CRD SAT generated by pCARGHO: A new efficient lectin-based affinity tag method for safe, simple and low-cost protein purification

Alexandre Kriznik, Melissa Yelehe-Okouma, Christophe Lec, Guillaume Groshenry, Hélène Le Cordier, Christophe Charron, Marc Quinternet, Hortense Mazon, Francois Talfournier, Sandrine Boschi-Muller, et al.

CRDSAT generated by pCARGHO: A new efficient lectin-based affinity tag method for safe, simple and low-cost protein purification

Journal:	*Biotechnology Journal*
Manuscript ID	biot.201800214.R2
Wiley - Manuscript type:	Biotech Method
Date Submitted by the Author:	n/a
Complete List of Authors:	kriznik, alexandre; Université de Lorraine, CNRS, IMoPA Yéléhé-Okouma, Mélissa; Université de Lorraine, CNRS, IMoPA; CHRU Nancy, Département de Pharmacologie Clinique et Toxicologie Lec, Jean-Christophe; Université de Lorraine, CNRS, IMoPA Groshenry, Guillaume; Université de Lorraine, CNRS, IMoPA Le Cordier, Hélène; Université de Lorraine, CNRS, IMoPA Charron, Christophe; Université de Lorraine, CNRS, IMoPA Quinternet, Marc; Unité Mixte de Service (UMS) 2008, Ingénierie Biologie Santé en Lorraine (IBSLor) Mazon, Hortense; Université de Lorraine, CNRS, IMoPA Talfournier, Francois; Université de Lorraine, CNRS, IMoPA Boschi-Muller, Sandrine; Université de Lorraine, CNRS, IMoPA Jouzeau, Jean-Yves; Université de Lorraine, CNRS, IMoPA; CHRU Nancy, Département de Pharmacologie Clinique et Toxicologie Reboul, Pascal; Université de Lorraine, CNRS, IMoPA
Primary Keywords:	Downstream processing
Secondary Keywords:	Chromatography, Protein purification
Additional Keywords:	galectin-3-derived CRD-tag, Protein purification based on new tag technology, Affinity chromatography

Point by point responses

Reviewer 1 Comment 1: The authors have not addressed my economic concerns. They state in their response that the cost of Lactose Sepharose would decrease if their methodology were widely adopted. Perhaps, but there are still no numbers. I took the time to look at TEV protease prices. MilliporeSigma sells 10 KU (about 2 mg) for about $450. The instruction sheet indicates that one adds this at a ratio of about 1:100, meaning that 2 mg will cleave about 200 mg of target protein (in an hour). So cleavage of 1 gm of protein would cost $2000. This is well beyond what most pharmaceutical companies would consider for cost of goods.

Author response 1:

The price for a wide spreaded Ni sepharose resin is around 1500 EUR for 100 mL (GE Healthcare), meanwhile a Lactose Sepharose column, customized by Galab Laboratories, is at 2000 EUR for 100 mL. As mentioned before, one can easily imagine a huge reduction of this cost when produced at industrial scale.

In fact, we realized that we were not precise enough in describing how to obtain the TEV. Dr Nick Ramalanjaona gave us an available plasmid (pRK793), (ref Kapust RB, *Protein Eng*. 2001 Dec, 14(12):993-1000) that allowed us to make our own TEV production in *E. coli* (see figure below). Modifications was made in the text (paragraph 2.6).

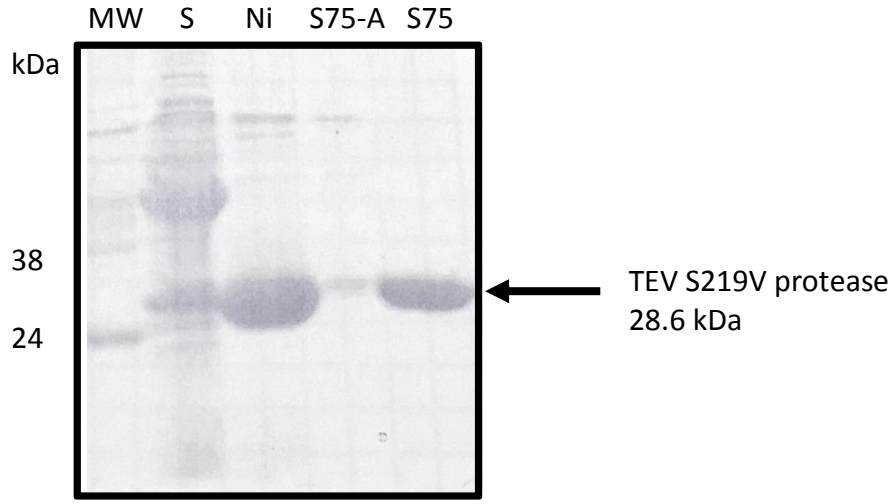


Fig: SDS Page electrophoresis of TEV protease purification steps, MW: molecular weight, S: supernatant, Ni: Ni-Sepharose column, S75-A: v0 fraction of Superdex 75 column, S75: fractions containing TEV protease, Superdex 75 column.

This experiment is easy to perform. Therefore, we are confident that other academic laboratories and small or big companies specialized in biotechnologies will be able to produce TEV as well.

This will considerably drop off the price of the cleavage step since, in our hands TEV production is about 50 mg/L in a classical LB medium, which represents according to the price given by MilliporeSigma a production of 11250 euros.

Here is an example of the estimated cost for **10 productions** (column lifetime):
A column of 100 ml = 1500 euros
Technical support: half a day (4h) at around 80 euros/h *10 = 3200
LB and buffers = 1000
Total: around 6000 euros

Each production is made from 6 L of bacterial culture, which gives a yield of 300 mg of TEV. Therefore, 10 productions will generate around 2000 to 3000 mg of TEV, if yields vary, for an estimated price of 2 to 3 euros/ mg of TEV.
In this case, the cleavage of 1 gm of protein would cost 20 to 30 euros rather than the 2000.

Biotech Method

CRD_{SAT} generated by pCARGHO: A new efficient lectin-based affinity tag method for safe, simple and low-cost protein purification

Alexandre Kriznik[1]
Mélissa Yéléhé-Okouma[1,2]
Jean-Christophe Lec[1]
Guillaume Groshenry[1]
Hélène Le Cordier[1]
Christophe Charron [1]
Marc Quinternet [3]
Hortense Mazon [1]
François Talfournier [1]
Sandrine Boschi-Muller[1]
Jean-Yves Jouzeau[1,2]
Pascal Reboul[1]

[1]Université de Lorraine, CNRS, Ingénierie Moléculaire et Pathologie Articulaire (IMoPA), F-54000 Nancy, France
[2]Département de Pharmacologie Clinique et Toxicologie, Centre Hospitalier Régional Universitaire, Nancy, France
[3]Unité Mixte de Service (UMS) 2008, Ingénierie Biologie Santé en Lorraine (IBSLor), Nancy, France

Correspondence: Pascal Reboul, PhD, UMR 7365 CNRS-Université de Lorraine, IMoPA, Biopôle de l'Université de Lorraine, Campus Biologie-Santé, 9 avenue de la forêt de Haye, BP20199, 54505 Vandœuvre-lès-Nancy cedex, France.

E-mail: pascal.reboul@univ-lorraine.fr

Keywords: Downstream processing, galectin-3-derived CRD-tag, Affinity chromatography, Protein purification based on new tag technology.

Abbreviations: **CDC25B$_{cd}$,** catalytic domain of the cell division cycle 25B; **CRD,** carbohydrate recognition domain; **CV,** column volume; **ESRRA,** estrogen related receptor alpha; **ITC,** isothermal titration calorimetry; **ESI-MS,** electrospray ionization mass spectrometry; **Gal-3,** galectin-3; **GST,** gluthathione S transferase; **MBP,** maltose binding protein; **pCARGHO,** plasmid containing a derived CArbohydrate Recognition domain of Galectin 3 from HOmo sapiens; **POI**, protein(s) of interest; **TEV,** tobacco etch virus protease; **TREM,** triggering receptor expressed on myeloid cells 1; **Trx1,** thioredoxin

Abstract

Purification of recombinant proteins remains a bottleneck for downstream processing. We engineered a new galectin 3 truncated form (CRD_{SAT}), functionally and structurally characterized, with preserved solubility and lectinic activity. Taking advantage of these properties, we designed an expression vector (pCARGHO), suitable for CRD_{SAT}-tagged protein expression in prokaryotes. CRD_{SAT} binds to lactose-Sepharose with a high specificity and facilitates solubilization of fusion proteins. This tag is structurally stable and can be easily removed from fusion proteins using TEV protease. Furthermore, due to their basic isoelectric point (pI), CRD_{SAT} and TEV are efficiently eliminated using cationic exchange chromatography. When pI of the protein of interest and CRD_{SAT} were close, other chromatographic methods were successfully tested. Using CRD_{SAT} tag, we purified several proteins from prokaryote and eukaryote origin and demonstrated as examples, the preservation of both *Escherichia coli* Thioredoxin 1 and human $CDC25B_{cd}$ activities. Overall, yields of proteins obtained after tag removal were about 5 to 50 mg per litre of bacterial culture. Our purification method displays various advantages described herein that may greatly interest academic laboratories, biotechnology and pharmaceutical companies.

