programs are involved.

Stage 3 is a model selection of the achievements of stage 2 with a following refinement of the predicted models.

Stage 4 announces the function of the predicted molecule by its structural conformation comparing with already known proteins of PDB database.

The predicted models are all evaluated by the C-Score, which assesses the quality of the

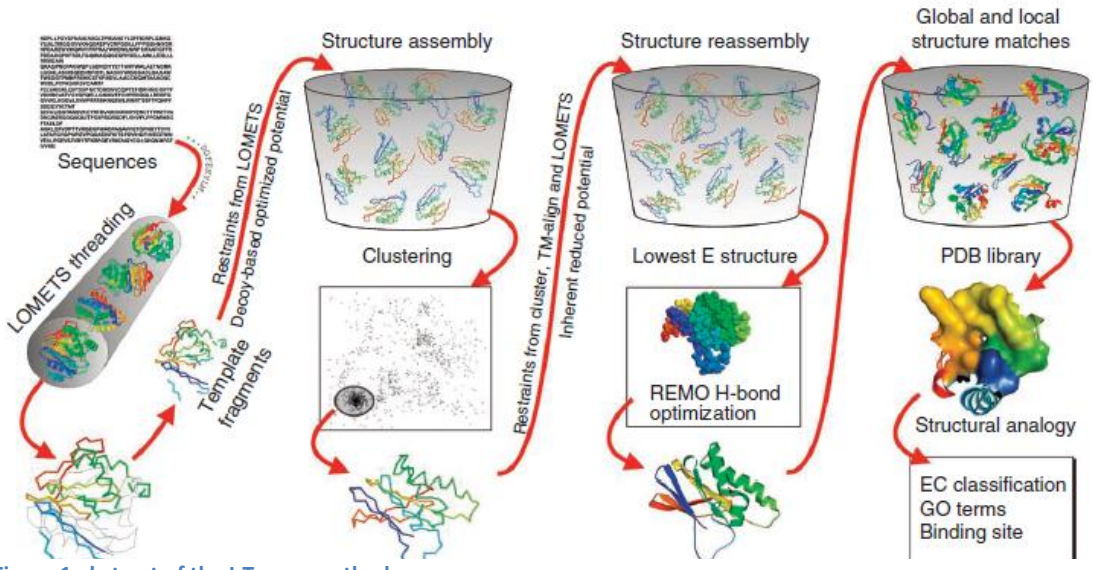


Figure 1 abstract of the I-Tasser method

prediction. It has a value from -5 to 2; the more positive the C-Score the merrier and more plausible the predicted structure is.

Lab in a Cell model

Experimental design:

The modeling was done in order to realize what parameters are crucial and which experiments need to be conducted to determine those.

Abbreviations:

Precipitator protein (P)

Column (C)

His-tagged Protein (H)

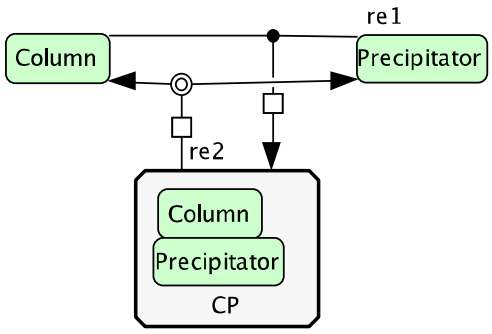
Assumptions:

[N] ∞

[I] ∞

If Nickel binds P or H the binding halftime ∞

Step 1



The Precipitator protein (P) binds the plastic column (C) by its plastic binding domain.

In reality the parameter “column” is not a soluble reagent, but the surface of the serological pipette that offers binding sites for the precipitator. We omitted the calculation for a surface reaction, since the simplified model is sufficient for our needs.

$$C + P \underset{k_0}{\overset{k_1}{}} CP$$

$$\frac{d\ CP}{dt} = C\ \ P\ k_1 - CP\ k_0 = 0$$

$$[CP] = C\ \ P\ \frac{k_1}{k_0}$$

All loose P are washed away. k_1 and k_0 are further assumed as constant.

dy/dt = cpk1-yk0
c-c0=p-p0=y0-y

wir haben uns das Leben bloss nicht "Einfach" genug gemacht da y0=0, und dazu kommt dass man sagen kann das p0=1 ist (100% der säule frei)

dan haben wir die gleichung:

dy/dt= (c0-y)(1-y)k1 - yk0 = k1y^2-(c0*k1+k1+k0)y+c0

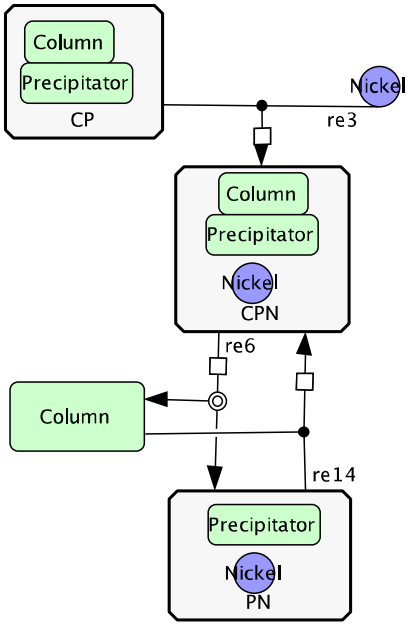
die Lösung dieser Gleichung ist zwar immer noch nicht einfach, aber sie hat 2 mögliche stable points wenn die Parameter "reasonable" sind. Daher wenn du eine value für c0 einsetzt und sagst das y(0) = 0, kriegest du eine Gleichung mit k1 und k0, also wenn du 2 Gleichungen mit 2 verschiedenen c0's hasst und die gemessene Proportion von der Säule die besetzt ist (dafür brauchst du die Sättigung Messung) kannst du k1 und k0 ausrechnen lassen...

