II. Results and Discussion

II.2 Side project

X-ray crystal structure and ligand-binding surface of REF1-II, a splice variant of the protein REF/Aly

1) Structural characterization of REF1-II

Despite the structural and biochemical data available to date, it is still not clear how the structure of REF proteins is related their biological functions. In an approach to understanding the structure-function relation of REF proteins, I have crystallized a stable REF construct and investigated its binding surface of interaction with two putative ligands, UAP56 and single stranded (ss) RNA, by means of NMR-based titration experiments. Two protein constructs (amino acids 1-103 and 1-128) have been used during this study, which are derived from the murine REF splicing variant, REF1-II. REF1-II is 161 amino acids long, from which REF1-103 spans the conserved N-box and the RBD motif and REF1-128 comprises the N-box, the RBD motif and most of the C-terminal RGG motif (figure 34).

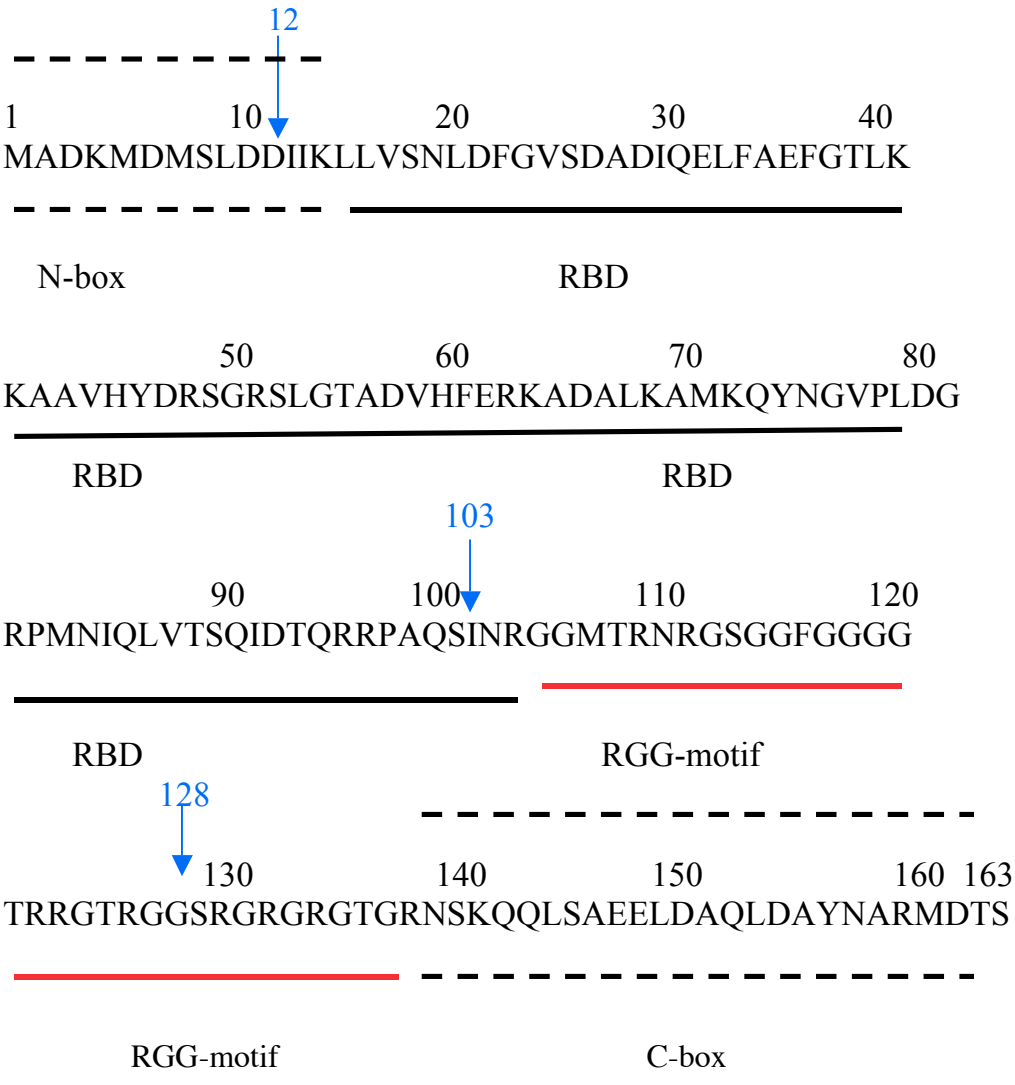

Figure 34. Amino acid sequence of REF1-II.

