II. Results and Discussion

II.1 Main project

Structural characterization of a protein/RNA complex: human protein TAP/retroviral CTE RNA

1) Optimization of an *in vitro* TAP/CTE reconstitution protocol

Overview

As a first step, the *in vitro* reconstitution of the TAP-N/CTE-B complex was optimized. For this, a minimal functional CTE RNA construct was designed, based on previous data available in the laboratory, in addition to *in vitro* transcription yields and RNA stability. Also, a minimal construct of the TAP-N protein was engineered, based on limited proteolysis, NMR experiments and electrophoretic band shift assays. The final conditions for complex assembly were based on the molar protein/RNA ratio, $MgCl_2$ molarity, pH and temperature conditions that most favoured a homogeneous 1:1 complex, necessary for structural studies.

1.1) RNA minimal construct

As mentioned in the introduction, the CTE RNA is an extended RNA stem loop with two conserved internal loops, exhibiting an 180° inverse symmetry (Ernst *et al.*, 1997b). This RNA construct interacts *in vitro* with hTAP/NXF1 full-length and with an N-terminal region of TAP/NXF1, which spans residues 96-372 (TAP96-372). Two molecules of TAP96-372 are necessary to titrate one molecule of the CTE RNA, suggesting that the CTE RNA displays a duplicated binding site (figure 15). In accordance with this, a single loop of the CTE element is sufficient to promote nuclear export (Grüter *et al.*, 1998). Based on this information, a minimal fragment (CTE-B) that comprehends approximately half of the full-length CTE motif has been used as an initial CTE minimal domain. This fragment has been identified previously in our lab, as the minimal part of CTE that still forms a stable complex with TAP96-372. The CTE-B RNA encompasses the internal loop B, the external hairpin loop and part of the stem of the CTE RNA. Initially, only 3 mg of pure CTE-B RNA were obtained out of a 10 ml *in vitro* transcription reaction. In addition, the RNA construct was unstable in solution. A modified CTE-B construct was engineered, to contain two extra base pairs (two GC on both flanks of the sequence) on the stem loop. This construct yields 5-6 mg per *in vitro* transcription reaction and is a far more stable RNA transcript (figure 15).

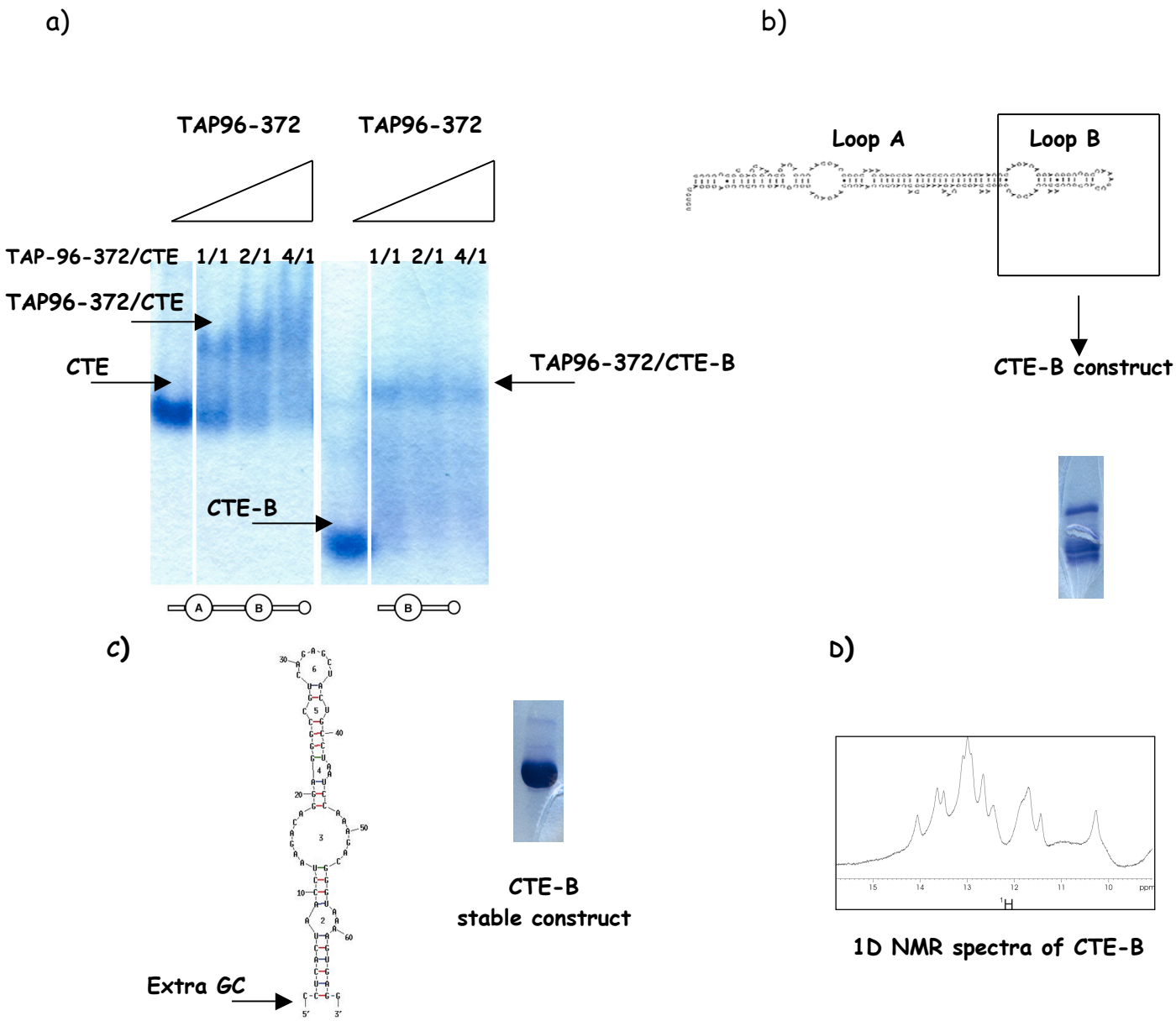

Figure 15. Identification of an optimal CTE RNA construct.
a) Increasing amounts of TAP96-372 were used to titrate a fixed concentration of CTE full-length and CTE-B. The titration assays were analysed with an 8% native-PAGE gel. b) An extra GC pair of nucleotides were added to the CTE-B construct. As expected for a structured RNA moiety, the final construct gave a homogeneous band on c) an 15% SDS-PAGE gel and had well-defined peaks on a d) 1D NMR spectrum.

1.2) Optimization of the RNA purification protocol

One-dimensional (1D) ^{1}H NMR experiments were used to evaluate the chemical integrity and purity of the optimized CTE-B RNA, as compared with the previously defined, unstable CTE-B minimal domain (figure 16). The peaks on the 1D NMR spectrum are well dispersed, which means that the CTE-B RNA is a structured macromolecule. An extensive purification protocol is however necessary, as traces of an undefined organic compound could be observed (figure 16).

