



HAL
open science

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, H el ene Le Cordier, Christophe Charron, Marc Quinternet, Hortense Mazon, Francois Talfournier, Sandrine Boschi-Muller, et al.

► To cite this version:

Alexandre Kriznik, Melissa Yelehe-Okouma, Christophe Lec, Guillaume Groshenry, H el ene Le Cordier, et al.. CRD SAT generated by pCARGHO: A new efficient lectin-based affinity tag method for safe, simple and low-cost protein purification. *Biotechnology Journal*, 2019, 14 (4), pp.e1800214. 10.1002/biot.201800214 . hal-01892441

HAL Id: hal-01892441

<https://hal.science/hal-01892441v1>

Submitted on 1 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



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

Journal:	<i>Biotechnology Journal</i>
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 Grosheury, 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).

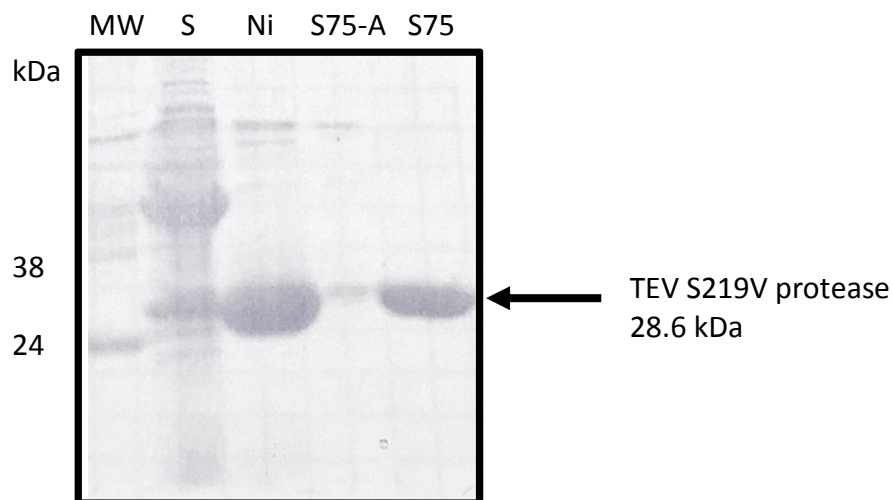


Fig: SDS Page electrophoresis of TEV protease purification steps, MW: molecular weight, S: supernatant, Ni: Ni-Sepharose column, S75-A: v0 fraction of Superdex 75 column, S75: fractions containing TEV protease, Superdex 75 column.

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.

1
2
3 This will considerably drop off the price of the cleavage step since, in our hands TEV production
4 is about 50 mg/L in a classical LB medium, which represents according to the price given by
5 MilliporeSigma a production of 11250 euros.
6
7

8
9 Here is an example of the estimated cost for **10 productions** (column lifetime):

10 A column of 100 ml = 1500 euros

11
12 Technical support: half a day (4h) at around 80 euros/h *10 = 3200

13
14 LB and buffers = 1000

15
16 Total: around 6000 euros
17
18

19 Each production is made from 6 L of bacterial culture, which gives a yield of 300 mg of TEV.
20 Therefore, 10 productions will generate around 2000 to 3000 mg of TEV, if yields vary, for an
21 estimated price of 2 to 3 euros/ mg of TEV.
22
23

24 In this case, the cleavage of 1 gm of protein would cost 20 to 30 euros rather than the 2000.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Peer Review

1
2
3
4 Biotech Method
5
6
7

8 **CRD_{SAT} generated by pCARGHO: A new efficient lectin-based affinity tag method for**
9 **safe, simple and low-cost protein purification**
10
11
12

13
14 Alexandre Kriznik¹

15 Mélissa Yéléhé-Okouma^{1,2}

16 Jean-Christophe Lec¹

17 Guillaume Groshenry¹

18 Hélène Le Cordier¹

19 Christophe Charron¹

20 Marc Quinternet³

21 Hortense Mazon¹

22 François Talfournier¹

23 Sandrine Boschi-Muller¹

24 Jean-Yves Jouzeau^{1,2}

25 Pascal Reboul¹
26
27
28
29
30
31
32

33 ¹Université de Lorraine, CNRS, Ingénierie Moléculaire et Pathologie Articulaires (IMoPA), F-
34 54000 Nancy, France

35 ²Département de Pharmacologie Clinique et Toxicologie, Centre Hospitalier Régional
36 Universitaire, Nancy, France

37 ³Unité Mixte de Service (UMS) 2008, Ingénierie Biologie Santé en Lorraine (IBSLor),
38 Nancy, France
39
40
41
42
43

44 **Correspondence:** Pascal Reboul, PhD, UMR 7365 CNRS-Université de Lorraine, IMoPA,
45 Biopôle de l'Université de Lorraine, Campus Biologie-Santé, 9 avenue de la forêt de Haye,
46 BP20199, 54505 Vandœuvre-lès-Nancy cedex, France.
47
48
49

50 **E-mail:** pascal.reboul@univ-lorraine.fr
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Keywords:** Downstream processing, galectin-3-derived CRD-tag, Affinity
5 chromatography, Protein purification based on new tag technology.
6
7
8
9

10 **Abbreviations:** **CDC25B_{cd}**, catalytic domain of the cell division cycle 25B; **CRD**,
11 carbohydrate recognition domain; **CV**, column volume; **ESRRA**, estrogen related receptor
12 alpha; **ITC**, isothermal titration calorimetry; **ESI-MS**, electrospray ionization mass
13 spectrometry; **Gal-3**, galectin-3; **GST**, glutathione S transferase; **MBP**, maltose binding
14 protein; **pCARGHO**, plasmid containing a derived CARbohydrate Recognition domain of
15 Galectin 3 from HOmo sapiens; **POI**, protein(s) of interest; **TEV**, tobacco etch virus
16 protease; **TREM**, triggering receptor expressed on myeloid cells 1; **Trx1**, thioredoxin
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4 binding proteins devoid of enzymatic activity or immune recognition. This new tag has
5 also been biophysically, biochemically and structurally characterized.
6
7
8
9

