SUPPLEMENTARY DATA

Single-stranded nucleic acids promote SAMHD1 complex formation

Victoria Tüngler [1,7], Wolfgang Staroske [2,7], Barbara Kind [1], Manuela Dobrick [1], Stefanie Kretschmer [1], Franziska Schmidt [1], Claudia Krug [1], Mike Lorenz [3], Osvaldo Chara [4,5], Petra Schwille [2,6], Min Ae Lee-Kirsch [1*]

[1] Children´s Hospital, Technical University Dresden, 01307 Dresden, Germany
[2] Biotechnology Center, Technical University Dresden, 01307 Dresden, Germany
[3] Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[4] Center for Information Services and High Performance Computing, Technical University Dresden, 01062 Dresden, Germany
[5] Institute of Physics of Liquids and Biological Systems (IFLYSIB), CONICET, National University of La Plata, B1900BTE La Plata, Argentina
[6] Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

* Corresponding author: Min Ae Lee-Kirsch, phone: +49-351-458 6887, fax: +49-351-458 6333, email: minae.lee-kirsch@uniklinikum-dresden.de.

[7] Joint first authorship

Supplementary Methods

FCCS data acquisition

Fluorescence cross-correlation spectroscopy (FCCS) and laser scanning microscopy were carried out on a commercial confocal microscope Zeiss LSM780 with an attached Confocor 3. The 488 nm Ar-laser-line was used to excite EGFP and the 561 nm line was used to excite 6-TAMRA / mCherry. Both laser lines were attenuated by an acousto-optical tunable filter to an intensity in the focal plane of 1.6 kW/cm² (GFP; FCCS with oligonucleotides and intracellular), 4.1 kW/cm² (GFP: FCCS with mCherry-SAMHD1 in Lysates), 0.71 kW/cm² (6-TAMRA), 7.5 kW/cm² (mCherry; FCCS in lysates) or 2.7 kW/cm² (mCherry; intracellular), respectively. Both excitation laser lines were directed by a 488 / 561 dichroic mirror (MBS) onto the back aperture of a Zeiss C-Apochromat 40x, N.A. = 1.2, water immersion objective. The fluorescence light was collected by the same objective, separated from the excitation light by the MBS, passing a confocal pinhole (35 µm in diameter) and split into two spectral channels by a second dichroic (NFT, LP565). After removing residual laser light by a 495-555 nm bandpass and 580 nm longpass emission filter, respectively, the fluorescence light was recorded by avalanche photodiode detectors (APDs). Before each experiment the setup was adjusted using a dye mixture of Alexa488 and CF568 and a 198 bp long Alexa488 / Atto565 double labelled DNA, yielding cross-correlation amplitudes of 80 % (± 5 %). The discrepancy to 100 % is due to imperfect overlap of the detection volumes and/ or imperfect labelling of the DNA (1;2;3;4). Therefore, even in presence of a perfect interaction, cross-correlation values would not reach 100 %.

For each lysate sample, FCCS was performed 20 µm above the cover slip in the sample solution. For each measurement, 12 runs, each 10 seconds long, or 24 runs each 5 seconds long were collected.

For intracellular FCCS, a confocal image was acquired and the FCCS focus was positioned inside the nucleus approximately 3 µm above the cover slip. For each measurement, 24 runs, each 10 seconds long were collected. Cells were kept in air-buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 15 mM glucose, 150 µg/ml BSA, 20 mM trehalose, 5.4 mM KCl, 0.85 mM MgSO4, 0.6 mM $CaCl_2$) at room temperature during measurement (5). Confocal imaging was performed using the 32 channel GaAsP detector of the Zeiss LSM780 in photon counting mode.

FCCS data analysis

The fluorescence signals of each run were software correlated by the Zeiss ZEN Software following the definition of auto- and cross-correlation (1;2;3)

$$G_{ij}(\tau) = \frac{\langle \delta F_i(t) \cdot \delta F_j(t+\tau) \rangle}{\langle F_i(t) \rangle \langle F_j(t) \rangle}$$.

Runs showing diffusion of remaining cellular compartments were discarded from the data evaluation.

