A new mRNA structure prediction based approach to identifying improved signal peptides for bone morphogenetic protein 2

Piers Wilkinson[1*], Brian Jackson[2], Hazel Fermor[3] and Robert Davies[4]

Abstract

Background Signal peptide (SP) engineering has proven able to improve production of many proteins yet is a laborious process that still relies on trial and error. mRNA structure around the translational start site is important in translation initiation and has rarely been considered in this context, with recent improvements in *in silico* mRNA structure potentially rendering it a useful predictive tool for SP selection. Here we attempt to create a method to systematically screen candidate signal peptide sequences *in silico* based on both their nucleotide and amino acid sequences. Several recently released computational tools were used to predict signal peptide activity (SignalP), localization target (DeepLoc) and predicted mRNA structure (MXFold2). The method was tested with Bone Morphogenetic Protein 2 (BMP2), an osteogenic growth factor used clinically for bone regeneration. It was hoped more effective BMP2 SPs could improve BMP2-based gene therapies and reduce the cost of recombinant BMP2 production.

Results Amino acid sequence analysis indicated 2,611 SPs from the TGF-β superfamily were predicted to function when attached to BMP2. mRNA structure prediction indicated structures at the translational start site were likely highly variable. The five sequences with the most accessible translational start sites, a codon optimized BMP2 SP variant and the well-established hIL2 SP sequence were taken forward to in vitro testing. The top five candidates showed non-significant improvements in BMP2 secretion in HEK293T cells. All showed reductions in secretion versus the native sequence in C2C12 cells, with several showing large and significant decreases. None of the tested sequences were able to increase alkaline phosphatase activity above background in C2C12s. The codon optimized control sequence and hIL2 SP showed reasonable activity in HEK293T but very poor activity in C2C12.

Conclusions These results support the use of peptide sequence based *in silico* tools for basic predictions around signal peptide activity in a synthetic biology context. However, mRNA structure prediction requires improvement before it can produce reliable predictions for this application. The poor activity of the codon optimized BMP2 SP variant in C2C12 emphasizes the importance of codon choice, mRNA structure, and cellular context for SP activity.

Keywords Signal peptides[1], Bone morphogenetic protein[2], RNA structure[3], Protein engineering[4], Synthetic biology[5], Gene therapy[6], Regenerative medicine[7]

*Correspondence:
Piers Wilkinson
mnpw@leeds.ac.uk
Full list of author information is available at the end of the article

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 2 of 16

Background

Signal peptides (SPs), also known as signal sequences, are the region of secreted proteins that target them to a secretion pathway. In eukaryotes most secreted proteins use the signal recognition particle (SRP) secretion pathway, relying on short N-terminal signal peptides that induce co-translational translocation into the endoplasmic reticulum. The SRP pathway and its SPs have seen a great deal of basic research over several decades and much is understood of their function (reviewed elsewhere [1]), however SP sequences are highly diverse and it is not understood why some function better than others in specific contexts.

A rarely considered aspect of SP behaviour is the influence of its nucleotide sequence on translation initiation. The SP nucleotide sequence lies at the 5' end of coding sequence and contains the start codon and the last nucleotide of the Kozak sequence. Crucially, it is likely to form secondary structures with the latter segment of the 5' UTR, the region containing the ribosomal attachment site. mRNA structure around the ribosomal attachment site is now established as an important mediator of translation initiation, with stable secondary structure in the region known to inhibit the process [2–6]. It therefore seems prudent to consider mRNA structure alongside amino acid sequence when selecting SP candidates. However, the large majority of SP optimization or substitution approaches do not do so, likely due to difficulties in predicting structure *in silico* or establishing it empirically. There have been a handful of attempts to integrate these factors in the past using *in silico* methods, though these were held back by difficulties in predicting mRNA structure and saw little success [7, 8]. The years since these attempts have seen rapid improvement in computational techniques due to advances in machine learning [9–13]. It was hoped the new generation of tools using these methods might offer improved performance for both amino acid sequence based SP prediction and mRNA structure prediction.

