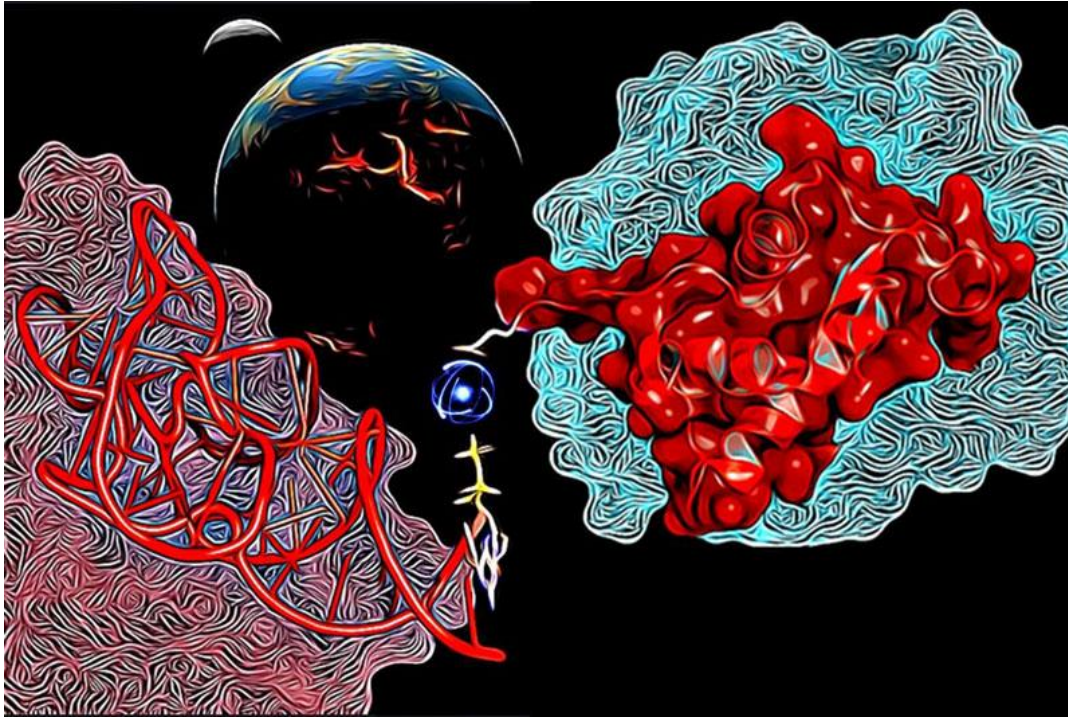


Sarah Krautwurst, FSU Jena

# RNA-protein interaction predictions – An evolutionary approach using co-alignments

38th TBI Winterseminar, Bled, Feb 16, 2023

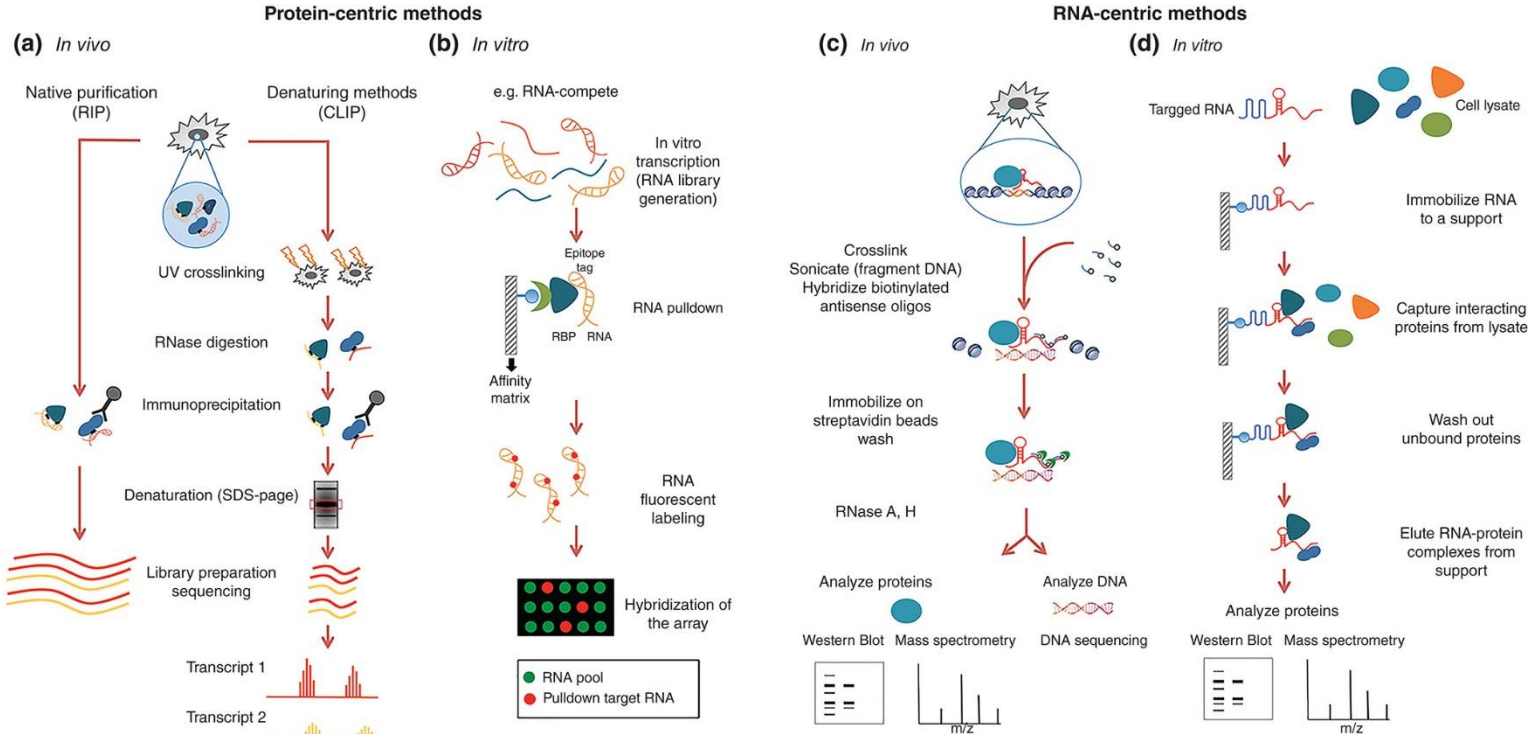
# RNA-protein interactions (RPIs) are ubiquitous



- RPIs are crucial for a variety of molecular processes, intra- and inter-organisms
- already played important roles in origins of life

Giacobelli *et al.* (2022), <https://doi.org/10.1093/molbev/msac032>  
<https://www.hfsp.org/hfsp-news-events/new-perspective-protein-rna-interaction-and-its-role-origins-life>

# Wet lab approaches for RPI detection



Marchese et al. (2016), <https://doi.org/10.1002/wrna.1378>

# Computational approaches for RPI prediction

Muppurala et al. BMC Bioinformatics 2011, 12:489  
<http://www.biomedcentral.com/1471-2105/12/489>



RESEARCH ARTICLE

Open Access

## Predicting RNA-Protein Interactions Using Only Sequence Information

Usha K Muppurala<sup>1,2\*</sup>, Vasant G Honavar<sup>1,3</sup> and Drena Dobbs<sup>1,2</sup>

W72-W79 Nucleic Acids Research, 2021, Vol. 49, Web Server issue  
<https://doi.org/10.1093/nar/gkab393>

Published online 4 June 2021

## catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions

Alexandros Armaos<sup>1,1\*</sup>, Alessio Colantoni<sup>2,1</sup>, Gabriele Proietti<sup>1,3</sup>, Jakob Rupert<sup>1,2</sup> and Gian Gaetano Tartaglia<sup>1,2,4,\*</sup>

<sup>1</sup>Center for Human Technology, Fondazione Istituto Italiano di Tecnologia (IIT), Genoa 16152, Italy, <sup>2</sup>Department of Biology and Biotechnology Charles Darwin, Sapienza University of Rome, Rome 00185, Italy, <sup>3</sup>Dipartimento di Neuroscienze, University of Genova, Genoa 16126, Italy and <sup>4</sup>Center for Life Nano- & Neuro-Science, Fondazione Istituto Italiano di Tecnologia (IIT), Rome 00161, Italy

Received March 08, 2021; Revised April 26, 2021; Editorial Decision April 27, 2021; Accepted April 29, 2021

ISMB 2016

## RCK: accurate and efficient inference of sequence- and structure-based protein-RNA binding models from RNAcompete data

Yaron Orenstein<sup>1</sup>, Yuhao Wang<sup>1</sup> and Bonnie Berger<sup>1,2,\*</sup>

<sup>1</sup>Computer Science and Artificial Intelligence Laboratory and <sup>2</sup>Math Department, MIT, Cambridge, MA, USA

\*To whom correspondence should be addressed.