1.1) Sample preparation

REF protein constructs were obtained as GST-fusion proteins, by heterologous expression in *E. coli* BL21 DE3 strains, as described in materials and methods. All genes had been sub-cloned into pGEXcs vectors and kindly provided by Elisa Izaurralde. Initially, three REF1-II constructs where expressed and their stability tested: amino acids 1-103, 14-103 and 1-128.

1.1.1) Purification strategy

A three-step purification protocol was established for REF1-II (1-103), (14-103) and REF1-II (1-128). Each recombinant fusion protein was enriched from the cell lysate by affinity chromatography, with reduced glutathione resin. The enriched fusion protein was cleaved with TEV protease and cleaned from the GST, protease and genomic RNA traces by ion exchange. REF1-II (1-103) and (14-103) were subject to anion exchange chromatography with a HiQ column, whereas REF1-II (1-128) was subject to cation exchange chromatography with a HiS column. Next, pure protein samples were subject to gel filtration prior to use. Typically, 3 to 4 mg of pure protein was obtained per Lt of heterologous expression (figure 35).

1.1.2) Characterization of the REF1-II constructs: stability in solution

I next investigated the stability of the three REF1-II protein constructs. From these, only REF1-II (1-103) and (1-128) were stable enough to proceed with further structural studies. Recombinant REF1-II (1-103) is stable in solution at room temperature and 4°C, REF1-II (14-103) degraded immediately during its purification, and REF1-II (1-128), which features all the functional motifs attributed to the REF family of proteins (e.g. an N-terminal box, RBD motif and RGG-motif) can be kept integral if a considerable amount of protease inhibitors is present during all the purification steps and experiments (figure 36).