The method was tested with Bone Morphogenetic Protein 2 (BMP2), an osteogenic growth factor in the TGF-β family that plays important roles in bone development, homeostasis and healing [14–19]. BMP2 has seen extensive use in regenerative approaches for bone healing and other orthopaedic applications, both as a recombinant protein and a gene therapy [20, 21]. The aim was to ameliorate methodologies to improve BMP2 secretion, hopefully improving the effectiveness of BMP2 gene therapies and reduce the cost of recombinant BMP2 production. Only one prior publication has attempted to identify improved SPs for BMP2 using a non-systematic approach, failing to identify any SPs offering improved secretion [22].

Here we present a new method to identify SPs to improve BMP2 secretion. The method employs various up-to-date computational tools from other authors to predict SP activity based on amino acid and nucleotide sequences, with particularly attention paid to the predicted mRNA structure at the ribosomal attachment site. A subset of the top results and two manually selected sequences of interest were then tested in vitro in two cells lines, with protein secretion and osteogenic activity investigated. We report the strengths and weaknesses of the technique and suggest several approaches to improve the method.

Methods

Acquisition of signal peptide amino acid sequences and creation of the *in silico* fusion peptide library

The protein sequences for all identified members of the TGF-β superfamily in all species were downloaded from the UniProt database as an XML file on 13/10/2022 (URL: https://www.uniprot.org/uniprotkb/?query=family:%22TGF-beta+family%22&sort=score) [23] . The SP amino acid sequences of proteins annotated with SPs were isolated and attached to the hBMP2 propeptide amino acid sequence, creating an *in silico* library of fusion proteins. Nucleotide sequences corresponding to the SP amino acid sequences were retrieved by isolating source database information and accession numbers from the Uni-Prot data. The various nucleotide sequence databases cited (NCBI nucleotide, EBI-ENA, Ensembl, WormBase ParaSite [24–27]) were then accessed by their application programming interfaces to retrieve the relevant sequence data. After retrieval, the nucleotide sequence dataset was subjected to an availability and validity check using the following criteria: nucleotide sequence successfully retrieved, sequence in frame, sequence begins with a start codon, and translated nucleotide sequence matches the UniProt protein sequence.

Prediction of signal peptide function and localisation with signalP and deepLoc

The *in silico* fusion proteins were analysed with signal peptide prediction software SignalP 6.0 (Technical University of Denmark) to predict if they would be recognised by secretion machinery [11]. The fast model was chosen due to substantially reduced completion time for the large data set used here. SignalP gives a score for the predicted likelihood of the sequence containing a functional eukaryotic SRP pathway signal peptide (called the Sec/SPI score) between 0 and 1, with sequences given higher scores considered more likely to contain an SRP signal peptide. An exclusion threshold of Sec/SPI < 0.5 was set.

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 3 of 16

The signal peptide targeting predictor DeepLoc 2.0 (Technical University of Denmark) was used to predict the localisation targets of the SPs [12]. DeepLoc provides predictions of localisation to various intracellular targets and the extracellular space, giving a score of 0 to 1 for each location with higher values indicating increased confidence in the prediction. An exclusion threshold of a < 0.5 for extracellular localisation was set.

Manually selected sequences

The expression optimisation tool "Translation Initiation coding region designer" (TISIGNER, https://tisigner.com/tisigner) was used to design a codon optimised version of the endogenous hBMP2 SP. TISIGNER prioritises minimal mRNA 5' opening energy rather than using a more standard tRNA availability based approach (see Bhandari et al. for a full description of the method [28]). Host organism was specified as "Other" and the following promoter and 5' UTR sequence from the intended pVax plasmid vector was entered to allow mRNA structure prediction (5' – 3' orientation):

GTGATGCGGTTTTGGCAGTACATCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAA TGTCGTAACAACTCCGCCCCATTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCTCTGGCTAACTAGAGAACCCACT GCTTACTGGCTTATCGAAATTAATACGACTCAC TATAGGGAGACCCAAGCTGGCTAGCGTTTAAAC TTAAGCTTGGTACCGAGCTCGGATCCACTAGTC CAGTGTGGTGGAATTCGGCTTGCCACC.

The full hBMP2 mature transcript nucleotide sequence from the NCBI nucleotide database (Accession: NM_001200.4) was then entered for optimisation.

The signal peptide nucleotide sequence from the hIL2 mRNA was taken from the hIL2 references sequence on NCBI nucleotide (Accession: NM_000586.4).