Bioinformatics, 33(19), 2017, 3636-3642  
doi: 10.1093/bioinformatics/btx350  
Advance Access Publication Date: 31 May 2017  
Original Paper

Structural bioinformatics

## DeepSite: protein-binding site predictor using 3D-convolutional neural networks

J. Jiménez<sup>1</sup>, S. Doerr<sup>1</sup>, G. Martínez-Rosell<sup>1</sup>, A. S. Rose<sup>2</sup> and G. De Fabritiis<sup>1,3,\*</sup>

<sup>1</sup>Computational Biophysics Laboratory (GRIB-MIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), 08003 Barcelona, Spain, <sup>2</sup>San Diego Supercomputer Center, UC San Diego, MC 0505, 9500 Gilman Drive, La Jolla, CA 92093-0505, USA and <sup>3</sup>ICREA, 08010 Barcelona, Spain

## Optimal Protein-RNA Area, OPRA: A propensity-based method to identify RNA-binding sites on proteins

Laura Pérez-Cano and Juan Fernández-Recio<sup>\*</sup>

Department of Life Sciences, Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain

## Protein-RNA interactions: structural characteristics and hotspot amino acids

DENNIS M. KRÜGER<sup>1,2</sup>, SASKIA NEUBACHER<sup>2</sup> and TOM N. GROSSMANN<sup>1,2</sup>

<sup>1</sup>Chemical Genomics Centre of the Max Planck Society, 44227 Dortmund, Germany

<sup>2</sup>Department of Chemistry and Pharmaceutical Sciences, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands

W530-W534 Nucleic Acids Research, 2021, Vol. 49, Web Server issue  
<https://doi.org/10.1093/nar/gkab294>

Published online 5 May 2021

## PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA

Melissa F. Adasme<sup>1,1</sup>, Katja L. Linnemann<sup>1,1</sup>, Sarah Naomi Bolz<sup>1,1</sup>, Florian Kaiser<sup>2</sup>, Sebastian Salentin<sup>1</sup>, V. Joachim Haupt<sup>2</sup> and Michael Schroeder<sup>1,1,\*</sup>

<sup>1</sup>Biotechnology Center (BIOTEC), CMCB, Technische Universität Dresden, Tatzberg 47-49, 01307 Dresden, Germany and <sup>2</sup>PharmaI GmbH, 01307 Dresden, Germany

Received January 29, 2021; Revised March 24, 2021; Editorial Decision April 08, 2021; Accepted April 13, 2021

INFORMATICS ORIGINAL PAPER

Vol. 29 no. 20 2013, pages 2588-2595  
doi:10.1093/bioinformatics/btt447

Structural bioinformatics

Advance Access publication August 23, 2013

## Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment

Jianyi Yang<sup>1</sup>, Ambrish Roy<sup>1</sup> and Yang Zhang<sup>1,2,\*</sup>

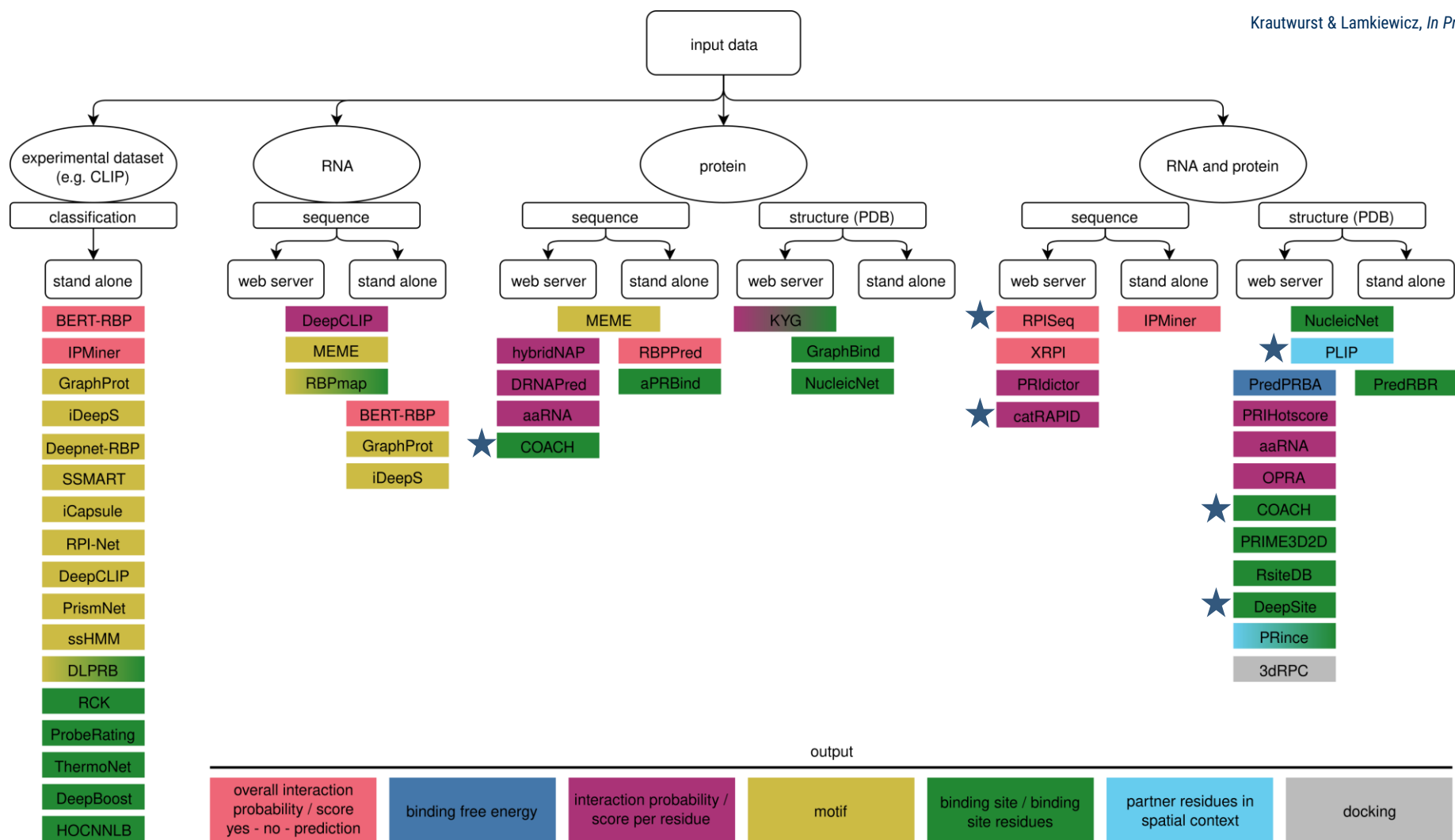
<sup>1</sup>Department of Computational Medicine and Bioinformatics and <sup>2</sup>Department of Biological Chemistry, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218, USA

Associate Editor: Anna Tramontano

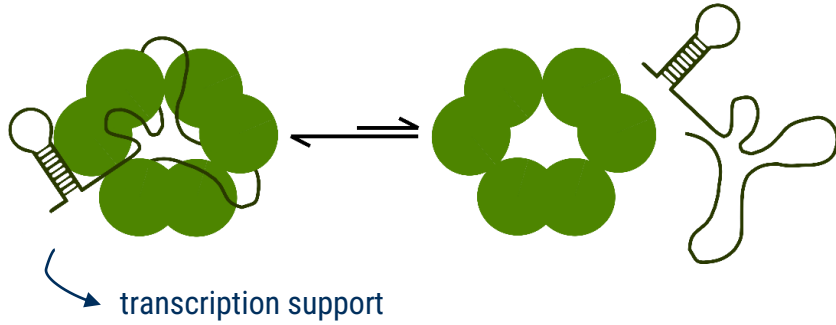
## Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains

Jian Zhang, Zhiqiang Ma and Lukasz Kurgan

Corresponding author: Lukasz Kurgan, Department of Computer Science, Virginia Commonwealth University, Richmond 23284, USA. Tel.: +1-804-827-3986; E-mail: lkurgan@vcu.edu