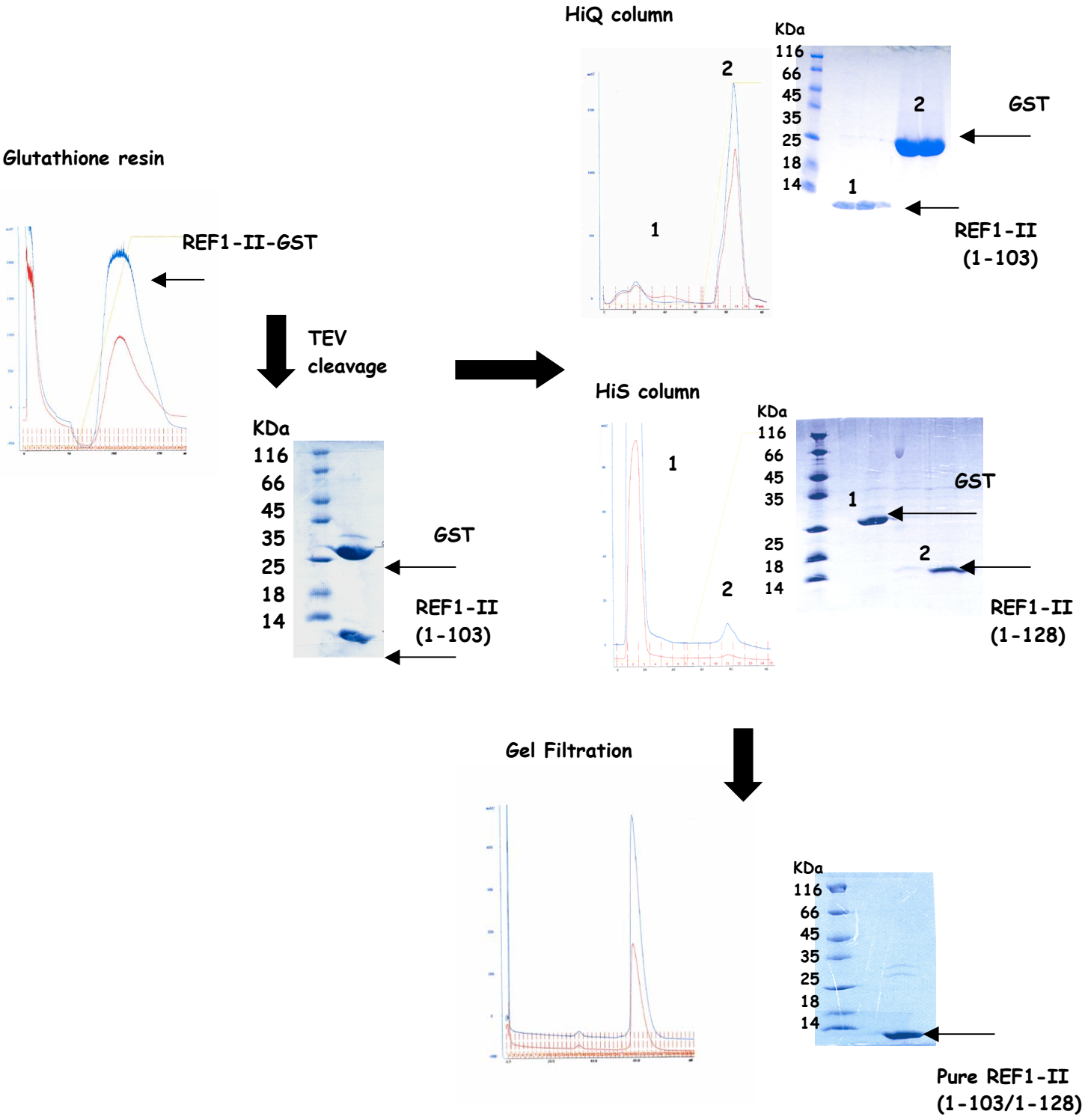

Figure 35. Protein purification. REF 1-II protein constructs were purified on the basis of a three-step purification protocol. The figure above presents a representative chromatogram for each step, along with a 15% SDS-PAGE gel. On each chromatogram, the blue curve refers to the absorbance at 280, whereas the red curve refers to the absorbance at 260 nm. The yellow line indicates the conductivity.