1 Introduction

Expression and purification of recombinant proteins remain a great challenge for various applications in life sciences. The purification step is often the rate-limiting and time-consuming process, although it has been greatly facilitated by the development of affinity tags since decades [1, 2]. Affinity tags are peptide sequences fused with a protein of interest (POI) that allow its purification from a crude biological extract using affinity chromatography methods.

The most commonly used affinity tags are His-tag, Glutathione S transferase (GST) and Maltose-binding protein (MBP) [3, 4]. His-tagged proteins can be easily purified using the immobilized metal affinity chromatography, especially with chelated nickel, and the fused protein can be eluted with an imidazole containing buffer. However, His-tag can prevent the POIs to properly fold, especially those with solubility issues [5]. Furthermore, purification yields are poor from mammalian cell extracts and moderate from *E. coli* extracts [3], due to co-purification of bacterial proteins [6]. MBP and GST are larger tags that allow purification of fully folded proteins based on their affinities for amylose- and glutathione-resins, respectively. Those tags enhance expression, solubility and folding of the target protein [4, 5] and may provide higher amounts of recombinant protein from *E. coli* than His-tag does [3]. However, high-level expression of MBP or GST fusion proteins in *E. coli* may result in accumulation of insoluble protein aggregates in inclusion bodies [7]. In addition, MBP steric hindrance may prevent the cleavage of the fused protein whereas GST forms dimers under oxidative conditions, which may also affect the structure and function of the POI [7]. Thus, if reducing conditions are not guaranteed, the fusion protein can undergo oxidative aggregation [8].

In the present study, we developed an easy, safe and low-cost protein purification method based on an engineered part of human galectin-3 carbohydrate recognition domain (CRD), namely CRD_{SAT}, having an affinity for lactose. Lectins are identified as carbohydrate

binding proteins devoid of enzymatic activity or immune recognition. This new tag has also been biophysically, biochemically and structurally characterized.

2 Materials and methods

2.1 Subcloning of 3 truncated galectin 3 (CRDs)

Truncated forms of Gal-3 (figure 1S) were obtained from the pTrcHisGal3 plasmid, which was previously used to produce recombinant human Gal-3 [9]. These CRDs were obtained by PCR using primers listed in Table 1S, then digested with *Nco*I and *Eco*RI and cloned into the same vector after deletion of the poly-histidine sequence. These forms namely CRD_{LITIL}, CRD_{GVVP} and CRD_{SAT}, correspond to the amino acid sequence 131-250, 124-250 and 96-250 of whole Gal-3, respectively.

2.2 Production and purification of CRD_{LITIL}, CRD_{GVVP} and CRD_{SAT}.

Each vector was introduced into Rosetta (DE3) *E. coli* strain and a positive clone was pre-cultured overnight at 37°C in 3 mL of Luria Broth (LB) medium containing 50 µg/mL of ampicillin. Pre-culture was upscaled to 100 mL of LB medium and incubated at 37°C until the absorbance at 600 nm reached 0.6, then 1 mM isopropyl-D-thiogalactoside was added and the culture was pursued overnight at 37°C.

After centrifugation (4000 g for 20 min) of culture, each bacterial pellet was suspended in 10 mL of lysis buffer. The slurry was sonicated at 4°C for 2 cycles of 1 min at 80 watts then, centrifuged at 20, 000 g for 30 min and the supernatant was filtered onto a 0.45 µm PVDF membrane. The filtrate was poured onto a 1 mL lactose-Sepharose column (Galab Laboratories, Hamburg, Germany) previously equilibrated. The column was washed with 10 column volumes (CV) and elution was obtained with 1 CV (see supplemental file for buffer composition).

2.3 Design of pCARGHO

pCARGHO (CArbohydrate Recognition domain of Galectin 3 from HOmo sapiens) derives from the pET20b(+) (Novagen, Lille, France) and contains additional sequences coding for the CRD_{SAT} (242-736) and a TEV cleavage site (221-241), both located in 5' of the multiple cloning sites (159-215), where the POI sequence can be cloned (Figure 2A). Sequences dedicated to protein expression were optimized based on *E. coli* codon-usage table and the plasmid was synthesized by GeneArt Life Technologies SAS (Thermo Fisher Scientific, France).

2.4 Biochemical and biophysical characterization of CRD_{SAT}

2.4.1 Native mass spectrometry

Purified CRD_{SAT} was buffer-exchanged against 50 mM ammonium acetate buffer pH 6.8 using microcentrifuge gel filtration columns (Zeba 7 K MWCO, Thermo Scientific). Electrospray ionization mass spectrometry (ESI-MS) measurements were performed in positive-ion mode on a MicrOTOF-Q instrument (Bruker Daltonics, France). Samples were injected in the mass spectrometer at a flow rate of 6 µL/min. For the native mass spectrometry, desalted sample was diluted in 50 mM ammonium acetate buffer pH 6.8 at a final concentration of 10 µM. The ESI needle voltage was set to 4.5 kV, nebulization gas

(Ar) pressure was 2.0 bars, drying gas flow was 6 L/min, source temperature was 200°C and the capillary exit voltage was set to 173 V. The acquisition range was 1000 to 8000 m/z. For the measurements, an external calibration standard was used (TuningMix solution). The Bruker Data Analysis Software (version 4.0) was used to analyze the data and deconvolute the mass spectra.

2.4.2 Crystallization and 3D structure

Crystals of CRD_{SAT} were grown by vapor diffusion in hanging drops at 293 K. Crystals were flash frozen in liquid nitrogen in the mother liquor containing 25% glycerol as cryoprotectant. A native data set was collected at 100 K on beamline PROXIMA-1 at the Synchrotron SOLEIL (Saclay, France), with incident radiation at a wavelength of 0.9786 Å and a crystal-to-detector distance of 296 mm. Diffraction spots were recorded on a Pilatus 6M detector (Dectris Ltd.) with a 0.2° oscillation per image. Data were indexed and scaled using XDS (Table 2S), and indexed intensities were converted to structure factors using TRUNCATE in the CCP4 suite [10] without any σ cut-off. The crystal structure of CRD_{SAT} was solved by molecular replacement with the program PHASER [11] using the PDB entry 4R9B [12]. A single solution was obtained with LLG = 5668 and TFZ = 28.8. Building of the model was performed using Coot [13], and the crystal structure was refined in the range of 10-1.8 Å using REFMAC5 [14] and to the final R_{factor} of 20.1% and R_{free} of 23.8%. Coordinates of the CRD_{SAT} structure have been deposited in the Protein Data Bank (accession number 6H64).

2.4.3 Measurement of CRD_{SAT} affinity for lactose

Affinity was measured by ITC and NMR (see supplemental file for details).

2.5 Subcloning of sequences coding several proteins of interest into pCARGHO

Coding sequences of thioredoxin (Trx1), a bacterial enzyme whose activity is commonly measured in the laboratory [15], the catalytic domain of the human cell division cycle 25B ($CDC25B_{cd}$), a phosphatase [16], the human nuclear receptor estrogen related receptor alpha (ESRRA) [17], and the extracellular part of Triggering receptor expressed on myeloid cells 1 (TREM1) [18] were produced in shuttle plasmids (pMAt, GeneArt Life Technologies). Sequences were subcloned into pCARGHO using *Nco*I / *Xho*I sites and generated vectors used to produce fused recombinant proteins with the CRD_{SAT} located at the N terminal position.