Creation of *in silico* predicted mRNAs and structure, stability and opening energy prediction

To predict the structure and stability of the fusion proteins' mRNAs when expressed from the intended pVax_BMP2 based plasmid vector (sequence available at https://www.ncbi.nlm.nih.gov/nuccore/MK433563.1), *in silico* predicted mRNAs were created. In 5' to 3' orientation these contained the 5' UTR of pVax_BMP2, the SP of interest, the hBMP2 propeptide, the pVax_BMP2 3' UTR, the BGH poly-A sequence to 20 bp downstream of the AATAA polyadenylation site, then a 150 bp poly A tail. The resulting predicted mRNA sequences were ~ 1.7 kb in length.

The RNA secondary structure predictor MXFold2 (Department of Biosciences and Informatics, Keio University, Japan) was used to predict the secondary structures and calculate the stability of the predicted mRNAs [13]. Stability was quantified with minimum free energy (MFE) values, which account for the predicted energetic contribution of every feature in the predicted secondary structure [13]. To calculate opening energy the MXFold2-generated secondary structures were submitted to RNAeval (see ViennaRNA package [29]) to provide the predicted energetic contribution of each bond in the MXFold2 generated structure. Opening energy was predicted by extracting the energy values for the ± 15nt segment surrounding the start codon (previously established as a relevant window by others [3–5]) using a python script. A simple metric was devised to simultaneously consider the predicted opening energy and total mRNA stability values, generating a metric termed SP score.

$$SPscore = \frac{MFE}{opening\ energy}$$

Higher SP scores indicate larger predicted MFE and smaller predicted opening energy; the hypothetically favourable combination of properties for maximising expression.

Additional *in silico* analysis

RNA structure visualisation was performed with the Forna web app. (http://rna.tbi.univie.ac.at/forna/). See Kerpedjiev et al. for details [30]. Nucleotide and amino acid sequence alignments were performed in SnapGene (SnapGene software, www.snapgene.com). The final 7 candidates and the endogenous hBMP2 SP were submitted to the SignalP 6 web app (https://services.healthtech.dtu.dk/services/SignalP-6.0/) to provide predictions of region boundaries [11]. "Organism: Eukarya", "Output format: Long" and "Model mode: Slow" were specified prior to running the model.

Plasmids

The plasmids pVax_BMP2 and pVax_BMP7 were kindly provided by Dr Georg Feichtinger (see Feichtinger et al. for details [31]). The pVax_GFP control plasmid was produced from pVax_BMP2 by restriction cloning. Briefly, the BMP2 CDS was removed and replaced with EGFP excised from the pCAG_GFP plasmid previously purchased from Addgene (https://www.addgene.org/11150/). Novel SP fragment synthesis, restriction cloning into pVax_BMP2 and endotoxin free maxi-prep scale up were outsourced using the Genewiz TurboGENE 7 Day service (Azenta life sciences).

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 4 of 16

Cell culture

HEK293T cells were kindly provided by Dr Brian Jackson. The cells were maintained in DMEM high glucose (D6429, Sigma), 10% FBS (EU-000-F, Seralab), 2 mM L-Glutamine (G7513, Sigma), 1% v/v penicillin/streptomycin. C2C12 cells were purchased from ATCC (CRL-1772). The cells were maintained in DMEM high glucose with 4mM L-Glutamine and 5% FBS, without antibiotics. Both lines were maintained at 37^0C in a 5% CO2 atmosphere and passaged twice a week during maintenance periods.

HEK293T cells were seeded in 24-well plates at 3×10^5 cells/well in 1 ml of complete DMEM, 24 h prior to transfection. Cells were transfected using Lipofectamine 3000 (L3000001, Thermo Fisher Scientific) at the manufacturer recommended high concentration. After 24 h an additional 1 ml of complete DMEM was added to each well to ensure the media was not exhausted before harvesting. Media was harvested 48 h after transfection. SP panel experiments contained 9 groups with each transfected with one of the following plasmids: the top 5 results from the in silico screen, the two manually selected SP sequences (SP hBMP2 TISIGNER and SP hIL2), a pVax_BMP2 positive control and a pVax_GFP negative control. Transfections were always performed in technical duplicate with two wells per group.

