



Sep 02, 2021

Analysis of CRAC Datasets

✓ Book Chapter

📁 In 1 collection

Clémentine Delan-Forino¹, David Tollervey¹

¹Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

1 *Works for me*

🔗 Share

dx.doi.org/10.17504/protocols.io.bntjmekn

Springer Nature Books

satyavati Kharde

ABSTRACT

The RNA exosome complex functions in both the accurate processing and rapid degradation of many classes of RNA in eukaryotes and Archaea. Functional and structural analyses indicate that RNA can either be threaded through the central channel of the exosome or more directly access the active sites of the ribonucleases Rrp44 and Rrp6, but in most cases, it remains unclear how many substrates follow each pathway *in vivo*. Here we describe the method for using an UV cross-linking technique termed CRAC to generate stringent, transcriptome-wide mapping of exosome–substrate interaction sites *in vivo* and at base-pair resolution.

We present a protocol for the identification of RNA interaction sites for the exosome, using UV cross-linking and analysis of cDNA (CRAC) [1, 2]. A number of related protocols for the identification of sites of RNA–protein interaction have been reported, including HITS-CLIP, CLIP-Seq, iCLIP, eCLIP, and others [3, 4, 5, 6]. These all exploit protein immunoprecipitation to isolate protein–RNA complexes. CRAC is distinguished by the inclusion of tandem affinity purification and denaturing purification, allowing greater stringency in the recovery of authentic RNA–protein interaction sites.

To allow CRAC analyses, strains are created that express a “bait” protein with a tripartite tag. This generally consists of His6, followed by a TEV-protease cleavage site, then two copies of the z-domain from Protein A (HTP). The tag is inserted at the C terminus of the endogenous gene within the chromosome. The fusion construct is the only version of the protein expressed and this is under the control of the endogenous promoter. Several alternative tags have been successfully used, including a version with N-terminal fusion to a tag consisting of 3× FLAG-PreScission protease (PP) cleavage site-His6 (FPH) [7]. This is a smaller construct and is suitable for use on proteins with structures that are incompatible with C-terminal tagging. An additional variant is the insertion of a PP site into a protein that is also HTP tagged. This allows the separation of different domains of multidomain proteins. Importantly, the intact protein is cross-linked in the living cell, with domain separation *in vitro*. This has been successfully applied to the exosome subunit Rrp44/Dis3 to specifically identify binding sites for the PIN endonuclease domain [8].

Briefly, during standard CRAC analyses, covalently linked protein–exosome complexes are generated *in vivo* by irradiation with UV-C (254 nm). This generates RNA radicals that rapidly react with proteins in direct contact with the affected nucleotide (zero length cross-linking). The cells are then lysed and complexes with the bait protein are purified using an IgG column. Protein–RNA complexes are specifically eluted by TEV cleavage of the fusion protein and cross-linked RNAs trimmed using RNase A/T1, leaving a protected “footprint” of the protein binding site on the RNA. Trimmed complexes are denatured using 6 M Guanidinium, immobilized on Ni-NTA affinity resin and washed under denaturing conditions to dissociate copurifying proteins and complexes. The subsequent enzymatic steps are all performed on-column, during which RNA 3' and 5' ends are prepared, labeled with ³²P (to allow RNA–protein complexes to be followed during gel separation) and linkers ligated. Note, however, that alternatives to using ³²P labeling have been reported (e.g., [6]). The linker-ligated, RNA–protein complexes are eluted from the Ni-NTA resin and size selected on a denaturing SDS-PAGE gel. Following elution, the bound RNA is released by degradation of the bait protein using treatment with Proteinase K. The recovered RNA fragments are identified by reverse transcription, PCR amplification and sequencing using an Illumina platform.

