



- (51) International Patent Classification:
A61K 38/00 (2006.01) *C07K 14/47* (2006.01)
- (21) International Application Number:
PCT/DK2016/050317
- (22) International Filing Date:
30 September 2016 (30.09.2016)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
PA2015 00605 2 October 2015 (02.10.2015) DK
62/324,257 18 April 2016 (18.04.2016) US
- (71) Applicants: UNIVERSITY OF COPENHAGEN
[DK/DK]; Nørregade 10, 1165 Copenhagen K (DK). ME-
MORIAL SLOAN-KETTERING CANCER CENTER
[US/US]; 1275 York Avenue, New York, New York 10065
(US).
- (72) Inventors: GROTH, Anja; c/o BRIC - University of
Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N

(DK). SAREDI, Giulia; c/o BRIC - University of Copen-
hagen, Ole Maaløes Vej 5, 2200 Copenhagen N (DK).
HUANG, Hongda; c/o - Memorial Sloan-Kettering Cancer
Center, 1275 York Avenue, New York, New York 10065
(US). HAMMOND, Colin; c/o BRIC - University of
Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N
(DK). PATEL, Dinshaw; c/o - Memorial Sloan-Kettering
Cancer Center, 1275 York Avenue, New York, New York
10065 (US).

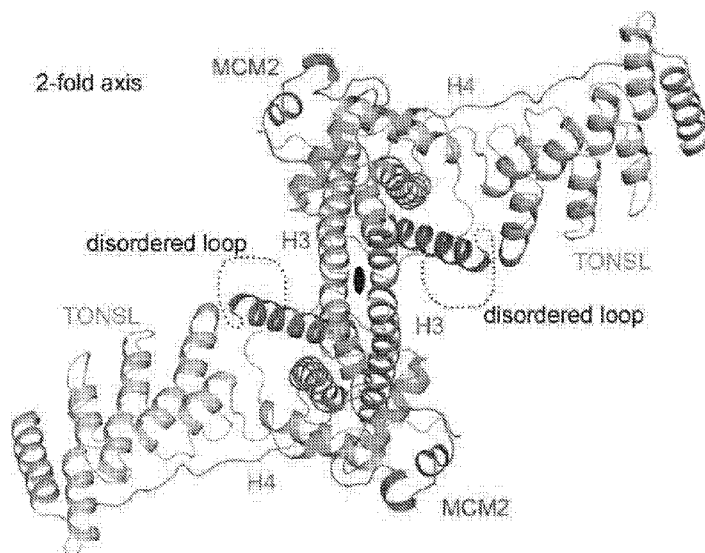
(74) Agent: HØIBERG P/S; Adelgade 12, 1304 Copenhagen
K (DK).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,

[Continued on next page]

(54) Title: SMALL MOLECULES BLOCKING HISTONE READER DOMAINS

Figure 1



(57) Abstract: The present invention relates to small molecules interfering with the conformational space of the TONSL ARD occupied by the histone H4 tail. These small molecules targets the binding pocket of TONSL encompassing the H4 residues K12- R23 and act by preventing or disrupting the binding of the H4 tail K12- R23 with the TONSL ARD via direct competition or via allosteric disruption of the binding pocket. The present inventors have identified and solved the structure of a histone reader domain of TONSL termed the ARD (ankyrin repeat domain).



TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
— with sequence listing part of description (Rule 5.2(a))

SMALL MOLECULES BLOCKING HISTONE READER DOMAINS

FIELD OF THE INVENTION

The present invention relates to novel inhibitors of the TONSL protein required for genome stability. TONSL functions as an obligate heterodimer with MMS22L in DNA repair of double strand breaks (DSBs) and stabilization of replicating DNA intermediates via error-free homologous recombination (HR).

The present inventors have solved the structure of the TONSL ARD in complex with its histone substrate, histone H4 tails, providing grounds for structure-based rational drug design of small molecules, peptides or polypeptides capable of preventing or disrupting the binding of the H4 tail with the TONSL ARD via direct competition or via allosteric disruption of the binding pocket.

BACKGROUND OF THE INVENTION

Cells are continuously exposed to diverse sources of DNA damage, and to face this challenge they have devised an array of DNA repair strategies. One of the key repair pathways for DNA double strand breaks, damaged replication forks and key for survival of proliferating cells is homologous recombination (HR).

Many cancer cells experience high load of replication stress, making them vulnerable to HR inhibition.

Cancer cells often also depend on this pathway to repair DNA damage generated by conventional chemotherapy; therefore, HR targeting drugs are clinically attractive due to synthetic lethality of affected cells.

Development of HR inhibitors is also of great clinical interest because PARP inhibitors kill HR defective cells. Therefore blocking two alternative pathways will deliver a much bigger impact than targeting either pathway alone. Additionally, some tumours cells are defective in other DNA repair pathways and therefore dependent on HR, making them sensitive to inhibition of HR repair.

TONSL (Tonsoku-like, NFKBIL2) is a crucial HR regulator that functions in a heterodimeric complex with MMS22L to promote repair of DSBs and damaged replication forks by HR. In HR, TONSL-MMS22L is proposed to regulate Rad51 loading

onto ssDNA. Depletion of MMS22L or TONSL results in a pronounced decrease of cell proliferation and marked hypersensitivity to the topoisomerase I poison camptothecin (CPT), which is most likely caused by an inability to promote RAD51-mediated repair of broken replication forks. However, the molecular mechanisms regulating TONSL-
5 MMS22L function in DNA repair are not known.

TONSL-MMS22L forms a complex with histones H3-H4, the histone chaperone ASF1 and MCM2, 4, 6, 7 that depend on the TONSL Ankyrin Repeat Domain (ARD), suggesting that the ARD could be important for TONSL function.

10

ARD is a common protein-protein interaction motif consisting of tandemly repeated modules of about 33 amino acids, occurring in a large number of functionally diverse proteins including transcriptional initiators, cell cycle regulators, cytoskeletal proteins, ion transporters and signal transducers.

15

Crystal structures of several ARD proteins are solved, including G9a/GLP ARD that binds specifically histone H3 methylated at lysine 9 (Collins et al., Nat Struct Mol Cell Biol 15:245, 2008). Yet the specificity of most ARDs remain unknown.

20 By sequence similarity, TONSL ARD is highly similar to the BARD1 ARD. BARD1 is a well-characterized HR factor, obligate partner of BRCA1, and is required for most cellular and tumour-suppressor functions of BRCA1 (Westermarck et al., Mol Cell Biol 23:7926, 2003).

25 The crystal structure of BARD1 ARD is available (Fox et al., J Biol Chem 283:21179, 2008) but the binding specificity of the BARD1 ARD is unknown, and thus does not provide grounds for inhibitor design.

Therefore, it is desirable to identify target molecules bound by TONSL ARD and obtain
30 crystals of these protein complexes in order to solve their structures at atomic resolution. This provides basis to determine whether the binding is important for TONSL-MMS22L function, thus making it an attractive drug target. Further, high-resolution structures of binding pockets with their substrates can enable the skilled addressee in designing small molecules interfering with the binding of TONSL ARD to
35 its target molecule and subsequently verifying the effect in biological assays in a rapid and reproducible manner.

SUMMARY OF THE INVENTION

The present inventors have discovered a molecular mechanism for recruitment of TONSL and its partner MMS22L to post-replicative chromatin that opens an avenue to target the HR repair pathway in cancer treatment. This mechanism relies on recognition of histone H4 unmodified at K20 (H4K20me0) by the TONSL Ankyrin Repeat Domain (ARD) domain that functions as a novel histone reader domain.

The highlights of this the discovery work include

1. Structural and biochemical data identifying the TONSL ARD as the first histone reader specific for H4 tails unmethylated at K20 (H4K20me0). H4K20me0 is specific to newly synthesized histones incorporated during DNA replication, marking post-replicative chromatin until late G2/M when K20me1 is established.
2. Functional and structural data revealing that TONSL via ARD recognition of H4K20me0 binds soluble new histones in complex with MCM2 and ASF1, proving a means to deliver TONSL-MMS22L to replicating chromatin.
3. Functional data that TONSL binds nucleosomal histones in post-replicative chromatin via ARD recognition of H4K20me0 and demonstration that this is required for TONSL-MMS22L accumulation at damaged replication forks and DNA lesions.
4. Functional data that TONSL ARD mutant protein titrates MMS22L away from chromatin and is toxic to cells.
5. The TONSL ARD mutant proteins induce G2/M arrest, replication-associated DNA damage (53BP1 foci) and reduce viability in the presence and absence of camptothecin (CPT).
6. By removing TONSL from chromatin, the function of TONSL-MMS22L complex in DNA repair is disabled.
7. Inhibitors of TONSL, e.g inhibitors of TONSL capable of associating with the TONSL ARD.
7. Histone H4 peptides or functional homologues thereof could act as inhibitors of TONSL

Taken together, this shows that DNA replication leaves a mark, in the form of deposited new histones unmodified at H4K20, which are recognized by the TONSL-MMS22L HR repair complex to differentiate pre- and post-replicative chromatin.

Furthermore, this work provides a breakthrough in understanding how the post-replicative chromatin state is read by a key HR factor TONSL, which mediates loading of Rad51 at damaged replication forks and DSBs. Small molecule inhibitors targeting TONSL ARD are thus capable to disable efficient HR repair of DSBs and damaged
5 replication forks.

The present invention thus relates to small molecules interfering with the conformational space of the TONSL ARD occupied by the histone H4 tail. These small molecules targets the binding pocket of TONSL encompassing the H4 residues K12-
10 R23 and act by preventing or disrupting the binding of the H4 tail K12- R23 with the TONSL ARD via direct competition or via allosteric disruption of the binding pocket.

Therefore, in one aspect the present invention relates to small molecules that prevent or disrupt the binding of a substrate in the binding pocket of the BARD1 ARD (defined
15 by its high structural similarity to the H4 tail K12-R23 binding pocket in TONSL ARD) via direct competition or via allosteric disruption of the binding pocket.

DETAILED DESCRIPTION OF THE INVENTION

The present inventors have identified and solved the structure of a histone reader domain of TONSL termed the ARD (ankyrin repeat domain). Importantly, the present
20 inventors have solved the structure of the TONSL ARD in complex with its histone substrate, histone H4 tails, providing grounds for structure-based rational drug design of TONSL inhibitors. Given the requirement of TONSL for HR of DSBs and repair of damaged replication forks, TONSL inhibitors can be efficient in compromising HR and be used either alone or in combination with conventional chemotherapy in killing
25 cancer cells. Thus, any of the TONSL inhibitors described herein may be useful in the treatment of cancer.

The crystal structure of the complex, recombinant TONSL ARD, cell lines expressing GFP-TONSL (wild type and histone H4 binding mutants) of the present invention are
30 all instrumental to design and identify specific small molecule inhibitors that interfere with TONSL ARD recognition of H4K20me0.

Below the present inventors, also disclose the development of a pipeline of biochemical and cell-based assays useful for TONSL inhibitor development.

35

These data and information are indispensable for the design of novel inhibitors suitable as anti-cancer therapeutics. The invention thus provides the means for designing *in silico* and subsequently synthesizing and testing in assays also provided by the invention, small molecule inhibitors that are capable of inhibiting the binding
5 between TONSL and its histone substrate, thereby interfering with HR, in turn resulting in direct anti-tumour effects and/or increased efficacy of available cytostatic compounds.

The present inventors have identified the key determinant for binding of H4 tails to
10 TONSL using interfacial mutants. Furthermore, the present inventors have demonstrated that TONSL ARD mutants that disrupt the H4 tail binding sites are no longer recruited to post-replicative chromatin, damaged replication forks, or DNA lesions (e.g., DNA double strand breaks). The TONSL ARD mutant proteins induce G2/M arrest, replication-associated DNA damage (53BP1 foci) and reduce viability in
15 the presence and absence of camptothecin (CPT).

As described in more detail below, said TONSL ARD mutants may phenocopy the effect of inhibitors of TONSL, in particular the effect of inhibitors which inhibit binding of TONSL ARD to histone H4. Thus, the invention provides that such inhibitors are
20 useful for inducing G2/M arrest, replication-associated DNA damage (53BP1 foci) and/or reduce cell viability in the presence and/or absence of camptothecin (CPT). This render the inhibitors useful in the treatment of cancer. The inhibitor may be any of the inhibitors described herein.

The present inventors also note that several TONSL mutants in the ARD domain
25 (residues 512-692) have been identified in cancer tissues of multiple organs including lung, skin, stomach, large intestine, biliary tract, prostate, endometrium, ovary, pancreas, oesophagus, urinary tract and central nervous system (COSMIC, Catalogue of somatic mutations in cancer, <http://grch37-cancer.sanger.ac.uk/cosmic/search?q=TONSL>). The term TONSL ARD refers to amino
30 acids 512 to 692 of SEQ ID NO:16.

The present inventors have determined that the TONSL ARD is highly similar to the ARD of BARD1 for which a structure has been published, but since the binding specificity of BARD1 ARD is unknown, this again does not provide grounds for inhibitor
35 design.

However, the details of TONSL ARD binding to H4 tails predict that inhibitors that disrupt TONSL binding to histones may also bind the BARD1 ARD and disrupt interaction with its substrate (e.g. histones). This would be desirable and probably increase drug efficacy as BARD1 and TONSL operate in separate steps within the same DNA repair pathway (HR). The present inventors also note that an established tumour suppressor mutation in BARD1 found in breast cancer targets a key residue predicted to be required for histone binding on the basis of the TONSL ARD-H4 structure.

In one aspect, the present invention relates to designing and developing drugs, small molecule inhibitors that would prevent recruitment of TONSL, a HR protein, to sites of DNA damage. Because TONSL together with its partner protein MMS22L is required for HR, interfering with TONSL recruitment to DNA lesion will impair HR repair and kill cancer cells or cells dependent on HR repair pathway either alone or in combination treatment with conventional chemotherapy.

The crystal structure

The present invention relates to the use of structure-based drug design methods to identify compounds that interfere with TONSL and TONSL-MMS22L complex recruitment to DNA lesions and DNA replication forks.

The invention discloses the crystal structure of the TONSL ARD in complex with its histone substrate, histone H4 tails. The TONSL ARD targets the H4 tail spanning residues Lys12 to Arg23, primarily through intermolecular hydrogen-bonding, electrostatic and van der Waals interactions (Figures 1-9).

The invention relates to the atomic details of the TONSL ARD binding pocket architecture and the intermolecular interactions with the H4 tail peptide that account for recognition specificity between the H4 tail (K4-R23) and residues lining the TONSL binding pocket.

In a preferred embodiment, the present invention relates to a crystal structure having the atomic coordinates available in the PDB Protein Databank under the PDB ID 5JA4 as deposited 11 April 2016, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or having a structure in which the atomic coordinates vary by less than 3Å in any direction from those set out in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb).

The variation of the atomic coordinates of the present invention should preferably be less than 3 Å, such as less than 2,75 Å, such as less than 2,5 Å, such as less than 2,25 Å, such as less than 2,0 Å, such as less than 1,75 Å, such as less than 1,5 Å, such as less than 1,25 Å, such as less than 1,0 Å, such as less than 0,75 Å, such as less than 0,5 Å or such as less than 0,25 Å.

In one embodiment, the invention relates to at least part of the atomic co-ordinate data available in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or data derivable therefrom.

In one embodiment, the present invention relates to a crystal comprising at least part of the crystal structure of the TONSL ARD in complex with its histone substrate, histone H4 tails.

In another embodiment, the invention relates to the distances between the atomic co-ordinates in the crystal structure available in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or any variation thereof that defines the binding surface of TONSL interacting with the histone H4 tail residues Lys12-Arg23 (Figures 1-9).

The histone H4(K12 to R23) peptide is positioned within a channel on the surface of the ankyrin repeat domains of TONSL. This surface constitutes an acidic patch and contains shallow binding pockets for the His18 and Lys20 side chains of H4. Further, the side chain of H4 R17 is aligned by being sandwiched between the side chains of Tyr572 and Cys608. Hence, any inhibitor must target the two shallow pockets and be compatible with the electrostatic nature of the binding surface.

Methods of forming a crystal structure

In one embodiment, the invention relates to a three dimensional crystal of a complex between

- a polypeptide comprising:
 - TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue thereof sharing at least 90% sequence identity therewith; and

- optionally MCM2 HBD consisting of amino acids 61 to 130 of SEQ ID NO:24 or a functional homologue thereof sharing at least 90% sequence identity therewith;
- Histone H4 of SEQ ID NO:23 or a functional homologue thereof sharing at least 90% sequence identity therewith; and
- optionally Histone H3 or a fragment thereof, e.g. histone H3.3 Δ 56 consisting of amino acids 57–135 of SEQ ID NO:25 or a functional homologue thereof sharing at least 90% sequence identity therewith.

10 In particular, the polypeptide may comprise said TONSL ARD or a functional homologue thereof linked to said MCM2 HBD or a functional homologue thereof by way of a peptide linker. Such a peptide linker typically consists of in the range 4 to 20 amino acids, e.g. 4 to 12 amino acids. Said amino acids may be Gly, and thus the linker may consist of in the range of 4 to 20 Gly residues, such as in the range of 4 to 15 12 Gly-residues.

In general, the complex comprises one polypeptide comprising TONSL ARD, whereas histone H4 and optionally histone H3 typically are present as separate peptides. Preferably, the polypeptide comprising TONSL ARD also comprise MCM2 HBD, wherein TONSL ARD may be linked to MCM2 HBD by way of a linker, e.g. a peptide linker.

20 Said fragment of Histone H3 may be a fragment of histone H3.3, e.g. amino acids 57–135 of SEQ ID NO:25. This fragment may also be a fragment of Histone H3.1 or Histone H3.2, which contains an identical fragment.

The present invention may in particular be based on the use of a unique fusion protein, MCM2 HBD–G₄–TONSL ARD (SEQ ID NO: 15), allowing crystallization of the TONSL ARD–MCM2 HBD–H3–H4 tetramer complex. The fusion protein consists of the human TONSL Ankyrin Repeat Domain (ARD, residues 512–692 of SEQ ID NO:16) and MCM2 Histone-binding Domain (HBD, fragments 61–130 of SEQ ID NO:24) covalently linked through a four-Glycine linker (G₄ linker) into one expression cassette.

30 Useful isolated polynucleotide or amino acid sequences of the present invention is disclosed in the SEQUENCE DATA listing below having at least 90% sequence identity, such as 91% sequence identity, 92% sequence identity, 93% sequence identity, 94% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, 99% sequence identity or 100% sequence identity.

35

All sequence listed are embodiments of the present invention. There are many ways to define the sequence identity of genes, nucleotides, or protein sequences. One way is through a direct comparison of two aligned sequences. Such a comparison is technically straightforward.

5

In particular, the term "sequence identity" as used herein may refer to the % of identical amino acids or nucleotides between a candidate sequence and a reference sequence following alignment. Thus, a candidate sequence sharing 80% amino acid identity with a reference sequence requires that, following alignment, 80% of the amino acids in the candidate sequence are identical to the corresponding amino acids in the reference sequence. Identity according to the present invention is preferably determined by aid of computer analysis, such as, without limitations, the ClustalW computer alignment program (Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680), and the default parameters suggested therein. The ClustalW software is available from as a ClustalW WWW Service at the European Bioinformatics Institute <http://www.ebi.ac.uk/clustalw>. Using this program with its default settings, the candidate amino acid sequence and the reference amino acid sequence are aligned. The number of fully conserved residues are counted and divided by the length of the reference sequence. Thus, sequence identity is preferably determined over the entire length of the reference sequence. The ClustalW algorithm may similarly be used to align nucleotide sequences. Sequence identities may be calculated in a similar way as indicated for amino acid sequences. Sequence identity as provided herein is calculated over the entire length of the reference sequence.

Functional homologues of a reference peptide or polypeptide according to the invention may be peptides or polypeptides sharing at least 70%, such as at least 80%, for example at least 90% sequence identity, such as at least 95% sequence identity, for example at least 98% sequence identity with said reference peptide or polypeptide.

In another embodiment sequences that hybridize to any of the isolated polynucleotide of the present invention is disclosed in the SEQUENCE DATA listing below under conditions of low stringency, are embodiments of the present invention.

35

In the present context, sequence that hybridize under low stringency is defined as a sequence which specifically hybridizes to said e.g. DNA or said complementary sequence in a hybridization solution containing 0.5M sodium phosphate buffer, pH 7.2, containing 7% SDS, 1 mM EDTA at 65° C. for 16 hours and washing twice at 65° C.
5 for twenty minutes in a washing solution containing 0.5×SSC and 0.1% SDS.

Crystallization of TONSL ARD in complex with a H4 tail or H3–H4 tetramer failed even with extensive screening. Because an additional binding protein may help to stabilize the whole complex and help crystallization, crystallization of TONSL ARD in complex
10 with the MCM2 HBD and H3–H4 tetramer was tested.

Very tiny crystals were obtained for this complex, but failed to give big and well-diffracted crystals. The whole complex of TONSL ARD with MCM2 HBD and H3–H4 tetramer might be destabilized by the harsh crystallization conditions and form sub
15 complexes thus hindering the optimization of the crystals.

Covalently linkage of TONSL ARD and MCM2 HBD into one cassette was thus tested through different length of Glycine linker (G_x linker). The G_{12} , G_{11} , G_{10} , G_9 , G_8 , G_7 , G_6 , G_5 and G_4 linkers had been tried and all these cassettes could be crystallized.
20

One of the constructs with a G_4 linker (MCM2 HBD– G_4 –TONSL ARD cassette) gave well diffracted crystals in complex with H3.3($\Delta 56$)–H4 (see Example 1), herein denoted as TONSL ARD–MCM2 HBD–H3–H4 tetramer complex (Figure 1).

25 Thus, the present invention relates to a composition comprising a protein assembly of TONSL ARD–MCM2 HBD–H3–H4 in its crystalline form and/or the details of the structure of the complex deduced from structural analysis of this crystal.

Specific data defining this crystal are detailed in Table 1, Example 1. Specifically, this
30 aspect of the invention relates to the TONSL ARD binding pocket architecture and the intermolecular interactions with the H4 tail peptide deduced from the atomic co-ordinates that define the binding surface of TONSL ARD with the histone H4 tail residues Lys12–Arg23 (e.g. in the crystal structure available in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or variation thereof).

The binding pocket data

The Lys12-Gly13-Gly14-Ala15 segment of H4 is positioned within a narrow surface channel of the TONSL ARD scaffold (Figures 2-4). The amino acid number provided in relation to histone H4 are general provided in relation to histone H4 of SEQ ID NO:34.

5

The intermolecular contacts spanning the Lys12-Gly13-Gly14-Ala15 segment of H4 include hydrophobic interactions between residues Gly13, Gly14 and Ala15 of H4 and residues Asn507, Cys508, Trp641, Tyr645 and Leu649 of ARD, as well as hydrogen bonds between the main-chain O of H4 Gly14 and N ϵ 1 of ARD Trp641, and between

10

the main-chain N of H4 Ala15 and O η of ARD Tyr645 (Figure 3, 8).

The main-chain O of H4 Lys16 hydrogen bonds with the N δ 2 of ARD Asn571, while the side-chain of H4 Lys16 forms contacts with ARD Asn607 and electrostatic interactions with the side-chain of ARD Glu597 (Figure 3).

15

The side-chain of H4 Arg17 stacks over the side-chains of ARD Tyr572 and Cys608, while its N η 1 atom forms two hydrogen bonds with main-chain O and O δ 1 of ARD Asn571 (Figures 3, 5).

20

The side-chain of H4 H18 penetrates into a pocket lined by four strictly conserved residues (Trp563, Glu568, Asn571 and Asp604) and is positioned over His567 of ARD. The side chain of H4 His18 is stacked between Trp563 and Asn571 and forms hydrogen bonds to Glu568 and Asp604 of ARD (Figures 3, 6).

25

The main-chain O of H4 Arg19 forms a hydrogen bond with N ϵ 1 of Trp563 and its side-chain forms contacts with Cys561 and Gly595 of ARD (Figure 3).

The H4 Lys20 residue is bound within an acidic surface pocket on ARD adjacent to the H4 His18 binding pocket. The side-chain of H4 Lys20 interacts with the side-chain of Met528 and contacts the edge of Trp563 of ARD, while the main-chain atoms of H4 Lys20 packs against Cys561 of ARD. The N ζ atom of H4 Lys20 forms three strong hydrogen bonds (distance < 3 Å) with the side-chains of strictly conserved residues Glu530, Asp559 and Glu568 of ARD, which surround H4 Lys20 within a regular triangle-like alignment (Figures 3, 7).

35

The intermolecular contacts spanning the Val21-Leu22-Arg23 segment of H4 include contacts between side-chains of H4 Val21 with Tyr560 and Cys561 of ARD, while H4 Leu22 interacts with Asp527 and Met528 of ARD. The main-chain N of H4 Arg23 forms a hydrogen bond with the main-chain O of Asp527 of ARD, while the side-chain packs against the side-chain of Tyr560 of ARD (Figures 3, 9).

From the structure descriptions above, it is apparent that the H4 tail (residues 12-23) with an extended β -strand like conformation lies in an elongated channel on the concave surface of the TONSL ARD. This channel is primarily acidic, which provides an electrostatic complementary fit with the positively charged H4 tail. Most importantly, the side chains of H4 His18 and Lys20 are accommodated within two adjacent pockets (Figure 4).

Substitution of H4 His18 with the larger Trp residue totally disrupts binding with TONSL ARD (Figure 13), underscoring the importance of fitting His18 in the pocket. As the N ζ atom of H4 Lys20 forms three strong hydrogen bonds within its binding pocket, it can be predicted that methylation on H4K20 should break these critical interactions.

Both isothermal titration calorimetry (ITC) and H4 tail peptide pull-downs of recombinant ARD and full length TONSL from cell extracts confirmed that H4K20me1/2 is incompatible with TONSL binding (Figures 13, 14, 15, 17, 18, 19).

Further, in vitro pull-down of full-length recombinant TONSL-MMS22L with modified reconstituted mononucleosomes showed that H4K20me2 significantly reduced TONSL-MMS22L binding (Figure 16).

The term "H4 tail binding surface of TONSL ARD" as used herein refers to the part of TONSL ARD binding the H4 tail. Amino acids of TONSL ARD involved in the binding of the H4 tail are described in detail above in this section, and the "H4 tail binding surface of TONSL ARD" may comprise any of these amino acids. In particular, the "H4 tail binding surface of TONSL ARD" may comprise amino acids Asp527, Met528, Glu530, Asp559, Tyr560, Cys561, Trp563, Glu568, Asn571, Tyr572, Gly595, Glu597, Asp604, Asn607, Cys608, Trp641, Tyr645 and Leu649 of SEQ ID NO:16.

Our data establish that the methylation of H4K20 hinders binding with TONSL ARD.

A previous study found that the ARD of the G9a/GLP methyltransferases specifically recognize H3K9me1/2, but the TONSL ARD showed no binding to H3K9me1 peptides.

Consistent with the structural data, histone H4 mutations H18A and K20A disrupted
5 binding to TONSL in cell extracts (Figure 11). Conversely, mutation of 6 conserved
TONSL residues lining the H4 His18 and Lys20 binding pockets disrupted binding to
H3-H4 and MCM2, both in vivo and in vitro without affecting binding to MMS22L,
previously shown to bind to the C-terminal part of TONSL (Figures 10, 12). *In vivo*,
these mutants abrogated binding to soluble H3-H4 and, consequentially, association
10 with ASF1 and MCM2 was lost without affecting MMS22L binding to the C-terminal part
of TONSL(Fig. 12).

Taken together, these results show that TONSL binds to free histones and
nucleosomes via ARD recognition of H4 tails, with the important residues His18 and
15 Lys20 fitting within two adjacent pockets. It is also shown that aa 9 to 25 of SEQ ID
NO:23 can bind TONSL ARD (aa512-692 of SEQ ID NO:16), but that the mutants are
strongly impaired in H4 tail peptide binding.

It is apparent that a small molecule that could bind to the His18-binding pocket or the
20 Lys20-binding pocket, as well as either single or covalently-linked small molecules
that target both pockets on TONSL ARD, would disrupt the interactions between H4
and TONSL.

Drug design

25 In the most basic sense, the drug design of the present invention involves the design
of single or linked small molecules that are complementary in shape and charge to the
molecular target of the present invention. The invention also relates to peptides or
polypeptides designed to bind to the target.

30 The drug design of the present invention relies on the knowledge of the three-
dimensional structure presented herein. In addition to small molecules,
biopharmaceuticals are an increasingly important class of drugs and computational
methods for improving the affinity, selectivity, and stability of these protein-based
therapeutics are also embodiments of the present invention.

35

The small molecules of the present invention are preferably designed so as not to affect any other important "off-target" molecules (often referred to as anti-targets), since drug interactions with off-target molecules may lead to undesirable side effects.

- 5 The term "small molecule" as used herein refers to any molecule with a molecular weight below 1000 Da, for example below 500 Da. Preferably, the "small molecule" may be an organic molecule having a molecular weight below 500 Da.

10 Another class of small molecules that constitute an embodiment of the invention are aptamers, which can be made up of DNA, RNA, or peptide units. Aptamers bind to a specific target molecule, and may be designed as described herein in this section. Alternatively, they may be created by selecting them from a large random sequence pool. Thus, aptamer may be any aptamer having the binding properties described herein.

15 In contrast to traditional methods of drug discovery (known as forward pharmacology), which rely on trial-and-error testing of chemical substances on cultured cells or animals, and matching the apparent effects to treatments, rational drug design (also called reverse pharmacology) begins with a hypothesis that
20 modulation of a specific biological target may have therapeutic value.

In order for a small molecule to be selected as a drug target according to the present invention, one essential piece of information is required, namely the evidence that
25 modulation of the target by the small molecule of the present invention will be disease modifying. This knowledge comes from disease linkage studies that show an association between e.g. mutations in the biological target and certain disease states.

The search for small molecules *may begin* by screening libraries of potential drug compounds. This may be done by using a virtual screen of existing and available small
30 molecule libraries. The atomic co-ordinate data of the present invention or data derivable therefrom are useful in selecting and/or designing small molecules capable of interfering with the adjacent histone H4H18 and H4K20 binding pockets on the surface of the Ankyrin repeats of TONSL.

Thus, in one aspect, the present invention relates to methods for selecting and/or designing such small molecules, wherein for e.g. the method involves the use of a computer modelling or a computer program to model all or part of the structure disclosed in the PDB Protein Databank under the PDB ID 5JA4, DOI:

- 5 [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or data derivable therefrom, identifying a potential small molecule based on its likely ability to interact with the modelled structure, synthesising or obtaining the small molecule from a commercial source and carrying out *in vitro* testing of the functionality.
- 10 In one embodiment, the present invention relates to a computer-based method for identifying a small molecule capable of interfering with the adjacent histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, comprising the steps of
- 15 a) providing a 3D structural representation of the histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL in a storage medium on a computer, wherein the 3D structural representation is derived from the atomic co-ordinates available in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or a variant thereof in which the
- 20 r.m.s. deviation of the x, y and z co-ordinates for all heavy atoms is less than 1.0 Å, and
- b) using the computer to apply structure-based drug design techniques to the structural representation.
- 25 In one embodiment, the structure-based drug design includes one or more steps of docking, such as but not limited to docking steps which screens members of a structural library.
- 30 In another embodiment, the invention relates to a computer-based method for identifying a small molecule capable of interfering with the adjacent histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, comprising the steps of
- 35 a) providing a 3D structural representation of the histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL in a storage medium on a computer, wherein the 3D structurel representation is derived

from the atomic co-ordinates of available in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or a variant thereof in which the r.m.s. deviation of the x, y and z co-ordinates for all heavy atoms is less than 1.0 Å, and

5

b) using the computer to apply structure-based drug design techniques to the structural representation, and

10

c) providing a compound identified by said structure based drug design technique, and

d) contacting said compound with the binding pocket on the surface of the Ankyrin repeats of TONSL and assaying the interaction between them.

15 In another embodiment, the invention relates to method of identifying a small molecule capable of interfering with the histone H4H18 and H4K18 binding pocket on the surface of the Ankyrin repeats of TONSL, or a fragment or variant thereof, said method comprising the steps of:

20

a. generating the spatial structure of the pocket on a computer screen using atomic coordinates as presented in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb), data derivable therefrom, or by a root mean square deviation over protein backbone atoms of not more than 1.0 Å,

25

b. generating potential small molecules with their spatial structure on the computer screen, and

c. selecting small molecules that can bind to at least one amino acid residue of the set of binding interaction sites.

30 In another embodiment, the invention relates to a computer-assisted method for identifying a small molecule capable of interfering with the histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, or a fragment or variant thereof, using a programmed computer comprising a processor, a data storage system, a data input device and a data output device, comprising the following steps:

35

a. inputting into the programmed computer through said input device data comprising; atomic coordinates of a subset of the atoms according to the present invention, thereby generating a criteria data set; wherein said atomic

coordinates are selected from the atomic coordinates as presented in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb), data derivable therefrom, or by a root mean square deviation over protein backbone atoms of not more than 1.0 Å,

5

b. comparing, using said processor, the criteria data set to a computer data base of low-molecular weight organic chemical structures and peptide fragments stored in the data storage system, and

10

c. selecting from said data base, using computer methods, a chemical structure having a portion that is structurally complementary to the criteria data set.

15

In another embodiment, the invention relates to a method for identifying a ligand, comprising the steps of:

20

a. selecting a potential ligand using atomic coordinates in conjunction with computer modelling, wherein said atomic coordinates are the atomic coordinates as presented in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb), data derivable therefrom, or by a root mean square deviation over protein backbone atoms of not more than 1.0 Å, by docking potential ligands into a set of binding interaction sites, said binding interaction generated by computer modelling and selecting a potential ligand capable of binding to at least one amino acid in said set of binding interaction sites of TONSL,

25

b. providing said potential ligand and said TONSL,

30

c. contacting the potential ligand with said TONSL, and

d. detecting binding of said potential ligand with TONSL.

A non-limiting Example of methods for identifying small molecules or inhibitors according to the invention is provided in Example 16 herein below.

35

SCHRODINGER analysis

The small molecule drug discovery suite by SCHRODINGER contains a comprehensive set of programs aimed at state-of-the-art support of every step of drug discovery and optimization. This represents one type of structure-based docking programs for small molecule inhibitor design, other could also be applied. The essence of drug discovery is in (1) finding key adjacent pocket(s) on a target protein that affects or controls its function and (2) finding covalently linked small molecules with high-affinity binding to the adjacent pocket(s) in the known databases followed by iterative optimization through addition/deletion of substituents while maximizing shape, hydrogen bonding and electrostatic complementarity.

The most effective definition of a binding pocket is possible when a crystal structure of a protein-ligand complex is known. In this case, the high precision GLIDE algorithm is applied to calculate the grid, which is subsequently used for high-throughput docking of commercially available small molecules. This grid represents the three-dimensional spatial information about essential components of binding surfaces such as all lipophilic atoms, all hydrogen-bonding atoms with a score accounting for orientation of hydrogen bonds, metal atoms if present in the protein, distribution of charges, as well as van der Waals atomic interactions. The ligand conformations are also pre-computed in SCHRODINGER through application of LigPrep procedure that filters out the least probable configurations and accounts for various pH-dependent ionization states of the prospective ligands.

Docking computations for two interacting components thus represented are highly precise and can be performed on high scale and with various degrees of controlled flexibility within the protein component. The protein-ligand complexes with ligands screened from databases can next be ranked using advantages of post-docking Embrace minimizations and prime MM GBSA of the SCHRODINGER suite. These procedures give estimates of free binding energy and thus are suitable tools for ligand ranking based on the computed binding affinity.

When the crystal structure of a drug-protein complex is not known, as is the case in TONSL-H4 tail complex, the protein binding pocket(s) and corresponding grid(s) can be computed from the assumption that binding of the prospective ligand(s) should outcompete the most critical aspects for interacting such as for e.g. amino acid residues H18 and K20 on histone H4 tail or any of the above listed interactions.

- In this case, the docking grid(s) are computed for the TONSL pocket(s) as defined by the cavities that incorporate these two residues: individually or through being covalently connected together. The scoring function for the competing small molecules
- 5 in this case is compared with that of the original binding targets. Optimization of identified small molecule leads is achieved by introduction/deletion of R-groups and computing the resulting effect on their affinity by isothermal titration calorimetry, prior to chemical synthesis of the most promising candidates.
- 10 Prospective drug candidates are generated in an iterative process of x-ray structure determination and ligand optimization, with the goal of identifying covalently linked optimized ligands that simultaneously target the H4H18 and H4K20 pockets on the TONSL scaffold.
- 15 The route to effective compounds according to the present invention is to
1. Determine by the use of the structure scaffold/co-ordinates in-silico and standard binding/scoring algorithms in-silico molecules with high binding affinity to the TONSL/H4 crystal structure disclosed herein
 - 20 2. Modify the small molecule structures (with R-substitutions) in silico to represent drug-like compounds
 3. Synthesize such structures by standard chemistry
 - 25 4. Test the resulting synthesized structures in the assays of the invention
 5. Select, modify and optimize the most promising compounds.
- 30 An alternative to this approach is to screen existing libraries in the assays, and optionally optimize these so as bind in the space as described.
- 35 Thus, in one embodiment the present invention relates to a method of selecting or designing a small molecule capable of interfering with the histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, said method

comprises use of at least part of the atomic co-ordinates data contained in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb) or data

derivable therefrom, wherein said method involves use of a computer modelling package or a computer program to model all or part of the structure of MCM2 HBD–

5 G4–TONSL ARD in complex with H3 (57–135) and H4, identifying such small molecule based on its likely ability to prevent or disrupt the H4 tail with the Ankyrin repeats of TONSL in the modelled structure.

In one embodiment, the invention relates to a fusion protein, MCM2 HBD–G₄–TONSL
10 ARD (SEQ ID NO: 15) or variants thereof (e.g. G_x linker = G₁₂, G₁₁, G₁₀, G₉, G₈, G₇, G₆, or G₅), which are instrumental for determining the structural details of TONSL binding small molecule inhibitors.

In another embodiment, the invention relates to the TONSL and H4 mutants that
15 disrupt interaction between TONLS and TONSL ARD with histone H4 (Figures 10–13, Examples 2–4). The mutant data identify key intermolecular interactions contributing to the specificity and stability of complex formation. The verified TONSL histone-binding mutants are instrumental for biochemical and biological assay design to assess the efficiency of screened small molecule inhibitors in blocking TONSL-histone
20 interaction and TONSL function.

In one embodiment the invention relates to compounds having a 3-[(3-Aminocyclopentyl)carbonyl]-1H-quinolin-4-one core. Such compounds may be any compound, preferably any small organic compound comprising the 3-[(3-
25 Aminocyclopentyl)carbonyl]-1H-quinolin-4-one, wherein one or more positions may be substituted with a substituent. Thus, the compound may be a 3-[(3-Aminocyclopentyl)carbonyl]-1H-quinolin-4-one, wherein one or more –H have been substituted for another group.

30 *Peptide inhibitors of TONSL*

The invention also provides inhibitors of TONSL, in particular such inhibitors, which can inhibit binding of TONSL ARD to histone H4. In preferred embodiments the inhibitor may be a compound binding the histone H4 tail binding surface of TONSL ARD.

Methods for identifying such inhibitors are provided herein elsewhere.

The inhibitor may be any useful compound for example a peptide inhibitor or a small molecule. In one embodiment of the invention, the inhibitor is a peptide inhibitor. The peptide inhibitor may for example be useful in the treatment of cancer.

As used herein the term "peptide inhibitor" refers to a compound comprising a peptide or a polypeptide optionally linked to a conjugated moiety. The peptide or polypeptide part of the peptide inhibitor is in the following referred to as "peptide".

10

The peptide inhibitor may thus be a peptide, which is capable of binding the TONSL ARD. Thus, the peptide inhibitors may be any of the peptides described herein below, wherein said peptide is capable of binding to TONSL ARD (e.g. to a peptide consisting of amino acids 512 to 692 of SEQ ID NO:16) with a K_d of at the most 10 μM , preferably a K_d of at the most 5 μM , such as with a K_d of at the most 3 μM . Said K_d may for example be determined as described in Example 10 and 15 below.

15

In one embodiment the peptide may comprise the sequence motif I: Arg-His-Xaa-Lys (SEQ ID NO:26), wherein Xaa may be any amino acid.

20

In one embodiment the peptide may comprise the sequence motif II: Arg-His-Xaa-Lys-Val-Leu (SEQ ID NO:27), wherein Xaa may be any amino acid.

In one embodiment the peptide may comprise the sequence motif III: Val-Leu-Arg.

In one embodiment the peptide may comprise the sequence motif IV: Arg-His-Xaa-Lys-Val-Leu-Arg (SEQ ID NO:28), wherein Xaa may be any amino acid.

25

In particular, the peptide may be a peptide consisting of in the range of 4 to 40 amino acids comprising the sequence Arg-His-Xaa-Lys, and/or one or more of the sequence motifs I, II, III or IV. The peptide may also be a peptide consisting of in the range of 4 to 25 amino acids, such as in the range of 4 to 15 amino acids, preferably in the range of 7 to 15 amino acids, such as in the range of 7 to 12 amino acids, such as in the range of 4 to 10 amino acids, for example in the range of 6 to 9 amino acids comprising the sequence Arg-His-Xaa-Lys, and/or one or more of the sequence motifs I, II, III or IV. The peptide may also be a peptide consisting of at the most 25 amino acids, preferably at the most 15 amino acids, such as at the most 12 amino acids, for

30

example at the most 9 amino acids, wherein the peptide comprises one or more of the sequence motifs I, II, III or IV. Said Lys is preferably unmethylated. Said Xaa may be any amino acid, for example it may be selected from the group consisting of Ala and Arg. In one embodiment of the invention Xaa is not Arg.

