

Challenges in RNA Regulation in Huntington's Disease: Insights from Computational Studies

Oscar Palomino-Hernandez,^[a, b, c, d] Michael A. Margreiter,^[a, b] and Giulia Rossetti^{*,[a, e, f]}

Abstract: Novel therapeutic approaches are being developed to tackle neurodegenerative diseases, due to the lack of efficiency of the known druggable targets. For Huntington's disease, a promising approach is the regulation of the RNA product. This target would allow for a selective and effective inhibition of the toxic effects exerted by the final nucleic product and the coded protein. In this review, the current state of the art of RNA regulation is discussed, with a brief

but insightful view on novel plausible targets. After this, an emphasis on successful computational and experimental approaches tailored in modeling and regulating RNA aberrant behavior are extensively presented. Finally, the application and limitations of current computational methods are discussed, and possible avenues for improvement are outlined.

Keywords: Huntington's disease • RNA • computational approaches • molecular dynamics

1. Introduction

Huntington's disease (HD) was first recognized as a distinct disease in 1872.^[1] A century later, the identification of a marker on chromosome 3 led to the subsequent elucidation of an unstable CAG trinucleotide expansion located in exon-1 of the huntingtin gene (*HTT*) as the single cause of the disease.^[2] These mutant mRNA transcripts are then translated into a prolonged glutamine tract (polyQ) in the mutant huntingtin protein (HTT). In turn, the polyQ tract is responsible for an increased protein aggregation propensity of mutant HTT,^[3] which can interconvert its monomeric, oligomeric, and fibrillar forms *in vitro* and *in vivo*.^[4]

In addition to the HTT protein-dependent pathogenicity, a toxic gain of function of the mutant *HTT* mRNA has also been observed.^[5,6] This is not unique to HD. During the last decade, it has become increasingly evident that abnormalities in the functionality of protein-coding and non-coding RNAs, as well as in RNA-binding proteins function, represent a common feature among many neurodegenerative diseases.^[7,8] Recent progress in next-generation sequencing has enabled researchers to study the most abundant transcripts, alternative splicing processes, and non-coding RNA molecules of nervous tissue,^[9] thus furthering our knowledge of RNA pathways.

In this review, we will describe the aberrant RNA behavior in HD and highlight the recent computational advances towards modeling in RNA regulation in HD.

2. RNA Behavior in HD

The concept of "RNA toxicity" denotes the direct ability of the mutant transcript to induce pathogenesis. In particular, this effect has been described as the result of aberrant interactions between RNA and other proteins. The RNA can also induce

toxicity indirectly by being translated into toxic proteins.^[10] This is the case in HD, where RNA toxic effects coexist with the ones at the protein level, thus rendering them more challenging to assess and to understand.

Before we delve into the many ways RNA contributes to progressive neurodegenerative illnesses, we give a brief overview of RNAs and focus on those typically implicated in HD.

[a] O. Palomino-Hernandez, M. A. Margreiter, Jun.-Prof. Dr. Dr. G. Rossetti

Computational Biomedicine, Institute of Neuroscience and Medicine (INM-9)/Institute for advanced simulations (IAS-5), Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: g.rossetti@fz-juelich.de

[b] O. Palomino-Hernandez, M. A. Margreiter
Faculty I, RWTH Aachen, 52425 Aachen, Germany

[c] O. Palomino-Hernandez
Computation-based Science and Technology Research Center, The Cyprus Institute, Nicosia 2121, Cyprus

[d] O. Palomino-Hernandez
Institute of Life Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

[e] Jun.-Prof. Dr. Dr. G. Rossetti
Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, 52425 Jülich, Germany

[f] Jun.-Prof. Dr. Dr. G. Rossetti
Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation University Hospital Aachen, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

2.1 The Versatile Functional Roles of RNA

The human transcriptome consists of coding messenger RNA (mRNA) and multiple classes of non-coding RNAs (ncRNA), including ribosomal, transfer and a myriad of other different types.^[11] While about 90% of the human genome is transcribed to RNA, only approximately 1.2% is protein-coding and thus suggests that ncRNA has functions in the regulation of physiological processes in all cell types.^[12]

Transcription of mRNA occurs initially as pre-mRNA (Figure 1). Mature mRNA, ready for translation, is formed by splicing pre-mRNA and hereby removing intronic sequences.^[13] Alternative splicing patterns result in different mature mRNA originating from the same pre-mRNA. To export mature mRNA from the nucleus into the cytoplasm, the ribonucleoprotein particle (mRNP) is formed. In the cytoplasm, mRNP associates with the ribosomal machinery and facilitates together with the transfer RNA (tRNA) the translation into the corresponding protein. During transcription, processing, transport, and eventual degradation, mRNA is handled not only by a host of RNA binding proteins but also by several ncRNAs.

Besides defective mRNA, mutations in ribosomal RNA (rRNA) and tRNA play roles in mammalian brain development, neurological syndromes, and neurodegeneration,^[14] which highlights the importance of RNA in biological processes.

2.2 MicroRNA Dysregulation in HD

MicroRNAs (miRNAs) are a class of small ncRNA (typically 20–30 nucleotides long) and have received renewed interest recently in the context of aging and neurodegeneration.^[9] They regulate gene expression post-transcriptionally, e.g. reducing

fluctuations in protein expression^[15] and are crucial for neuronal function and disease.^[16] For a description in depth of biogenesis of miRNAs, the reader is referred to other works.^[17–20]

HD appears to affect both miRNA biogenesis and expression, since: 1. Reduced levels of ribonucleases Droscha and Dicer were noted in mouse models of HD.^[21] 2. *Post-mortem* cortex samples of HD patients showed impairment of specific miRNAs, such as upregulation of miR-29a and miR-330 and downregulation of miR-132, affecting the RNA-induced silencing complex (RISC) activity.^[22,23] Some works indicate that mutant HTT is also involved in the regulation of RNA processes.^[24]

All these aspects can be further affected by the presence of oxidative stress in HD,^[25] which is tightly intertwined with miRNA machinery and can further exacerbate disease progression.^[26,27]

2.3 mRNA CAG Expansion as a Primary Driver in HD

Several evidences demonstrate that the RNA toxicity in HD is triggered by expanded CAG repeats in mRNA and not by other trinucleotides repeats encoding also for glutamine (i.e. CAA). Experiments in *Drosophila* showed that CAG repeats interrupted with CAA clearly manifest less pronounced toxic effects.^[28] A similar result was found in human neuronal cells.^[29,30] Moreover, while both CAG and CAA encode for glutamine, the secondary RNA structures linked to each of them differ. Whereas expanded CAG repeats produce a hairpin-like structure, CAA or CAG stretches interrupted with CAA do not fold into a hairpin.^[31] Moreover, CAG hairpin structures were shown to bind to proteins in a length-dependent manner,^[32] therefore contributing to HD pathogenicity.

Indeed a feature of HD is the presence of nuclear RNA



Oscar Palomino-Hernandez obtained a master's degree from the National Autonomous University of Mexico, and is currently an early stage researcher under the STIMULATE Marie Skłodowska-Curie Innovative Training Network. His research interests include cheminformatics, machine learning and molecular modeling applied to drug discovery. His research focus is exploring the druggability of highly flexible regions in nucleic acids and intrinsically disordered proteins.



Michael A. Margreiter obtained a master's degree from the University of Innsbruck, Austria, and is currently a Ph.D. candidate at RWTH University Aachen, Germany. His research focuses on the structural and dynamical characterization of alternative targets for polyglutamine disorders, in particular RNA processing enzymes, and the development of novel ligands towards treatment options for these diseases.



Jun.-Prof. Dr. Dr. Giulia Rossetti is currently a professor at the medical faculty of RWTH University in Aachen, Germany and group leader at the Institute of Neuroscience and Medicine (INM-9) and Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich, in Germany. She is deeply involved in several EU-wide projects, such as the Human Brain Project, Helmholtz Information & Data Science School, STIMULATE-ITN, among others. Her research is mostly focused on neuropathologies such as neurodegenerative diseases, neuropathic pain disorders and Schizophrenia. Her group mostly uses HPC-based molecular simulation, ranging from coarse-grain to classical up to ab-initio representation, as well as free-energy calculations, bioinformatics and cheminformatics.