Relative to CLIP-related protocols, CRAC offers the advantages of stringent purification, that substantially reduces background, and on-bead linker ligation that simplifies separation of reaction constituents during successive enzymatic steps. It also avoids the necessity to generate high-affinity antibodies needed for immunoprecipitation. Potential disadvantages are that, despite their ubiquitous use in yeast studies, tagged constructs may not be fully functional. This can be partially mitigated by confirming the ability of the tagged protein to support normal cell growth and/or RNA processing, or by comparing the behavior of N- and C-terminal tagged constructs. Additionally, because linkers are ligated to the protein–RNA complex, a possible disadvantage is that UV-cross-linking of the RNA at, or near, the 5' or 3' end it may sterically hinder on-column (de)phosphorylation and/or linker ligation. With these caveats, CRAC has been successfully applied to >50 proteins in budding yeast, and in other systems ranging from pathogenic bacteria to viral infected mouse cells [7, 9].

DOI

dx.doi.org/10.17504/protocols.io.bntjmekn

PROTOCOL CITATION

Clémentine Delan-Forino, David Tollervey 2021. Analysis of CRAC Datasets. [protocols.io](https://dx.doi.org/10.17504/protocols.io.bntjmekn)
<https://dx.doi.org/10.17504/protocols.io.bntjmekn>

COLLECTIONS ⓘ



Mapping Exosome–Substrate Interactions In Vivo by UV Cross-Linking

KEYWORDS

RNA degradation, Protein–RNA interaction, RNA-binding sites, UV cross-linking, Yeast, Exosome, RNA processing

LICENSE

This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Oct 22, 2020

LAST MODIFIED

Sep 02, 2021

OWNERSHIP HISTORY

Oct 22, 2020  Lenny Teytelman protocols.io

Jul 05, 2021  Emma Ganley protocols.io

Aug 24, 2021  Satyavati Kharde

Aug 26, 2021  satyavati Kharde

PROTOCOL INTEGER ID

43595

PARENT PROTOCOLS

Part of collection

[Mapping Exosome–Substrate Interactions In Vivo by UV Cross-Linking](#)

GUIDELINES

Here, we will describe the main steps of processing and the most commonly used modules of the pyCRAC software for our analysis.

MATERIALS TEXT

Yeast Strains and Culture Media

Yeast Strains

Purification of the RNA–protein complex requires that the protein of interest is tagged, generally with the HTP (His × 6–TEV protease cleavage site–Protein A × 2) tandem affinity tag [1,2]. In order to study RNA targets of the exosome, strains were prepared carrying tagged, intact Rrp44 and versions that lacked exonuclease or endonuclease activity, expressed from the chromosomal *RRP44* locus or from a single copy plasmid in *rrp44Δ* strains. Both were studied by CRAC to confirm that recovered RNAs are similar [10]. Then, strains expressing mutant and wild-type versions of Rrp44 from a single copy plasmid were used for CRAC.

We also tagged genomic copies of the nuclear exosome exonuclease Rrp6, the exosome core subunits Csl4 (exosome cap) and Rrp41 (exosome channel), and both wild-type and mutated components of the TRAMP complex (exosome cofactors) Mtr4, Mtr4-arch, Air1, Air2, Trf4 and Trf5. The untransformed, parental yeast strain (BY4741) was used as a negative control throughout the analyses.

Growth Media

Tryptophan absorbs 254 nm light, potentially interfering with cross-linking, and should be omitted from growth media. We use Yeast Nitrogen Base (YNB, Formedium) supplemented with 2% glucose and amino acids without tryptophan, unless other amino acids need to be omitted for plasmid maintenance.

Buffers and Solutions

To avoid potential contamination, check pH of buffers by pipetting a small volume onto pH paper.