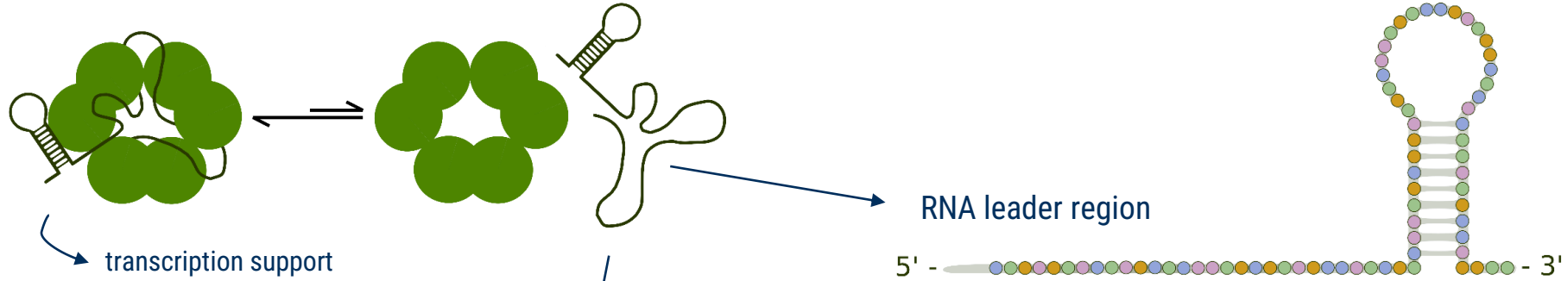


# Ebolavirus VP30 interacts with viral RNA leader region



John *et al.* (2007), <https://doi.org/10.1128/JVI.02523-06>  
Biedenkopf *et al.* (2016), <https://doi.org/10.1128/JVI.00271-16>  
Schlereth *et al.* (2017), <https://doi.org/10.1080/15476286.2016.1194160>

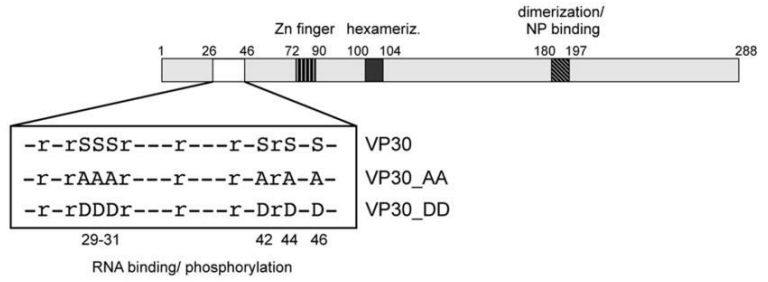
# Ebolavirus VP30 interacts with viral RNA leader region



1            10            20            30            40            50            60  
 MEASYERGRPRAARQHSRDGHDHHV**RARSSSR**ENYRGEYRQSRSSASQVRVPTVFHKKRVE

61            70            80            90            100            110            120  
 PLTVPPAPKDI**C**PTLKKGF**L**C**DSSFC**KKD**H**QLES**L**TDRE**L**LL**L**IAR**K**TCGS**V**EQ**Q**LN**I**T**A**

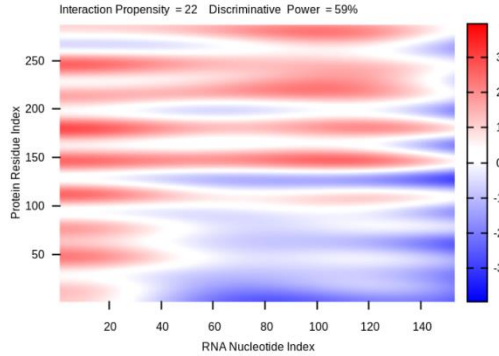
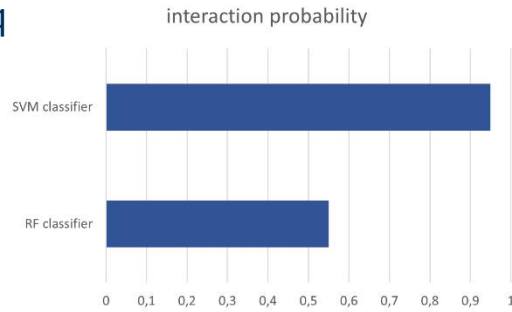
⇒



John et al. (2007), <https://doi.org/10.1128/JVI.02523-06>  
 Biedenkopf et al. (2016), <https://doi.org/10.1128/JVI.00271-16>  
 Schlereth et al. (2017), <https://doi.org/10.1080/15476286.2016.1194160>

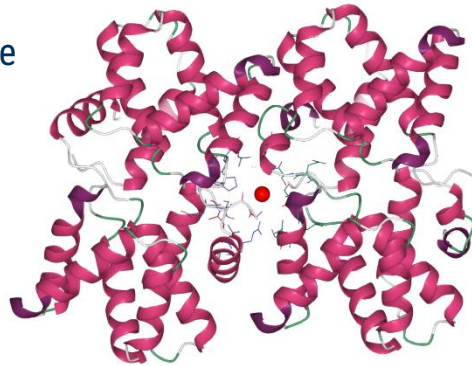
# Performance of selected tools on VP30 + viral RNA

RPISeq

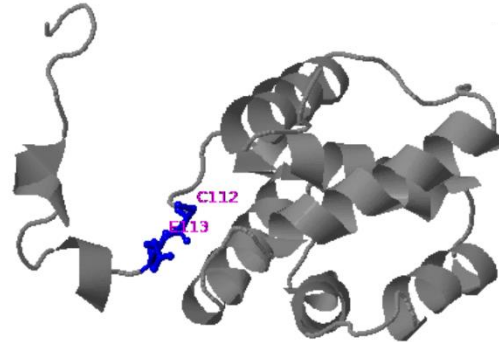


catRAPID

DeepSite

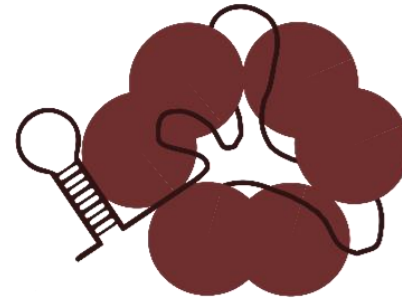
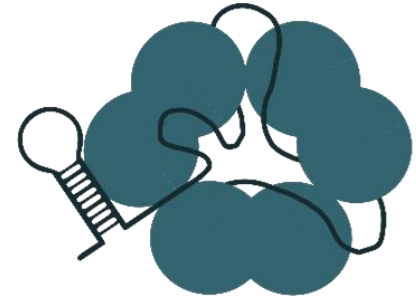
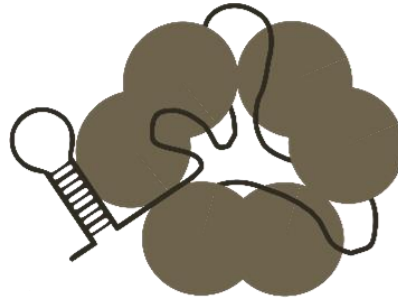


COACH

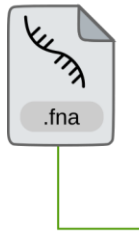




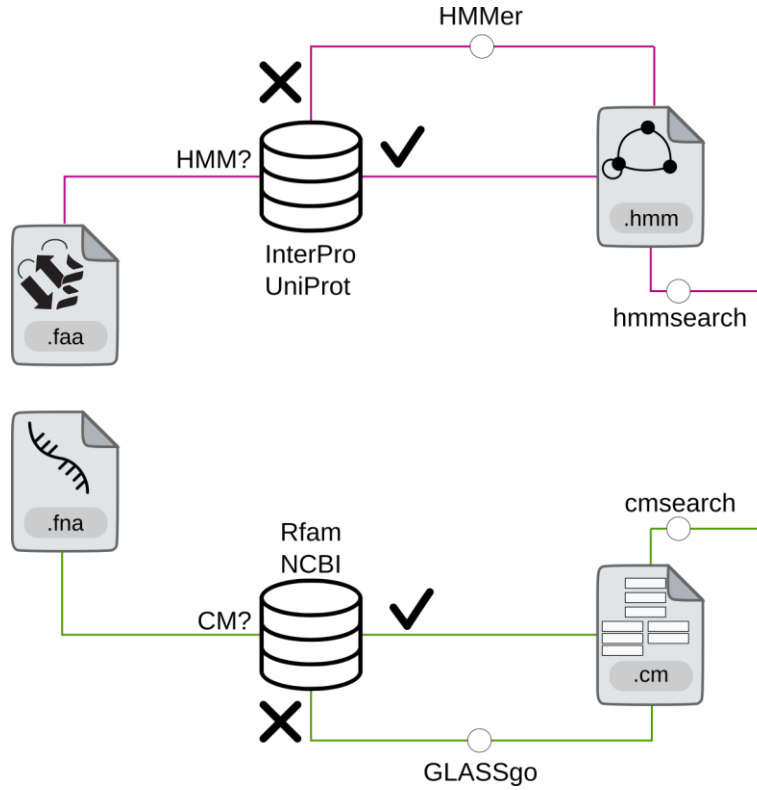
# Using homologs for RPI prediction?