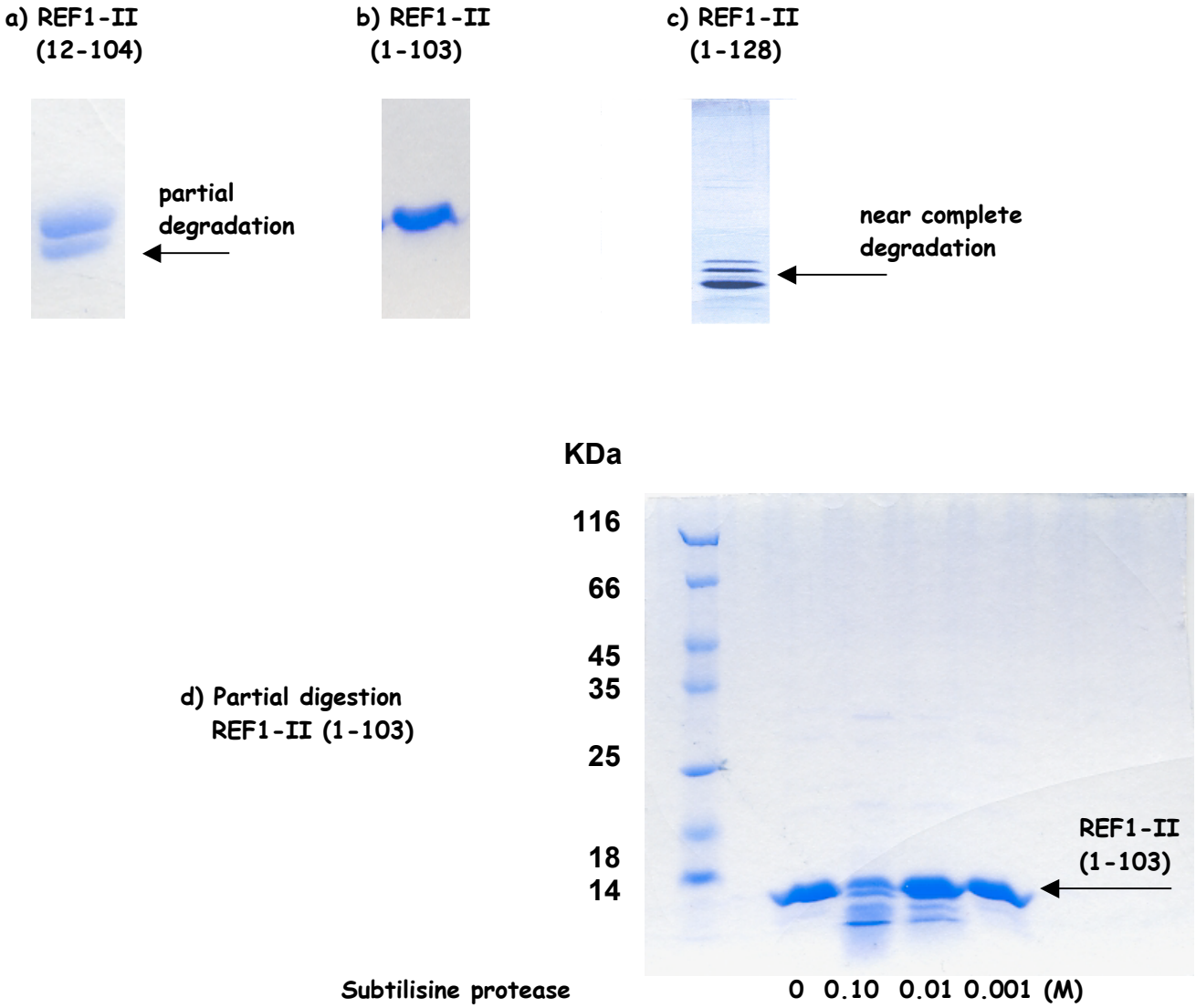

Figure 36. Protein stability. The stability of three REF1-II constructs was tested: a) 12-104; b) 1-103 and c) 1-128. d) REF-1-II (1-103) is stable in the presence of subtilisine protease at limited activity (0.001 M). Pure samples of the three REF constructs were left at 4°C for several days, from which several aliquots taken systematically and loaded onto a SDS-PAGE gel. REF103 was stable even after 1 week at 4°C. Conversely, REF12-104 degraded immediately and REF1-128 could remain stable at least for one week, when protease inhibitors were added to all the buffer solutions used to manipulate the protein.

1.2) Crystallographic experiments

REF1-II (1-103), the more stable protein construct was used for crystallographic trials. An exhaustive screen was performed, based on a sparse matrix approach. For this, several commercially available kits were employed. Two temperatures (12, 4°C) and two concentrations (5, 10 mgl/ml) were tested. Table 2 below, describes the conditions in which crystals grew. Crystals diffracting beyond 2 Å were obtained with 3.4 M 1,6 Hexanediol (figure 37).

TABLE 2

Sample	Conditions	Cryoprotectant	Outcome
REF1-103 10 mg/ml	10% v/v isopropanol 0.1 M imidazol, pH 8.0 4°C	Mother liquor + 25%v/v glycerol	Crystals dissolved while handling
REF1-103 10mg/ml	Several hits on PEG-conditions 4°C	Mother liquor	Needles Diffracted weak
REF1-103 10mg/ml	3.4 M 1,6 Hexanediol 0.2 M $MgCl_2 \bullet 6H_20$ 0.1 M Tris pH 8.5 12°C	Mother liquor	Diffracted to 1.9 Å at the Swiss Light Source (SLS)