10 **2 Materials and methods**

11 12 13 14 **2.1 Subcloning of 3 truncated galectin 3 (CRDs)**

15 Truncated forms of Gal-3 (figure 1S) were obtained from the pTrcHisGal3 plasmid, which
16 was previously used to produce recombinant human Gal-3 [9]. These CRDs were obtained
17 by PCR using primers listed in Table 1S, then digested with *NcoI* and *EcoRI* and cloned into
18 the same vector after deletion of the poly-histidine sequence. These forms namely CRD_{LITIL},
19 CRD_{GVVP} and CRD_{SAT}, correspond to the amino acid sequence 131-250, 124-250 and 96-250
20 of whole Gal-3, respectively.
21
22
23
24
25
26
27
28
29

30 **2.2 Production and purification of CRD_{LITIL}, CRD_{GVVP} and CRD_{SAT}.**

31 Each vector was introduced into Rosetta (DE3) *E. coli* strain and a positive clone was pre-
32 cultured overnight at 37°C in 3 mL of Luria Broth (LB) medium containing 50 µg/mL of
33 ampicillin. Pre-culture was upscaled to 100 mL of LB medium and incubated at 37°C until
34 the absorbance at 600 nm reached 0.6, then 1 mM isopropyl-D-thiogalactoside was added
35 and the culture was pursued overnight at 37°C.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 After centrifugation (4000 g for 20 min) of culture, each bacterial pellet was suspended in
5
6 10 mL of lysis buffer. The slurry was sonicated at 4°C for 2 cycles of 1 min at 80 watts
7
8 then, centrifuged at 20, 000 g for 30 min and the supernatant was filtered onto a 0.45 µm
9
10 PVDF membrane. The filtrate was poured onto a 1 mL lactose-Sepharose column (Galab
11
12 Laboratories, Hamburg, Germany) previously equilibrated. The column was washed with
13
14 10 column volumes (CV) and elution was obtained with 1 CV (see supplemental file for
15
16 buffer composition).
17
18
19

20 **2.3 Design of pCARGHO**

21
22 pCARGHO (CARbohydrate Recognition domain of Galectin 3 from HOmo sapiens) derives
23
24 from the pET20b(+) (Novagen, Lille, France) and contains additional sequences coding for
25
26 the CRD_{SAT} (242-736) and a TEV cleavage site (221-241), both located in 5' of the multiple
27
28 cloning sites (159-215), where the POI sequence can be cloned (Figure 2A). Sequences
29
30 dedicated to protein expression were optimized based on *E. coli* codon-usage table and the
31
32 plasmid was synthesized by GeneArt Life Technologies SAS (Thermo Fisher Scientific,
33
34 France).
35
36
37

38 **2.4 Biochemical and biophysical characterization of CRD_{SAT}**

39 **2.4.1 Native mass spectrometry**

40
41
42 Purified CRD_{SAT} was buffer-exchanged against 50 mM ammonium acetate buffer pH 6.8
43
44 using microcentrifuge gel filtration columns (Zeba 7 K MWCO, Thermo Scientific).
45
46 Electrospray ionization mass spectrometry (ESI-MS) measurements were performed in
47
48 positive-ion mode on a MicrOTOF-Q instrument (Bruker Daltonics, France). Samples were
49
50 injected in the mass spectrometer at a flow rate of 6 µL/min. For the native mass
51
52 spectrometry, desalted sample was diluted in 50 mM ammonium acetate buffer pH 6.8 at a
53
54 final concentration of 10 µM. The ESI needle voltage was set to 4.5 kV, nebulization gas
55
56
57
58
59
60

1
2
3
4 (Ar) pressure was 2.0 bars, drying gas flow was 6 L/min, source temperature was 200°C
5
6 and the capillary exit voltage was set to 173 V. The acquisition range was 1000 to 8000
7
8 m/z. For the measurements, an external calibration standard was used (TuningMix
9
10 solution). The Bruker Data Analysis Software (version 4.0) was used to analyze the data
11
12 and deconvolute the mass spectra.
13
14
15

16 2.4.2 Crystallization and 3D structure

17
18 Crystals of CRD_{SAT} were grown by vapor diffusion in hanging drops at 293 K. Crystals were
19
20 flash frozen in liquid nitrogen in the mother liquor containing 25% glycerol as
21
22 cryoprotectant. A native data set was collected at 100 K on beamline PROXIMA-1 at the
23
24 Synchrotron SOLEIL (Saclay, France), with incident radiation at a wavelength of 0.9786 Å
25
26 and a crystal-to-detector distance of 296 mm. Diffraction spots were recorded on a Pilatus
27
28 6M detector (Dectris Ltd.) with a 0.2° oscillation per image. Data were indexed and scaled
29
30 using XDS (Table 2S), and indexed intensities were converted to structure factors using
31
32 TRUNCATE in the CCP4 suite [10] without any σ cut-off. The crystal structure of CRD_{SAT}
33
34 was solved by molecular replacement with the program PHASER [11] using the PDB entry
35
36 4R9B [12]. A single solution was obtained with LLG = 5668 and TFZ = 28.8. Building of the
37
38 model was performed using Coot [13], and the crystal structure was refined in the range
39
40 of 10-1.8 Å using REFMAC5 [14] and to the final R_{factor} of 20.1% and R_{free} of 23.8%.
41
42 Coordinates of the CRD_{SAT} structure have been deposited in the Protein Data Bank (accession
43
44 number 6H64).
45
46
47

48 2.4.3 Measurement of CRD_{SAT} affinity for lactose

49
50 Affinity was measured by ITC and NMR (see supplemental file for details).
51
52
53
54
55
56
57
58
59
60

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 (pMA_t, 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 μ M Tsa1, H₂O₂ (100 μ M) as a substrate and a coupled system based on *E.coli* Trx1 (50 μ M) / Thioredoxin reductase (1 μ M) / NADPH (200 μ M) 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 μ M 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 (Iac1). This suggests that CRD_{LITIL} is not soluble and remains in the