5 In one embodiment the peptide comprises or consists of:

- I. a sequence consisting of amino acid 12 to 23 of SEQ ID NO:23 or of SEQ ID NO:34; or
- 10 II. a functional homologue thereof consisting of a sequence of amino acid 12 to 23 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids, for example up to 4 amino acids, for example up to 3 amino acids, such as up to 2 amino acids, for example 1 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:23 or of SEQ ID NO:34,
- 15 III. a functional homologue thereof consisting of a sequence of amino acid 12 to 23 of SEQ ID NO:23, wherein up to 6 amino acids, such as up to 5 amino acids, for example up to 4 amino acids, such as up to 3 amino acids, for example up to 2 amino acids may be substituted may be substituted,
- 20 with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

In one embodiment the peptide comprises or consists of:

- IV. a sequence consisting of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34; or
- 25 V. a functional homologue thereof consisting of a sequence of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids, for example up to 4 amino acids, for example up to 3 amino acids, such as up to 2 amino acids, for example 1 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:23 or of SEQ ID NO:34,
- 30 VI. a functional homologue thereof consisting of a sequence of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 6 amino acids, such as up to 5 amino acids, for example up to 4 amino acids, such as up to 3 amino acids, for example up to 2 amino acids may be substituted
- 35

with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

In one embodiment the peptide comprises or consists of:

- 5 VII. a sequence consisting of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34; or
- 10 VIII. a functional homologue thereof consisting of a sequence of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids, for example up to 4 amino acids, for example up to 3 amino acids, such as up to 2 amino acids, for example 1 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:23,
- 15 IX. a functional homologue thereof consisting of a sequence of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 6 amino acids, such as up to 5 amino acids, for example up to 4 amino acids, such as up to 3 amino acids, for example up to 2 amino acids may be substituted

with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

20

In embodiments of the invention where the peptide comprises a functional homologue as described under III., VI. or IX herein above it may be preferred that:

- 25 I. the amino acid corresponding to amino acid 17 of SEQ ID NO:23 or of SEQ ID NO:34 is a charged amino acid, such as a positively charged amino acid, for example an amino acid selected from the group consisting of Lys and Arg
- II. the amino acid corresponding to amino acid 18 of SEQ ID NO:23 or of SEQ ID NO:34 is a charged amino acid, such as a positively charged amino acid, for example His
- 30 III. the amino acid corresponding to amino acid 20 of SEQ ID NO:23 or of SEQ ID NO:34 is a charged amino acid, such as a positively charged amino acid, for example an amino acid selected from the group consisting of Lys and Arg.

- 35 In one embodiment, the peptide inhibitors may be identified based on the sequence of histone H4 of SEQ ID NO:23 or a fragment thereof, i.e. by systematically or randomly

replacing one or more amino acids of an H4 peptide, or certain parts of H4 native peptide, with other amino acids.

In general the peptide inhibitor should not be too large. Accordingly, it may be preferred that the peptide consists of at the most 40 amino acids, such as at the most 25 amino acids, for example at the most 20 amino acids, such as at the most 15 amino acids, for example at the most 12 amino acids. For example, the peptide may comprise or consist of in the range of 4 to 40, for example in the range of 4 to 25, for example in the range of 4 to 20, such as in the range of 4 to 15, such as in the range of 4 to 12, for example in the range of 7 to 15 amino acids, such as in the range of 7 to 12 amino acids, such as in the range of 5 to 10 or such as in the range of 6 to 9 consecutive amino acids of SEQ ID NO:23.

In one embodiment, the peptide may comprise or consist of in the range of 4 to 40, for example in the range of 4 to 25, for example in the range of 4 to 20, such as in the range of 4 to 15, such as in the range of 4 to 12, for example in the range of 7 to 15 amino acids, such as in the range of 7 to 12 amino acids, such as in the range of 5 to 10 or such as in the range of 6 to 9 consecutive amino acids of SEQ ID NO:23, wherein up to 3, such as up to 2, for example at the most one amino acid may have been substituted for another amino acid, and wherein the peptide comprises one or more of the sequence motifs I, II, III or IV.

Said functional homologues of fragments of SEQ ID NO:23 are preferably peptides, comprising above defined sequence, and which are capable of binding TONSL.

As described above the peptide may comprise one or more Lys residues. It may be preferred that at least one Lys residue, for example all Lys residues are unmethylated. In particular, when the peptide comprises a fragment of SEQ ID NO:23 it may be preferred that the amino acid corresponding to Lys20 of SEQ ID NO:23 is unmethylated.

In some embodiments of the invention, the C-terminal of the peptide is amidated, i.e. having the chemical structure $-C(O)NH_2$.

In some embodiments of the invention, the C-terminal of the peptide is alkylated e.g. methylated i.e. having the chemical structure $-C(O)OCH_3$.

In some embodiments of the invention, the N-terminal of the peptide is acetylated i.e. having the chemical structure $\text{CH}_3\text{C}(\text{O})\text{N}(\text{H})-$.

5 In some embodiments of the invention, the N-terminal of the peptide is formylated i.e. having the chemical structure $\text{HC}(\text{O})\text{N}(\text{H})-$.

It is understood that a peptide consisting of a given sequence may have a C-terminal, which is amidated or alkylated and/or an N-terminal, which is acetylated or formylated.

10

The peptide inhibitor may in particular be a peptide comprising or consisting of a peptide selected from the group consisting of:

Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂,

Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂,

15 Lys-Gly-Gly-Ala-Lys-Arg-His-Ala-Lys-Val-Leu-Arg-NH₂ and

Lys-Gly-Gly-Ala-Ala-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂.

The moiety "-NH₂" indicated in the sequence above indicates that the C-terminal of the peptides are amidated. The invention however also encompass peptides which are not amidated. Accordingly, in one embodiment, the invention relates to peptide inhibitors comprising or consisting of at the most 40, for example at the most 25, for example at the most 20, such as at the most 15, such as at the most 12 amino acids, wherein the peptide comprises a sequence selected from the group consisting of SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32 and SEQ ID NO:33.

25

In some embodiments of the invention, the peptide is a TONSL mutant polypeptide, for example any of the TONSL mutant polypeptides described herein below in the section "TONSL mutant polypeptides". For example, the peptide may be selected from the group consisting of polypeptides of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, 30 SEQ ID NO:20, SEQ ID NO:21 and SEQ ID NO:22.

In some embodiments of the invention the inhibitor is a fragment of TONSL, in particular a fragment of TONSL capable of binding histone H4, but incapable of binding at least one other binding partner of TONSL, e.g. incapable of binding MMS22L. Said 35 fragment of TONSL, may for example be a fragment comprising the H4 tail binding surface of TONSL. In one embodiment, the fragment comprises or consists of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue thereof sharing at least

70%, such as at least 80%, for example at least 85%, such as at least 90%, for example at least 95% sequence identity therewith.

5 The peptide described herein may in some embodiments be linked to a conjugated moiety, for example the peptide may be covalently linked to a conjugated moiety. Said conjugated moiety may for example be any of the moieties described herein below.

10 In one embodiment the conjugated moiety may be a peptide, a sugar, a lipid, a polymeric molecule or any other chemical group that can be covalently linked to a peptide. For example, the conjugated moiety may improve the physical properties of the peptide, such as its solubility, stability or half-life.

15 In one embodiment the conjugated moiety is a polymeric molecule, such as polyethylene glycol (PEG) and polyvinylpyrrolidone (PVP). The polymeric molecule may also be a modified PEG, for example NPEG.

Assays validating the small molecules of the present invention

20 Molecules identified according to the present invention are tested in biochemical and cell biology assays listed below. Molecules that show activity in these assays towards inhibiting TONSL function similar to TONSL ARD mutants are further tested in high-throughput biological assay for efficacy in killing cancer cells and ultimately in pre-clinical and clinical trials.

Biochemical screening and biophysical testing (assays)

25 The key components in the assays are:

- a) Recombinant human TONSL ARD (residues 512-692, point 1)
- b) Recombinant human TONSL ARD mutants E530A, D559A, W563A,
30 E568A, N571A, D604A (point 2)
- c) Histone H4 peptide-containing K20me0 with or without acetylation at K16 (H4 tail covering the critical residues for TONSL binding, for example 9-25 or 14-33) coupled to fluorescent molecule or biotin

d) Histone H4 peptide containing K20me2 with or without acetylation at K16 as a control for no binding to TONSL ARD H4 tail covering the critical residues for TONSL binding, for example 9-25 or 14-33) coupled to fluorescent molecule or biotin

e) Human cells expressing tagged TONSL ARD wild type (WT) or histone-binding mutant (e.g. any of the mutants described herein in the section "TONSL mutant polypeptide", for example any of E530A, D559A, W563A, E568A, N571A, or D604A,). We generated and verified expression constructs for GFP-fusion proteins for mutants and inducible cell lines for WT, E568A and N571A

Assay 1: High-throughput (HT) measurements of small molecule dependent disruption of the TONSL ARD – H4 tail interaction detailed in point 2.

First line screening may be based on Fluorescence anisotropy measurements using histone H4 peptides conjugated with a fluorescent probe. Assay conditions are as detailed below in our demonstrated ITC assay (Example 2), with the critical criteria that ARD binding being observed to H4 (residues 9-25), but not H4K20me2 (residues 9-25, dimethylated at K20) (Figure 13).

Assay 2: Measurement of binding properties of small molecule inhibitors to ARD

Second line screening may involve Isothermal titration calorimetry (ITC) to determine binding affinity/association constants (K_a) and Surface plasmon resonance (SPR) assays to determine rate constants. The resulting molecules yielding higher binding affinity to TONSL than unmodified histone H4 peptide (Example 2) are further shortlisted to identify those with μM to nM effective range.

Assay 3. Crystallization and optimization

Candidate small molecules may be crystallized with TONSL ARD (using our demonstrated crystallization conditions outlined in point 1, Example 1), providing further possibility for drug optimization.

Assay 4. Competition assay based on peptide pulldown

a. Selected small molecules are validated for their ability to compete out binding of recombinant ARD to H4K20me0 peptides (H4K20me2 and ARD mutants are used as negative controls, assay conditions as in Example 3).

5

b. Selected small molecules are validated for their ability to compete out binding of GFP-TONSL-MMS22L to H4K20me0 peptides in cell extracts.

10 Grow cells expressing tagged TONSL WT or ARD mutant (as control for no histone H4 binding). Prepare whole-cell extracts according to standard protocols (documented conditions for this assay in Example 4). Incubate the cell extracts with surface-bound biotinylated-histone H4 peptides (e.g. H4K20me0 vs. H4K20me2 as a control for no binding to WT TONSL). Dose-response analysis with small molecule inhibitors assaying loss of WT TONSL binding to H4K20me0 peptides by ELISA, western blotting or similar
15 approaches.

HT biological assays to screen for small molecule inhibitors

The present inventors have demonstrated that TONSL binds to chromatin via ARD recognition of the tails of nucleosomal histone H4 unmethylated at K20 (Figure 21).
20 The biological function of a small molecule inhibitor is thus to prevent TONSL from binding histone H4 and displace TONSL-MMS22L complex from chromatin.

The present inventors have developed high-content microscopy based assays, which are used for drug screening (Example 5).

25

1) Immunofluorescence staining of endogenous TONSL shows binding to post-replicative chromatin, which is lost upon depletion of TONSL with RNAi (Figure 20).

30

2) binding of GFP-TONSL to chromatin is disrupted by ARD mutants that disrupt H4 binding (Figure 21).

The present invention covers small molecule inhibitors that displace TONSL and TONSL-MMS22L complex from post-replicative chromatin and mimic TONSL ARD
35 mutants.

Key components:

- a) Anti-TONSL antibody (Sigma, ref. nr. HPA0244679), tested in our high-content assay (Example 5)
- 5 b) Pre-extraction conditions to remove soluble TONSL (Example 5)
- c) Condition for TONSL depletion (negative control) (Example 5)
- d) GFP-TONSL WT and ARD mutant cell lines

Rationale of HT assays:

- 10 1. High-content imaging to identify small molecules that remove endogenous TONSL and TONSL-MMS22L complex from chromatin (Figure 21)
- 2. High-content imaging of TONSL binding to chromatin using cells expressing GFP-TONSL WT and ARD mutant (N571A, point 2) upon incubation with TONSL inhibitors by high-content imaging

15

Assay 1: HT imaging of chromatin-bound endogenous TONSL

Cells are grown in a HT format (96 or 384 well-plates). Incubated with small molecule TONSL inhibitors in a dose-response and time-course set-up. Soluble proteins are removed by cell pre-extraction followed by cell fixation according to standard protocols. Detect chromatin-bound TONSL with commercial antibody (Sigma, ref. nr. HPA0244679) documented by us for specificity (Example 5). TONSL siRNA are used as control. Chromatin-bound TONSL may be quantified by conventional methods, for example using standard high content imaging (Example 5) or alternatively by FACS analysis.

25

Assay 2: HT imaging of GFP-TONSL

Grow cells with expression of GFP-TONSL WT and ARD mutant (N571A, point 2: negative control for binding) in a HT format (96 or 384 well-plates). Incubate with small molecule TONSL inhibitors in a dose-response and time-course set-up. Remove soluble proteins by pre-extraction cells and fix cells according to standard protocols. Quantify GFP-TONSL levels on chromatin by standard high content imaging (Example 5).

30

Biological assays to measure HR inhibition

TONSL is required for HR of damaged replication forks and DNA double strand breaks (Duro et al. 2010 Mol Cell 40:619; O'Donnell et al. 2010 Mol Cel 40:619; O'Connell et al. 2010 Mol Cell 40:645; Piwko et al. EMBO J 2010 29:4210). Thus, small molecule
5 inhibitors preventing TONSL or MMS22L recruitment to chromatin phenocopy the loss of TONSL or MMS22L.

To test the capacity of small molecule inhibitors to compromise HR, published assays to monitor recruitment of repair proteins (e.g., Duro et al. 2010 Mol Cell 40:619;
10 O'Donnell et al. 2010 Mol Cel 40:619; O'Connell et al. 2010 Mol Cell 40:645; Piwko et al. EMBO J 2010 29:4210) and HR reporter cell lines are used (e.g., Pierce et al., Genes Dev 13:2633, 1999). In those assays, the cells are treated with small molecule TONSL inhibitors in a time and dose-dependent manner and efficiency of HR repair is assayed by high-content microscopy

15

TONSL mutant polypeptide

In one embodiment the invention relates to a TONSL mutant polypeptide. Said TONSL mutant polypeptide is preferably a TONSL mutant polypeptide, which is incapable of
20 binding histone H4, but which retains other TONSL functions. Preferably, said TONSL mutant polypeptide is capable of binding MMS22L and wherein said TONSL mutant does not bind histone H4. Such TONSL mutant polypeptides will function as dominant negative mutants, because they will bind MMS22L and other members of the complex, preferably with the same or similar affinity as the wild type TONSL polypeptide, but
25 they will not bind the post-replicative chromatin to any significant extent.

In one embodiment the TONSL mutant polypeptide is a polypeptide of SEQ ID NO:16 carrying one or more mutations, preferably one or more mutations in amino acid(s) contributing to the H4 tail binding surface of TONSL ARD. Thus, the TONSL mutant
30 polypeptide may comprise at least one mutation in an amino acid of the histone H4 tail binding surface of the TONSL ARD, wherein said TONSL mutant polypeptide apart from said mutation is identical to SEQ ID NO:16 or shares a sequence identity with SEQ ID NO:16 of at least 70%, wherein said TONSL mutant polypeptide is capable of binding MMS22L and wherein said TONSL mutant does not bind histone H4.

35 In particular said TONSL mutant polypeptide may be a polypeptide of SEQ ID NO:16 carrying a mutation in at least one amino acid selected from the group consisting of

amino acid number 527, 528, 530, 559, 560, 561, 563, 568, 571, 572, 595, 597, 604, 607, 608, 641, 645 and 649 of SEQ ID NO:16.

5 In one embodiment the TONSL mutant polypeptide carries a mutation in one or more of the amino acids selected from the group consisting of: Glu530, Asp559, Trp563, Glu568, Gln571 and D604 of SEQ ID NO:16.

10 In one embodiment TONSL mutant polypeptide may comprise or even consist of a sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21 and SEQ ID NO:22.

15 The TONSL mutant polypeptide may be linked to a conjugated moiety, for example the TONSL mutant polypeptide may be linked to a detectable marker. Said detectable marker may be a peptide or a polypeptide, e.g. a fluorescent protein, such as GFP.

A method for predicting the effect of inhibition of TONSL

The invention also provides methods for predicting the effect of inhibition of TONSL. As described herein above, the TONSL mutant polypeptides described herein may be dominant negative mutant. Expression of such mutants may phenocopy
20 TONSL/MMS22L depletion, and thus mimic the actions of an inhibitor of TONSL binding to histone H4.

Thus, in one embodiment the invention relates to methods for predicting the effect of inhibition of TONSL, said method comprising the steps of

- 25
- expressing a TONSL mutant polypeptide in an organism and/or cells of an organism, wherein said polypeptide optionally is expressed conditionally,
 - determining the effect of said polypeptide in said organism and/or cells wherein said effect of expressing said polypeptide in said organism and/or cells indicates the effect of inhibition of TONSL in said organism and/or cells.

30 In particular, said TONSL mutant polypeptide may be any of the TONSL mutant polypeptides described herein above in the section "TONSL mutant polypeptide". The organism and/or cells may comprise a heterologous nucleic acid encoding any of said TONSL mutant polypeptides.

Thus, the effect obtained after expression of a dominant negative TONSL mutant
35 polypeptide may mimic the effect of inhibition of TONSL in said organism and/or cells. Thus, the effect observed is indicative of the effect expected to be obtained, if the

organism/cells were treated with an inhibitor of TONSL, in particular with an inhibitor inhibiting binding between TONSL ARD and the histone H4.

The organism or the cells may represent a disease model, and thus the method may be used to predict the efficacy of treating said disease with an inhibitor or TONSL, e.g. an inhibitor inhibiting binding between TONSL ARD and the histone H4.

The organism may be any organism, e.g. a mammal, for example a mouse, a rat or a rabbit. The organism may be genetically modified to contract a particular disease, for example cancer. Thus, the organism may be a mouse genetically engineered to contract cancer.

The cells may also represent a disease model. The cells may be any cells, such as mammalian cells, e.g. human cells. The cells may be cultivated using any conventional method. For example, the cells may be cultured in a 3D culture, which more closely may mimic in vivo conditions.

Since expression of TONSL mutant polypeptide may be toxic to cells, it may be preferred that said TONSL mutant polypeptide is expressed only conditionally. Thus, the TONSL mutant polypeptide may be expressed only in specific cells of said organism and/or said TONSL mutant polypeptide may be induced to be expressed at specific times. Thus, the organisms and/or cells may comprise a heterologous nucleic acid encoding any of the TONSL mutant polypeptides described herein above in the section "TONSL mutant polypeptide" operably linked to an inducible promoter. In that manner expression of said TONSL mutant polypeptide may be induced at any specific time. Inducible promoters are well known in the art and the skilled person will be able to select a useful promoter. The organism may also comprise a heterologous nucleic acid encoding any of said TONSL mutant polypeptides operably linked to a promoter directing expression in only some cell types. Such promoters are also known in the art.

Method for determining TONSL inhibitory effect

The invention also provides methods of identifying the TONSL inhibitory effect of a putative inhibitor of TONSL. In particular said methods can be used to determine whether a compound is capable of inhibiting binding between TONSL and histone H4. Such methods are useful for screening for inhibitors of TONSL. However, the methods are also useful for validating whether a putative inhibitor in fact is capable of inhibiting TONSL. Thus, the methods can be used for determining biosimilarity between a known

TONSL inhibitor and a similar compound, which is a putative TONSL inhibitor. The methods may also be used in quality control of TONSL inhibitors. The methods may also be used to validate the inhibitory effect of an inhibitor identified by computer aided techniques using the atomic coordinates provided in the PDB Protein Databank under the PDB ID 5JA4, DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb).

Said methods may comprise the steps of

- a) providing a test compound, which is a putative inhibitor of TONSL
 - b) providing an host organism expressing TONSL of SEQ ID NO:16, TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue of any of the aforementioned sharing at least 90% sequence identity to amino acids 512 to 692 of SEQ ID NO:16;
 - c) contacting said host organism with said putative inhibitor
 - d) detecting chromatin associated TONSL, TONSL ARD or functional homologue thereof in said host organism,
- wherein reduction of chromatin associated TONSL, TONSL ARD or a functional homologue thereof is indicative of said test compound being an inhibitor of TONSL.

Said reduction of chromatin associated TONSL, TONSL ARD or a functional homologue thereof is preferably a reduction compared to said level in said host organism in the absence of said inhibitor. In particular, said reduction may be a reduction to less than 50%, such as a reduction to less than 30%, such as a reduction to less than 20%, for example a reduction to less than 10% of the chromatin associated TONSL, TONSL ARD or a functional homologue thereof in said host organism in the absence of said inhibitor. In one embodiment, the reduction is that there is no detectable chromatin associated TONSL, TONSL ARD or a functional homologue thereof.

The host organism may be any useful host organism including cells, e.g. mammalian cells maintained in a tissue culture. Thus, the term "cells of the host organism" may refer to part of a multicellular host organism or it may refer to the host organism per se, when the host organism is cells. The host organism may endogenously express TONSL, in which case the method may involve detecting chromatin associated endogenous TONSL. Endogenous TONSL may be detected by various means, but frequently, endogenous TONSL is detected by means of an antibody, a binding fragment of an antibody or another binding molecule specifically recognising and binding TONSL. The antibody, fragment or binding molecule may be directly labelled with a detectable label. It is also possible that the methods involve use of a secondary

antibody, fragment or binding molecule binding the first antibody, fragment or binding molecule, wherein the secondary antibody, fragment or binding molecule is labelled with a detectable label. Other methods for detecting binding of an antibody, fragment or binding molecule are also available and may be used with the methods.

5

The detectable label may be any detectable label including but not limited to fluorophores, bioluminescents, chemoluminescents, dyes, enzymes, heavy metals, or radioactive compounds. In preferred embodiments the detectable label is a fluorophore, i.e. a fluorescent moiety.

10

It is also comprised within the invention that said host organism comprises a heterologous nucleic acid encoding said TONSL, TONSL ARD or functional homologue thereof. In such embodiments the methods may involve detecting chromatin associated heterologous TONSL, TONSL ARD or functional homologue thereof. The

15

heterologous TONSL, TONSL ARD or functional homologue thereof may be linked to a detectable label, for example to a fluorescent polypeptide, such as GFP. Thus, the step of detection, may be a step of detecting the detectable label. Heterologous TONSL, TONSL ARD or functional homologue thereof may also be detected used the same methods useful for detection of endogenous TONSL.

20

The detectable label may be detected using any technical means useful for detecting the particular detectable label. In embodiments of the invention, where the detectable label is fluorescent, then the detectable label may be detected by means of FACS, fluorescent microscopy or high content microscopy.

25

Step d) of the method may comprise a step of extracting soluble proteins from the cells of said host organism. This step may ensure that all TONSL, TONSL ARD or functional homologues thereof, which is not chromatin associated is extracted from the cells of said host organism. Following extraction of soluble proteins all remaining

30 TONSL, TONSL ARD or functional homologues thereof may be regarded as chromatin associated TONSL, TONSL ARD or functional homologues thereof .

35

Extracting soluble protein may be done in any conventional manner, e.g. by permeabilising cell membranes, e.g. by use of a detergent optionally followed by one or multiple washing steps. Any detergent may be used for permeabilisation, e.g. Triton (e.g. Triton X-100), NP-40, saponin, Tween (e.g. Tween-20), Digitonin or Leucoperm. After extraction the cells may be fixed prior to detection. Fixation

reagents are well known in the art and includes for example aldehydes, e.g. formaldehyde, paraformaldehyde or glutaraldehyde.

One useful method for determining chromatin associated TONSL, TONSL ARD or functional homologues thereof is described in Example 5 herein. Other useful methods for determining chromatin associated TONSL, TONSL ARD or functional homologues thereof are described in the section "High-throughput (HT) biological assays for efficacy of small molecule inhibitors in killing cancer cells", for example in Assay 1 and 2 of that section.

The methods for determining TONSL inhibitory effect may also be any of the methods described herein in the section "Assays validating the small molecules of the present invention", for example any of the assays 1, 2, 3 or 4. Thus, these methods may not only be used for validating small molecules, but can also generally be used for determining TONSL inhibitory effect of a compound.

In one embodiment the method comprises the steps of:

- a) Providing histone H4 or a fragment thereof comprising at least amino acids 17 to 20, such as at least amino acids 12 to 23, for example at least amino acids 9 to 25, such as at least amino acids 14 to 33 of SEQ ID NO:23, wherein said histone H4 or fragment thereof optionally may be linked to a detectable label;
- b) providing TONSL of SEQ ID NO:16, TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue of any of the aforementioned sharing at least 90% sequence identity to amino acids 512 to 692 of SEQ ID NO:16
- c) Incubating said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof in the presence of a putative inhibitor of TONSL
- d) Determining whether binding between said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof

wherein reduction in binding of said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof is indicative of that said putative inhibitor of TONSL is capable of inhibiting TONSL.

Said reduction in binding may in particular be a reduction to at the most 50%, such as a reduction to less than 30%, such as a reduction to less than 20%, for example a reduction to less than 10% of the binding of said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof observed in the absence of
5 said putative inhibitor. In one embodiment, the reduction is that there is no detectable binding between said histone H4 or fragment thereof and said TONSL, TONSL ARD or a functional homologue thereof.

10 *High-throughput (HT) biological assays for efficacy of small molecule inhibitors in killing cancer cells*

Loss of TONSL leads to accumulation of replication associated DNA damage and sensitizes cells to camptothecin (CPT)-induced DNA damage (Duro et al. 2010 Mol Cell 40:619; O'Donnell et al. 2010 Mol Cell 40:619; O'Connell et al. 2010 Mol Cell 40:645;
15 Piwko et al. EMBO J 2010 29:4210). Further it is well established that loss of HR sensitizes cells to PARP inhibition. The prediction is thus that TONSL inhibition alone or in combination with CPT-like drugs or PARP inhibitors will be toxic to cancer cells.

Key components:

- 20 a) Panel of cancer cell lines and primary cells
 b) Camptothecin, PARP inhibitors, conventional chemotherapy drugs targeting DNA replication

Rationale of HT assays:

- 25 i. Assay the cell cytotoxicity of TONSL inhibitors by screening a panel of cancer cell lines and primary cells.
 ii. Assay the synergy of TONSL inhibitors and other cancer drugs (conventional chemotherapy, CPT and derivatives, PARP inhibitors) in killing cancer cells.

30

Typical cell toxicity assay:

Grow cells in any HT format (e.g. 96 or 384 plates) and incubate with small molecule inhibitors in a dose and time dependent manner. Impact on cell proliferation and cell death may be analysed using using any suitable method available, e.g. colorimetric
35 assays or HT microscopy.

In vivo models for drug testing

Pre-clinical animal testing of TONSL inhibitors with the aim of starting clinical trials.

5 *Small molecules of the invention*

The small molecules of the present invention are useful for impairing homologous recombination. These could also involve both short linear and cyclic peptides and peptidomimetics, whose low molecular weights allows cell permeability.

- 10 By use of the structural data presented in the present application and the proof of concept assays above, the present inventors have identified pharmacophore targeted molecules, including small molecules, covalently linked small molecules, peptides or cyclic counterparts, interfering with the adjacent histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL.

15

Structure properties of the small molecule inhibitor

The present invention relates to small molecule inhibitors, which target the conformational space of the TONSL ARD occupied by the histone H4 tail encompassing residues K12-R23 and act to disrupt the binding of the H4 tail K12-R23 with the
20 TONSL ARD via direct competition or via allosteric disruption of the binding pocket.

- In one embodiment, the present invention relates to a small molecule, which targets or interferes with the conformational space of the TONSL ARD occupied by the histone H4 tail encompassing residues K12-R23 and acting to prevent or disrupt the binding of
25 the H4 tail K12-R23 with the TONSL ARD via direct competition or via allosteric disruption of the binding pocket.

- As the skilled addressee would know, it's apparent that a small molecule that could bind to the His18-binding pocket or the Lys20-binding pocket or both pockets on
30 TONSL ARD, would disrupt the interactions between H4 and TONSL.

In one embodiment, the present invention relates to a small molecule that targets the H4 tail spanning residues Lys12 to Arg23 through intermolecular hydrogen-bonding, electrostatic and/or van der Waals interactions.

In a more preferred embodiment, the present invention relates to a small molecule, wherein the molecule targets the intermolecular contacts spanning the Lys12-Gly13-Gly14-Ala15 segment of H4.

5

In another preferred embodiment the present invention relates to a small molecule, wherein the molecule targets the hydrophobic interactions between residues Gly13, Gly14 and Ala15 of H4 and residues Asn507, Cys508, Trp641, Tyr645 and Leu649 of ARD.

10

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets the hydrogen bonds between the main-chain O of H4 Gly14 and N ϵ 1 of ARD Trp641, and between the main-chain N of H4 Ala15 and O η of ARD Tyr645.

15

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets the main-chain O of H4 Lys16 hydrogen bonds with the N δ 2 of ARD Asn571.

20

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the side-chain of H4 Arg17, which stacks over the side-chains of ARD Tyr572 and Cys608, while its N η 1 atom forms two hydrogen bonds with main-chain O and O δ 1 of ARD Asn571. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of associating with the side-chains of ARD Tyr572 and Cys608, and forming hydrogen bonds with the main-chain O and O δ 1 of ARD Asn571.

25

30

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the side-chain of H4 H18, which penetrates into a pocket lined by four strictly conserved residues (Trp563, Glu568, Asn571 and Asp604) and is positioned over His567 of ARD. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of associating with Trp563, Glu568, Asn571, Asp604 and His567 of TONSL ARD.

35

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD

occupied by the side chain of H4 His18, which is stacked between Trp563 and Asn571 and forms hydrogen bonds to Glu568 and Asp604 of ARD. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of associating with Trp563 and Asn571 and forming hydrogen bonds to Glu568 and Asp604 of
5 TONSL ARD.

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the main-chain O of H4 Arg19 that forms a hydrogen bond with Nε1 of
10 Trp563 and its side-chain forms contacts with Cys561 and Gly595 of ARD. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of associating with the Nε1 of Trp563 and with Cys561 and Gly595 of TONSL ARD.

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the H4 Lys20 residue, which is bound within an acidic surface pocket on
15 ARD adjacent to the H4 His18 binding pocket.

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by side-chain of H4 Lys20, which interacts with the side-chain of Met528 and contacts the edge of Trp563 of ARD, while the main-chain atoms of H4 Lys20 packs against Cys561 of ARD. Thus, in one embodiment the invention relates to a
20 compound, e.g. a TONSL inhibitor capable of associating with the side-chain of Met528 and Trp563 and Cys561 of TONSL ARD.
25

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the Nζ atom of H4 Lys20, which forms three strong hydrogen bonds
30 (distance < 3 Å) with the side-chains of strictly conserved residues Glu530, Asp559 and Glu568 of ARD, which surround H4 Lys20 within a regular triangle-like alignment. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of forming hydrogen bonds with the side chains of Glu530, Asp559 and Glu568 of of TONSL ARD.

35

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets the intermolecular contacts spanning the Val21-Leu22-Arg23 segment of H4, which include contacts between side-chains of H4 Val21 with Tyr560 and Cys561 of ARD, while H4 Leu22 interacts with Asp527 and Met528 of ARD.

5

In one embodiment, the present invention relates to a small molecule, wherein the molecule targets and/or associates with the conformational space of the TONSL ARD occupied by the main-chain N of H4 Arg23, which forms a hydrogen bond with the main-chain O of Asp527 of ARD, while the side-chain packs against the side-chain of Tyr560 of ARD. Thus, in one embodiment the invention relates to a compound, e.g. a TONSL inhibitor capable of forming hydrogen bond(s) with the main-chain O of Asp527 of ARD, while the binding the side-chain of Tyr560 of TONSL ARD.

10

In one embodiment, the present invention relates to a small molecule according to the present invention capable of blocking histone reader domains in a protein selected from the group consisting of TONSL, BARD1 and ANKRD11.

15

Such molecules are typically identified and validated by the SCHRODINGER package to identify and validate covalently linked small molecule drugs targeted to the adjacent histone H4H18 and H4K20 binding pockets on the surface of the Ankyrin repeats of TONSL.

20

Relationship with BARD1

The topology of TONSL ARD domain including the histone-binding surface is highly similar to the ARD of BARD1 (Figure 23). BARD1 is an obligate binding partner of BRCA1 and is required for most cellular and tumour-suppressor functions of BRCA1. The key residues involved in histone H4 binding are conserved in ARD of BARD1. Furthermore, N571 residue of TONSL ARD that is essential for histone H4 binding (Figures 10, 12) and recruitment of TONSL to chromatin (Figure 21) and sites of DNA damage (Figure 22) and viability (Figure) corresponds to the BARD1 N470S cancer mutation (Figure 23). Therefore, small molecule inhibitors against TONSL ARD are predicted to target BARD1 ARD domain as well. Given that both TONSL and BARD1 function in the same DNA repair pathway, small molecule inhibitors designed based on our invention that would target both TONSL and BARD ARD would predictably inhibit HR more efficiently and show more potent anti-cancer activity as compared to inhibitors specific for TONSL ARD only.

25

30

35

General

It should be understood that any feature and/or aspect discussed above in connections with the compounds according to the invention apply by analogy to the
5 crystal structures and/or methods described herein.

The atomic coordinate's variation of the crystal structures of the present invention such e.g. 1Å up to 3Å are used interchangeably.

10 The following sequence data, structures, figures and examples are provided below to illustrate the present invention. They are intended to be illustrative and are not to be construed as limiting in any way.

BRIEF DESCRIPTION OF THE FIGURES

15 *Figure 1*

View of the overall structure of the TONSL-ARD-MCM2 HBD-H3-H4 tetramer complex showing the relative positions of two TONSL ARDs, two MCM2 HBDs and an H3-H4 tetramer.

Figure 2

20 TONSL ARD consists of four ANK repeats and uses its elongated concave surface to target the H4 tail spanning residues 12 to 23.

Figure 3

Intermolecular interactions between TONSL ARD and the H4 tail (12-23).

Figure 4

25 The electrostatic potential surface of ARD showing the acidic concave surface-binding site for the H4 tail.

Figure 5

Highlight of the inter-molecular interactions of H4 Arg 17 with TONSL ARD.

Figure 6

30 Highlight of the inter-molecular interactions of H4 His18 with TONSL ARD.

Figure 7

Highlight of the inter-molecular interactions of H4 Lys20 with TONSL ARD.

Figure 8

Molecular details of the interactions of TONSL ARD with H4 tail region residues 12-15.

Figure 9

- 5 Molecular details of the interactions of TONSL ARD with H4 tail region residues 21-23.

Figure 10

In vitro pull-down of recombinant histones H3-H4 with recombinant GST-TONSL ARD wild type or indicated mutants.

Figure 11

- 10 Immunoprecipitation of HA-SNAP-H4 wild-type (WT) and indicated mutants transiently transfected into GFP-TONSL U-2-OS cells.

Figure 12

Immunoprecipitation of GFP-TONSL WT or indicated mutants from transiently transfected HeLa S3 cells.

- 15 *Figure 13*

ITC analysis of TONSL ARD binding to H4 tail peptides with the indicated modifications.

Figure 14

In vitro pull-down of recombinant TONSL ARD with biotinylated H4 tail peptides.

- 20 *Figure 15*

In vivo pull-down of GFP-TONSL from cell extracts with biotinylated H4 tail peptides

Figure 16

- 25 In vitro pull-down of recombinant TONSL-MMS22L heterodimer with biotinylated recombinant mononucleosomes, unmodified or di-methylated at K20. * indicates an unspecific band. (bottom) TONSL binding quantified relative to histones. Unpaired t-test: *, $P < 0.05$; mean of 6 independent experiments is shown; whiskers, outliers.

Figure 17

Pull-down of GFP-TONSL from cell extracts with biotinylated H4 tail peptides

Figure 18

Immunoprecipitation of GFP-TONSL from solubilized chromatin of GFP-TONSL U-2-OS cells

Figure 19

5 Immunoprecipitation of GFP-TONSL from solubilized chromatin of GFP-TONSL U-2-OS cells

Figure 20

High-content quantitative imaging of TONSL in pre-extracted U-2-OS fibroblasts. Plots show total TONSL and DAPI intensities in cells treated with control or TONSL siRNAs, confirming the specificity of TONSL antibody. Dots represent measured intensities of
10 individual cell nuclei

Figure 21

TONSL recruitment to chromatin requires recognition of H4K20me0. (a) U-2-OS cells conditional for GFP-TONSL WT or N571A were directly fixed or pre-extracted to remove soluble proteins. Scale bar, 20 μ m. (b) (left) Analysis of GFP-TONSL WT and
15 mutants by cellular fractionation. (right) The chromatin-bound fraction (C) was quantified by western blotting relative to total GFP-TONSL (error bar, SD; $n \geq 3$). S – soluble fraction, C- chromatin bound fraction. Staining with MemCode (ThermoFisher) was used to control the protein loading

Figure 22

20 TONSL recruitment to DNA lesions requires recognition of H4K20me0. (top) U-2-OS cells conditional for GFP-TONSL WT or N571A were laser-irradiated and co-stained for 53BP1 and cyclin B to mark DNA damage and identify S/G2 cells, respectively. Full arrowheads, GFP-TONSL recruitment; empty arrowheads, no recruitment. Scale bar, 10 μ m. (bottom) Quantification of GFP-TONSL cells showing recruitment to laser
25 tracks (error bars, SD; $n=3$)

Figure 23

Superposition of the structure of TONSL ARD and BARD1 ARD. The main residues involved in TONSL ARD interactions with the H4 tail are compared to the corresponding residues of BARD1 ARD. The two ARDs show highly similar topology
30 and conservation of the histone-binding surface

Figure 24

Co-localization analysis of chromatin-bound GFP-TONSL with MCM2 (**A**) and EdU (**B**) in inducible cell lines. Cells were either pulsed with EdU (left) or released into S phase in

continuous presence of EdU (right). Co-localization analysed by deconvolution microscopy and measurement of Pearson coefficient in single cells ($n > 15$, two independent experiments). Representative images,

Figure 25

- 5 Chromatin-binding of GFP-TONSL WT and ARD mutants analysed across the cell cycle (representative of 2 experiments).

Figure 26

Chromatin-binding of GFP-TONSL ARD WT and mutant analysed by high-content imaging of pre-extracted inducible U-2-OS cells treated with CPT (3 hours, 1 μ M CPT) as indicated (representative of 3 experiments)

- 10

Figure 27

- 15 A) Co-immunoprecipitation analysis of TONSL-MMS22L with Flag-HA-MCM2 WT or histone binding mutant of MMS22L(Y81A, Y90A) isolated from solubilized chromatin from HU-treated cells (2 hours, 3mM) (representative of 2 experiments).
- B) Analysis of GFP-TONSL ARD WT and mutant recruitment to CPT-challenged replication forks by NCC as illustrated (CPT, 1 μ M). (-), no biotin-dUTP.

Figure 28 and 29

- 20 Colony formation upon GFP-TONSL expression induced by tetracycline in TONSL depleted (**fig. 28**) and control cells (**fig. 29**). Cells were siRNA treated and induced to express siRNA-resistant GFP-TONSL ARD WT and mutant by tetracycline (tet) and treated with CPT (24 hours, 50nM) as indicated. Rescue efficiency determined by colony forming efficiency of TONSL depleted cells (A). Dominant negative effect of TONSL ARD mutant determined by colony forming efficiency of cells treated with control siRNA (**B**). Data points represent two different cell concentrations in technical triplicate from two or four (28, left panel) independent biological replicates.
- 25

Figure 30

- A)** Colony formation upon GFP-TONSL expression induced by tetracycline in TONSL depleted and control cells as shown in Fig. 4f but including additional ARD mutants. Two cell concentrations in technical triplicate from two (E568A, D559A) or four (WT, N571A) biological replicates are shown. **B)** Representation of the complementation analysis from Fig. 4f in a single panel including both CPT treated and untreated cells.
- 30

This illustrates that the toxicity of the TONSL ARD mutant is comparable to CPT treatment of cells expressing WT TONSL.

Figure 31

Cells were treated as in Fig. 29 and analysed for cell cycle distribution using EdU and Dapi (left) and 53BP1 foci (right). Error bars, mean with S.D (four (left) or five (right) biological replicates in technical triplicates).

Figure 32

Chromatin-binding of MMS22L. The chromatin-bound fraction was analyzed by western blotting (**B**) and quantified relative to total MMS22L (**A**) (untreated: error bar, SD, n=3; CPT: error bar, mean with range, n=2).

Figure 33

Pull-down of recombinant histones H3-H4 and MCM histone binding domain (HBD) with GST-TONSL ARD WT or indicated mutants.

Figure 34

Circular dichroism (CD) analysis of TONSL ARD WT and the indicated ARD mutants. The indicated ARD point mutations do not destabilize the overall structure of the ARD.

Figure 35

ITC analysis of TONSL acidic stretch and ARD (aa. 450-692) with H3K9me1 (aa. 1-21) and H4 (aa. 9-25).

Figure 36

Analysis of TONSL chromatin-binding in MOF-depleted cells. Chromatin-bound TONSL was quantified by high content imaging of pre-extracted cells stained for endogenous TONSL. Mean TONSL intensity is shown. A.U., arbitrary units. Knock-down efficiency and expected effect on histone modification were confirmed by western blotting (representative of 2 experiments).

Figure 37

H4K20 methylation levels measured by mass spectrometry in synchronized TIG3 cells. Cell were treated with HU (3 mM) or CPT (1 μ M) from 3 – 6 hours after release or left untreated (6 hours) (error bars, range; n=2).

EXAMPLES

Example 1 – The Crystal Structure

Protein production: All proteins described in the Examples herein, unless otherwise indicated, were expressed in BL21(DE3)-RIL cell strain (Stratagene).