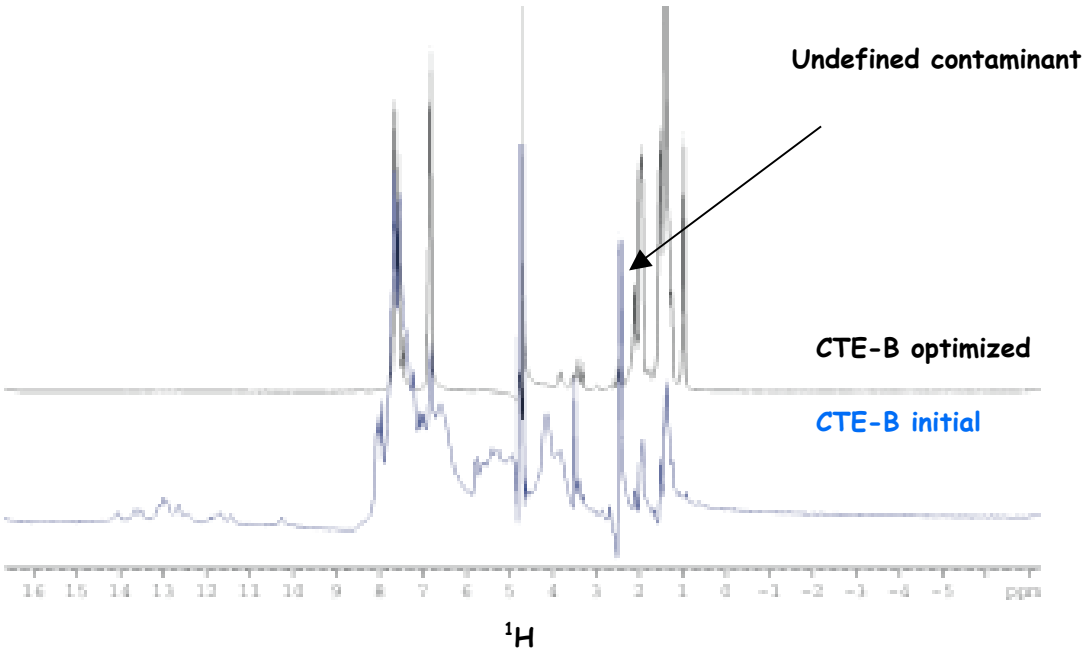

Figure 16. 1D NMR spectrum of CTE-B constructs.
blue: initial CTE-B; black: optimized CTE-B.

These observations led to the establishment of an optimal purification protocol, where the RNA was free of trace contaminants, stable and structured. As described in the section of materials and methods, the RNA transcription product was phenol-extracted from the reaction mixture and purified by 8M urea denaturing-PAGE electrophoresis. A pure RNA transcript was obtained after two rounds of ethanol precipitation and extensive dialysis with a 1M-to-100mM NaCl gradient (figure 17).

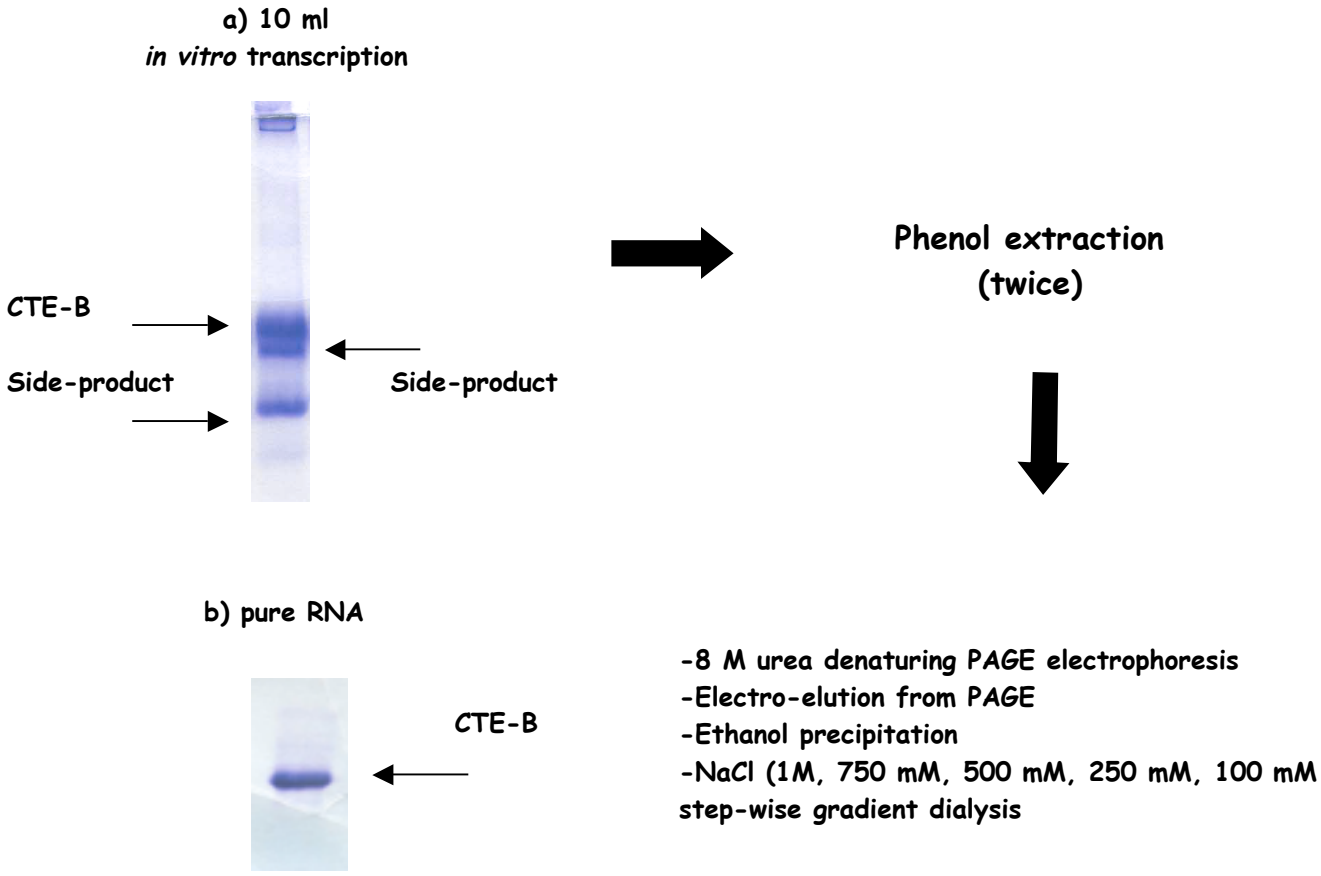

Figure 17. Optimized purification protocol of the *in vitro* transcribed RNA. a) a 15% SDS PAGE gel showing the crude reaction mixture. b) a 15% SDS PAGE gel showing an aliquot of pure CTE-B RNA.

1.3) Protein minimal construct

As mentioned above, the full-length human protein TAP and a fragment spanning residues 96-372 (TAP96-372) have been shown to interact with the CTE RNA and to form a stable complex, even in the presence of high concentrations of an unspecific RNA competitor (Liker *et al*., 2000). Recently, the structure of TAP102-372 has been elucidated. It encompasses an N-terminal flexible region (residues 102-118), followed by a non-canonical RNP-type RNA-binding domain (RBD, residues 119-198), a short linker (residues 199-202) and a leucine rich repeat domain (LRR, residues 203-362) (Liker *et al*., 2000). In the crystal structure of TAP102-372, amino acids 102-117 were flexible and could not be traced on the electron density map. Previous attempts to engineer a shorter functional TAP protein construct where hindered because the construct TAP102-372 did not bind to the CTE-B, whereas TAP96-372 (a shorter construct) did have CTE-binding activity. The reason for this was that there was a point mutation on the TAP102-372 protein construct, which prevented it from interacting with the CTE RNA *in vitro*. In addition, a construct beginning in amino acid 118, TAP (118-372), did not bind to the CTE RNA either (data not shown). Based on two-dimensional (2D) $^{1}H^{N}$-^{15}N HSQC-TROSY NMR spectra (referred to as fingerprint NMR experiments below) (figure 18) and limited proteolysis experiments (figure 19), I have engineered a shorter protein construct (TAP 112-372, referred to as TAP-N below) that interacts *in vitro* with the CTE RNA. TAP-N lacks an N-terminal flexible region (residues 96-111), that contrary to previous results, is not necessary for binding to the CTE RNA.