1
2
3
4 inclusion bodies. CRD_{GGVVP} (14.6 kDa) was not efficiently produced, as no corresponding
5
6 band was observed either in the pellet (P2), the soluble (S2) or the elution (lac2) fraction.
7
8 Finally, CRD_{SAT} (around 20 kDa) was recovered in the elution fraction (lac3), indicating a
9
10 preservation of its solubility and lectinic activity. Therefore, CRD_{SAT} was a suitable
11
12 candidate for developing a new tag based on lactose-Sepharose purification. CRDs
13
14 produced above derived from the human sequence of galectin-3 without any
15
16 consideration for using codon bias between bacteria and human, explaining probably the
17
18 weak production of CRD_{SAT}. In order to optimize the use of CRD as a tag, a R224K mutation
19
20 was introduced to reduce the 3D hindrance and to provide more flexibility between
21
22 CRD_{SAT} and the POI. The numbering refers to the whole human galectin 3 sequence. Finally
23
24 the extra G at position 2 of CRD_{SAT} (figure 1S) was removed. Considering these
25
26 modifications for the construction of pCARGHO (figure 2A), it became possible to obtain
27
28 100 mg of soluble CRD_{SAT} from 3 litres of culture (not shown). This significant yield
29
30 allowed convenient downstream characterization experiments.
31
32
33

3.2 Biophysical and biochemical characterisation of CRD_{SAT}

34
35
36 As expected from the deletion of the N-terminal non-lectin domain of Gal-3 [19, 20],
37
38 CRD_{SAT} was overwhelmingly detected in the monomeric state. Indeed, under non-
39
40 denaturing conditions, the mass spectrum showed two narrow charge state distributions,
41
42 one with a high abundant ion series (8+ to 10+), and one with a low abundant ion series
43
44 (12+ to 14+) (figure 2B). Mass determination of the major ion series was 18684 Da, which
45
46 is in agreement with the theoretical mass of the monomer of CRD_{SAT} without its N-terminal
47
48 Met. The low abundant ion series was identified to be the non-covalent dimer of CRD_{SAT}
49
50 with a determined mass of 37370 Da (Inset Figure 2B). The monomer state of CRD_{SAT} was
51
52 confirmed with gel filtration (not shown). The crystal structure of CRD_{SAT} in complex with
53
54 lactose was solved at 1.8 Å resolution. The asymmetric unit contained 6 monomers of
55
56
57
58
59
60

1
2
3
4 CRD_{SAT}, 6 molecules of lactose, 9 sulfate ions, and 458 water molecules. Over 95% of the
5
6 residues were within the most favoured regions in a Ramachandran plot, as defined by
7
8 PROCHECK [21]. As shown in figure 2C, the CRD_{SAT} structure displayed typical Gal-3 CRD
9
10 structure [12, 22] with root-mean-square deviation values in the range 0.46-0.53 Å. A
11
12 molecule of lactose bound to the carbohydrate-binding site of each monomer of CRD_{SAT}.
13
14 The dissociation constant (K_D) of CRD_{SAT} for lactose was determined by ITC and NMR.
15
16 First, the ITC data were efficiently fitted to a single-site model and led to a K_D value of 135
17
18 μM . The 1:1 binding stoichiometry obtained was in agreement with the presence of one
19
20 lactose molecule interacting with a single CRD domain (figure 2D). Second, several methyl
21
22 groups of the CRD_{SAT} were identified on a METHYL-SOFAST-HMQC spectrum (figure 2SA),
23
24 as expected for a well-folded protein domain. Fortunately, 4 of these isolated methyl
25
26 groups undergo ^1H chemical shift modifications upon lactose addition (figure 2SB),
27
28 probably due to their spatial proximity with the ligand. The perturbation data were
29
30 properly derived into a dissociation constant of $122 \pm 6 \mu\text{M}$, in very nice agreement with
31
32 the ITC measurement (figure 2SC). This K_D value in the range of that of Gal-3₁₁₅₋₂₅₀
33
34 measured by Diehl et al [23] but lower than that of Gal-3₁₀₇₋₂₅₀ form ($K_D = 1\text{mM}$) [24].
35
36
37

38 **3.3 Affinity purification of CRD fusion proteins, cleavage and isolation of the** 39 **proteins of interest and biochemical characterisation**

40
41
42 To validate the potency of CRD_{SAT} as a tag favouring protein purification, pCARGHO was
43
44 used to produce and purify four CRD_{SAT}-fused proteins from bacterial (Trx1) or human
45
46 (CDC25B_{cd}, ESRRA and TREM1) origin. After elution with lactose, only one or a major band
47
48 was detected by SDS-PAGE corresponding to the CRD_{SAT} fusion protein (figure 3A-D, Lac).
49
50 We firstly checked efficiency of TEV protease activity on one CRD_{SAT} fusion protein.
51
52 Results indicated that the digestion of CRD-Trx was complete within 8 h at room
53
54 temperature (figure 3SA) or overnight at 4°C (figure 3SB). Depending on the biochemical
55
56
57
58
59
60