- 5 The GST-tagged TONSL ARD and its mutants including E530A, D559A, W563A, E568A, N571A and D604A were cloned into pGEX-6P-1 vector (GE Healthcare). The expressed proteins were first purified using Glutathione Sepharose 4B, then further purified by gel-filtration step. In some case, the GST-tag was removed with 3C protease before gel-filtration step. For purification of GST-H3 tail and GST-H4 tail proteins, the human
10 histones H3 fragment 1–59 and H4 fragment 1–31 were cloned into pGEX-6P-1 vector respectively. The proteins were expressed and purified in the same way.

- For production of recombinant full-length TONSL-MMS22L heterodimer, the sequence coding for full-length MMS22L was fused with a MBP tag at the 5' end and 10×His tag at the 3' end. The sequence coding for full-length TONSL was fused with
15 GST tag at the 5' end. Both MMS22L and TONSL constructs were cloned into a pFastBac1 vector. The complex was expressed in Sf9 cells by co-infection with both recombinant baculoviruses according to manufacturer's recommendation (Invitrogen). The proteins were extracted from Sf9 cells and purified similarly as described previously for Sgs1. Briefly, the complex was purified on amylose resin, and MBP and
20 GST tags were subsequently cleaved with PreScission protease. The heterodimer was then further purified using a Ni-NTA affinity resin. Washes were performed with 300 mM NaCl buffer.

- For crystallization: The human TONSL Ankyrin Repeat Domain (ARD, residues 512–
25 692) and MCM2 Histone-binding Domain (HBD, residues 61–130) were covalently linked through a four-Glycine linker (G₄ linker) into a single expression cassette (SEQ ID NO: 15).

- The MCM2 HBD–G₄–TONSL ARD expression cassette was cloned into a modified
30 RSFDuet-1 vector (Novagen), with an N-terminal His6-SUMO tag. The resulting plasmid was coexpressed with a pETDuet plasmid harboring human histone genes H3.3(57–135) and H4(1–102) in BL21(DE3)-RIL cell strain (Stratagene).

- The E. coli was cultured at 37 °C using LB media with 50 µg/ml Kanamycin, 100 µg/ml
35 Ampicillin and 34 µg/ml Chloramphenicol.

When the *E. coli* reached cell density of $OD_{600} \sim 1.0$, 0.5 mM IPTG was added into the LB media which was further incubated at 20 °C overnight.

The expressed protein complex was first purified on HisTrap HP column (GE Healthcare). After removing the His₆-SUMO tag by using Ulp1 (SUMO protease), the protein complex was further purified on HiLoad 16/600 Superdex 200 column (GE Healthcare) in the buffer of 20 mM Tris pH 7.5 and 500 mM NaCl.

The purified G4 linker complex, MCM2 HBD-G₄-TONSL ARD cassette-H3.3(57-135)-H4(1-102) complex (herein designated as TONSL ARD-MCM2 HBD-H3-H4 tetramer complex) with a concentration of 23 mg ml⁻¹ in the buffer of 20 mM Tris pH 7.5 and 1 M NaCl, was crystallized in the condition of 100 mM MES pH 5.6-6.6, 5-10% isopropanol using setting-drop vapor-diffusion method at 20 °C. All the crystals were soaked in a cryoprotectant made from the mother liquor supplemented with 25% glycerol before flash freezing in liquid nitrogen.

The data set for the TONSL ARD-MCM2 HBD-H3-H4 tetramer complex was collected at 0.979 Å on 24-ID-C/E NE-CAT (Advanced Photo Source, Argonne National Laboratory). The data was processed using the HKL 2000 program. The initial structure for the complex was solved by molecular replacement in PHASER with our previous structure of the MCM2 HBD-H3-H4 tetramer complex as a search model and manually refined and built using Coot. The final structure of this complex was refined to 2.43 Å resolution using PHENIX. Table 1 summarizes the statistics for data collection and structural refinement

Table 1. Data collection and refinement statistics

TONSL ARD-MCM2 HBD-H3-H4 Tetramer Complex	
Data collection	
Space group	P3 2 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	139.5, 139.5, 72.9

α, β, γ (°)	90, 90, 120
Resolution (Å)	50-2.43 (2.95-2.43) ^a
R_{pim} (%)	3.8 (46.8)
$I/\sigma I$	23.1 (1.8)
Completeness (%)	99.8 (99.7)
Redundancy	5.5 (5.5)

Refinement

Molecules per asymmetric unit	1
No. reflections (total/unique)	171,308/31,146
$R_{\text{work}}/ R_{\text{free}}$ (%)	20.1/24.6
No. atoms	
Protein	2,908
MES	12
Glycerol	12
Water	87
B-factors	
Protein	81.8
MES	108.5
Glycerol	92.6
Water	59.8
R.m.s deviations	

Bond lengths (Å)	0.009
Bond angles (°)	1.316
Ramachandran plot ^b	
Favored (%)	95.9
Allowed (%)	4.1

^a Highest resolution shell is shown in parenthesis.

^b Calculated using MolProbity in PHENIX.

5 The structural data obtained are described below in Annex 1.

One molecule of each protein MCM2 HBD–TONSL ARD cassette, H3 and H4 is present in the asymmetric unit. The crystallographic symmetric operation reconstitutes a tetramer of H3–H4, thus resulting in formation of an intact complex with two copies of

10 MCM2 HBD–TONSL ARD cassette in complex with an H3–H4 tetramer (named TONSL ARD–MCM2 HBD–H3–H4 tetramer complex), which is consistent with our previous finding that MCM2 HBD binds and stabilizes an H3–H4 tetramer under physiological conditions. The TONSL ARD–MCM2 HBD–H3–H4 tetramer complex is also highly similar to our previous structure of the MCM2 HBD–H3–H4 complex, with a pair of

15 MCM2 HBDs wrapping around the lateral surface of the H3–H4 tetramer, while the two TONSL ARDs interact with each of the H4 tails. The TONSL ARD forms no intermolecular interactions with the MCM2 HBD, consistent with the H3–H4 tetramer bridging the interaction between TONSL and MCM2 in cells. The TONSL ARD forms extensive contacts with a segment of the H4 tail (residues 12–23), but shows only

20 minimal contacts with the core of the H3–H4 tetramer. The TONSL ARD could be modeled to bind the H4 tail in the context of the nucleosome without steric clashes and a conserved positive patch may interact with the nucleosomal DNA, suggesting that the extensive interactions between TONSL ARD and the H4 tail could describe TONSL binding to both soluble non-nucleosomal histones H3–H4 together with MCM2

25 and also to H3–H4 in nucleosomes.

The Ankyrin (ANK) repeat fold contains two antiparallel helices named the inner and outer helix respectively, followed by a hairpin loop named the finger. The TONSL ARD consists of four ANK repeats, three of which adopt the canonical ANK repeat fold

30 (ANK1-3), while the remaining one is an atypical and capping repeat (ANK4). Besides

the internal fingers 1-3, the TONSL ARD contains an extra loop preceding ANK1, designated finger 0. The TONSL ARD uses its elongated concave surface composed of inner helices (α A1, α B1, α C1 and α D1) and fingers 0-4 to form extensive intermolecular contacts with the extended β -strand like conformation of the H4 tail. It is notable that 15 out of 18 residues that constitute the H4 tail-binding surface of TONSL ARD are highly conserved. The TONSL ARD targets the H4 tail spanning residues Lys12 to Arg23, primarily through intermolecular hydrogen-bonding, electrostatic and van der Waals interactions.

10 *Example 2 - Determination of TONSL ARD binding to histone H4 peptides by Isothermal Titration Calorimetry (ITC)*

All the ITC titrations were performed on a Microcal ITC 200 calorimeter at 25 °C. The peptides of histone H4 (residues 9-25) and its modified peptides K16ac (with acetylation on Lys16), H18W (with His18 mutated to Trp18), H4K20me1 (mono-methylation on Lys20) and H4K20me2 (di-methylation on Lys20), and peptide of histone H3(1-19)K9me1 (mono-methylation on Lys9) were all synthesized at Tufts University Core Facility. The exothermic heat of the reaction was measured by sequential 2.2 μ l injections of the peptides (1.41 mM in buffer 20 mM Tris pH 7.5 and 0.5 M NaCl) into 200 μ l of the TONSL ARD solution (145 μ M in the same buffer), spaced at intervals of 150 s. The data were processed with Microcal Origin software and the curves were fit to a single site binding model.

Example 3 – Assaying binding of recombinant TONSL ARD to histone H4 peptides in vitro

25 Purified recombinant TONSL ARD (residues 512-692) was stored at 400 μ M concentration in 1 M NaCl, 20 mM Tris HCl pH 7.5 at -80 °C. For each pull-down, 400 pmol of the ARD stock (1 μ l, 400 μ M) was diluted with 99 μ l of binding buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 5 % Glycerol, 0.25 % NP-40, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 1 mM Leupeptin, 1 mM Pepstatin). ARD input material was scaled to the number of pull-downs performed. For each pull-down, a histone H4 peptide (JPT Peptide Technologies GmbH) spanning residues 14-33 (2.5 μ l, 250 μ M) with a C-terminal biotinoyl-lysine residue or biotin (2.5 μ l, 400 μ M) was added to 1.1 ml of binding buffer in addition to 100 μ l of the ARD input material and the mixture incubated overnight rotating at 4 °C. The next day 25 μ l of MyOne Streptavidin C1 beads (Life Technologies) was washed in binding buffer (3x 500 μ l) for each pull-down

removing the final wash from the beads. The ARD/peptide or ARD/biotin mixture was added to an aliquot of pre-washed MyOne Streptavidin C1 beads and incubated with rotation at 4 °C for 3 hours. Finally the beads were washed (2x 300 µl and 1 x 200 µl of 300 mM NaCl, 50 mM Tris HCl pH 7.5, 5 % Glycerol, 0.25 % NP-40, 0.2 mM EDTA, 5 0.5 mM DTT, 0.2 mM PMSF, 1 mM Leupeptin, 1 mM Pepstatin) and pulldown material visualized by Coomassie staining after SDS PAGE separation of proteins on a NuPAGE 4-12 % gel.

10 *Example 4 - Measuring GFP-TONSL binding to histone H4K20me0 peptides in cell extracts*

For detergent-soluble extracts (NP40/NaCl), U-2-OS cells expressing GFP-TONSL WT or ARD mutants were washed with cold PBS, scraped and incubated for 15 min on ice in HS buffer supplemented with trichostatin A (TSA) and protease and phosphatase inhibitors (5 mM sodium fluoride, 10 mM µ-glycerolphosphate, 0.2 mM sodium 15 vanadate, 10 µg/ml leupeptin, 10 µg/ml pepstatin, 0.1 mM PMSF, Sigma). After centrifugation at 16.000 g for 15 min at 4°C, the supernatant was collected. For pull-downs from cell extracts, MyOne T1 beads were incubated O/N with 1 µg of biotinylated histone peptides in High Salt (HS; 300mM NaCl, 0.5% NP40, Tris HCl, EDTA, 5% glycerol) buffer and subsequently washed 2 times with PBS. 1 mg of 20 NP40/NaCl extract from GFP-TONSL U-2-OS cells was added to the beads and incubated for 2 hrs rotating at 4°C. The beads were then washed 5 times with HS buffer, 2 min rotating at 4°C. After washing, the beads were resuspended in 1XLSB and boiled for 10 min. The eluted proteins were loaded on a 4-12% Bis-Tris NuPage gel (LifeTechnologies). Proteins were then transferred to a 0.2 µm nitrocellulose 25 membrane by O/N wet transfer at 20V and detected by western blotting.

Example 5 - Quantifying chromatin-bound TONSL in human cells by HT microscopy

U-2-OS and TIG3 cells were grown on 6-well plates (1x10⁵ cells seeded/well 1 day 30 prior to analysis). To analyse only chromatin-bound TONSL, soluble proteins were removed (pre-extracted) by 5 min incubation on ice with CSK buffer 0.5% Triton (CSK buffer: 10 mM PIPES pH 7, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂ plus protease and phosphatase inhibitors 1 mM DTT, 10 ug/ml leupeptin, 10 ug/ml pepstatin, 0.1 mM PMSF, 0.2 mM sodium vanadate, 5 mM sodium fluoride, 10 mM beta-

glycerolphosphate). Cells were then rinsed with CSK and PBS before fixation in 4% formaldehyde for 10 min and staining with TONSL antibody.

For detection, cells were blocked with PBS containing 5% BSA and 0.1% Tween20 for 1 h, and incubated with TONSL antibody (Sigma, ref. nr. HPA0244679) overnight at 4°C (1:400 in blocking buffer). After washing 3 times with PBS containing 5% BSA and 0.1% Tween 20, anti-rabbit-Alexa488 (LifeTechnologies 1:1000 in blocking buffer) was applied and let to incubate for 30 min. Cells were washed 3 times and DNA was counterstained with DAPI (Sigma). Images were acquired on ScanR high-content imaging system (Olympus) and analysed using ScanR software. Relative fluorescence intensity TONSL was quantified relative to cells depleted for TONSL using the specific siRNAs (O'Donnell et al. 2010 Mol Cel 40:619, synthesized by Sigma) for 30 h at 100 nM concentration.

Example 6

Given that TONSL-MMS22L binds histones in a pre-deposition complex with ASF1 and MCM2, TONSL-MMS22L could be loaded onto replicating DNA together with new histones. It is shown herein that in nascent chromatin, new histones were exclusively unmethylated at H4K20 (98% H4K20me0), while old recycled histones were almost fully methylated at H4K20 (me1, 7%; me2, 88%; me3, 2%). New histones became methylated in late G2/M, rendering G1 chromatin devoid of H4K20me0. This identifies H4K20me0 on new histones as a signature of post-replicative chromatin, implying that TONSL-MMS22L can bind H4 tails on new histones at replication forks and sister chromatids until late G2/M. Confirming this prediction, TONSL accumulated on chromatin in S phase, remained chromatin-bound in a population of G2 cells and was excluded from chromatin in G1 (Fig. 20). To discriminate pre- and post-replicative chromatin, we labeled replicating DNA with EdU (pulse to mark ongoing replication, continuous labeling to identify post-replicative chromatin) and marked pre-replicative chromatin with MCM2, and analyzed co-localization with TONSL. TONSL staining was mutually exclusive with MCM2 (Fig. 24A), but co-localized with EdU pulse labeling in very early S phase and with replicated DNA (continuous EdU labeling) throughout S phase (Fig. 24B). TONSL was present at sites of ongoing DNA replication throughout S phase, but the degree of co-localization declined in mid/late S (Fig. 24B, left panel), consistent with TONSL binding to post-replicative chromatin also after fork passage (Fig. 24B, right panel). Mutation of the TONSL ARD abrogated recruitment of TONSL to chromatin, including DNA replication sites (Fig. 25,). Together, this demonstrates

that TONSL is recruited to replication forks and post-replicative chromatin via ARD recognition of H4K20me0 on new histones. TONSL ARD recognition of the H4 tail is required for binding to post-replicative chromatin and recruitment of damaged forks and DNA lesions.

5 Mutation of TONSL ARD also abrogated chromatin binding (Fig. 26) and recruitment to replication forks in the presence of replication poisons like camptothecin (CPT) and hydroxyurea (HU) (Fig. 27A and B). Furthermore, ARD mutation prevented accumulation of TONSL at site-specific DSBs (Fig. 28) and microlaser-generated DNA damage. Co-staining with cell cycle markers confirmed that TONSL is recruited to DNA
10 repair sites only in S and G2 cells as expected. H4K20me0 binding is required for TONSL accumulation at damaged forks and DNA lesions in post-replicative chromatin. However, this was not due to increased H4K20me0, suggesting that unmasking of H4 tails upon chromatin decompaction and/or interaction with repair factors contribute to TONSL-MMS22L accumulation at repair sites. Consistent with the latter, MMS22L
15 interaction with Rad51 can stabilize the complex at challenged forks (P. Cejka and M. Peter, personal communication). Our data argue that this is subsequent to H4K20 binding (Fig. 26 to 28). In complementation analysis, TONSL WT partially rescued viability of TONSL depleted cells in the presence and absence of CPT (Fig. 28, Fig. 30A and B), whereas TONSL ARD mutants were highly toxic (Fig. 28, Fig. 30A and B). In
20 control cells, TONSL ARD mutants also reduced viability and caused G2/M arrest accompanied by replication-associated DNA damage (Fig. 29 and Fig. 31). Further, the TONSL ARD mutant titrated MMS22L away from chromatin, explaining the dominant negative phenotype that mimics TONSL-MMS22L depletion (Fig. 32A and B). Collectively, this indicates that recognition of H4K20me0 is central to TONSL-MMS22L
25 function in safeguarding genome stability.

The cell viability was assayed by clonogenic assay (see Example 8 herein below for details of the assay) using inducible U2OS cells that express siRNA resistant GFP-TONSL WT or ARD mutant (D559A, E568A, N571A), in the presence or absence of CPT. Aforementioned results shows that GFP-TONSL WT can partially complement the
30 growth defect and CPT sensitivity of TONSL depleted cells, while expression of GFP-TONSL ARD mutants further impairs cell viability and survival to CPT. Also, it show that expression of GFP-TONSL ARD mutants in control cells has a dominant negative effect on cell growth.

Accumulation of the genome instability marker 53BP1 and cell cycle arrest in G2/M in
35 cells expressing GFP-TONSL ARD mutants was determined. Inducible U2OS cells were

- induced to express resistant GFP-TONSL WT or ARD mutant (D559A, E568A, N571A) for 24 hours and then fixed and stained for 53BP1 and EdU after an additional 24 hours. Images were acquired by high-throughput microscopy. This shows that expression of the TONSL ARD mutants phenocopies TONSL/MMS22L depletion.
- 5 Accordingly, these mutants are useful for testing the effect of inhibitors of TONSL in various disease models. The demonstrated titration of MMS22L away from chromatin in cells expressing GFP-TONSL ARD mutants, both in the absence and presence of CPT may explain why expression of TONSL ARD phenocopies TONSL/MMS22L depletion.
- 10 It is revealed that post-replicative chromatin has a distinct histone modification signature, read by the TONSL-MMS22L effector protein. This opens a new avenue to understand how DNA repair and other chromosomal transactions can be directly linked to the replication state of a genomic locus. Intriguingly, it is the new histones that make post-replicative chromatin distinct. It is presented that TONSL-MMS22L is
- 15 delivered to nascent chromatin with new histones via the pre-deposition complex with MCM2 and ASF1. TONSL may thus have a dual function as a histone chaperone and histone reader. The structural work proposes that TONSL acts in a histone chaperone-like capacity by sequestering the H4 tail to prevent spurious contacts with DNA during H3-H4 deposition. Further, TONSL ARD may counteract chromatin compaction by
- 20 preventing association of the H4 tail with the H2A-H2B acidic patch on neighboring nucleosomes. Thus, TONSL changes our perception of a histone chaperone by binding both soluble and nucleosomal histones. In its function as a histone reader, TONSL localizes MMS22L to post-replicative chromatin via H4K20me0 and allows TONSL-MMS22L to accumulate at damaged forks and DNA lesions. We envision that
- 25 H4K20me0 works as an affinity trap, making TONSL-MMS22L readily available to support Rad51 loading during HR. This provides a new angle to understand the role of H4K20 in DNA repair, complementing the well-described role of H4K20me1/2 in recruiting 53BP1 to promote NHEJ in competition with BRCA1-BARD1. In post-replicative chromatin, H4K20me1/2 on old histones will support 53BP1 recruitment.
- 30 Whether H4K20me0 on new histones also influences DNA repair pathway choice will be of interest to future investigations. It is notable that the structure of the TONSL ARD, including the histone-binding surface, is highly similar to the ARD of BARD1, required for BRCA1 tumor suppressor function and HR. Multiple mutations in the TONSL ARD are reported in cancer (C608G, COSM4879909; P557S, COSM4565032;
- 35 E597K, COSM3382163) and the N571 residue, key to H4 binding, corresponds to the BARD1 N470S cancer mutation. This underscores that the tumor suppressor function

of H4K20me0 recognition and the possibilities it brings for targeted cancer therapy should be explored in the future.

Circular dichroism analysis of TONSL ARD WT or mutants stability (fig. 34), show that the indicated ARD point mutations do not destabilize the overall structure of the ARD.

5 ITC analysis of TONSL acidic stretch and ARD (450-692) with H4 tail (9-25) and H3K9me1 (1-21) peptides (Fig. 35) show that TONSL (450-692) does not bind to H3K9me1, but recognizes H4 peptides.

The methods used for immunofluorescence, microscopy and laser microirradiation are described below in Example 7.

10

Example 7

Immunofluorescence, microscopy and laser microirradiation

U-2-OS, HeLa S3, and TIG-3 cells were grown in DMEM (Gibco) containing 10% FBS (Hyclone) and 1% penicillin/streptomycin and drugs for selection. The construct for
15 siRNA resistant GFP-TONSL was described (O'Donnell et al., Mol Cell **40**, 619-631 (2010)) and ARD mutation were introduced in this construct by site-directed mutagenesis. Cells inducible for GFP-TONSL WT and ARD mutants were generated in Flp-In T-Rex U-2-OS cells (Invitrogen) by transfection of pcDNA5/FRT/TO-GFP-TONSL plasmids with Lipofectamine 2000, according to the manufacturer's protocol, and
20 selection with hygromycin (200 µg/ml). All cell lines were authenticated by western blotting and immunofluorescence. Expression of GFP-TONSL was induced by addition of 1 µg/ml of tetracycline for 24 hours. U-2-OS and TIG3 cells were synchronized by a single thymidine block (2mM) and released into S phase in the presence of 24 □M dCTP. For transient expression of GFP-TONSL, expression plasmids were introduced by
25 transfection with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol and cells harvested 24 hours after transfection. siRNA transfection was performed with RNAiMax reagent (Invitrogen) according to the manufacturer's protocol.

siRNA sequences: siSET8#1: 5'-GUACGGAGCGCCAUGAAGU-3'; siSET8#2: 5'-
30 ACUUCAUGGCGCUCCGUACUU-3'; siMOF#1: 5'-GUGAUCCAGUCUCGAGUGA-3'; siMOF#2: 5'-GUGAUCCAGUCUCGAGUGA-3'; siTONSL: 5'-GAGCUGGACUUAAGCAUGA-3'. All cell lines used in this study tested negative for mycoplasma contamination.

U-2-OS cells conditional for GFP-TONSL were grown on glass coverslips and either directly fixed in 4% paraformaldehyde (PAF) for 10 mins or washed in CSK, pre-
35 extracted 5 min with CSK/0.5% Triton X-100 and rinsed with CSK and PBS before fixation in 4% PAF for 10 mins. Coverslips were mounted on glass slides with Mowiol

mounting medium (Sigma Aldrich) containing DAPI. Fluorescence images were collected on a DeltaVision system with a 40× or 60× oil immersion objective. For deconvolution microscopy, z-stacks were acquired (step of 0.2 μ m), deconvolved and analyzed by SoftWoRX 5.0.0. Pearson coefficient correlation analysis was performed on single cells using SoftWoRX 5.0.0. Brightness and contrast were adjusted using Adobe Photoshop CS6. For high content quantitative analysis, fluorescence images were acquired using an Olympus ScanR high-content microscope and processed on the ScanR analysis software. More than 5000 cells per sample were analyzed. Graphs were generated with TIBCO Spotfire software. For microirradiation experiments, cells grown on glass coverslips were fixed in 4% formaldehyde for 15 min, permeabilized with PBS containing 0.2% Triton X-100 for 5 min and incubated with primary antibodies diluted in DMEM for 1 hour at room temperature. Following staining with secondary antibodies (Alexa Fluor 488, 568 and 647; Life Technologies) for 30 min, coverslips were mounted on glass slides in Vectashield mounting medium (Vector Laboratories) containing the nuclear stain DAPI. For detection of nucleotide incorporation during DNA replication, an EdU-Plus labeling kit (Life Technologies) was used according to the manufacturer's instructions. Confocal images were acquired on an LSM-780 (Carl Zeiss) mounted on a Zeiss-AxioObserver Z1 equipped with a Plan-Neofluar 40×/1.3 oil immersion objective. Image acquisition and analysis was carried out with LSM-ZEN software. Laser microirradiation of cells was performed essentially as described.

Example 8

Clonogenic assay

U-2-OS inducible for GFP-TONSL ARD WT and mutant were transfected with siRNA, trypsinized 24 hours later and seeded in technical triplicates of 1000 or 3000 cells in the presence or absence of tetracycline. After 24 hours, the cells were washed and left in fresh medium for 12-15 days before fixation and staining with MeOH/Crystal Violet. Colony formation efficiency was determined by manual colony counting or quantification of Crystal Violet staining by ImageJ software and normalized to non-induced control.

Example 9

Nascent Chromatin Capture (NCC)

The NCC protocol from (Alabert et al., *Nat Cell Biol* **16**, 281-293 (2014)) was adjusted for adherent U-2-OS cells. CPT (1 μ M) was added 5 minutes prior to b-dUTP labelling and was included in all steps until fixation. Cells were incubated for 5 minutes in a

hypotonic buffer (50 mM KCl, 10 mM Hepes) containing biotin-dUTP and resuspended into fresh cell culture medium for an additional 15 minutes. Cells were fixed 15 minutes in 1% formaldehyde, rinsed twice in PBS and collected by scraping in cold room. Nuclei were mechanically isolated in sucrose buffer (0.3 M sucrose, 10 mM

5 HEPES-NaOH at pH 7.9, 1% Triton X-100 and 2mM MgOAc). Chromatin was solubilized by 28 cycles 30 sec ON, 90 sec OFF in sonication buffer (10 mM HEPES-NaOH at pH 7.9, 100 mM NaCl, 2 mM EDTA at pH 8, 1 mM EGTA at pH 8, 0.2% SDS, 0.1% sodium sarkosyl and 1 mM phenylmethanesulphonylfluoride) using a Bioruptor at 4° C. Solubilized chromatin was pre-cleared using streptavidin-coated magnetic beads

10 (MyC1 Streptavidin beads) pre-incubated with biotin. b-dUTP labelled chromatin was next purified over night at 4° C using streptavidin-coated magnetic beads. Beads were washed 5 times for 2 minutes in wash buffer (10 mM HEPES-NaOH pH 7.9; 200 mM NaCl; 2 mM EDTA pH 8; 1 mM EGTA pH 8; 0.1% SDS; 1 mM PMSF). Total chromatin (input) and isolated nascent chromatin were boiled 40 min on beads in LSB 1X (50

15 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 8% Glycerol, Bromophenol blue) and separated by SDS-PAGE for western blotting. An alternative method which may be used in place of NCC is iPOND (Sirbu et al., Nat Protoc 3, 594-605 (2012)).

Example 10

20 **Native MS analysis of protein-peptide complexes**

Various peptides were obtained including the peptides described in Examples 11, 12, 13, 14 and 15. If the peptide was in the form of a trifluoroacetate salt (TFA salt), it was desalted on a polymeric weak cation exchange sorbent (Strata X-CW, Phenomenex, Sartrouville, France): Dry peptide (1-3 µmol) were solubilized in 200 µL

25 pure water (Sigma-Aldrich, Lyon, France). The pH of the solution was adjusted to 7-7.5 by addition of 1 mM aqueous ammonium hydroxide solution (solution freshly prepared from 20% concentrated ammonium hydroxide). The exchange column was conditioned with 1 mL methanol (Sigma-Aldrich, Lyon, France) followed by 1 mL water. The solution was injected and the column was washed twice with 500 µL pure

30 water. Finally the peptide was eluted by gravity with 2 times 500 µL of a 20:80:5 acetonitrile/water/formic acid mixture. The elution fraction containing the expected peptide was identified by mass spectrometry. The TFA-free solution was freeze-dried to get a white solid. The solid was dissolved in 300 µL pure water and the solution was freeze-dried for a second time.

35 A stock solution of the peptide was prepared by adding 300 µL of pure water to the resulting solid and the concentration of the solution was measured by UV at a wavelength of 205 nm. The extinction coefficient for the corresponding peptide

sequence was calculated using the Nick Anthis online protein parameter calculator script (<http://nickanthis.com/tools/a205.html>). The absence of TFA adducts was confirmed by mass spectrometry with infusion of the protein-peptide mixture at low acceleration voltage (V_c 25 V) in native conditions.

- 5 Purified TONSL ARD (i.e. amino acid 512-692 of TONSL of SEQ ID NO:16) was buffer exchanged against 500 mM NH_4OAc pH 7.5 using a NAP 5 column (NAPTM-5, GE Healthcare). Bradford protein quantitation was performed on buffer exchanged protein. Protein was diluted in water to a final concentration of 10 μM and kept on ice until native MS analyses were performed.

10

The mass spectrometry (MS) analysis were carried out on an electrospray time-of-flight mass spectrometer (LCT, Waters, Manchester, UK) equipped with an automated chip-based nanoESI device (Triversa Nanomate, Advion Biosciences, Ithaca, NY). External calibration was done in the positive ion mode over the mass range m/z 200-
15 6000 using the multiply charged ions produced by 0.4 μM horse heart myoglobin solution diluted in water/acetonitrile 50/50 mixture acidified with 0.5% (v/v) formic acid. Homogeneity and purity of TONSL-ARD was first checked under denaturing conditions by diluting the protein to 1 μM in 50/50 water/acetonitrile mixture acidified with 0.5% (v/v) formic acid. Mass measurement revealed mainly the presence of a
20 species with a molecular weight of 20747.18 ± 0.34 Da. Characterization of peptide binding to TONSL-ARD under native conditions was performed in 50 mM NH_4OAc pH 7.5 keeping a constant 5% amount of EtOH (v/v). The protein concentration was set to 10 μM and different peptide concentrations ranging from 1 to 5 molar equivalents were added. Incubations were performed at room temperature for 15 min. Mass
25 spectra were recorded using reduced cone voltage ($V_c = 50$ V) and elevated interface pressure ($P_i = 5$ mbar) which correspond to fine-tuned instrumental settings providing sufficient ion desolvation while preserving the integrity of weak non-covalent complexes in the gas phase. Micromass MassLynx 4.1 was used for data acquisition and processing. Native MS analysis of protein-peptide complexes were attempted by
30 incubation of 10 μM TONSL-ARD with several peptide concentrations (10, 20, 30, 40, 50 μM).

35

For data processing under native conditions, protein-peptide complex abundance (%PL) was estimated from the peak heights of 9^+ and 8^+ charge states of free and bound TONSL-ARD, assuming that compound binding does not alter protein response factor. The proportion of 1:1 stoichiometry complex is calculated according to the following equation:

$$\% P1L1 = \frac{I_{P1L1}}{I_{P1L1} + I_P}$$

To estimate the binding affinities (K_d) of peptides for TONSL-ARD, a mixture of protein/peptide was analyzed at a constant peptide to protein ratio with decreasing protein concentration until the lowest detection limit was reached. For data processing under native conditions, protein-peptide complex abundance (%PL) was estimated from the peak heights of 9^+ and 8^+ charge states of free and bound TONSL-ARD. This protein proportion is used to calculate the bound protein, the free protein and the free peptide concentrations. The K_d is calculated by using the following equation for each protein/peptide concentration. The final K_d for each peptide is the average of the K_d at different protein/peptide ratio.

$$K_d = \frac{[P_{free}][L_{free}]}{[PL]}$$

Test compounds

15

Example 11

Leu-Gly-Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-Asp-Asn-Ile-NH₂

(Histone H4 peptide)

Purchased from JPT Peptide Technologies GmbH, Berlin, Germany

20

The following peptides were obtained from Schafer-N Aps, Copenhagen, Denmark and were prepared by using Fmoc-chemistry on chlorotrityl resins, a method known to those skilled in the art.

25 *Example 12* Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂

Example 13 Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂

Example 14 Lys-Gly-Gly-Ala-Lys-Arg-His-Ala-Lys-Val-Leu-Arg-NH₂

30

Example 15 Lys-Gly-Gly-Ala-Ala-Arg-His-Arg-Lys-Val-Leu-Arg-NH₂

Using the assays described in Examples 10, the following values were obtained.

Example	%P1L1	Apparent K_d (μ M)	Titration K_d (μ M)
11			0.077
12	62.6	2.2	1.95
13	89.9	0.1	0.31
14	73.3	1.0	0.96
15	81.4	0.4	not determined*

* A K_d value could not be measured by titration

Example 16

Structure based design of small molecule and peptide inhibitors:

- 5 Using the structural information of the present invention, we conducted a large scale virtual screening of the collective repertoire of world wide vendor chemical libraries of available screening compounds (small molecules and peptides) directly related to drug discovery applications of ligands, which can be developed into potent and efficacious TONSL inhibitors. The screened (in silico) chemical libraries and the virtual screening
10 protocol is described below.

Chemical vendor libraries:

- The 'In Stock' subset of the ZINC database containing 12,782,590 biologically relevant screening molecules, that are stripped for counter ions and assigned tautomers, protonation states, charges and 3D conformations, was obtained from <http://zinc.docking.org> and stored in SDF files.
15

Preparation of the TONSL structure for docking:

- a. The histone H4 tail (K12-D24) and crystallographic water molecules was removed from the TONSL ARD complex structure (see Example 1).
- b. All hydrogen atoms were built and optimized using ICM (Molsoft L.L.C., San Diego, CA, USA)
20
- c. A box of interactions grids (20x20x20 Å) centered around E530, D559, E568 and D604 was calculated using ICMs (Molsoft L.L.C., San Diego, CA, USA) docking tools.
- d. Full flexible 'ligand' docking of the 12,782,590 screening molecules (default parameters) was performed using ICM (Molsoft L.L.C., San Diego, CA, USA).

- e. The chemical structures and predicted binding conformations of the top 0.001% scored compounds with MW < 500 Da, 0.2 < drug-like score < 1, clogP < 5, number of rotatable bonds < 12, compound strain < 10 were manually assessed.
- f. Initially, a total of 49 high scored compounds were acquired and of those 27 were tested for TONSL ARD binding using a standard native MS binding experiment. A protocol for an MS binding experiment is outlined in Example 10.

Discovery of TONSL-ARD hits in primary binding screen

Among the compounds tested in competition binding experiments with an unmodified histone H4 peptide tail (SEQ ID NO), weak binding to TONSL ARD (amino acids 512-692 of SEQ ID NO:16) were observed for at least one chemo type based on a "3-[(3-Aminocyclopentyl)carbonyl]-1H-quinolin-4-one" scaffold. AG100021 (MW 317) was the best binding compound in this first binding experiment forming ~20 % complex at 20 µM.

STRUCTURAL DATA

Annex 1 – Structure data of TONSL-GFP

The structure data including the coordinates of the "Crystal structure of Human TONSL and MCM2 HBDS binding to a histone H3-H4 tetramer" are provided in the PDB database under the PDB ID 5JA4. As used herein the term "PDB ID 5JA4" refers to the PDB ID 5JA4 as deposited with PDB on 11 April 2016. The PDB ID 5JA4 has the DOI: [10.2210/pdb5ja4/pdb](https://doi.org/10.2210/pdb5ja4/pdb), and is accessible at <http://www.rcsb.org/pdb/explore/explore.do?structureId=5JA4>. The structure data including the same coordinates as PDB ID 5JA4 are also provided in Annex 1 of Danish patent application PA 2015 00605 and in Annex 1 of US patent application 62/324,257.

SEQUENCE DATA

SEQ ID NO: 1 - TONSL WT / GFP

5 atggtgagcaagggcgaggagctgttcaacgggggtggtgcccacccctgggtcgagctggacggcgacgtaaacy
gccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctg
caccaccggcaagctgcccgtgccctggcccacccctcgtgaccacctgacctacggcggtgcagtgcttcagc
cgctaccccgaccacatgaagcagcaagcacttcttcaagtcggccatgcccgaaggctacgtccaggagcgca
ccatcttcttcaaggacgagggcaactacaagaccccgcccgaggtgaagttcgagggcgacaccttggtgaa
ccgcatcgagctgaagggcatcgacttcaaggagggaaggcaacatcctggggcacaagctggagtacaactac
10 aacagccacaacgtctatcatatggtccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccaca
acatcgaggacggcagcgtgcagctcgccgacccactaccagcagaacaccccccatcgggcgacggccccgtgct
gctgcccgcacaaccactacctgagcaccacagtcggccctgagcaagacccccaacgagaagcgcgatcacatg
gtcctgctggagttcgtgacggcgccggggatcactctcgcatggacgagctgtacaagggcgcgccaATGA
15 GCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAGCCAAGGCGCAGAGGGCCGGCCAGCGCGCGAAGA
GGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCAC
TGGCAGGAGCTGCAGCTTCGGGGAGCGCGCTGACGACCCCTCTGGGCTGTGCCGTGGCCCCACCGCAAGATCGGAG
AGCGCCTGGCCGAGATGGAGGACTACCGGGCTGCCTTGACGACCCAGCACCAGTACCTGGAGCTGGCACATTC
CCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCCGCACCCACCTGGACATCTATGACCAC
20 TGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGC
TGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGGACCCGCTCTATCTCAACCTGGGCGCTCACCTT
TGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCTTGGCGAGCAGAACCAC
CTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCCAGG
CTATGCGCTGCTTGGAGGGTGCCTGGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCCAGTGCCTG
CGTGGTTATTGCACAGGTCTCTCAAGACCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTAC
25 AGGCTGGGCTCCCAGAAGCCTGTGCAGAGGGCAGCCATCTGTGCAACCTCCAGCATGTGCTGGCAGTGGTCC
GGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCCCTCAGGGTGCCATGGTTCATCTGTGAGCAGCTAGG
GGACCTCTTCTCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTGCTGAG
CTGCTGGACAGACCGGGTGTGAGCGGGCCATCATCCAGTGTCCCTGGCCACCACACTGGCAGACATGAAGG
ACCACCATGGGGCCGTGCGCCACTATGAGGAGGAAGTGAAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAA
30 GACCTGGCTGAACATTGCACTGTCCCGCGAGGAGGCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTCCAG
AAGCGCTCAGCTGTGCCCAGCAGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGCAGCATCTCCATACCCG
TGCAGCTGAGGCTGCAGCCCAGGAGGCCCTTGAGACCGAAACAGACTGCGGGAGCTCAGTGTAGCTGAAGA
TGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCCCGCGAGGTG
35 GAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCC
ACCTGGGGCCGGCGAAGGGGAGCAAGTGAACCGGCGAAACGACATGGGGCGACCCCTGCTGCACCGAGCCTG
CATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGCCACCCCCCTTAACCCCTCGGGACTACTGT
GGCTGGACACCTCTGCACGAGGCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGG
40 CCGCAGTGGACGACCCAGGTGCCCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGGGCTCAACTTGGCCA
CTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTCACCCCTCCGCACTCGAAAAGGGCCTCAGCCCCG
CTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCCGAGGGACCTGGACCTGGAGACGCGGCAGAAAGGCCAGGG
CCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAAGATCCCCACAGCTCCCAGGCCTTCCACACCCCCAAG
CAGCCTTCTGTTTGACCCGAGACCTCTCCTCCTTTGAGCCCCCTGCCCAGAACCCCCCTCTAATAGCACTAGA
CTCCCAGAGGCCCTCTCAGGTCCATGTTCAGGCTCTCCCCAGGCGAGCGGCACCAAGCCATGGCCAGGCTCGGA
45 GGAGCAGGATGGGCGAGCCAGCAGCAGCAGCTCAGAAAGGCGAGGACGCGAGGACGCGCCAGCCGCGCTC
CCAGAAGAGGCCCTCGGTGCTCGGCCACAGCACAACGGGTGGCAGCCTGGACGCTGGCCCCCGCCAGCAACAGG
GAAGCAGCCACAGCCAGCACCAGCCGGGCGAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGTGCTCAGAGCC
GGCTGGGGCCTGGGCCACCGCGGGGCCACAGCAAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGAGGAGGA
GTGCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGAGCCGCGGGCCCCCGCCCCGGGGCAC
50 GGAGACAACCGCAGGCCCCAGTAGTACCTCTGGGTGCGACAGTGAAGAGAGCAGGCCCCGTGCCCGAGCCAAAGC
AGGTCCGCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCGAGGGCCAGCAGCCTGGCTTCAGAACCC
TCCAGGGAGCCCCAGCACCCTCAGAGGCTCTCAGAGCCAGTGGGGACAGCTCTGCGGCAGGCCAGCCCTTGGGT
CCGGCCCCCGCCCCCTCCCATCCGGTTCCAGTTCAAGTTACAGATCATCTCTTCTCATCCCTGTCCCACACA
GCAGTGACACCCACTCTGTGGCCTGGCTGCGCGAGCAGCGCGGCCAGCGCTACTACCAGACCTGCGGGCTGCT
55 GCCCAGGCTCACCTACGGAAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGC
AATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCCGTTGACTGACCGCTACCGCAGGGCCTGCC
AGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTGCGAGCGCTGGAGCTCCAGGGCTGGGCTCTCGTTTCAG
CGCCTGCTCCCTGGCCCTGGACAGGCCCCAGCTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTC
CGGGAGCTGCGCCTGGCAGGGAACCGGCTGGGGGACAAGTGTGTGCTGAGCTGGTGGCTGCCCTGGGCACCA
TGCCAGCCTGGCCCTCCTTGACCTCTCCTCCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGG
60 GCTCCCAGGCCAAGCCACCTTGACAGATTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGT
GGCCAGTCCCTGGCCTCCCTCCTGCACGCTGCCCTTACTCAGCACCCCTGCGCCTGCAGGCGTGTGGCTTCG
GCCCCAGCTTCTTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCC

5 GTCCCTGTCTTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTC
CTGCACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTCCGACCTCATGGAGCCTGTATTCCGATACC
TGGCCAAGGAAGGCTGTGCTCTAGCCACCTGACCCTGTCTGCAAACACCTGGGGGACAAGGCTGTTAGAGA
CCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCTCAGATCAGCTGT
GCCAGCTTGGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTTGGCCTGTCAAGGCT
GCGCCGTCCAGGGTCCCTTGGCCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAAGTGCAGCTGTG
CAGCAGACGCCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCCAGTCCGCGGGCCCCGGCGAG
TGCACGCTGGACCACGGCTCCAAGCTCTTCTTTGGCGCCTCTAG