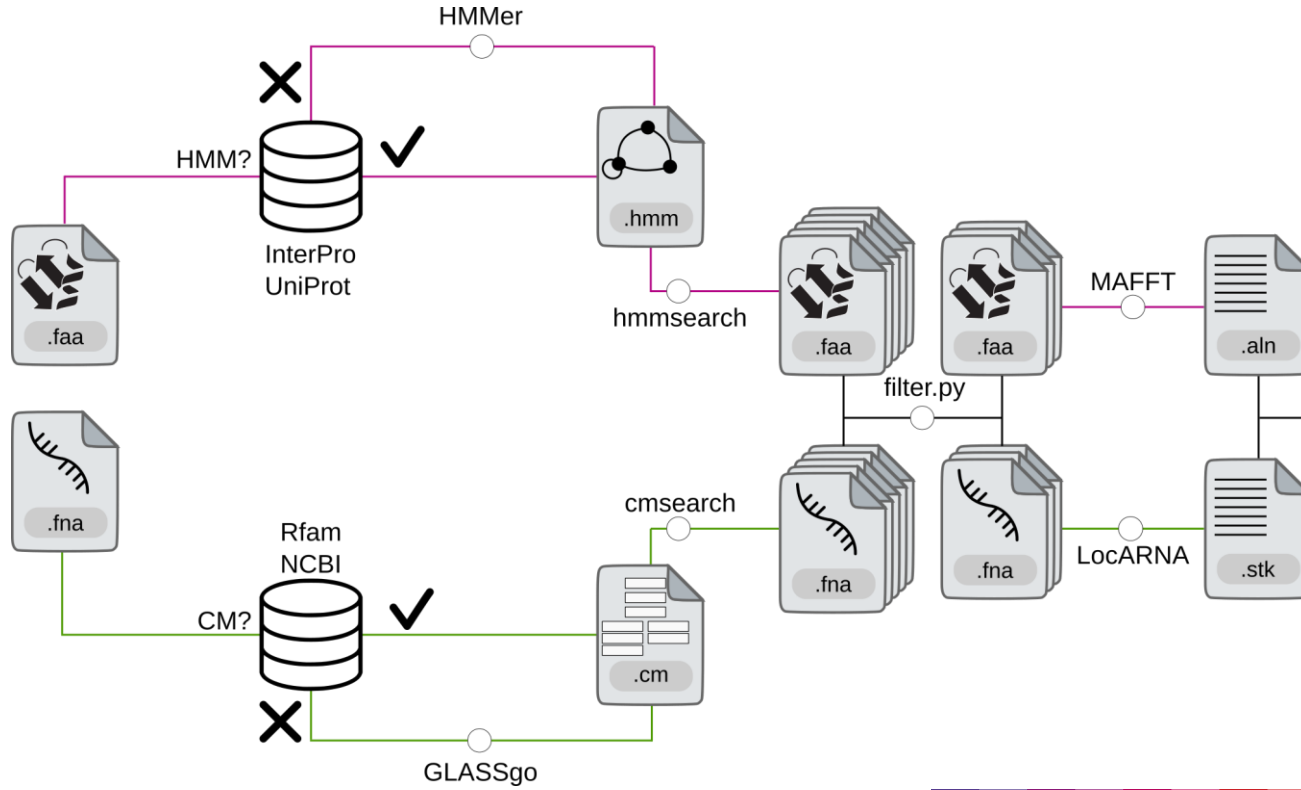
# Pipeline: Infer RPI-involved residues from co-evolution



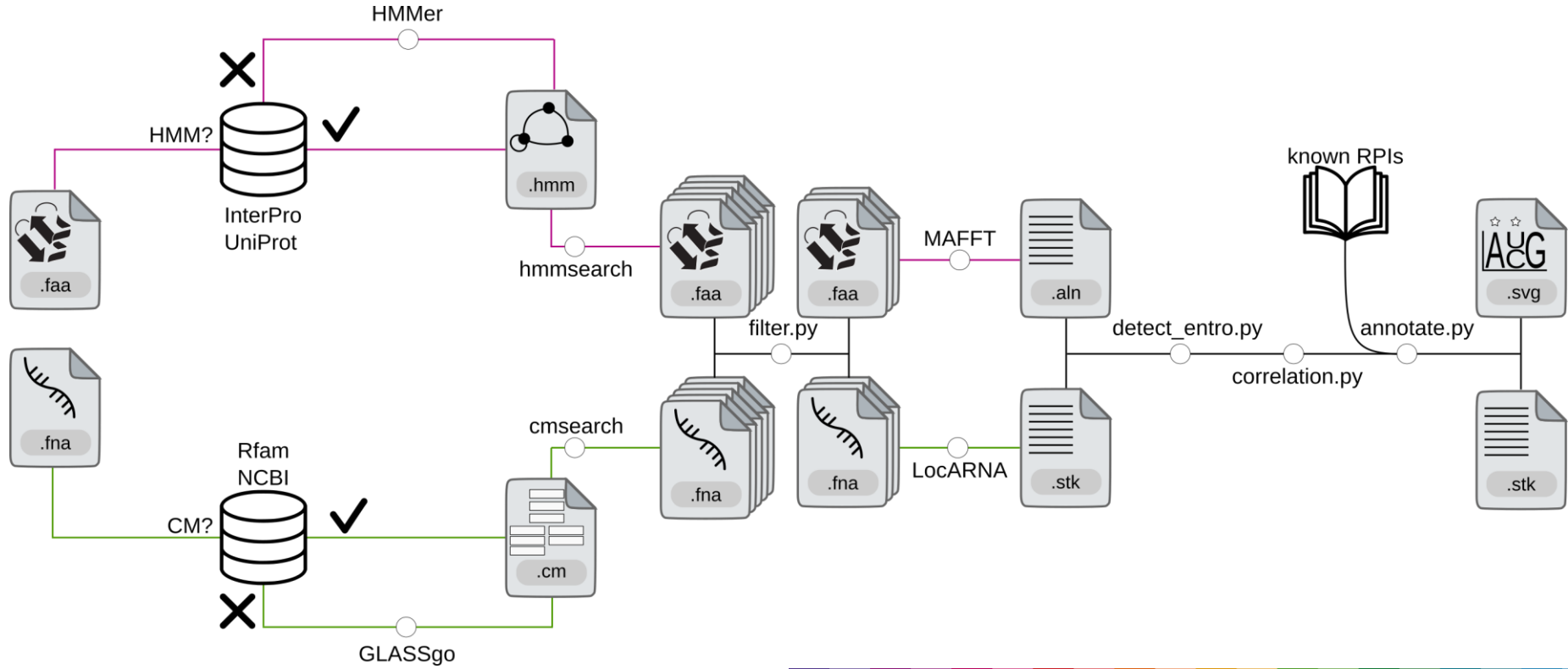
# Pipeline: Infer RPI-involved residues from co-evolution



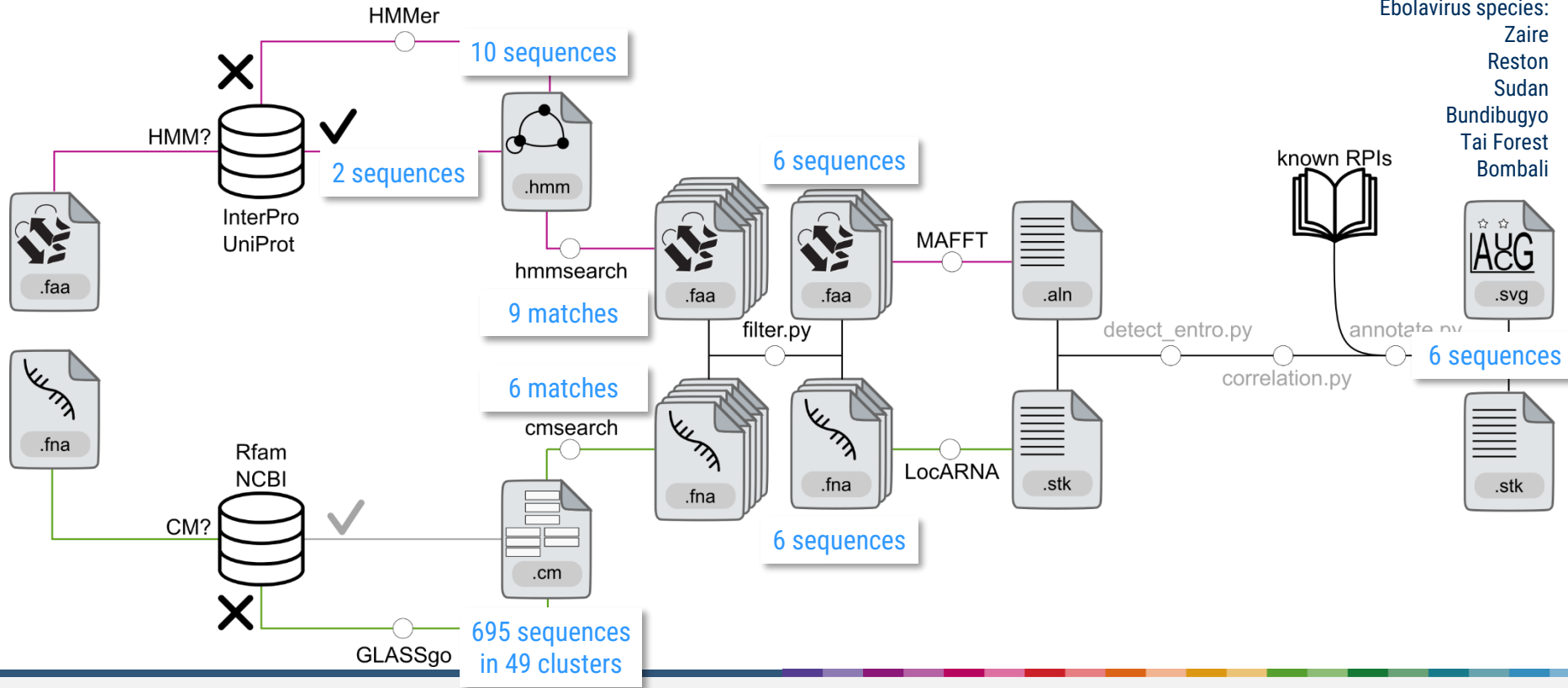
# Pipeline: Infer RPI-involved residues from co-evolution