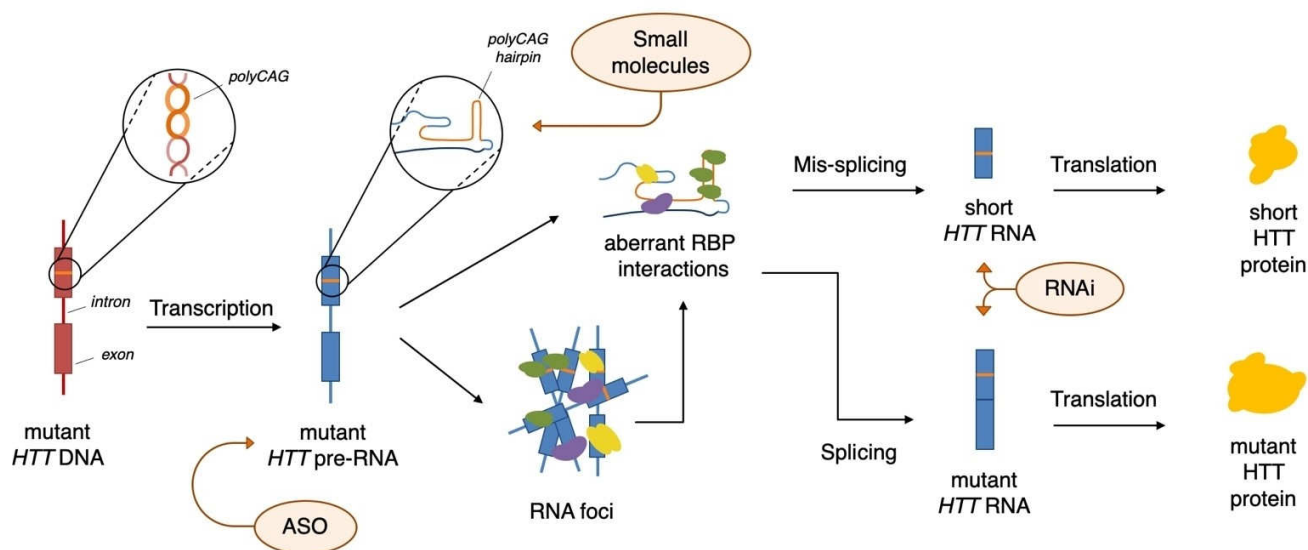


Figure 1. Scheme for the Huntingtin protein pathway from the mutant DNA. Red boxes show successful approaches towards regulating RNA, and the corresponding RNA target.

inclusions (nuclear foci) loaded with several proteins having affinity for the CAG mRNA hairpin.^[33] The number of foci per nucleus correlates with CAG repeat expansion.^[34] Presumably, RNA repeat inclusions are trapped in the nucleus as a consequence of RNA overloading with proteins. This, in turn, decreases the amount of available RNA-binding proteins while increasing the likelihood of alterations in splicing and expression of other mRNA. In this way, the expanded CAG repeats have effects on the subcellular localization of the mutant transcript, as well as in the binding dynamics of specific proteins.

2.4 RNA Binding Proteins Show Aberrant Interactions with Mutant HTT mRNA

Several evidences demonstrate that the RNA toxicity in HD is triggered by expanded CAG repeats in mRNA and not by other trinucleotides repeats encoding also for glutamine (i.e. CAA). Experiments in *Drosophila* showed that CAG repeats interrupted with CAA clearly manifest less pronounced toxic effects.^[28] A similar result was found in human neuronal cells.^[29,30] Moreover, while both CAG and CAA encode for glutamine, the secondary RNA structures linked to each of them differ. Whereas expanded CAG repeats produce a hairpin-like structure, CAA tracts or CAG stretches interrupted with CAA do not fold into a hairpin.^[31] Moreover, CAG hairpin structures were shown to bind to proteins in a length-dependent manner,^[32] therefore contributing to HD pathogenicity.

In *HTT* mRNA with multiple CAG repeats, the formation of the hairpin-like structure leads to abnormal protein interactions, impeding their normal functions. Researchers

probed the mutant *HTT* RNA interactome using mass spectrometry and found that dysregulated splicing is pivotal for RNA-induced neurotoxicity in HD. Accordingly, they found that the majority of proteins captured by mutant *HTT* RNA are part of the spliceosome pathway.^[35]

In this section, we discuss a few key RNA-protein interactions relevant to HD recently described in the literature.

1. MID1, when acting as an E3 ubiquitin ligase, catalyzes the degradation of target proteins such as PP2A, and also allows the activity of mTOR, a PP2A-opposing kinase. Thus, mTOR enhances the activity of S6K, a kinase translational regulator. In this way, the activation of MID1 is linked to an increase of translation.^[32] It has been described that in HD, the complex MID1 binds to *HTT* mRNA in a length-dependent manner.^[36]
2. Nuclear RNA foci co-localizing with muscleblind-like splicing regulator 1 (MBNL1) were found in fibroblasts of patients with HD and SCA3.^[37] Furthermore, alternative splicing defects of MBNL1 target genes were detected in HD fibroblasts and other human cell lines expressing expanded CAG repeats,^[37] suggesting a role for MBNL1 sequestration and alternative splicing defects in polyQ diseases.
3. Serine/arginine-rich splicing factor 6 (SRSF6/SRp55) binding to the CAG repeats is suggested to lead to SRSF6 loss of function and subsequent splicing defects in specific targets, although this mechanism has not been yet completely clarified.^[38,39]
4. Nucleolin, a protein of the nucleolus in control of rRNA transcription, shows impaired activity in HD, which in turn triggers apoptosis via the p53 pathway.^[40] It was shown by Tsoi et al.^[41,42] that CAG-repeat RNA induces nucleolar

stress, and is able to bind and sequester the nucleolar protein nucleolin, thus disturbing ribosome homeostasis.

The examples above illustrate that CAG mRNA modifies in a significant way several biological pathways. Thus, approaches towards blocking these aberrant interactions can be pursued as therapeutic strategies towards restoring normal protein function.

In the past years, two main strategies were followed to reach this aim: either targeting directly the CAG mRNA to hamper its aberrant interactions; or inhibiting RNA binding proteins by directly targeting their RNA binding motifs. In the subsequent pages, we will discuss advances and achievements following both strategies.

3. The CAG Hairpin: Insights into Structure, Stability, and Dynamics

Although the complete structural characterization of the *HTT* mRNA remains elusive, various RNA structural motifs have been identified for this macromolecule. Studies have consistently demonstrated that *HTT* mRNA transcripts containing CAG repeats can assume a stable secondary structure and fold into duplex helices and hairpins,^[43,44] and that interruptions in the normal *HTT* gene destabilize the smaller hairpin structure,^[45] suggesting a relevant role for the RNA structure towards inducing a toxic effect.

Within a normal CAG repeat number, in the *HTT* mRNA structure, a nearby CCG polymorphic trinucleotide repeat pairs with the polyCAG section, forming contiguous and small

bulges with an A–C mismatch in a hairpin. However, for mutant *HTT* mRNA, the CAG region interacts both with the polyCCG section, as well as with itself, forming a secondary bulge.^[46] The number of repeats is linked to the evolution from the bulge into a proper secondary hairpin, the latter thus being the one associated with the toxicity of RNA in HD.^[6]

Notably, sections of the *HTT* mRNA covering duplex helices have been experimentally obtained, and the isolated sections have been analyzed via NMR and X-ray crystallography (Table 1).

The first crystallized structure by Kilisek and coworkers^[52] in 2010 described the purine-purine mismatch due to the A–A interaction for the sequence r(–GG(CAG)₂CC–)₂, with weak hydrogen-bonding interaction between the two nucleobases. Pure electrostatic interactions were suggested as the main responsible for the affinity of RNA binding proteins towards RNA, such as MBNL1. Later studies^[50] also highlighted that the A–A mismatch widens the major groove, changing from a pure A-type RNA to a hybrid between A- and B-type RNA (with a C1'–C1' distance from complementary riboses of around 10.3 Å for C–G contacts, but 11.3 for A–A contacts).^[50] This widening was suggested to increase the accessibility of binding proteins or small molecules.^[50]

Tawani and Kumar also analyzed the latter sequence via NMR and restrained molecular dynamics (MD) studies.^[50] Unlike the previous report of Yildirim et al.,^[51] the internal CAG A–A pairs displayed the anti-anti conformation, but the external CAG A–A pairs displayed the anti-syn conformation (Figure 2B).