1.4) Purification of TAP-N to homogeneity

As mentioned in the section of materials and methods, recombinant TAP-N was expressed as a GST fusion under standard conditions. The recombinant fusion protein was enriched by affinity chromatography with a reduced gluthatione column and cleaved with TEV protease at 4°C. Contaminant genomic RNA, protease traces and the cleaved GST fusion were removed by ion exchange chromatography. The pure protein was subject to gel filtration, prior to use (figure 20).

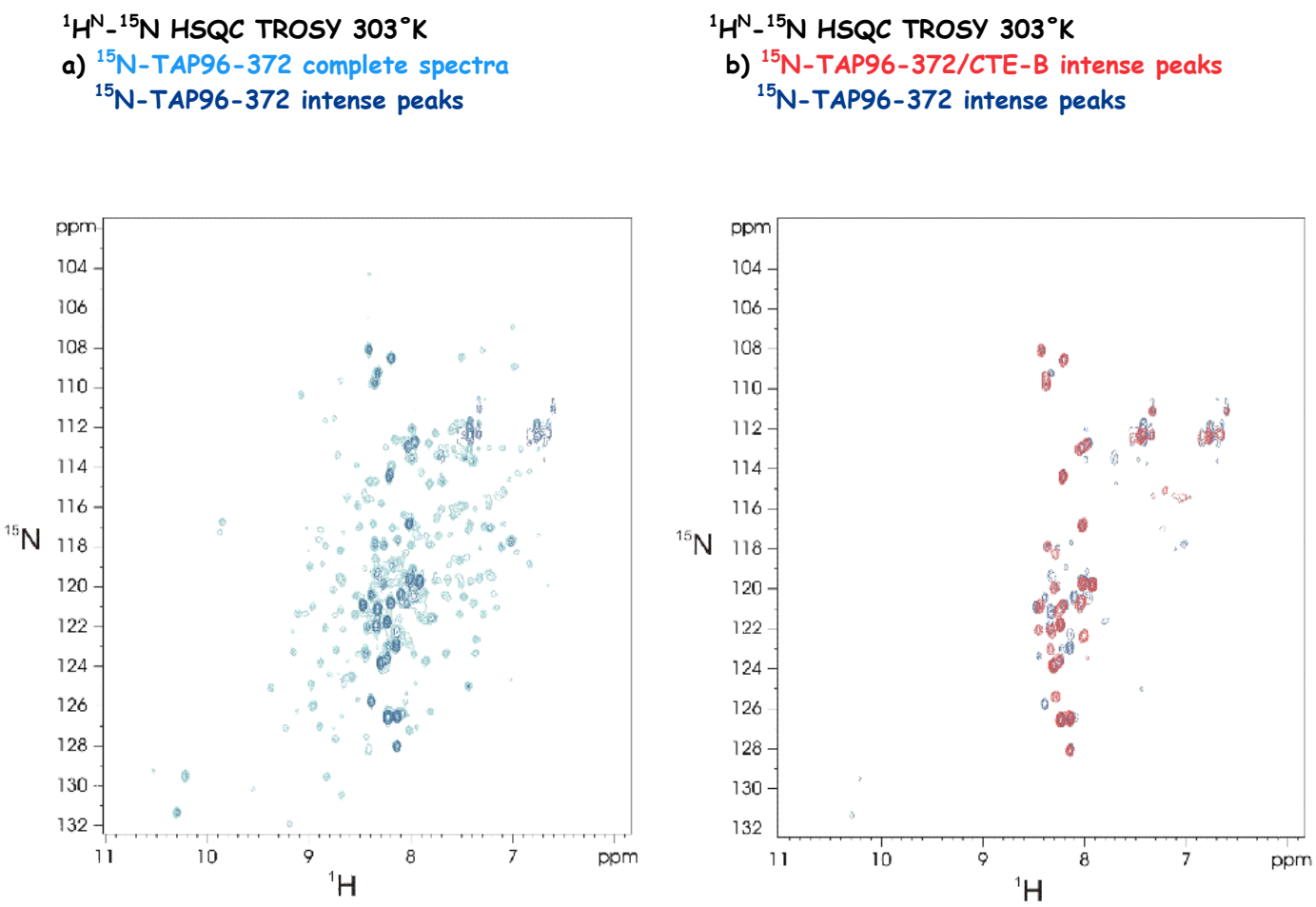

Figure 18. Two $^{1}H^{N}$-^{15}N HSQC TROSY spectra are presented for TAP96-372 unbound and TAP96-372/CTE-B. Spectra a) is an overlay of the same experiment at two signal levels. A spectrum where only the intense signals can be visualized is shown in dark blue. In overlay and coloured in light blue, are all the visible peaks. Spectrum b) is an overlay of two spectra that show only the intense peaks for TAP96-372 unbound (blue) and TAP96-372/CTE-B (red). Spectrum a) tells us that TAP96-372 contains a flexible region, which gives intense signals in comparison with the core of the structure. One can learn from the overlay on b) that there is minimal signal perturbation of the intense signals when TAP96-372 is bound to CTE-B. This information suggests that TAP96-372 contains a flexible region that does not participate in CTE RNA recognition.