C2C12 cells were seeded in 24 well plates at 6×10^4 cells/well in 1 ml of complete C2C12 media. The cells were transfected after 24 h as described above. Experimental groups were the same as those in the HEK293T experiment with the addition of a co-transfection pVax_BMP2 + pVax_BMP7 osteogenesis positive control group (effectiveness previously established by others [31]). In this group 250ng of each plasmid was used per well to match the 500ng used in the other groups. 24 h after lipofection the cells were media changed with 1 ml of complete media supplemented with 10 µg/ml heparin (H3149, Sigma Aldrich) to reduce interaction with heparan sulphate proteoglycans that would otherwise rapidly clear hBMP2 from solution [32, 33]. After a further 24 h (48 h after lipofection) the media was harvested for BMP2 ELISA and new heparin-supplemented media added. The plates were maintained in heparin-supplemented media until 7 days post transfection, with media changes every 2/3 days. On day 7 the wells were washed with PBS and the plates stored at -80^0C for up to 48 h prior to quantitative alkaline phosphatase (ALP) assay. Transfections were always performed in technical duplicate, with two wells per condition. Each well was considered a separate sample for ELISA and ALP assays.

hBMP2 ELISA

hBMP2 capture and biotin-conjugated detection antibodies, CHO-derived rhBMP2 standard and streptavidin-HRP working solution from the hBMP2 DuoSet ELISA system (DY355, R&D systems) were used for all experiments. Nunc MaxiSorp ELISA plates (M9410, Merck), 10% BSA ELISA reagent diluent/blocking solution concentrate (DY995, R&D systems) and TMB substrate (421,501, Biolegend) were purchased separately and used for all experiments. ELISAs were performed according to the hBMP2 DuoSet ELISA system manufacturer recommendations. rhBMP2 standard curves were made up in the appropriate complete media. A Wellwash microplate washer (5,165,000, Thermo Fisher Scientific) was used for all washes. ELISAs were always performed in technical duplicate (two wells per sample). The mean value from the two wells was used for data analysis.

Quantitative ALP assay

Nitrophenol phosphate (NPP) quantitative ALP assays were performed to assess osteogenesis in the C2C12s seven days after transfection. Cells were washed with PBS then lysed with 100ul lysis solution (0.5% Triton X-100 in deionized water) per well with 250RPM radial shaking for 1 h. 10ul from each well was transferred to a new 96 well plate, then a 90ul of NPP working solution (5mM NPP (4876, EMD Millipore), 0.5 M 2-AMP (A9199, Sigma Aldrich), 2mM MgCl2 (25108.260, VWR chemical), pH 10.3 in deionised water) was added to each well. A 4-Nitrophenol (4NP) end-product standard curve of 10–200µM (4NP to concentration(1048, Sigma Aldrich), 0.5 M 2-AMP, 2mM $MgCl_2$, pH 10.3 buffer) was run simultaneously to allow calculation of the final 4NP concentration in each well. Plates were incubated at room temperature for 45 min then absorbances at 405 nm and a 600 nm wavelength control were measured using a microplate spectrophotometer (Multiskan GO microplate reader, Thermo Fisher). ALP assays were performed in technical triplicate (three wells per sample). Mean values were used for analysis.

Statistics

ELISA and ALP data were normalised to the positive control group prior to analysis. For ELISAs the positive control was the pVax_BMP2 group, while for ALP assays the positive control was the pVax_BMP2 + pVax_BMP7 group. For HEK293T ELISA data $N=7$, for the C2C12 ELISA and ALP data $N=3$. The normalised data were then tested for normality using Shapiro-Wilk tests. All data were found to be normally distributed, and significance was tested using one-way ANOVAs with selected

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 5 of 16

Dunnett's multiple comparison tests between the positive control and other groups to improve power. All statistical tests were performed in GraphPad Prism 9.3.1 (Graph-Pad Software).