Later, Pan et al.^[53] suggested that spurious conformations could occur in the external CAG A–A pairs because they were

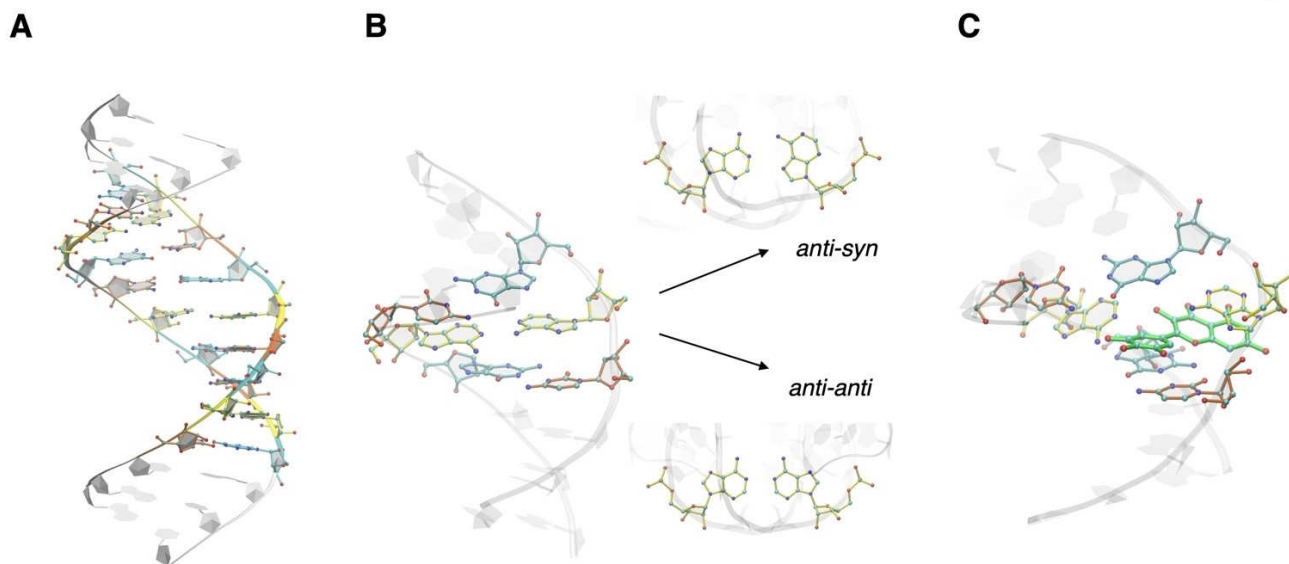


Figure 2. The CAG motif from experimental structures. The colors encode Cytidine (orange), Adenine (yellow) and Guanosine (cyan). The CAG can be found as a triplet (A, PDB ID: 5VH7),^[49] a single motif (B, PDB ID: 2MS5),^[50] or single motif co-crystallized with Myricetin (C, PDB ID: 5X11)^[48] are shown. The main conformations for the A–A mismatch are also highlighted (B, PDB ID: 4J50).^[51]

Table 1. RNA structures covering CAG motifs or repeats.

PDB ID ^[a]	Description ^[b]	Related entries	Year
6QIS ^[47]	Synthetic macrocycles binding to 2 CAG repeats ^[a]	6QIV, 6QIT, 6QIR, 6QIQ	2019
5X1I ^[48]	Flavonoids binding to a single CAG motif ^[b]	–	2018
5VH7 ^[49]	Double helical RNA with three CAG repeats ^[b]	–	2017
4YN6 ^[50]	Double helical RNA with three CAG repeats ^[b]	–	2015
2MS5 ^[50]	Double helical RNA with CAG motif ^[a]	–	2014
4J50 ^[51]	Double helical RNA with three CAG repeats ^[a]	–	2013
3NJ6 ^[52]	Double helical RNA with two CAG repeats ^[a]	3NJ7	2010

[a] X-ray solved structure [b] NMR solved structure.

flanked by CC or GG nucleotides, which were shown to be the least stable of the 10 dinucleotide steps.^[54] Because these steps were never present in a genuine (CAG)_n, its presence could bias the conformation of adjacent A–A pairs.

It was found that anti-anti conformations, followed by anti-syn conformations, were energetically preferred, by a very small energetic difference (~1 kcal/mol), which could explain all the previous findings. Transition rates of the A–A conformations were also found to be frequent. They also suggested that anti-anti interactions are indeed connected with a wider major groove, and a substantial decrease of the inclination angle, while anti-syn interactions favors the canonical A–RNA form.^[53]

Chen and coworkers^[49] published independently a study with NMR and restrained MD, which indicated that one- and zero-hydrogen bond states are the most frequent structures observed in the CAG A–A interactions, and that the A–A and U–U pairs induced distortions to A-form RNA of their helices, therefore corroborating the previous results (Figure 2A).

4. Small Molecules Binding to CAG Hairpins: Experiments and Computations

In this section, computational and experimental approaches regarding CAG-small molecule interactions will be described. In this respect, molecular simulation has revealed itself as an increasingly powerful tool to predict RNA/ligands adducts structures.^[55–58] In particular, all-atom molecular dynamics (MD) simulations reproduce rather accurately the structure, (sub-) microsecond conformational fluctuations and energetics of RNA oligos alone^[59] and in complex with ligands.^[60,61]

4.1 Probing the Druggability of CAG Triplet

Kumar and coworkers^[62] probed small molecules binding to the CAG-motif, specifically at the AA mismatch, with a combination of virtual screening and chemical similarity searching, based on known nucleic acid binders. A small molecule, 4-guanidinophenyl-4-guanidinobenzoate (hereafter

D6), was found to inhibit the CAG-MBNL1 complex in vivo, with D6 binding to the CAG hairpin with high affinity.

In its 2013 study, Yildirim et al.^[51] also suggested that the anti-syn conformation is quite flexible, and showed a plausible binding pocket near AA base pairs, thus reinforcing the viability of a binding pocket in the CAG motif.

The first structure confirming experimentally the interaction between the CAG motif and a small molecule was obtained by Khan and coworkers,^[48] showing the interaction of r(–CCG(CAG)CGG–)₂ with Myricetin using NMR spectroscopy and restrained MD (Figure 2C). The structure shows the flavonoid located in the A–A mismatch region, binding as an intercalator. As a result, one of the adenines is forced towards the solvent. Further π – π stacking and hydrogen bond interactions stabilize the bound pose.

Studies by Mukherjee et al.^[47] showed the binding of cyclic bis-naphthyridines to r(–G(CAG)₂C–)₂ by biophysical and X-ray studies. Interestingly, the results showed a proportion of two ligands per CAG triplet, where the ligands are located inside the RNA helix and mimic nucleobases. The ligands are located in the A–A mismatch, as in the flavonoid example, but in this case the immediate guanosine and cytosine are flipped out, widening and kinking the helix.

Overall, the two crystallized complexes are posed in the A–A mismatch but inducing quite different binding conformations. These results suggest a large plasticity of the CAG hairpin, which in turn can be used as an advantage when designing new small molecule binders towards CAG-motifs.

4.2 Docking Compounds into CAG Triplet Pockets

The analysis of the atomic details of CAG RNA-ligand interactions was greatly advanced by the availability of high-resolution structures of CAG RNA-ligand complexes discussed above. However, the experimental structure determination for RNA and its complexes is challenging and currently cannot be accomplished in a high-throughput manner. Therefore, the development and the implementation of computer software for modeling RNA-ligand complexes is becoming a key technology in this field.

Based on the results obtained by Myricetin, Khan and coworkers^[63] performed a chemical similarity search, in order

to find bioactive compounds towards the CAG motif. After identifying the top candidates, docking studies suggested the respective binding poses and the docking binding strength. Further testing analyzed the activity of these compounds towards alleviating the polyQ-mediated pathogenesis in HD cellular models and cells derived from HD patients.

In another study,^[64] a pyridocoumarin family of derivatives was also tested against the CAG motif. Docking studies were also used to model the putative binding mode, and further ITC and fluorescence studies were performed.

Taken together, when modeling the interaction between a small molecule and RNA, docking should be taken as a complimentary procedure or as an educated guess, due to the lack of proper force field parameterization (*vide infra*).

4.3 Molecular Dynamics Simulations and Free Energy Calculations for CAG Repeats Binding Ligands

Bochicchio and coworkers^[65] used computational tools towards predicting the binding pose and affinity of D6 (*vide supra*) towards a CAG-repeat oligo $r(-GG(CAG)_2CC-)_2$. The free energy surface (FES) was calculated with MD by well-tempered metadynamics^[66] as a function of three collective variables: distance between the centers of mass of the RNA and D6 (d_{CM}), the number of intermolecular hydrogen bonds (n_{HB}), and the number of intermolecular hydrophobic contacts (n_{HC}).

Although the starting pose was obtained with a docking procedure, the authors noticed that the initial pose changed significantly, indicating that the docked pose was far from the real minimum. The obtained *in silico* model suggested a non-intercalating binding mode for D6, with salt bridges between the guanidine tails of D6 and the phosphate backbone, and parallel (β ring) and T-shaped (α ring) π -stacking interactions with two adenines of the CAG repeats.

Based on this information, Matthes et al.^[67] identified a set of CAG-repeat binder candidates by *in silico* methods. One of those, Furamidine, decreases the protein level of HTT in an HD cell line model, demonstrating for the first time the activity of an RNA binder against mutant HTT protein in living cells. Free energy calculations were also performed on furamidine-CAG interaction, which agreed with experimental results. Moreover, furamidine reduces MID1 binding, as well as the binding of other RNA binding proteins, to HTT mRNA *in vitro*.^[67] A comparison of the two binding poses is shown in Figure 3.

Up to now, the affinity of ligands binding to RNA CAG repeats has been measured for oligos of nonpathogenic length.^[62,64] Knowledge of the binding poses of these ligands may help design new molecules, which eventually bind to CAG repeats with pathogenic length.