# Pipeline: Infer RPI-involved residues from co-evolution

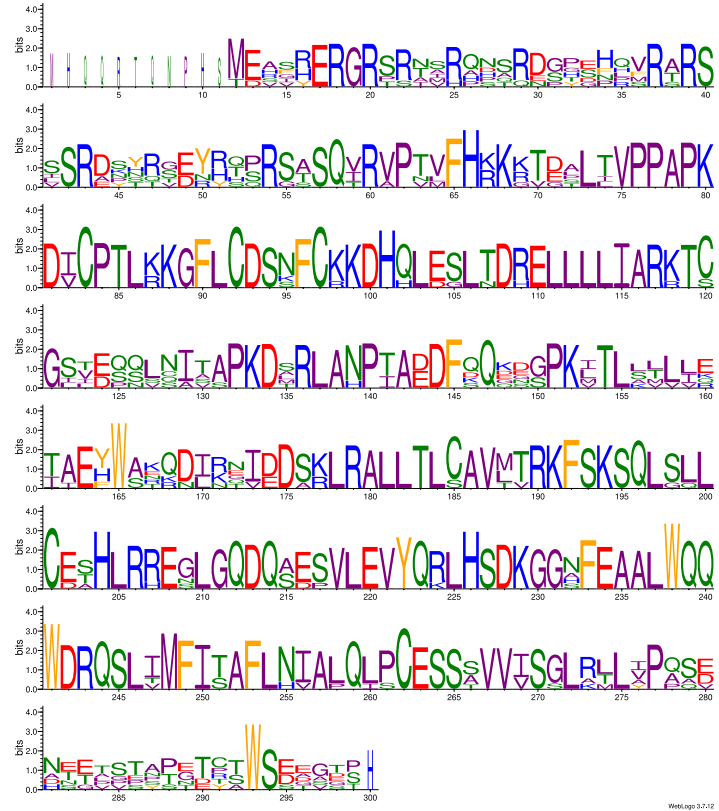
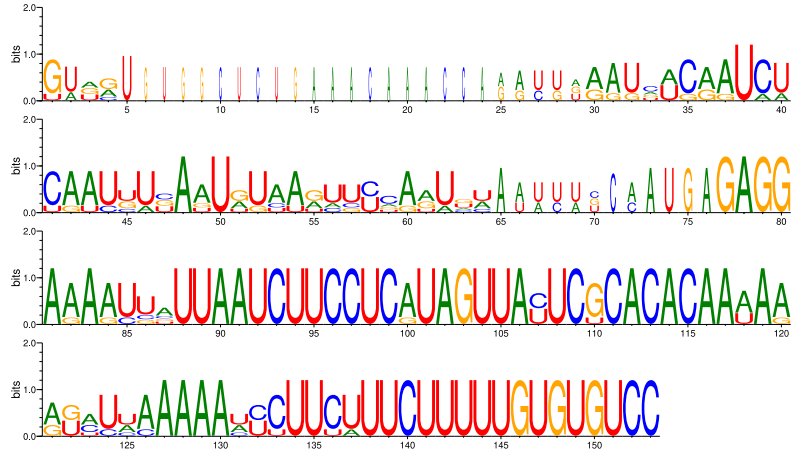


# Pipeline: Infer RPI-involved residues from co-evolution



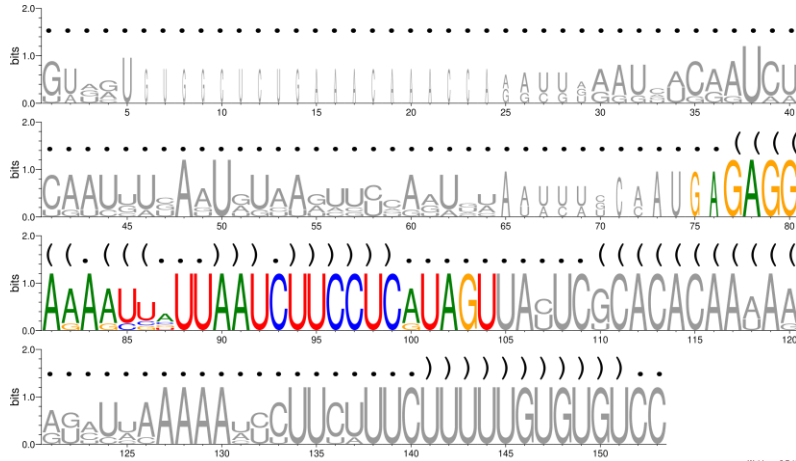
# VP30 + leader RNA co-alignments

6 Ebolavirus species:  
 Zaire                      Reston  
 Sudan                      Bundibugyo  
 Tai Forest                Bombali



# VP30 + leader RNA co-alignments

6 Ebolavirus species:  
 Zaire                      Reston  
 Sudan                      Bundibugyo  
 Tai Forest                Bombali



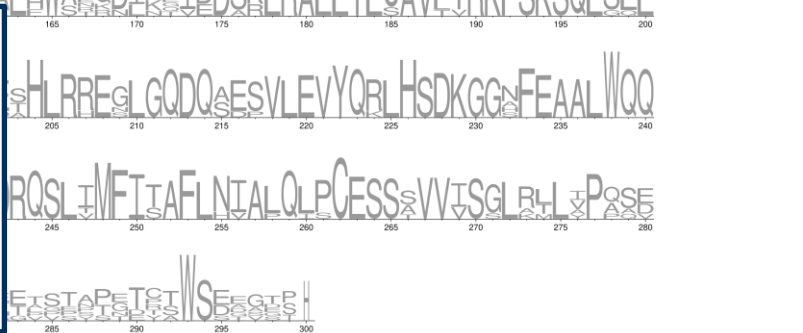
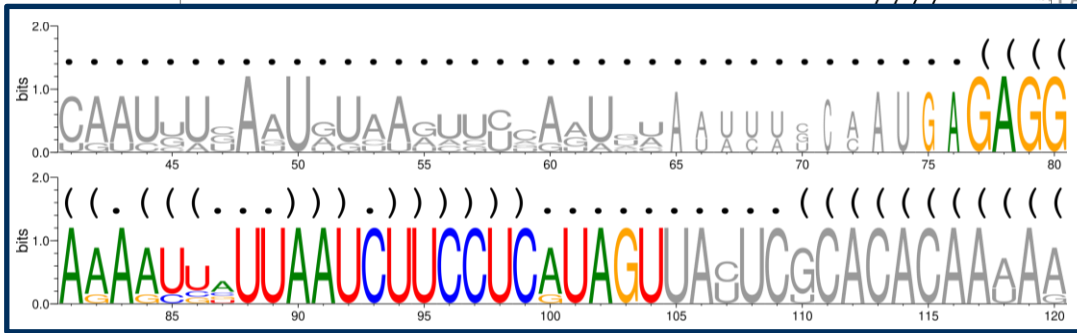
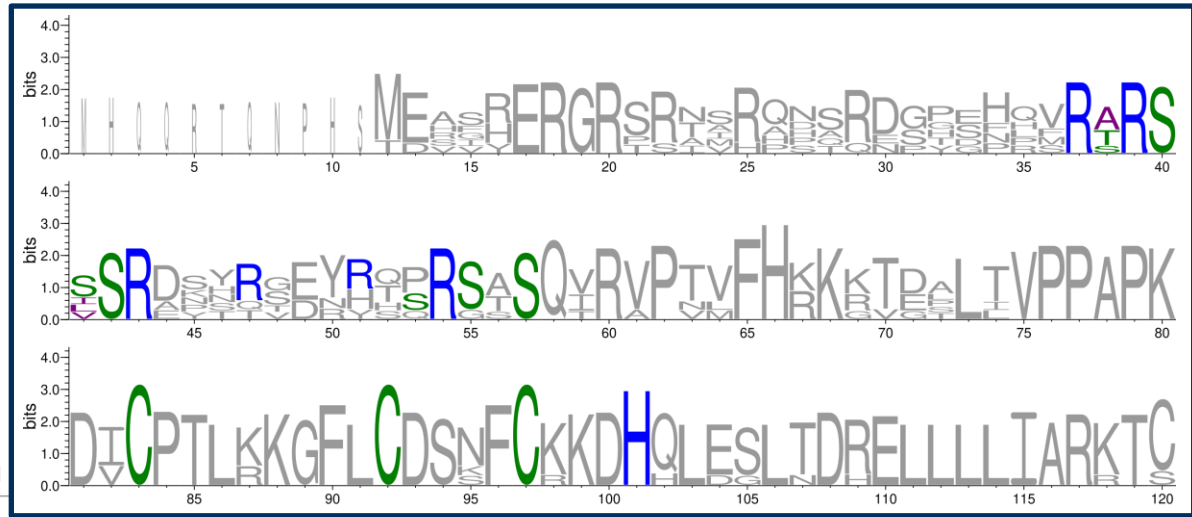
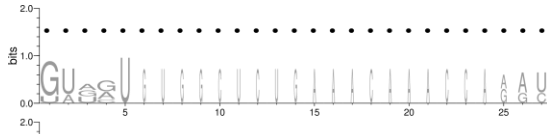
WebLogo 3.7.12