1. Phosphate-buffered saline (PBS).
2. TN150-Lysis buffer: 50 mM Tris-HCl pH 7.8, 150 mM sodium chloride, 0.1% Nonidet P-40 substitute (Roche), 5 mM β -mercaptoethanol, one tablet of EDTA-free cOmplete protease inhibitor cocktail (Roche, 11697498001) per 50 ml solution.
3. TN1000 buffer: 50 mM Tris-HCl pH 7.8, 1 M sodium chloride, 0.1% Nonidet P-40 substitute (Roche), 5 mM β -mercaptoethanol.
4. TN150 buffer: 50 mM Tris-HCl pH 7.8, 150 mM sodium chloride, 0.1% Nonidet P-40 substitute (Roche), 5 mM β -mercaptoethanol.
5. Wash buffer I: 6 M guanidine hydrochloride, 50 mM Tris-HCl pH 7.8, 300 mM sodium chloride, 10 mM imidazole pH 8.0, 0.1% Nonidet P-40 substitute (Roche), and 5 mM β -mercaptoethanol.
6. Wash buffer II: 50 mM Tris-HCl pH 7.8, 50 mM sodium chloride, 10 mM imidazole pH 8.0, 0.1% Nonidet P-40 substitute (Roche), and 5 mM β -mercaptoethanol.
7. 1 \times PNK buffer: 50 mM Tris-HCl pH 7.8, 10 mM magnesium chloride, 0.1% Nonidet P-40 substitute (Roche), 5 mM β -mercaptoethanol
8. 5 \times PNK buffer: 250 mM Tris-HCl pH 7.8, 50 mM magnesium chloride, 25 mM β -mercaptoethanol.
9. Elution buffer: 50 mM Tris-HCl pH 7.8, 50 mM sodium chloride, 150 mM imidazole pH 8.0, 0.1% Nonidet P-40 substitute (Roche), 5 mM β -mercaptoethanol.
10. Proteinase K buffer: 50 mM Tris-HCl pH 7.8, 50 mM sodium chloride, 0.1% Nonidet P-40 substitute (Roche), and 5 mM β -mercaptoethanol, 1% sodium dodecyl sulfate (v/v), 5 mM EDTA.
11. 1 M Tris-HCl pH 7.8.
12. 0.5 M EDTA [Ethylenediaminetetraacetic acid disodium salt dihydrate] pH 8.0.
13. Guanidine HCl [Guanidinium].
14. 5 M sodium chloride.
15. 2.5 mM imidazole pH 8.0.
16. Trichloroacetic acid (TCA).
17. Acetone.
18. Methanol.
19. Proteinase K solution (20 mg/ml).
20. 3 M sodium acetate pH 5.2.
21. 25:24:1 phenol-chloroform-isoamyl alcohol mixture.
22. 100% and 70% ethanol (stored at -20 °C).
23. 10 \times TBE buffer: 890 mM Tris base, 890 mM boric acid, 20 mM EDTA.
24. Deionized water.

Enzymes and Enzymatic Reaction Components

1. TEV protease (do not use His-tagged TEV as this will be recovered on the Ni column).
2. Thermosensitive alkaline phosphatase (TSAP) (Promega, M9910).
3. RNasin RNase inhibitor (Promega, N2511, red cap).
4. T4 RNA ligase 1 (New England Biolabs, M0204S).
5. [γ -³²P] ATP (6000 Ci/mmol, Hartmann Analytic).
6. 10 mM deoxyribonucleotides (10 mM each) (Sigma-Aldrich, D7295).
7. Superscript III and accompanying 5 \times first strand buffer (Invitrogen, 18080044).
8. 100 mM DTT (Invitrogen, accompanies 18080044).
9. RNase H (New England Biolabs, M0297S).
10. LA Taq polymerase (TaKaRa, RR002M).
11. 10 \times LA Taq PCR Buffer (TaKaRa, accompanies RR002M).
12. RNase-IT (Agilent) RNase A+T1, working stock prepared by diluting 1:100 in water, store long term at -20 °C.
13. ATP, 100 mM and 10 mM solutions in water, aliquot and store at -20 °C, avoid repeated freezing and thawing.
14. T4 PNK, T4 Polynucleotide Kinase (New England BioLabs, M0201L).
15. Proteinase K (Roche Applied Science), prepare 20 mg/ml stock in deionized water, aliquot and store at -20 °C.