2.6 Production and purification of proteins of interest using pCARGHO

Ten mL of an overnight pre-culture of transformed C41 (DE3) *E. coli* strain were added to 1L of LB medium supplemented with 50 µg/mL of ampicillin, then incubated at 37°C and protein production was induced as described in section 2.2 at 37°C or 20°C. First purification step was performed as described for CRD_{SAT} but with a column of 25 mL coupled to an ÄKTA Avant chromatographic device (GE Healthcare). Fusion proteins were cleaved overnight at 4°C using TEV S219V mutant protease produced in house (corresponding plasmid pRK793 was a gift from Dr. Nick Ramalanjaona, IMoPA lab) at a 1:100 w/w ratio. Fused POIs were separated from CRD_{SAT} and TEV protease with an appropriate second step (see supplemental file for details). POI was then concentrated onto an Amicon-YM 10 kDa concentrator (Merck Millipore, Molsheim, France) until a final concentration of 2 to 7.5 mg/mL. The identity of the protein was confirmed by ESI-MS for TREM1, whereas other proteins were checked using SDS PAGE analysis.

2.7 Biochemical characterization of purified enzymes

2.7.1 *E. coli* Trx1

After CRD_{SAT} removal, Trx1 activity was indirectly measured as follows. *Saccharomyces cerevisiae* Tsa peroxiredoxin activity was detected using 0.5 µM Tsa1, H_2O_2 (100 µM) as a substrate and a coupled system based on *E.coli* Trx1 (50 µM) / Thioredoxin reductase (1 µM) / NADPH (200 µM) and by monitoring the decrease of absorbance at 340 nm on a SAFAS UVmc2 spectrophotometer [15].

2.7.2 *Human* $CDC25B_{cd}$

$CDC25B_{cd}$ activity was assayed without removing CRD_{SAT}-tag in the presence of para-nitrophenyl phosphate (pNPP). The hydrolysis of pNPP was followed at 30° in Tris-HCl 50 mM, glycerol 10%, NaCl 50 mM, pH 8.5 by measuring the absorbance at 405 nm of p-nitrophenolate released during the reaction on a SAFAS UVmc2 spectrophotometer. Corrections were made for spontaneous hydrolysis of the pNPP under the same experimental conditions. Concentrations used were: pNPP from 0 to 50 mM and 1 µM of CRD_{SAT}-$Cdc25B_{cd}$. Experimental data were fitted to the Michaelis-Menten equation using least-squares regression analysis (SciDAVis 1.D013).

3 Results and discussion

3.1 Production of the truncated forms of Gal-3 and characterisation of their lectinic activity

We tested the length of the CRD required to produce a soluble form with lectinic capacity. The maximum production was obtained for the CRD_{LITIL} (figure 1, P1, see * at around 15 kDa) but it was neither recovered in the supernatant after centrifugation (S1) nor in the eluted fraction (lac1). This suggests that CRD_{LITIL} is not soluble and remains in the

inclusion bodies. CRD_{GGVVP} (14.6 kDa) was not efficiently produced, as no corresponding band was observed either in the pellet (P2), the soluble (S2) or the elution (lac2) fraction. Finally, CRD_{SAT} (around 20 kDa) was recovered in the elution fraction (lac3), indicating a preservation of its solubility and lectinic activity. Therefore, CRD_{SAT} was a suitable candidate for developing a new tag based on lactose-Sepharose purification. CRDs produced above derived from the human sequence of galectin-3 without any consideration for using codon bias between bacteria and human, explaining probably the weak production of CRD_{SAT}. In order to optimize the use of CRD as a tag, a R224K mutation was introduced to reduce the 3D hindrance and to provide more flexibility between CRD_{SAT} and the POI. The numbering refers to the whole human galectin 3 sequence. Finally the extra G at position 2 of CRD_{SAT} (figure 1S) was removed. Considering these modifications for the construction of pCARGHO (figure 2A), it became possible to obtain 100 mg of soluble CRD_{SAT} from 3 litres of culture (not shown). This significant yield allowed convenient downstream characterization experiments.

3.2 Biophysical and biochemical characterisation of CRD_{SAT}

As expected from the deletion of the N-terminal non-lectin domain of Gal-3 [19, 20], CRD_{SAT} was overwhelmingly detected in the monomeric state. Indeed, under non-denaturing conditions, the mass spectrum showed two narrow charge state distributions, one with a high abundant ion series (8+ to 10+), and one with a low abundant ion series (12+ to 14+) (figure 2B). Mass determination of the major ion series was 18684 Da, which is in agreement with the theoretical mass of the monomer of CRD_{SAT} without its N-terminal Met. The low abundant ion series was identified to be the non-covalent dimer of CRD_{SAT} with a determined mass of 37370 Da (Inset Figure 2B). The monomer state of CRD_{SAT} was confirmed with gel filtration (not shown). The crystal structure of CRD_{SAT} in complex with lactose was solved at 1.8 Å resolution. The asymmetric unit contained 6 monomers of

CRD_{SAT}, 6 molecules of lactose, 9 sulfate ions, and 458 water molecules. Over 95% of the residues were within the most favoured regions in a Ramachandran plot, as defined by PROCHECK [21]. As shown in figure 2C, the CRD_{SAT} structure displayed typical Gal-3 CRD structure [12, 22] with root-mean-square deviation values in the range 0.46-0.53 Å. A molecule of lactose bound to the carbohydrate-binding site of each monomer of CRD_{SAT}. The dissociation constant (K_D) of CRD_{SAT} for lactose was determined by ITC and NMR. First, the ITC data were efficiently fitted to a single-site model and led to a K_D value of 135 µM. The 1:1 binding stoichiometry obtained was in agreement with the presence of one lactose molecule interacting with a single CRD domain (figure 2D). Second, several methyl groups of the CRD_{SAT} were identified on a METHYL-SOFAST-HMQC spectrum (figure 2SA), as expected for a well-folded protein domain. Fortunately, 4 of these isolated methyl groups undergo ^1H chemical shift modifications upon lactose addition (figure 2SB), probably due to their spatial proximity with the ligand. The perturbation data were properly derived into a dissociation constant of 122 ± 6 µM, in very nice agreement with the ITC measurement (figure 2SC). This K_D value in the range of that of Gal-$3_{115-250}$ measured by Diehl et al [23] but lower than that of Gal-$3_{107-250}$ form (K_D = 1mM) [24].

3.3 Affinity purification of CRD fusion proteins, cleavage and isolation of the proteins of interest and biochemical characterisation

To validate the potency of CRD_{SAT} as a tag favouring protein purification, pCARGHO was used to produce and purify four CRD_{SAT}-fused proteins from bacterial (Trx1) or human ($CDC25B_{cd}$, ESRRA and TREM1) origin. After elution with lactose, only one or a major band was detected by SDS-PAGE corresponding to the CRD_{SAT} fusion protein (figure 3A-D, Lac). We firstly checked efficiency of TEV protease activity on one CRD_{SAT} fusion protein. Results indicated that the digestion of CRD-Trx was complete within 8 h at room temperature (figure 3SA) or overnight at 4°C (figure 3SB). Depending on the biochemical

and biophysical properties of the protein various second step purification methods were employed. After cleavage with the TEV protease (TEV), Trx1 was segregated with a cation exchange chromatography. As an acidic protein, Trx1 did not bind to the column (figure 3A, SP1), whereas CRD_{SAT} (SP2) and TEV (not detectable), which are basic proteins, did and were eluted in the presence of a high ionic strength buffer. We obtained pure and homogenous Trx1 with a yield of 50 mg per litre of culture. ESRRA, was purified using size exclusion chromatography (figure 3B) with a yield of 20 mg per litre of culture. TREM1 was separated from CRD_{SAT} and TEV with a hydrophobic interaction chromatography (figure 3C), using an ammonium sulphate decreasing gradient that eluted TREM-1 prior to TEV or CRD_{SAT}. Four mg per litre of culture were obtained. Finally, $CDC25B_{cd}$ was obtained directly from the lactose-Sepharose column with a minor contaminant at around 65 kDa, eliminated with a SP-Sepharose (figure 3D). The yield was 2.5 mg of CRD_{SAT}-$CDC25B_{cd}$ per litre of culture. As expected, all fused proteins remained able to bind to the lactose Sepharose column and were obtained with no or a low level of contaminants. This single step procedure facilitated the process of the tag elimination through a single additional step, chosen among different available technics according to the biophysical or biochemical properties of POIs. Indeed, the affinity of CRD_{SAT} for lactose was strong enough to allow the first step of purification in high saline buffers (300 to 500 mM), which prevented the non-specific binding of contaminant proteins to the resin. In addition, lactose is cheap and effluents of lactose chromatography are easily manageable. As the lectinic property of Gal-3 is divalent cation independent, the use of EDTA to prevent proteolysis was favoured. Reducing experimental conditions were also tested with success (not shown). CRD_{SAT} tag promoted solubilization of POI as shown for TREM1 recovered in the soluble fraction after bacteria lysis in contrast to other published methods [25, 26]. Moreover, TREM1 produced with our system was already used in another scientific study [18]. Other groups such as Tielker *et al* [27] used a bacterial lectin with affinity for fucose

and mannose, while Li *et al* [28] used a mushroom lectin with affinity for lactose. When we analysed the activity of Trx1 after CRD removal, the Trx1 steady state constant k_{obs} value was 0.87 ± 0.07 s^{-1}, which is close to 0.94 ± 0.05 s^{-1} described previously, with another method of purification [29]. As expected, CRD_{SAT}-$CDC25B_{cd}$ exhibited a classical Michaelis-Menten behavior with the following parameters: k_{cat} = 0.109 ± 0.002 s^{-1} and K_M = 4.9 ± 0.4 mM (figures 3E, F). These parameters are consistent with those already published (k_{cat} ~ 0.3 s^{-1} and K_M ~ 9 mM) [16, 30] for $CDC25B_{cd}$ alone indicating that in this case, the presence of the CRD_{SAT} domain did not modify the enzymatic activity. Preliminary data also demonstrate the potential usefulness of the CRD_{SAT} system in eukaryotic cells via the polyribosomal machinery (figure 4SA). Moreover, the system was not sensitive to endogenous proteases during the purification (fractions S, Ft and lac in figure 4SB & C).