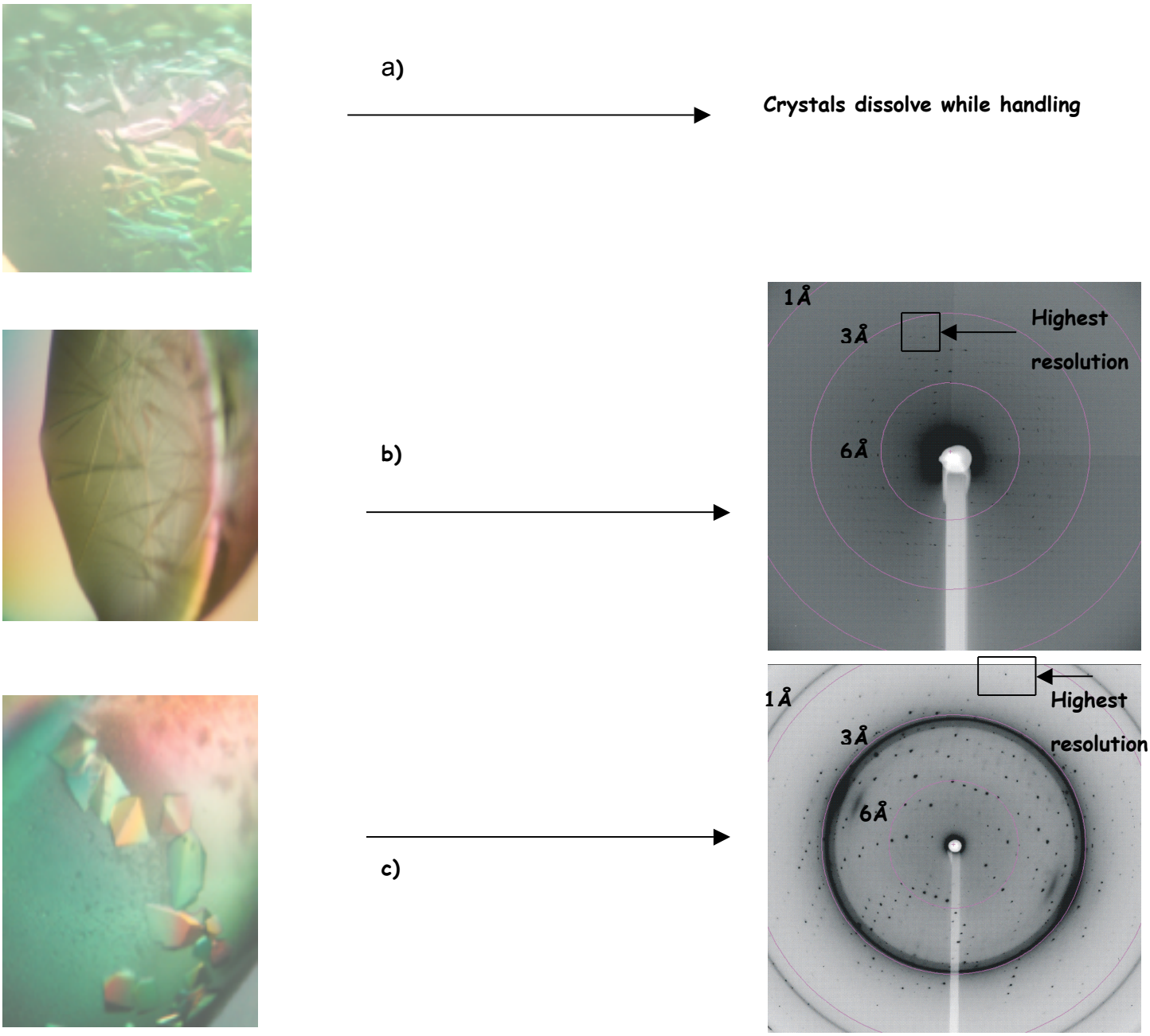

Figure 37. Crystalization of REF1-II (1-103). REF1-II (1-103) crystallized in the presence of organic precipitants: isopropanol; 1,6 Hexanediol and several PEGs. a) Crystals obtained with ispopropanol dissolved while handling. b) several PEGs gave needle crystals that diffracted weekly. c) 1,6 Hexandiol gave crystals that diffracted beyond 2.0 Å.