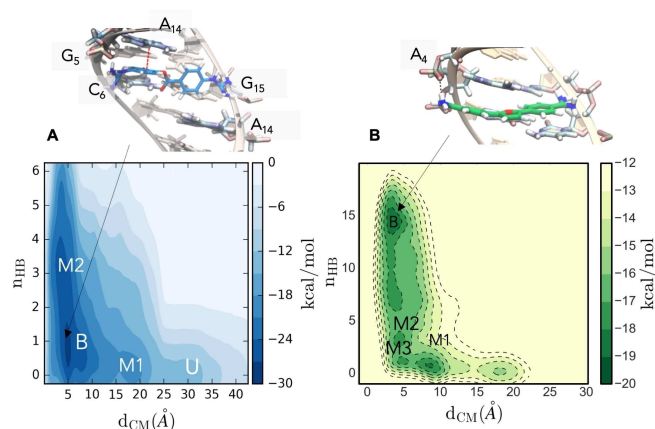


Figure 3. 2D free energy minima from metadynamics calculations for the compound D6 (A), and Furamidine (B).^[67]

5. Modulating Protein-RNA Interactions

Coding and non-coding RNA are handled and processed by several RNA-binding proteins (RBPs), which often feature evolutionarily conserved RNA binding domains (RBDs), e.g. RNA recognition motifs (RRMs).^[68] A recently compiled evolutionary-oriented database of RRBs (RRMdb) was put forward by Nowacka et al.^[69] Also, high throughput RNA interactome capture (RIC) experiments performed by Hentze et al. have significantly expanded our knowledge of existing RBPs.^[70] Interestingly, they found that many proteins detected in screens for RBPs lack a structured Pfam domain, while detected RBDs were enriched in intrinsically disordered regions (IDRs). This suggests that many RNA-protein interactions in RBPs occur via non-structured proteins or regions.^[71]

5.1 *In Silico* Approaches to Detect RNA Binding Proteins

Encouragingly, experimental methods have been augmented with computational approaches to reliably predict RBPs. Such algorithms can be classified as either sequence-, or template-based. The former primarily depend on evolutionary information and physiochemical similarity, e.g. RBPPred,^[72] whereas template-based methods also incorporate structural information of RBPs, e.g. focusing on electrostatic complementarity.^[73] In this regard, support vector machine models (SVM) were successfully used in both mentioned examples.

SVMs were trained using large-scale mass spectrometry data, exploiting the strong tendency of RBPs to interact with each other in an algorithm known as SONAR (Support vector machine Obtained from Neighborhood Associated RBPs) and showed good predictive ability.^[74] SPOT-Seq-RNA represents an alternative template-based technique that predicts protein-RNA complex structures and their associated binding affinities

using a knowledge-based energy function, DRNA, derived from known protein-RNA complexes.^[75,76]

Apart from RBPs, there is considerable interest in the computational development of algorithms that map the RNA-interacting surface of proteins to predict the RNA binding residues that facilitate protein-RNA interactions.^[77,78] Miao and Westhof reported a large scale assessment of 19 web servers and 3 stand-alone programs on 41 datasets including more than 5000 proteins derived from 3D structures of protein-nucleic acid complexes.^[79] An updated overview is presented by Wan et al.^[80]

5.2 Predicting RNA-protein Complexes

Experimental RBP prediction in combination with computational approaches will further our understanding of mutant *HTT* synthesis, regulation, modification, and degradation. In turn, these insights can guide the rational design of agents that hamper the correct interaction of mutant *HTT* RNA with HD-relevant RBPs, thus representing an intriguing therapeutic strategy.

In this respect, the development of RNA-protein docking algorithms has been hampered by many obstacles. Although initially, researchers just extended algorithms for protein-protein docking to protein-nucleic acids complexes, the high flexibility, often associated with RNA molecules, rendered searching-algorithms using only static RNA structures questionable at best. Additionally, while small molecule protein docking programs benefited from the ever growing amount of structural protein data, the number of experimentally elucidated RNA-protein complexes still lags behind.^[77] Several programs have been developed to mitigate this situation (see footnote¹).

Due to the growth of high-throughput experimental technologies, like Chip-Chip and Chip-Seq, a lot of RNA-protein interactions have been revealed, which enables data-driven inference of RNA-protein associations.^[91] Various machine learning methods have been employed for the prediction of RBP-RNA interactions. In this regard, methods based on SVM such as BindN/BindN+^[92,93] the PiRaNhA web server,^[94] or PSSMs-SVMs^[95] have been developed. Other algorithms rely on decision trees, e.g. NAPS,^[96] Naïve Bayes classifiers, e.g., RNABindR^[97] or random forests, e.g. PRBR.^[98]

Notable recent advancements include DisoRDPbind, a method to predict intrinsically disordered protein interactions with RNA^[99] and long short-term memory (LSTM) neural networks approaches for generative models to construct single-stranded nucleic acids which target proteins with high affinity.^[100]

6. Challenges and Opportunities of RNA Modeling

To date, only short segments of the pathogenic RNA have been obtained by structural studies, which leaves important questions regarding the behaviour of the toxic RNA in more than four repeats, or the interaction of this with other relevant binding proteins. In this way, computational approaches in the form of molecular dynamics or molecular docking have been used for several years in many applications to successfully model the RNA structure and dynamics.

In the next lines, some of the current options for RNA modeling are discussed, along with current limitations and challenges.

6.1 RNA-small Molecules Docking Approaches

After the development of the first scoring function specific for RNA-ligand complexes, by Morley et al.^[101] for their proprietary high-throughput docking program RiboDock, several novel methodologies tailored for ligand-RNA docking were proposed.

Among these, we can find: (i) Knowledge-based scoring functions for evaluating RNA-ligand complexes like DrugScoreRNA,^[102] or LigandRNA,^[103] where both methods use the analysis of RNA-ligand contact distances for evaluating the stability of the complexes. (ii) Grid-based scoring functions, like the one used in DOCK6,^[56] or MORDOR,^[104] that implement the nonbonded terms of the AMBER and other molecular mechanics force field.

6.2 Atomistic Force Fields

Atomistic molecular dynamics (MD) simulation is a powerful tool for characterizing, at atomistic level, the conformational changes undergone by proteins. The application of such tools to RNA structures, however, has proven more challenging, mostly due to the fact that the physical models (force fields) available for MD simulations of RNA molecules are considered substantially less accurate in many aspects than those currently available for proteins.^[105]

The stability of nucleic acids depends on a delicate balance of electrostatic and Van der Waals forces, which has not been properly treated in current force fields. Current problems that have been commonly mentioned^[105–109] are the following:

- torsional parameters on the RNA backbone and glycosidic bonds.

¹Currently, popular programs and web servers include RosettaDock server,^[81] 3D-Garden,^[82] HADDOCK,^[83] HEX server,^[84] SwarmDock,^[85] ZDOCK server,^[86] ATTRACT,^[87] pyDOCKSAXS,^[88] InterEvDOCK,^[89] and HDOCK.^[90]

- overstabilization of nucleobase stacking by electrostatic and van der Waals (vdW) interactions.
- underestimation of base-pairing strength, which can lead to a destabilization of the proper native state.
- excessive stabilization of the unfolded single-stranded RNA ensemble by intramolecular base-phosphate and sugar-phosphate interactions
- conformational sampling of the 2-hydroxyl group on the sugar.
- charged species are not properly handled, such as phosphate groups, polyvalent ions, etc.

The AMBER force field^[107,110] is currently the most widely used force field for MD simulations of RNA systems. Initial efforts to improve the accuracy of the AMBER RNA force field have aimed largely at improving the description of canonical double-helical RNA structure, by refining backbone and glycosidic torsion parameters on the basis of quantum mechanical (QM) calculations^[107,111–115] or experimental data.^[116] Also, van der Waals parameters were optimized to avoid overstabilization of nucleobase stacking and the underestimation of base-pairing strength.^[117–119] This improved the correct modeling of RNA tetraloops, but still did not optimize the accuracy of other RNA systems, like single strand RNA.

Recently, the D. E. Shaw's group further modified the electrostatic, vdW, and torsional parameters of the AMBER ff14 RNA force field^[110] by using a combination of *ab initio* and empirical methods. The new RNA force field showed a more accurate reproduction of the energetics of nucleobase stacking, base pairing, and key torsional conformers for extended and structured RNA systems, including short and long single strand RNAs, RNA duplexes, tetraloops, and riboswitches.^[105]

Kuhrova et al.^[106] suggested a different approach, by selectively fine-tuning H-bonds. The gHBfix potential introduced highlights that conventional reparameterization of dihedral potentials or non-bonded terms can lead to major undesired side effects, and that the addition of these extra terms improves the force field performance while avoiding introduction of major new imbalances.