Results

Results from the *in silico* pipeline

The initial TGF-β family dataset retrieved from UniProt contained 34,861 records. 19,250 protein sequences were found to be annotated with a confirmed or likely signal peptide and were taken forward. Nucleotide sequence retrieval and quality control checks (sequence in frame, sequence begins with a start codon, translated nucleotide sequence matches the UniProt protein sequence) were then performed. 1,344 sequences failed one of these criteria and were removed, leaving 17,906 to be taken forward. An SP nucleotide sequence duplicate removal step was then performed, with 7,545 sequences found to be duplicates and removed. 10,361 unique and matched nucleotide/amino acid sequence pairs remained and were used to create the *in silico* fusion protein and predicted mRNA sequence libraries used for further analysis. The *in silico* fusion protein library was subjected to SP activity and localisation prediction. SignalP analysis indicated that 3,203 of the fusion proteins were predicted to have a low chance of secretion and were excluded, leaving 7,158 sequences remaining. DeepLoc analysis indicated all the fusion proteins were predicted to show extracellular secretion, therefore none were excluded. Nucleotide sequences for these fusion proteins were then analysed for weak Kozak sequences. 4,547 sequences were found to have non-optimal Kozak sequences and were excluded. This left a final set of 2,611 sequences for further analysis. See Fig. 1 for an overview of the filtering process.

The secondary structures of the predicted mRNA sequences for the remaining SP-fusions were predicted with MXFold2. Predicted opening energies varied substantially across the final data set (mean: -343.3 kcal/mol, SD: 106.6), while MFE was much less variable (mean: -451.9 kcal/mol, SD: 11.2). The most promising 5 results from the in silico screen according to their calculated SP score are displayed in Table 1. The sequences were all derived from different vertebrate species. Three sequences were derived from ARTN or closely related

unknown protein orthologs, in all cases from mammalian species. The two additional identified proteins were BMP15 and GDF-10 orthologs, from an avian and piscine species respectively. There was modest variation in SP length, ranging from 22 to 44 residues in length (note the native hBMP2 SP is 23 residues long). The MFEs of the 5 sequences showed little variation (mean: -466.7 kcal/mol, SD: 19.6). Opening energies were more variable but covered only a small fraction of the range seen across the whole dataset (mean: -72.6 kcal/mol, SD: 13.6). Sequence alignments revealed that there was limited similarity between sequences from distantly related species and proteins. There was unsurprisingly strong sequence similarity between the 3 artemin orthologs (see Figs. 2A and 3A), with all displaying a ~ 25 bp insertion. All three sequences were predicted to show the same mRNA secondary structure in the ribosomal binding site window (see Fig. 2B). A loose leucin-rich motif was conserved in all sequences (see Fig. 3A).

The endogenous hBMP2 SP and the two manually selected sequence, SP hBMP2-TIS and SP hIL2, showed similar MFEs but large opening energies compared to the *in silico* pipeline top 5, with IL2 being particularly high (see Table 2) due to predicted low self-complementarity (see Fig. 2B). Interestingly the hBMP2-TISIGNER sequence, supposedly codon optimised for lowered opening energy, was predicted to have a marginally increased opening energy when compared to the native sequence. This likely results from differences in the methods employed by TISIGNER and MXFold2/RNAEval when predicting structure and assigning bond energy values. All three sequences showed low SP scores, resulting in them acting as useful counterpoints to the top in silico candidates when testing the usefulness of the metric in vitro.

The final seven candidates to be taken forward to in vitro testing were subjected to additional *in silico* analysis. The amino acid sequences were re-submitted to SignalP using the slow model for an analysis of predicted SP region boundaries. The four sequences that were noticeably longer than the native hBMP2 SP (GFD-10 from *P. ranga* & the artemin orthologs) were predicted to have much longer N-regions than the other sequences (see Fig. 3B). The N-region of hBMP2

(See figure on next page.)