Oligonucleotides

All oligonucleotides were supplied by Integrated DNA Technologies (IDT) and are listed in Table 1. The forward and reverse PCR primers introduce sequences that allow binding of the PCR product to an Illumina flow cell. Illumina compatible adapters, RT and PCR primers: miRCat-33 Conversion Oligos Pack (miRCat-33 adapter and miRCat-33 RT primer, IDT), other oligonucleotides synthesized by custom order.

A	B	C
Illumina 5' adapter	L5Aa	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrArGrC-OH
	L5Ab	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrArGrC-OH
	L5Ac	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrCrGrCrArGrC-OH
	L5Ad	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrCrCrUrUrArGrC-OH
	L5Ba	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrArGrArGrC-OH
	L5Bb	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrUrGrArGrC-OH
	L5Bc	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrArCrUrArGrC-OH
	L5 Bd	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrCrUrCrUrArGrC-OH
	L5Ca	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrUrArGrC-OH
	L5Cb	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrGrGrArGrC-OH
	L5Cc	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrArCrUrCrArGrC-OH
	L5Cd	invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrArCrUrUrArGrC-OH
Illumina 3' adapter	miRCAT 33	AppTGGAATTCTCGGGTGCCAAG/ddC/
RT primer	miRCat RT	CCTTGGCACCCGAGAATT
PCR primers	P5_Fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
	PE-miRCat_Rev.	CAAGCAGAAGACGGCATACGACCTTGGCACCCGAGAATTCC

Table 1. Oligonucleotides used in CRAC experiments

After dissolving, prepare aliquots of adapters and store at -80 °C.

Laboratory Equipment

1. Incubator with orbital shaker.
2. UV cross-linker (Megatron, UV03). Megatron parts were purchased from UV03 (<http://www.uvo3.co.uk>).
3. Refrigerated centrifuge for 1 l bottles.
4. Refrigerated centrifuge for 50 ml and 15 ml centrifuge tubes.
5. Temperature controlled dry block (with range 16–65 °C) with shaking (preferentially two blocks).
6. Refrigerated microcentrifuge.
7. SDS-PAGE tank XCell SureLock Mini-Cell for NuPAGE gels.
8. Mini Trans Blot Electrophoretic Transfer Cell (wet-transfer apparatus for Western blotting) (Bio-Rad).
9. Phosphorimaging cassette.
10. Film developer.
11. Bunsen burner.
12. Thermocycler for cDNA synthesis.
13. Magnetic stirrer/hot plate.
14. Apparatus for agarose gel electrophoresis.
15. Gel scanner attached to printer, able to print gel scan in its original size.
16. Qubit 3.0 Fluorometer (Thermo Scientific).
17. Vortexer.
18. Geiger counter.
19. Laboratory room with authorization to work with radioactivity.

Other Consumables and Labware

1. Culture materials: 50 ml and 500 ml flasks for preculture, 4 l flasks for culture.
2. Filter units for buffer sterilization with pore size 0.2 µm.
3. RNase-free filter pipette tips.
4. SD medium: CSM -Trp and CSM -Trp -Leu (Formedium) for strains requiring plasmid maintenance with Leucine auxotrophic marker with 2% glucose and yeast nitrogen base (3 l of medium per sample).
5. 0.1 mm Zirconia beads.
6. IgG Sepharose®6 Fast Flow (GE Healthcare, 17-0969-01).
7. Spin columns (Pierce, Snap Cap).
8. Ni-NTA resins (Qiagen, 30210).
9. 1.5 ml microcentrifuge tubes.
10. GlycoBlue (Ambion, AM9515) or glycogen for RNA/Protein precipitation.
11. NuPAGE bis-Tris 4–12% precast gradient gels (Invitrogen, NP0322BOX). This system is essential due to its high pH stability through the run.