1
2
3
4 and biophysical properties of the protein various second step purification methods were
5
6 employed. After cleavage with the TEV protease (TEV), Trx1 was segregated with a cation
7
8 exchange chromatography. As an acidic protein, Trx1 did not bind to the column (figure
9
10 3A, SP1), whereas CRD_{SAT} (SP2) and TEV (not detectable), which are basic proteins, did
11
12 and were eluted in the presence of a high ionic strength buffer. We obtained pure and
13
14 homogenous Trx1 with a yield of 50 mg per litre of culture. ESRRA, was purified using size
15
16 exclusion chromatography (figure 3B) with a yield of 20 mg per litre of culture. TREM1
17
18 was separated from CRD_{SAT} and TEV with a hydrophobic interaction chromatography
19
20 (figure 3C), using an ammonium sulphate decreasing gradient that eluted TREM-1 prior to
21
22 TEV or CRD_{SAT}. Four mg per litre of culture were obtained. Finally, CDC25B_{cd} was obtained
23
24 directly from the lactose-Sepharose column with a minor contaminant at around 65 kDa,
25
26 eliminated with a SP-Sepharose (figure 3D). The yield was 2.5 mg of CRD_{SAT}-CDC25B_{cd} per
27
28 litre of culture. As expected, all fused proteins remained able to bind to the lactose
29
30 Sepharose column and were obtained with no or a low level of contaminants. This single
31
32 step procedure facilitated the process of the tag elimination through a single additional
33
34 step, chosen among different available technics according to the biophysical or
35
36 biochemical properties of POIs. Indeed, the affinity of CRD_{SAT} for lactose was strong
37
38 enough to allow the first step of purification in high saline buffers (300 to 500 mM), which
39
40 prevented the non-specific binding of contaminant proteins to the resin. In addition,
41
42 lactose is cheap and effluents of lactose chromatography are easily manageable. As the
43
44 lectinic property of Gal-3 is divalent cation independent, the use of EDTA to prevent
45
46 proteolysis was favoured. Reducing experimental conditions were also tested with success
47
48 (not shown). CRD_{SAT} tag promoted solubilization of POI as shown for TREM1 recovered in
49
50 the soluble fraction after bacteria lysis in contrast to other published methods [25, 26].
51
52 Moreover, TREM1 produced with our system was already used in another scientific study
53
54 [18]. Other groups such as Tielker *et al* [27] used a bacterial lectin with affinity for fucose
55
56
57
58
59
60

1
2
3
4 and mannose, while Li *et al* [28] used a mushroom lectin with affinity for lactose. When we
5
6 analysed the activity of Trx1 after CRD removal, the Trx1 steady state constant k_{obs} value
7
8 was $0.87 \pm 0.07 \text{ s}^{-1}$, which is close to $0.94 \pm 0.05 \text{ s}^{-1}$ described previously, with another
9
10 method of purification [29]. As expected, CRD_{SAT}-CDC25B_{cd} exhibited a classical Michaelis-
11
12 Menten behavior with the following parameters: $k_{\text{cat}} = 0.109 \pm 0.002 \text{ s}^{-1}$ and $K_{\text{M}} = 4.9 \pm 0.4$
13
14 mM (figures 3E, F). These parameters are consistent with those already published ($k_{\text{cat}} \sim$
15
16 0.3 s^{-1} and $K_{\text{M}} \sim 9 \text{ mM}$) [16, 30] for CDC25B_{cd} alone indicating that in this case, the
17
18 presence of the CRD_{SAT} domain did not modify the enzymatic activity. Preliminary data
19
20 also demonstrate the potential usefulness of the CRD_{SAT} system in eukaryotic cells via the
21
22 polyribosomal machinery (figure 4SA). Moreover, the system was not sensitive to
23
24 endogenous proteases during the purification (fractions S, Ft and lac in figure 4SB & C).
25
26
27

28 **4 Concluding remarks**

29
30 We engineered a novel protein purification single-step system, based on the lectinic
31
32 property of CRD_{SAT}, which extends the arsenal of affinity tag chromatographic methods.
33
34 This tag can function under reducing conditions and in the presence of divalent cation
35
36 chelators. It enables the segregation of the POIs thanks to its high specificity for lactose, a
37
38 small compound easily removable and without biological toxicity. The size of CRD_{SAT} does
39
40 not prevent the cleavage of the fused protein by TEV. Finally, we showed that our
41
42 innovative system is also appropriate for eukaryotic cells at least by using polyribosomes.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Synchrotron 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.

- 1
2
3
4 [16] E. B. Gottlin, X. Xu, D. M. Epstein, S. P. Burke, J. W. Eckstein, D. P. Ballou, J. E. Dixon, *J*
5 *Biol Chem.* **1996**, *271*, 27445.
6
7
8 [17] E. Bonnelye, P. Reboul, N. Duval, M. Cardelli, J. E. Aubin, *Arthritis Rheum.* **2011**, *63*,
9 2374.
10
11
12 [18] K. Carrasco, A. Boufenzer, L. Jolly, H. Le Cordier, G. Wang, A. J. Heck, A. Cerwenka, E.
13 Vinolo, A. Nazabal, A. Kriznik, P. Launay, S. Gibot, M. Derive, *Cell Mol Immunol.* **2018**.
14
15
16 [19] A. Lepur, E. Salomonsson, U. J. Nilsson, H. Leffler, *J Biol Chem.* **2012**, *287*, 21751.
17
18 [20] C. Atmanene, C. Ronin, S. Teletchea, F. M. Gautier, F. Djedaini-Pilard, F. Ciesielski, V.
19 Vivat, C. Grandjean, *Biochem Biophys Res Commun.* **2017**, *489*, 281.
20
21
22 [21] R. A. Laskowski, M. W. McArthur, D. S. Moss, J. M. Thornton, *J Appl Cryst.* **1993**, *26*,
23 283.
24
25
26 [22] K. Saraboji, M. Hakansson, S. Genheden, C. Diehl, J. Qvist, U. Weininger, U. J. Nilsson, H.
27 Leffler, U. Ryde, M. Akke, D. T. Logan, *Biochemistry.* **2012**, *51*, 296.
28
29
30 [23] C. Diehl, O. Engstrom, T. Delaine, M. Hakansson, S. Genheden, K. Modig, H. Leffler, U.
31 Ryde, U. J. Nilsson, M. Akke, *J Am Chem Soc.* **2010**, *132*, 14577.
32
33
34 [24] J. Seetharaman, A. Kanigsberg, R. Slaaby, H. Leffler, S. H. Barondes, J. M. Rini, *J Biol*
35 *Chem.* **1998**, *273*, 13047.
36
37
38 [25] S. Radaev, M. Kattah, B. Rostro, M. Colonna, P. D. Sun, *Structure.* **2003**, *11*, 1527.
39
40
41 [26] M. S. Kelker, T. R. Foss, W. Peti, L. Teyton, J. W. Kelly, K. Wuthrich, I. A. Wilson, *J Mol*
42 *Biol.* **2004**, *342*, 1237.
43
44
45 [27] D. Tielker, F. Rosenau, K. M. Bartels, T. Rosenbaum, K. E. Jaeger, *Biotechniques.* **2006**,
46 *41*, 327.
47
48
49 [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.
50 Zhao, *Protein Expr Purif.* **2016**, *119*, 51.
51
52
53 [29] A. Olry, S. Boschi-Muller, M. Marraud, S. Sanglier-Cianferani, A. Van Dorsselear, G.
54 Branlant, *J Biol Chem.* **2002**, *277*, 12016.
55
56
57
58
59
60