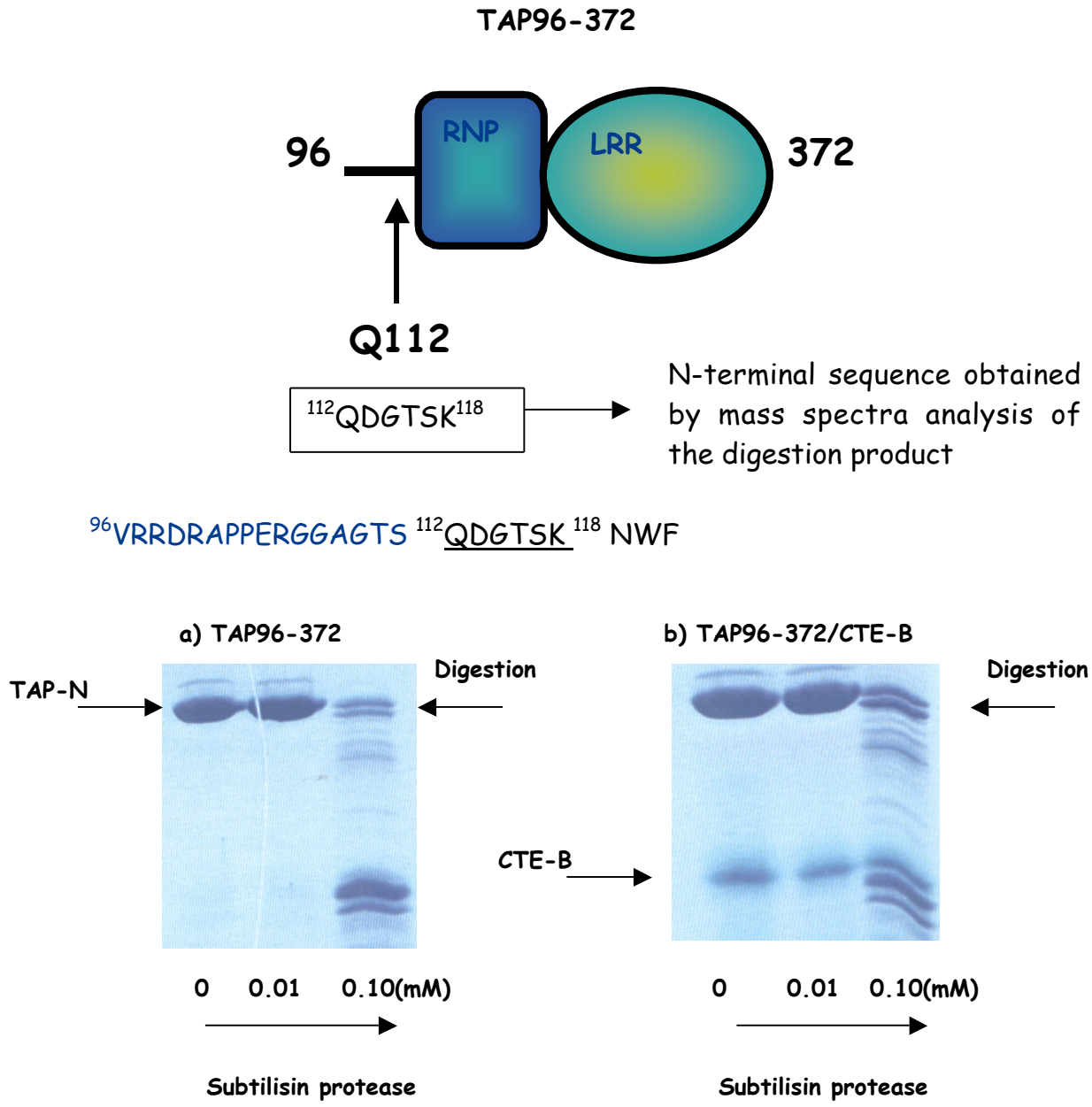

Figure 19. Limited proteolysis of TAP96-372 unbound and TAP96-372/CTE-B.
a) proteolytic pattern of TAP96-372 unbound.
b) proteolytic pattern of TAP96-372/CTE-B complex.

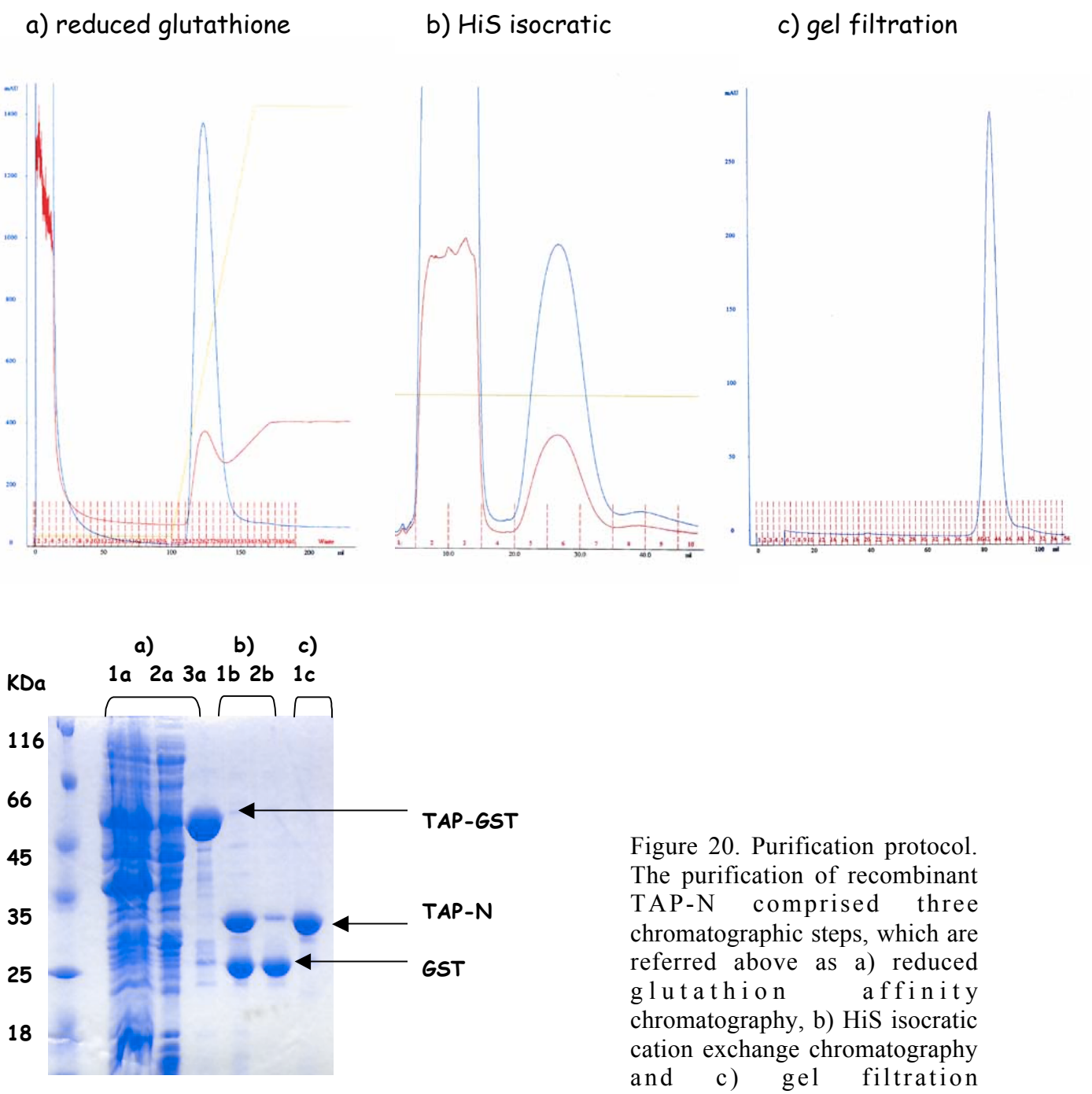

Figure 20. Purification protocol. The purification of recombinant TAP-N comprised three chromatographic steps, which are referred above as a) reduced glutathion affinity chromatography, b) HiS isocratic cation exchange chromatography and c) gel filtration chromatography. A 15% SDS-PAGE gel with aliquots taken at each step is also shown. On each chromatogram, the curve in blue corresponds to the absorbance at 280 nm, whereas the curve in red refers to the absorbance at 260 nm. The yellow line indicates the conductivity.

1.5) Preparation of spin-labelled samples for NMR experiments

Spin-labelled protein samples were prepared for NMR experiments, by heterologous expression in *E. coli*. For this, a minimal medium (M9) was supplemented with spin labelled elements, $^{15}NHCl_4$ and /or $^{2}H_20$, depending on the case. The protein yields were not high, probably due to the presence of rare-codons in the gene sequences used. In addition, the amount of protein obtained decreased with the use of spin-labelled elements during bacterial growth. In particular, a decrease in protein yields correlated with the addition of more than 50% v/v of $^{2}H_20$ to the growth media. This fact became a limiting factor, when multidimensional NMR experiments were performed, given the need for a ^{13}C-, ^{2}H-, ^{15}N- labelled sample. To overcome this limitation, a commercially available yeast extract with 95% $^{2}H_20$, 99% ^{13}C and 75% ^{15}N isotopes (Silantes) was used to prepare ^{13}C-, ^{2}H-, ^{15}N- TAP-N. This media has been described in materials and methods. Figure 21 shows a growth curve as plotted when the protein was expressed in minimal M9 media with different percentages of $^{2}H_2O$ and a constant percentage of ^{15}N, or with the commercial medium Silantes. As shown in the plot, cells grow at a relatively similar rate in M9 minimal medium supplemented with 70 or 85% of $^{2}H_20$. Significant improvement in cell density, and protein yields were obtained when cells where supplemented with Silantes. The growth curve with this special media was similar to the one for M9 $^{2}H_20$ 0.0%.