1.2.1) Structure determination

Well-formed pyramidal crystals grew at room temperature after 15 days, from a hanging drop equilibrated with 3.4 M 1,6-Hexanediol, 0.2 M $MgCl_2$ and 0.1 M Tris, at pH 8.5. A complete data set was recorded from a frozen crystal diffracting beyond 2 Å. The space group was determined as C222(1), with one molecule per asymmetric unit. The crystal structure was solved by molecular replacement with the program AMORE (Navaza 2001), using as a search model the protein CBP20 (Mazza *et al.*, 2001), and refined to 1.9 Å resolution with CNS (Brunger *et al.*, 1998). The missing Cα-traces and side chains on the electron density were built with the program O (Kleywegt and Jones 1999). The final model has working and free R-values of 27.7% and 30.0% respectively, and is fits well with the observed electron density, as shown by the Fo-Fc electron density map (figure 38).

Space group C222(1)

Resolution range Å = 30-1.9 (2.0-1.9)

I/σ = 8.4 (2.0)

Completeness = 99.7% (100%)

Multiplicity = 5.7 (5.1)

R_{merge}= 5.0% (27.7%)

R_{free}=30.0%

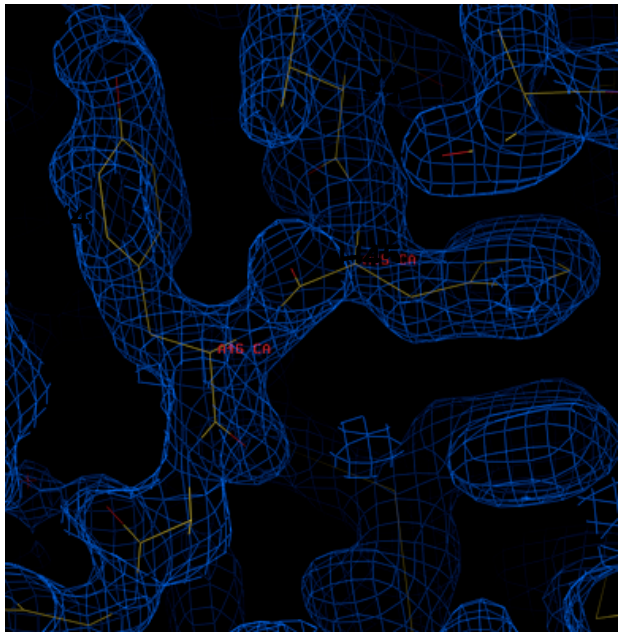

Figure 38. Structure determination of REF1-II (1-103).
Left: statistical values of the refined structure
Right: Snapshot of the Fo-Fc electron density map (contoured at σ=1) zoomed around amino acids H45-Y46.

The Rfree is slightly high, considering the resolution, which is 1.9 Å. The refined model however, has good stereochemistry, as shown by the Ramachandran plot, where all the amino acids are located within the allowed regions: 96.8% in the most favoured regions and 3.2% in the additional allowed regions (figure 39). Amino acids 1-11 and 91-103 are disordered in the structure and thus, could not be traced on the electron density map. This is very probably, the reason for such a high Rfree, in the presence of the otherwise, good statistical values of the experimental data and the refined model (figure 38, 39).

Conclusions

In agreement with the solution structure of REF/Aly, reported by Wright and colleagues (Perez-Alvarado *et al*., 2003), the structure of REF1-II (1-103) folds into a □-helix/□-strand sandwich, with a □□□□□□ topology, typical for an RNP-type RBD domain (figure 40). The X-ray structure presented here has however, an extra □-strand in loop 3 (L3) (figure 41). A region of REF's RBD, that Wright and colleagues could not define precisely by NMR spectroscopy (Perez-Alvarado *et al*., 2003). It is not clear though, if this extra □-strand is a consequence of crystal packing and/or the crystallization conditions. Further, the electrostatic surface potential of the structure was calculated and displayed as a surface cartoon with the program GRASP (Nicholls *et al*., 1991) (figure 42). Interestingly, the □-sheet platform of REF proteins does not feature hydrophobic amino acids (figure 40) and/or a positively charged surface patch on its surface (figures 40, 42), which is expected for a typical RNP-type RBD motif (Burd *et al*., 1994; Varani and Nagai 1998). This could explain the fact that the RNA-binding site of REF proteins has been mapped to the RGG variable regions and the N-/C-terminal boxes, but not to the RBD motif (Stutz *et al*., 2000).