1
2
3
4 [30] H. Bhattacharjee, J. Sheng, A. A. Ajees, R. Mukhopadhyay, B. P. Rosen, *Biochemistry*.
5
6 **2010**, *49*, 802.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

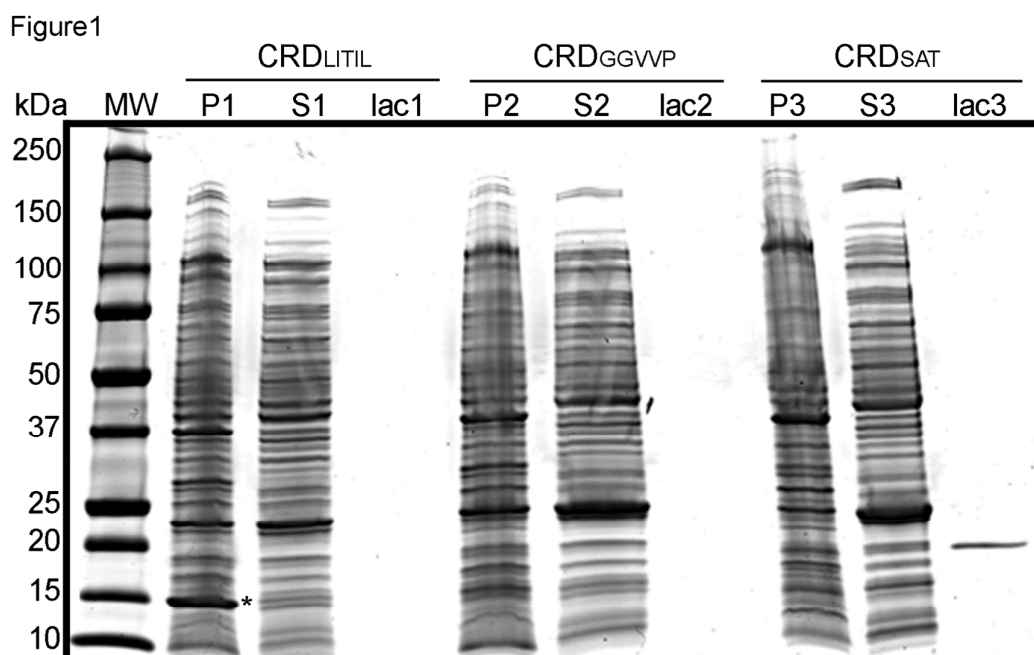


Figure 1: 4-20% SDS PAGE of the CRDs produced in bacteria and purified on lactose-Sepharose. MW: molecular weights; P1, P2, P3: resuspended pellet, S1, S2, S3: supernatant after bacterial lysis, lac1, lac2, lac3: lactose elution of CRD_{LITIL}, CRD_{GGWP} and CRD_{SAT}, respectively.

Figure 2

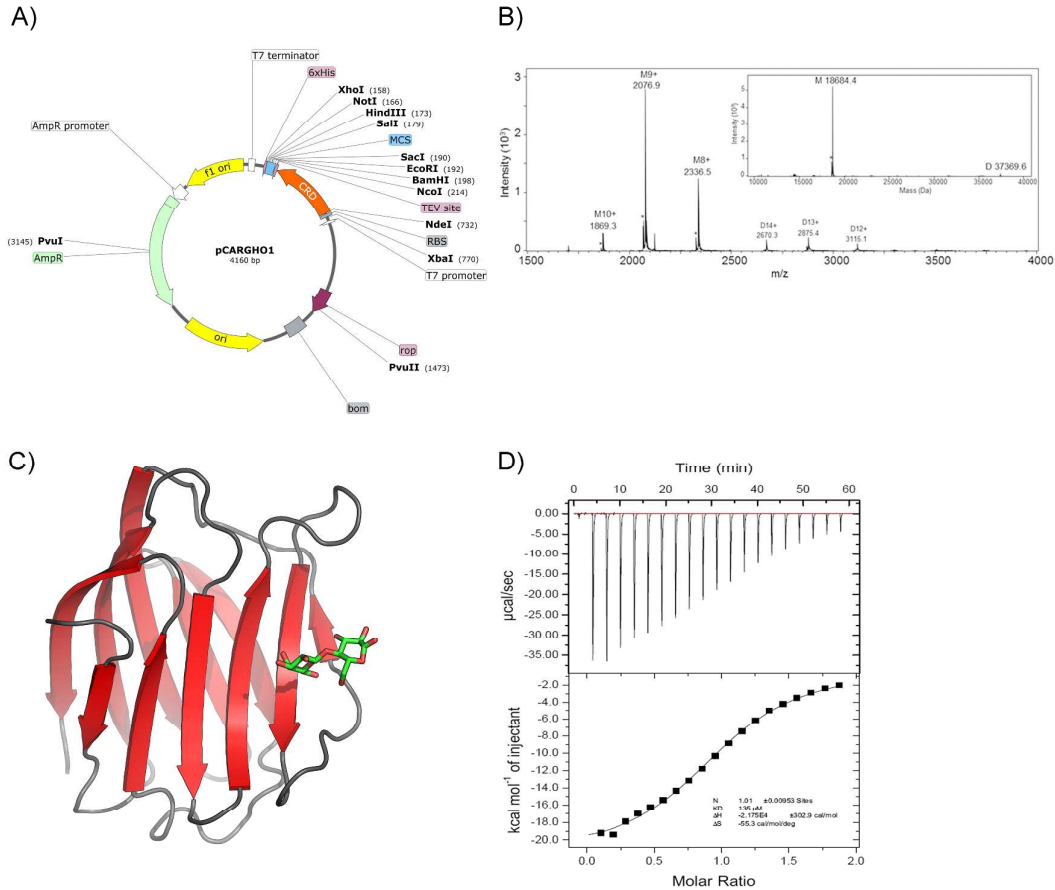


Figure 2: pCARGHO vector and characterisation of CRD_{SAT}. A) Map of the expression plasmid pCARGHO1. B) Native electrospray mass spectrum of CRD_{SAT}. Charge states are indicated with the m/z values for each peak. The inset shows the corresponding deconvoluted mass spectrum with the experimental masses 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.