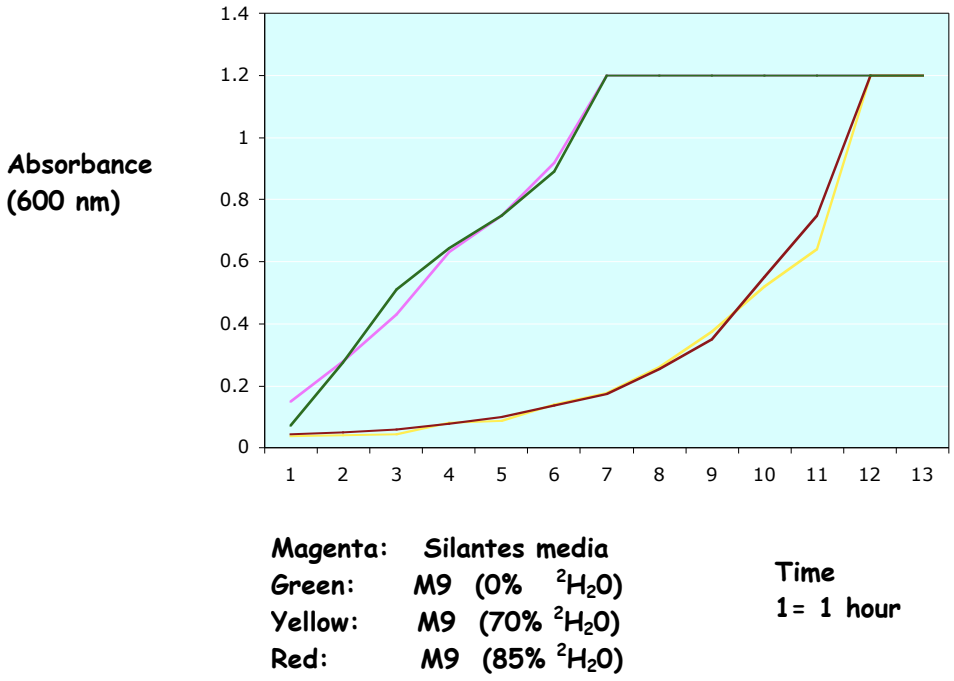

Figure 21. A plot of cell growth density over time. Cell density was measured at 600 nm in AU, and aliquots were taken every hour. Four different conditions were tested and their growth rate plotted in different colours. M9 minimal medium at constant ^{15}N percentage and $^{2}H_20$ 0%, 70% and 85%) and Silantes yeast extract ($^{2}H_20$ 95%). The colour code for each condition is shown above.

1.6) Determination of the molecular weight of TAP-N

Overview

The molecular weight of the protein TAP-N and its oligomeric state in solution were investigated by analytical centrifugation, mass spectrometry and ^{1}H 1D NMR T2 relaxation experiments. This information is important because it corroborates the chemical integrity of the protein and confirms whether or not the sample is a monomer in solution.

1.6.1) Analytical Ultracentrifugation Experiments

Analytical ultracentrifugation (UTC) experiments provide an efficient means for determining the molecular mass and characterizing the oligomeric state of macromolecules in solution. It stands in advantage over gel filtration experiments, because it is independent of the shape of the molecule. Thus, giving a more accurate molecular weight value. For this purpose, milligram amounts of TAP-N were prepared for UTC experiments, which were performed and analysed by Dr Arie Geerlof, at the EMBL in-house protein expression facility. The estimated molecular weight for TAP-N could be calculated to a rough value of 30 KDa, which suggests that TAP-N is a monomer in solution.

1.6.2) Mass Spectrometry

Mass Spectrometry was useful in corroborating the results obtained by UTC and 1D NMR T2 relaxation measurements. It also provided a means for investigating the purity and integrity of the protein samples used during this study. A small aliquot of the sample prepared for the structural studies, was given to the EMBL in-house proteomics core facility. The ionic dispersion of the sample gave a major peak at 29.9 KDa, with almost no contaminants (figure 22).

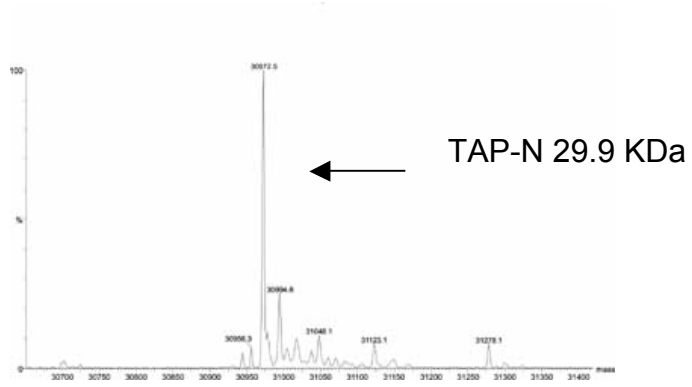

Figure 22.The mass spectrometric pattern of pure TAP-N is shown above. The protein is relatively pure, as one major peak appears on the spectrum. Also, the molecular weight is in agreement with the theoretical value calculated for TAP-N.

1.6.3) NMR spectroscopic characterization of TAP-N

^{1}H 1D NMR experiments were used to calculate a rough molecular weight of the protein based on T2 relaxation experiments (figure 23). In accordance with UTC experiments and the value obtained by mass spectrometry, TAP-N is a monomer in solution, with a molecular weight of approximately 30 KDa.

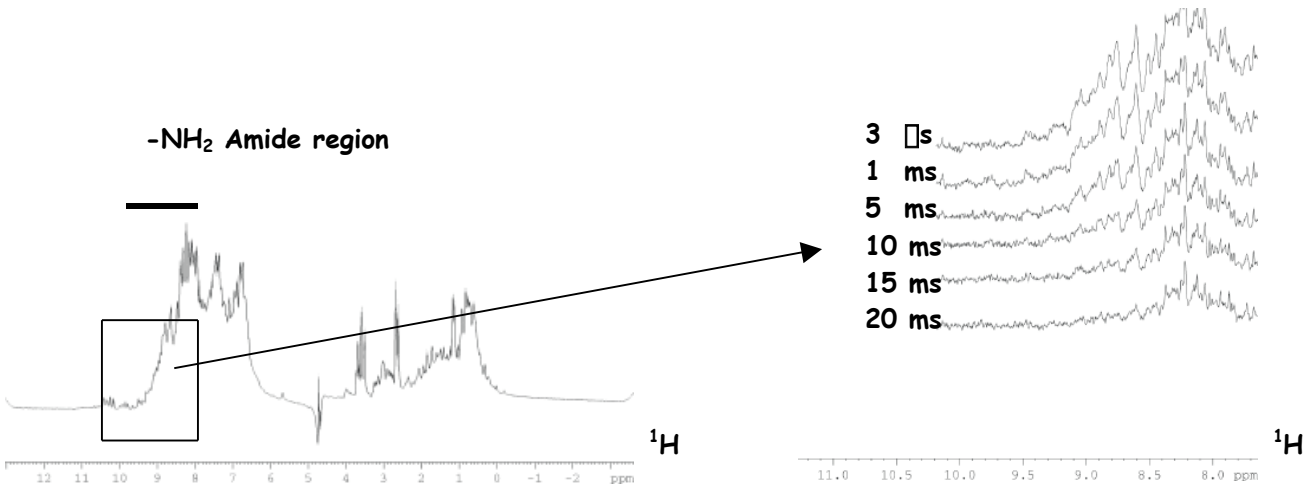

Figure 23. NMR based characterization of a 0.1 mM TAP-N sample. T2 relaxation measurements were taken for several time points. An inset on the left shows the region of 10-8 ppm, where the signals corresponding to the protons of the backbone amides are observed. The signal on this region decays around 20 ms, which corresponds to an approximate molecular weight of 25-30 KDa. This estimation is based on a plot of average T2 values for several proteins with a known molecular weight.

1.7) Investigation of the factors that influence the *in vitro* reconstitution of TAP-N/CTE-B: the role of pH, temperature, salt and protein/RNA molar ratio

Overview

Electrophoretic band shift assays were used to investigate the influence of temperature, pH, $MgCl_2$ concentration and protein/RNA molar ratio, on the *in vitro* reconstitution of the TAP-N/CTE-B complex (figure 24). Further, 2D NMR $^{1}H^{N}$-^{15}N HSQC TROSY experiments were performed to monitor the tendency of TAP-N to oligomerize upon binding to the CTE-B RNA (figure 25).