4 Concluding remarks

We engineered a novel protein purification single-step system, based on the lectinic property of CRD_{SAT}, which extends the arsenal of affinity tag chromatographic methods. This tag can function under reducing conditions and in the presence of divalent cation chelators. It enables the segregation of the POIs thanks to its high specificity for lactose, a small compound easily removable and without biological toxicity. The size of CRD_{SAT} does not prevent the cleavage of the fused protein by TEV. Finally, we showed that our innovative system is also appropriate for eukaryotic cells at least by using polyribosomes.

Acknowledgement

This work was supported by la Société d'Accélération du Transfert de Technologies, SAYENS. We thanks the Platform of Biophysics and Structural Biology of UMS 2008 IBSLor, Université de Lorraine, CNRS, INSERM. We are grateful to the Synchroton SOLEIL (Saclay, France) for access to the Proxima-1 beamline. The authors acknowledge Serge Zotoglo and Jocelyne Charbonnel for their efficient technical helpful.

Conflict of interest

A.K., M.Y.O., and P.R. are listed as inventors on a patent application describing the CRD_{SAT} and pCARGHO purification system technology. (WO2017194888). License agreements are distributed by SAYENS. All other authors have no conflict of interest.

5 References

[1] K. Terpe, *Appl Microbiol Biotechnol.* **2003**, *60*, 523.

[2] M. Uhlen, B. Nilsson, B. Guss, M. Lindberg, S. Gatenbeck, L. Philipson, *Gene.* **1983**, *23*, 369.

[3] X. Zhao, G. Li, S. Liang, *J Anal Methods Chem.* **2013**, *2013*, 581093.

[4] Y. Li, *Biotechnol Appl Biochem.* **2010**, *55*, 73.

[5] D. W. Wood, *Curr Opin Struct Biol.* **2014**, *26*, 54.

[6] E. Hochuli, *J Chromatogr.* **1988**, *444*, 293.

[7] M. E. Kimple, A. L. Brill, R. L. Pasker, *Curr Protoc Protein Sci.* **2013**, *73*, Unit 9 9.

[8] A. S. Pina, C. R. Lowe, A. C. Roque, *Biotechnol Adv.* **2014**, *32*, 366.

[9] A. Janelle-Montcalm, C. Boileau, F. Poirier, J. P. Pelletier, M. Guevremont, N. Duval, J. Martel-Pelletier, P. Reboul, *Arthritis Res Ther.* **2007**, *9*, R20.

[10] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr D Biol Crystallogr.* **2011**, *67*, 235.

[11] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J Appl Crystallogr.* **2007**, *40*, 658.

[12] J. Su, T. Zhang, P. Wang, F. Liu, G. Tai, Y. Zhou, *Acta Biochim Biophys Sin (Shanghai).* **2015**, *47*, 192.

[13] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr D Biol Crystallogr.* **2010**, *66*, 486.

[14] A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr.* **2004**, *60*, 2184.

[15] A. Bersweiler, B. D'Autreaux, H. Mazon, A. Kriznik, G. Belli, A. Delaunay-Moisan, M. B. Toledano, S. Rahuel-Clermont, *Nat Chem Biol.* **2017**, *13*, 909.

[16] E. B. Gottlin, X. Xu, D. M. Epstein, S. P. Burke, J. W. Eckstein, D. P. Ballou, J. E. Dixon, *J Biol Chem.* **1996**, *271*, 27445.

[17] E. Bonnelye, P. Reboul, N. Duval, M. Cardelli, J. E. Aubin, *Arthritis Rheum.* **2011**, *63*, 2374.

[18] K. Carrasco, A. Boufenzer, L. Jolly, H. Le Cordier, G. Wang, A. J. Heck, A. Cerwenka, E. Vinolo, A. Nazabal, A. Kriznik, P. Launay, S. Gibot, M. Derive, *Cell Mol Immunol.* **2018**.

[19] A. Lepur, E. Salomonsson, U. J. Nilsson, H. Leffler, *J Biol Chem.* **2012**, *287*, 21751.

[20] C. Atmanene, C. Ronin, S. Teletchea, F. M. Gautier, F. Djedaini-Pilard, F. Ciesielski, V. Vivat, C. Grandjean, *Biochem Biophys Res Commun.* **2017**, *489*, 281.

[21] R. A. Laskowski, M. W. McArthur, D. S. Moss, J. M. Thornton, *J Appl Cryst.* **1993**, *26*, 283.

[22] K. Saraboji, M. Hakansson, S. Genheden, C. Diehl, J. Qvist, U. Weininger, U. J. Nilsson, H. Leffler, U. Ryde, M. Akke, D. T. Logan, *Biochemistry.* **2012**, *51*, 296.

[23] C. Diehl, O. Engstrom, T. Delaine, M. Hakansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, M. Akke, *J Am Chem Soc.* **2010**, *132*, 14577.

[24] J. Seetharaman, A. Kanigsberg, R. Slaaby, H. Leffler, S. H. Barondes, J. M. Rini, *J Biol Chem.* **1998**, *273*, 13047.

[25] S. Radaev, M. Kattah, B. Rostro, M. Colonna, P. D. Sun, *Structure.* **2003**, *11*, 1527.

[26] M. S. Kelker, T. R. Foss, W. Peti, L. Teyton, J. W. Kelly, K. Wuthrich, I. A. Wilson, *J Mol Biol.* **2004**, *342*, 1237.

[27] D. Tielker, F. Rosenau, K. M. Bartels, T. Rosenbaum, K. E. Jaeger, *Biotechniques.* **2006**, *41*, 327.

[28] X. J. Li, J. L. Liu, D. S. Gao, W. Y. Wan, X. Yang, Y. T. Li, H. T. Chang, L. Chen, C. Q. Wang, J. Zhao, *Protein Expr Purif.* **2016**, *119*, 51.

[29] A. Olry, S. Boschi-Muller, M. Marraud, S. Sanglier-Cianferani, A. Van Dorsselear, G. Branlant, *J Biol Chem.* **2002**, *277*, 12016.

[30] H. Bhattacharjee, J. Sheng, A. A. Ajees, R. Mukhopadhyay, B. P. Rosen, *Biochemistry.* **2010**, *49*, 802.