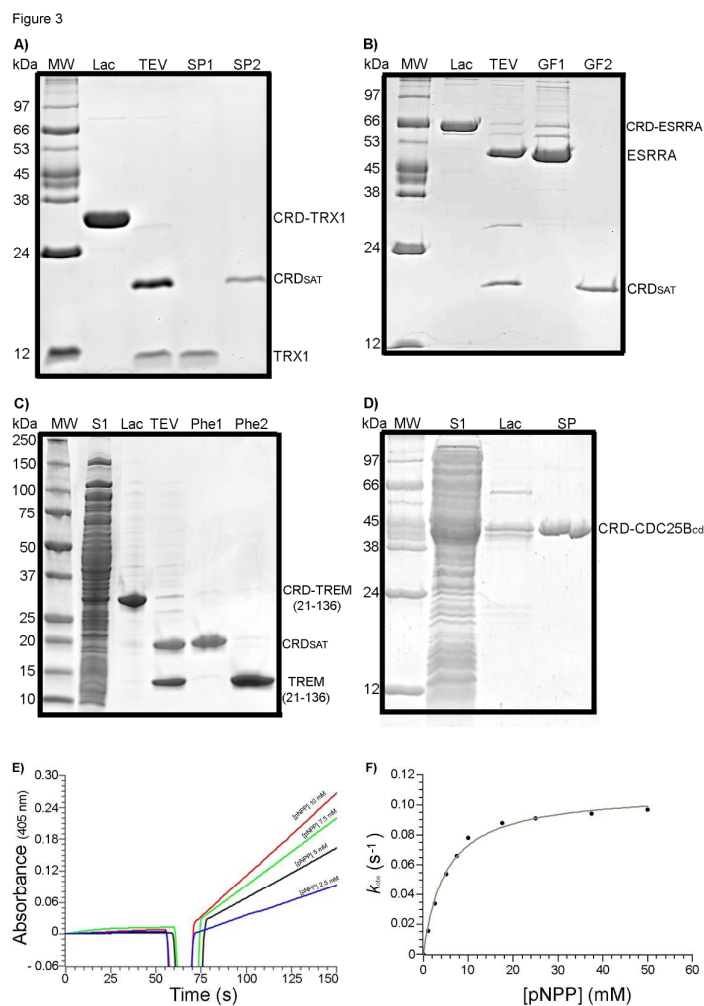


Figure 3: 4-20% SDS PAGE analyses of CRD_{SAT} fused protein and steady-state kinetic of CRD_{SAT}-CDC25B_{cd}. A) Trx1, B) ESRRA, C) a truncated form of TREM1 (AA 21-136) and D) CDC25B_{cd}. Lac: lactose elution, TEV: fused protein and CRD_{SAT} after TEV protease cleavage, SP1: flow through of the 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 Michaelis-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 μ M) was added to measure initial velocity. F) Experimental data (\bullet), representative of three independent experiments, were fitted to the Michaelis-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	MADN	FLHD	ALSG	SGNPN	PQGW	PAGW	NPQAGAGGYPGASYPGAYPGQAPP
CRD_{SAT}							
CRD_{GVVP}							
CRD_{LITIL}							
	70	80	90	100	110	120	130
Gal-3	YPGAP	PAGVY	PGPP	SGPGAY	PSSGQ	PSATG	AYPATGPY
CRD_{SAT}							
CRD_{GVVP}							
CRD_{LITIL}							
	140	150	160	170	180	190	200
Gal-3	KPNAN	RIALD	FQRG	NDVAF	HFNPR	FNENNR	RVIVC
CRD_{SAT}							
CRD_{GVVP}							
CRD_{LITIL}							
	210	220	230	240	250		
Gal-3	HFKVA	VNDA	HLLQ	YNHR	VKKL	NEISK	LGISG
CRD_{SAT}							
CRD_{GVVP}							
CRD_{LITIL}							

When the CRD_{GVVP} 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 _{GVVP}	S: TTAATACCATGGGGGAGTGGTGCC
	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 MgSO₄ 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 ESRRR 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)
$\langle I/\sigma(I) \rangle^a$	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
$\langle B\text{-factors} \rangle$ (Å ²)	
all atoms	21.36
protein atoms	20.78
lactose	23.20
Sulfate ions	36.21
Water	32.92

^aNumber in parentheses corresponds to the last resolution shell 1.80-1.91 Å.