1.7.1) Electrophoretic band shift assays

If mixed in a 1:1 molar ratio, a homogeneous TAP-N/CTE-B complex could be reconstituted at a broad range of pH and $MgCl_2$ molar values. Protein oligomerization however, occurred if TAP-N was mixed in excess, with respect to the CTE-B (figure 24).

1.7.2) NMR-based titration experiments: monitoring the tendency of TAP-N to oligomerize

To investigate further, the tendency of TAP-N to oligomerize upon CTE-B binding, I titrated a given molar concentration of ^{15}N- labelled TAP-N, with increasing amounts of CTE-B and recorded a $^{1}H^{N}$-^{15}N HSQC TROSY spectrum for each titration step (figure 25). As typically observed during protein oligomerization, many ^{1}H-^{15}N correlation peaks disappeared on the 2D NMR fingerprint, when the CTE-B was initially mixed with ^{15}N- labelled TAP-N, at a TAP-N/CTE-B:1/0.1 molar ratio (figure 25). Increasing the molar RNA ratio stepwise did not bring back the missing ^{1}H-^{15}N correlation peaks, on the 2D NMR fingerprint. In agreement with the electrophoretic band shift assays, these results suggested that, in order to avoid the oligomerization of TAP-N, the CTE-B has to be initially mixed in excess, with respect to TAP-N.

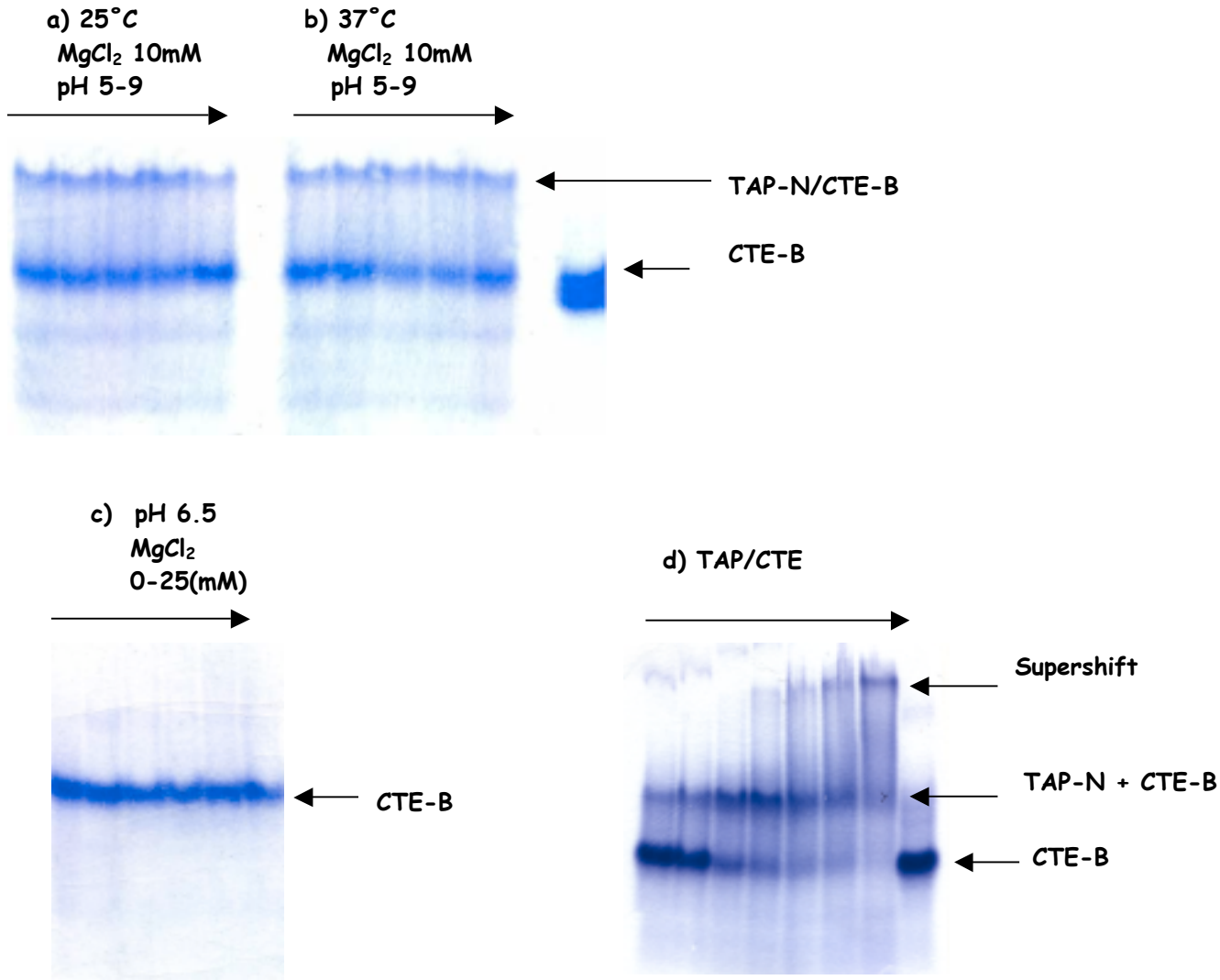

Figure 24. Electrophoretic band shift assays (8% native-PAGE) for the TAP-N/CTE-B complex and the CTE-B RNA. Influence of different pH values, on the TAP-N/CTE-B complex at a) 25°C and b) 37°C. c) Influence of $MgCl_2$ molarity on the homogeneity of the CTE-B RNA at pH 6.5; d) Titration of a constant molar value of CTE-B with increasing amounts of TAP-N.