A model including two diffusing species and a term accounting for the photophysics was fitted to the two auto-correlation and the cross-correlation curves, using a weighted Marquardt non-linear least square fitting algorithm. The cross-correlation curve was the average of the two cross-correlation curves (red vs. green and green vs. red channel).

$$G_{ij}(\tau) = G_{ij}(0) \cdot \left(1 + \frac{T}{1-T}\right) \cdot e^{\left(-\tau/\tau_T\right)} \cdot \left(\frac{F}{\left(1+\frac{\tau}{\tau_1}\right)\sqrt{1+\frac{\tau}{S^2 \cdot \tau_1}}} + \frac{1-F}{\left(1+\frac{\tau}{\tau_2}\right)\sqrt{1+\frac{\tau}{S^2 \cdot \tau_2}}} \right)$$

T represents the fraction of molecules in the dark state, τ_T the lifetime of the dark state, τ_1 and τ_2 are the diffusion times of the two species, F is the fraction of the two species and S is the form factor, which is the ratio of axial (ω_z) over the radial radius (ω_{xy}). During the fit the triplet fraction T of the cross-correlation curve was fixed to zero.

The amplitudes $G_{ij}(0)$ were corrected for spurious autofluorescent background of the cells, using the following algorithm:

$$\widetilde{G}_{ij}(0) = G_{ij}(0) \frac{F_i}{F_i - B_i} \cdot \frac{F_j}{F_j - B_j}$$

Here, F is the measured count rate and B the measured background count rate, which was measured in cell lysates from HEK293 cells transfected with an empty vector. The background in the green channel was between 1 and 2 % of the measured fluorescent signal and for the red channel below 1 %. In addition, background corrected amplitudes, $\widetilde{G}_{ij}(0)$, were corrected for spectral crosstalk. Spectral crosstalk was only encountered from the green into the red channel, with an amount of β = 8.8 %. Subsequently, the green amplitude is unaffected, while the red and the cross-correlation amplitude need to be corrected for spectral crosstalk:

$$\widehat{G}_g(0) = G_g(0)$$

$$\widehat{G}_r(0) = \frac{F_r^2 G_r(0) + \beta^2 F_g^2 G_g(0) - 2\beta F_r F_g G_{CC}(0)}{\left(F_r - \beta F_g\right)^2}$$

$$\widehat{G}_{CC}(0) = \frac{F_r F_g G_{CC}(0) - \beta F_g^2 G_g(0)}{\left(F_g F_r - \beta F_g\right)^2}$$

From the corrected amplitudes, the number of fluorescent particles in the detection volume was calculated, which is given by the following set of formulas:

$$\hat{G}_g(0) = \frac{1}{N_g + N_{gr}}; \hat{G}_r(0) = \frac{1}{N_r + N_{gr}}; \hat{G}_{CC}(0) = \frac{N_{gr}}{\left(N_g + N_{gr}\right)\left(N_r + N_{gr}\right)}$$

Here, N_g is the number of only green labelled particles, N_r the number of only red labelled particles and N_{gr} the number of particles, which carry both labels. The amount of cross-correlation was calculated by dividing the number of double labelled particles by all particles carrying a green label:

$$CC = \frac{N_{gr}}{N_g + N_{gr}}$$

Correction of measured brightness for fluorescent particles containing a dark fraction

Several red fluorescent proteins (RFP) harbor a fraction of non-fluorescent molecules, also referred to as dark fraction, that may arise due to formation of nonfunctional chromophors during the maturation process (4;6). In FCCS analysis, these non-fluorescent molecules lead either to a reduced concentration value for monomeric molecules or to a reduced brightness value for oligomeric particles. If the amount of the dark fraction is known, it is possible to correct the measured brightness value for the influence of the dark fraction by calculating the brightness, which would be measured if all RFP molecules were fluorescent. The derivation of the correction formula is outlined as follows:

The autocorrelation function for multiple diffusing species can be written as the sum of the single autocorrelation function weighted by the squared brightness weighing factor Λ (7).

$$G(\tau) = \sum_{i=1}^{n} \Lambda_i^2 \cdot G_i(\tau)$$

with

$$\Lambda_i = \frac{\eta_i \cdot N_i}{\sum_{i=1}^{n} \eta_i \cdot N_i}.$$

Where η_i is the molecular brightness and N_i is the number of particles of species i. The molecular brightness is the product of absorptivity, fluorescence quantum efficiency and experimental fluorescence collection efficiency. Therefore, the product $\eta_i \cdot N_i$ is the fluorescence signal S_i of all molecules from species i, and the sum over all species is the total measured signal $S_i = \sum_{i=1}^{n} \eta_i \cdot N_i$.