SEQ ID NO: 2 - TONSL WT

10 ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
CGCGAAGAGGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCC
CACC GCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGCAGCACCAGCACCAG
15 TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
ACCCACCTGGACATCTATGACCACTGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
TTCAGGAAGAGCATCTTCCTTGGCGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
20 CTGGGCACCACTCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGCCCGGGAG
TGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGGTTATTGCACAGGTCTCTCAA
GACCTGGGAGACTTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
GTGCAGAGGGCAGCCATCTGTGCAACCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
GAAGAGGCTGAGGGCAGAGACCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTTGTGAGCTGCTGGAC
25 AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
CATGGGGCCGTGCGCCACTATGAGGAGGAACCTGAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
ACCTGGCTGAACATTGCACTGTCCCGCAGGAGGCGCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTTC
CAGAAAGCGCTCAGCTGTGCCACAGCGGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGCAGCATCTC
CATAACCGTGAGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGT
30 GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTG
GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGGAACCGGCGAAACGACATGGGG
GCGACCCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGC
CACCCTCTTAACCTTCGGGACTACTGTGGCTGGACACCTCTGCACGAGGCCTGCAACTACGGGCATCTA
35 GAAATTTGTCTTCTTCTGCTGGAGCAGCGGGCCAGCTGGAGACGCCAGGTGGCCAGGGCTGCGAAGGC
ATCACCCCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
GCGTCCGTCAACCTCCGCACTCGAAAGGGCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
GCCTCGGGCCAAAGATCCCCACAGCTCCAGGCCTTCCACACCCCCAAGCAGCCTTCTGTTTACCCCCGAG
40 ACCTCTCTCTCTTTGAGCCCCCTGCCAGAACCCCCCTCTAATAGCACTAGACTCCAGAGGCCTCTCAG
GTCCATGTGAGGCTCTCCCCAGGGCAGGCGGCACCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
CCAGCCAGCAGCAGCAGCAGCTCAGAAGGCGAGGACAGCGCAGGCCCGCACGGCCGTCCCAGAAGAGG
CCTCGGTGCTCGGCCACAGCACAACGGTGGCAGCTGAGCAGCCCTGGCCCCGAGCAACAGGGAAGCA
45 GCCACAGCCAGCACCAGCGGGCAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGTGAGAGCCGG
CTGGGGCCTGGCCACC GCGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGCGAGCCGCGGCCCCGCCCCCGG
GGCACTGGAGACAACCGCAGGCCAGTAGTACCTCTGGGTGCGACAGTGAGGAGAGCAGGCCCGCTGCC
CGAGCCAAGCAGGTCCGCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCAGGGCCCCAGCAGC
50 CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCCCAGGGTCTCAGAGCCCAGTGGGGACAGCTCTGCG
GCAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCTCCCATCCGGGTTTCAGTTCAAGTTCAGGATCATCTC
TTCTCATCCCTGTCCACACAGCAGTGACACCCACTCTGTGGCTGGCTGGCCGAGCAGGCGGCCAG
CGCTACTACCAGACCTGCGGGCTGCTGCCAGGCTCACCTACGGAAAGAGGGGGCCCTGCTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
55 CAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCGTTTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAG
CTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGG
CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC
CTCTCCTCCAATCACCTGGGTCCCGAAGGCCCTGCGCCAGCTTGCCATGGGGCTCCCAAGGCCAACCCACC
TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
60 TCCCTCCTGCACGCCTGCCCTTACTCAGCACCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCTGTCCCTG
TCCTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTCCTG

CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTTCGGACCTCATGGAGCCTGTATTCCGATAC
CTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCCCTGAG
ATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
5 GGCCTGTCAGGCTGCGCCGTCCAGGGTCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
CGGGAAGTGCAGCTGTGCAGCAGACGCCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
AGTCGGCCGGGCCCCGGCGAGTGCACGCTGGACCACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 3 - TONSL E530A

10 ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
CGCGAAGAGGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCC
CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGACAGCACCAGCACCAG
15 TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
ACCCACCTGGACATCTATGACCACTGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
TTCAGGAAGAGCATCTTCTTGCGGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
20 CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGCCCCGGGAG
TGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGGTTATTGCACAGGTCCTCCAA
GACCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
GTGCAGAGGGCAGCCATCTGTGCAAACTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
GAAGAGGCTGAGGGCAGAGACCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTGCTGAGCTGTGGAC
25 AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
CATGGGGCCGTGCGCCACTATGAGGAGGAACTGAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGGCAAG
ACCTGGCTGAACATTGCACTGTCCCAGGAGGAGGCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTC
CAGAAAGCGCTCAGCTGTGCCAGCAGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGCAGCATCTC
CATACCGTGACAGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGT
30 GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTG
GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGAACCGGCGAAACGACATGGGG
GCGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGC
CACCCCTTAACCCTCGGGACTACTGTGGCTGGACACCTCTGCACGAGGCCTGCAACTACGGGCATCTA
35 GAAATTGTCCGCTTCCCTGCTGGACCACGGGGCCGAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGC
ATCACCCCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
GCGTCCGTCACCCCTCCGCACTCGAAAGGGCCTCAGCCCCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
GCCTCGGGCCAAGATCCCCACAGCTCCAGGCCCTTCCACACCCCCAAGCAGCCTTCTGTTGACCCCGAG
40 ACCTCTCCTCCTTTGAGCCCCCTGCCAGAACCCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAG
GTCCATGTGAGGGTCTCCCCAGGGCAGGCGGACAGCCATGGCCAGGCCCTCGGAGGAGCAGGCATGGG
CCAGCCAGCAGCAGCAGCAGCTCAGAAAGCGAGGACAGCGCAGGCCCCGACGGCCGTCCCAGAAGAGG
CCTCGGTGCTCGGCCACAGCACACGGGTGGCAGCCTGGACGCTGGCCCCGCCAGCAACAGGGAAGCA
GCCACAGCCAGCACCAGCCGGGCAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGTCTCAGAGCCGG
45 CTGGGGCCTGGCCACCGCGGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCCTGACCCGCAGCCGCGGCCCGCCCCCGG
GGCACTGGAGACAACCGCAGGCCCAGTAGTACCTCTGGGTGCGACAGTGAGGAGAGCAGGCCCCGTGCC
CGAGCCAAGCAGGTCCGCCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGAGGGGCCAGCAGC
CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCCCAGGGTCTCAGAGCCCACTGGGGACAGCTCTGCG
50 GCAGGCCAGCCCTTGGGTCCGGCCCCCGCCCCCTCCCATCCGGGTTCGAGTTCAAGTTCAGGATCATCTC
TTCCTCATCCCTGTCCACACAGCAGTGACACCACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAG
CGCTACTACCAGACCTGCGGGCTGCTGCCAGGCTCACCTACGGAAGAGGGGGCCCTGCTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
55 CAGGCCGTGGAGCTCCAGGGCTTGGGCCTTCGTTTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCCA
CTTACACCCCTGTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCTGGCAGGGAACCGG
CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC
CTCTCCTCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCAGGCCAAGCCACC
TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
60 TCCCTCCTGCACGCTGCCCTTACTCAGCACCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCTGTCCCTG

TCCTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTCCTG
CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTTCGGACCTCATGGAGCCTGTATTCCGATAC
CTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCCCTGAG
5 ATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
GGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
CGGGAACCTGCAGCTGTGCAGCAGACGCCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
AGTCGGCCGGGCCCCGGCGAGTGCACGCTGGACCACGGCTCCAAGCTCTTCTTTGGCGCCTCTAG

10 *SEQ ID NO: 4 - TONSL E530A / GFP*

atggtgagcaagggcgaggagctgttcacccgggtggtgcccacccctggtcgagctggaacggcgacgta
aacggccacaagttcagcgtgtccggcgaggcgaggcgatgccacctacggcaagctgacctgaag
ttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacctgaacctacggcgtg
cagtgtctcagccgtaccocgaccacatgaagcagcagcacttcttcaagtccgccatgccgaaggc
15 taegtccaggagcgacccatcttcttcaaggacgaacggcaactacaagaccgcgcgcgaggtgaagttc
gagggcgacacccctggtgaaccgcacatcgagctgaagggcactcgacttcaaggaggacggcaacatcctg
gggcacaagctgagtaacaactacaacagccacaacgtctatatcatgcccgaacaagcagaagaacggc
atcaaggtgaacttcaagatccgccaacacatcgaggacggcagcgtgcagctcgcgaccactaccag
cagaacaccccccatcggcgacggccccgtgctgctgcccagacaaccactacctgagcaccacagtcgcgc
20 ctgagcaaaagaccccaacgagaagcgcgacacacatggctcctgctgaggttcgtgaccgcgcgcgggac
actctcggcatggagcagctgtacaagggcgcgccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGGCGGAAGAGGGCCGCGCTGTGCCACCAGCTGGGG
GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
GAGCGCGCTGACGACCCCTCTGGGCTGTGCCGTGGCCACCGCAAGATCGGAGAGCGCCTGGCCGAGATG
25 GAGGACTACCCGGCTGCCTTGACGACACCAGCACAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGACCCACCTGGACATCTATGACCACTGCCAGTCG
AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
GGGACACTGGCCAGGGAGAGCTGAATGAGATGAGGACCCGCTCTATCTCAACCTGGGCCTCACCTTT
GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCCTTGCGGAGCAGAAC
30 CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCAC
TCCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAG
AGCGAGTGTGCTGCTGGTTATTGCACAGGTCTTCCAAGACCTGGGAGACTTTTGGCTGCCAAGCGAGCC
CTGAAGAAGGCCTACAGGCTGGGCTCCCGAGAAGCCTGTGTCAGAGGGCAGCCATCTGTGAGAACCTCCAG
CATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCCCTCAGGGTGCC
35 ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT
TACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGACAGACCGGGTGTGAGCGGGCCATCATCCACGTG
TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAAGT
AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCAGGAG
GCCCGCGATGCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCAG
40 CGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTCCATACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
GCCCCCTGAGACCGGAAACAGACTGCGGGAGCTCAGTGTAGCTGAAGATGAAGATGAGGAGAGGAGGCG
GAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCGGCGAGGTGGAGCTCTCAGAGGGCGAG
GACGACACCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCCACCTGGGCCGGCGG
AAGGGGAGCAAGTGGAACCGGCGAAACGACATGGGG**GCG**ACCCTGCTGCACCGAGCCTGCATCGAGGGC
45 CAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGCCACCCCCCTTAACCCCTCGGGACTACTGTGGCTGG
ACACCTCTGCACGAGGCCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGGCC
GCAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCCTCAACTGTGGC
CACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTCACCCCTCCGACTCGAAAGGGCCCTC
AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
50 AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
TTCCACACCCCAAGCAGCCTTCTGTTTGACCCGAGACCTCTCCTCCTTTGAGCCCTGCCAGAACCC
CCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAGGTCCATGTGAGGTCTCCCCAGGGCAGGCGGCA
CCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGGCCAGCCAGCAGCAGCAGCTCAGAAGGCGAG
GACAGCGCAGGCCCCGCACGGCCGTCCAGAAGAGGCCTCGGTGCTCGGCCACAGCACAACGGGTGGCA
55 GCCTGGACGCCTGGCCCCCGCCAGCAACAGGAAGCAGCCAGCCAGCAGCCAGCCGGGCGAGCCTACCA
CGAGCTACCGGGGTGGGTCAGTGTCTCAGAGCCGGCTGGGGCCTGGGCCACCGGGGCCACAGCAAA
GCCCTTGCCCCCAGGCAGCGCTCATCCCGAGGAGGAGTGCCTGGCTGGGGACTGGCTGGAGCTGGAC
ATGCCCTGACCCGCAGCCGCCGGCCCCCGCCCCGGGGCACTGGAGACAACCGCAGGCCAGTAGTACC
TCTGGGTGCGACAGTGAGGAGAGCAGGCCCCGTGCCCGAGCCAAGCAGGTCCGCTGACCTGCATGCAG
60 AGTTGCAGTGCGCCAGTTAACGCAGGGCCCAGCAGCCTGGCTTCAGAACCTCCAGGGAGCCCCAGCACC
CCCAGGGTCTCAGAGCCCAGTGGGGACAGCTCTGCGGCAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCT

CCCATCCGGGTTTCGAGTTCAAGTTCAGGATCATCTCTTCTCATCCCTGTCCCACACAGCAGTGACACC
CACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAGCGCTACTACCAGACCTGCGGGCTGTGCCCAGG
CTCACCCCTACGGAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGCAAT
5 GACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCGTTGACTGACCGCTACCGCAGGGCCTGC
CAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTGCAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCG
TTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAGCTTACACCCCTGTGCGGGCCCTCAAGCTGCAC
ACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGGCTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCT
GCCCTGGGCACCATGCCCAGCCTGGCCCTCCTTGACCTCTCTCCAATCACCTGGGTCCCGAAGGCCTG
10 CGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACCTTGACAGAGTTTGGAGgaattagatctatcgatg
aACCCCTGCGGGGACGGCTGTGGCCAGTCCCTGGCCTCCCTCCTGCACGCCTGCCCCCTACTCAGCACC
CTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTCTTTCTGAGCCACCAGACAGCACTGGGTAGTGCT
TTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTGTCTACAACGCCCTGGGAGCCCCCTGCCCTGGCC
AGGACCCTGCAGAGCCTGCCCGCCGGCACCCTCCTGCACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAG
GGTGATTTCGACCTCATGGAGCCTGTATTCCGATACCTGGCCAAGGAAGGCTGTGCTCTAGCCACCTG
15 ACCCTGTCTGCAAACCACTGGGGGACAAGGCTGTTAGAGACCTGTGCAGATGTCTCTCTGTGCCCC
TCACTCATCTCACTGGATCTGTCTGCCAACCCTGAGATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCC
ACCCTCCAAAAGCGGCCCCAAGGCCTTAGCTTCCTTGGCCTGTCAGGCTGCGCCGTCCAGGGTCCCCTG
GGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAAGTGCAGCTGTGCAGCAGACGCCTCTGC
GCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCAGTCGGCCGGGCCCCGGCGAGTGCACGCTGGAC
20 CACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 5 - TONSL D559A

ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
CGCGAAGAGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
25 CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCC
CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGACGACACCAGCACCAG
TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
ACCCACCTGGACATCTATGACCACTGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
30 ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
TTCAGGAAGAGCATCTTCTTGGCGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAG
TGTGCGCACACCAATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGGTTATTGCACAGGTCTCTCAA
GACCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
35 GTGCAGAGGGCAGCCATCTGTGCAACCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
GAAGAGGCTGAGGGCAGAGACCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGAC
AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
CATGGGGCCGTGCGCCACTATGAGGAGGAACAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
40 ACCTGGCTGAACATTGCACTGTCCCGCAGGAGGCGCGGATGCCTACGAGCTGCTGGCCCCGTGCTTC
CAGAAAGCGCTCAGCTGTGCCAGCAGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTTCAGCATCTC
CATACCGTGACGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGT
GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGAGCCACAGCGGAGAGCGAAGCCCTG
GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
45 GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGAACCGGCGAAACGACATGGGG
GAGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGC
GACCCCTTAACCTTCGGG**CC**TAAGTGTGGCTGGACACCTCTGCACAGAGGCTGCAACTACGGGCATCTA
CAAATTGCTCCGCTTTCCTGCTGGACCACGGGCGCAGTTGGACGACCCAGGTGGCGAGGGCTGCCAAGGC
ATCACCCCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTTGAACGGGGG
50 GCGTCCGTCAACCTCCGCACTCGAAAGGGCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGAGGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
GCCTCGGGCCAAGATCCCCACAGCTCCAGGCCTTCCACACCCCAAGCAGCCTTCTGTTTGACCCCGAG
ACCTCTCTCTCTTTGAGCCCCCTGCCAGAACCCCCCTCTAATAGCACTAGACTCCAGAGGCCTCTCAG
GTCCATGTGAGGCTCTCCCCAGGGCAGGCGGCACAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
55 CCAGCCAGCAGCAGCAGCTCAGAAGGCGAGGACAGCGCAGGCCCCGACGGCCGTCCCAGAAGAGCA
CCTCGGTGCTCGGCCACAGCACAAGGTTGGCAGCTGGACGCGCTGGCCCCGCGCAGCAACAGGGAAGCA
GCCACAGCCAGCACCAGCGGGGACGCTTACCAGGCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGG
CTGGGGCCTGGCCACCAGCGGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGAGCCGCGCGGCCCGCCCCCGG
60 GGCAGTGGAGACAACCGCAGGCCAGTAGTACCTCTGGGTGGACAGTGAGGAGAGCAGGCCCGCTGCC
CGAGCCAAGCAGGTCCGCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCAGGGCCCCAGCAGC

CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCCCAGGGTCTCAGAGCCCAGTGGGGACAGCTCTGCG
GCAGGCCAGCCCTTGGGTCCGGCCCCGCCCTCCCATCCGGGTTCGAGTTCAAGTTCAGGATCATCTC
TTCCTCATCCCTGTCCACACAGCAGTGACACCCACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAG
5 CGCTACTACCAGACCTGCGGGCTGCTGCCAGGCTCACCTACGGAAGAGGGGGCCCTGCTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
CCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
CAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCGTTTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAG
CTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGG
10 CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC
CTCTCCTCCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCAGGCCAAGCCACC
TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
TCCCTCCTGCACGCCTGCCCTTACTCAGCACCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCTGTCCCTG
15 TCCTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCTGCAGAGCCTGCCCGCCGGCACCCCTG
CACTTAGAGCTCAGCTCCGTGGCAGCCGCAAGGGTGATTTCGGACCTCATGGAGCCTGTATTCCGATAC
CTGGCCAAGGAAGGCTGTGCTCTAGCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCTTGAG
ATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
GGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
20 CGGGAACCTGCAGCTGTGCAGCAGACGCCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
AGTCGGCCGGCCCCGGCGAGTGACGCTGGACCACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 6 - TONSL D559A / GFP

atggtgagcaagggcgaggagctgttcacccgggtggtgcccacccctggtcgagctggacggcgacgta
25 aacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaag
ttcatctgcaccacccggcaagctgcccgtgccctggcccaccctcgtgaccacccctgacctacggcgtg
cagtgtctcagccgtaccccgaccacatgaagcagcagcacttcttcaagtcgccatgcccgaaggc
tacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccgcgcggaggtgaagttc
30 gagggcgacacccctggtgaaccgcacatcgagctgaagggcactgacttcaaggaggacggcaacatccctg
gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggc
atcaaggtgaacttcaagatccgcccacaacatcgaggacggcagcgtgcagctcgcgaccactaccag
cagaacaccccatcggcgacggccccgtgctgctgcccagacaaccactacctgagcaaccacgtccgcc
ctgagcaaaagaccccaacgagaagcgcgacacatggtcctgctggagttcgtgaccgcgcgggac
actctcggcatggacgagctgtacaagggcgcgccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
35 AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGCGCGAAGAGGCCGCGCTGTGCCACCAGCTGGGG
GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
GAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCCACCCGAAGATCGGAGAGCGCCTGGCCGAGATG
GAGGACTACCCGGCTGCCTTGACAGCACCAGCACCAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGCACCCACCTGGACATCTATGACCACTGCCAGTCG
40 AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
GGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGGACCCGCCTCTATCTCAACCTGGGCCTCACCTTT
GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCCTTGCGGAGCAGAAC
CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCAC
TCCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAG
45 AGCGAGTGCTGCGTGGTTATTGCACAGGTCTCCAAGACCTGGGAGACTTTTGGCTGCCAAGCGAGCC
CTGAAGAAGGCCTACAGGCTGGGCTCCCAAGCCTGTGAGAGGGCAGGCCATCTGTCAGAACCCTCCAG
CATGTGCTGGCAGTGCTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCCCTCAGGGTGCC
ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT
50 TACCAGAAGCAGCTGCGTTTTTGTGAGCTGCTGGACAGACCGGGTGTGAGCGGGCCATCATCCACGTG
TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAACTG
AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCGAGGAG
GCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCAG
CGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTCCATAACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
GCCCCGTGAGACCGAAACCAGACTGCGGGAGCTCAGTGTAGCTGAAGATGAAGATGAGGAGGAGGAGCG
55 GAGGAGCGGCAGCCACAGCGGAGAGCGCAAGCCCTGGAGGCCGGCAGGTGGAGCTCTCAGAGGGCGAG
GACGACACCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCCACCTGGGCCGGCGG
AAGGGGAGCAAGTGAACCGGGCGAAACGACATGGGGGAGACCCTGCTGCACCGAGCCTGCATCGAGGGC
CAGCTGCGCCGCTCCAGGACCTTGTGAGGCAGGGCCACCCCTTAACCCTCGGGCCTACTGTGGCTGG
ACACCTCTGCACGAGGCCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGGCC
60 GCAGTGAGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCCTCAACTGTGGC
CACTTCGAGGTGGCTGAGCTGCTGTTGAACGGGGGGCGTCCGTCACCCTCCGCACTCGAAAGGGCCTC

AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
 AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
 TTCCACACCCCAAGCAGCCTTCTGTTTGACCCGAGACCTCTCCTCCTTTGAGCCCTGCCAGAACCC
 CCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAGGTCCATGTCAGGGTCTCCCCAGGGCAGGCGGCA
 5 CCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGGCCAGCCAGCAGCAGCAGCTCAGAAGGCGAG
 GACAGCGCAGGCCCCGACAGGCCGTCCCAGAAGAGGCCTCGGTGCTCGGCCACAGCACAACGGGTGGCA
 GCCTGGACGCCTGGCCCCGCCAGCAACAGGGAAGCAGCCACAGCCAGCACCAGCCGGGCAGCCTACCAG
 GCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGGCTGGGGCCTGGCCCCACCGCGGGGCCACAGCAAA
 GCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAGGAGTGCTGGCTGGGGACTGGCTGGAGCTGGAC
 10 ATGCCCTTGACCCGCAGCCGCCGGCCCCGCCCGGGGCACTGGAGACAACCGCAGGGCCAGTAGTACC
 TCTGGGTGCGACAGTGAGGAGAGCAGGCCCGTGCCTGAGCCAAAGCAGGTCCGCTGACCTGCATGCAG
 AGTTGCAGTGCGCCAGTTAACGCAGGGCCCAGCAGCCTGGCTTCAGAACCTCCAGGGAGCCCCAGCACC
 CCCAGGGTCTCAGAGCCCAGTGGGGACAGCTCTGCGGCAGGCCAGCCCTTGGGTCCGGCCCCGCCCT
 CACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCACGCTACTACCAGACCTGCGGGCTGCTGCCAGG
 15 CTCACCCTACGGAAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGCAAT
 GACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCGTTGACTGACCGCTACCGCAGGGCCTGC
 CAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTGCAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCG
 TTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAGCTTACACCCCTGCTGCGGGCCCTCAAGCTGCAC
 20 ACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGGCTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCT
 GCCCTGGGCACCATGCCCAGCCTGGCCCTCCTTGACCTCTCCTCCAATCACCTGGGTCCCGAAGGCCTG
 CGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACCTTGACAGAGTTTGGAGGaattagatctatcgatg
 aACCCCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCCTCCCTCCTGTCACGCCTGCCCCCTACTCAGCACC
 CTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTCTTTCTGAGCCACCAGACAGCACTGGGTAGTGCT
 25 TTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTGTCTTACAACGCCCTGGGAGCCCCCTGCCCTGGCC
 AGGACCCTGCAGAGCCTGCCCCGCCGCCACCCTCCTGCACCTTAGAGCTCAGCTCCGTGGCAGCCGGCAAG
 GGTGATTGCGACCTCATGGAGCCTGTATTCCGATACCTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTG
 ACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTTAGAGACCTGTGCAGATGTCTCTCTGTGCCCC
 TCACTCATCTCACTGGATCTGTCTGCCAACCTTGAGATCAGCTGTGCCAGCTTGGAAAGAGCTCCTGTCC
 30 ACCCTCCAAAAGCGGGCCCAAGGCCTTAGCTTCCCTTGGCCTGTGAGGCTGCGCCCTCAGGGTCCCTG
 GGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAACTGCAGCTGTGCAGCAGAGCCTCTGC
 GCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCAGTCCGGCCGGGGCCCCGGCGAGTGCACGCTGGAC
 CACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

35 SEQ ID NO: 7- TONSL W563A

ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
 CGCGAAGAGGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
 CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCTCTGGGCTGTGCCGTGGCC
 40 CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGCAGCACCAGCACCAG
 TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGGCACCATCGGCCGC
 ACCCACCTGGACATCTATGACCACTGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
 AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
 ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
 45 TTCAGGAAGAGCATCTTCTTGGCGGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCGCTACAAC
 CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCCAGGCTATGCGCTGCTTGGAGGGTGCCCCGGAG
 TGTGCGCAGACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGTTATTGACACCTGCTCCAA
 50 TGGCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCGAGCTCCCA
 GTGCAGAGGGCAGCCATCTGTTCAGAACCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
 GAAGAGGCTGAGGGCAGAGACCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
 AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGAC
 AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
 CATGGGGCCGTGCGCCACTATGAGGAGGAACCTGAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
 ACCTGGCTGAACATTGCACTGTCCCAGGAGGAGGCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTC
 CAGAAAGCGCTCAGCTGTGCCAGCAGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTC
 55 CATACCGTGCAGCTGAGGCTGCAGCCCGAGGAGGCCCTGAGACCGAAACAGACTGCGGGAGCTCAGT
 GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGCAGCCACAGCGGAGAGCGAAGCCCTG
 GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
 GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGAACCGGGCAACGACATGGGG
 GAGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCGAGGGC
 60 CACCCCTTAACCTCGGGACTACTGTGGCGCGACACCTCTGCACGAGGCCTGCAACTACGGGCATCTA
 GAAATTGTCCGCTTCTGCTGGACCACGGGGCCGAGTGGACGACCAGGTGGCCAGGGCTGCGAAGGC

ATCACCCCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
GCGTCCGTCACCCTCCGCACTCGAAAGGGCCTCAGCCCCGTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
5 GCCTCGGGCCAAGATCCCCACAGCTCCAGGCCTTCCACACCCCCAAGCAGCCTTCTGTTGACCCCGAG
ACCTCTCCTCCTTTGAGCCCCCTGCCAGAACCCCCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAG
GTCCATGTGAGGTCTCCCCAGGGCAGGCGGCACCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
CCAGCCAGCAGCAGCAGCTCAGAAGGCGAGGACAGCGCAGGCCCGCACGGCCGTCCCAGAAGAGG
CCTCGGTGCTCGGCCACAGCACAACGGGTGGCAGCCTGGACGCCTGGCCCCGCCAGCAACAGGGAAGCA
10 GCCACAGCCAGCACCAGCCGGGCAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGTCTCAGAGCCGG
CTGGGGCCTGGCCACCAGCGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGCAGCCGCCGGCCCCGCCCGCG
GGCACTGGAGACAACCGCAGGCCAGTAGTACCTCTGGGTGCGACAGTGAGGAGAGCAGGCCCGCTGCC
CGAGCCAAGCAGGTCCGCCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCAGGGCCCCAGCAGC
15 CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCTCAGGGTCTCAGAGCCAGTGGGGACAGCTCTGCG
GCAGGCCAGCCCTTGGGTCCGGCCCCGCCCTCCCATCCGGGTTCGAGTTCAAGTTCAGGATCATCTC
TTCCTCATCCCTGTCCCACACAGCAGTGACACCCACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAG
CGCTACTACCAGACCTGCGGGCTGCTGCCAGGCTCACCCTACGGAAAGAGGGGGCCCTGCTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
20 CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
CAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCGTTACAGCGCTGCTCCCTGGCCCTGGACCAGGCCAG
CTTACACCCCTAGTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCCG
CTGGGGGCAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGTAC
CTCTCCTCCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCAGGCCAAGCCACC
25 TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
TCCCTCCTGCACGCCTGCCCTTACTCAGCACCCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTG
TCCTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTCCTG
CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTGCGACCTCATGGAGCCTGTATTCCGATAC
30 CTGGCCAAGGAAGGCTGTGCTCTAGCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTGATCTGTCTGCCAACCCCTGAG
ATCAGCTGTGCCAGCTTGGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
GGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
CGGGAACCTGCAGCTGTGCAGCAGACGCCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
AGTCGGCCGGGCCCGGCGAGTGACAGCTGGACCACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

35

SEQ ID NO: 8 - TONSL W563A / GFP

atggtgagcaagggcgaggagctgttcaccgggggtggtgcccattcctggtcgagctggacggcgacgta
aacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaag
ttcatctgcaccaccggcaagctgcccgtgccttgcccaccctcgtgaccacctgacctacggcgtg
40 cagtgttcagcgcgtaccccgaccacatgaagcagcagcacttctcaagtcgcccattgcccgaagggc
tacgtccaggagcgaccatcttctcaaggacgacggcaactacaagaccgcgcgaggtgaagttc
gagggcgacacctgtgaaccgcacatcgagctgaagggcactgacttcaaggaggacgggaacatcctg
gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggc
atcaaggtgaacttcaagatccgcccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
45 cagaacacccccatcgggcagcgccccgtgctgctgcccagacaaccactacctgagcaccacgtccgcc
ctgagcaaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcgcgcccggatc
actctcggcacatggacgagctgtacaagggcgcgcccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGCGCAAGAGGCCGCGCTGTGCCACCAGCTGGGG
GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
50 GAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCCACCAGCAAGATCGGAGAGCGCCTGGCCGAGATG
GAGGACTACCCGGCTGCCTTGCAGCACCAGCACCAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGCACCCACCTGGACATCTATGACCACTGCCAGTCG
AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
GGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGGACCCGCCTCTATCTCAACCTGGGCCTCACCTTT
55 GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCTTGGCGGAGCAGAAC
CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTTGGGCACCATCCACTGGCGCGCGGCCAGCAG
TCCCAGGCTATGCGCTGCTTGGAGGGTCCCCGGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAG
AGCGAGTGCTGCGTGGTTATTGCACAGGTCTTCCAAGACCTGGGAGACTTTTTGGCTGCCAAGCGAGCC
CTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCTGTGCAGAGGGCAGCCATCTGTGAGAACCTCCAG
60 CATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCTCAGGGTGCC
ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT

TACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGACAGACCGGGTGTGAGCGGGCCATCATCCACGTG
 TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAACTG
 AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCAGGAG
 5 GCCGGCGATGCCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCAG
 CGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTCCATACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
 GCCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGTGTAGCTGAAGATGAAGATGAGGAGGAGGAGGCG
 GAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCCGCGAGGTGGAGCTCTCAGAGGGCGAG
 GACGACACCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCCACCTGGGCGGCGG
 AAGGGGAGCAAGTGGAAACCGGCGAAACGACATGGGGGAGACCCTGCTGCACCGAGCCTGCATCGAGGGC
 10 CAGCTGCGCCGCTCCAGGACCTTGTGAGGCAGGGCCACCCCCCTTAACCCTCGGGACTACTGTGGC**GCG**
 ACACCTCTGCACGAGGCCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCCTGTGGACCACGGGGCC
 GCAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCCTCAACTGTGGC
 CACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTCAACCCTCCGCACTCGAAAGGGCCTC
 AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
 15 AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
 TTCCACACCCCAAGCAGCCTTCTGTTTGACCCGAGACCTCTCCTCCTTTGAGCCCCCTGCCAGAACCC
 CCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAGGTCCATGTCAGGGTCTCCCCAGGGCAGGCGGCA
 CCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGGCCAGCCAGCAGCAGCAGCAGCTCAGAAGGCGAG
 GACAGCGCAGGCCCCGCACGGCCGTCCAGAAAGAGGCCTCGGTGCTCGGCCACAGCACAAACGGGTGGCA
 20 GCCTGGACGCCTGGCCCCGCCAGCAACAGGGAAGCAGCCACAGCCAGCACCAGCCGGGCAGCCTACCAG
 GCAGCCATCCGGGTGTGGGCAGTGCTCAGAGCCGGCTGGGGCCTGGCCCACCGCGGGGCCACAGCAAA
 GCCCCGTGCCCCCCAGGCAGCCTCATCCCGAGGAGGAGTGCCTGGCTGGGGACTGGCTGGAGCTGGAC
 ATGCCCTGACCCGCAGCCGCGGCCCGCCCCCGGGGCACTGGAGACAACCGCAGGCCAGTAGTACC
 TCTGGGTGCGACAGTGAGGAGAGCAGGCCCCGTGCCCGAGCCAAGCAGGTCCGCTGACCTGCATGCAG
 25 AGTTGCAGTGCGCCAGTTAACGCAGGGCCCAGCAGCCTGGCTTCAGAACCTCCAGGGAGCCCCAGCACC
 CCCAGGGTCTCAGAGCCAGTGGGGACAGCTCTGCGGCAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCT
 CCCATCCGGGTTCGAGTTCAAGTTCAAGTTCAGGATCATCTCTTCTCATCCCTGTCCACACAGCAGTGACACC
 CACTCTGTGGCTGGCTGGCCGAGCAGGCGGCCAGCGCTACTACCAGACCTGCGGGCTGCTGCCCAGG
 CTCACCCTACGGAAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGCAAT
 30 GACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCCGTTGACTGACCGCTACCGCAGGGCCTGC
 CAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTGCAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCG
 TTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAGCTTACACCCCTGCTGCGGGCCCTCAAGCTGCAC
 ACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGGCTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCT
 GCCCTGGGCACCATGCCCAGCCTGGCCCTCCTTGACCTCTCCTCCAATCACCTGGGTCCCGAAGGCCTG
 35 CGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACCTTGCAGAGTTTGGAGgaattagatctatcgatg
 aACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCCTCCCTCCTGCACGCCTGCCCTTACTCAGCACC
 CTGCGCCTGCAGGCTGTGGCTTCGGCCCCAGCTTCTTTCTGAGCCACCAGACAGCAGCTGGGTAGTGCT
 TTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTGTCTACAACGCCCTGGGAGCCCCTGCCCTGGCC
 AGGACCCTGCAGAGCCTGCCCGCGGCACCCTCCTGCACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAG
 40 GGTGATTGCGACCTCATGGAGCCTGTATTCCGATACCTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTG
 ACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTTAGAGACCTGTGCAGATGTCTCTCTGTGCCCC
 TCACTCATCTCACTGGATCTGTCTGCCAACCCTGAGATCAGCTGTGCCAGCTTGAAGAGCTCCTGTCC
 ACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTTGGCCTGTCAGGCTGCGCCGTCCAGGGTCCCCTG
 45 GGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAACTGCAGCTGTGCAGCAGACGCCTCTGC
 GCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCAGTCGGCCGGGCCCCGGCGAGTGACAGCTGGAC
 CACGGCTCCAAGCTCTTCTTCGGCGCCTCTAG

SEQ ID NO: 9 - TONSL E568A

ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
 50 CGCGAAGAGGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
 CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTCCGTGGCC
 CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCGGCTGCCTTGCAGCACCAGCACCAG
 TACCTGGAGCTGGCACATTCCCTGCGCAACCACAGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
 ACCACCTGGACATCTATGACCAGTGCCAGTGCAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
 55 AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
 ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
 TTCAGGAAGAGCATCTTCTTGGCGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
 CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGCCCGGGAG
 TGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGGTTATTGCACAGGTCTCTCAA
 60 GACCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
 GTGCAGAGGGCAGCCATCTGTGCAACCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG

GAAGAGGCTGAGGGCAGAGACCCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
 AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGAC
 AGACCGGGTGCTGAGCGGGCCATCATCCACGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
 CATGGGGCCGTGCGCCACTATGAGGAGGAAGTGAAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
 5 ACCTGGCTGAACATTGCACTGTCCCAGGAGGAGGCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTC
 CAGAAAGCGCTCAGCTGTGCCCAGCAGGCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTC
 CATACCGTGCAGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACAGACTGCGGGAGCTCAGT
 GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTG
 GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
 10 GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGGAAACCGGCGAAACGACATGGGG
 GAGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGC
 CACCCCTTAACCTCGGGACTACTGTGGCTGGACACCTCTGCACGCGGCCCTGCAACTACGGGCATCTA
 GAAATTGTCCGCTTCTGCTGGACCACGGGGCCGAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGC
 ATCACCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
 15 GCGTCCGTCAACCTCCGCACTCGAAAGGGCCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
 TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
 GCCTCGGGCCAAGATCCCCACAGCTCCAGGCCCTTCCACACCCCAAGCAGCCTTCTGTTTGACCCCGAG
 ACCTCTCCTCCTTTGAGCCCCCTGCCAGAACCCCTCTAATAGCACTAGACTCCAGAGGCCTCTCAG
 GTCCATGTGAGGCTCTCCCCAGGGCAGGCGGCACAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
 20 CCAGCCAGCAGCAGCAGCAGCTCAGAAGGCGAGGACAGCGCAGGCCCGCACGGCCGTCCCAGAAGAGG
 CCTCGGTGCTCGGCCACAGCACAACGGGTGGCAGCCTGGACGCTTGGCCCCGCGCAGCAACAGGGAAGCA
 GCCAGCCAGCAGCAGCAGCGGGCAGCCTACAGGAGGCCATCCGGGGTGTGGGAGTGTCTCAGAGCCGG
 CTGGGGCTGGCCACCGCGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGAGGAG
 GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGCAGCCGCGGCCCGCCCCCGG
 25 GGCAGTGGAGACAACCGCAGGCCCAGTAGTACCTCTGGGTGCGACAGTGAAGGAGAGCAGGCCCGTGGC
 CGAGCCAAGCAGGTCCGCTGACCTGCATGCAGAGTTGAGTGCAGCAGTTAAGCAGAGGCCAGCAGC
 CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCCGAGGTCTCAGAGCCAGTGGGGACAGCTCTGCG
 GCAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCTCCCATCCGGGTTCGAGTTCAAGTTCAGGATCATCTC
 TTCCTCATCCCTGTCCCACACAGCAGTGCACCCACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAG
 30 CGTACTACAGACCTGCGGGCTGCTGCCAGGCTCACCCTACGGAAGAGAGGGGCCCTGCTGGCCCCCA
 CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
 CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
 CAGGCCGTGGAGCTCCAGGGCTTGGGCTCTCGTTAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAG
 CTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGG
 35 CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC
 CTCTCCTCCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACC
 TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
 TCCCTCCTGCACGCCTGCCCTTACTCAGCACCCCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
 TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCTGTCCCTG
 40 TCCTACAACGCCCTGGGAGCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTCCTG
 CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTGCGACCTCATGGAGCCTGTATTCCGATAC
 CTGGCCAAGGAAGGCTGTGCTCTAGCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
 AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCCCTGAG
 ATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
 45 GGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCTGGGCTGGGCTGGGCTGTGGGACAAGATAGCCGCGCAGCTC
 CGGAACTGCAGCTGTGCAGCAGACGCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCC
 AGTCGGCCGGGCCCGGCGAGTGCACGCTGGACCACGGCTCCAAGCTCTTTTCGGCGCCTCTAG

SEQ ID NO: 10 - TONSL E568A / GFP

50 atggtgagcaagggcgaggagctgttcacccgggtggtgcccacccctggtcgagctggaacggcgacgta
 aacggccacaagttcagcgtgtccggcgaggggcgagggcgatgccacctacggcaagctgaccctgaag
 ttcattctgcaaccacggcaagctgcccgtgcccgtgcccacccctcgtgaccacccctgaacctacggcgtg
 cagtgtctcagccgctaccccgaccacatgaagcagcagcacttcttcaagtcggccatgcccgaaggc
 tacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccgcgcccaggtgaagttc
 55 gagggcgacacccctggtgaaccgcacatcgactgaagggcatcgacttcaaggaggacggcaacatcctg
 gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggc
 atcaaggtgaacttcaagatccgcccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
 cagaacacccccatcgggcgacggccccgtgctgctgcccagacaaccactacctgagcaccacagtcggcc
 ctgagcaaagacccccaacgagaagcgcgatcacatggctcctgctggagttcgtgacggccgcccgggac
 60 actctcggcatggacgagctgtacaagggcgcgccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
 AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGGCGCGAAGAGGCGCGCTGTGCCACCAGCTGGGG

GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
GAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCCACCCGCAAGATCGGAGAGCGCCTGGCCGAGATG
GAGGACTACCCGGCTGCCTTGACAGCACCAGCACCAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
5 ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGCACCCACCTGGACATCTATGACCACTGCCAGTCG
AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
GGGACACTGGCCAGGGAGAGCTGAATGAGATGAGGACCCGCCTCTATCTCAACCTGGGCCTCACCTTT
GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCTTTGCGGAGCAGAAC
CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCAC
10 TCCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAG
AGCGAGTGCTGCGTGGTTATTGCACAGGTCCTCCAAGACCTGGGAGACTTTTTGGCTGCCAAGCGAGCC
CTGAAGAAGGCCTACAGGCTGGGCTCCCGAAGCCTGTGACAGAGGGCAGCCATCTGTGAGAACCTCCAG
CATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCTCAGGGTGCC
ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT
15 TACCAGAAGCAGCTGCGTTTTGCTGAGCTGCTGGACAGACCGGTGCTGAGCGGGCCATCATCCACGTG
TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAACCTG
AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCGAGGAG
GCCCGCGATGCCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCCCAG
CGTCCCCAGCTGCAGAGGCAGGTCTTGCAGCATCTCCATACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
20 GCCCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGTGTAGCTGAAGATGAAGATGAGGAGGAGGAGGCG
GAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCCGCGAGGTGGAGCTCTCAGAGGGCGAG
GACGACACCCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCCACCTGGGCGGGCG
AAGGGGAGCAAGTGGAAACCGGCGAAACGACATGGGGAGACCCCTGCTGCACCCAGGCTGCATCGAGGGC
CAGCTGCGCCGCTCCAGGACCTTGTGAGGCAGGGCCACCCCTTAACCCCTCGGAGCTACTGTGGCTGG
25 ACACCTCTGCACGCGGCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGGCC
GCAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCCTCAACTGTGGC
CACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTACCCCTCCGCACTCGAAAGGGCCTC
AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
30 TTCCACACCCCAAGCAGCCTTCTGTTTGACCCCGAGACCTCTCCTCCTTTGAGCCCTGCCCAGAACCC
CTCTTAATAGCACTAGACTCCCAGAGGCCTCTCAGGTCCATGTACGGGTCTCCCAGGGCAGGGCGGCA
CCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGGCCAGCCAGCAGCAGCAGCAGCTCAGAAGGCGAG
GACAGCGCAGGCCCCGCACGGCCGTCCAGAAAGAGGCCTCGGTGCTCGGCCACAGCACAACGGGTGGCA
GCCTGGACGCTGGCCCCGCCAGCAACAGGGAAGCAGCCACAGCCAGCACCAGCCGGGCAGCCTACCAAG
35 GCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGGCTGGGGCCTGGCCACCCGCGGGGCCACAGCAAA
GCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAGGAGTGCTGGCTGGGGACTGGCTGGAGCTGGAC
ATGCCCTGACCCGCAGCCGCCGGCCCCGCCCGGGGCACTGGAGACAACCGCAGGCCCAAGTAGTACC
TCTGGGTGCGACAGTGAGGAGAGCAGGCCCCGTGCCCGAGCCAAGCAGGTCCGCTGACCTGCATGCAG
AGTTGCAGTGCGCCAGTTAACGCAGGGCCAGCAGCTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCC
40 CCCAGGGTCTCAGAGCCAGTGGGGCAGCTCTGCGGCAGGCCAGCCCTTGGGTCCGGCCCCGCCCTT
CCCATCCGGGTTCGAGTTCAAGTTTCAAGATCATCTCTTCTCATCCCTGTCCCACACAGCAGTGACACC
CACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAGCGCTACTACCAGACCTGCGGGCTGTGCCCAGG
CTCACCTACGAAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGCAAT
GACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCCGTTGACTGACCGCTACCGCAGGGCCCTGC
45 CAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTGCAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCG
TTCAGCGCCTGCTCCCTGGCCCTGGACAGGCCACGCTTACACCCCTGCTGCGGGCCCTCAAGCTGCAC
ACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGCTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCT
GCCCTGGGCACCATGCCCAGCCTGGCCCTCCTTGACCTCTCCTCCAATCACCTGGGTCCCGAAGGCCTG
CGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACCTTGACAGAGTTTGGAGgaattagatctatcgatg
50 aACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCCTCCCTCCTGCACGCTGCCCCCTTACTCAGCACC
CTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTCTTTCTGAGCCACCAGACAGCACTGGGTAGTGCT
TTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTGTCTTACAACGCCCTGGGAGCCCCCTGCCCTGGCC
AGGACCCTGCAGAGCCTGCCCGCCGGCACCCCTCTGCACCTTAGAGCTCAGCTCCGTGGCAGCCGGCAAG
GGTGATTGCGACCTCATGGAGCCTGTATTCCGATACCTGGCCAAGGAAGGCTGTGCTCTAGCCACCTG
55 ACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTAGAGACCTGTGCAGATGTCTCTCTGTGCCCC
TCACTCATCTCACTGGATCTGTCTGCCAACCCTGAGATCAGCTGTGCCAGCTTGGAAAGAGCTCCTGTCC
ACCCTCCAAAAGCGGGCCCCAAGGCCTTAGCTTCTTGGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCTG
GGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAACTGCAGCTGTGCAGCAGACGCTCTGC
GCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCAGTCCGGCCGGGGCCCGGCGAGTGCACGCTGGAC
CACGGCTCCAAGCTCTTCTTTGCGCGCCTCTAG

SEQ ID NO: 11 - TONSL N571A

ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGCGAAAGCCAAGGCGCAGAGGGGCCGGGCAGCGG
CGCGAAGAGGCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGCCATGGCCGCTACGCCGAGGCT
5 CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCC
CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGCAGCACCAGCACCAG
TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
ACCCACCTGGACATCTATGACCACTGCCAGTCGAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
10 ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
TTCAGGAAGAGCATCTTCCTTGGCGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGCCCGGGAG
TGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGTTATTGCACAGGTCTCTCAA
GACCTGGGAGACTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
15 GTGCAGAGGGCAGCCATCTGTGAGAACCCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
GAAGAGGCTGAGGGCAGAGACCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTTGCTGAGCTGCTGGAC
AGACCGGGTGCTGAGCGGGCCATCATCCAGTGTCCTTGGCCACCACACTGGGAGACATGAAGGACCAC
CATGGGGCCGTGCGCCACTATGAGGAGGAAGTGAAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
20 ACCTGGCTGAACATTGCACTGTCCCGCAGGAGGCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTC
CAGAAAGCGCTCAGCTGTGCCAGCAGGCCCAGCGTCCCCAGCTGCAGAGGCAGGTCTTGCAGCATCTC
CATAACGTGCAGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGT
GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTG
GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
25 GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGGAAACCGGCGAAACGACATGGGG
GAGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCTCCAGGACCTTGTGAGGCAGGGC
CACCCCTTAACCTTCGGGACTACTGTGGCTGGACACCTCTGCACAGGGCCTGCGCTACGGGCATCTA
GAAATTGTCCGCTTCCTGCTGGACCACGGGGCCGAGTGAGGACGACCCAGGTGGCCAGGGCTGCGAAGGC
ATCACCCCCCTCCACGATGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
30 GCGTCCGTCAACCTCCGCACTCGAAAGGGCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
GCCTCGGGCCAAGATCCCCACAGCTCCCAGGCCTTCCACACCCCCAAGCAGCCTTCTGTTGACCCCGAG
ACCTCTCTCTCTTTGAGCCCCCTGCCAGAACCCCCCTCTAATAGCACTAGACTCCCAGAGGCCTCTCAG
GTCCATGTCAAGGTCTCCCCAGGGCAGGCGGCACCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
35 CACGACGACGAGCAGCTCAGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
CCTCGGTGCTCGGCCACAGCACAACGGGTGGCAGCCTGGACGCTGGCCCCGCCAGCAACAGGGAAGCA
GCCACAGCCAGCACCAGCGGGCAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGG
CTGGGGCCTGGCCACCAGCGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGCAGCCGCCGGCCCCGCCCGGG
40 GGCAGTGGAGACAACCGCAGGCCCAGTAGTACCTCTGGGTGCGACAGTGAGGAGAGCAGGCCCCGTGCC
CGAGCCAAGCAGGTCCGCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCAGGGCCCCAGCAGC
CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCTCAGGGTCTCAGAGCCAGTGGGGACAGCTCTGCG
GCAGCCACGCTTGGGTCCGGCCCCGCCCCCTCCATCCGGTTTCAGTTCAAGTTCAGGATCATCTC
TTCCTCATCCCTGTCCACACAGCAGTGACACCCACTCTGTGGCCTGGCTGGCCGAGCAGGCGGCCAG
45 CGCTACTACCAGACCTGCGGGCTGTGCCAGGCTCACCTACGGAAAGAGGGGGCCCTGTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
CAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCGTTTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAG
CTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGG
50 CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC
CTCTCCTCAATACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCAGGCCAAGCCACC
TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
TCCCTCCTGCACGCTGCCCTTACTCAGCACCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTG
55 TCCTACAACGCCCTGGGAGCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCGCCGGCACCTCCTG
CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTTCGGACCTCATGGAGCCTGTATTCCGATAC
CTGGCCAAGGAAGGCTGTGCTCTAGCCACCTGACCCTGTCTGCAAACACCTGGGGGACAAGGCTGTT
AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCACTCACTGGATCTGTCTGCCAACCCTGAT
ATCAGCTGTGCGAGCTTGAAGAGCTCCTGTCTCACCCTCCAAAAGCGGGCCCCAAGGCCTTAGCTTCCT
60 GGCCTGTGAGGCTGCGCCGTCCAGGGTCCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
CGGGAAGTGCAGCTGTGCAGCAGACGCTCTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
AGTCGGCCGGGCCCCGGCGAGTGACGCTGGACCACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 12 - TONSL N571A / GFP

atggtgagcaagggcgaggagctgttcacccgggtggtgcccacccctggtcgagctggacggcgacgta
aacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaag
5 ttcattctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacctgacctacggcgtg
cagtgtcttcagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgccatgccgaaggc
tacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccgcgcccagggtgaagttc
gagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg
gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggc
10 atcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
cagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcaccacagtcggcc
ctgagcaaaagaccccaacgagagcgaagcgcgatcacatggtcctgctggagttcgtgaccgcggccgggac
actctcggcatggacgagctgtacaaggcgcgcccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGGCGCGAAGAGGCCGCGCTGTGCCACCAGCTGGGG
15 GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
GAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCCACCGCAAGATCGGAGAGCGCCTGGCCGAGATG
GAGGACTACCCGGCTGCCTTGCAACACCAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGCACCCACCTGGACATCTATGACCACTGCCAGTGC
AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
20 GGGACACTGGCCAGGGAGAGCTGAATGAGATGAGGACCCGCTCTATCTCAACCTGGGCCCTCACCTTT
GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCCTTGCGGAGCAGAAC
CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCAC
TCCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAGTGTGCGCACACCATGAGGAAGCGTTTCATGGAG
AGCGAGTGCTGCGTGGTTATTGACACAGGTCTCCAAGACCTGGGAGACTTTTGGCTGCCAAGCGAGCC
25 CTGAAGAAGGCCCTACAGGCTGGGCTCCAGAAGCCTGTGCAGAGGGCAGCCATCTGTGCAACCTCCAG
CATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCTCAGGGTGCC
ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT
TACCAGAAGCAGCTGCGTTTTTGTGAGCTGCTGGACAGACCCGGGTGCTGAGCGGGCCATCATCCACGTG
TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAAGT
30 AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCAGGAG
GCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCAG
CGTCCCCAGCTGCAGAGGCAGGTCTTGACAGCATCTCCATACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
GCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGTGTAGCTGAAGATGAAGATGAGGAGGAGGAGGCG
GAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCCGCGAGGTGGAGCTCTCAGAGGGCGAG
35 GACGACACCGATGGCCTGACCCCGAGCTGGAGGAGGACGAGGAGCTTCAGGGACCTGGGCCGCGG
AAGGGGAGCAAGTGAACCGGCAACGACATGGGGGAGACCTGCTGCACCGAGCTGCATCGAGGGC
CAGCTGCGCCGCTCCAGGACCTTGTGAGGCAGGGCCACCCCTTAACCTCGGGACTACTGTGGCTGG
ACACCTCTGCACGAGGCCTGCGCCTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGGCC
GCAGTGAGCAGCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCCTCAACTGTGGC
40 CACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTCACCTCCGCACTCGAAAGGGCCTC
AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
TTCCACACCCCAAGCAGCCTTCTGTTTACCCCGAGACCTCTCCTCCTTTGAGCCCTGCCAGAACCC
CCCTCTAATAGCACTAGACTCCCAGAGGCTCTCAGGTCCATGTGAGGGTCTCCCCAGGGCAGGCGGCA
45 CCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGGCCAGCCAGCAGCAGCAGCTCAGAAGGCGAG
GACAGCGCAGGCCCCGCACGGCCGTCCAGAAGAGGCCTCGGTGCTCGGCCACAGCACACGGGTGGCA
GCCTGGACGCCTGGCCCCGCCAGCAACAGGGAAGCAGCCACAGCCAGCACCAGCCGGGCAGCCTACCAG
GCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGGCTGGGGCCTGGCCACCGCGGGGCCACAGCAAA
GCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAGGAGTGCCTGGCTGGGGACTGGCTGGAGCTGGAC
50 ATGCCCTGACCCGCGAGCCGCGGCCCGCCCGGGGCACTGGAGACAACCGCAGGCCCAGTAGTACC
TCTGGGTGCGACAGTGAGGAGAGCAGGCCCGCTGCCCGAGCCAAGCAGGTCCGCTGACCTGCATGCAG
AGTTGCAGTGCGCCAGTTAACGCAGGGGCCAGCAGCCTGGCTTCAGAACCTCCAGGGAGCCCCAGCACC
CCCAGGGTCTCAGAGCCAGTGGGGACAGCTCTGCGGCAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCT
CCCATCCGGGTTTCAGTTCAAGTTCAAGATCATCTCTTCTCATCCCTGTCCCACACAGCAGTGACACC
55 CACTCTGTGGCTGGCTGGCCGAGCAGGCGGCCAGCGCTACTACCAGACCTGCGGGCTGTGCCCAGG
CTCACCTACGGAAAGAGGGGGCCCTGCTGGCCCCACAGGACCTCATCCCTGATGTGCTGCAGAGCAAT
GACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTGCCCCCGTTGACTGACCGCTACCGCAGGGCCTGC
CAGAGCCTGGGGCAGGGGAGCACAACAGGTGCTGAGGCGGCTGGAGCTCCAGGGCTTGGGCTCTCG
TTACGCGCCTGCTCCCTGGCCCTGGACAGGCCCCAGCTTACACCCCTGCTGCGGGCCCTCAGCTGCAC
60 ACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGGCTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCT
GCCCTGGGCACCATGCCCAGCCTGGCCCTCCTTGACCTCTCTCCAATCACCTGGGTCCCGAAGGCCTG
CGCCAGCTTGCCATGGGGCTCCAGGCCAAGCCACCTTGACAGAGTTTGGAGgaattagatctatcgatg

aACCCCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCCTCCCTCCTGCACGCCTGCCCCCTTACTCAGCACC
CTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTCTTTCTGAGCCACCAGACAGCACTGGGTAGTGCT
TTCCAAGATGCTGAGCACCTGAAGACCCTGTCCCTGTCTTACAACGCCCTGGGAGCCCCCTGCCCTGGCC
5 AGGACCCTGCAGAGCCTGCCCGCCGGCACCCTCCTGCACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAG
GGTGATTCGGACCTCATGGAGCCTGTATTCCGATACCTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTG
ACCCGTGTCTGCAAACCACCTGGGGGACAAGGCTGTTAGAGACCTGTGCAGATGTCTCTCTGTGCCCC
TCACTCATCTCACTGGATCTGTCTGCCAACCCTGAGATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCC
ACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTTGGCCTGTCAGGCTGCGCCGTCCAGGGTCCCCTG
10 GGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTCCGGGAACTGCAGCTGTGCAGCAGACGCCTCTGC
GCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCCAGTCGGCCGGGCCCCGGCGAGTGCACGCTGGAC
CACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 13 - TONSL D604A

15 ATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGCAAGGGCGAAAGCCAAGGCGCAGAGGGCCGGGCAGCGG
CGCGAAGAGGGCCGCGCTGTGCCACCAGCTGGGGGAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCT
CTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGGGAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCC
CACCGCAAGATCGGAGAGCGCCTGGCCGAGATGGAGGACTACCCGGCTGCCTTGACAGCACCAGCACCAG
TACCTGGAGCTGGCACATTCCCTGCGCAACCACACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGC
20 ACCACCTGGACATCTATGACCACTGCCAGTGCAGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAG
AAGAGCTTGGCTATTGTGGATGAGGAGCTGGAGGGGACACTGGCCCAGGGAGAGCTGAATGAGATGAGG
ACCCGCCTCTATCTCAACCTGGGCCTCACCTTTGAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTAC
TTCAGGAAGAGCATCTTCTTTCGGGAGCAGAACCACCTTTACGAGGACCTATTCCGCGCCCCGCTACAAC
CTGGGCACCATCCACTGGCGCGCGGGCCAGCACTCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAG
TGTGCGCACACCATGAGGAAGCGGTTTCATGGAGAGCGAGTGCTGCGTGGTTATTGCACAGGTCCTCCAA
25 GACCTGGGAGACTTTTTTGGCTGCCAAGCGAGCCCTGAAGAAGGCCTACAGGCTGGGCTCCCAGAAGCCT
GTGCAGAGGGCAGCCATCTGTGCAACCTCCAGCATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTG
GAAGAGGCTGAGGGCAGAGACCCTCAGGGTGCCATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCC
AAGGCAGGAGACTTTCCAGGGCAGCTGAGGCTTACCAGAAGCAGCTGCGTTTTTGTGAGCTGCTGGAC
AGACCGGGTGCTGAGCGGGCCATCATCCAGTGTCCCTGGCCACCACACTGGGAGACATGAAGGACCAC
30 CATGGGGCCGTGCGCCACTATGAGGAGGAACAGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAG
ACCTGGCTGAACATTGCACTGTCCCGCAGGAGGCGGAGCTGCCTACGAGCTGCTGGCCCCGTGCTTC
CAGAAAGCGCTCAGCTGTGCCCAGCAGGGCCAGCGTCCCGAGCTGCAGAGGCAAGGTCTTGCAAGCATCTC
CATAACGTGCAGCTGAGGCTGCAGCCCCAGGAGGCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGT
GTAGCTGAAGATGAAGATGAGGAGGAGGAGGCGGAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTG
35 GAGGCCGGCGAGGTGGAGCTCTCAGAGGGCGAGGACGACACCGATGGCCTGACCCCGCAGCTGGAGGAG
GACGAGGAGCTTCAGGGCCACCTGGGCCGGCGGAAGGGGAGCAAGTGGAACCGGCGAAACGACATGGGG
GAGACCCTGCTGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGC
CACCCCTTAACCTTCGGGACTACTGTGGCTGGACACCTCTGCACGAGGCCTGCAACTACGGGCATCTA
GAAATTGTTCCGCTTCTGTGGTGGACACCGGGCCAGTGGACGACCCAGGTGGCCAGGCTGCCAAGGC
40 ATCACCCCCCTCCAC**GCT**GGCCCTCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGG
GCGTCCGTCAACCTCCGCACTCGAAAGGGCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTG
TACCGCAGGGACCTGGACCTGGAGACGCGGCAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCT
GCCTCGGGCCAAAGATCCCCACAGCTCCAGGCCTTCCACACCCCCAAGCAGCCTTCTGTTTACCCCCGAG
ACCTCTCTCTCTTTGAGCCCCCTGCCAGAACCCCCCTCTAATAGCACTAGACTCCAGAGGCCTCTCAG
45 GTCCATGTGAGGCTCTCCCCAGGGCAGGCGGCACCAGCCATGGCCAGGCCTCGGAGGAGCAGGCATGGG
CCAGCCAGCAGCAGCAGCTCAGAAGGGCGGAGACAGCGCAGGCCCCGACGGCCGTCCCAGAAAGAG
CTCGGTCAGTCGGCCACAGCACAAAGGTGGAGCTGAGCAGCTGGCCCTGGCCCCAGCAACAGGGAAGCA
GCCACAGCCAGCACCAGCCGGGCAGCCTACCAGGCAGCCATCCGGGGTGTGGGCAGTGTGAGAGCCGG
50 CTGGGGCCTGGCCACCAGCGGGGCCACAGCAAAGCCCTTGCCCCCAGGCAGCGCTCATCCCGGAGGAG
GAGTGCCTGGCTGGGGACTGGCTGGAGCTGGACATGCCCTGACCCGACGCGCGGCCCCGCCCCCGG
GGCACTGGAGACAACCGCAGGCCCCAGTAGTACCTCTGGGTGGACAGTGAGGAGAGCAGGCCCCGTGCC
CGAGCCAAGCAGGTCCGCTGACCTGCATGCAGAGTTGCAGTGCGCCAGTTAACGCAGGGCCCCAGCAGC
CTGGCTTCAGAACCTCCAGGGAGCCCCAGCACCCCCAGGGTCTCAGAGCCCAGTGGGGACAGCTCTGCG
GTAGGCCAGCCCTTGGGTCCGGCCCCGCCCCCTCCCATCCGGGTTTCAGGTTCAAGTTCAAGATCATCTC
55 TTCTCATCCCTGTCCACACAGCACTGACACCCACTCTGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG
CGCTACTACCAGACCTGCGGGCTGCTGCCAGGCTCACCTACGGAAAGAGGGGGCCCTGCTGGCCCCA
CAGGACCTCATCCCTGATGTGCTGCAGAGCAATGACGAGGTGTTGGCTGAGGTGACTTCGTGGGACCTG
CCCCCGTTGACTGACCGCTACCGCAGGGCCTGCCAGAGCCTGGGGCAAGGGGAGCACCAACAGGTGCTG
CAGGCCGTGGAGCTCCAGGGCTTGGGCCTCTCGTTTCAGCGCCTGCTCCCTGGCCCTGGACCAGGCCAG
60 CTTACACCCCTGCTGCGGGCCCTCAAGCTGCACACAGCACTCCGGGAGCTGCGCCTGGCAGGGAACCGG
CTGGGGGACAAGTGTGTGGCTGAGCTGGTGGCTGCCCTGGGCACCATGCCAGCCTGGCCCTCCTTGAC

CTCTCCTCCAATCACCTGGGTCCCGAAGGCCTGCGCCAGCTTGCCATGGGGCTCCCAGGCCAAGCCACC
 TTGCAGAGTTTGGAGgaattagatctatcgatgaACCCCTGGGGGACGGCTGTGGCCAGTCCCTGGCC
 TCCCTCCTGCACGCCTGCCCCCTTACTCAGCACCTGCGCCTGCAGGCGTGTGGCTTCGGCCCCAGCTTC
 TTTCTGAGCCACCAGACAGCACTGGGTAGTGCTTTCCAAGATGCTGAGCACCTGAAGACCTGTCCCTG
 5 TCCTACAACGCCCTGGGAGCCCCCTGCCCTGGCCAGGACCCTGCAGAGCCTGCCCCGCCGGCACCCCTCCTG
 CACTTAGAGCTCAGCTCCGTGGCAGCCGGCAAGGGTGATTTCGGACCTCATGGAGCCTGTATTCCGATAC
 CTGGCCAAGGAAGGCTGTGCTCTAGCCCACCTGACCCTGTCTGCAAACCACCTGGGGGACAAGGCTGTT
 AGAGACCTGTGCAGATGTCTCTCTGTGCCCCCTCACTCATCTCACTGGATCTGTCTGCCAACCCCTGAG
 10 ATCAGCTGTGCCAGCTTGGAAGAGCTCCTGTCCACCCTCCAAAAGCGGCCCAAGGCCTTAGCTTCCTT
 GGCCTGTCAGGCTGCGCCGTCCAGGGTCCCCTGGGCCTGGGCCTGTGGGACAAGATAGCCGCGCAGCTC
 CGGGAACCTGCAGCTGTGCAGCAGACGCCTTGTGCGCTGAGGACAGGGACGCCCTGCGCCAGCTGCAGCCC
 AGTCGGCCGGGCCCCGGCGAGTGACAGCTGGACCACGGCTCCAAGCTCTTCTTTTCGGCGCCTCTAG

SEQ ID NO: 14 - TONSL D604A / GFP

15 atggtgagcaagggcgaggagctgttcacccgggtggtgcccacccctggtcgagctggacggcgacgta
 aacggccacaagttcagcgtgtccggcgaggggcgaggcgatgccacctacggcaagctgacctgaag
 ttcacatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtagccacctgaacctacggcgtg
 cagtgtctcagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgccatgcccgaggc
 20 taagtccaggagcgacccatcttcttcaaggacgacggcaactacaagaccggcgccgagggtgaagttc
 gagggcgacacccctggtgaaccgcacgcagctgaagggcactcgacttcaaggaggacggcaacatcctg
 gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggc
 atcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
 cagaacacccccatcgggcagggccccgtgctgctgcccagacaaccactacctgagcaccacagtcggcc
 25 ctgagcaaaagacccccaacgagaagcgcgatcacatggtcctgctggtgagttcgtgaccggcgccgggatc
 actctcggcacatggacgagctgtacaagggcgcgcccaATGAGCCTGGAGCGCGAGCTTCGCCAGCTGAGC
 AAGGCGAAAGCCAAGGCGCAGAGGGCCGGGCGAGCGCGCGAAGAGGCGCGCTGTGCCACCAGCTGGGG
 GAGCTCCTGGCCGGCCATGGCCGCTACGCCGAGGCTCTGGAGCAGCACTGGCAGGAGCTGCAGCTTCGG
 GAGCGCGCTGACGACCCTCTGGGCTGTGCCGTGGCCACCCGCAAGATCGGAGAGCGCCTGGCCGAGATG
 GAGGACTACCCGGCTGCCTTGACGACACAGCACCAGTACCTGGAGCTGGCACATTCCCTGCGCAACCAC
 30 ACGGAGCTGCAGAGGGCCTGGGCCACCATCGGCCGCACCCACCTGGACATCTATGACCACTGCCAGTCG
 AGGGATGCTTTGCTGCAGGCACAGGCTGCCTTTGAGAAGAGCTTGGCTATTGTGGATGAGGAGCTGGAG
 GGGACACTGGCCAGGGAGAGCTGAATGAGATGAGGACCCGCCTCTATCTCAACCTGGGCCTCACCTTT
 GAGAGCCTGCAGCAGACAGCCCTGTGCAACGATTACTTCAGGAAGAGCATCTTCTTGGCGAGCAGAAC
 CACCTTTACGAGGACCTATTCCGCGCCCGCTACAACCTGGGCACCATCCACTGGCGCGCGGGCCAGCAC
 35 TCCCAGGCTATGCGCTGCTTGGAGGGTGGCCGGGAGTGTGCGCACACCATGAGGAAGCGGTTTCATGGAG
 AGCGAGTGCTGCGTGGTTATTGCACAGGTCCTCCAAGACCTGGGAGACTTTTGGCTGCCAAGCGAGCC
 CTGAAGAAGGCCTACAGGCTGGGCTCCAGAAAGCCTGTGCAGAGGGCAGCCATCTGTGAGAACCTCCAG
 CATGTGCTGGCAGTGGTCCGGCTGCAGCAACAGCTGGAAGAGGCTGAGGGCAGAGACCTCAGGGTGCC
 ATGGTCATCTGTGAGCAGCTAGGGGACCTCTTCTCCAAGGCAGGAGACTTTCCAGGGCAGCTGAGGCT
 40 TACCAAGACAGCTGCGTTTTGCTGAGCTGCTGGACAGACCGGGTGCTGAGCGGGCCATCATCCAGTG
 TCCCTGGCCACCACACTGGGAGACATGAAGGACCACCATGGGGCCGTGCGCCACTATGAGGAGGAAGT
 AGGCTGCGCAGCGGCAACGTGCTGGAGGAGGCCAAGACCTGGCTGAACATTGCACTGTCCCGCGAGGAG
 GCCGGCGATGCCTACGAGCTGCTGGCCCCGTGCTTCCAGAAAGCGCTCAGCTGTGCCAGCAGGCCCAG
 CGTCCCCAGCTGCAGAGGCAGGTCTTGACGATCTCCATACCGTGCAGCTGAGGCTGCAGCCCCAGGAG
 45 GCCCCCTGAGACCGAAACCAGACTGCGGGAGCTCAGTGATAGCTGAAGATGAAGATGAGGAGGAGGAGGCG
 GAGGAGGCGGCAGCCACAGCGGAGAGCGAAGCCCTGGAGGCGCGGAGGTGGAGCTCTCAGAGGGCGAG
 GACGACACCGATGGCCTGACCCCGCAGCTGGAGGAGGACGAGGAGCTTCAGGGCCACCTGGGCCGGCGG
 AAGGGGAGCAAGTGGAAACCGGCGAAACGACATGGGGGAGACCTGCTGCACCGAGCCTGCATCGAGGGC
 CAGCTGCGCCGCGTCCAGGACCTTGTGAGGCGAGGGCCACCCCTTAACCTTCGGGACTACTGTGGCTGG
 50 ACACCTCTGCACGAGGCCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCTGCTGGACCACGGGGCC
 GCAGTGAGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCAC**GCT**GGCCTCAACTGTGGC
 CACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGTCCGTCACCCTCCGCACTCGAAAGGGCCTC
 AGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGGCAG
 AAGGCCAGGGCCATGGAGATGCTGCTCCAGGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCC
 55 TTCCACACCCCAAGCAGCCTTCTGTTTGACCCCGAGACCTCTCCTCCTTTGAGCCCTGCCAGAACCC
 CCGCTCTAATAGCACTAGACTCCAGAGGCCTCTCAGGTCATGTGAGGCTCTCCAGGGCAGGCGGCA
 CCAGCCATGGCCAGGCCTCGGAGGAGCGGATGGGCCAGCCAGCAGCAGCAGCTCAGAAGGCGAG
 GACAGCGCAGGCCCCGCACGGCCGTCCAGAAAGAGGCCTCGGTGCTCGGCCACAGCACACGGGTGGCA
 GCCTGGACGCCTGGCCCCGCCAGCAACAGGGAAGCAGCCACAGCCAGCAGCCGGGCAGCCTACAG
 60 GCAGCCATCCGGGGTGTGGGCAGTGCTCAGAGCCGGCTGGGGCCTGGCCACCGCGGGGCCACAGCAAA
 GCCCTTGCCCCCAGGCAGCGCTCATCCCGAGGAGGAGTGCTGGCTGGGGACTGGCTGGAGCTGGAC

30
35

GGATCCGGGGCCCCCTGGAGGAAGAAGAGGATGGAGAGGAGCTCATTGGAGATGGCATGGAAAGGGACTACCGCG
CCATCCCAGAGCTGGACGCCTATGAGGCCGAGGGACTGGCTCTGGATGATGAGGACGTAGAGGAGCTGACGGC
CAGTCAGAGGGAGGCAGCAGAGCGGGCCATGCGGCAGCGTGACCGGGAGGCTGCCCCGGGGCCTGGGCCGAGGA
GGTGGAGGTGGCCACCTGGGGCCGGCGGAAGGGGAGCAAGTGAACCGGGCGAAACGACATGGGGGAGACCCTGC
TGCACCGAGCCTGCATCGAGGGCCAGCTGCGCCGCGTCCAGGACCTTGTGAGGCAGGGCCACCCCTTAACCC
TCGGGACTACTGTGGCTGGACACCTCTGCACGAGGCCTGCAACTACGGGCATCTAGAAATTGTCCGCTTCCTG
CTGGACCACGGGGCCGCAGTGGACGACCCAGGTGGCCAGGGCTGCGAAGGCATCACCCCCCTCCACGATGCCC
TCAACTGTGGCCACTTCGAGGTGGCTGAGCTGCTGCTTGAACGGGGGGCGCTCCGTCACCCCTCCGCACTCGAAA
GGGCCTCAGCCCGCTGGAGACGCTGCAGCAGTGGGTGAAGCTGTACCGCAGGGACCTGGACCTGGAGACGCGG
CAGAAGGCCAGGGCCATGGAGATGCTGCTCCAGCGGCTGCCTCGGGCCAAGATCCCCACAGCTCCCAGGCCT
TCCACACCCCAAGCAGCCTTCTGTTTGACCCCCGAGACCTCTTAACCTCGAG

40 >gi|187608777|ref|NP_038460.4| tonsoku-like protein [Homo sapiens]
MSLERELRQLSKAKAKAQRAGQRREEAALCHQLGELLAGHGRYAAEALQHWQELQLRERADDPLGCAVAHRKI
GERLAEMEDYPAALQHQQHYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
ELEGTLAQGELNEMRTRLYLNLGLTFESLQQTALCNDYFRKSI FLAEQNHLYEDLFRARYNLGTIHWRAGQHS
QAMRCLEGARECAHTMRKRFMSECCVVIAQVQLDGLDFLAAKRALKKAYRLGSGQKPVQRAAICQNQLHVLAV
45 VRLQGGLEEAEGRDPQGAMVCEQLDGLDFSKAGDFPRAAEAYQKRLFAELLDRPQAEARI IHVSLATTLGDM
KDHHGAVRHYEEELRLRSGNVLEEAKTLWNIALSREEAGDAYELLAPCFQKALSQAQQAPRQLQRQVLQLHLH
TVQLRLQPQEAPETETRLRELSVAEDEDEEEEAEEAAATAESEALEAGEVELSEGEDDDTGLTPQLEEDEELQ
GHLGRRKGSKWNRRNDMGE^ETL^LLRACIEGQLRRVQDLVRQGHPLNPR^DYCG^WTP^LHE^AC^NYGHLEIVRFLLDH
GAAVDDPGGQGCEGITPLH^DALNCGHFEVAELLLERGASVTLRTRKGLSPLET^LQQWVKLYRRDL^DLETRQKA
50 RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPETSPPLSPCPPEPPSNSTRLPEASQAHVRVSPGQAAPAMARF
RRSRHGPASSSSSEGEDSAGPARPSPQKRPRCSATARVAAWTPGPASNREAASTASRAAYQAAIRGVGSAQ
SRLGPGPPRGHSSKALPAQAA^LIP^EEE^ECLAGDWLELD^MPL^TRSRRPRPGTGDNRRPSSTSGSDSEESRPRARA
KQVRLTCMQSCSAPVNAGPSSLASPEPPGSPSTPRVSEPSGSDSSAAGQPLGPAPPPP IRVRVQVQDHLFLIPVF
HSSDTHSVAWLAEQAAQRYRYQT^CG^LL^PRL^TLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
55 CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRALKLHTALRELRLAGNRLGDKCVAELVAALG
TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSL^EE^LDL^SMN^PLGDGCGQSLASLLHACPLLSTLR^LQACG
FGPSFFLSHQ^TALGSAFQDAEHLK^TLS^LSYNALGAPALARTLQSLPAGTLLHLELSSVAAGKQSD^LMEPVFR
YLAKEGICALAHLT^LLSANHLGDKAVRDL^RCL^SLCPS^LIS^LDL^SANPE^ISCASLEELLSTLQKRP^GCL^SFLGLS
GCAVQGLGLGLWDK^LIAAQLRELQ^LCSRR^LCAEDRDALROLOPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO: 17 - Protein sequences of TONSL E530A single mutants (mutated alanins are highlighted)

MSLERELRQLSKAKAKAQRAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
 5 GERLAEMEDYPAALQHQQHYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
 ELEGTLAQGELNEMRTRLRYLNLGLTFESLQQTALCNDYFRKSI FLAEQNHLYEDLFRARYNLGTIHWRAQQHS
 QAMRCLEGARECAHTMRKRFMESECCVVIAQVLQDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAI IHVSLATTLGDM
 KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQQAQRQPQLQRQVLQHLH
 10 TVQLRLQPQEAPETETRLRELSVAEDEDEEEEAEEAAATAESEALEAGEVELSEGEDDDTGLTPQLEEDEELQ
 GHLGRRKGSKWNRNRNDMGATLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHEACNYGHLEIVRFLLDH
 GAAVDDPGGQCGEGITPLHDALNCGHFEVAELLERLGASVTLTRKGLSPLETQLQWVKLYRRDLLETRQKA
 RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPETSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
 RRSRHGPASSSSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQA AIRGVGSAQ
 15 SRLGPGPPRGHSHKALAPQAALIPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
 KQVRLTCMQSCSAPVNAGPSSSLASEPPGSPSTPRVSEPSGDSSAAGQPLGPAPPPP IRVRVQVQDHLFLIPVP
 HSSDTHSVAWLAEQAAQRYRYQTCGLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
 CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRALKLHTALRELRLAGNRLGDKCVAELVAALG
 TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLLEELDLSMNPLGDGCGQSLASLLHACPLLSTLRQLQACG
 20 FGPSFFLSHQ TALGSAFQDAEHLKTL SLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGDSDLMEPVFR
 YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
 GCAVQGPLGLGLWDKIAAQLRELQLCSRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO: 18 - Protein sequences of TONSL D559A

MSLERELRQLSKAKAKAQRAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
 25 GERLAEMEDYPAALQHQQHYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
 ELEGTLAQGELNEMRTRLRYLNLGLTFESLQQTALCNDYFRKSI FLAEQNHLYEDLFRARYNLGTIHWRAQQHS
 QAMRCLEGARECAHTMRKRFMESECCVVIAQVLQDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAI IHVSLATTLGDM
 KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQQAQRQPQLQRQVLQHLH
 30 TVQLRLQPQEAPETETRLRELSVAEDEDEEEEAEEAAATAESEALEAGEVELSEGEDDDTGLTPQLEEDEELQ
 GHLGRRKGSKWNRNRNDMGATLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHEACNYGHLEIVRFLLDH
 GAAVDDPGGQCGEGITPLHDALNCGHFEVAELLERLGASVTLTRKGLSPLETQLQWVKLYRRDLLETRQKA
 RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPETSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
 RRSRHGPASSSSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQA AIRGVGSAQ
 35 SRLGPGPPRGHSHKALAPQAALIPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
 KQVRLTCMQSCSAPVNAGPSSSLASEPPGSPSTPRVSEPSGDSSAAGQPLGPAPPPP IRVRVQVQDHLFLIPVP
 HSSDTHSVAWLAEQAAQRYRYQTCGLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
 CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRALKLHTALRELRLAGNRLGDKCVAELVAALG
 TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLLEELDLSMNPLGDGCGQSLASLLHACPLLSTLRQLQACG
 40 FGPSFFLSHQ TALGSAFQDAEHLKTL SLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGDSDLMEPVFR
 YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
 GCAVQGPLGLGLWDKIAAQLRELQLCSRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO: 19 - Protein sequences of TONSL W563A

MSLERELRQLSKAKAKAQRAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
 45 GERLAEMEDYPAALQHQQHYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
 ELEGTLAQGELNEMRTRLRYLNLGLTFESLQQTALCNDYFRKSI FLAEQNHLYEDLFRARYNLGTIHWRAQQHS
 QAMRCLEGARECAHTMRKRFMESECCVVIAQVLQDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAI IHVSLATTLGDM
 KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQQAQRQPQLQRQVLQHLH
 50 TVQLRLQPQEAPETETRLRELSVAEDEDEEEEAEEAAATAESEALEAGEVELSEGEDDDTGLTPQLEEDEELQ
 GHLGRRKGSKWNRNRNDMGATLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHEACNYGHLEIVRFLLDH
 GAAVDDPGGQCGEGITPLHDALNCGHFEVAELLERLGASVTLTRKGLSPLETQLQWVKLYRRDLLETRQKA
 RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPETSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
 RRSRHGPASSSSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQA AIRGVGSAQ
 55 SRLGPGPPRGHSHKALAPQAALIPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
 KQVRLTCMQSCSAPVNAGPSSSLASEPPGSPSTPRVSEPSGDSSAAGQPLGPAPPPP IRVRVQVQDHLFLIPVP
 HSSDTHSVAWLAEQAAQRYRYQTCGLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
 CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRALKLHTALRELRLAGNRLGDKCVAELVAALG
 TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLLEELDLSMNPLGDGCGQSLASLLHACPLLSTLRQLQACG
 60 FGPSFFLSHQ TALGSAFQDAEHLKTL SLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGDSDLMEPVFR

YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
GCAVQGPLGLGLWDKIAAQLRELQLCSRRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO: 20 - Protein sequences of TONSL E568A

5 MSLERELRQLSKAKAKAQAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
GERLAEMEDYPAALQHQQHLYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
ELEGTLAQGELNEMRTRLYLNLGLTFESLQQTALCNDYFRKSIPLAEQNHLIEDLFRARYNLGTIHWACQHS
QAMRCLEGARECAHTMRKRFMESECCVVIAQVLDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
10 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAIHVSLATTLGDM
KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQAQRPQLQRQVLQHLH
TVQLRLQPQEAPEPETETRLRELSVAEDEDEEEEAEEAAATAESEAEEAGEVELSEGEDDDTGLTPQLEEEDEELQ
GHLGRRKGSKWNRRNDMGETLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHAAACNYGHLEIVRFLLDH
GAAVDDPGGQCGEGITPLHDALNCGHFEVAELLERLGRASVTLRTRKGLSPLETQQVWKLYRRDLDLLETRQKA
RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPEPSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
15 RRSRHGPASSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQAIRGVGSAQ
SRLGPGPPRGHSHKALAPQAALIPPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
KQVRLTCMQSCSAPVNAGPSSSLASEPPGSPSTPRVSEPSGDSSAAGQPLGPAPPPPPIRVRVQVQDHLFLIPVP
HSSDTHSVAWLAEQAQRYYYQTGCLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRLKHLHTALRELRLAGNRLGDKCVAELVAALG
20 TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLLEELDLSMNLPGDGGQSLASLLHACPLLSLTLRLQACG
FGPSFFLSHQ TALGSFAQDAEHLKTLSLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGSDLMPEPVFR
YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
GCAVQGPLGLGLWDKIAAQLRELQLCSRRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO: 21 - Protein sequences of TONSL N571A

MSLERELRQLSKAKAKAQAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
GERLAEMEDYPAALQHQQHLYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
ELEGTLAQGELNEMRTRLYLNLGLTFESLQQTALCNDYFRKSIPLAEQNHLIEDLFRARYNLGTIHWACQHS
QAMRCLEGARECAHTMRKRFMESECCVVIAQVLDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
30 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAIHVSLATTLGDM
KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQAQRPQLQRQVLQHLH
TVQLRLQPQEAPEPETETRLRELSVAEDEDEEEEAEEAAATAESEAEEAGEVELSEGEDDDTGLTPQLEEEDEELQ
GHLGRRKGSKWNRRNDMGETLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHAAACNYGHLEIVRFLLDH
GAAVDDPGGQCGEGITPLHDALNCGHFEVAELLERLGRASVTLRTRKGLSPLETQQVWKLYRRDLDLLETRQKA
RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPEPSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
35 RRSRHGPASSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQAIRGVGSAQ
SRLGPGPPRGHSHKALAPQAALIPPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
KQVRLTCMQSCSAPVNAGPSSSLASEPPGSPSTPRVSEPSGDSSAAGQPLGPAPPPPPIRVRVQVQDHLFLIPVP
HSSDTHSVAWLAEQAQRYYYQTGCLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
40 CQSLGQGEHQVQLQAVELQGLGLSFSACSLALDQAQLTPLLRLKHLHTALRELRLAGNRLGDKCVAELVAALG
TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLLEELDLSMNLPGDGGQSLASLLHACPLLSLTLRLQACG
FGPSFFLSHQ TALGSFAQDAEHLKTLSLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGSDLMPEPVFR
YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
GCAVQGPLGLGLWDKIAAQLRELQLCSRRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