Fig. 1 Overview of the filtering process employed in the *in silico* pipeline. Initially ~ 35,000 TGF-β superfamily sequences were retrieved, though only ~ 19,000 were predicted to contain SPs. Of these, ~ 10,000 were found to be valid records and to have unique nucleotide sequences. Existing computational methods indicated ~ 7,000 were predicted to still function and to target the extracellular space when attached to BMP2. ~2,500 were found to have strong Kozak sequences and were taken forward to mRNA structure prediction. Sequences with the least structure at the translational start site but high global stability were thought to be preferable. The top 5 sequences from the pipeline and two manually selected alternatives were taken forward to in vitro work
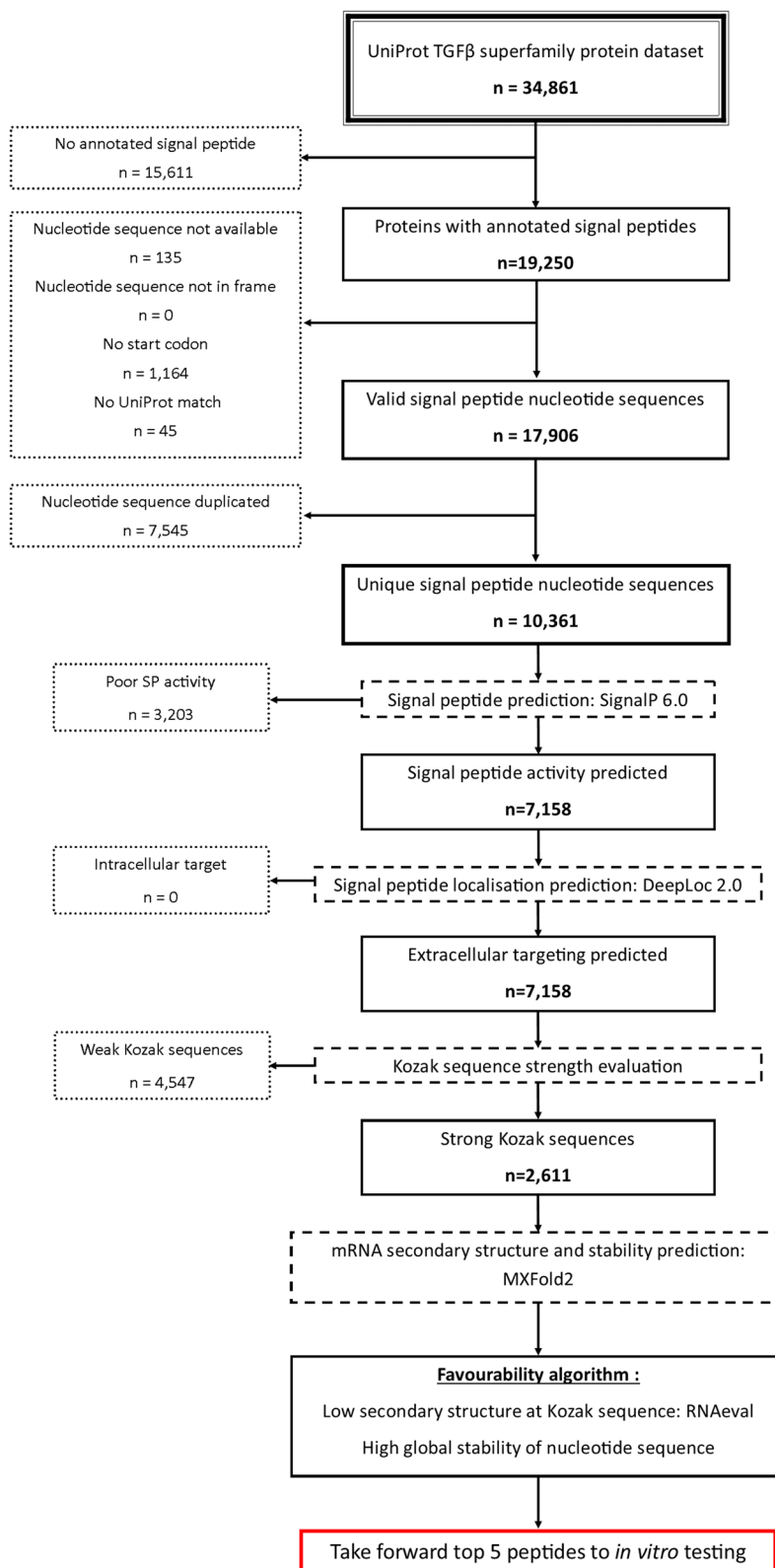
Wilkinson *et al. BMC Biotechnology* (2024) 24:34

Page 6 of 16

Fig. 1 (See legend on previous page.)