Putting this together and using the fact that the amplitude at $\tau = 0$ is $G(\tau = 0) = 1/N$ gives the following relations

$$G(0) = \sum_{i=1}^{n} \frac{\eta_i^2 \cdot N_i^2}{S^2} G_i(0) = \sum_{i=1}^{n} \frac{\eta_i^2 \cdot N_i^2}{S^2} \frac{1}{N_i} = \sum_{i=1}^{n} \frac{\eta_i^2 \cdot N_i}{S^2} = \frac{1}{N}.$$

N is the average number of all particles in the focal volume. The average molecular brightness of all particles is calculated by dividing the total fluorescence signal S by the average number of particles N, which leads to the following formula:

$$\eta_{aver} = \frac{S}{N} = S \cdot G(0) = S \cdot \sum_{i=1}^{n} \frac{\eta_i^2 \cdot N_i}{S^2} = \frac{\sum_{i=1}^{n} \eta_i^2 \cdot N_i}{\sum_{i=1}^{n} \eta_i \cdot N_i}$$

For the case of a protein which oligomerizes, the different species i are now considered as different oligomerization states, where the index i can be used to indicate the number of monomers, which are forming that oligomerization state.

If not all of these monomers are fluorescent, a dark fraction of this fluorescent protein can be assumed. In the case, that the fluorescent protein has such a dark fraction df, each single higher-ordered oligomerization state $N_i, i \geq 2$ has a different brightness. The probability that an oligomer with the size i has the brightness η_i with $j \leq i$ is given by the binomial distribution

$$p(j \mid i) = \binom{i}{j} \cdot df^{i-j} \cdot (1 - df)^{j}$$

A mixture of different oligomerization states may contain oligomers, which have the same brightness, but are different in size. For example, a mixture of trimers and tetramers contains

a fraction of trimers with two fluorescent monomers and a fraction of tetramers with also two fluorescent monomers. Then, both fractions have the same brightness, but different sizes.

Now the number of particles, having all the same brightness η_i, is calculated by the following sum

$$N_j = \sum_{i=1}^{n} p(i \mid j) \cdot N_i \ .$$

With the definition of the probability p_i for the oligomerization state i by $p_i = {N_i}/{N} = N_i / \sum_{i=1}^{n} N_i$ and under the assumption that the brightness of j monomers is j times the brightness of a single monomer $\eta_1; \eta_j = j \cdot \eta_1$ it is now possible to give a formula for the average molecular brightness of all particles

$$\eta_{aver} = \frac{\sum_{j=1}^{n} \eta_j^2 \cdot N_j}{\sum_{j=1}^{n} \eta_j \cdot N_j} = \frac{\sum_{j=1}^{n} \eta_j^2 \cdot \sum_{i=1}^{n} p(j \mid i) \cdot p_i \cdot N}{\sum_{j=1}^{n} \eta_j \cdot \sum_{i=1}^{n} p(j \mid i) \cdot p_i \cdot N} = \eta_1 \frac{\sum_{i=1}^{n} \sum_{j=1}^{i} j^2 \cdot p(j \mid i) \cdot p_i}{\sum_{i=1}^{n} \sum_{j=1}^{i} j \cdot p(j \mid i) \cdot p_i} .$$

With the definitions and values of the expected value and variance of the binomial distribution

expected value: $\bar{j} = \sum_{j=1}^{i} j \cdot p(j \mid i) = i \cdot (1 - df)$

variance: $\sigma_j^2 = \overline{j^2} - \bar{j}^2 = i \cdot df \cdot (1 - df)$

$$\sum_{j=1}^{i} j^2 \cdot p(j \mid i) = \bar{j}^2 = \sigma_j^2 + \overline{j^2} = i \cdot df \cdot (1 - df) + i^2 \cdot (1 - df)^2$$

follows

$$\eta_{aver} = \eta_1 \frac{\sum_{i=1}^{n} \left(i \cdot df \cdot (1 - df) + i^2 \cdot (1 - df)^2 \right) \cdot p_i}{\sum_{i=1}^{n} i \cdot (1 - df) \cdot p_i} = \eta_1 \frac{df \cdot \sum_{i=1}^{n} i \cdot p_i + (1 - df) \cdot \sum_{i=1}^{n} i^2 \cdot p_i}{\sum_{i=1}^{n} i \cdot p_i}$$

$$\eta_{aver} = \eta_1 \frac{df \cdot \bar{n} + (1 - df) \cdot \overline{n^2}}{\bar{n}} = \eta_1 \left(df + (1 - df) \frac{\overline{n^2}}{\bar{n}} \right)$$