Figure1

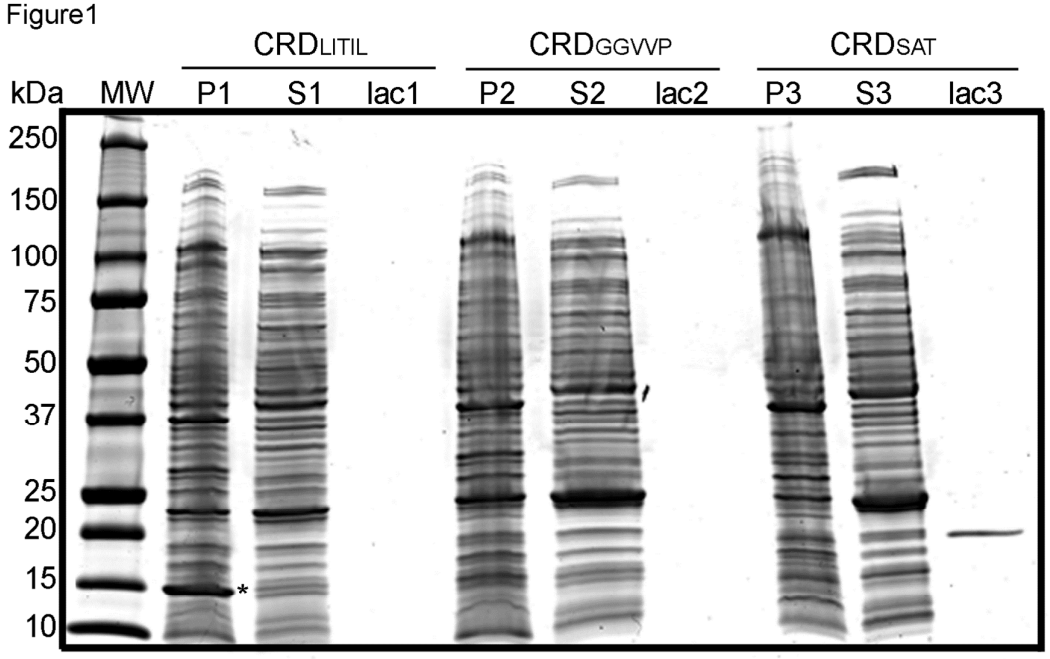


Figure 1: 4-20% SDS PAGE of the CRDs produced in bacteria and purified on lactose-Sepharose. MW: molecular weights; P1, P2, P3: resuspended pellet, S1, S2, S3: supernatant after bacterial lysis, lac1, lac2, lac3: lactose elution of CRD_{LITIL}, CRD_{GGVVP} and CRD_{SAT}, respectively.

Figure 2

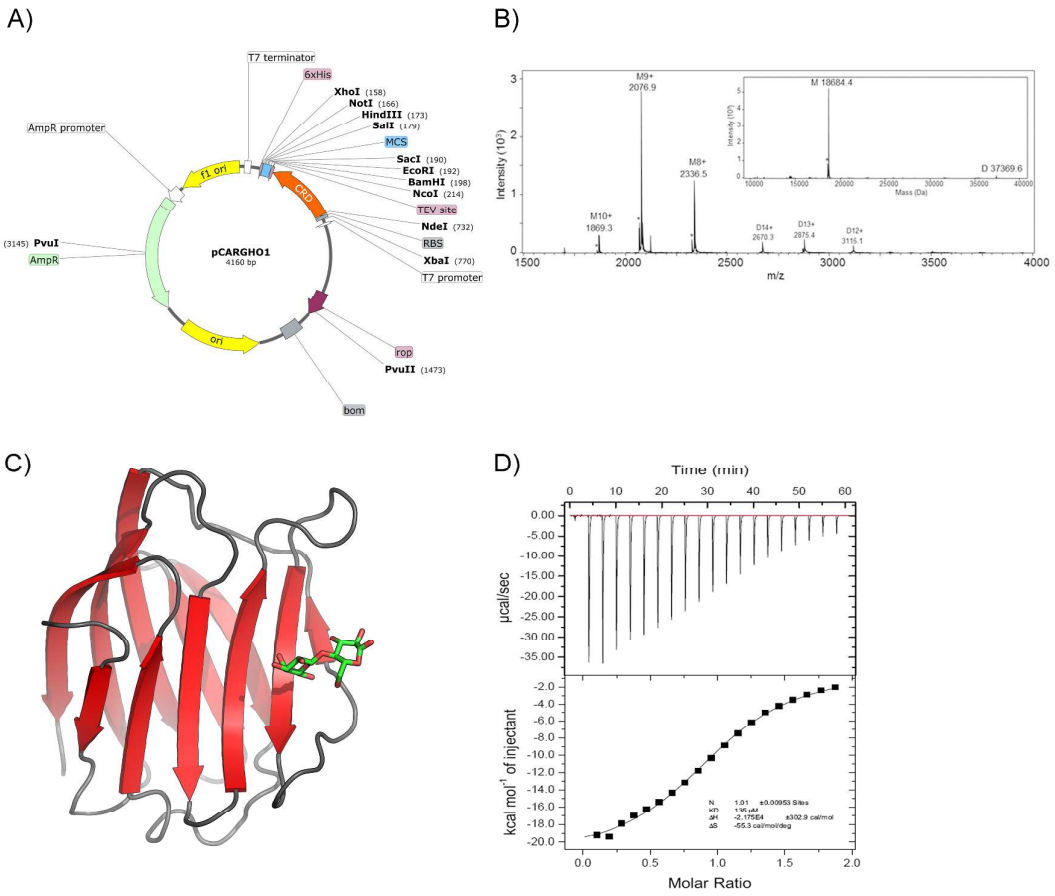


Figure 2: pCARGHO vector and characterisation of CRD_{SAT}. A) **Map of the expression plasmid pCARGHO1.** B) **Native electrospray mass spectrum of CRD_{SAT}.** Charge states are indicated with the m/z values for each peak. The inset shows the corresponding deconvoluted mass spectrum with the experimental masses in Da. M=monomer without N-terminal Met; D=dimer. * indicates a minor species that could correspond to the monomer M without the N-terminal Ser amino acid. C) **Ribbon representation of the crystal structure of CRD_{SAT}.** The lactose molecule is shown in red and green. D) **ITC data characterizing complex formation between CRD_{SAT} and lactose**. Upper panel presents the experimental ITC data and lower panel the extracted heats of binding, ΔQ, as a function of added lactose together with the fitted binding curves.

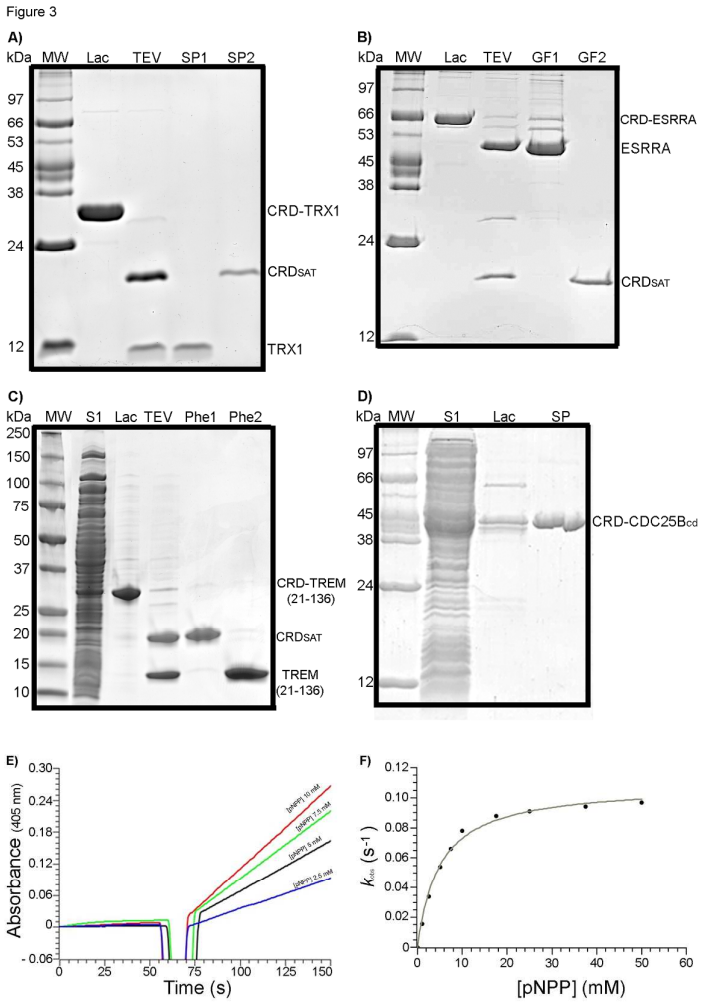


Figure 3: 4-20% SDS PAGE analyses of CRD_{SAT} fused protein and steady-state kinetic of CRD_{SAT}-$CDC25B_{cd}$. A) **Trx1,** B) **ESRRA,** C) **a truncated form of TREM1** (AA 21-136) and D) **$CDC25B_{cd}$**. Lac: lactose elution, TEV: fused protein and CRD_{SAT} after TEV protease cleavage, SP1: flow through of the sulphopropyl-Sepharose column, SP2: peak at 1 M NaCl; GF1: first peak on Superdex 75, GF2: second peak; Phe1: flow through Phenyl-Sepharose, Phe2: peak at 300 mM ammonium sulfate; S1: supernatant CRD_{SAT}-$CDC25B_{cd}$ after bacterial lysis, Phe: peak at 275 mM KCl. E) Spontaneous pNPP hydrolysis was measured and then CRD_{SAT}-$CDC25B_{cd}$ (1 µM) was added to measure initial velocity. F) Experimental data (●), representative of three independent experiments, were fitted to the Michelis-Menten equation.