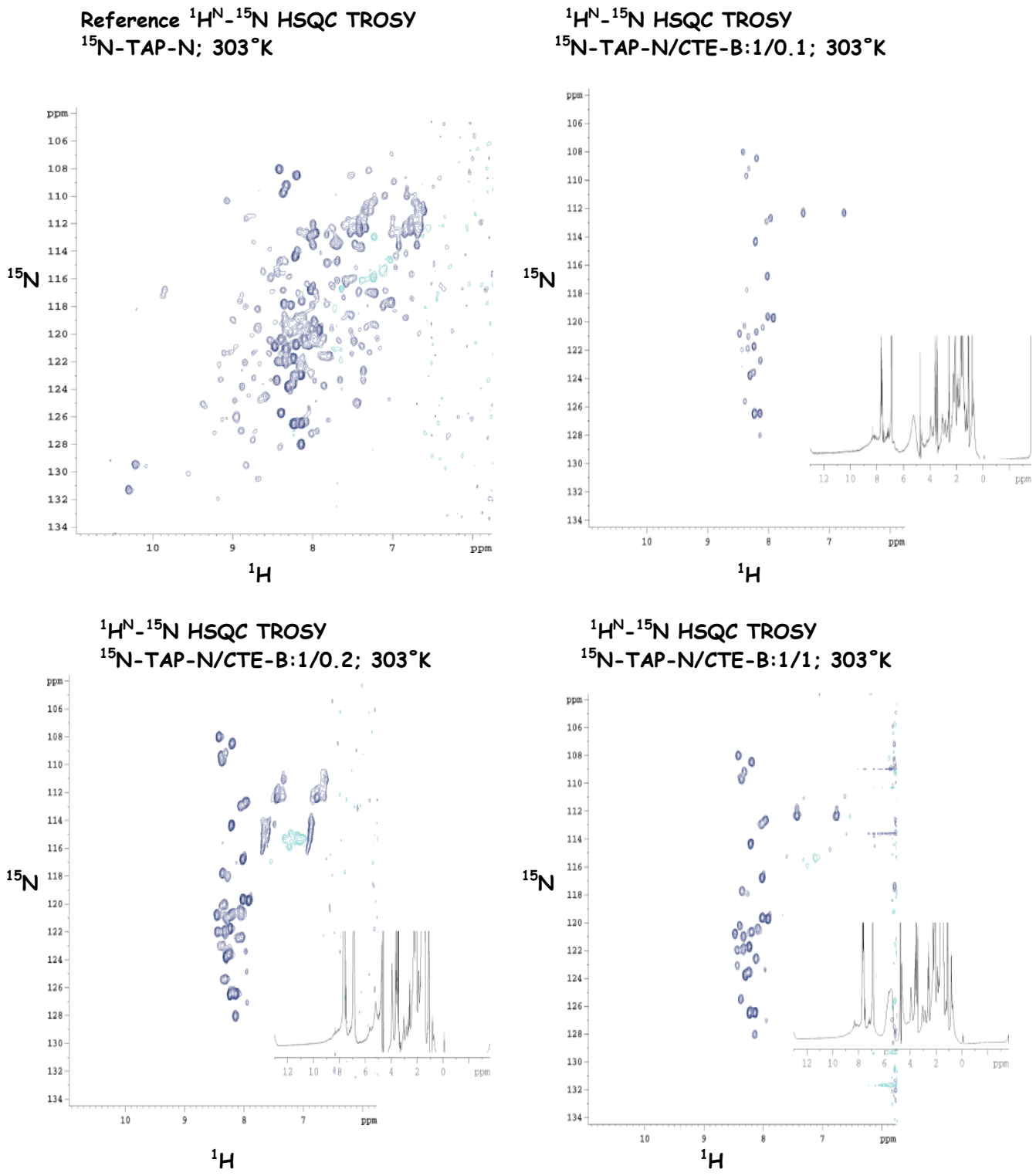

Figure 25. Influence of the TAP-N/CTE-B molar ratio, on the oligomeric state of TAP-N, as monitored by $^{1}H^{N}$-^{15}N HSQC TROSY spectra**.** An inset of a corresponding ^{1}H 1D spectrum is shown, along with each spectra. The fact that all 1D NMR spectra are similar, suggests that TAP-N dos not degrade or unfold, upon CTE-B addition.

1.8) Chromatographic characterization of TAP-N, CTE-B and TAP/CTE-B

The chromatographic properties of TAP-N, CTE-B and the TAP-N/CTE-B complex where investigated, to evaluate an appropriate *in vitro* reconstitution protocol. Even if equal molar amounts of protein and RNA were optimally mixed, the inactive, unbound fraction of either protein or RNA had to be removed. Further, the purification of a homogeneous 1:1 complex was not straightforward, because TAP-N, CTE-B and the TAP-N/CTE-B complex have similar gel filtration chromatographic patterns (figure 26), and because the complex dissociates, when subjected to ion exchange chromatography (data not shown). A homogeneous complex could be obtained, only when the RNA was mixed in slight excess, and the complex purified by gel filtration. This allowed the separation of the elution peaks corresponding to the TAP-N/CTE-B complex and the trace amounts of CTE-B RNA (figure 27).

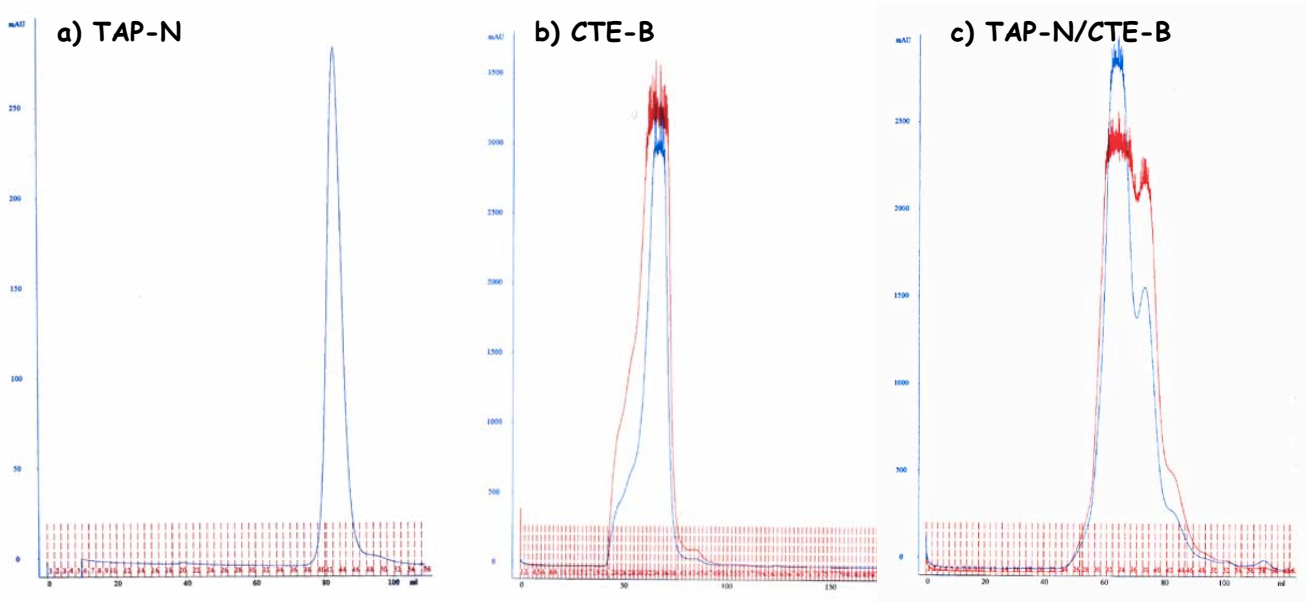

Figure 26. Gel filtration chromatograms for TAP-N, CTE-B RNA and the TAP-N/CTE-B complex. The three samples elute at about 60-80 ml and have broad elution curves. On each chromatogram, the curve in blue corresponds to the absorbance at 280 nm, whereas the curve in red refers to the absorbance at 260 nm. The yellow line indicates the conductivity.

1.9) Complex *in vitro* reconstitution

Given the information presented in the previous sections, I have designed a protocol for the *in vitro* reconstitution and purification of a 1:1 TAP-N/CTE-B complex. Since $MgCl_2$ and pH did not disturb proper RNA conformation or protein/RNA complex formation, I decided to use a compromised pH value and $MgCl_2$ molarity (pH 6.5 and 5-10mM $MgCl_2$), suitable for structural studies. For both crystallographic and NMR experiments, a low ionic strength is desirable. The reason for this is that a high salt content increases the chances of having false positives in crystallographic screens on the one hand, and masks magnetization transfer events during NMR experiments on the other. Finally, a low pH stabilizes RNA integrity in solution and slows-down amide proton exchange during NMR experiments.

To avoid unspecific interactions, the annealed CTE-B RNA was added drop-wise to a 20 fold diluted solution of TAP-N. The mixture was incubated in ice for at least 2 hours, concentrated to approximately 1 ml and homogenized by gel filtration. The crucial point here was to add an excess of RNA. Having the RNA in excess prevented protein oligomerization and allowed the removal of unbound RNA molecules by gel filtration. Figure 27 is a scheme of the *in vitro* reconstitution protocol.

1.10) Stability of the complex in solution

To investigate the stability of TAP-N/CTE-B in solution, aliquots of the protein/RNA complex were taken over the course of a week and analyzed by SDS-PAGE and native-PAGE electrophoresis (figure 28). If handled correctly, mg amounts of TAP-N/CTE-B complex remain stable at 4°C, for up to one week.