2) Chemical shift perturbation experiments

To test the ligand recognition properties of REF proteins, a series of chemical shift perturbation experiments were performed. On each experiment, a ^{15}N labelled sample of either REF1-II (1-103) or REF1-II (1-128) was titrated with an excess of putative protein and/or ssRNA ligand (table 3). A reference $^{1}H^{N}$-^{15}N HQSC NMR spectrum of the corresponding REF1-II construct was compared with the $^{1}H^{N}$-^{15}N HSQC NMR spectrum of each titration experiment (figures 43a-c). Next, the amino acids, whose backbone amides were perturbed upon ligand addition, were docked onto the crystal structure (figure 44). The assignment of the corresponding ^{1}H-^{15}N correlations was based on the backbone chemical shift assignments available from the NMR structure of REF/Aly (Perez-Alvarado *et al*., 2003).

Initially, the chemical shift perturbation experiments were recorded for titrations of REF1-II (1-103) with the core domain of hUAP56 (figure 43a), a single stranded RNA, chosen at random (uacagg) (not shown), and the protein construct TAP96-372 (figure 43b), which has been reported not to interact with REF proteins (Stutz *et al*., 2000), and was used as a negative control during these experiments. To corroborate the titration experiments of REF1-II (1-103) with RNA, chemical shift perturbation experiments were also recorded for REF1-II (1-128), when titrated with an excess of uacagg ssRNA (figure 43c). The reason being that REF1-II (1-128) features, as mentioned before a RGG region, which is meant to participate in REF's ssRNA-binding.

TABLE 3

^{15}N- REF1-103		^{15}N- REF1-128	
10x hUAP56	+	5-10x uacagg RNA	+
10x TAP(96-372)	-		
5-10x uacagg RNA	+		

All the perturbed ^{1}H-^{15}N correlations localize to loops 3 and 5, and in part to the α-helix 2 of REF1-II (1-103) (Table 4, figure 44). These results highlight a region of REF1-II where backbone amides are perturbed in the presence of an ssRNA oligomer, uacagg and the core domain of the human protein UAP56.

TABLE 4

hUAP56 (amides perturbed)	uacagg RNA (amides perturbed)
V24	L20
T38	D21
L39	G23
K71	S25
Y73	D26
G75	R81
K 40	F36

Conclusions

As mentioned above, the amino acids perturbed upon the addition of RNA localize to loop 3 and loop 5 and very probably the RGG box (this is not known because the NMR structure of REF/Aly, from which the backbone chemical shifts have been retrieved, does not feature an RGG motif) of REF1-II (1-128), whereas those affected by the presence of UAP56 locate mostly to α2 (figure 44). It is worth to mention that the observed ^{1}H-^{15}N perturbation values are minor, which questions their biological significance, as similar ^{1}H-^{15}N perturbation shifts, can eventually be driven by protein

aggregation or fluctuations of pH and/or salt molarity in the sample. The important message, to take from these results, is that REF's RBD β-sheet platform, does not seem to interact with ssRNA, as would be expected for a typical RNP-type RBD. At least as it has been reported to date (Burd *et al*., 1994; Varani and Nagai 1998). In this context, several RNP-type RBD moieties have been described whose functional role is other (e.g. protein-recognition), rather than RNA recognition, but still have a conserved RNP-type RBD β-sheet platform, as shown from a sequence alignment comparison (figure 45), and unspecific ssRNA-binding activity (e.g.; CBP20, Mazza *et al*., 2001; U2AF35, Kielkopf *et al*., 2001, U2AF65, Selenko *et al*., 2003 and Y14, Fribourg *et al*., 2003).

The results I present here, allow for the suggestion that REF proteins are the first example where the β-sheet platform of an RNP-type RBD, does not recognize ssRNA. At least to the extent of the information presented here and the biochemical studies that have been reported for REF proteins before (Stutz *et al*., 2000). To corroborate these results further, a sound ligand-binding investigation should be performed for REF proteins, and compared to the RNA-binding pattern of at least one canonical RNP-type RBD protein. Different RNA constructs and if possible, several experimental strategies should be used, such as a combination of surface plasmon resonance, electrophoretic band-shift assays and chemical shift perturbation assays.