Figure legends

Figure 1: 4-20% SDS PAGE of the CRDs produced in bacteria and purified on lactose-Sepharose. MW: molecular weights; P1, P2, P3: resuspended pellet, S1, S2, S3: supernatant after bacterial lysis, lac1, lac2, lac3: lactose elution of CRD_{LITIL}, CRD_{GGVVP} and CRD_{SAT}, respectively.

Figure 2: pCARGHO vector and characterisation of CRD_{SAT}. A) **Map of the expression plasmid pCARGHO1.** B) **Native electrospray mass spectrum of CRD_{SAT}.** Charge states are indicated with the m/z values for each peak. The inset shows the corresponding deconvoluted mass spectrum with the experimental masses in Da. M=monomer without N-terminal Met; D=dimer. * indicates a minor species that could correspond to the monomer M without the N-terminal Ser amino acid. C) **Ribbon representation of the crystal structure of CRD_{SAT}.** The lactose molecule is shown in red and green. D) **ITC data characterizing complex formation between CRD_{SAT} and lactose**. Upper panel presents the experimental ITC data and lower panel the extracted heats of binding, ΔQ, as a function of added lactose together with the fitted binding curves.

Figure 3: 4-20% SDS PAGE analyses of CRD_{SAT} fused protein and steady-state kinetic of CRD_{SAT}-$CDC25B_{cd}$. A) **Trx1,** B) **ESRRA,** C) **a truncated form of TREM1** (AA 21-136) and D) **$CDC25B_{cd}$**. Lac: lactose elution, TEV: fused protein and CRD_{SAT} after TEV protease cleavage, SP1: flow through of the sulphopropyl-Sepharose column, SP2: peak at 1 M NaCl; GF1: first peak on Superdex 75, GF2: second peak; Phe1: flow through Phenyl-Sepharose, Phe2: peak at 300 mM ammonium sulfate; S1: supernatant CRD_{SAT}-$CDC25B_{cd}$ after bacterial lysis, Phe: peak at 275 mM KCl. E) Spontaneous pNPP hydrolysis was measured and then CRD_{SAT}-$CDC25B_{cd}$ (1 µM) was added to measure initial velocity. F) Experimental data (●), representative of three independent experiments, were fitted to the Michelis-Menten equation.

Supporting data

Figure 1S: Optimization of the different forms of CRD obtained from human galectin 3 gene.

Amino acid sequences.

```
          1         10         20         30         40         50         60
Gal-3     MADNFSLHDALSGSGNPNPQGWPGAWGNQPAGAGGYPGASYPGAYPGQAPPGAYPGQAPPGAYPGAPGA
CRDSAT
CRDGVVP
CRDLITIL
          70         80         90        100        110        120        130
Gal-3     YPGAPAPGVYPGPPSGPGAYPSSGQPSATGAYPATGPYGAPAGPLIVPYNLPLPGGVVPRMLITILGTV
CRDSAT                              MGSATGAYPATGPYGAPAGPLIVPYNLPLPGGVVPRMLITILGTV
CRDGVVP                                                    MGGVVPRMLITILGTV
CRDLITIL                                                               MGLITILGTV
           140        150        160        170        180        190        200
Gal-3     KPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTKLDNNWGREERQSVFPFESGKPFKIQVLVEPD
CRDSAT    KPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTKLDNNWGREERQSVFPFESGKPFKIQVLVEPD
CRDGVVP   KPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTKLDNNWGREERQSVFPFESGKPFKIQVLVEPD
CRDLITIL  KPNANRIALDFQRGNDVAFHFNPRFNENNRRVIVCNTKLDNNWGREERQSVFPFESGKPFKIQVLVEPD
            210        220        230        240        250
Gal-3     HFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASYTMIStop
CRDSAT    HFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASYTMIStop
CRDGVVP   HFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASYTMIStop
CRDLITIL  HFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASYTMIStop
```

When the CRD_{GGVVP} was cloned an ATG codon was inserted whereas CRD_{LITIL} and CRD_{SAT} cloning introduced a supplementary codon between the ATG and the natural sequence of gal-3. This new codon codes for a glycine. Yellow highlighting: conserved Arg224 residue mutated into Lys.

Table 1S: ***PCR primers used for producing the 3 truncated forms of galectin-3.***

CRD_{LITIL}	S: TTAATACCATGGGTCTGATAACAATTCTGGG
	AS : GGTGGGAATTCAGATTCTTATATCATGGTATAT
CRD_{GGVVP}	S: TTAATACCATGGGGGGAGTGGTGCC
	AS : GGTGGGAATTCAGATTCTTATATCATGGTATAT
CRD_{SAT}	S : TTAATACCATGGGTAGTGCCACCGGA
	AS : GGTGGGAATTCAGATTCTTATATCATGGTATAT

Underlined scripts are restriction sites for *Nco*I and *Eco*RI in sense or antisense primers, respectively.

Chemical composition of buffers for purifications

Lactose-Sepharose column

Lysis buffer: Tris 20 mM pH 8, NaCl 300 mM, EDTA 2 mM containing 4 mM $MgSO_4$ and 5 U/mL Benzonase® (Merck, Darmstadt, Germany).

Equilibration buffer: Tris 20 mM, pH 8 containing NaCl 300 mM, EDTA 2 mM.

Washing buffer: Tris 20 mM, pH 8, NaCl 300 mM.

Elution buffer: Tris 20 mM, pH 8 containing NaCl 300 mM and lactose 100 mM.

All columns, used in a second step of purification, were purchased from GE Healthcare.

Sulphopropyl-Sepharose column

a) Separation of Trx1 from CRD_{SAT} and TEV

After cleavage by TEV, the elution fraction of lactose-Sepharose was dialysed overnight against 20 mM of MES pH 6.1

Equilibration buffer: Mes 20 mM, pH 6.1.

Buffer A: Equilibration buffer (Trx is retrieved in flow through).

Buffer B: Mes 20 mM, pH 6.1 containing NaCl 1 M (TEV and CRD_{SAT} are eluted).

b) Separation of $CDC25B_{cd}$ from CRD_{SAT} and TEV

After cleavage by TEV, the elution fraction of lactose-Sepharose is dialysed overnight against 50 mM Tris-HCl, pH 7.2 containing 10% glycerol.

Equilibration buffer: Tris-HCl 50 mM, pH 7.2 containing glycerol 10%.

Elution buffer: Tris-HCl 50 mM, pH 7.2 containing glycerol 10% and KCl 1 M.

Elution: isocratic gradient from equilibrium buffer to elution buffer.

$CDC25B_{cd}$ is eluted at 275 mM KCl.

Superdex 75 HR column

Separation of ESRRA from CRD_{SAT} and TEV

Used buffer: Tris 20 mM, pH 8 containing NaCl 150 mM.

Phenyl-Sepharose column

Separation of TREM1 from CRD_{SAT} and TEV

After cleavage by TEV, ammonium sulfate concentration is adjusted to 1 M to the reaction mix.

Equilibration buffer: Tris-HCl 20 mM, pH 8 containing ammonium sulfate 1 M.

Elution buffer: Tris-HCl 20 mM, pH 8.

Elution: isocratic gradient from equilibration to elution buffer.

TREM1 is eluted at 300 mM ammonium sulfate.

Table 2S: Crystallographic statistics

Data collection

Space group	$P6_2$
Cell parameters (Å)	a= b=118.1, c=137.5
Wavelength (Å)	0.9786
Resolution (Å)	50.0-1.8
R_{sym} (%)[a,b]	4.9 (27.4)
Completeness (%)[a]	99.9 (99.3)
<I/σ(I)>[a]	33.3 (8.7)
Multiplicity[a]	10.3 (10.0)

Refinement

Resolution (Å)	10.0-1.8
No. reflections	99705
No. free reflections	1043
R_{factor}(%)[c]	20.1
R_{free} (%)	23.8
r.m.s.d. Bonds (Å)	0.023
r.m.s.d. Angles (°)	2.387
<*B*-factors> (Å^2)	
all atoms	21.36
protein atoms	20.78
lactose	23.20
Sulfate ions	36.21
Water	32.92

[a]Number in parentheses corresponds to the last resolution shell 1.80-1.91 Å.