To calculate the correction of the average molecular brightness to the expected values, if the fluorescent molecules would not have a dark fraction, the formula above is used for $df = 0$

$$\eta_{aver,0} = \eta_{aver}(df = 0) = \eta_1 \frac{\overline{n^2}}{\overline{n}} \quad \Rightarrow \quad \frac{\overline{n^2}}{\overline{n}} = \frac{\eta_{aver,0}}{\eta_1}$$

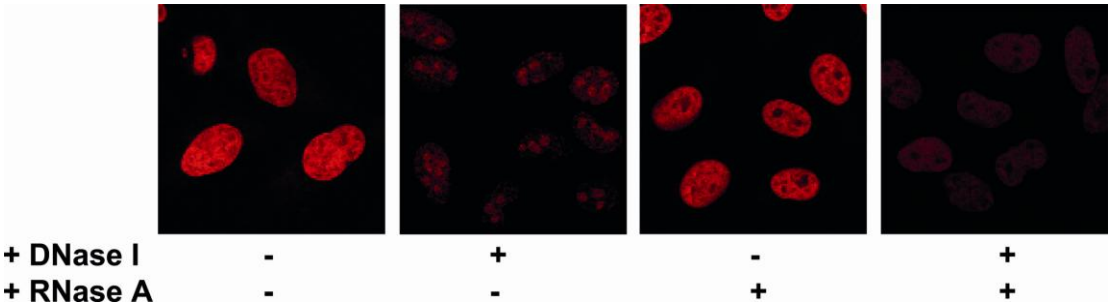
and the result is put back in the formula

$$\frac{\eta_{aver}}{\eta_1} = df + (1 - df)\frac{\eta_{aver,0}}{\eta_1} \quad \Rightarrow \quad \frac{\eta_{aver,0}}{\eta_1} = \frac{1}{1 - df} \cdot \left(\frac{\eta_{aver}}{\eta_1} - df \right).$$

Supplementary Table S1. Sequences of oligonucleotides used for FCCS and pull-down assay

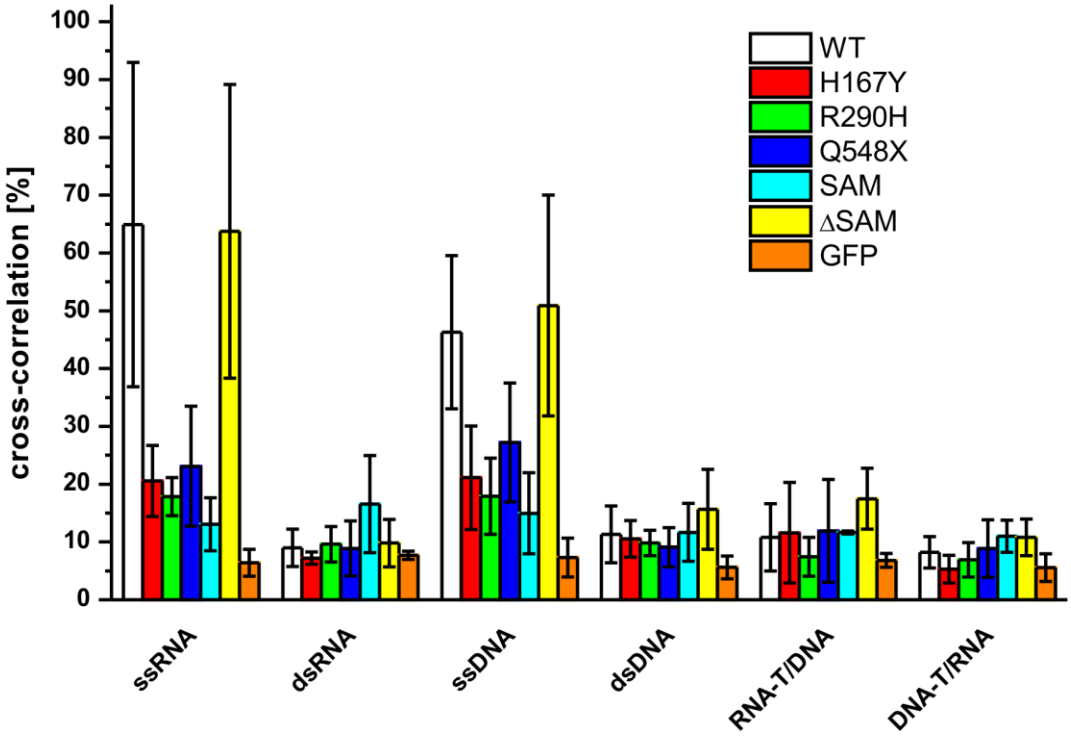
Name	Sequence (5´-3´)	Comments
ssRNA sense	GCA UGC GAC CUC UGU UUG A-(6-TAMRA or biotin)	6-TAMRA label used for FCCS; biotin label used for pull-down
ssRNA antisense	UCA AAC AGA GGU CGC AUG C	used for generation of dsRNA and RNA:DNA hybrid with labeled ssDNA
ssDNA sense	GCA TGC GAC CTC TGT TTG A-(6-TAMRA or biotin)	
ssDNA antisense	TCA AAC AGA GGT CGC ATG C	used for generation of dsDNA and RNA:DNA hybrid with labeled ssRNA