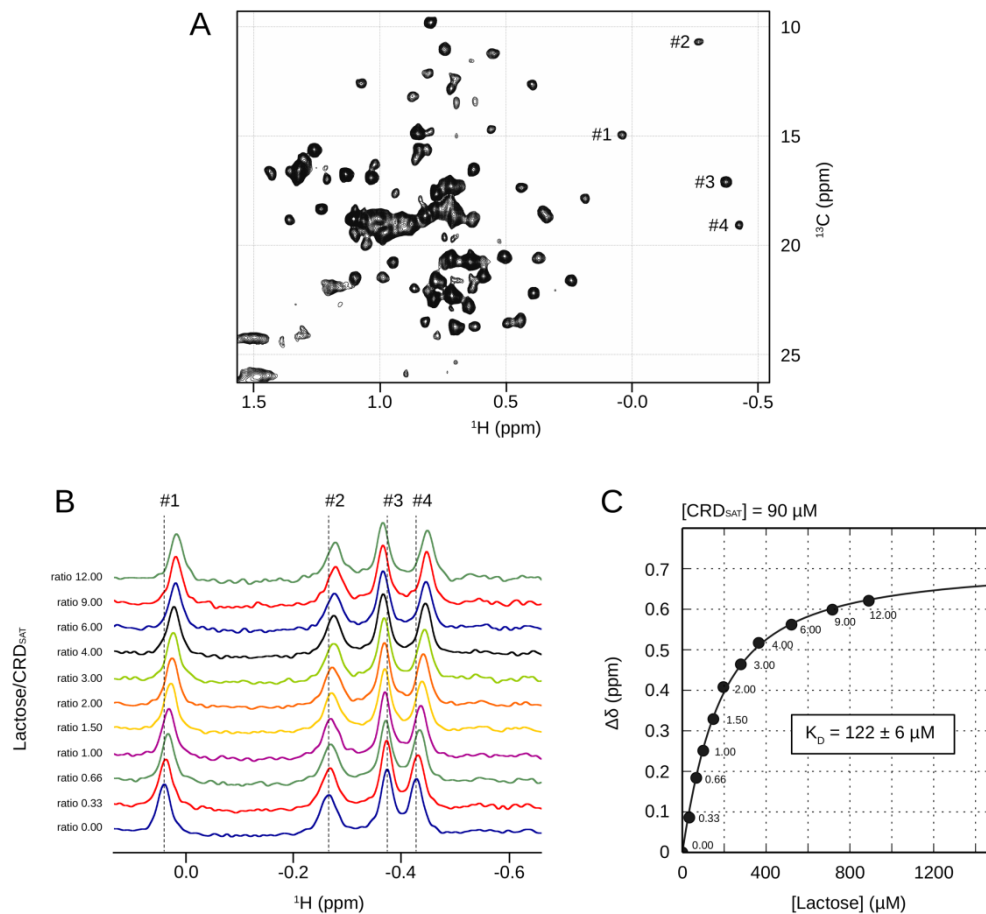
^b $R_{\text{sym}} = \Sigma |I - \langle I \rangle| / \Sigma I$

^c $R_{\text{factor}} = \Sigma ||F_{\text{obs}}| - |F_{\text{calc}}|| / \Sigma |F_{\text{obs}}|$

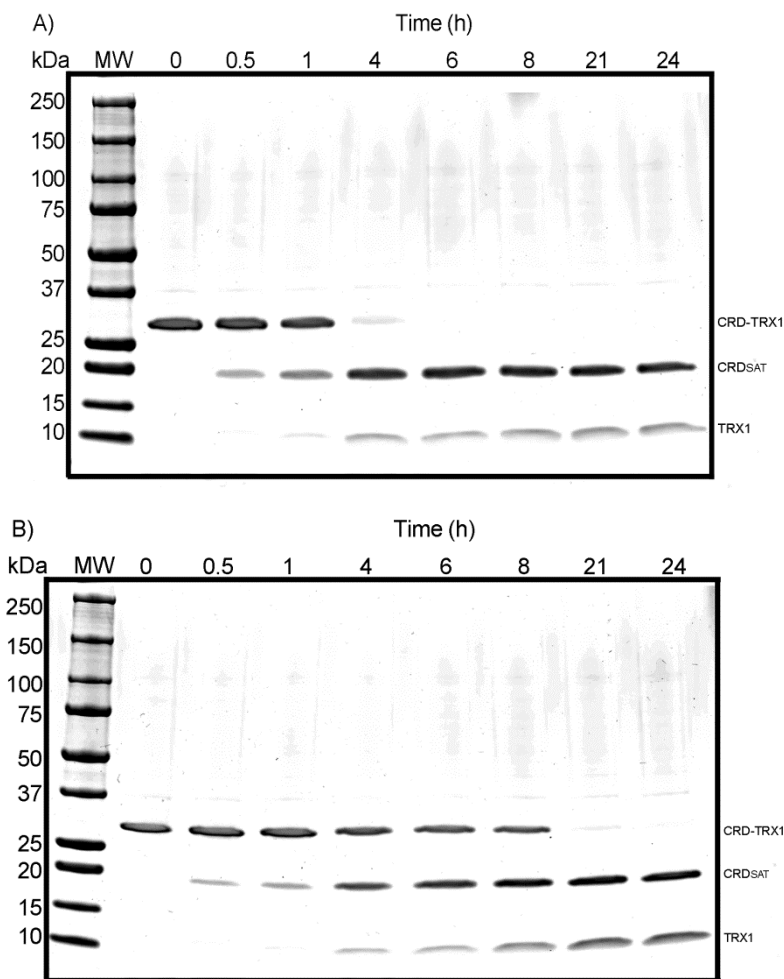
Figure 2D and 2S: ITC and NMR-derived K_D calculation of CRD for lactose.*Materials and methods:*

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 ^1H 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 ^1H HET-SOFAST spectra centered on methyl protons. At each point of the titration, amplitude of the ^1H displacement of each peak was calculated as follow: $\Delta\delta = |\delta^1\text{H}_i - \delta^1\text{H}_j|$, where $\delta^1\text{H}_i$ was the initial position of the peak (in ppm) and $\delta^1\text{H}_j$ was the position of the peak at molar ratio j (in ppm). For an averaging the results, $\Delta\delta$ of the four isolated methyl peaks were added together for each molar ratio tested. Finally, $\Delta\delta$ 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.



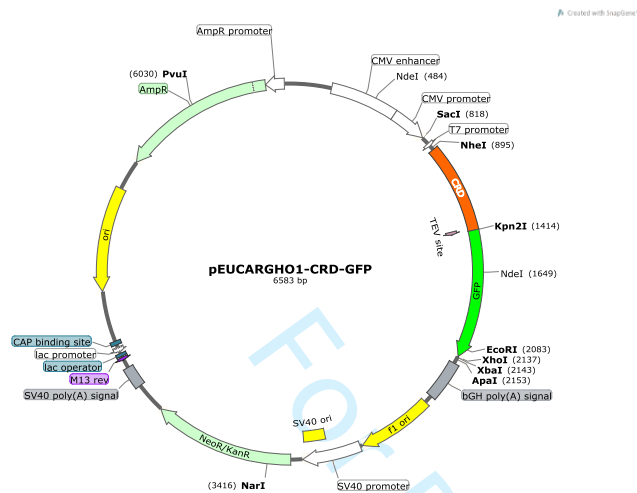
(A) METHYL-SOFAST-HMQC spectrum of the CRD_{SAT} domain (concentrated at 1 mM) recorded at 600 MHz and 298 K in NaPi 50 mM, pH 6.4. Isolated peaks #1 to #4 are highlighted. (B) Series of ¹H HET-SOFAST spectra of the CRD domain (concentrated at 90 μM) recorded under different molar ratios (Lactose/CRD_{SAT}) 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 ¹H 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.