[b] $R_{sym} = \Sigma|I-<I>|/\Sigma I$

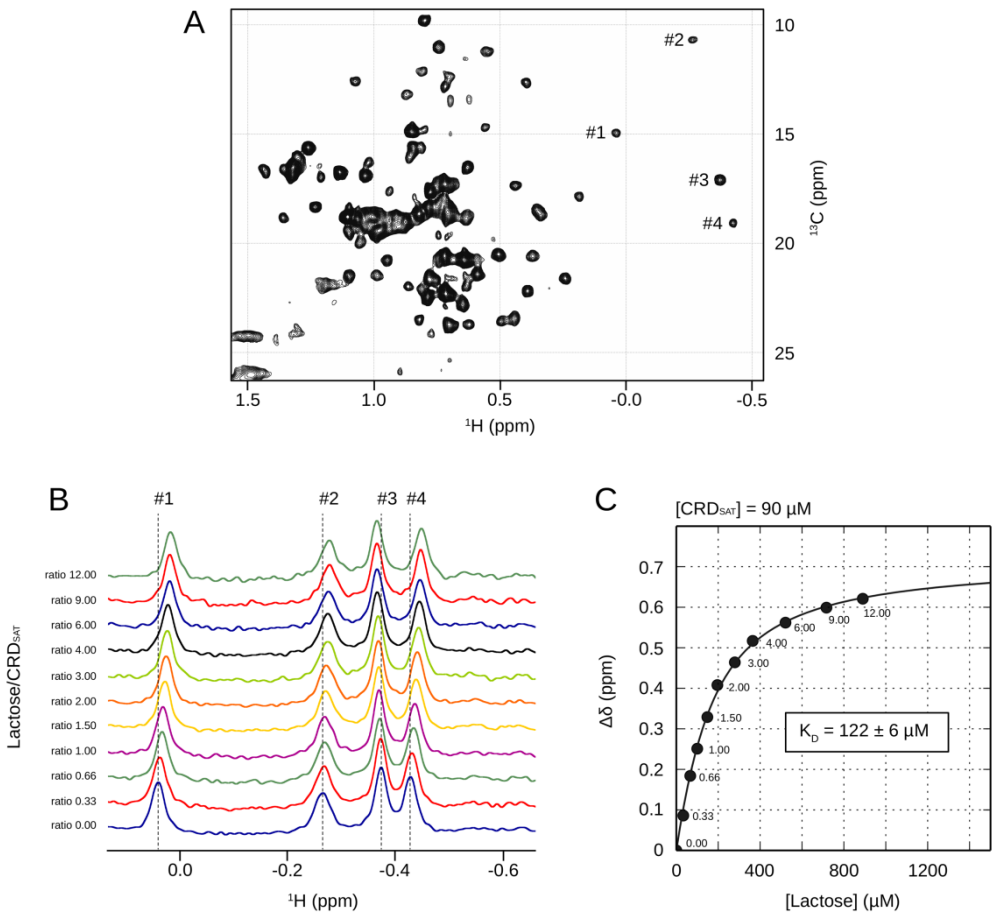
[c] $R_{factor} = \Sigma||F_{obs}|- |F_{calc}||/\Sigma|F_{obs}|$

Figure 2D and 2S: ITC and NMR-derived K_D calculation of CRD for lactose.

Materials and methods:

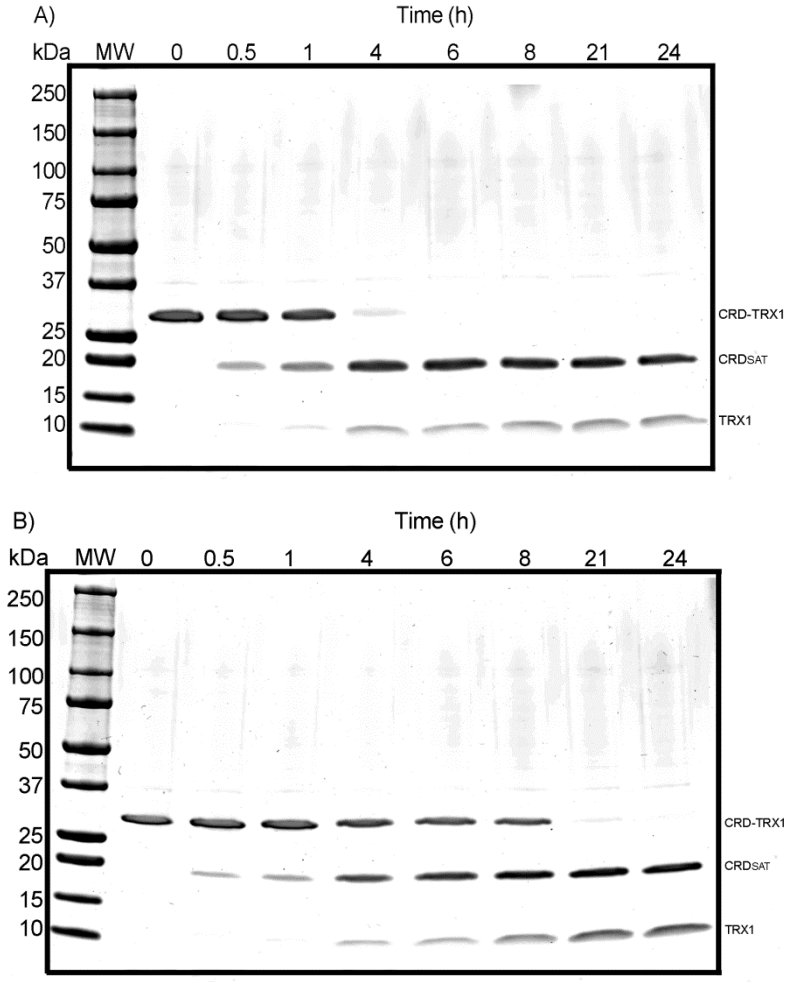
ITC was performed using a MicroCal iTC 200 (GE Healthcare, Velizy-Villacoublay, France) at 293 K. Two hundred µL of CRD_{SAT} at 150 µM were placed in the calorimeter cell, and a lactose solution at 1.5 mM was added as follows: 0.4 µL for the first injection, then 19 injections of 2 µL over 4 s at 120 s intervals. The heat of reaction per injection was determined by integration of the peak areas using the Origin software (MicroCal), which provides the values for heat of binding (ΔH), the stoichiometry of binding (n) and the dissociation constant (K_D).

For NMR experiments, CRD_{SAT} and lactose were placed into NaPi 50 mM, pH 6.4. All data were recorded at 298 K on a 600 MHz spectrometer equipped with a TCI-cryoprobe (Bruker). A METHYL-SOFAST-HMQC was recorded using the ^{13}C natural abundance of a 1 mM CRD sample. The ^{1}H chemical shift perturbations of four isolated methyl peaks were monitored upon lactose addition on a 90 µM sample of CRD_{SAT}, making it possible the determination of the K_D value (*Quinternet M, et al Chemistry, 2012 Mar 26;18(13):3969-74. doi: 10.1002/chem.201101983)*. 11 molar ratios were tested using one-dimensional ^{1}H HET-SOFAST spectra centered on methyl protons. At each point of the titration, amplitude of the ^{1}H displacement of each peak was calculated as follow: $\Delta\delta = | \delta^{1}H_i - \delta^{1}H_j |$, where $\delta^{1}H_i$ was the initial position of the peak (in ppm) and $\delta^{1}H_j$ was the position of the peak at molar ratio j (in ppm). For an averaging the results, Δδ of the four isolated methyl peaks were added together for each molar ratio tested. Finally, Δδ values were plotted against the lactose concentration. A Two-parameter nonlinear curve fitting was performed using an in-house Python script to derive the K_D value. The statistical error was estimated from a Monte Carlo simulation.



(A) METHYL-SOFAST-HMQC spectrum of the CRD$_{SAT}$ domain (concentrated at 1 mM) recorded at 600 MHz and 298 K in NaPi 50 mM, pH 6.4. Isolated peaks #1 to #4 are highlighted. (B) Series of ^{1}H HET-SOFAST spectra of the CRD domain (concentrated at 90 µM) recorded under different molar ratios (Lactose/CRD$_{SAT}$) and in the experimental conditions mentioned above. The initial positions of peaks #1 to #4 are indicated using dash lines. (C) Curve fitting and K_D value derived from cumulative ^{1}H chemical perturbations of the CRD domain. Δδ of each peak were added together to get a more accurate result. The Lactose/CRD$_{SAT}$ ratio value is reported next to each point.