Supplementary Figure S1. Digestion of endogenous nucleic acids with DNase I or RNase A



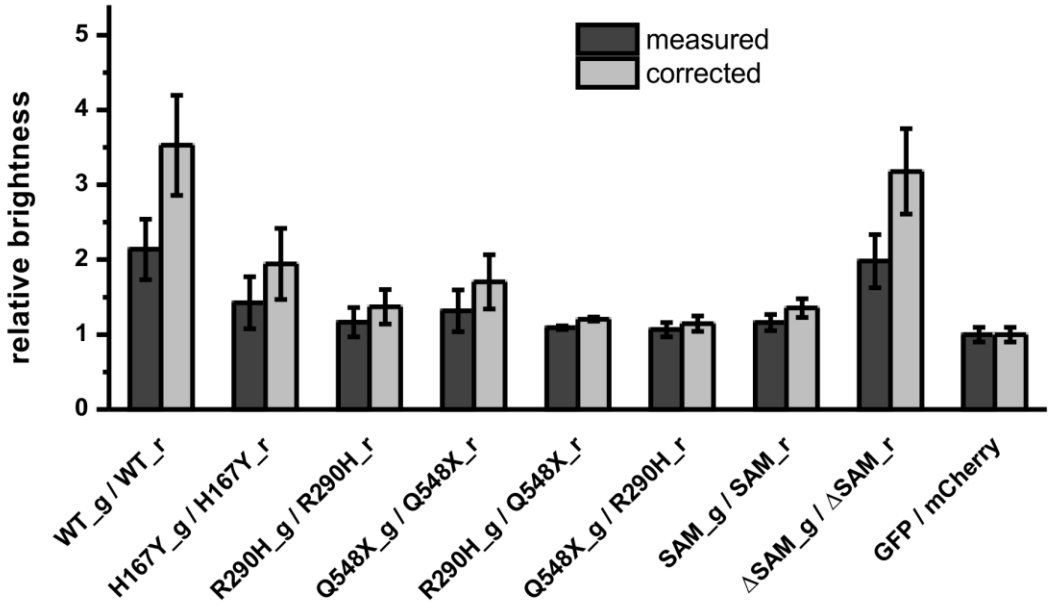
Supplementary Figure S1. Visualisation of nucleic acid digestion by DNAse I and RNase A in HeLa cells after staining with Sytox Orange as used for FLIM experiments. DNase I leads to a strong decrease of Sytox Orange fluorescence within the hetero- and euchromatin regions of the nucleus, while the nucleoli, which are the sites of rRNA transcription and rRNA assembly, appear as prominent red fluorescent dots. Following RNase A treatment of cells, Sytox Orange staining of nucleoli is lost, while the chromatin fluorescence remains unaltered. The combined treatment of cells with DNase I and RNase A leads to a strong reduction of Sytox Orange fluorescence across the nucleus.