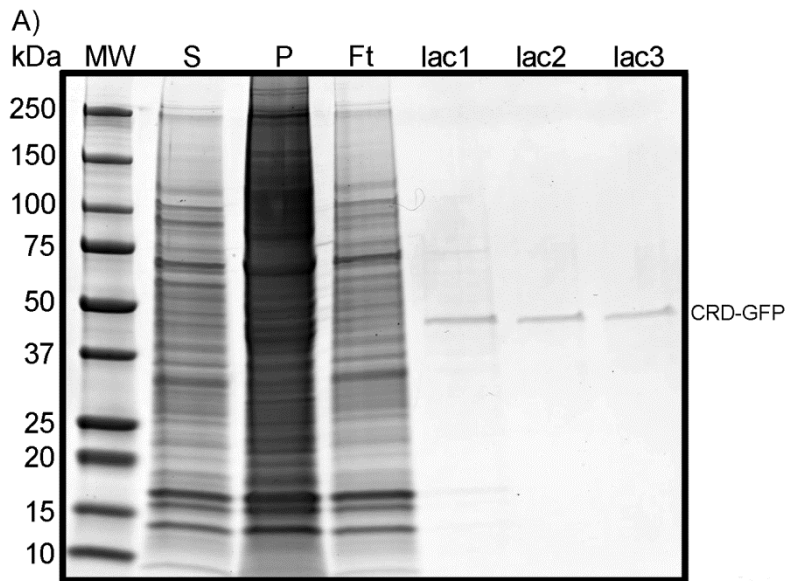
The cleavage reaction was carried out either at room temperature A) or 4°C B) with a fusion protein (CRD-TRX1) to enzyme ratio of 100:1 as described in the material and method section.

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).

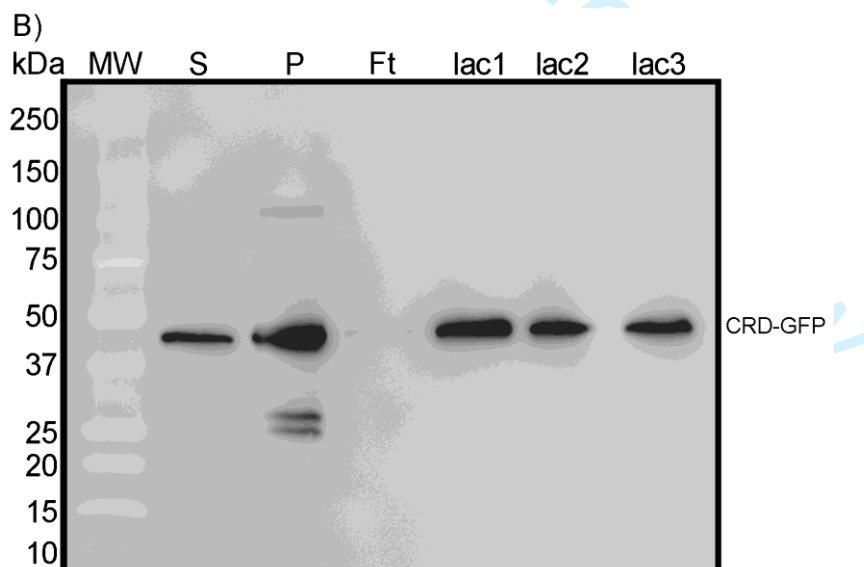


The HEK293 cell line was routinely cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and an antibiotic mixture (100 units/mL penicillin base, 100 µg/mL streptomycin base) at 37°C in a humidified atmosphere of 5% CO₂/95% air. For transfection assays, 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.



22
23
24
25
26
27
28
29
30

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).



50
51
52
53
54
55
56
57
58
59
60

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

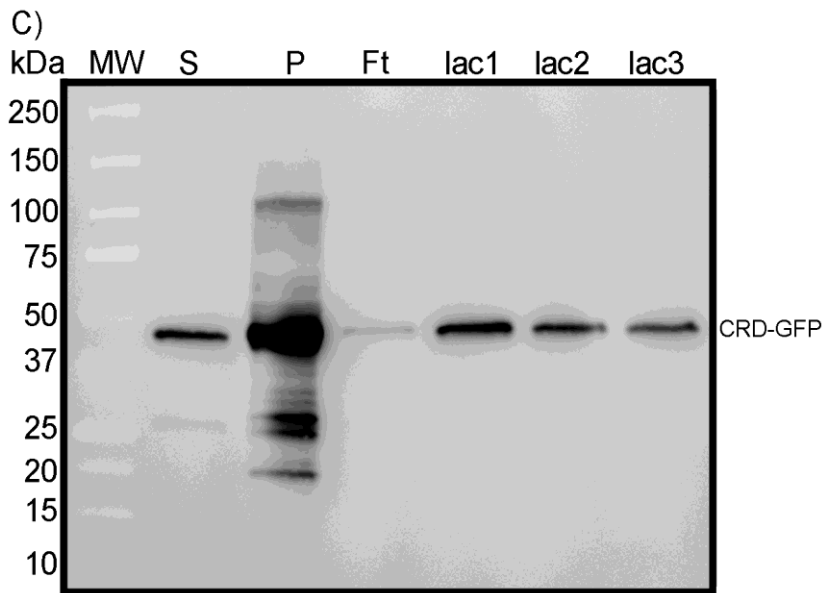


Figure 4SC: Immunodetection of the CRD-GFP with anti GFP antibody. Fractions are described in legend of the figure 4SA.