WebLogo 3.7.12

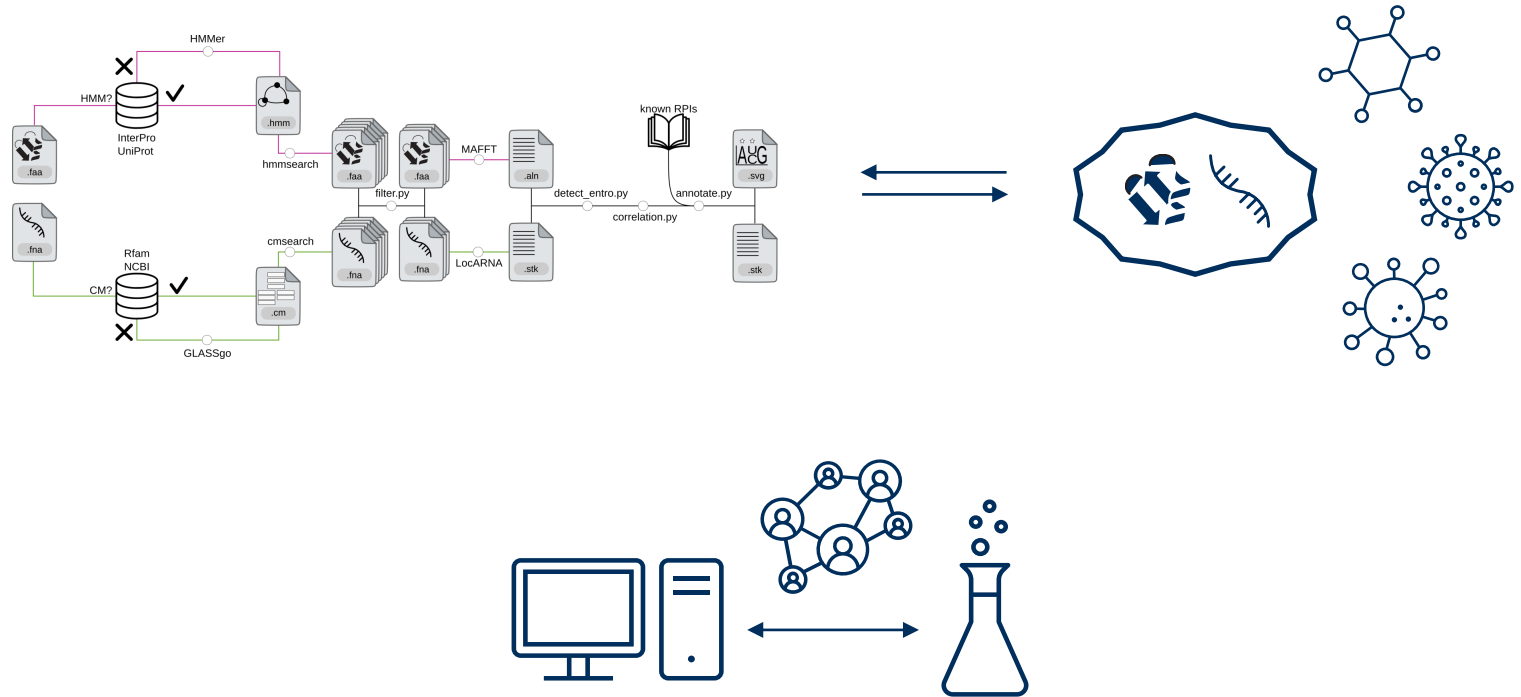


# VP30 + leader RNA co-alignments

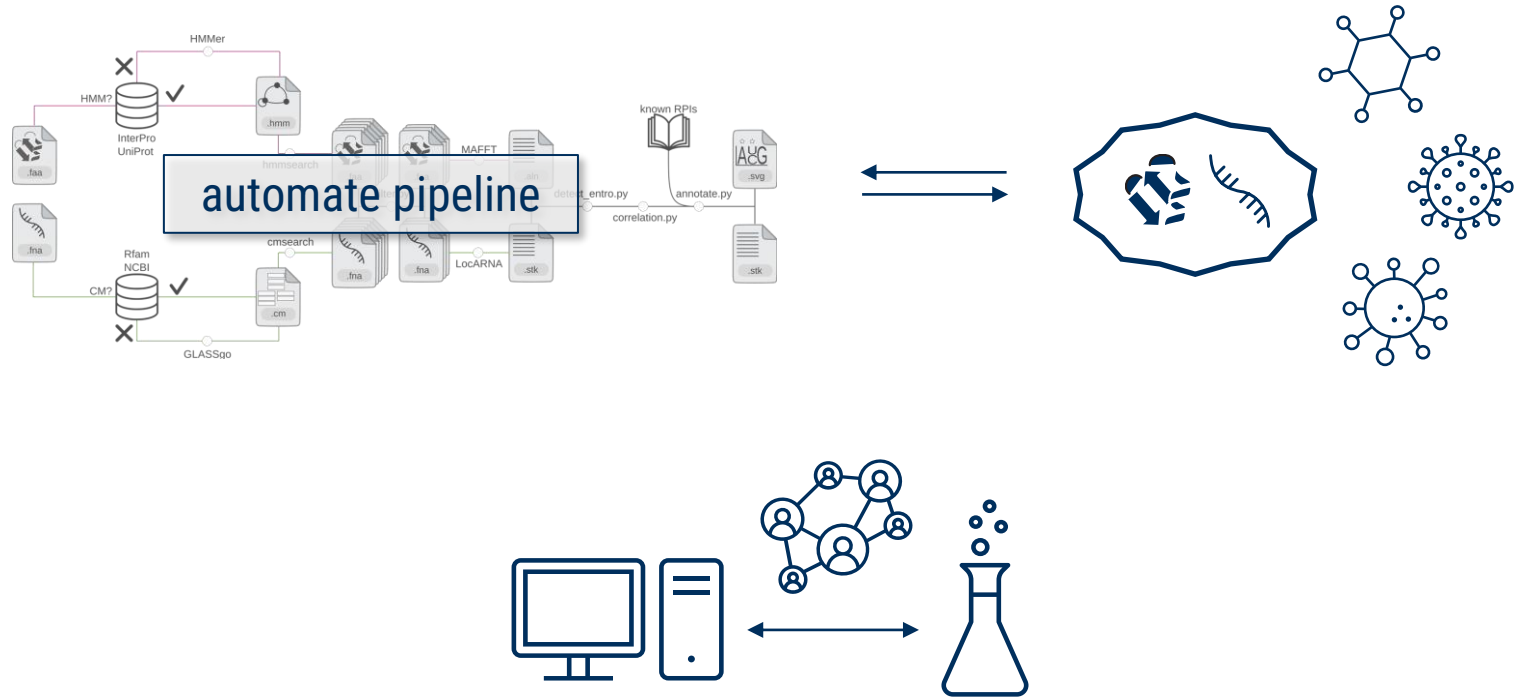


WebLogo 3.7.12

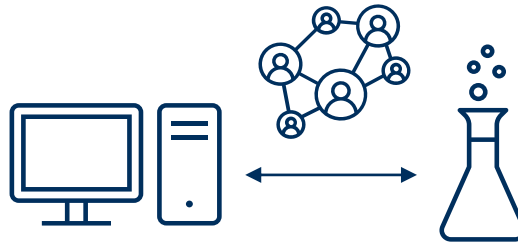
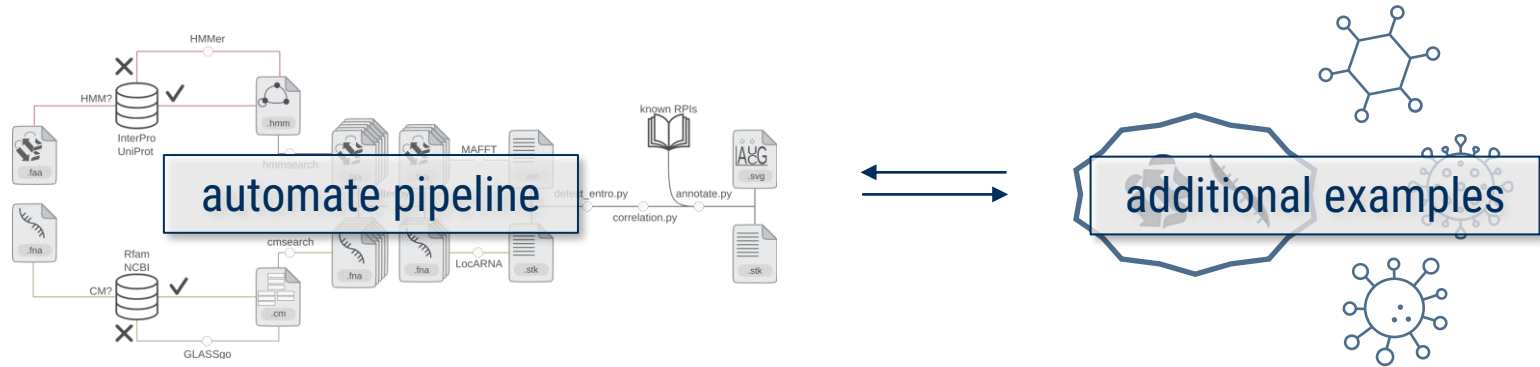
# Further directions



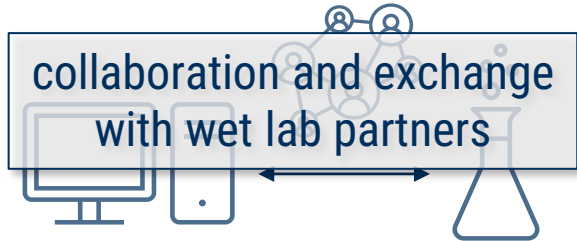
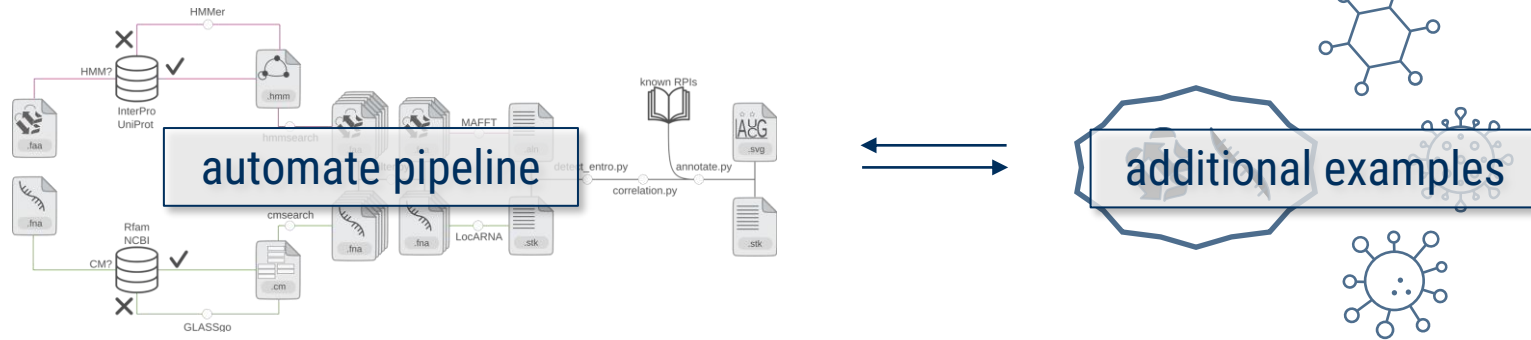
# Further directions



# Further directions



# Further directions





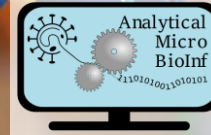
RNA  
BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS

FRIEDRICH-SCHILLER-  
UNIVERSITÄT  
JENA

NFDI4  
MICROBIOTA



BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS



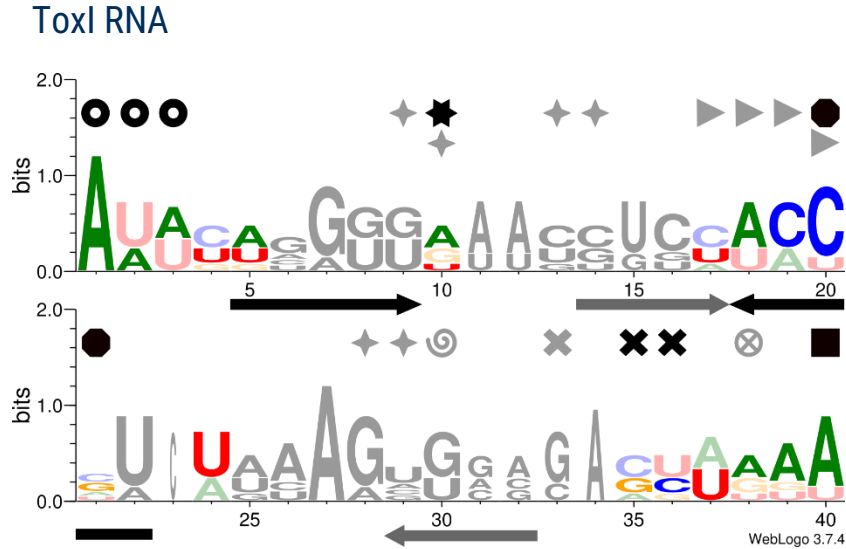
**contact:**

[sarah.krautwurst@uni-jena.de](mailto:sarah.krautwurst@uni-jena.de)

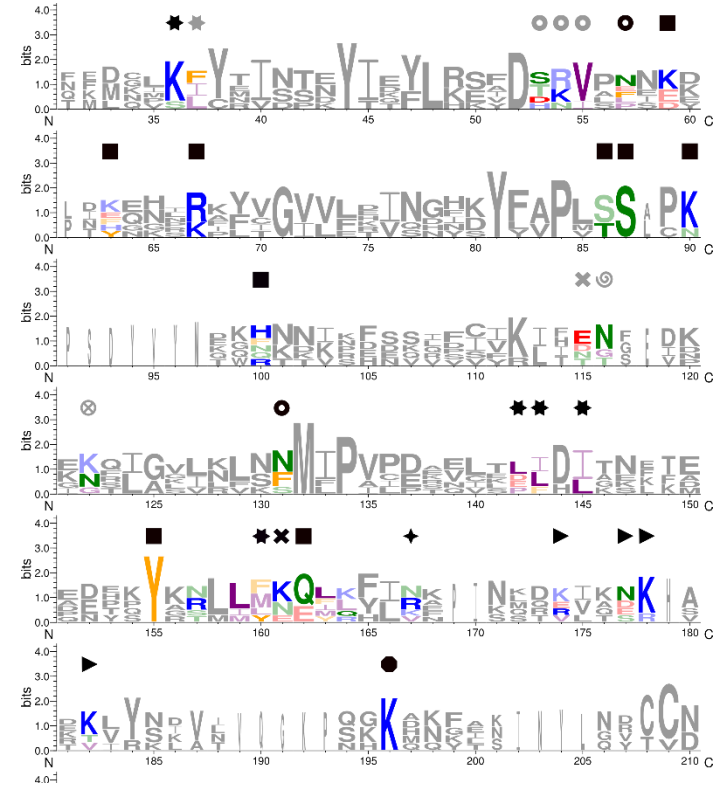
[kevin.lamkiewicz@uni-jena.de](mailto:kevin.lamkiewicz@uni-jena.de)