Supplementary Figure S2. Cross-correlation of wild type and mutant GFP-SAMHD1 with double-stranded nucleic acids



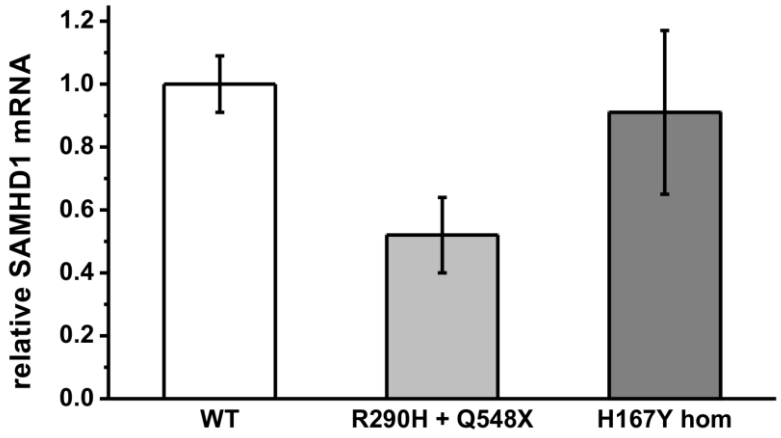
Supplementary Figure S2. Association of nucleic acid species with SAMHD1 visualized by FCCS. Summary of cross-correlation data of GFP-SAMHD1 and GFP-tagged mutants with different single-stranded (ssRNA, ssDNA) and double-stranded (dsRNA, dsDNA, RNA:DNA hybrids) oligonucleotides. Two types of RNA:DNA hybrids in which either the RNA-strand (RNA-T/DNA) or the DNA-strand (DNA-T/RNA) was fluorescently labelled, were examined.

Supplementary Figure S3. Brightness of mCherry-SAMHD1



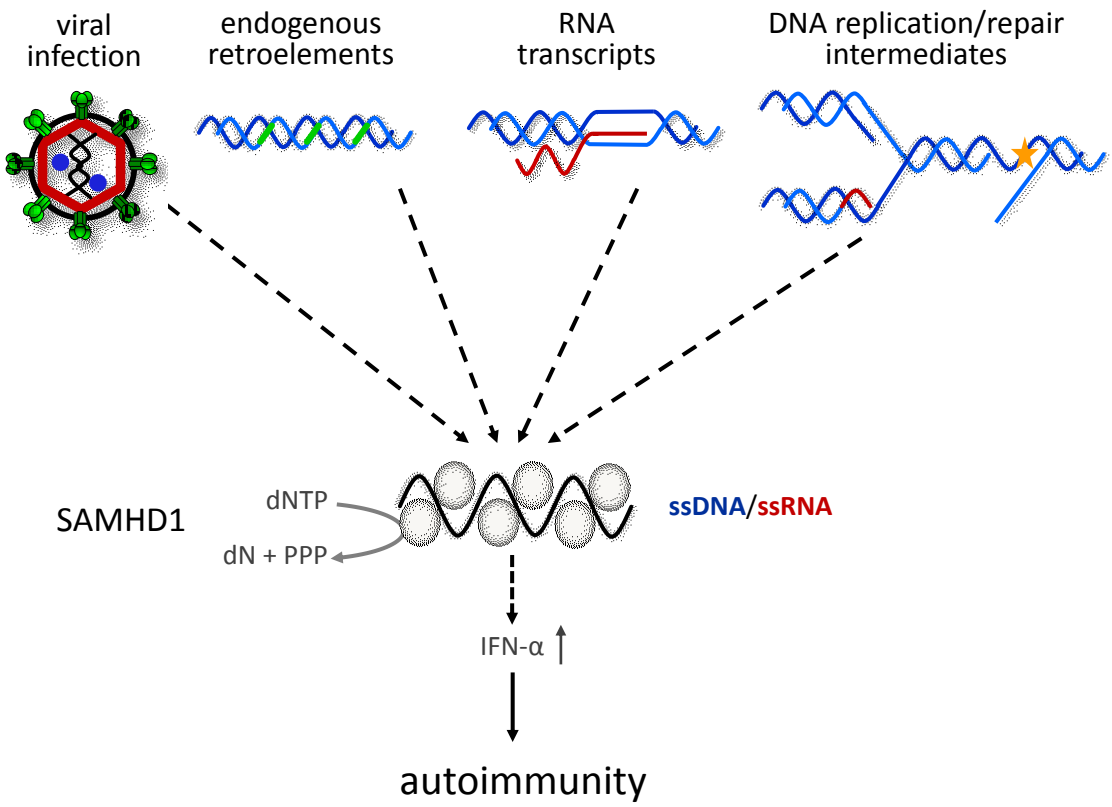
Supplementary Figure S3. Summary of cross-correlation data in cell lysates containing GFP-SAMHD1 and mCherry-SAMHD1 showing the measured and corrected brightness values for mCherry-SAMHD1 assuming a dark fraction of 55 % for mCherry.