Force fields usually consist of several empirical energy terms, including short-range bonded interactions and non-bonded interactions such as dispersion and electrostatics. In particular, non-polarizable force fields use fixed point charges to represent electrostatic interactions. The main limitation of said force fields is the absence of polarization, *i.e.*, the response of the charge distribution to the environment. This is particularly relevant when dealing with small chemicals interacting with highly charged biomolecules, like RNA. Also, atom-centered point charge models are very different from the realistic charge distributions, which are usually smooth and anisotropic. Moreover, these models cannot reproduce charge penetration effects that occur when atomic electron clouds overlap.^[120,121] Such effects are critical for determining the equilibrium geometry and energy of molecular complexes.^[122–124]

With respect to the previous shortcomings, some strategies have been followed: (i) Charge anisotropy can be reproduced by the adoption of higher-order multipolar electrostatics models or the addition of off-center charge sites.^[125–127] In this regard, atomic multipoles truncated at the quadrupole term have been shown to model phenomena such as σ -holes, lone pairs, and π -bonding. (ii) Charge penetration can be modeled as a softening of the electrostatic interaction at short ranges, usually with the aid of empirically determined damping functions.^[128] Charge penetration models can be combined with the above anisotropic electrostatic model, by placing charge densities on bonds or lone pairs,^[129] or by using damping functions for atomic multipoles.^[128]

Advances in incorporating electron polarization in RNA simulations are represented by the polarizable Drude force field for RNA^[108] (an extension of the Drude-2017 force field),^[130] or the AMOEBA force field for nucleic acids.^[131] It is worth mentioning that albeit the description of the biomolecular system often increases, the computational demand for the computation also rises. In this respect, both Drude and AMOEBA have their own OpenMM version which enables GPU computing^[132–134] and AMOEBA also has its Tinker-HP version for parallel computing.^[135]

Overall, atomistic force fields are suitable for modeling short RNA-paired sequences, RNA-small molecule interactions, and RNA protein interactions.

6.3 Coarse-grained Force Fields

Coarse-grained force fields have gained attention as potential functions that can sample conformational spaces of larger molecules in a reduced computational time, at the cost of a less defined atomic description. In these models, more than an atom is encoded into a bead (pseudo-atom), which thus reduces the total number of particles in the system.

These potentials are able to study biologically interesting processes such as the folding of RNA structures, as in the TAR hairpin,^[136] which usually develop over very long time scales, where normal atomistic force fields would require prohibitive amounts of computational power.

Several force fields exist, which can go from very coarse definitions, using 1 (NAST)^[137] or 3^[138,139] pseudo-atoms per nucleotide, to more defined models, ranging from 5 to 7 pseudo-atoms (RACER,^[136] SimRNA,^[140] HiRE-RNA,^[141] MARTINI^[142]) per nucleotide. In general, the more detailed CG models agree that higher resolution in the backbone or in the nucleobases improves the overall behaviour of the simulations.

However, one of the major issues regarding coarse-grained force fields in RNA is the lack of directionality on key molecular interactions such as hydrogen bonds, or the lack of an explicit description of non-canonical base pairs, which are critical for RNA structures.^[143]

7. Current Therapeutic Approaches and Perspectives

Here we give a brief overview of the many therapeutics which are currently approved or under clinical trial for HD, as well as their biological rationale.^[144,145]

7.1 Small Molecule Approaches

Some small molecules targeting excitotoxicity have shown efficacy in humans, such as Tetrabenazine, Deutetrabenazine and Memantine.^[146] Another strategy is hampering HTT proteolysis, by inhibiting caspase-1 and -3 mRNA upregulation, e.g. with Minocycline.^[147]

Since mitochondrial dysfunction plays a major role already at an early stage of HD pathogenesis,^[148] researchers investigated Meclizine and Cystamine with success in mouse/fly models.^[149,150] Furthermore, targeting transcriptional dysregulation with histone deacetylase inhibitors, e.g. sodium phenylbutyrate or suberoylanilide hydroxamic acid proved efficient in mice models.^[151,152]

Also, as shown by Matthes et al.^[67] and Kumar and coworkers,^[62] it is possible to decrease mRNA toxicity, as well as hamper its aberrant binding with proteins, by directly targeting the CAG hairpin in the *HTT* mRNA with small molecules, like D6 or Furamidine.

7.2 Small Biologics

Yamamoto et al. showed with a conditional mouse model the potential for silencing mutant *HTT* expression as a putative treatment option for HD.^[153] Moreover, they demonstrated that a continuous influx of mutant protein is essential to maintain inclusions and symptoms. Thus, the silencing of the mutant *HTT* gene is an interesting therapeutic approach.

In general, RNA-targeted treatments in HD are based on RNA interference (RNAi), antisense oligonucleotides (ASO), or small-molecule splicing modulators.

RNA interference. RNAi relies on either duplex RNAs (dsRNAs) or chemically modified single-stranded RNAs (ssRNAs),^[154–156] which are introduced into the body for robust and sustained suppression of the target gene product.^[157,158] Since these miRNA are known to produce only a partial knockdown in the transduced region, further elucidation of the precise mechanism behind miRNA-based *HTT* silencing in the brain with computational biology provides an intriguing line of research.^[159–161]

Antisense Oligonucleotides (ASO). Single-stranded artificial DNA molecules either bind to mRNA resulting in mRNA degradation, or interfere with ribosomal attachment, modulating RNA splicing.^[144,162] However, some disadvantages such as complicated administration procedures are present.^[163]

HD patients with two mutant alleles are rare and limited data suggested a similar disease onset age but a more severe clinical progression,^[164] while *HTT* has been associated with neuroprotective effects.^[165] Therefore, allele-selective ASOs, targeting only mutant *HTT*, were investigated by Carroll et al. Their ASOs exploit single-nucleotide polymorphisms (SNPs) to increase selectivity for the mutant allele.^[166] They showed furthermore, that therapies targeting as few as three SNPs may benefit 85 % of patients with HD of northern European and indigenous South American descent.

On the other hand, Tominersen (formerly known as RG6042 and IONIS-HTTRx) is a non-allele-selective ASO, consisting of a synthetic 20-nucleotide DNA-like sequence, where oxygen-to-sulphur backbone substitutions, and 2'-O-methoxyethyl modifications improve efficacy and the pharmacokinetic profile.^[167] Tominersen was the first agent to be assessed in clinical trials and researches estimated an 55–85 % reduction in cortical mHTT and a 20–50 % reduction in caudate mHTT while no dose-limiting toxicity was noted.^[144,168]

7.3 RNA-processing Enzymes

Voßfeldt et al. performed a near genome-wide RNAi screen for modulators of polyglutamine-induced neurotoxicity in *Drosophila*.^[169] In this work, the authors grouped putative modifiers into three categories: 1. Protein turnover/quality control (*Trp2*, *DnaJ-1*, *Hop*, *Hsc70Cb*, *Hsc70-4*, *Prosp2*, etc) 2. Nuclear import/export (*emb*, *Ntf-2* and *CG5738*) and 3. mRNA transport/editing and translation (*orb*, *Nelf-E*, *Prp8*, etc). Silencing the fly ortholog (dTRMT2a) of human tRNA methyltransferase homolog 2a (TRMT2a) potentially decreased polyQ-induced photoreceptor degeneration, polyQ aggregates, and toxicity. Moreover, a reduction in the amount of polyQ aggregates and polyQ-toxicity in other disease models was reported. Therefore, inhibition of TRMT2a enzyme in patients affected by polyglutamine disorders, including HD, might be a promising therapeutic strategy.

8. Outlook

HD is caused by CAG repeat expansions in the *HTT* gene, encoding the huntingtin protein. Although no disease-curing or disease-slowing treatment currently exists for HD, the development of therapies to target *HTT* transcription and the translation of its mRNA is of great interest, and it is currently under intense investigation.

Here we have reviewed the current knowledge of RNA pathogenicity in HD and how this was exploited, with both experimental and computational approaches, to propose novel therapeutic strategies. Up to now, three main methods to reduce *HTT* mRNA toxicity emerged: These are ASOs, RNAi, and small-molecules either targeting splicing modulators or

directly the *HTT* mRNA, i.e. specifically the toxic CAG hairpin.

Clinical trials using ASOs targeting *HTT* mRNA are currently underway, and future ASO trials are planned, including agents that seek to lower mutant *HTT* selectively. On the other hand, the absence of human trials for RNAi reflects the distribution and delivery challenges associated with RNAi: RNA does not distribute well throughout brain tissue, so stereotactic surgery is needed to deliver the agent via a viral vector, at the cost of increased invasiveness, and risks associated with long-term toxicity.

For these reasons, small molecule RNA-targeting compounds are a viable strategy: a brain-penetrant, orally bioavailable small molecule acting on protein manufacture to lower *HTT* expression or on the toxic CAG hairpin to lower its aberrant ability to recruit proteins, is appealing, specifically in terms of easiness of delivery.

Computational methods are becoming key tools in advancing the understanding of the biomolecular interactions and the development of novel, potent modulators of biological pathways. Overall, some advances have been made towards inhibiting protein-RNA interactions through small molecules, peptides, and nucleic acids. However, more research is required to improve the quality of structural predictions via force fields or statistical learning tools.

We expect that as the development of new modulators increases, novel data-intensive tools from the machine learning field will further contribute here, and the resulting models will benefit both computational and experimental research, in order to manage to tackle these challenging neurodegenerative diseases.

Acknowledgements

O. P.-H. has been supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 765048. We are grateful to Prof. Paolo Carloni, and to the reviewers for critical reading of this manuscript and the valuable suggestions.