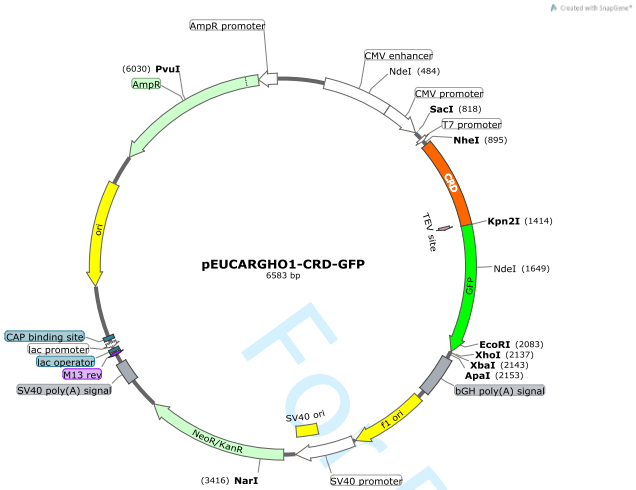
Figure 3S: 4-20% SDS-PAGE analysis of CRD-tag removal by TEV cleavage.



The cleavage reaction was carried out either at room temperature A) or 4°C B) with a fusion protein (CRD-TRX1) to enzyme ratio of 100:1 as described in the material and method section.

Figure 4S: Production and purification of CRD-GFP in HEK293 cells.

Materials and methods: pEUCARGHO derived from the pcDNA3.1-Flag was synthesized by GeneArt Life Technologies SAS (Thermo Fisher Scientific, France) and contains the following supplemental elements: CRD_{SAT} (907-1414), a sequence coding a TEV cleavage site (1396-1414), both located in 5' of the multiple cloning sites (1414-1523) where GFP was cloned (see map below).



The HEK293 cell line was routinely cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and an antibiotic mixture (100 units/mL penicillin base, 100 µg/mL streptomycin base) at 37°C in a humidified atmosphere of 5% CO_2/95% air. For transfection assays, wells of six well plates were seeded with 500,000 cells, which were left to recover overnight and then transfected with 1 µg of pEUCARGHO-CRD-GFP and 4 µl of DharmaFECT kb (Dharmacon), according to the manufacturer's instructions for 6 h. Then, medium was replaced with fresh complete DMEM and cells were harvested after 48 h. Cells were washed twice in PBS, scraped in PBS and centrifuged at 500 g for 10 min. Pellet was homogenized in PBS containing 300 mM NaCl and frozen in liquid nitrogen. After defrosting, the homogenate was sonicated at 4°C for 3 cycles of 1 min at 60 watts then, centrifuged at 20, 000 g for 1 h. The supernatant was incubated with 1 mL of lactose-Sepharose for overnight at 4°C. and purification was processed as for the prokaryotic production of the Gal-3 truncated forms. Elution was sequentially performed thrice with 1 mL of PBS containing 200 mM of lactose. Fraction from several steps of purification were analysed either by 4-12SDS-PAGE (A) or immunoblotting (B and C). Proteins were electrophoresed in 4-12% SDS-PAGE and transferred to nitrocellulose membranes using a semi dry supply (BioRad). After blotting, immunodetection was performed using a polyclonal anti-CRD antibody (dilution 1/10000) produced with our CRD_{SAT} by Proteogenix or a polyclonal anti-GFP antibody (dilution 1/10000, Genetex). Revelation was done with an anti rabbit HRP-conjugated antibody (dilution 1/20000, Thermoscientific) and LumiGLO (Eurobio) as chemiluminescent substrate.

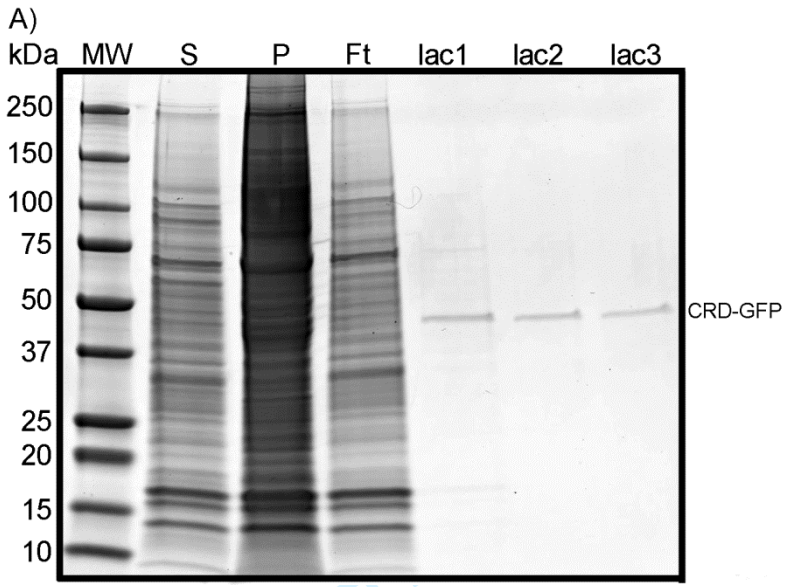


Figure 4SA: 4-20% SDS PAGE of the CRD-GFP produced in HEK cells and purified on lactose-Sepharose. MW: molecular weights, S1: supernatant after cell lysis, P: cell pellet after centrifugation at 20, 000 g. Ft: flow through of the lactose-Sepharose column, lac1, lac2, lac3: lactose elution (1mL for each fraction).

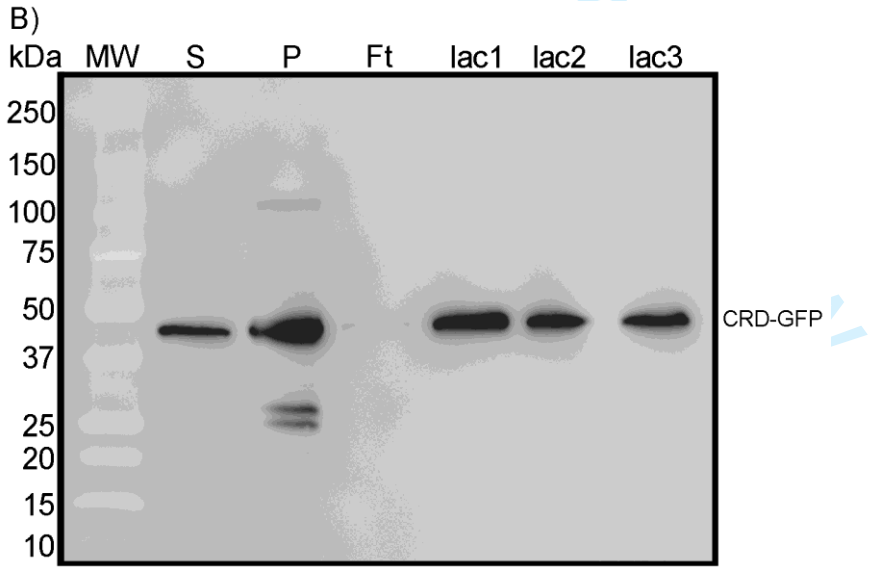


Figure 4SB: Immunodetection of the CRD-GFP with anti CRD antibody. Fractions are described in legend of the figure 4SA.

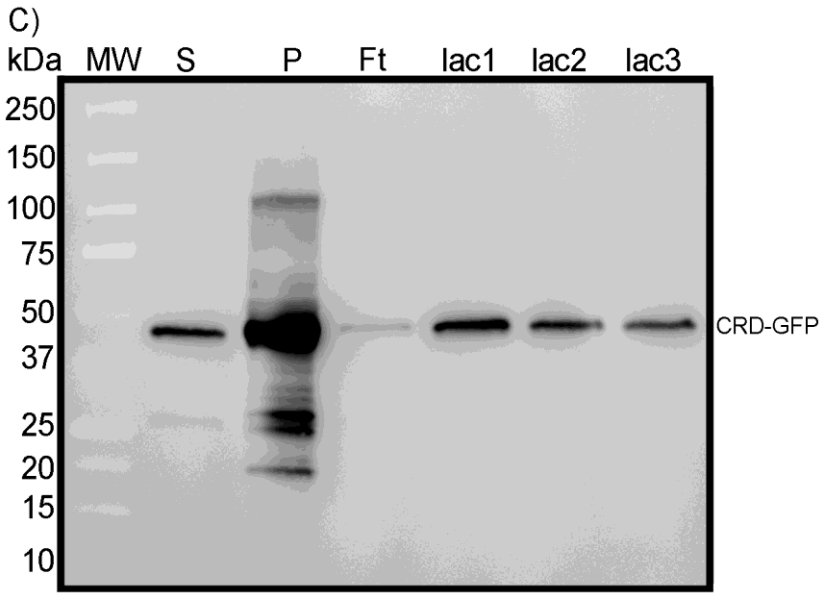


Figure 4SC: Immunodetection of the CRD-GFP with anti GFP antibody. Fractions are described in legend of the figure 4SA.