Supplementary Figure S4. Expression of SAMHD1 mRNA in AGS patients



Supplementary Figure S4. SAMHD1 expression was determined by quantitative real-time RT-PCR of total RNA extracted from lymphoblastoid cells using Taqman Universal PCR Master Mix (Applied Biosystems) and the following oligonucleotides (SAMHD1-F 5'-GGTCATGGGCCATTTTCTCACA-3', SAMHD1-R 5'-GCCTTGTTCATGCGTCCATTTC-3', SAMHD1-Probe 5'FAM-ACGATTTATTCCACTTGCTCGCCCGGA-3'TAMRA). SAMHD1 gene expression was normalized to GAPDH mRNA using the following oligonucleotides (GAPDH-F 5'-GAAGGTGAAGGTCGGAGTC-3', GAPDH-R 5'-GAAGATGGTGATGGGATTTC-3', GAPDH-Probe FAM5'-CAAGCTTCCCGTTCTCAGCC-3'TAMRA). SAMHD1 mRNA expression in the patient homozygous for H167Y is similar to wild type cells, while it is reduced by 50 % in cells from the patient compound heterozygous for R290H and Q548X suggesting nonsense-mediated decay of the mRNA transcribed from the Q548X allele.

Supplementary Figure S5. Hypothetical model of SAMHD1 complex formation



Supplementary Figure S5. Hypothetical model of SAMHD1 dimers forming complexes of higher order along single-stranded nucleic acids. Putative SAMHD1 substrates may be viral nucleic acids or endogenous single-stranded nucleic acids originating from retroelements, transcription or from DNA replication and repair. SAMHD1 degrades deoxynucleoside triphosphates (dNTP), the building blocks of DNA synthesis, into the constituent deoxynucleoside (dN) and inorganic triphosphate (PPP). Loss of SAMHD1 activity could therefore contribute to the synthesis of nucleic acid species that could trigger an autoimmune response via activation of interferon-α (IFN-α). Dashed lines indicate unknown pathways.

Supplementary References

1. Schwille,P., Meyer-Almes,F.J. and Rigler,R. (1997) Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. Biophys. J., 72, 1878-1886.

2. Bacia,K., Kim,S.A. and Schwille,P. (2006) Fluorescence cross-correlation spectroscopy in living cells. Nat. Methods, 3, 83-89.

3. Bacia, K., Petrášek, Z. & Schwille, P. (2012) Correcting for Spectral Cross-Talk in Dual-Color Fluorescence Cross-Correlation Spectroscopy. Chemphyschem, *13*, 1221-1231.

4. Foo,Y.H., Naredi-Rainer,N., Lamb,D.C., Ahmed,S. and Wohland,T. (2012) Factors affecting the quantification of biomolecular interactions by fluorescence cross-correlation spectroscopy. Biophys. J., 102, 1174-1183.

5. Ohrt T., Staroske W., Muetze J., Crell K., Landthaler M., Schwille P. (2011) Fluorescence cross-correlation spectroscopy reveals mechanistic insights into the effect of 2'-O-methyl modified siRNAs in living cells. Biophys. J. 100, 2981-2990.

6. Padilla-Parra, S., Audugé, N., Lalucque, H., Mevel, J.-C., Coppey-Moisan, M. and Tramier, M. (2009) Quantitative Comparison of Different Fluorescent Protein Couples for fast FRET-FLIM Acquisition. Biophys. J., 2009, 97, 2368-2376.

7. Thompson, N. L. (1991) Fluorescence Correlation Spectroscopy, Topics in Fluorescence Spectroscopy, Volume 1: Techniques, Lakowicz, J. R. (Ed.), Plenum Press, New York, 1991, 337-378.