45

SEQ ID NO: 22 - Protein sequences of TONSL D604A

MSLERELRQLSKAKAKAQAGQRREEAALCHQLGELLAGHGRYAEALEQHWQELQLRERADDPLGCAVAHRKI
GERLAEMEDYPAALQHQQHLYLELAHSLRNHTELQRAWATIGRTHLDIYDHCQSRDALLQAQAAFEKSLAIVDE
ELEGTLAQGELNEMRTRLYLNLGLTFESLQQTALCNDYFRKSIPLAEQNHLIEDLFRARYNLGTIHWACQHS
QAMRCLEGARECAHTMRKRFMESECCVVIAQVLDLGDFLAAKRALKKAYRLGSQKPVQRAAICQNLQHVLA
50 VRLQQQLEEAEGRDPQGAMVICEQLGDLFSKAGDFPRAAEAYQKQLRFAELLDRPGAERAIHVSLATTLGDM
KDHHGAVRHYEEELRLRSGNVLEEAKTWNIALSREEAGDAYELLAPCFQKALSQAQAQRPQLQRQVLQHLH
TVQLRLQPQEAPEPETETRLRELSVAEDEDEEEEAEEAAATAESEAEEAGEVELSEGEDDDTGLTPQLEEEDEELQ
GHLGRRKGSKWNRRNDMGETLLHRACIEGQLRRVQDLVRQGHPLNPRDYCGWTPLHAAACNYGHLEIVRFLLDH
GAAVDDPGGQCGEGITPLHAAALNCGHFEVAELLERLGRASVTLRTRKGLSPLETQQVWKLYRRDLDLLETRQKA
RAMEMLLQAAASGQDPHSSQAFHTPSSLLFDPEPSPPLSPCEPPSNSTRLPEASQAHVRVSPGQAAPAMARP
55 RRSRHGPASSSSSEGEDSAGPARPSQKRPRCSATAQRVAAWTPGPASNREAATASTSRAYQAIRGVGSAQ

5 SRLGPGPPRGHSKALAPQAALIPEEECLAGDWLELDMPLTRSRPRPRGTGDNRRPSSTSGSDSEESRPRARA
 KQVRLTCMQSCSAPVNAGPSSLAEPGSPSTPRVSEPSGDSSAAGQPLGPAPPPPIRVRVQVQDHLFLIPVP
 HSSDTHSVAWLAEQAAQRYYYQTGGLLPRLTLRKEGALLAPQDLIPDVLQSNDEVLAEVTSWDLPLPLTDYRRA
 CQSLGQGEHQVVLQAVELQGLGLSFSACSLALDQAQLTPLLRLKLHTALRELRLAGNRLGDKCVAELVAALG
 TMPSLALLDLSSNHLGPEGLRQLAMGLPGQATLQSLEELDLSMNPLGDGCGQSLASLLHACPLLSTLRLQACG
 FGPSFFLSHQ TALGSAFQDAEHLKTL SLSYNALCAPALARTLQSLPAGTLLHLELSSVAACKGSDLMPEVFR
 YLAKEGCALAHLTLSANHLGDKAVRDLRCRLSLCPSLISLDLSANPEISCASLEELLSTLQKRPQGLSFLGLS
 GCAVQGPLGLGLWDKIAAQLRELQLCSRRLCAEDRDALRQLQPSRPGPGECTLDHGSKLFFRRL

SEQ ID NO:23 – protein sequence of Histone H4

10 >sp|P62805|H4_HUMAN Histone H4 OS=Homo sapiens

MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK
 VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRITLYGFGG

SEQ ID NO:24 – protein sequence of MCM2

15 >sp|P49736|MCM2_HUMAN DNA replication licensing factor MCM2 OS=Homo sapiens

MAESSESFTMASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTE
 GPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQR
 DREAGRGLGRMRRLLYDSDEEDEERPARRRQVERATEDGEEDEEMIESIENLEDLKGH
 SVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAR
 20 EHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQL
 HLNQLIRTSGVVTSTCTGVLPQLSMVKYCNKCNFVLGPFQSQNQEVKPGSCPECQSAGP
 FEVNMEETIYQNYQIRIQESPGKVAAGRLPRSKDAILLADLVDCKPGDEIELTGIYHN
 NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFAS
 IAPSIYGHEDIKRLALALFGGEPKNPGGKHKVRGDI NVLLCGDPGTAKSQFLKYIEKVS
 25 SRAIFTTGQASAVGLTAYVQRHPVSREWLEAGALVLADRGVCLIDEFDKMNDQDRTSI
 HEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDIL
 CVVRDVTDPVQDEMLARFVVGSHVRHHPNKEEEGLANGSAAEPAMPNTYGEVPLPQEV
 KKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMAEAHARI
 HLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFFARYLSFRDNNELLLFILKQLV
 30 AEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNL SAFYDSELFMRNKF SHDLKRKMI
 LQQF

SEQ ID NO:25 – protein sequence of Histone H3.3

>sp|P84243|H33_HUMAN Histone H3.3 OS=Homo sapiens

35 MARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTE
 LLIRKLFPQRLVREIAQDFKTDLRFQSAAGALQEASEAYLVGLFEDTNLCAIHAKRVTI
 MPKDIQLARRIRGERA

SEQ ID NO:26

40 RHXX

SEQ ID NO:27

RHXKVL

SEQ ID NO:28

5 RHXKVLR

SEQ ID NO:29

Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg

10 SEQ ID NO:30

Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg

SEQ ID NO:31

Lys-Gly-Gly-Ala-Lys-Arg-His-Ala-Lys-Val-Leu-Arg

15 SEQ ID NO:32

Lys-Gly-Gly-Ala-Ala-Arg-His-Arg-Lys-Val-Leu-Arg

SEQ ID NO:33

Leu-Gly-Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Val-Leu-Arg-Asp-Asn-Ile

20

SEQ ID NO:34 --protein sequence of Histone H4

SGRGKGGKGLGKGGAKRHRKVL RDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK

VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGR TLYGFEGG

25

ITEMS OF THE INVENTION

Some aspects of the invention may further be identified by the following items:

- 5 1. A small molecule, which targets the conformational space of the TONSL ARD occupied by the histone H4 tail encompassing residues K12-R23 and acting to prevent or disrupt the binding of the H4 tail K12-R23 with the TONSL ARD via direct competition or via allosteric disruption of the binding pocket.
- 10 2. A small molecule according to item 1, that targets the H4 tail spanning residues Lys12 to Arg23 through intermolecular hydrogen-bonding, electrostatic and/or van der Waals interactions.
3. A small molecule according to any of items 1-2, wherein the molecule targets the
15 intermolecular contacts spanning the Lys12-Gly13-Gly14-Ala15 segment of H4.
4. A small molecule according to any of items 1-3, wherein the molecule targets the hydrophobic interactions between residues Gly13, Gly14 and Ala15 of H4 and residues Asn507, Cys508, Trp641, Tyr645 and Leu649 of ARD.
- 20 5. A small molecule according to any of items 1-4, wherein the molecule targets the hydrogen bonds between the main-chain O of H4 Gly14 and N ϵ 1 of ARD Trp641, and between the main-chain N of H4 Ala15 and O η of ARD Tyr645.
- 25 6. A small molecule according to any of items 1-5, wherein the molecule targets the main-chain O of H4 Lys16 hydrogen bonds with the N δ 2 of ARD Asn571.
7. A small molecule according to any of items 1-6, wherein the molecule targets the side-chain of H4 Arg17, which stacks over the side-chains of ARD Tyr572 and Cys608,
30 while its N η 1 atom forms two hydrogen bonds with main-chain O and O δ 1 of ARD Asn571.
8. A small molecule according to any of items 1-7, wherein the molecule targets the side-chain of H4 H18, which penetrates into a pocket lined by four strictly conserved
35 residues (Trp563, Glu568, Asn571 and Asp604) and is positioned over His567 of ARD.

9. A small molecule according to any of items 1-8, wherein the molecule targets the side chain of H4 His18, which is stacked between Trp563 and Asn571 and forms hydrogen bonds to Glu568 and Asp604 of ARD.
- 5 10. A small molecule according to any of items 1-9, wherein the molecule targets the main-chain O of H4 Arg19, which forms a hydrogen bond with N ϵ 1 of Trp563 and its side-chain forms contacts with Cys561 and Gly595 of ARD.
- 10 11. A small molecule according to any of items 1-10, wherein the molecule targets the H4 Lys20 residue that is bound within an acidic surface pocket on ARD adjacent to the H4 His18 binding pocket.
- 15 12. A small molecule according to any of items 1-11, wherein the molecule targets side-chain of H4 Lys20 which interacts with the side-chain of Met528 and contacts the edge of Trp563 of ARD, while the main-chain atoms of H4 Lys20 packs against Cys561 of ARD.
- 20 13. A small molecule according to any of items 1-12, wherein the molecule targets the N ζ atom of H4 Lys20 that forms three strong hydrogen bonds (distance < 3 Å) with the side-chains of strictly conserved residues Glu530, Asp559 and Glu568 of ARD, which surround H4 Lys20 within a regular triangle-like alignment.
- 25 14. A small molecule according to any of items 1-13, wherein the molecule targets intermolecular contacts spanning the Val21-Leu22-Arg23 segment of H4, which includes contacts between side-chains of H4 Val21 with Tyr560 and Cys561 of ARD, while H4 Leu22 interacts with Asp527 and Met528 of ARD.
- 30 15. A small molecule according to any of items 1-14, wherein the molecule targets the main-chain N of H4 Arg23 which forms a hydrogen bond with the main-chain O of Asp527 of ARD, while the side-chain packs against the side-chain of Tyr560 of ARD.
- 35 16. A small molecule according to any of items 1-15, capable of blocking histone reader domains in a protein selected from the group consisting of TONSL, BARD1 and ANKRD11.
17. A small molecule according to any of items 1-16 for use as a medicament.

18. A small molecule according to any of items 1-17 for use in treatment of cancer.
19. A method of selecting or designing a small molecule capable of interfering with the histone H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, said method comprises use of at least part of the atomic co-ordinates data contained in PDB ID 5JA4 or data derivable therefrom, wherein said method involves use of a computer modelling package or a computer program to model all or part of the structure of MCM2 HBD-G4-TONSL ARD in complex with H3 (57-135) and H4, thereby identifying said molecule by designing or selecting the molecule based on its likely ability to interact with a modelled structure.
20. An isolated polynucleotide or amino acid sequence having at least 90% sequence identity to any of SEQ ID NO 1-22.
21. A crystal comprising covalently linked MCM2 HBD-G₄-TONSL ARD in complex with H3 (57-135) and H4 (that diffracted to 2.43 Å resolution).
22. A crystal structure having the atomic coordinates or a subset hereof set out in PDB ID 5JA4 or having a structure in which the atomic coordinates vary by less than 3Å in any direction from those set out therein.

CLAIMS

1. An inhibitor of TONSL, wherein said inhibitor inhibits binding of TONSL ARD to
5 histone H4.
2. The inhibitor of TONSL according to claim 2, wherein the inhibitor binds the
conformational space of the TONSL ARD occupied by the histone H4 tail comprising
residues K12-R23, wherein said inhibitor prevents and/or disrupts the binding of the
10 histone H4 with the TONSL ARD via direct competition or via allosteric disruption of
the binding pocket.
3. The inhibitor according to any one of the preceding claims, wherein the inhibitor
binds the histone H4 tail binding surface of TONSL ARD.
15
4. A small molecule, which targets the conformational space of the TONSL ARD
occupied by the histone H4 tail encompassing residues K12-R23 and acting to prevent
or disrupt the binding of the histone H4 tail K12-R23 with the TONSL ARD via direct
competition or via allosteric disruption of the binding pocket.
20
5. A small molecule according to claim 4 or the inhibitor according to any one of
claims 1 to 3, that targets the TONSL ARD interaction with the histone H4 tail
spanning residues Lys12 to Arg23 through intermolecular hydrogen-bonding,
electrostatic and/or van der Waals interactions.
25
6. A small molecule according to any one of claims 4 to 5 or the inhibitor according to
any one of claims 1 to 3, wherein the molecule targets the intermolecular contacts
between TONSL ARD and the Lys12-Gly13-Gly14-Ala15 segment of histone H4.
7. A small molecule according to any of claims 4 to 6 or the inhibitor according to any
one of claims 1 to 3, wherein the molecule targets the hydrophobic interactions
30 between residues Gly13, Gly14 and Ala15 of histone H4 and residues Asn507,
Cys508, Trp641, Tyr645 and Leu649 of TONSL ARD.
8. A small molecule according to any of claims 4 to 7 or the inhibitor according to any
one of claims 1 to 3, wherein the molecule targets the hydrogen bonds between the
35

main-chain O of histone H4 Gly14 and Nε1 of TONSL ARD Trp641, and between the main-chain N of H4 Ala15 and Oη of TONSL ARD Tyr645.

9. A small molecule according to any of claims 4 to 8 or the inhibitor according to any one of claims 1 to 3, wherein the molecule targets the main-chain O of histone H4 Lys16 hydrogen bonds with the Nδ2 of TONSL ARD Asn571.
10. A small molecule according to any of claims 4 to 9 or the inhibitor according to any one of claims 1 to 3, capable of blocking histone reader domains in a protein selected from the group consisting of TONSL, BARD1 and ANKRD11.
11. The inhibitor according to any one of claims 1 to 3 and 5 to 10, wherein the inhibitor is a peptide or a polypeptide optionally linked to a conjugated moiety.
12. The inhibitor according to claim 11, wherein the inhibitor is a peptide comprising the sequence Arg-His-Xaa-Lys, wherein Xaa may be any amino acid.
13. The inhibitor according to claim 11, wherein the inhibitor is a peptide comprising the sequence Arg-His-Xaa-Lys-Val-Leu.
14. The inhibitor according to any one of claims 11 to 13, wherein the inhibitor is a peptide comprising the sequence Val-Leu-Arg.
15. The inhibitor according to any one of claims 11 to 14, wherein the inhibitor is a peptide comprising the sequence Arg-His-Xaa-Lys-Val-Leu-Arg.
16. The inhibitor according to any one of claims 11 to 15, wherein the peptide or polypeptide comprises or consists of:
- I. a sequence consisting of amino acid 12 to 23 of SEQ ID NO:23 or of SEQ ID NO:34; or
 - II. a functional homologue thereof consisting of a sequence of amino acid 12 to 23 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:34,
 - III. a functional homologue thereof consisting of a sequence of amino acid 12 to 23 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 6 amino acids may be substituted,

with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

17. The inhibitor according to any one of claims 11 to 13, wherein the peptide or polypeptide comprises or consists of:

- a sequence consisting of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34; or
- a functional homologue thereof consisting of a sequence of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:34,
- a functional homologue thereof consisting of a sequence of amino acid 9 to 25 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 6 amino acids may be substituted

with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

18. The inhibitor according to any one of claims 11 to 14, wherein the peptide or polypeptide comprises or consists of:

- a sequence consisting of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34; or
- a functional homologue thereof consisting of a sequence of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 5 amino acids may be substituted, and wherein said peptide comprises at least Arg17, His18 and Lys20 of SEQ ID NO:34,
- a functional homologue thereof consisting of a sequence of amino acid 14 to 33 of SEQ ID NO:23 or of SEQ ID NO:34, wherein up to 6 amino acids may be substituted

with the proviso that the inhibitor is different to histone H4 of SEQ ID NO:23.

19. The inhibitor according to any one of claims 16 to 18, wherein at the most 4, preferably at the most 3, such as at the most 2, for example at the most one amino acid may be substituted in said functional homologue.

20. The inhibitor according to claim 11, wherein the inhibitor comprises or consists of a peptide selected from the group consisting of SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32 and SEQ ID NO:33.
- 5 21. The inhibitor according to any one of claims 11 to 20, wherein said peptide consists of at the most 40 amino acids, such as at the most 25 amino acids, for example at the most 20 amino acids.
- 10 22. The inhibitor according to any one of claims 11 to 16 and 21, wherein said peptide consists of at the most 15 amino acids, for example at the most 12 amino acids.
23. The inhibitor according to any one of claims 11 to 22, wherein said peptide comprises or consists of in the range of 4 to 40 consecutive amino acids of SEQ ID NO:23.
- 15 24. The inhibitor according to any one of claims 11 to 23, wherein the amino acid corresponding to Lys20 of SEQ ID NO:23 is unmethylated.
25. The inhibitor according to claim 11, wherein said peptide comprises or consists of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21 or SEQ ID NO:22.
- 20 26. The inhibitor according to claim 11, wherein said peptide comprises or consists of amino acids 512 to 692 of SEQ ID NO:16.
- 25 27. The inhibitor according to any one of claims 11 to 26, wherein the C-terminal of the peptide is amidated or alkylated.
28. The inhibitor according to any one of claims 11 to 27, wherein the N-terminal of the peptide is acetylated or formylated.
- 30 29. The inhibitor according to any one of claims 11 to 28, wherein said conjugated moiety is a polymeric molecule, such as polyethylene glycol (PEG) and polyvinylpyrrolidone (PVP).
- 35 30. The inhibitor according to any one of claims 1 to 3 or the small molecule according to any one of claims 4 to 10, wherein the inhibitor or the small molecule is a compound having a 3-[(3-Aminocyclopentyl)carbonyl]-1H-quinolin-4-one core.

31. The inhibitor according to any one of the preceding claims, wherein the TONSL ARD is identified as amino acids 512-692 of TONSL of SEQ ID NO:16.

5 32. The inhibitor according to any one of the preceding claims, wherein histone H4 is histone H4 of SEQ ID NO:23 or histone H4 of SEQ ID NO:34.

33. The inhibitor according to any one of claims 2 to 32, wherein the K12-R23 of the histone H4 tail corresponds to amino acids 12 to 23 of SEQ ID NO:34.

10

34. A small molecule according to any of claims 4 to 10 or the inhibitor according to any one of claims 1 to 3 and 5 to 33 for use as a medicament.

15 35. A small molecule according to any of claims 4 to 10 or the inhibitor according to any one of claims 1 to 3 and 5 to 34 for use in treatment of cancer.

36. Use of a small molecule according to any of claims 4 to 10 or the inhibitor according to any one of claims 1 to 3 and 5 to 33 for use in the preparation of medicament for treatment of cancer.

20

37. A three dimensional crystal of a complex between

- a polypeptide comprising:
 - TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue thereof sharing at least 90% sequence identity therewith; and
 - optionally MCM2 HBD consisting of amino acids 61 to 130 of SEQ ID NO:24 or a functional homologue thereof sharing at least 90% sequence identity therewith;
- Histone H4 of SEQ ID NO:23 or a functional homologue thereof sharing at least 90% sequence identity therewith; and
- Optionally Histone H3.3 Δ 56 consisting of amino acids 57–135 of SEQ ID NO:25 or a functional homologue thereof sharing at least 90% sequence identity therewith.

25

30

38. The crystal according to claim 37, wherein the polypeptide comprises or consists of TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 linked to MCM2 HBD consisting of amino acids 61 to 130 of SEQ ID NO:24 via a linker of in the range of 4 to 20 Gly residues, such as in the range of 4 to 12 Gly-residues.

5

39. The crystal according to any one of claims 37 to 38, wherein the crystal is of space group P3 2 1.

10 40. The crystal according to any one of claims 37 to 39, wherein the unit cell parameters are $a=b=139.5 \pm 3\text{\AA}$, $c=72.9 \pm 3\text{\AA}$, $\alpha=\beta=90^\circ \pm 2^\circ$ and $\gamma=120^\circ \pm 2^\circ$.

41. The crystal according to any one of claims 27 to 40, wherein the unit cell parameters are $a=b=139.5$, $c=72.9$, $\alpha=\beta=90^\circ$ and $\gamma=120^\circ$.

15 42. A method of selecting or designing a small molecule capable of interfering with the histone H4R17, H4H18 and H4K20 binding pocket on the surface of the Ankyrin repeats of TONSL, said method comprises use of at least part of the atomic co-ordinates data contained in PDB ID 5JA4, or data derivable therefrom, wherein said method involves use of a computer modelling package or a computer program to
20 model all or part of the structure of MCM2 HBD-G4-TONSL ARD in complex with H3 (57-135) and H4, identifying such small molecule based on its likely ability to prevent or disrupt the H4 tail with the Ankyrin repeats of TONSL in the modelled structure.

25 43. A method of identifying an inhibitor of TONSL inhibiting binding of TONSL ARD to histone H4, wherein the method comprises the steps of

- 30
- a) obtaining a crystal according to any one of claims 37 to 41
 - b) using the crystal obtained in step a) to obtain an x-ray diffraction pattern;
 - c) solving the three dimensional structure of the crystal from the diffraction pattern obtained in step b), thereby obtaining the three dimensional structure of the complex
 - d) identifying one or more compounds that binds to the H4 tail binding surface of TONSL ARD based on the obtained three dimensional structure.

44. A computer-based method for identifying an inhibitor of TONSL inhibiting binding of TONSL ARD to histone H4, said method comprising the steps of

- 5 a) generating a spatial structure of the histone H4 tail-binding surface of TONSL ARD in a computer based on the atomic coordinates provided in PDB ID 5JA4, or atomic coordinates of a three-dimensional structure that deviates from the three-dimensional structure presented in PDB ID 5JA4 by a root mean square deviation over the protein backbone atoms of not more than 3Å, preferably not more than 1Å
- 10 b) generating a spatial structure of potential inhibitors in a computer
- c) selecting potential inhibitors having a structure which can bind at least 1 amino acid residue of the histone H4 tail binding surface of TONSL ARD.

15 45. The method according to claim 44, wherein said potential inhibitor has a structure, which can bind at least 2, preferably at least 3, such as at least 5, for example at least 10 of the amino acids of the histone H4 tail binding surface of TONSL ARD.

20 46. The method according to claim 44 to 45, wherein the histone H4 tail binding surface of TONSL ARD comprise at least amino acids Asp527, Met528, Glu530, Asp559, Tyr560, Cys561, Trp563, Glu568, Asn571, Tyr572, Gly595, Glu597, Asp604, Asn607, Cys608, Trp641, Tyr645 and Leu649.

25 47. An isolated polynucleotide or amino acid sequence having at least 90% sequence identity to any of SEQ ID NO 1-22.

30 48. A TONSL mutant polypeptide comprising at least one mutation in an amino acid of the histone H4 tail binding surface of the TONSL ARD, wherein said TONSL mutant polypeptide apart from said mutation is identical to SEQ ID NO:16 or shares a sequence identity with SEQ ID NO:16 of at least 70%, wherein said TONSL mutant polypeptide is capable of binding MMS22L and wherein said TONSL mutant does not bind histone H4.

35 49. The polypeptide according to claim 48, wherein said TONSL mutant polypeptide carries a mutation in one or more of the amino acids selected from the group consisting of: Glu530, Asp559, Trp563, Glu568, Gln571 and D604.

50. The polypeptide according to any one of claims 48 to 49, wherein the polypeptide comprises or consists of a sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21 and SEQ ID NO:22.

5 51. A method for predicting the effect of inhibition of TONSL, said method comprising the steps of

- expressing the polypeptide according to any one of claims 11 to 33 and 48 to 50 in an organism and/or cells of an organism, wherein said polypeptide optionally is expressed conditionally,

10 • determining the effect of said polypeptide in said organism and/or cells

wherein said effect of expressing said polypeptide in said organism and/or cells indicates the effect of inhibition of TONSL in said organism and/or cells.

52. The method according to claim 51, wherein the effect is indicative of the effect of inhibition of TONSL with an inhibitor of TONSL, wherein said inhibitor inhibits binding
15 of TONSL ARD to histone H4, for example an inhibitor according to any one of claims 1 to 21.

53. The method according to any one of claims 51 to 52, wherein said organism or said cells represent a disease model.

20

54. The method according to any one of claims 51 to 53, wherein the organism is a mouse, for example a mouse genetically engineered to contract cancer.

55. The method according to any one of claims 51 to 53, wherein the cells are
25 cultured in a 3D culture.

56. A method for identifying an inhibitor of TONSL and/or determining inhibitory effect of a putative inhibitor of TONSL, said method comprising the steps of

a) providing a test compound, which is a putative inhibitor of TONSL

30 b) providing an host organism expressing TONSL of SEQ ID NO:16, TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue of any of the aforementioned sharing at least 90% sequence identity to amino acids 512 to 692 of SEQ ID NO:16;

c) contacting said host organism with said putative inhibitor

d) detecting chromatin associated TONSL, TONSL ARD or functional homologue thereof in said host organism,

5 wherein reduction of chromatin associated TONSL, TONSL ARD or a functional homologue thereof is indicative of said test compound being an inhibitor of TONSL.

57. The method according to claim 56, wherein the test compound is the TONSL mutant polypeptide according to any one of claims 48 to 50.

10 58. The method according to claim 56, wherein the test compound is a peptide according to any one of claims 11 to 33.

59. The method according to any one of claims 56 to 58, wherein said host organism comprises a heterologous nucleic acid encoding said TONSL, TONSL ARD or functional homologue thereof.

15 60. The method according to any one of claims 56 to 59, wherein the TONSL, TONSL ARD or functional homologue thereof is linked to a detectable linker, for example to a fluorescent polypeptide.

61. The method according to any one of claims 56 to 60, wherein step d) may comprise the steps of:

20 d.1) extracting soluble proteins from the cells of said host organism

d.2) detecting TONSL, TONSL ARD or functional homologue thereof in cells of said host organism, thereby detecting chromatin associated TONSL, TONSL ARD or functional homologue thereof.

25 62. A method for determining inhibitory effect of a putative inhibitor of TONSL, said method comprising the steps of

e) Providing histone H4 or a fragment thereof comprising at least amino acids 17 to 20 of SEQ ID NO:23,

- f) providing TONSL of SEQ ID NO:16, TONSL ARD consisting of amino acids 512 to 692 of SEQ ID NO:16 or a functional homologue of any of the aforementioned sharing at least 90% sequence identity to amino acids 512 to 692 of SEQ ID NO:16
- 5 g) Incubating said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof in the presence of a putative inhibitor of TONSL
- h) Determining whether binding between said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof
- 10 wherein reduced binding of said histone H4 or fragment thereof with said TONSL, TONSL ARD or functional homologue thereof is indicative of that said putative inhibitor of TONSL is capable of inhibiting TONSL.
- 15 63. The method according to claim 62, wherein the inhibitor is an inhibitor or a small molecule according to any one of claims 1 to 33.