Table 1 Overview of species, protein, sequence data and calculated RNA properties for the top results from the in silico screen

Species	Protein	Abbreviated name	UniProt accession	Nucleotide database & accession	Signal peptide nucleotide sequence (5'–3')	Signal peptide amino acid sequence (N–C)	SignalP score	mRNA MFE (kcal/mol)	mRNA opening energy (kcal/mol)	SP score	SP Z-score
Vireo altiloquus	BMP 15	SP1	A0A7K5LQ98	EBI ENA: NWT20625.1	ATGGCTATGCCCTACTCTTTTGCCAGCCTCCTCCTCCTCCTCCTTGTTGTGCCCCTTTCCCAGGCT	MAMPYSFA-SLLLLLLV-VPLSQA	0.9997	-438.2	-47	9.32	10.5
Parambassis ranga	GDF 10-like	SP2	A0A6P7IG99	NCBI Nucleotide: XM_028407950.1	ATGGAATCACTTTTCAGATCCTCAGCTATGCTCAGCCGGCTTTTTTTCATCCTGTGTATATTGGTGATCCTGGAGTCCAGCTGGGCT	MESLFRSSA-MLSRLFFILCIL-VILESSWA	0.9997	-448.6	-70	6.41	6.6
Muntiacus muntjak	Artemin related	SP3	A0A5N3WKB5	EBI ENA: KAB0362191.1	ATGGAGCCTGGACGTGGAGGCCCTTCTGTGCTGCCCCTCCGGGCCGGGCCTAGGAGGCAGCAGCCTGCCCTGTGGCACACCCTGGCCGCTCTGGCCCTGCTGAGCAGCGTCGCCGAGGGT	MEPGRG-GPSVLPLRAG-PRRQQPAL-WHTLAALA-LLSSVAEG	0.5494	-488.7	-82	5.96	6.0
Phyllostomus discolor	Artemin	SP4	A0A7E6DRJ2	EBI ENA: KAF6107540.1	ATGGAGCCTGGACGTGGAGGCCCTTCTGTGCTGCCCCGCTGGCCCCTGCCTAGGCGGCAGCCTGCCCTGTGGCCAACCCTGGCTACTCTGGTTCTGCTGAGCAGTGTCGCTGAGGCC	MEPGRGGPS-VLPRWPLPRRQ-PALWPTLATLV-LLSSVAEA	0.7878	-480.1	-82	5.85	5.9

Wilkinson *et al. BMC Biotechnology*   *(2024) 24:34*

Page 8 of 16

Table 1 (continued)

Species	Protein	Abbreviated name	UniProt accession	Nucleotide database & accession	Signal peptide nucleotide sequence (5′ –3′)	Signal peptide amino acid sequence (N –C)	SignalP score	mRNA MFE (kcal/ mol)	mRNA opening energy (kcal/ mol)	SP score	SP Z-score
Molossus molossus	Artemin	SP5	A0A7J8F7B8	EBI ENA: KAF6443102.1	ATGGAGCCT GGATGTGGA GGCCCTCCT GTGCTGCTC CGCTGGACC CCACCTACG CGGCAGCCT GCCCTGTGG CCAACCCTG GCCACTCTG GTCCTGCTG AGCAGTGTC GCTGAGGCC	MEPGCGGPPV- LLRWTPPTRQ- PALWPTLATLV- LLSSVAEA	0.6976	-477.9	-82	5.83	5.8