References

- [1] G. Huntington, *J. Neuropsychiatry Clin. Neurosci.* **2003**, *15*, 109–112.
- [2] J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, *Nature* **1983**, *306*, 234–238.
- [3] C. A. Ross, S. J. Tabrizi, *Lancet Neurol.* **2011**, *10*, 83–98.
- [4] B. Sahoo, I. Arduini, K. W. Drombosky, R. Kodali, L. H. Sanders, J. T. Greenamyre, R. Wetzel, *PLoS One* **2016**, *11*, e0155747.
- [5] E. Marti, *Brain Pathol.* **2016**, *26*, 779–786.
- [6] R. Nalavade, N. Griesche, D. P. Ryan, S. Hildebrand, S. Krauss, *Cell Death Dis.* **2013**, *4*, e752.
- [7] V. V. Belzil, T. F. Gendron, L. Petrucelli, *Mol. Cell. Neurosci.* **2013**, *56*, 406–419.
- [8] J.-M. Gallo, P. Jin, C. A. Thornton, H. Lin, J. Robertson, I. D'Souza, W. W. Schlaepfer, *J. Neurosci.* **2005**, *25*, 10372–10375.
- [9] J. C. Gomez-Verjan, E. R. Vazquez-Martinez, N. A. Rivero-Segura, R. H. Medina-Campos, *Hum. Genet.* **2018**, *137*, 865–879.
- [10] B. Swinnen, W. Robberecht, L. Van Den Bosch, *EMBO J.* **2020**, *39*, e101112.
- [11] T. A. Cooper, L. Wan, G. Dreyfuss, *Cell* **2009**, *136*, 777–793.
- [12] ENCODE Project Consortium, E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, et al., *Nature* **2007**, *447*, 799–816.
- [13] M. E. Wilkinson, C. Charenton, K. Nagai, *Annu. Rev. Biochem.* **2019**, DOI 10.1146/annurev-biochem-091719-064225.
- [14] F. Tuorto, R. Parlato, *J. Mol. Biol.* **2019**, *431*, 1763–1779.
- [15] J. M. Schmiedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Blüthgen, D. S. Marks, A. van Oudenaarden, *Science* **2015**, *348*, 128–132.
- [16] H.-I. Im, P. J. Kenny, *Trends Neurosci.* **2012**, *35*, 325–334.
- [17] R. F. Ketting, S. E. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, R. H. Plasterk, *Genes Dev.* **2001**, *15*, 2654–2659.
- [18] Y. Zeng, R. Yi, B. R. Cullen, *EMBO J.* **2005**, *24*, 138–148.
- [19] V. N. Kim, J. Han, M. C. Siomi, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 126–139.
- [20] D. P. Bartel, *Cell* **2018**, *173*, 20–51.
- [21] S.-T. Lee, K. Chu, W.-S. Im, H.-J. Yoon, J.-Y. Im, J.-E. Park, K.-H. Park, K.-H. Jung, S. K. Lee, M. Kim, et al., *Exp. Neurol.* **2011**, *227*, 172–179.
- [22] R. Johnson, C. Zuccato, N. D. Belyaev, D. J. Guest, E. Cattaneo, N. J. Buckley, *Neurobiol. Dis.* **2008**, *29*, 438–445.
- [23] K. Pircs, R. Petri, S. Madsen, P. L. Brattås, R. Vuono, D. R. Ottosson, I. St-Amour, B. A. Hersbach, M. Matusiak-Brückner, S. H. Lundh, et al., *Cell Rep.* **2018**, *24*, 1397–1406.
- [24] J. N. Savas, A. Makusky, S. Ottosen, D. Baillat, F. Then, D. Krainc, R. Shiekhata, S. P. Markey, N. Tanese, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10820–10825.
- [25] Z. Liu, T. Zhou, A. C. Ziegler, P. Dimitrion, L. Zuo, *Oxid. Med. Cell. Longev.* **2017**, *2017*, 2525967.
- [26] A. Emde, E. Hornstein, *EMBO J.* **2014**, *33*, 1428–1437.
- [27] J. Konovalova, D. Gerasymchuk, I. Parkkinen, P. Chmielarz, A. Domanskyi, *Int. J. Mol. Sci.* **2019**, *20*, DOI 10.3390/ijms20236055.
- [28] L.-B. Li, Z. Yu, X. Teng, N. M. Bonini, *Nature* **2008**, *453*, 1107–1111.
- [29] M. Bañez-Coronel, S. Porta, B. Kagerbauer, E. Mateu-Huertas, L. Pantano, I. Ferrer, M. Guzmán, X. Estivill, E. Martí, *PLoS Genet.* **2012**, *8*, e1002481.
- [30] X. Sun, P. P. Li, S. Zhu, R. Cohen, L. O. Marque, C. A. Ross, S. M. Pulst, H. Y. E. Chan, R. L. Margolis, D. D. Rudnicki, *Sci. Rep.* **2015**, *5*, 12521.
- [31] K. Sobczak, G. Michlewski, M. de Mezer, E. Kierzek, J. Krol, M. Olejniczak, R. Kierzek, W. J. Krzyzosiak, *J. Biol. Chem.* **2010**, *285*, 12755–12764.
- [32] N. Griesche, J. Schilling, S. Weber, M. Rohm, V. Pesch, F. Matthes, G. Auburger, S. Krauss, *Front. Cell. Neurosci.* **2016**, *10*, 226.
- [33] M. Wojciechowska, W. J. Krzyzosiak, *Hum. Mol. Genet.* **2011**, *20*, 3811–3821.
- [34] R.-J. Hsu, K.-M. Hsiao, M.-J. Lin, C.-Y. Li, L.-C. Wang, L.-K. Chen, H. Pan, *PLoS One* **2011**, *6*, e16417.