# Co-alignments with interaction annotation for ToxIN RPI



ToxN protein



# Using co-evolution for RPI prediction

predict sequence preferences for poorly studied RBPs


→ infer preference information from its homologous RBPs using PWMs as sequence preference representations + KNN algorithm

→ binding data not needed


Yang et al. *BMC Bioinformatics* (2018) 19:96  
<https://doi.org/10.1186/s12859-018-2091-8>

BMC Bioinformatics

RESEARCH ARTICLE Open Access

 CrossMark

## Inferring RNA sequence preferences for poorly studied RNA-binding proteins based on co-evolution

Shu Yang<sup>1\*</sup> , Junwen Wang<sup>2</sup> and Raymond T. Ng<sup>1</sup>

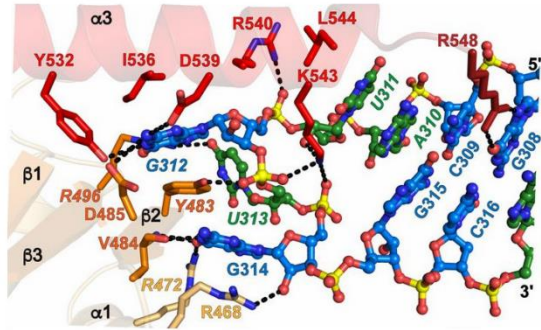
**Abstract**  
**Background:** Characterizing the binding preference of RNA-binding proteins (RBP) is essential for us to understand the interaction between an RBP and its RNA target, and to decipher the mechanisms of post-transcriptional regulation.