Also theoretisch brauchst du 2 Messungen in dem Bereich wo der well noch nicht gesättigt ist und eine Messung mit der gesättigten Säule.

Step 2

The Precipitator has 3 different modi, depending on how many Nickel (between one and four are possible, depending on the precipitator version can bind. However, by adding a Nickel concentration much higher than protein concentration, it can be assumed, that all possible Nickel binding positions are occupied. Therefore we simplify the model assuming all positions are occupied and that there are no differences between the binding affinities in the 3 modi.

There is again a certain chance that the complex falls off the column. We further simplify the model by assuming that the on/off rates of the complex to the column remain the same although H is bound.



k_2 >>1

$$\boldsymbol{CP + N \overset{k_2}{} CPN \underset{k_0}{\overset{k_1}{}} C + PN}$$

$$\frac{\boldsymbol{d\ CPN}}{\boldsymbol{dt}} = \boldsymbol{CP\ \ N\ k_2 + \ C\ [PN]k_0 - \ CPN\ k_1\ = 0}$$

$$\boldsymbol{CPN\ = \frac{CP\ \ N\ k_2 + \ C\ [PN]k_0}{k_1}}$$

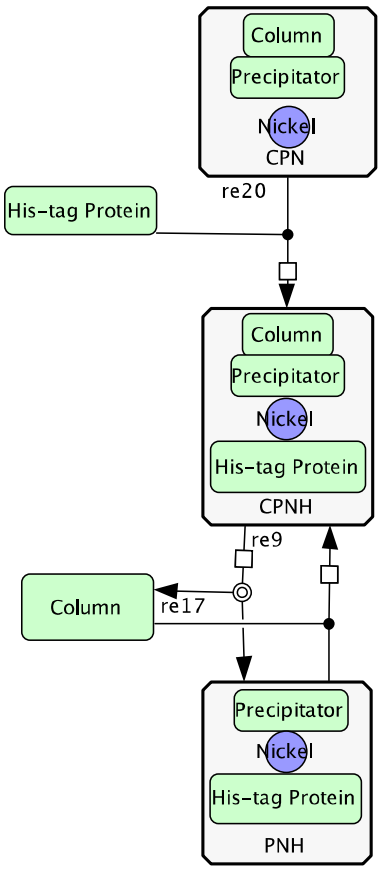
All non-column bound complexes are washed off.

Step 3

The Precipitator in its three different modi binds His-tagged Protein (H).

There is a certain chance that the complex falls of the column.

We simplify the model by assuming that the k_1 and k_0 rates of the complex binding to the column remain the same although H is bound. In reality they probably did change due to the increased mass of the complex after the His-tagged protein is bound



$$\boldsymbol{CPN + H \overset{k_5}{} CPNH \underset{k_0}{\overset{k_1}{}} C + PNH}$$

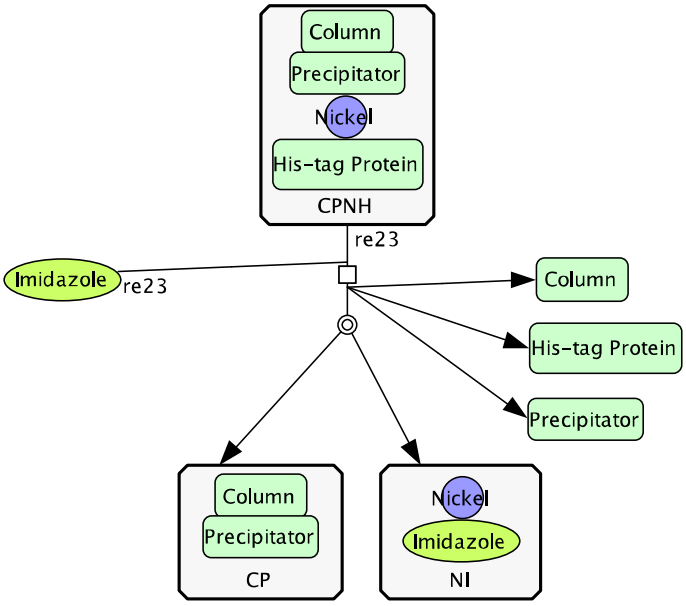
$$\frac{\boldsymbol{d\ CPNH}}{\boldsymbol{dt}} = \boldsymbol{CPN\ \ H\ k_5 + \ C\ [PNH]k_0 - \ CPNH\ k_1\ = 0}$$

$$\boldsymbol{CPNH\ = \frac{CPN\ \ H\ k_5 + \ C\ [PNH]k_0}{k_1}}$$

All non-column bound complexes are washed off.

Step 4:

By adding Imidazole (I) all Nickel get solved from the complexes.



$$CPNH \xrightarrow{k_8} CP + IN + H + C + P$$

$$\frac{d\,H}{dt} = CPNH\ \ I\ k_8$$

$$H = \frac{[CPNH][I]k_8}{[CP][IN]k_9}$$

All His-tagged Proteins are in solution now, purified from cell debris.

Conclusion:

1. To determine k_1 and k_0 , experiments to find out the binding affinity of the plastic binding domain are necessary. To get a direct access to these values, we planned to clone the plastic binding domain in front of a GFP. Then, dilution and washing assays are performed on polystyrene microtiter plates, red out by a fluorescence plate reader and thus, k_1 and k_0 can be calculated.

2. A qualitative experiment to prove that Nickel is binding the Precipitator is sufficient, since k_2 >> 1 and does not play a significant role in our setup. This experiment could be
3. In the end we want to give an estimate, how often one column can be reused until the efficiency goes too low.
4. We want to predict the amount of eluted His tagged protein depending on the size of the column.
5.