- [35] J. Schilling, M. Broemer, I. Atanassov, Y. Duernberger, I. Vorberg, C. Dieterich, A. Dagane, G. Dittmar, E. Wanker, W. van Roon-Mom, et al., *J. Mol. Biol.* **2019**, *431*, 1869–1877.
- [36] S. Krauss, N. Griesche, E. Jastrzebska, C. Chen, D. Rutschow, C. Achmüller, S. Dorn, S. M. Boesch, M. Lalowski, E. Wanker, et al., *Nat. Commun.* **2013**, *4*, 1511.
- [37] A. Mykowska, K. Sobczak, M. Wojciechowska, P. Kozłowski, W. J. Krzyzosiak, *Nucleic Acids Res.* **2011**, *39*, 8938–8951.
- [38] K. Sathasivam, A. Neueder, T. A. Gipson, C. Landles, A. C. Benjamin, M. K. Bondulich, D. L. Smith, R. L. M. Faull, R. A. C. Roos, D. Howland, et al., *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2366–2370.
- [39] M. Fernández-Nogales, J. R. Cabrera, M. Santos-Galindo, J. J. M. Hoozemans, I. Ferrer, A. J. M. Rozemuller, F. Hernández, J. Avila, J. J. Lucas, *Nat. Med.* **2014**, *20*, 881–885.
- [40] K. Kalita, D. Makonchuk, C. Gomes, J.-J. Zheng, M. Hetman, *J. Neurochem.* **2008**, *105*, 2286–2299.
- [41] H. Tsoi, T. C.-K. Lau, S.-Y. Tsang, K.-F. Lau, H. Y. E. Chan, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13428–13433.
- [42] H. Tsoi, H. Y. E. Chan, *Cerebellum* **2013**, *12*, 310–312.
- [43] B. A. Lenzmeier, C. H. Freudenreich, *Cytogenet. Genome Res.* **2003**, *100*, 7–24.
- [44] C. J. Cummings, H. Y. Zoghbi, *Hum. Mol. Genet.* **2000**, *9*, 909–916.
- [45] K. Sobczak, W. J. Krzyzosiak, *J. Biol. Chem.* **2005**, *280*, 3898–3910.
- [46] S. Busan, K. M. Weeks, *Biochemistry* **2013**, *52*, 8219–8225.
- [47] S. Mukherjee, L. Błaszczyk, W. Rypniewski, C. Falschlunger, R. Micura, A. Murata, C. Dohno, K. Nakatani, A. Kiliszek, *Nucleic Acids Res.* **2019**, *47*, 10906–10913.
- [48] E. Khan, A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra, A. Kumar, *ACS Chem. Biol.* **2018**, *13*, 180–188.
- [49] J. L. Chen, D. M. VanEtten, M. A. Fountain, I. Yildirim, M. D. Disney, *Biochemistry* **2017**, *56*, 3463–3474.
- [50] A. Tawani, A. Kumar, *PLoS One* **2015**, *10*, e0131788.
- [51] I. Yildirim, H. Park, M. D. Disney, G. C. Schatz, *J. Am. Chem. Soc.* **2013**, *135*, 3528–3538.
- [52] A. Kiliszek, R. Kierzek, W. J. Krzyzosiak, W. Rypniewski, *Nucleic Acids Res.* **2010**, *38*, 8370–8376.
- [53] F. Pan, V. H. Man, C. Roland, C. Sagui, *Biophys. J.* **2017**, *113*, 19–36.
- [54] D. Svozil, P. Hobza, J. Sponer, *J. Phys. Chem. B* **2010**, *114*, 1191–1203.
- [55] S. Fulle, H. Gohlke, *J. Mol. Biol.* **2009**, *387*, 502–517.
- [56] P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James, I. D. Kuntz, *RNA* **2009**, *15*, 1219–1230.
- [57] P. Daldrop, F. E. Reyes, D. A. Robinson, C. M. Hammond, D. M. Lilley, R. T. Batey, R. Brenk, *Chem. Biol.* **2011**, *18*, 324–335.
- [58] L. Chen, G. A. Calin, S. Zhang, *J. Chem. Inf. Model.* **2012**, *52*, 2741–2753.
- [59] F. Musiani, G. Rossetti, L. Capece, T. M. Gerger, C. Micheletti, G. Varani, P. Carloni, *J. Am. Chem. Soc.* **2014**, *136*, 15631–15637.
- [60] S. Fulle, N. A. Christ, E. Kestner, H. Gohlke, *J. Chem. Inf. Model.* **2010**, *50*, 1489–1501.
- [61] T. N. Do, E. Ippoliti, P. Carloni, G. Varani, M. Parrinello, *J. Chem. Theory Comput.* **2012**, *8*, 688–694.
- [62] A. Kumar, R. Parkesh, L. J. Szajder, J. L. Childs-Disney, K. Sobczak, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 496–505.
- [63] E. Khan, S. K. Mishra, R. Mishra, A. Mishra, A. Kumar, *Sci. Rep.* **2019**, *9*, 16872.
- [64] E. Khan, S. Biswas, S. K. Mishra, R. Mishra, S. Samanta, A. Mishra, A. Tawani, A. Kumar, *Biochimie* **2019**, *163*, 21–32.
- [65] A. Bochicchio, G. Rossetti, O. Tabarrini, S. Krauß, P. Carloni, *J. Chem. Theory Comput.* **2015**, *11*, 4911–4922.
- [66] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 020603.
- [67] F. Matthes, S. Massari, A. Bochicchio, K. Schorpp, J. Schilling, S. Weber, N. Offermann, J. Desantis, E. Wanker, P. Carloni, et al., *ACS Chem. Neurosci.* **2018**, *9*, 1399–1408.
- [68] T. Glisovic, J. L. Bachorik, J. Yong, G. Dreyfuss, *FEBS Lett.* **2008**, *582*, 1977–1986.
- [69] M. Nowacka, P. Boccaletto, E. Jankowska, T. Jarzynka, J. M. Bujnicki, S. Dunin-Horkawicz, *Database (Oxford)* **2019**, *2019*, DOI 10.1093/database/bay148.
- [70] M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 327–341.
- [71] S. Calabretta, S. Richard, *Trends Biochem. Sci.* **2015**, *40*, 662–672.
- [72] X. Zhang, S. Liu, *Bioinformatics* **2017**, *33*, 854–862.
- [73] S. Shazman, Y. Mandel-Gutfreund, *PLoS Comput. Biol.* **2008**, *4*, e1000146.
- [74] K. W. Brannan, W. Jin, S. C. Huelga, C. A. S. Banks, J. M. Gilmore, L. Florens, M. P. Washburn, E. L. Van Nostrand, G. A. Pratt, M. K. Schwinn, et al., *Mol. Cell* **2016**, *64*, 282–293.
- [75] H. Zhao, Y. Yang, Y. Zhou, *Nucleic Acids Res.* **2011**, *39*, 3017–3025.
- [76] Y. Yang, H. Zhao, J. Wang, Y. Zhou, *Methods Mol. Biol.* **2014**, *1137*, 119–130.
- [77] S. Dvir, A. Argoetti, Y. Mandel-Gutfreund, *Curr. Opin. Struct. Biol.* **2018**, *53*, 124–130.
- [78] S. Jones, *Biophys. Rev. Lett.* **2016**, *8*, 359–367.
- [79] Z. Miao, E. Westhof, *PLoS Comput. Biol.* **2015**, *11*, e1004639.
- [80] H. Wan, J.-M. Li, H. Ding, S.-X. Lin, S.-Q. Tu, X.-H. Tian, J.-P. Hu, S. Chang, *Protein Pept. Lett.* **2019**, DOI 10.2174/0929866526666191028162302.
- [81] S. Lyskov, J. J. Gray, *Nucleic Acids Res.* **2008**, *36*, W233–8.
- [82] V. I. Lesk, M. J. E. Sternberg, *Bioinformatics* **2008**, *24*, 1137–1144.
- [83] S. J. de Vries, M. van Dijk, A. M. J. J. Bonvin, *Nat. Protoc.* **2010**, *5*, 883–897.
- [84] G. Macindoe, L. Mavridis, V. Venkatraman, M.-D. Devignes, D. W. Ritchie, *Nucleic Acids Res.* **2010**, *38*, W445–9.
- [85] M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, P. A. Bates, *Bioinformatics* **2013**, *29*, 807–809.
- [86] B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, Z. Weng, *Bioinformatics* **2014**, *30*, 1771–1773.
- [87] S. J. de Vries, C. E. M. Schindler, I. Chauvot de Beauchêne, M. Zacharias, *Biophys. J.* **2015**, *108*, 462–465.
- [88] B. Jiménez-García, C. Pons, D. I. Svergun, P. Bernadó, J. Fernández-Recio, *Nucleic Acids Res.* **2015**, *43*, W356–61.
- [89] J. Yu, M. Vavrusa, J. Andreani, J. Rey, P. Tufféry, R. Guerois, *Nucleic Acids Res.* **2016**, *44*, W542–9.
- [90] Y. Yan, D. Zhang, P. Zhou, B. Li, S.-Y. Huang, *Nucleic Acids Res.* **2017**, *45*, W365–W373.
- [91] X. Pan, Y. Yang, C.-Q. Xia, A. H. Mirza, H.-B. Shen, *Wiley Interdiscip. Rev. RNA* **2019**, *10*, e1544.
- [92] L. Wang, S. J. Brown, *Nucleic Acids Res.* **2006**, *34*, W243–8.
- [93] L. Wang, C. Huang, M. Q. Yang, J. Y. Yang, *BMC Syst. Biol.* **2010**, *4 Suppl 1*, S3.
- [94] Y. Murakami, R. V. Spriggs, H. Nakamura, S. Jones, *Nucleic Acids Res.* **2010**, *38*, W412–6.
- [95] M. Kumar, M. M. Gromiha, G. P. S. Raghava, *Proteins* **2008**, *71*, 189–194.