Figure 1

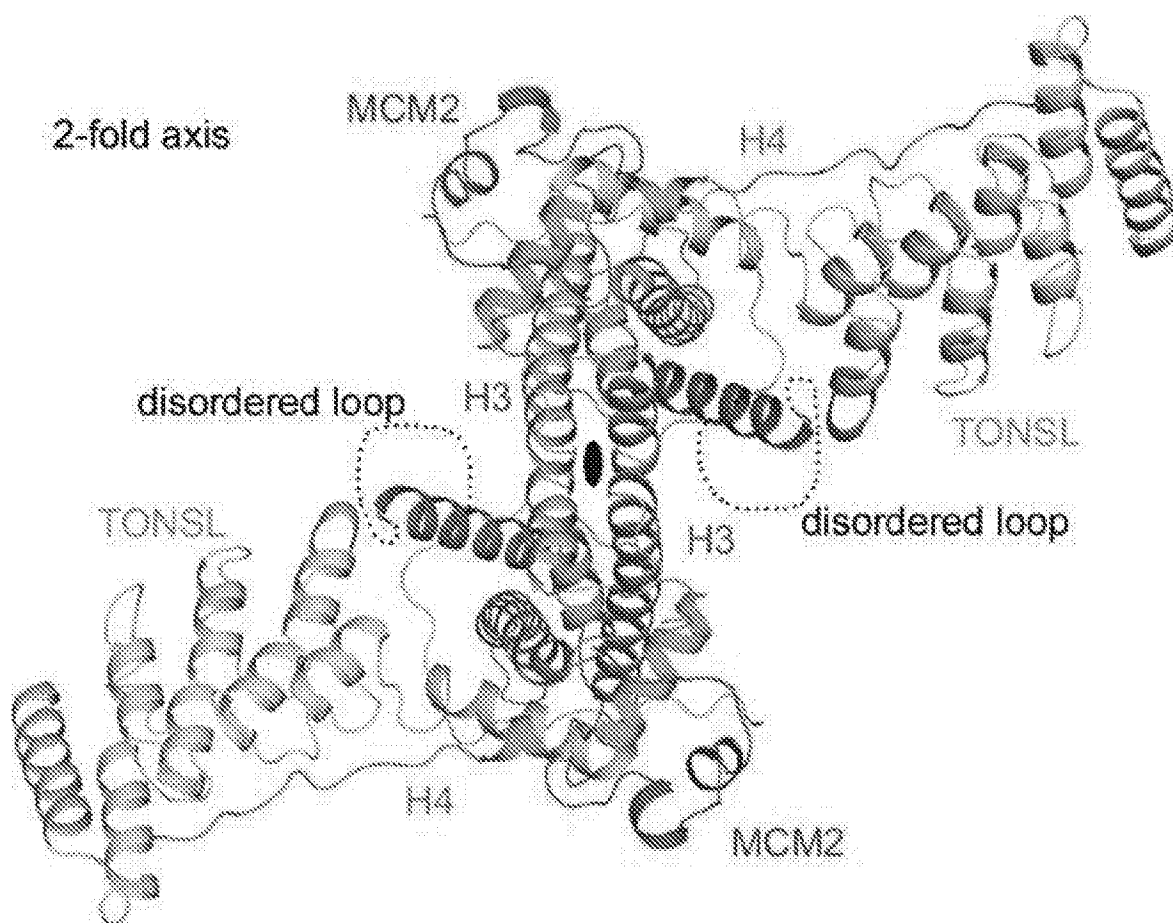


Figure 2

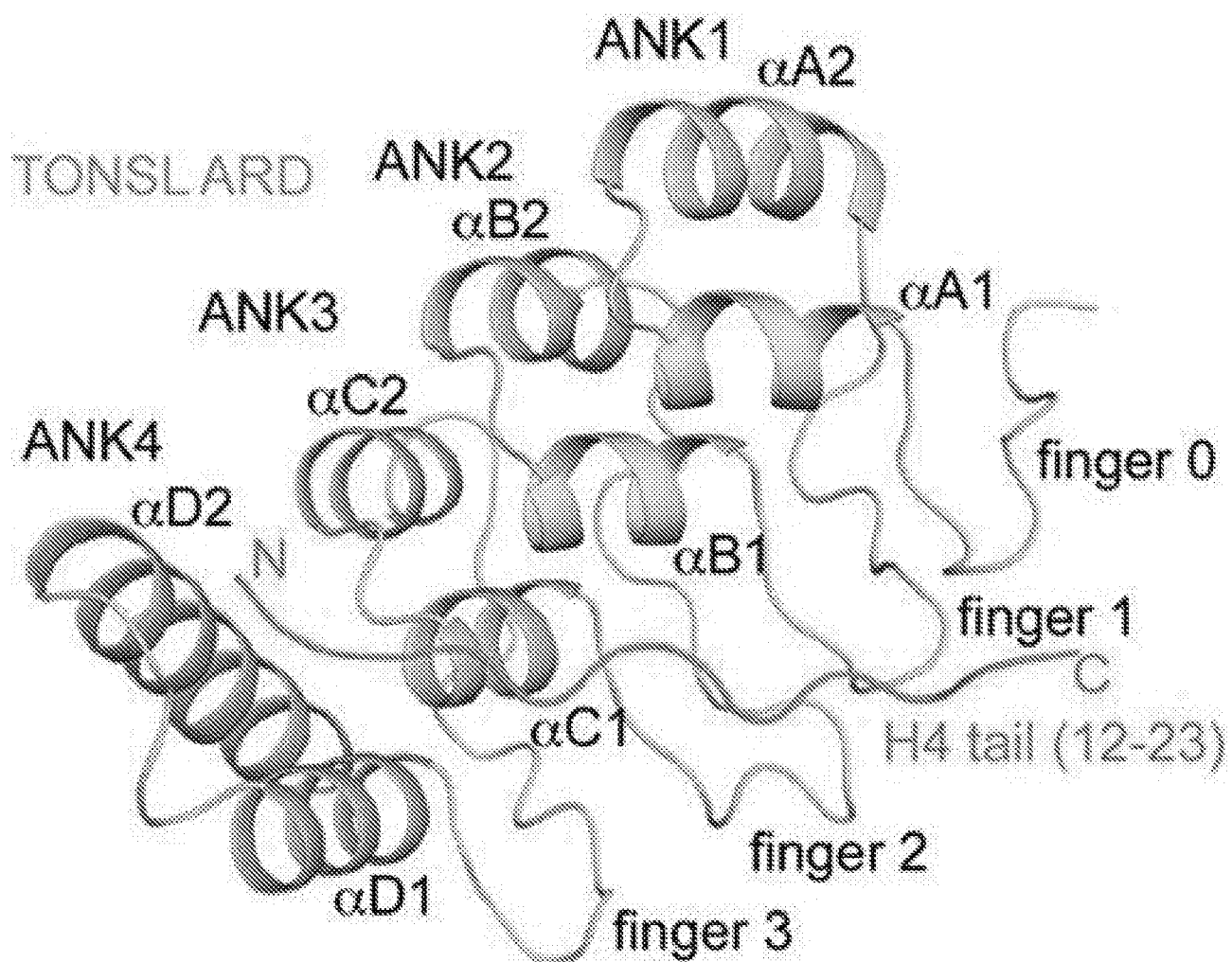


Figure 3

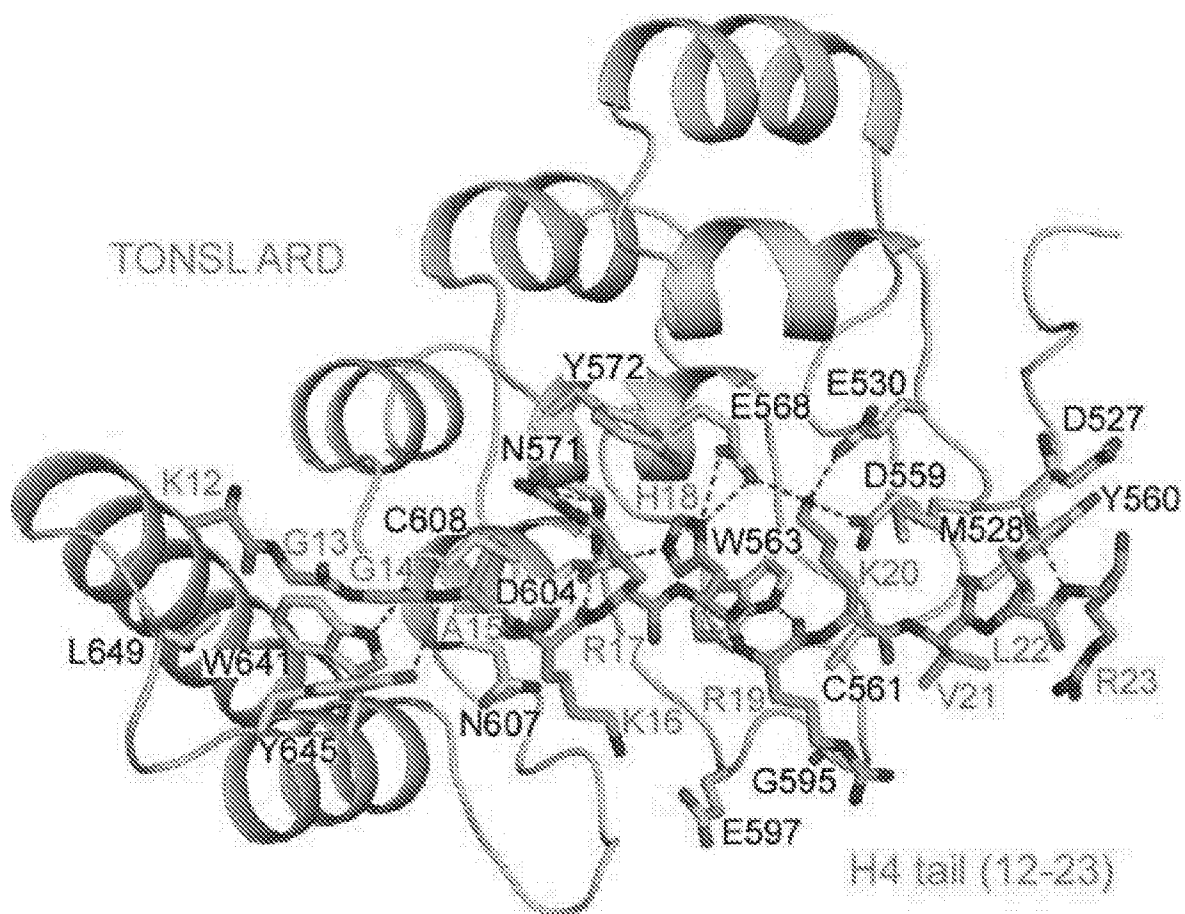


Figure 4

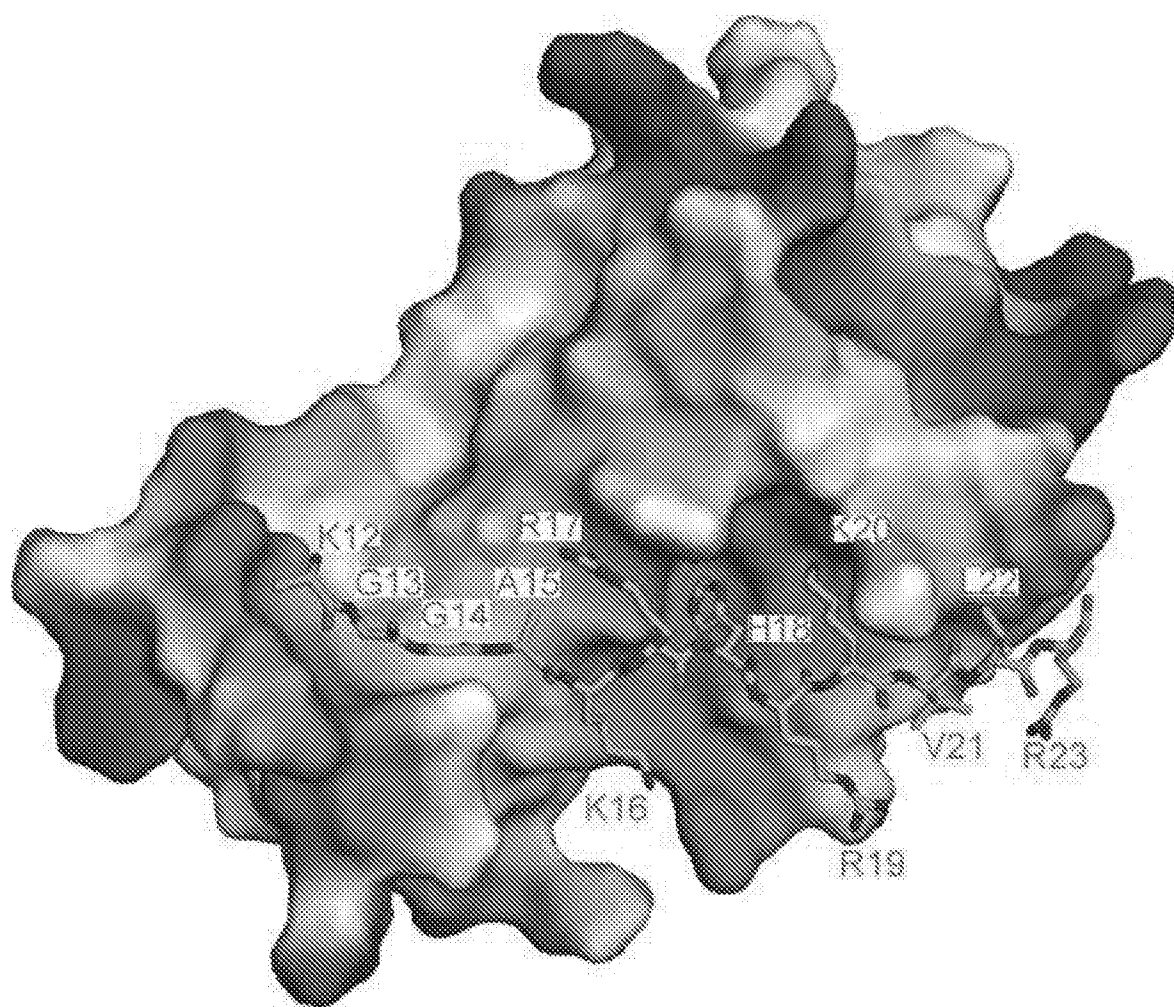


Figure 5

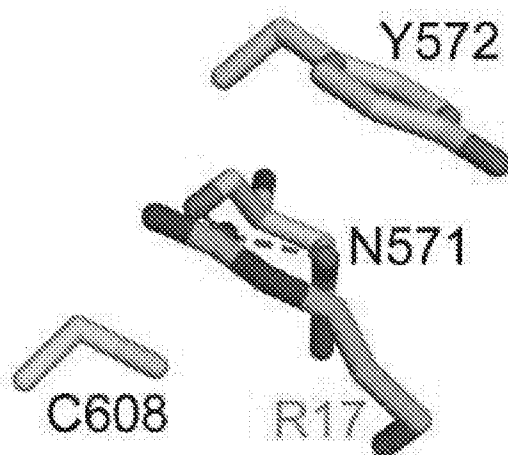


Figure 6

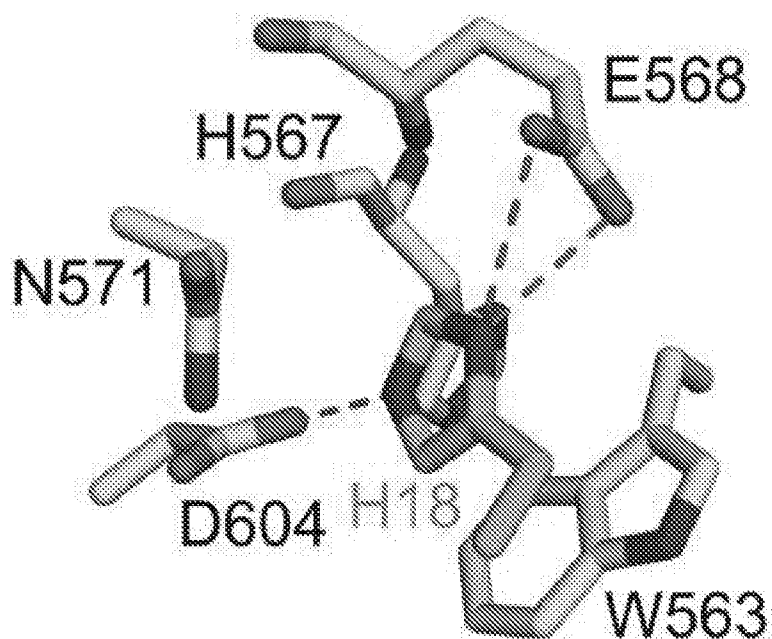


Figure 7

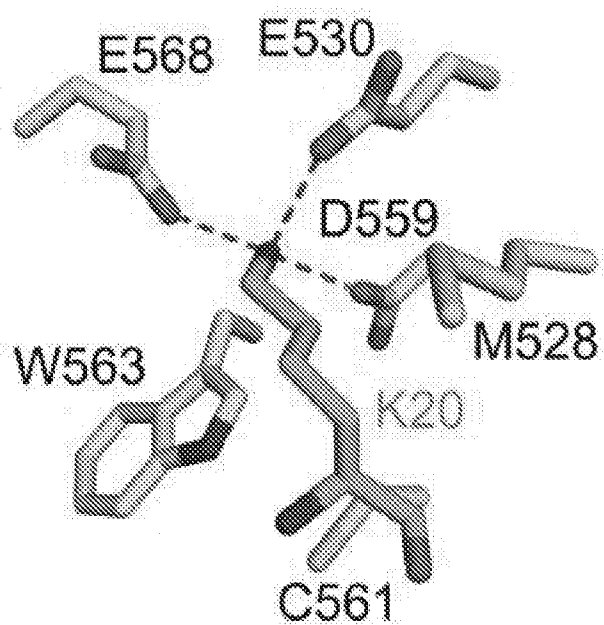


Figure 8

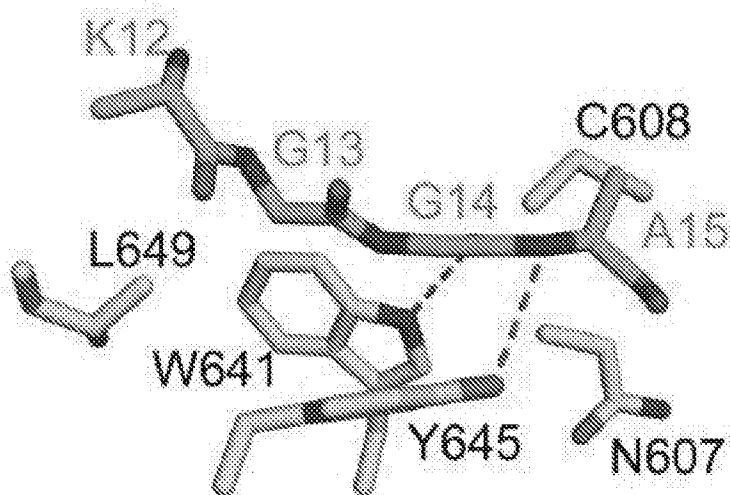


Figure 9

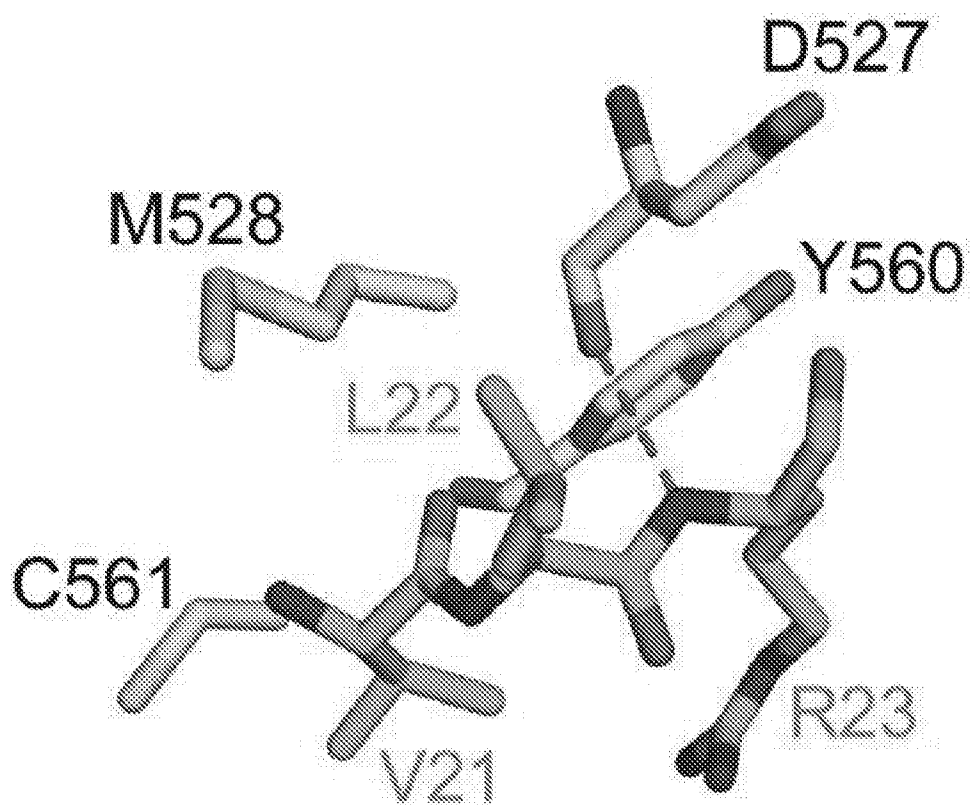


Figure 10

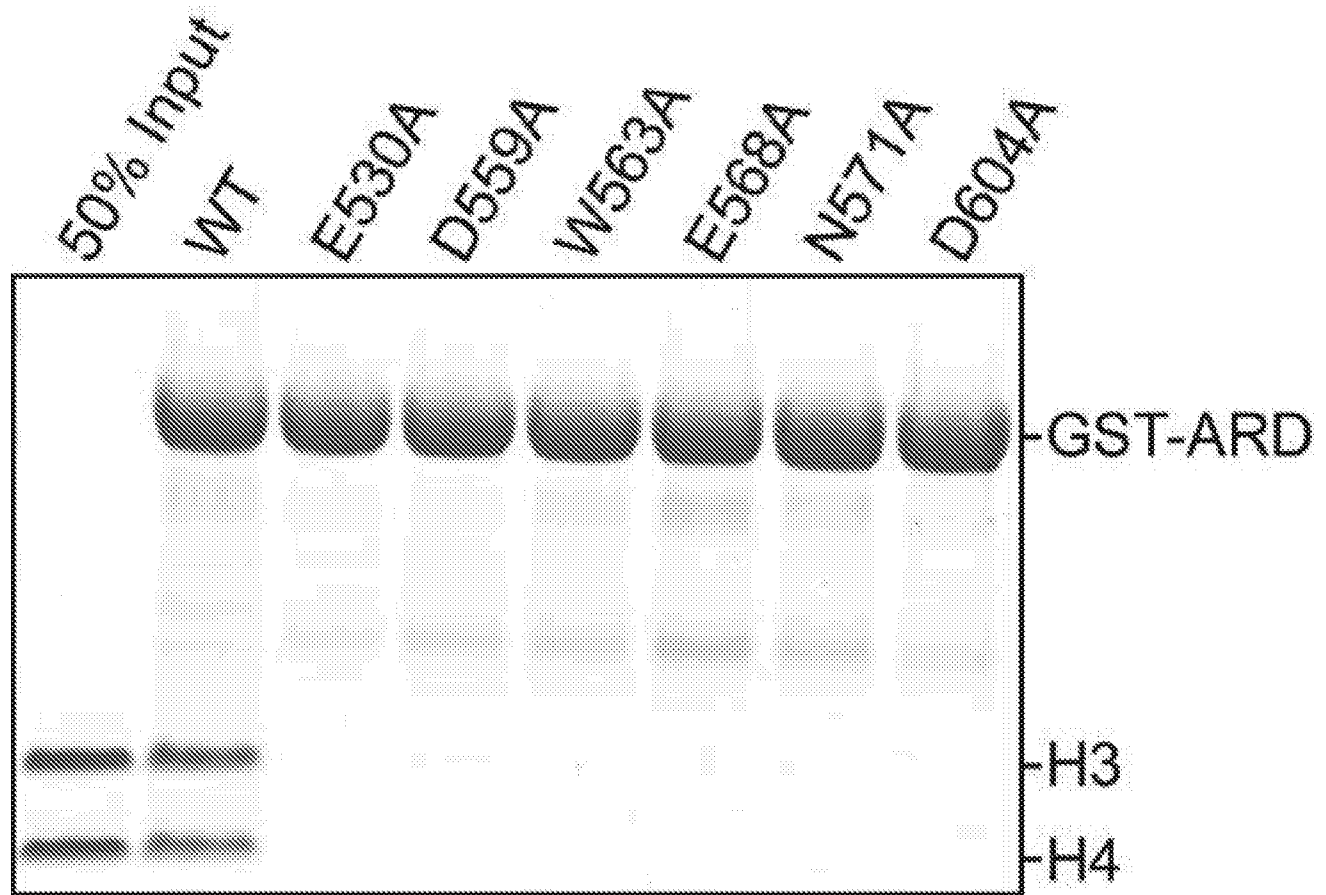


Figure 11

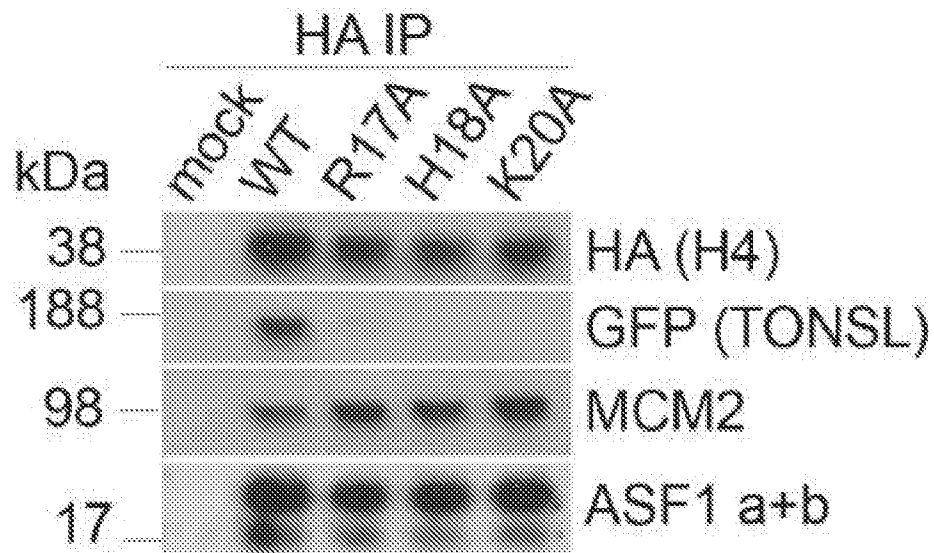
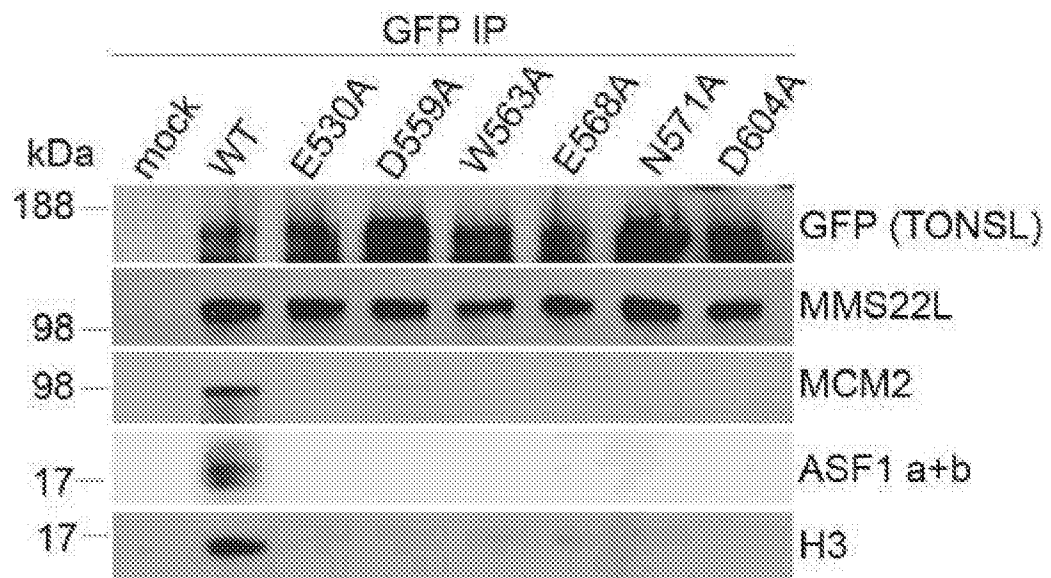


Figure 12



10/26

Figure 13

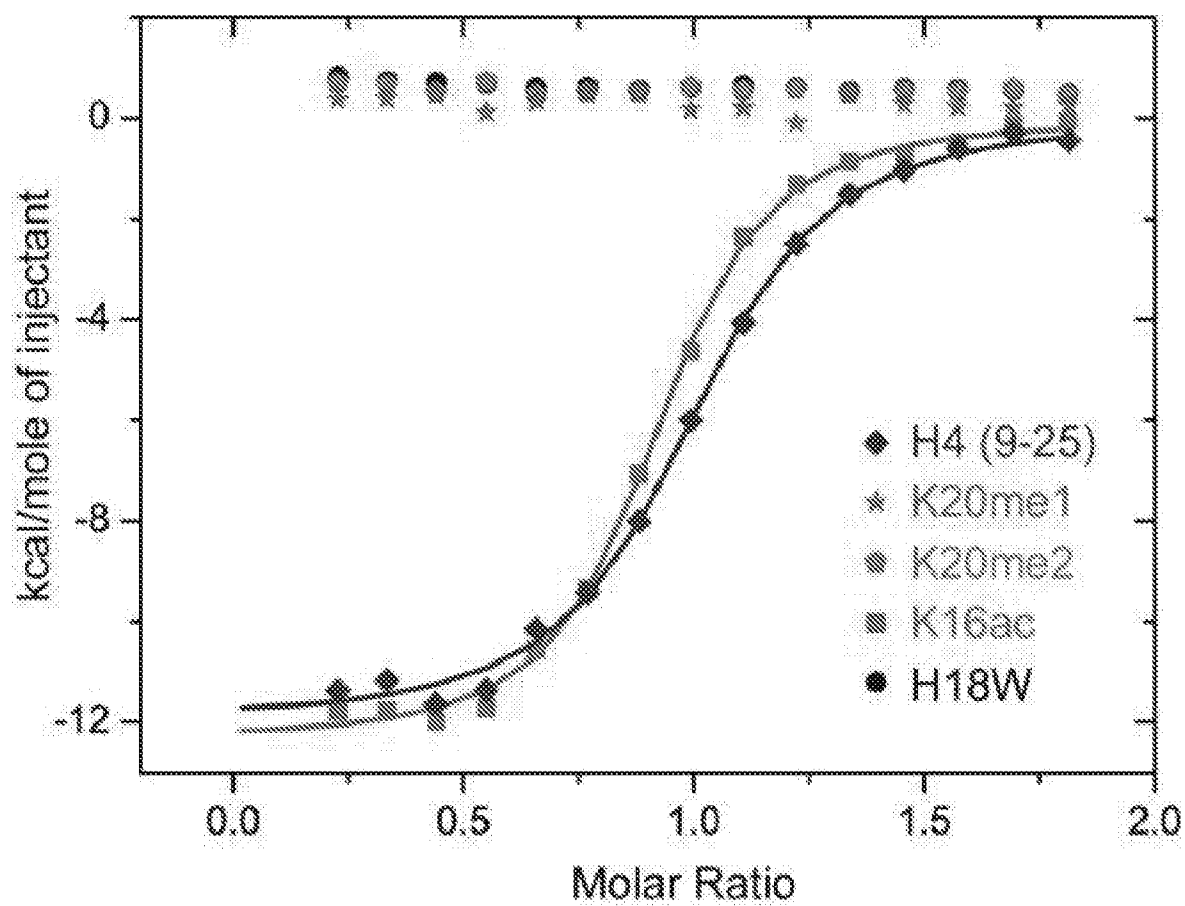


Figure 14

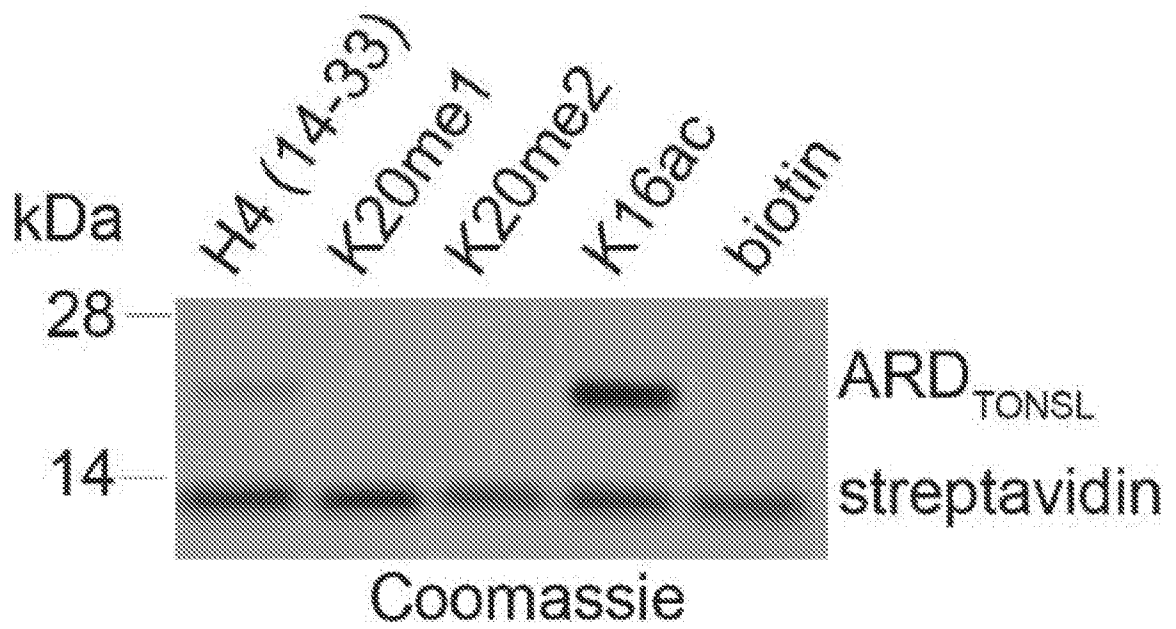


Figure 15

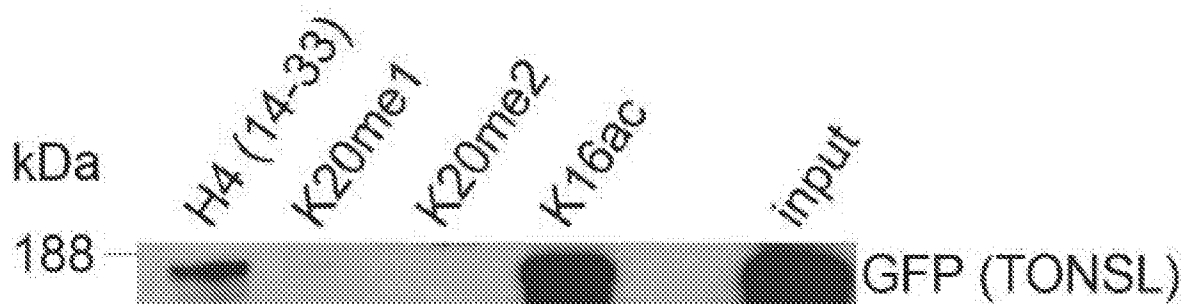


Figure 16

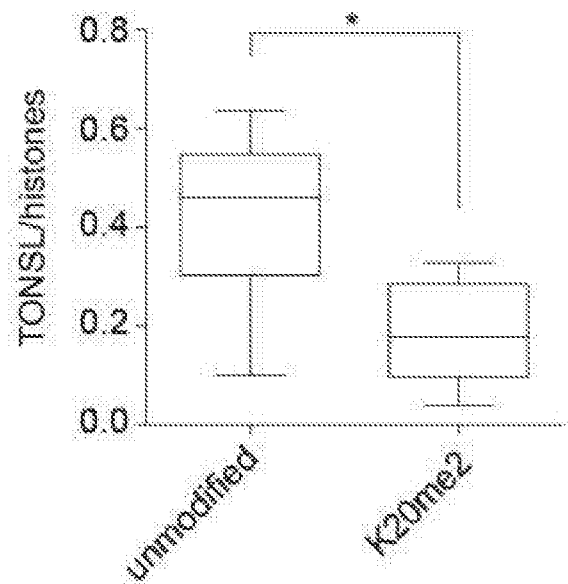
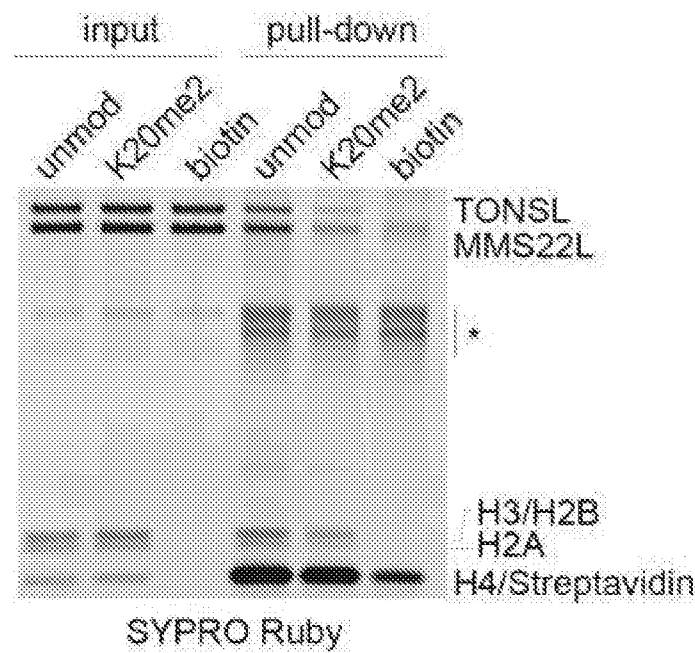


Figure 17

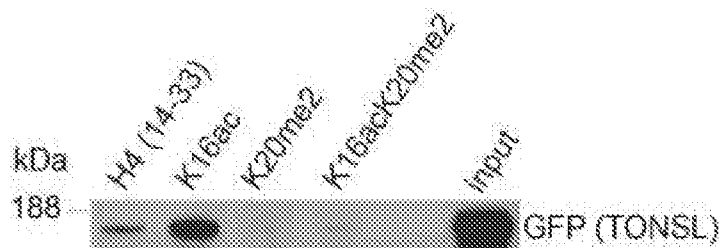


Figure 18

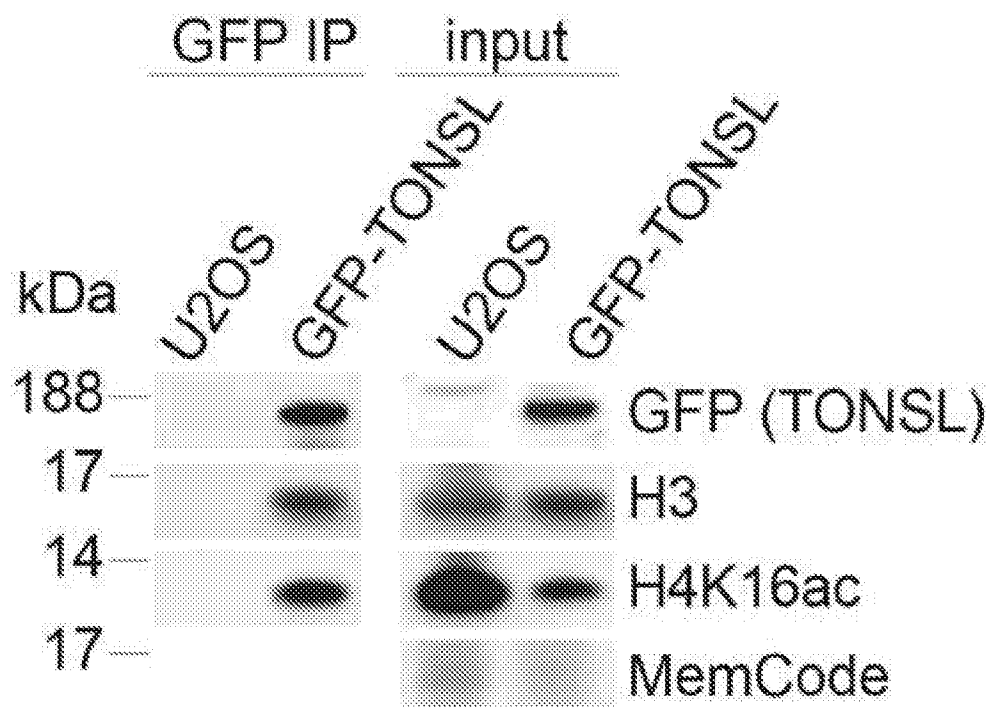


Figure 19

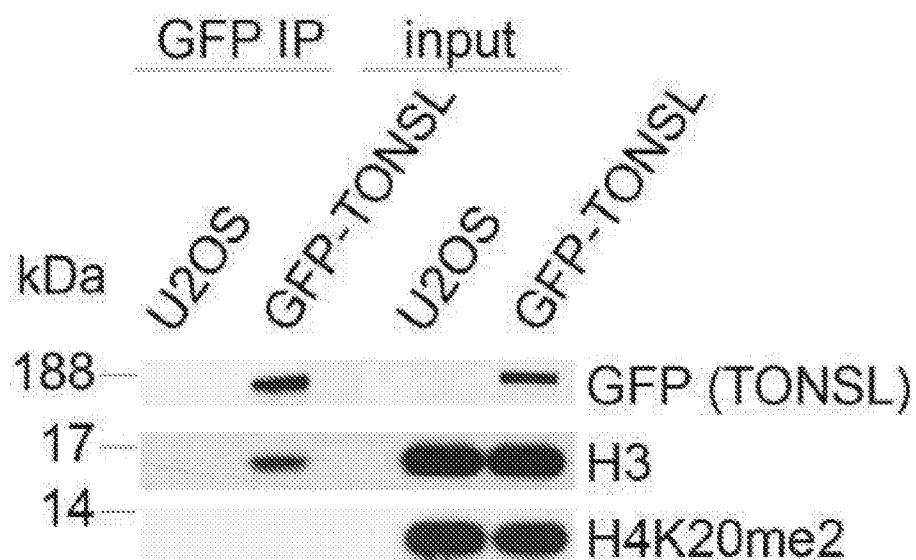


Figure 20

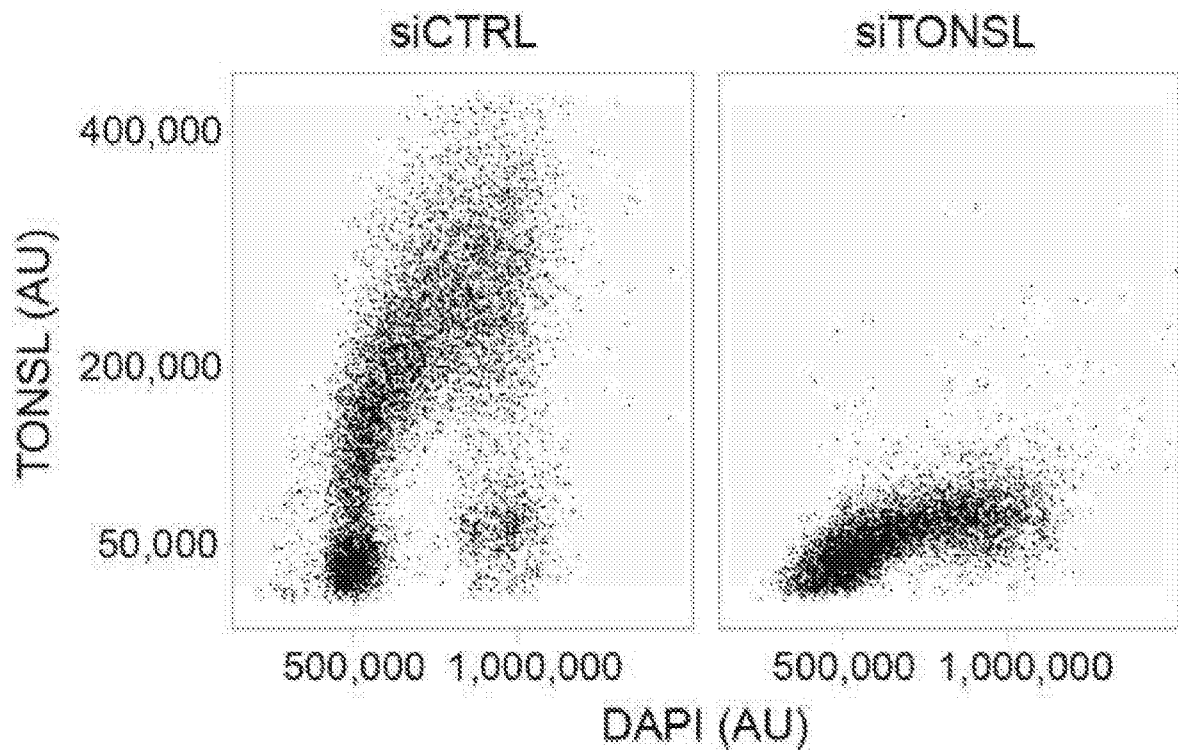


Figure 21

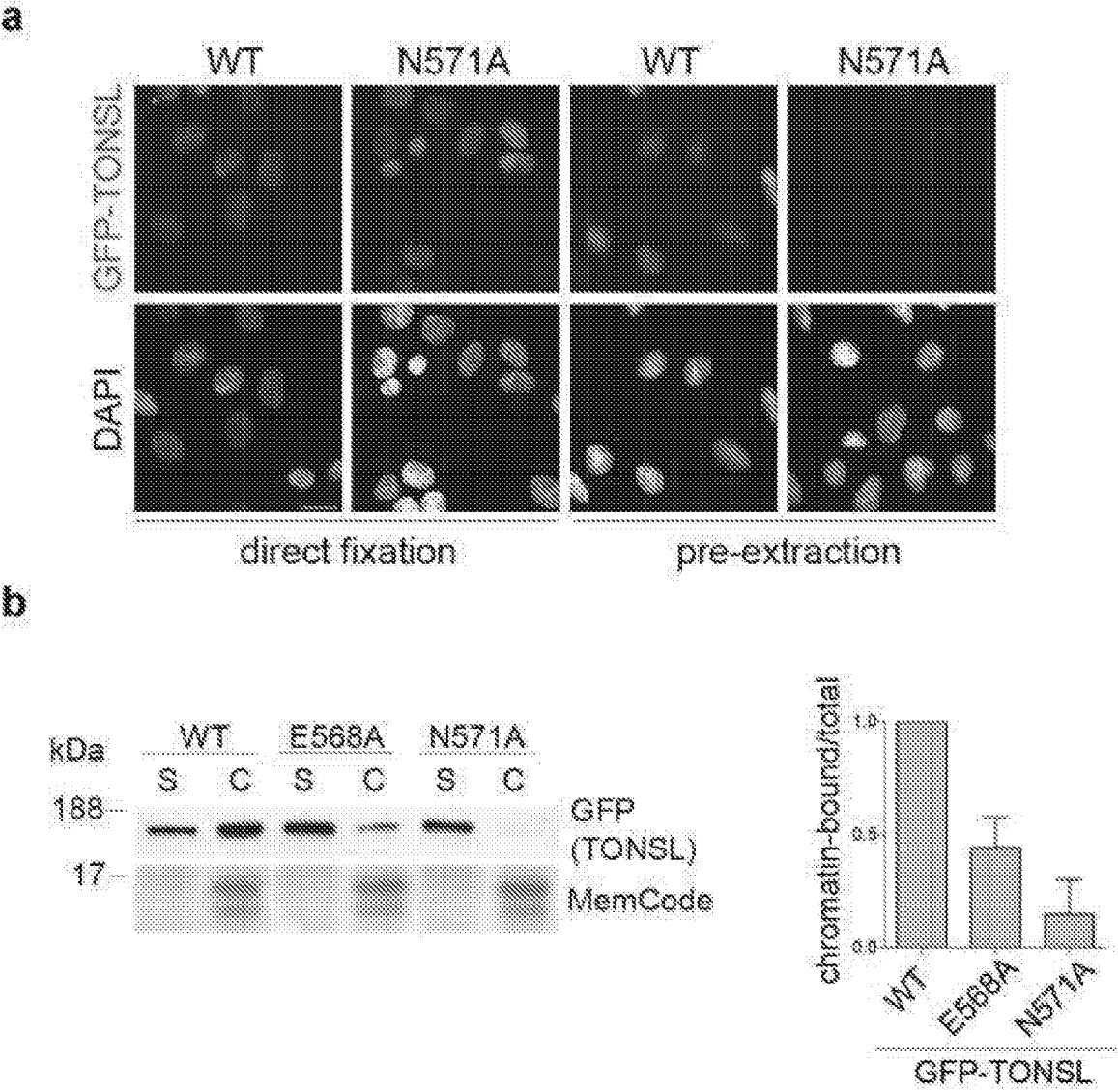


Figure 22

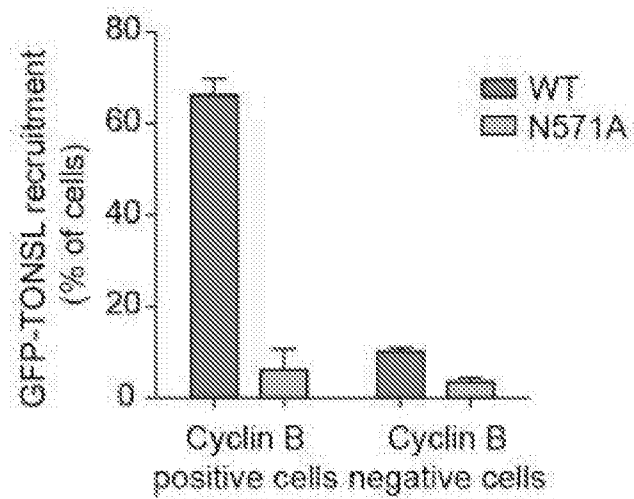
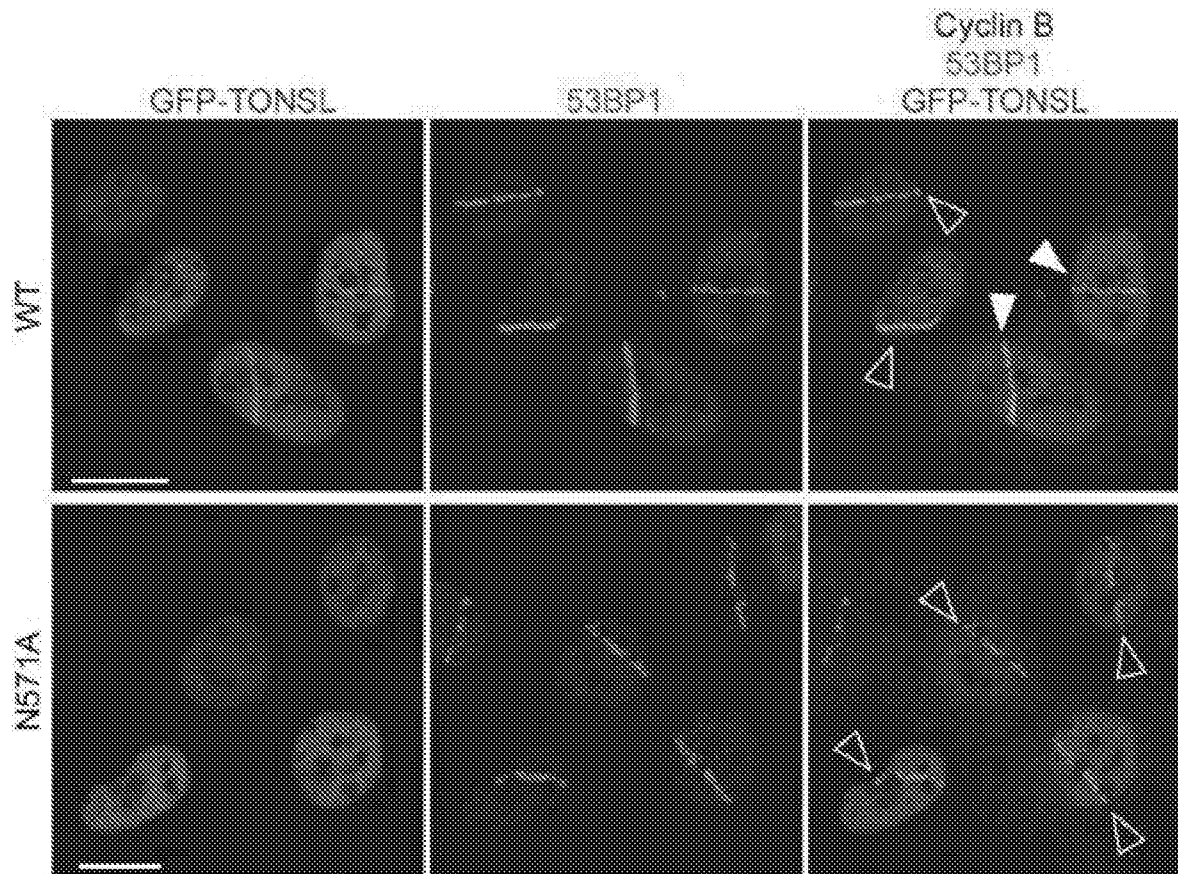


Fig. 23

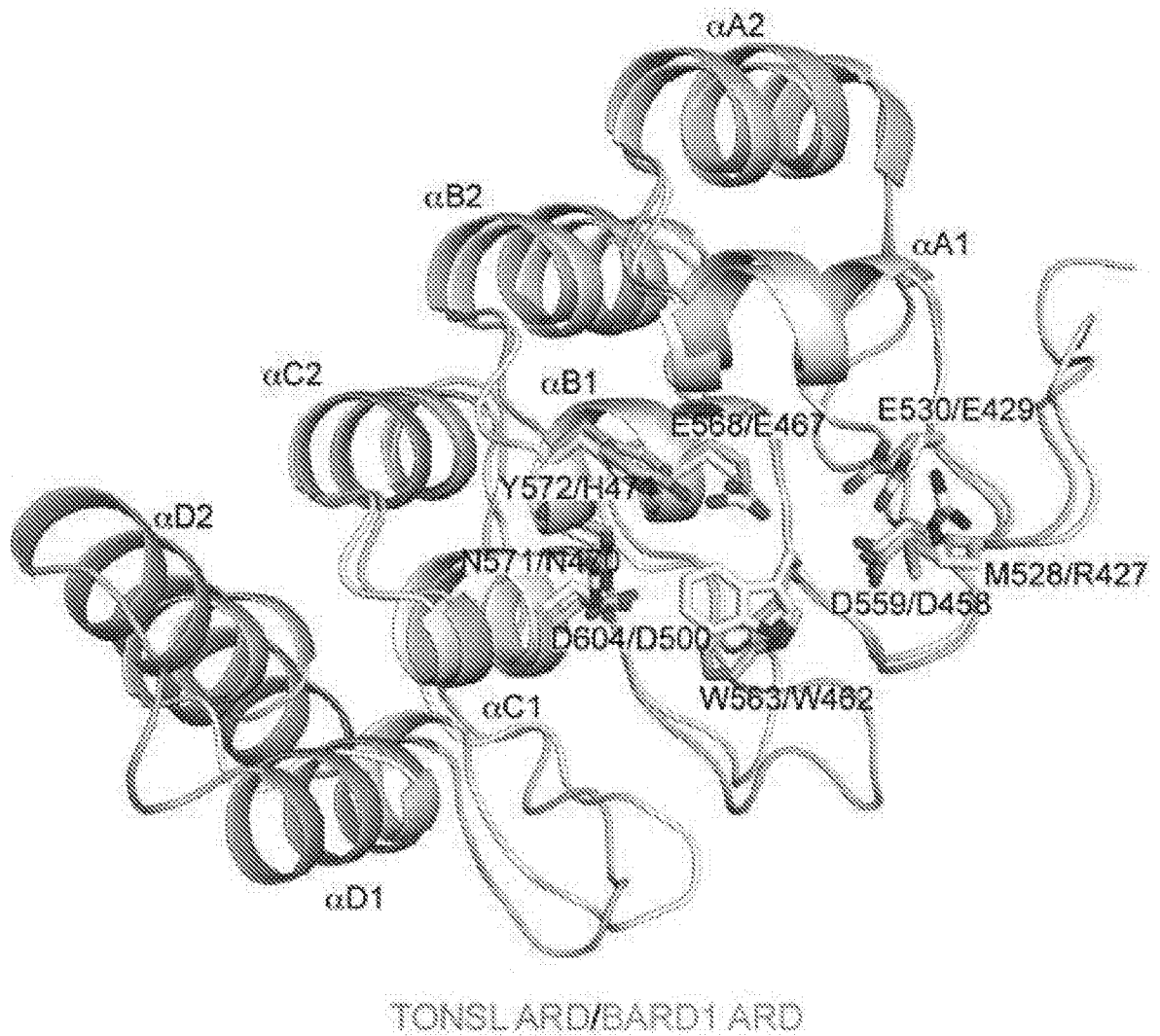
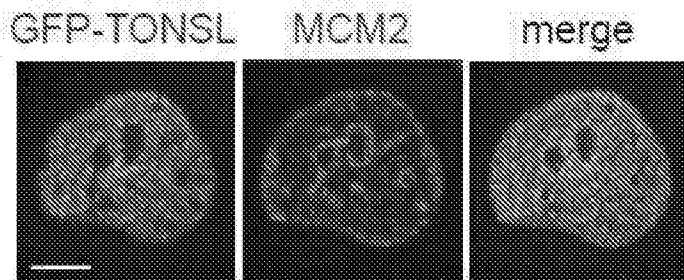