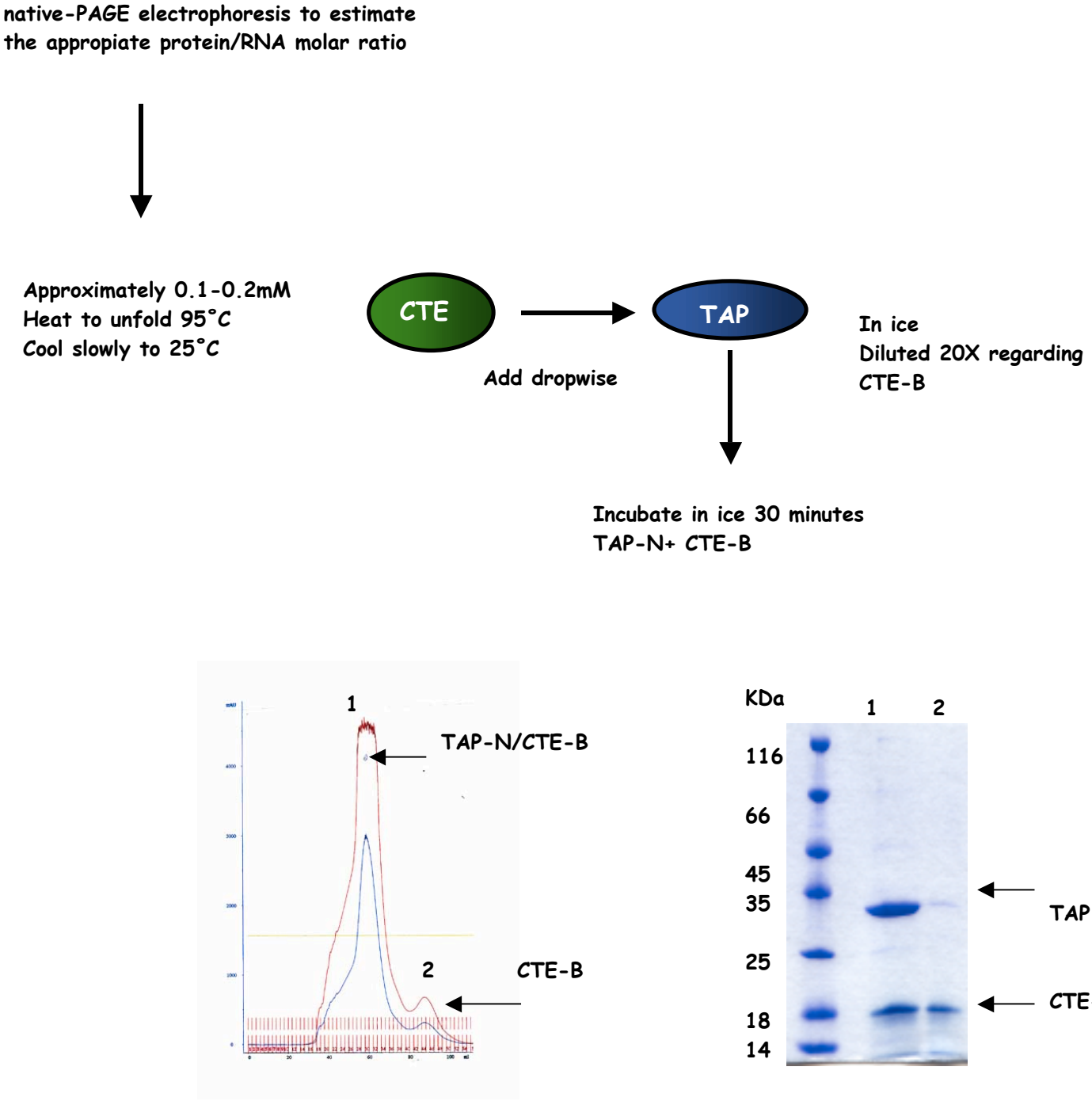

Figure 27. *In vitro* reconsitution and purification protocol

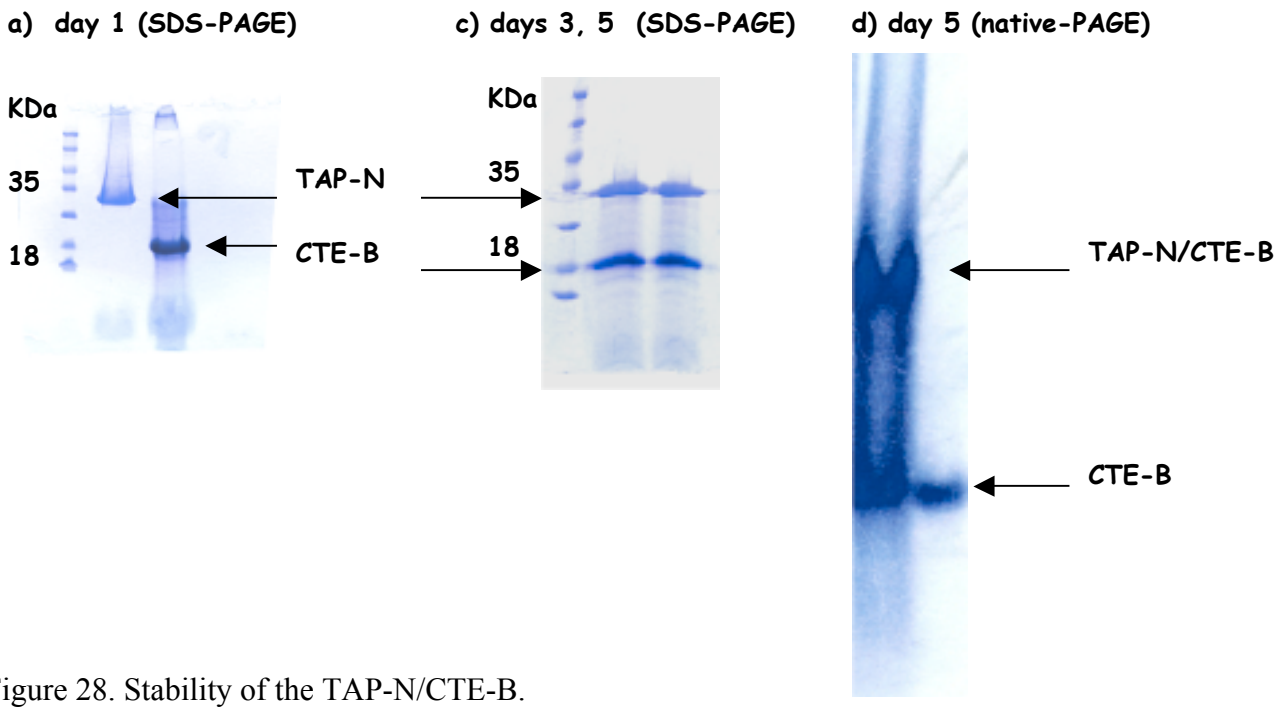

Figure 28. Stability of the TAP-N/CTE-B.
a) 15% SDS-PAGE of TAP-N and CTE-B before in vitro reconstitution of the complex. b) 15% SDS-PAGE of TAP-N/CTE-B after 3 and 5 days at 4°C. c) 8% native-PAGE of TAP-N/CTE-B after 5 days. It can be observed on the native-PAGE in d), that some RNA is unbound with time, but it is not degraded.

Conclusions

The interaction of TAP-N with the CTE-B RNA is stable, as indicated by the fact that a homogeneous complex can be obtained by gel filtration. Further, the *in vitro* reconstitution of the TAP-N/CTE-B complex is not affected by pH and salt concentrations, but depends on the molar protein/RNA ratio. In this sense, the supershift band observed during an electrophoretic band shift assay (Figure 23d), on the one hand and the disappearance of ^{1}H-^{15}N correlation peaks on a 2D NMR fingerprint (figure 24), on the other, suggest a tendency of TAP-N to undergo conformational changes, which most probably lead to protein oligomerization. This only occurred if the protein was mixed in molar excess with respect to the RNA, as a drop-wise addition of CTE-B RNA, to a diluted solution of TAP-N prevented both the formation of a supershift on a native-PAGE and the disappeance of ^{1}H-^{15}N correlation peaks on a 2D NMR fingerprint.