- [96] M. B. Carson, R. Langlois, H. Lu, *Nucleic Acids Res.* **2010**, *38*, W431–5.
- [97] M. Terribilini, J. D. Sander, J.-H. Lee, P. Zaback, R. L. Jernigan, V. Honavar, D. Dobbs, *Nucleic Acids Res.* **2007**, *35*, W578–84.
- [98] X. Ma, J. Guo, J. Wu, H. Liu, J. Yu, J. Xie, X. Sun, *Proteins* **2011**, *79*, 1230–1239.
- [99] C. J. Oldfield, Z. Peng, L. Kurgan, *Methods Mol. Biol.* **2020**, *2106*, 225–239.
- [100] J. Im, B. Park, K. Han, *BMC Genomics* **2019**, *20*, 967.
- [101] S. D. Morley, M. Afshar, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 189–208.
- [102] P. Pfeffer, H. Gohlke, *J. Chem. Inf. Model.* **2007**, *47*, 1868–1876.
- [103] A. Philips, K. Milanowska, G. Lach, J. M. Bujnicki, *RNA* **2013**, *19*, 1605–1616.
- [104] C. Guilbert, T. L. James, *J. Chem. Inf. Model.* **2008**, *48*, 1257–1268.
- [105] D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1346–E1355.
- [106] P. Kührová, V. Mlýnský, M. Zgarbová, M. Krepl, G. Bussi, R. B. Best, M. Otyepka, J. Šponer, P. Banáš, *J. Chem. Theory Comput.* **2019**, *15*, 3288–3305.
- [107] J. Šponer, P. Banáš, P. Jurečka, M. Zgarbová, P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer, M. Otyepka, *J. Phys. Chem. Lett.* **2014**, *5*, 1771–1782.
- [108] J. A. Lemkul, A. D. MacKerell, *J. Comput. Chem.* **2018**, *39*, 2624–2646.
- [109] E. J. Denning, A. D. MacKerell, *J. Am. Chem. Soc.* **2012**, *134*, 2800–2806.
- [110] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- [111] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817–3829.
- [112] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [113] M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T. E. Cheatham, P. Jurečka, *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- [114] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, et al., *Nat. Methods* **2016**, *13*, 55–58.
- [115] I. Yildirim, H. A. Stern, S. D. Kennedy, J. D. Tubbs, D. H. Turner, *J. Chem. Theory Comput.* **2010**, *6*, 1520–1531.
- [116] A. Gil-Ley, S. Bottaro, G. Bussi, *J. Chem. Theory Comput.* **2016**, *12*, 2790–2798.
- [117] C. A. Morgado, P. Jurečka, D. Svozil, P. Hobza, J. Šponer, *J. Chem. Theory Comput.* **2009**, *5*, 1524–1544.
- [118] A. A. Chen, A. E. García, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16820–16825.
- [119] P. Banáš, A. Mládek, M. Otyepka, M. Zgarbová, P. Jurečka, D. Svozil, F. Lankaš, J. Šponer, *J. Chem. Theory Comput.* **2012**, *8*, 2448–2460.
- [120] J. A. Rackers, Q. Wang, C. Liu, J.-P. Piquemal, P. Ren, J. W. Ponder, *Phys. Chem. Chem. Phys.* **2016**, *19*, 276–291.
- [121] T. M. Parker, C. D. Sherrill, *J. Chem. Theory Comput.* **2015**, *11*, 4197–4204.
- [122] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, P. Ren, *Annu. Rev. Biophys.* **2019**, *48*, 371–394.
- [123] R. G. Huber, M. A. Margreiter, J. E. Fuchs, S. von Grafenstein, C. S. Tautermann, K. R. Liedl, T. Fox, *J. Chem. Inf. Model.* **2014**, *54*, 1371–1379.
- [124] A. N. Bootsma, A. C. Doney, S. E. Wheeler, *J. Am. Chem. Soc.* **2019**, *141*, 11027–11035.
- [125] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, et al., *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- [126] P. Ren, J. Chun, D. G. Thomas, M. J. Schnieders, M. Marucho, J. Zhang, N. A. Baker, *Q. Rev. Biophys.* **2012**, *45*, 427–491.
- [127] G. A. Cisneros, M. Karttunen, P. Ren, C. Sagui, *Chem. Rev.* **2014**, *114*, 779–814.
- [128] Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J.-P. Piquemal, P. Ren, *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.
- [129] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, T. A. Darden, *J. Chem. Phys.* **2006**, *124*, 104101.
- [130] A. Savelyev, A. D. MacKerell, *J. Phys. Chem. B* **2014**, *118*, 6742–6757.
- [131] C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, P. Ren, *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.
- [132] J. Huang, J. A. Lemkul, P. K. Eastman, A. D. MacKerell, *J. Comput. Chem.* **2018**, *39*, 1682–1689.
- [133] M. Harger, D. Li, Z. Wang, K. Dalby, L. Lagardère, J.-P. Piquemal, J. Ponder, P. Ren, *J. Comput. Chem.* **2017**, *38*, 2047–2055.
- [134] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, et al., *J. Chem. Theory Comput.* **2013**, *9*, 461–469.
- [135] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, et al., *Chem. Sci.* **2018**, *9*, 956–972.
- [136] D. R. Bell, S. Y. Cheng, H. Salazar, P. Ren, *Sci. Rep.* **2017**, *7*, 45812.
- [137] M. A. Jonikas, R. J. Radmer, A. Laederach, R. Das, S. Pearlman, D. Herschlag, R. B. Altman, *RNA* **2009**, *15*, 189–199.
- [138] N. A. Denesyuk, D. Thirumalai, *J. Phys. Chem. B* **2013**, *117*, 4901–4911.
- [139] Y.-Z. Shi, F.-H. Wang, Y.-Y. Wu, Z.-J. Tan, *J. Chem. Phys.* **2014**, *141*, 105102.
- [140] M. J. Boniecki, G. Lach, W. K. Dawson, K. Tomala, P. Lukasz, T. Soltysinski, K. M. Rother, J. M. Bujnicki, *Nucleic Acids Res.* **2016**, *44*, e63.
- [141] S. Pasquali, P. Derreumaux, *J. Phys. Chem. B* **2010**, *114*, 11957–11966.
- [142] J. J. Uusitalo, H. I. Ingólfsson, S. J. Marrink, I. Faustino, *Biophys. J.* **2017**, *113*, 246–256.
- [143] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, et al., *Chem. Rev.* **2018**, *118*, 4177–4338.
- [144] K. M. Shannon, *CNS Drugs* **2020**, DOI 10.1007/s40263-019-00695-3.
- [145] A. Kumar, V. Kumar, K. Singh, S. Kumar, Y.-S. Kim, Y.-M. Lee, J.-J. Kim, *Brain Sci.* **2020**, *10*, DOI 10.3390/brain-sci10010043.
- [146] A. Beister, P. Kraus, W. Kuhn, M. Dose, A. Weindl, M. Gerlach, *J. Neural Transm. Suppl.* **2004**, 117–122.
- [147] R. M. Bonelli, A. K. Hödl, P. Hofmann, H.-P. Kapfhammer, *Int. Clin. Psychopharmacol.* **2004**, *19*, 337–342.
- [148] C. Carmo, L. Naia, C. Lopes, A. C. Rego, *Adv. Exp. Med. Biol.* **2018**, *1049*, 59–83.
- [149] V. M. Gohil, N. Offner, J. A. Walker, S. A. Sheth, E. Fossale, J. F. Gusella, M. E. MacDonald, C. Neri, V. K. Mootha, *Hum. Mol. Genet.* **2011**, *20*, 294–300.
- [150] Z. Mao, Y. S. Choo, M. Lesort, *Eur. J. Neurosci.* **2006**, *23*, 1701–1710.

- [151] G. Gardian, S. E. Browne, D.-K. Choi, P. Klivenyi, J. Gregorio, J. K. Kubilus, H. Ryu, B. Langley, R. R. Ratan, R. J. Ferrante, et al., *J. Biol. Chem.* **2005**, *280*, 556–563.
- [152] E. Hockly, V. M. Richon, B. Woodman, D. L. Smith, X. Zhou, E. Rosa, K. Sathasivam, S. Ghazi-Noori, A. Mahal, P. A. S. Lowden, et al., *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2041–2046.
- [153] A. Yamamoto, J. J. Lucas, R. Hen, *Cell* **2000**, *101*, 57–66.
- [154] S. Q. Harper, P. D. Staber, X. He, S. L. Eliason, I. H. Martins, Q. Mao, L. Yang, R. M. Kotin, H. L. Paulson, B. L. Davidson, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5820–5825.
- [155] M. DiFiglia, M. Sena-Esteves, K. Chase, E. Sapp, E. Pfister, M. Sass, J. Yoder, P. Reeves, R. K. Pandey, K. G. Rajeev, et al., *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17204–17209.
- [156] D. Yu, H. Pendergraft, J. Liu, H. B. Kordasiewicz, D. W. Cleveland, E. E. Swayze, W. F. Lima, S. T. Crooke, T. P. Prakash, D. R. Corey, *Cell* **2012**, *150*, 895–908.
- [157] B. L. Davidson, A. M. Monteys, *Cell* **2012**, *150*, 873–875.
- [158] L. M. Stanek, S. P. Sardi, B. Mastis, A. R. Richards, C. M. Treleaven, T. Taksir, K. Misra, S. H. Cheng, L. S. Shihabuddin, *Hum. Gene Ther.* **2014**, *25*, 461–474.
- [159] X. Dong, S. Cong, *Front. Mol. Neurosci.* **2019**, *12*, 156.
- [160] L. Mégret, S. S. Nair, J. Dancourt, J. Aaronson, J. Rosinski, C. Neri, *BioRxiv* **2020**, DOI 10.1101/2020.01.24.918540.
- [161] R. Roshan, T. Ghosh, V. Scaria, B. Pillai, *Drug Discovery Today* **2009**, *14*, 1123–1129.
- [162] C. Rinaldi, M. J. A. Wood, *Nat. Rev. Neurosci.* **2018**, *14*, 9–21.
- [163] R. S. Geary, D. Norris, R. Yu, C. F. Bennett, *Adv. Drug Delivery Rev.* **2015**, *87*, 46–51.
- [164] S. Finkbeiner, *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, DOI 10.1101/cshperspect.a007476.
- [165] C. Buren, L. Wang, A. Smith-Dijak, L. A. Raymond, *J. Huntingtons Dis* **2014**, *3*, 365–376.
- [166] J. B. Carroll, S. C. Warby, A. L. Southwell, C. N. Doty, S. Greenlee, N. Skotte, G. Hung, C. F. Bennett, S. M. Freier, M. R. Hayden, *Mol. Ther.* **2011**, *19*, 2178–2185.
- [167] E. J. Wild, S. J. Tabrizi, *Lancet Neurol.* **2017**, *16*, 837–847.
- [168] F. B. Rodrigues, E. J. Wild, *J. Huntingtons Dis* **2018**, *7*, 89–98.
- [169] H. Vossfeldt, M. Butzlaff, K. Prüssing, R.-A. Ni Chárthaigh, P. Karsten, A. Lankes, S. Hamm, M. Simons, B. Adryan, J. B. Schulz, et al., *PLoS One* **2012**, *7*, e47452.

Manuscript received: March 6, 2020

Revised manuscript received: May 22, 2020

Version of record online: June 16, 2020