12. NuPAGE LDS Sample Buffer, 4× (Life Technologies).
13. MOPS running buffer (Invitrogen, NP0001).
14. NuPAGE transfer buffer (Invitrogen, NP0006).
15. Nitrocellulose membranes (Thermo Scientific or GE Healthcare).
16. Phosphorescent rulers for autoradiography.
17. Kodak BioMax MS Autoradiography Film.
18. DNA Gel extraction kit with low elution volumes (e.g., MinElute Gel extraction kit (Qiagen)).
19. Transparency film.
20. MetaPhor high resolution agarose (Lonza, 50181).
21. SYBR Safe (Life Technologies, S33102).
22. 50 bp DNA ladder (e.g., GeneRuler 50) and loading dye (e.g., GeneRuler DNA Ladder Mix by Thermo Scientific, SM0331).
23. Prestained protein standard SeeBlue Plus2 (Life Technologies, LC5925).
24. Scalpels.
25. Qubit dsDNA HS Assay Kit (Life Technologies, Q32851).

SAFETY WARNINGS

Please refer to the Safety Data Sheets (SDS) for health and environmental hazards.

BEFORE STARTING

Analysis of sequences obtained from exosome subunits CRAC experiments is done using custom scripts and software packages. The pyCRAC [11] software, a suite of python scripts which can be used to analyze sequencing data obtained from protein–RNA UV cross-linking protocols, includes most of the necessary tools. Here, we will describe the main steps of processing and the most commonly used modules of the pyCRAC software for our analysis.

Preprocessing Step: Demultiplexing, Quality Filtering, Trimming of Adapters

1

The 5' adapters mentioned in previous sections contain barcodes allowing multiplexing of several samples in a sequencing lane. In addition to barcodes, 5' adapters contain three random nucleotides allowing removal of PCR duplicates. This allows detection of reads with the same start and end positions that arise from PCR duplication of a single cDNA rather than independent linker ligation events.

2 For multiplexed samples, first split the output file from sequencing by barcodes, using pyCRAC package:

```
$ pyBarcodeFilter.py -b barcodes.list -f multiplexed_input.fastq
```

where barcodes.list is a tab-delimited text file containing the list of barcodes used in the experiment with corresponding names of samples, used in output files names.



Here is an example of how the file should appear:

A	B
NNNTAAGC	Rrp44-HTP_L5Aa
NNNATTAGC	Rrp6-HTP_L5Ab
NNNGTGAGC	Rrp44-exo-HTP_L5Bb

The random nucleotides will be stripped in this step and will be placed into the header of each sequence of the output fastq files. Later steps can make use of this information in order to collapse PCR duplicates (see step 5).

It is important to note that the standard version of this script requires the adapters to be designed as shown in

Table 1.

- 3 Sequencing data can then be quality filtered and adapters trimmed using [Flexbar](#) [12] with parameters `-at 1 -ao 4`:

```
$ flexbar -r input.fastq -f solexa -as TGGGAATTCTCGGGTGCCAAGG -at 1 -ao 4 -u 3 -m 7 -n 16 -t
flexbar.fastq
```

where `input.fastq` and `flexbar.fastq` are the input and output fastq files names respectively.

4

When useful, for instance when proportion of 3' oligoadenylated reads must be calculated (see step 9), “-g” parameter can be added to tag reads with 3' adapter. Then “grep” can be used to retain only these reads:

```
$ grep -A 3 --no-group-separator removal flexbar.fastq > flexbar_adaptercontaining.fastq;
done &
```

Collapsing

5

Then, sequences can be collapsed, thanks to the random nucleotides present in 5' linker as mentioned in step 2, using `pyFastqDuplicateRemover.py` script from `pyCRAC` software, so that reads having identical ends and identical random nucleotides in the 5' barcode are counted as one:

```
$ nohup pyFastqDuplicateRemover.py -f flexbar.fastq -o flexbar_comp.fasta &
```

where `flexbar_comp.fasta` is the collapsed output file.

This step can be skipped if the analysis aims to study ribosomal RNA. Indeed, with the linkers mentioned above, collapsing allows to keep only 64 alternatives sequences (3 random nucleotides = 4^3 possibilities); since the exosome strongly binds to pre-rRNA, collapsing would lead to flattening exosome binding peaks across pre-rRNA. However, this step is essential for study of exosome binding on RNA polymerase II transcripts.