Fig. 24

A)



Pearson coefficient, 0.42 ± 0.06

B)

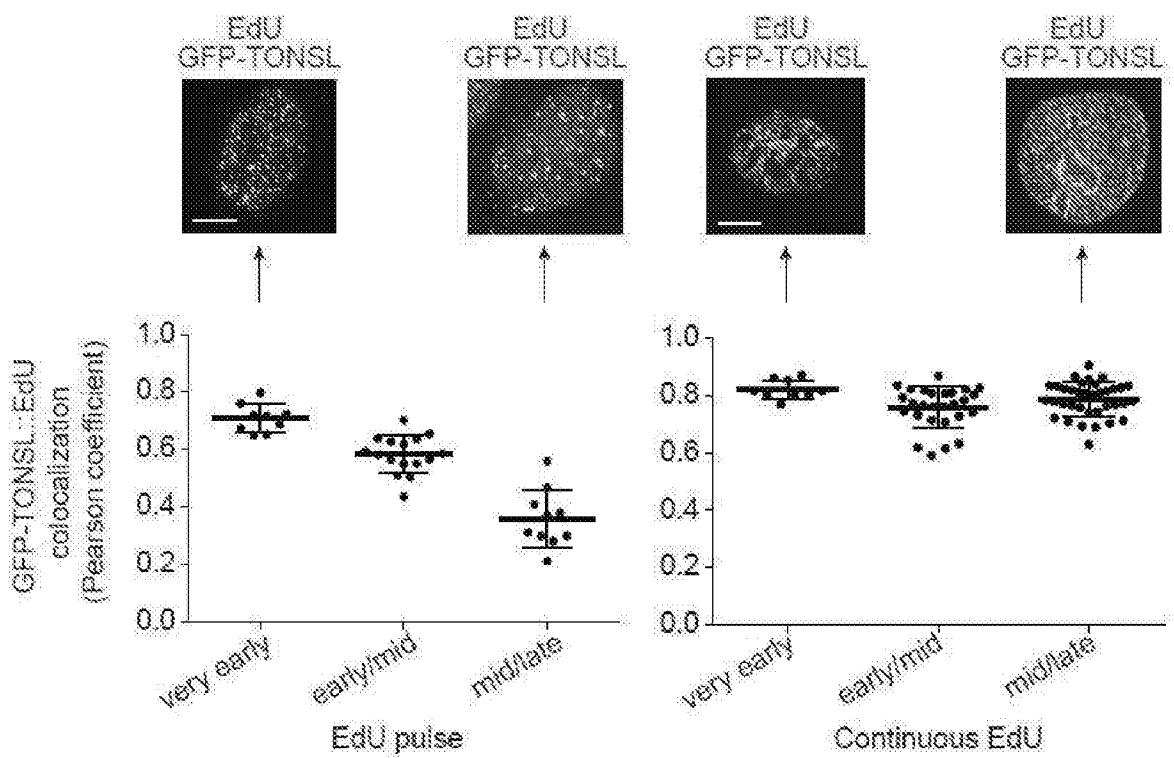


Fig. 25

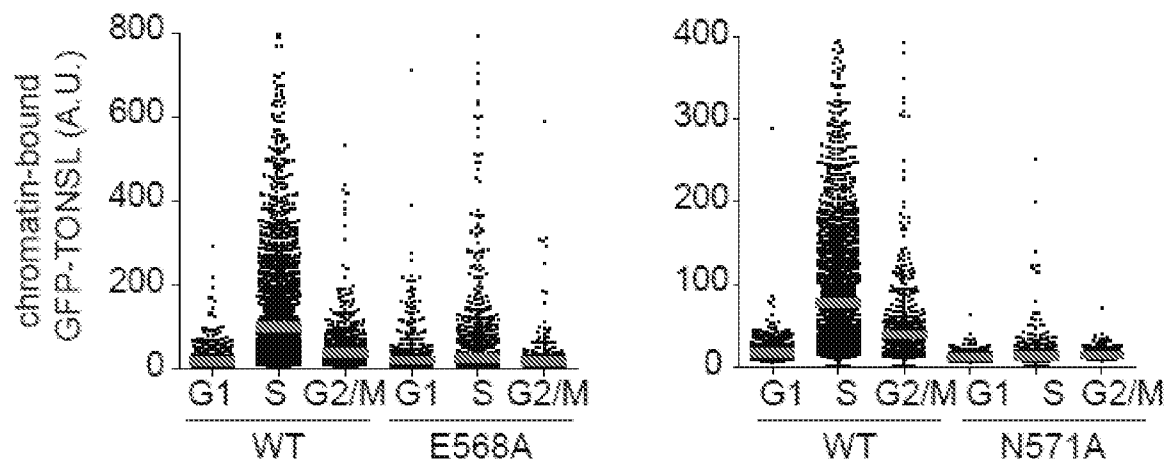


Fig. 26

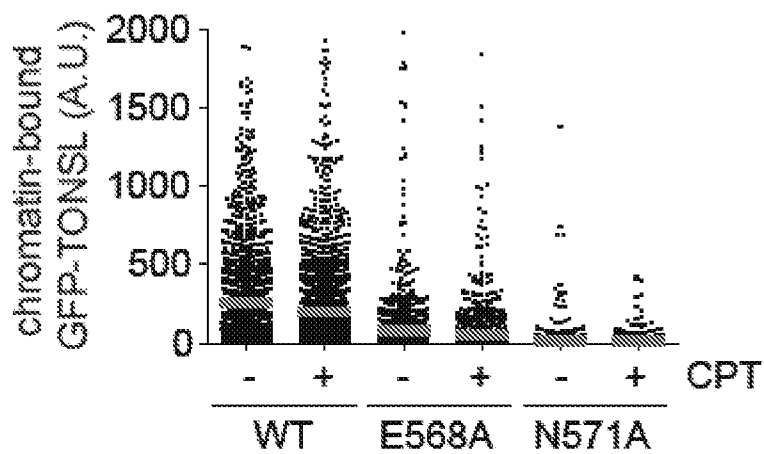
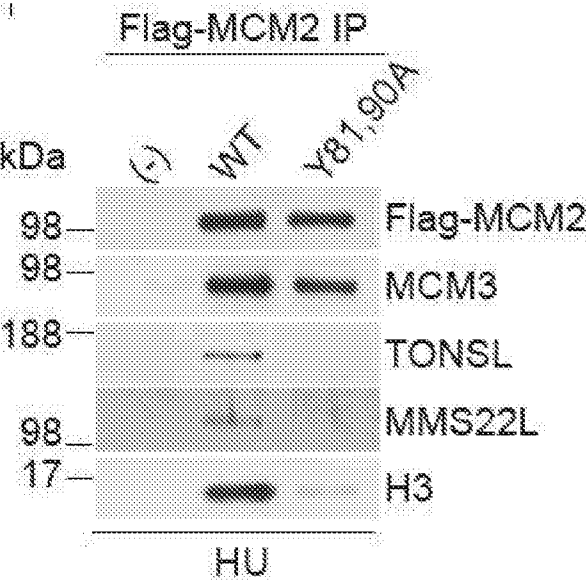


Fig. 27

A)



B)

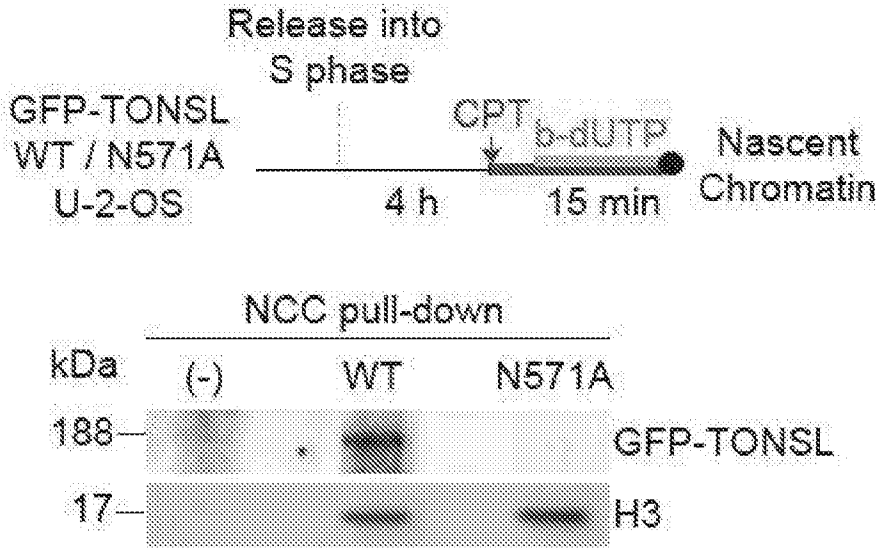


Fig. 28

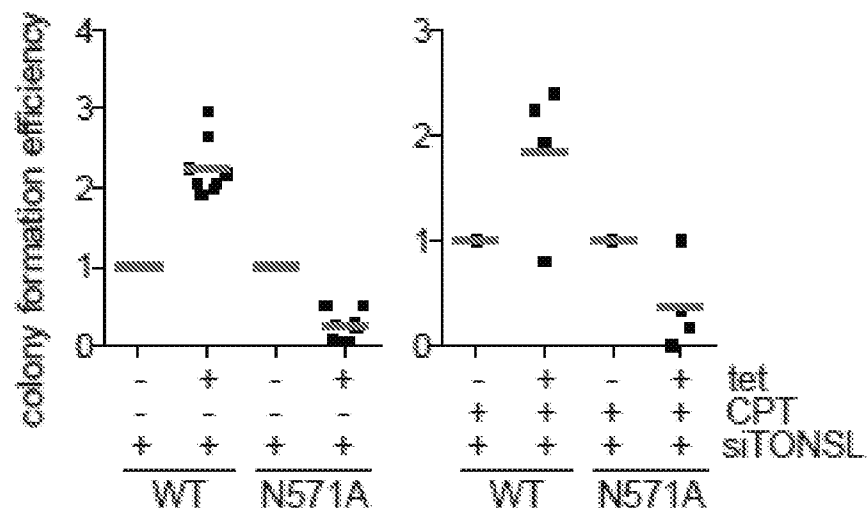


Fig. 29

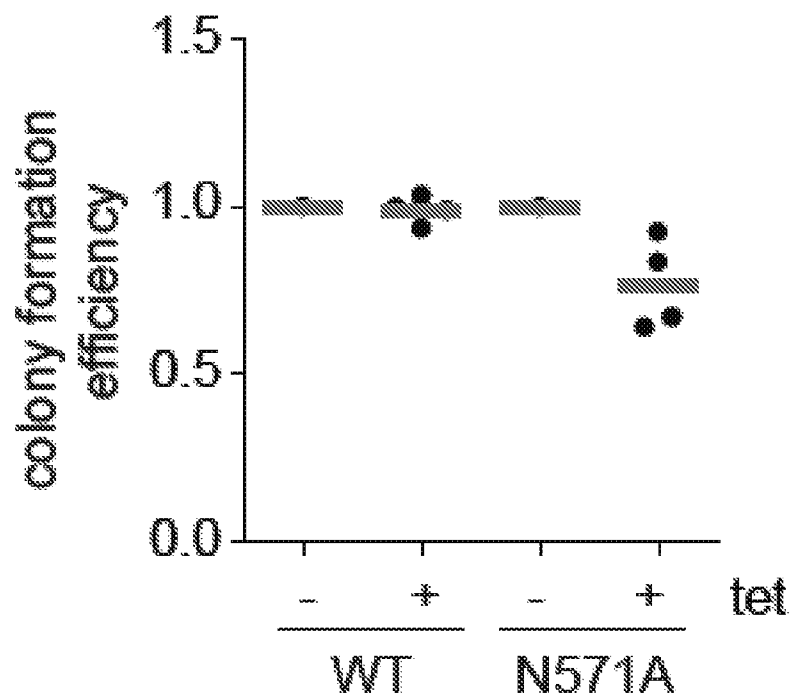
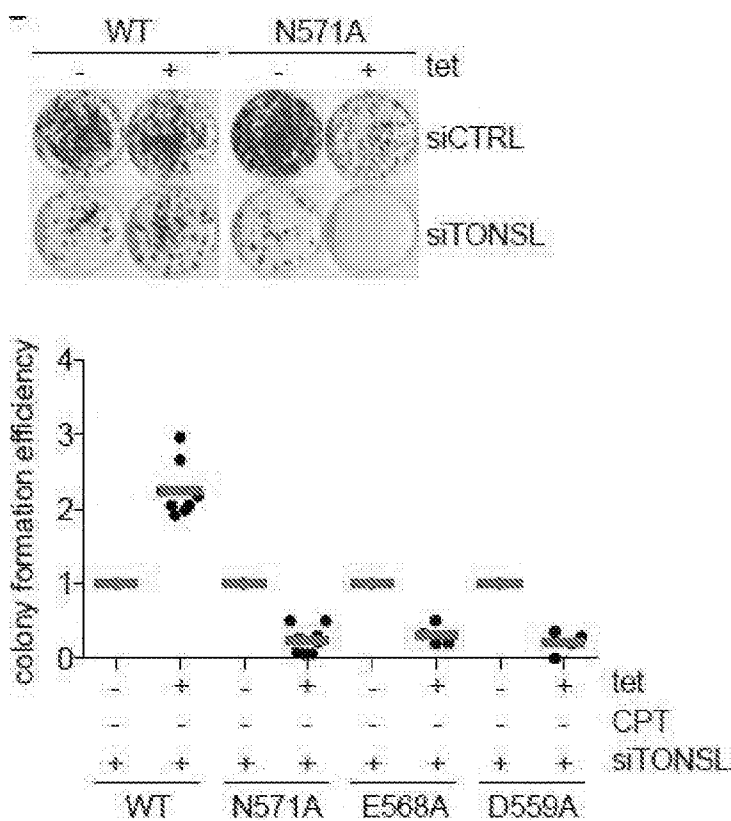


Fig. 30

A)



B)

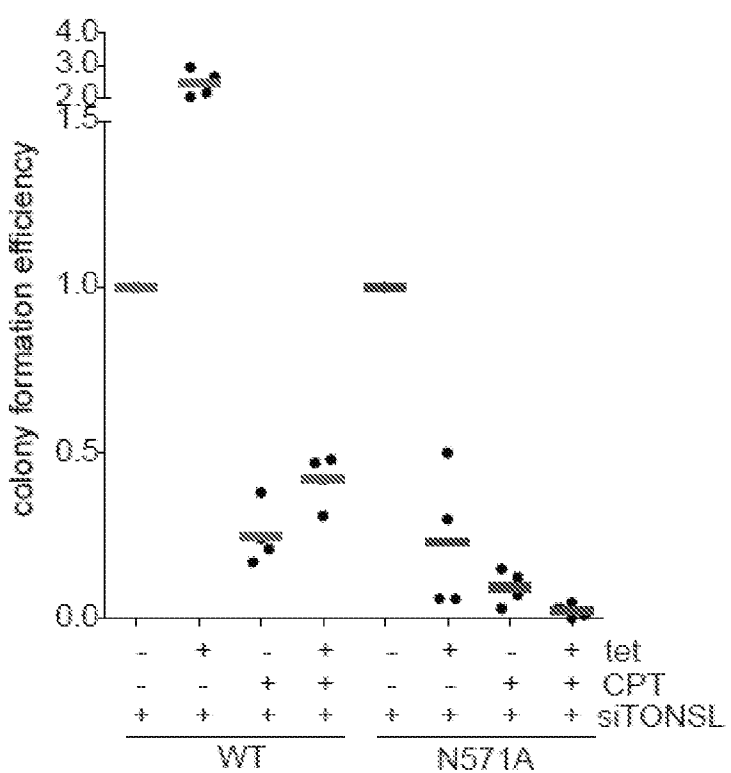


Fig. 31

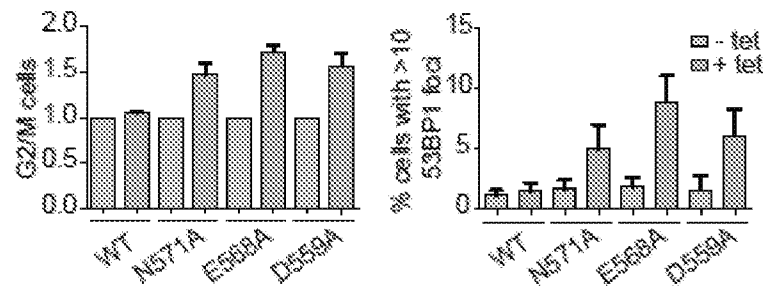
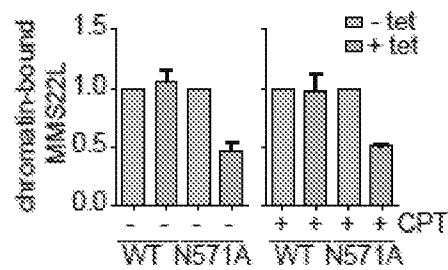


Fig. 32

A)



B)

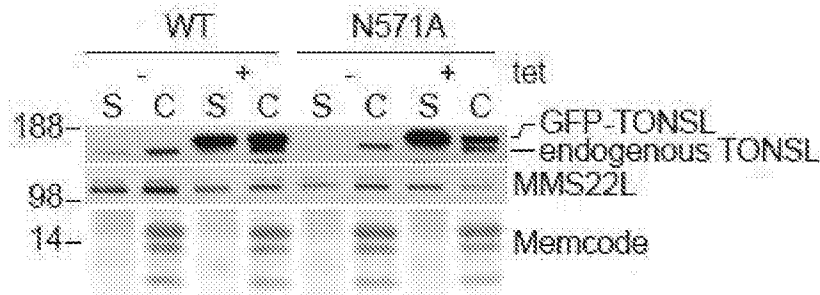


Fig. 33

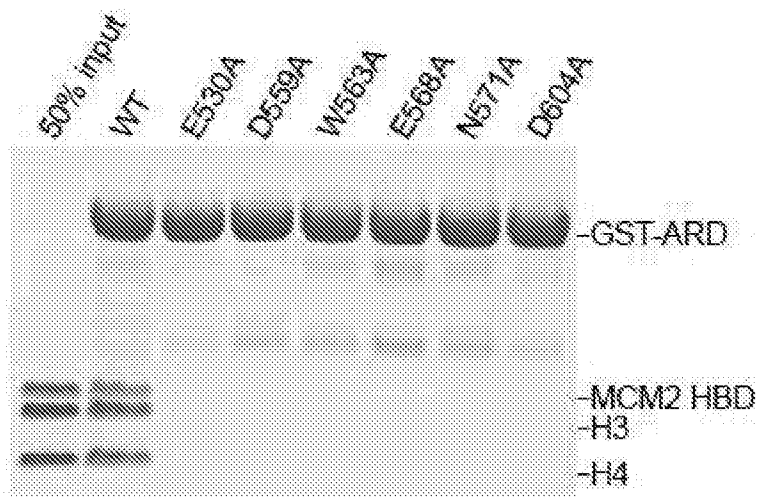


Fig. 34

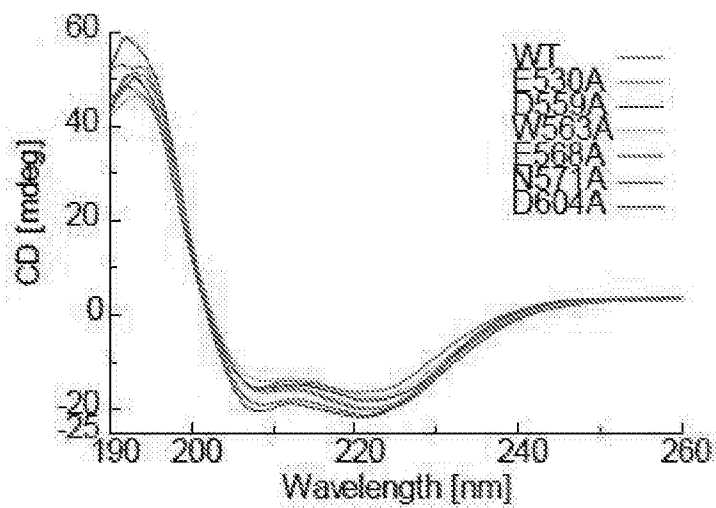


Fig. 35

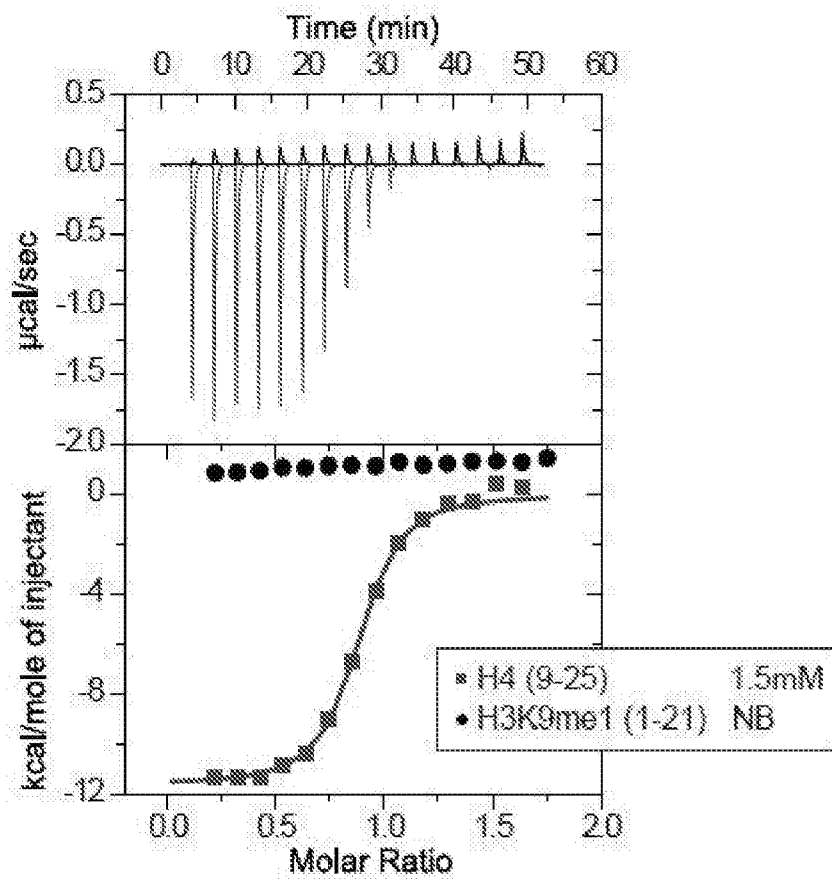


Fig. 36

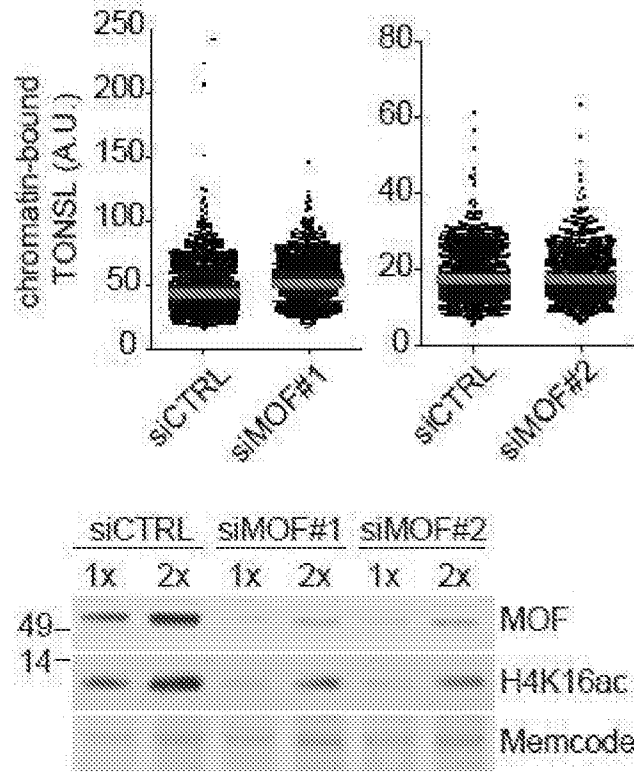
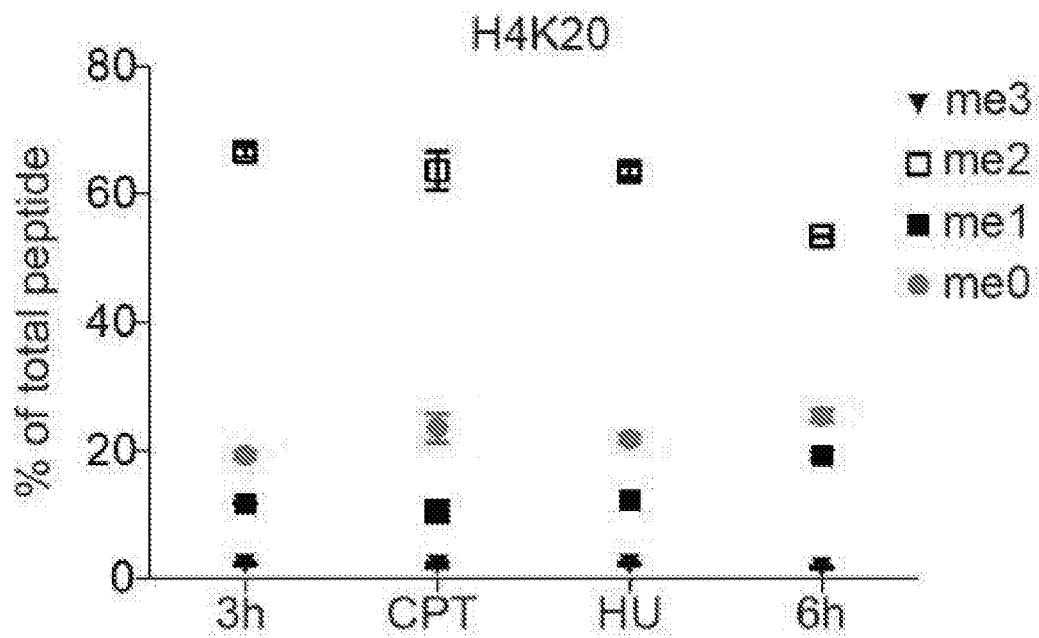


Fig. 37



INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2016/050317

A. CLASSIFICATION OF SUBJECT MATTER
 INV. A61K38/00 C07K14/47
 ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

A61K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MELANIE A. ADAMS-CIOABA ET AL: "Structural basis for the recognition and cleavage of histone H3 by cathepsin L", NATURE COMMUNICATIONS, vol. 2, 1 February 2011 (2011-02-01), page 197, XP055326311, United Kingdom ISSN: 2041-1723, DOI: 10.1038/ncomms1204 Supplementary material; fragment H4 12-27 and 15-27. -& MELANIE A. ADAMS-CIOABA ET AL: "Supplementary Information.-Structural basis for the recognition and cleavage of histone H3 by cathepsin L", NATURE COMMUNICATIONS, vol. 2, 1 February 2011 (2011-02-01), page 197, XP055326313, DOI: 10.1038/ncomms1204 ----- -/--	1-16,20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

1 February 2017

Date of mailing of the international search report

21/02/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

López García, F

INTERNATIONAL SEARCH REPORT

International application No.

PCT/DK2016/050317

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PANAGIS FILIPPAKOPOULOS ET AL: "Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family", CELL, vol. 149, no. 1, 1 March 2012 (2012-03-01), pages 214-231, XP055326309, US ISSN: 0092-8674, DOI: 10.1016/j.cell.2012.02.013 Fig. 6, Fig. 6S -----	1-16,20
X	STEPHEN M FUCHS ET AL: "Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition", CURRENT BIOLOGY, CURRENT SCIENCE, GB, vol. 21, no. 1, 23 November 2010 (2010-11-23), pages 53-58, XP028150458, ISSN: 0960-9822, DOI: 10.1016/J.CUB.2010.11.058 [retrieved on 2010-11-25] Supplementary material, Table S1 -& STEPHEN M FUCHS ET AL: "Supplemental Information.-Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition", CURRENT BIOLOGY, vol. 21, no. 1, 23 November 2010 (2010-11-23), pages 53-58, XP055326393, -----	1-16,20
X	HIROSHI KIMURA ET AL: "What happens to nucleosomes during transcription?", THE JOURNAL OF CELL BIOLOGY : JCB, vol. 267, no. 7, 25 June 2001 (2001-06-25), pages 2837-1354, XP055326388, US ISSN: 0021-9525, DOI: 10.1038/46567 Conjugate H4-GFP -----	1-20
X	WO 2011/098400 A1 (HOFFMANN LA ROCHE [CH]; HESSE FRIEDERIKE [DE]; HOESS EVA [DE]; MUELLER) 18 August 2011 (2011-08-18) PEG-RHAK (SEQ. ID. No.:2) -----	1-12
X	US 2013/274295 A1 (PUJOL ONOFRE AURORA [ES]) 17 October 2013 (2013-10-17) Peptide RHKK (SEQ. ID. No.:7) ----- -/--	1-12

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MIRAL DIZDAROGLU ET AL: "Identification and Quantification of DNA Repair Proteins by Liquid Chromatography/Isotope-Dilution Tandem Mass Spectrometry Using Their Fully 15 N-Labeled Analogues as Internal Standards", JOURNAL OF PROTEOME RESEARCH., vol. 10, no. 8, 5 August 2011 (2011-08-05) , pages 3802-3813, XP055326449, US ISSN: 1535-3893, DOI: 10.1021/pr200269j Table 2, sequence VLR	1-11,14
X	----- WO 2013/148178 A1 (BROAD INST INC [US]) 3 October 2013 (2013-10-03) GKGGAKRHRKVL (SEQ. ID. No.: 261); RHRK (SEQ. ID. No.: 327)	1-13,16
X	----- DATABASE UniProt [Online] 21 July 1986 (1986-07-21), "RecName: Full=Histone H4;", XP002765705, retrieved from EBI accession no. UNIPROT:P62805 Database accession no. P62805 the whole document	1-20
X	----- DATABASE UniProt [Online] 26 November 2014 (2014-11-26), "RecName: Full=Histone H4 {ECO:0000256 RuleBase:RU000528}; Flags: Fragment;", XP002765706, retrieved from EBI accession no. UNIPROT:A0A091VRC9 Database accession no. A0A091VRC9 the whole document ----- -/--	1-20

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 8 April 2008 (2008-04-08), "RecName: Full=Tonsoku-like protein; AltName: Full=Inhibitor of kappa B-related protein; Short=I-kappa-B-related protein; Short=IkappaBR; AltName: Full=NF-kappa-B inhibitor-like protein 2; AltName: Full=Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 2;", XP002765707, retrieved from EBI accession no. UNIPROT:Q96HA7 Database accession no. Q96HA7 the whole document -----	1-10,26,47
X	WO 2012/123119 A1 (CELLZOME AG [DE]; DREWES GERARD [DE]; JOBERTY GERARD [DE]) 20 September 2012 (2012-09-20) SEQ. ID. No.:4 -----	1-23
X	WO 2015/023796 A2 (SHEA LONNIE D [US]; MILLER STEPHEN D [US]; YAP JONATHAN WOON TECK [US]) 19 February 2015 (2015-02-19) SGRGKGGKGLGKGGAKRHRKVLRDNIQGI (SEQ. ID. No.: 3773) and LGKGGAKRHRKVLRD (SEQ. ID. No: 4149) -----	1-20
X	WO 2011/010715 A1 (UNIV TOKYO [JP]; KODAMA TATSUHIKO [JP]; TANAKA TOSHIYA [JP]; KAWAMURA) 27 January 2011 (2011-01-27) AKRHRKVLRD (paragraph 64) -----	1-15,20-22
X	HUA YUAN ET AL: "MYST protein acetyltransferase activity requires active site lysine autoacetylation", EMBO JOURNAL., vol. 31, no. 1, 4 January 2012 (2012-01-04), pages 58-70, XP055327190, GB ISSN: 0261-4189, DOI: 10.1038/emboj.2011.382 GKGGAKRHRKIL -----	1-12,16
X	WO 2013/040142 A2 (IOGENETICS LLC [US]; BREMEL ROBERT D [US]; HOMAN JANE [US]) 21 March 2013 (2013-03-21) GFSGAKRDRPEL (SEQ. ID. No.: 1943695) -----	1-11,16
X	WO 2012/082752 A2 (MOREHOUSE SCHOOL OF MEDICINE [US]; LILLARD JAMES W [US]; SINGH SHAILES) 21 June 2012 (2012-06-21) FYLPKRHRKVCGRP (SEQ. ID. No.: 123) -----	1-12,16
	-/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/031072 A1 (LA ROSA THOMAS J [US] ET AL) 12 February 2004 (2004-02-12) SEQ. ID. No: 213350 -----	1-11,14, 16-18
X	US 2011/214206 A1 (LA ROSA THOMAS J [US] ET AL) 1 September 2011 (2011-09-01) SEQ. ID. No.: 224649 -----	1-11,14, 16-18
X	DATABASE UniProt [Online] 1 November 1996 (1996-11-01), "SubName: Full=Histone H4-1 {ECO:0000313 EMBL:CAA66645.1}; Flags: Fragment;", XP002765708, retrieved from EBI accession no. UNIPROT:Q27821 Database accession no. Q27821 the whole document -----	1-12, 16-18
X	WO 2012/050963 A2 (GEN HOSPITAL CORP [US]; BERNSTEIN BRADLEY [US]; GOREN ALON [US]) 19 April 2012 (2012-04-19) SEQ. ID. No.: 56 -----	1-17
X	WO 2012/158122 A1 (UNIV NANYANG TECH [SG]; LIU CHUAN-FA [SG]; LI FUPENG [SG]) 22 November 2012 (2012-11-22) Example 5, Fig. 8 -----	1-16
X	US 2011/152122 A1 (BERGER SHELLEY L [US] ET AL) 23 June 2011 (2011-06-23) Fig. 4D -----	1-12,16, 17
X	US 2009/183270 A1 (ADAMS THOMAS R [US] ET AL) 16 July 2009 (2009-07-16) SEQ. ID. Nos.: 27014 and 8717 -----	1-18
X	WO 2015/001383 A1 (INST RECH DEVELOPPEMENT IRD [FR]; UNIV PERUANA CAYETANO HEREDIA [PE]) 8 January 2015 (2015-01-08) SEQ. ID. No.: 33 -----	1-11,14, 18
X	WO 2007/132177 A1 (UNIV BIRMINGHAM [GB]; NIGHTINGALE KARL PETER [GB]) 22 November 2007 (2007-11-22) SEQ. ID. No.: 123, 124, 126, 127, 128, 129, 130 ----- -/--	1-11,14, 18

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ERIS DURO ET AL: "Identification of the MMS22L-TONSL Complex that Promotes Homologous Recombination", MOLECULAR CELL, vol. 40, no. 4, 1 November 2010 (2010-11-01), pages 632-644, XP055076349, ISSN: 1097-2765, DOI: 10.1016/j.molcel.2010.10.023 SEQ. ID. No.: 6815	1-11,14, 16-18
A	----- SOPHIA B. CHERNIKOVA ET AL: "Inhibiting homologous recombination for cancer therapy", CANCER BIOLOGY & THERAPY, vol. 13, no. 2, 15 January 2012 (2012-01-15), pages 61-68, XP055326067, US ISSN: 1538-4047, DOI: 10.4161/cbt.13.2.18872 the whole document	1-63
A	----- WO 2008/129239 A2 (ISIS INNOVATION [GB]; HELLEDAY THOMAS [GB]) 30 October 2008 (2008-10-30) the whole document	1-63
A	----- LUCA LIVRAGHI ET AL: "PARP inhibitors in the management of breast cancer: current data and future prospects", BMC MEDICINE, BIOMED CENTRAL LTD., LONDON, GB, vol. 13, no. 1, 13 August 2015 (2015-08-13), page 188, XP021227264, ISSN: 1741-7015, DOI: 10.1186/S12916-015-0425-1 the whole document	1-63
A	----- WO 2005/012524 A1 (UNIV SHEFFIELD [GB]; HELLEDAY THOMAS [SE]) 10 February 2005 (2005-02-10) the whole document	1-63
X,P	----- GIULIA SAREDI ET AL: "H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex", NATURE, vol. 534, no. 7609, 22 June 2016 (2016-06-22), pages 714-718, XP055326014, United Kingdom ISSN: 0028-0836, DOI: 10.1038/nature18312 the whole document	1-63
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2016/050317

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	CAMPOS ERIC I ET AL: "Analysis of the Histone H3.1 Interactome: A Suitable Chaperone for the Right Event", MOLECULAR CELL, vol. 60, no. 4, 29 October 2015 (2015-10-29), pages 697-709, XP029306640, ISSN: 1097-2765, DOI: 10.1016/J.MOLCEL.2015.08.005 the whole document	1-63
X	----- DATABASE UniParc [Online] 26 May 2015 (2015-05-26), XP002765726, Database accession no. UPI000290A5B1 the whole document	47,48
X	----- DATABASE UniProt [Online] 9 January 2013 (2013-01-09), "SubName: Full=Tonsoku-like, DNA repair protein {ECO:0000313 EMBL:JAA24846.1};" , XP002765725, retrieved from EBI accession no. UNIPROT:K7B3J4 Database accession no. K7B3J4 the whole document -----	47,48

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/DK2016/050317

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011098400 A1	18-08-2011	CA 2786340 A1	18-08-2011
		CN 102740896 A	17-10-2012
		EP 2533813 A1	19-12-2012
		ES 2587574 T3	25-10-2016
		HK 1171375 A1	05-12-2014
		JP 5749744 B2	15-07-2015
		JP 6014194 B2	25-10-2016
		JP 2013518922 A	23-05-2013
		JP 2015145402 A	13-08-2015
		KR 20120137364 A	20-12-2012
		MX 339559 B	31-05-2016
		RU 2012138553 A	20-03-2014
		US 2012309679 A1	06-12-2012
		WO 2011098400 A1	18-08-2011
US 2013274295 A1	17-10-2013	NONE	
WO 2013148178 A1	03-10-2013	NONE	
WO 2012123119 A1	20-09-2012	NONE	
WO 2015023796 A2	19-02-2015	AU 2014306603 A1	18-02-2016
		CA 2918823 A1	19-02-2015
		CN 105555301 A	04-05-2016
		EP 3033102 A2	22-06-2016
		JP 2016528254 A	15-09-2016
		KR 20160042079 A	18-04-2016
		US 2015190485 A1	09-07-2015
		US 2015283218 A1	08-10-2015
		WO 2015023796 A2	19-02-2015
WO 2011010715 A1	27-01-2011	NONE	
WO 2013040142 A2	21-03-2013	EP 2771349 A2	03-09-2014
		WO 2013040142 A2	21-03-2013
WO 2012082752 A2	21-06-2012	CN 103619351 A	05-03-2014
		CN 106177946 A	07-12-2016
		EP 2651443 A2	23-10-2013
		JP 2014508117 A	03-04-2014
		US 2012128688 A1	24-05-2012
		WO 2012082752 A2	21-06-2012
US 2004031072 A1	12-02-2004	NONE	
US 2011214206 A1	01-09-2011	US 2011214206 A1	01-09-2011
		US 2012216318 A1	23-08-2012
WO 2012050963 A2	19-04-2012	EP 2621942 A2	07-08-2013
		US 2013217027 A1	22-08-2013
		US 2016068885 A1	10-03-2016
		WO 2012050963 A2	19-04-2012
WO 2012158122 A1	22-11-2012	CN 103534235 A	22-01-2014
		EP 2707356 A1	19-03-2014
		SG 194727 A1	30-12-2013
		US 2014316105 A1	23-10-2014
		WO 2012158122 A1	22-11-2012

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/DK2016/050317

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011152122	A1	23-06-2011	NONE
US 2009183270	A1	16-07-2009	US 2009183270 A1 16-07-2009 US 2012227131 A1 06-09-2012 US 2016264980 A1 15-09-2016
WO 2015001383	A1	08-01-2015	EP 3017305 A1 11-05-2016 US 2016146809 A1 26-05-2016 WO 2015001383 A1 08-01-2015
WO 2007132177	A1	22-11-2007	EP 2016420 A1 21-01-2009 JP 2009536336 A 08-10-2009 US 2009149343 A1 11-06-2009 WO 2007132177 A1 22-11-2007
WO 2008129239	A2	30-10-2008	NONE
WO 2005012524	A1	10-02-2005	AT 516353 T 15-07-2011 AU 2004261779 A1 10-02-2005 BR PI0412909 A 26-09-2006 CA 2533423 A1 10-02-2005 CN 1856313 A 01-11-2006 CN 1856572 A 01-11-2006 CN 102935230 A 20-02-2013 DK 1649017 T3 17-10-2011 EP 1649017 A1 26-04-2006 ES 2367433 T3 03-11-2011 GB 2419882 A 10-05-2006 HK 1094006 A1 28-03-2014 JP 5519097 B2 11-06-2014 JP 5547164 B2 09-07-2014 JP 5848728 B2 27-01-2016 JP 2006528618 A 21-12-2006 JP 2012102099 A 31-05-2012 JP 2014001210 A 09-01-2014 KR 20060052877 A 19-05-2006 NO 338398 B1 15-08-2016 NZ 545307 A 28-08-2009 PT 1649017 E 30-09-2011 US 2007179160 A1 02-08-2007 US 2015005327 A1 01-01-2015 WO 2005012524 A1 10-02-2005 ZA 200601386 B 25-04-2007