Yang et al. (2018), <https://doi.org/10.1186/s12859-018-2091-8>

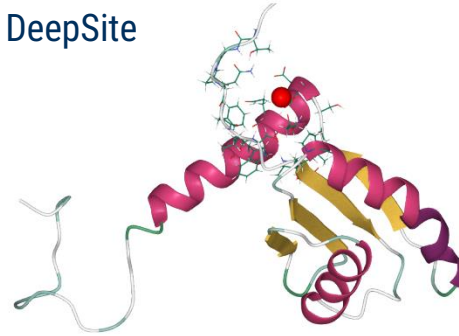


# Performance of selected tools on 7SK RNA + LARP7 protein

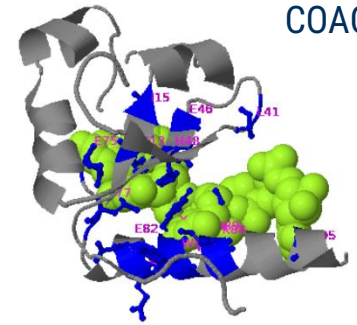
literature  
reference



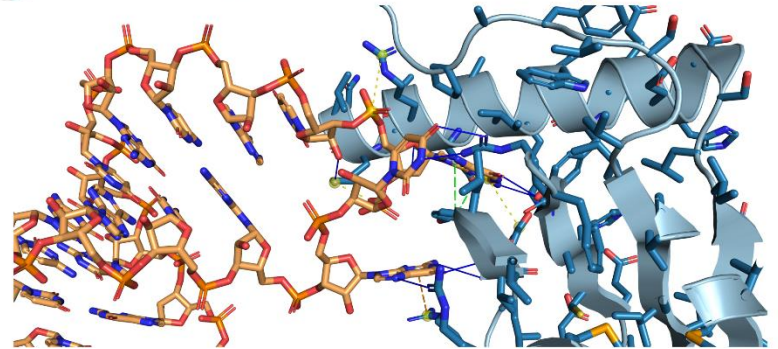
DeepSite



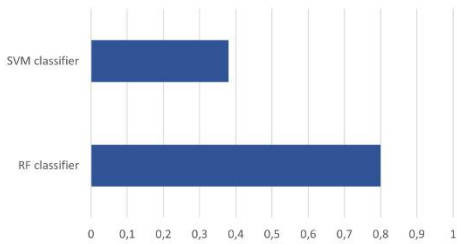
COACH



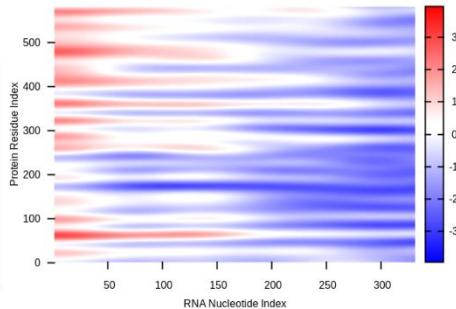
PLIP



interaction probability



Interaction Propensity = 84 Discriminative Power = 98%



RPISeq

catRAPID

Eichhorn et al. (2018), <https://doi.org/10.1073/pnas.1806276115>

# Ebolavirus VP30 interacts with viral RNA leader region

American Society for Microbiology  
Journal of Virology  
Volume 81, Issue 17, 1 September 2007, Pages 8967-8976  
<https://doi.org/10.1128/JVI.02523-06>

Genome Replication and Regulation of Viral Gene Expression

## Ebola Virus VP30 Is an RNA Binding Protein

Sinu P. John<sup>1</sup>, Tan Wang<sup>2</sup>, Scott Steffen<sup>3,†</sup>, Sonia Longhi<sup>4</sup>, Connie S. Schmaljohn<sup>3</sup>, and Colleen B. Jonsson<sup>2,\*</sup>

<sup>1</sup>Graduate Program in Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294

<sup>2</sup>Department of Biochemistry and Molecular Biology, Southern Research Institute, Birmingham, Alabama 35205

<sup>3</sup>U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland 21702

<sup>4</sup>Architecture et Fonction des Macromolécules Biologiques, UMR 6098 and Universités Aix-Marseille I et II, Campus de Luminy, 13288 Marseille Cedex 09, France

### ABSTRACT

The Ebola virus (EBOV) genome encodes for several proteins that are necessary and sufficient for replication and transcription of the viral genome *in vitro*; NP, VP30, VP35, and L. VP30 acts in *trans* with an RNA secondary structure upstream of the first transcriptional start site to modulate transcription. Using a bioinformatics approach, we identified a region within the N terminus of VP30 with sequence features that typify intrinsically disordered regions and a potential RNA binding site. To experimentally assess the ability of VP30 to directly interact with the viral RNA, we purified recombinant EBOV VP30 to >90% homogeneity and assessed RNA binding by UV cross-linking and filter-binding assays.

American Society for Microbiology  
Journal of Virology  
Volume 90, Issue 16, 15 August 2016, Pages 7481-7496  
<https://doi.org/10.1128/JVI.00271-16>

Genome Replication and Regulation of Viral Gene Expression

## RNA Binding of Ebola Virus VP30 Is Essential for Activating Viral Transcription

Nadine Biedenkopf<sup>b</sup>, Julia Schlereth<sup>a</sup>, Arnold Grünweller<sup>a</sup>, Stephan Becker<sup>b</sup>, and Roland K. Hartmann<sup>a</sup>

<sup>a</sup>Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Germany

<sup>b</sup>Institut für Virologie, Philipps-Universität Marburg, Marburg, Germany

### ABSTRACT

The template for Ebola virus (EBOV) transcription and replication is a viral RNA nucleocapsid composed of the viral negative-sense (–) RNA genome, which is complexed by the nucleoprotein (NP), VP35, polymerase L, and VP30. While viral replication is exerted by polymerase L and its cofactor VP30, mRNA synthesis is regulated by the viral nucleocapsid protein VP30. VP30 is an EBOV-specific transcription factor. VP30 is a homohexamer containing a nonconventional zinc finger. The transcription of the viral RNA genome by VP30 is strongly influenced by its phosphorylation state. RNA binding contributed to VP30's function in transcription. We studied the RNA binding of VP30 using a novel mobility shift assay and the 3-terminal 154 nucleotides of the viral RNA genome.

RNA BIOLOGY  
2016, VOL. 13, NO. 9, 783–798  
<http://dx.doi.org/10.1080/15476286.2016.1194160>

RESEARCH PAPER

## RNA binding specificity of Ebola virus transcription factor VP30

Julia Schlereth<sup>a</sup>, Arnold Grünweller<sup>a</sup>, Nadine Biedenkopf<sup>b</sup>, Stephan Becker<sup>b</sup>, and Roland K. Hartmann<sup>a</sup>

<sup>a</sup>Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marburg, Germany; <sup>b</sup>Institut für Virologie, Philipps-Universität Marburg, Marburg, Germany

### ABSTRACT

The transcription factor VP30 of the non-segmented RNA negative strand Ebola virus balances viral transcription and replication. Here, we comprehensively studied RNA binding by VP30. Using a novel VP30:RNA electrophoretic mobility shift assay, we tested truncated variants of 2 potential natural RNA substrates of VP30 - the genomic Ebola viral 3'-leader region and its complementary antigenomic counterpart (each ~155 nt in length) - and a series of other non-viral RNAs. Based on oligonucleotide interference, the major VP30 binding region on the genomic 3'-leader substrate was assigned to the internal expanded single-stranded region (~ nt 125–80). Best binding to VP30 was obtained with ssRNAs of optimally ~ 40 nt and mixed base composition; underrepresentation of purines or pyrimidines was tolerated, but homopolymeric sequences impaired binding. A stem-loop structure, particularly at the 3'-