Alignment

- 6 Reads should then be aligned to the *Saccharomyces cerevisiae* genome (SGD v64) using `Novoalign` (Novocraft) with genome annotation from Ensembl (EF4.74) [13], supplemented with noncoding sequences as described [14], with parameters `-r Random`:

```
$ novoalign -f flexbar_comp.fasta -s 1 -r Random -d
Saccharomyces_cerevisiae.EF4.74.novoindex > flexbar_comp.novo
```

where `Saccharomyces_cerevisiae.EF4.74.novoindex` is the genome-specific index file generated by `novoindex`, and `flexbar_comp.novo` is the output file name.

The “-r Unique” or “-r All” parameters are useful especially for study of exosome binding across tRNAs which share common sequences [10]. “-r Unique” will lead to preferential loss of a subset of sequences (e.g., ribosomal sequences which are represented by two identical *RDN37* sequences in the yeast reference genome).

By default, `NovoAlign` filters out all reads shorter than 17 nt (as shorter reads are unlikely to map uniquely to the yeast genome). For datasets obtained from `Rrp44 CRAC`, it was useful to align shorter sequences [15] enriched for

species targeted to Rrp44 exonuclease site and bypassing the exosome channel (Rrp44 protects 9 nt while exosome + Rrp44 protects 31–33 nt). In some analyses, we then used “-l 9” parameter (instead of -l 17 default).

Counting Overlaps with Genomic Features

- 7 To study distribution of reads across the genome, use `pyReadCounters.py` from the `pyCRAC` package. A GTF format file for genome annotation is required by the `pyCRAC` software and is critical to the interpretation of the output of the `pyCRAC` pipeline. `pyCRAC` is sensitive to the formatting within the GTF file and we find it useful to check the annotated GTF file using the `pyCheckGTFfile.py` command to ensure that the GTF file is suitable for use with the `pyCRAC` software:

```
$ pCheckGTFfile.py --gtf annotation.gtf --o annotation_checked.gtf
```

where `annotation.gtf` is a GTF format file of the genome annotation.

```
$ pyReadCounters.py -f flexbar_comp.novo --gtf=annotation_checked.gtf --rpkm
```



The output files are

1. a `gtf` file that can be used as input files in numerous analyses within `pyCRAC` package.
2. a hit table file presenting the counts of reads mapped to each genomic feature within each defined RNA class in absolute value and read number normalized per kilobase per millions (if `-rpkm` parameter is specified in the command line).

Distribution along Genes

- 8 To observe binding distribution of exosome subunits across individual genes, use `pyPileup.py` from the `pyCRAC` package. The output is a tab-delimited file that can be plotted to obtain a visual overview of binding along the gene of interest. This gives particularly good quality plots for RNAs that are strongly targeted by the exosome:

```
$ pyPileup.py -f flexbar_comp.novo --gtf=annotation_checked.gtf --tab=sequence.tab -g gene.list & -r 0
```

where `sequence.tab` is a tab-delimited file with genes name and sequences and `gene.list` is a text file with the names of genes for which you want to generate output files.

`-r` parameter allows the user to indicate the length of flanks to be added on 5' and 3' ends of genes.

To study binding across a particular class of RNA, metagene plots are generated. We used custom-made scripts, still not available online. However, the `computeMatrix`, `plotProfile`, and `plotHeatmap` modules of the `deepTools` software allow for similar analyses [16].

Oligo-A Reads

- 9 Selection of reads containing 3' nonencoded A tracks, allows identification of targets oligoadenylated by TRAMP prior binding of the exosome. We use custom-made scripts giving as output files:
1. a `fasta` file containing only oligo-A reads, used for downstream analyses
 2. a text file with the ratio of oligo-A to total reads, and
 3. a text file with the list of nonencoded 3' tails.