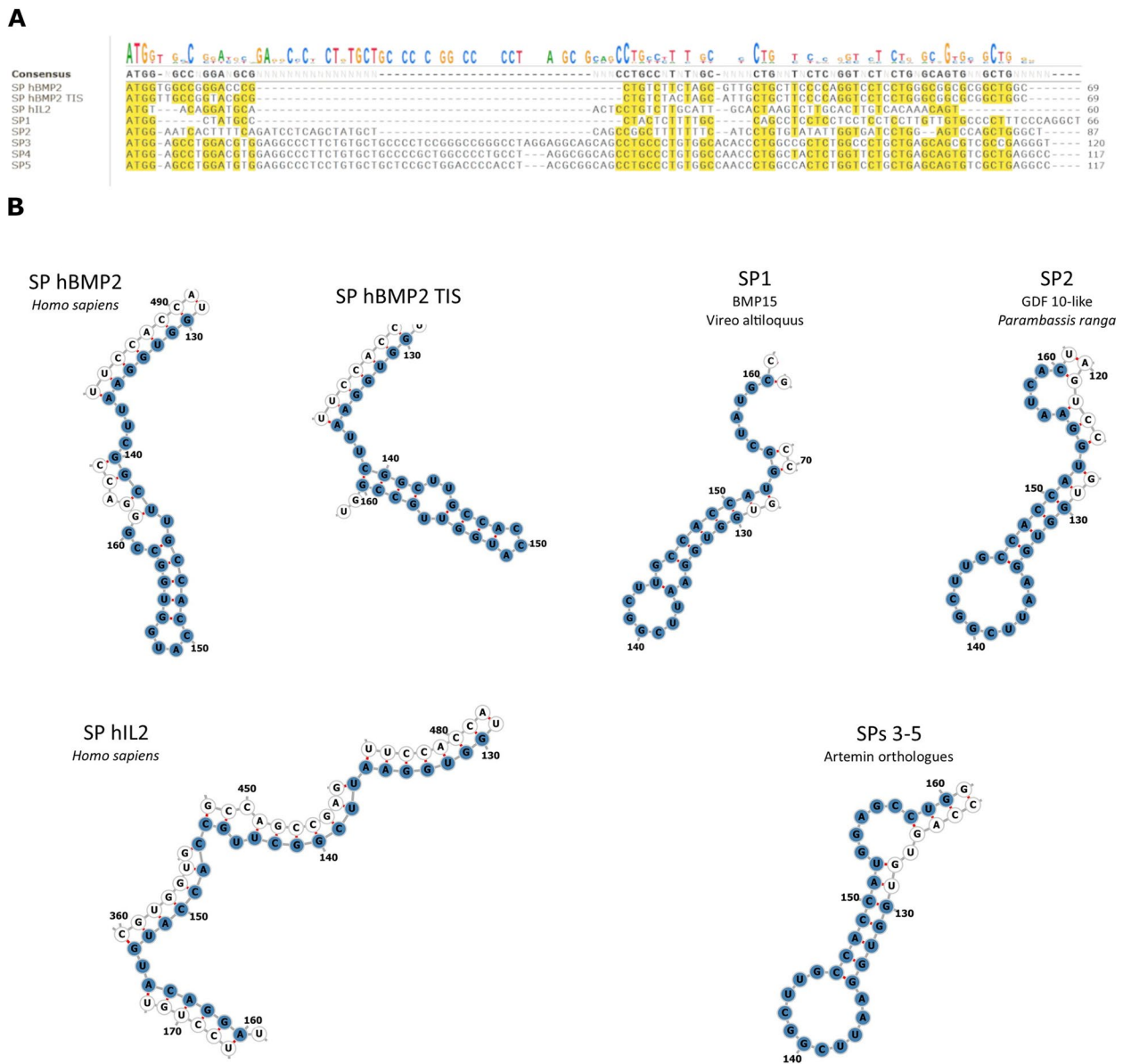Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 9 of 16

Fig. 2 Further nucleotide sequence analysis (**A**) Nucleotide sequence alignment of the final 7 SP sequences. Note the moderate variety with few positions being strongly conserved across the final set. **A** ~ 25 bp insertion can be seen in the SPs from Artemin related genes (SPs 3–5), making them noticeably longer than the other sequences. **B** Predicted mRNA secondary structures at the ribosomal attachment site. The ± 15 bp window is highlighted in blue, with start codon found at positions 151–153

is only four residues long, while that of *P. ranga* GDF-10 was 12 residues and the artemin orthologs were 21–22 residues. Additionally, the SP cleavage site for the hIl2 SP was predicted with less confidence than the other sequences and was 3 residues downstream of the true SP/preprotein sequence boundary (see Fig. 3B). The leucine motif identified in the alignment was predicted to be part of the H-region(see Fig. 3B), an established loose feature of mammalian SPs [1].

***In vitro* validation in two cell lines**

ELISA results from HEK293T conditioned media showed that the signal peptides identified by the *in silico* work and the manually selected sequences of interest did not induce a significant increase in hBMP2 secretion versus the pVax_BMP2 positive control (see Fig. 4A). The *in silico* selected signal peptides all showed small but non-significant increases versus the pVax_BMP2 positive control, while the manually selected SPs showed small
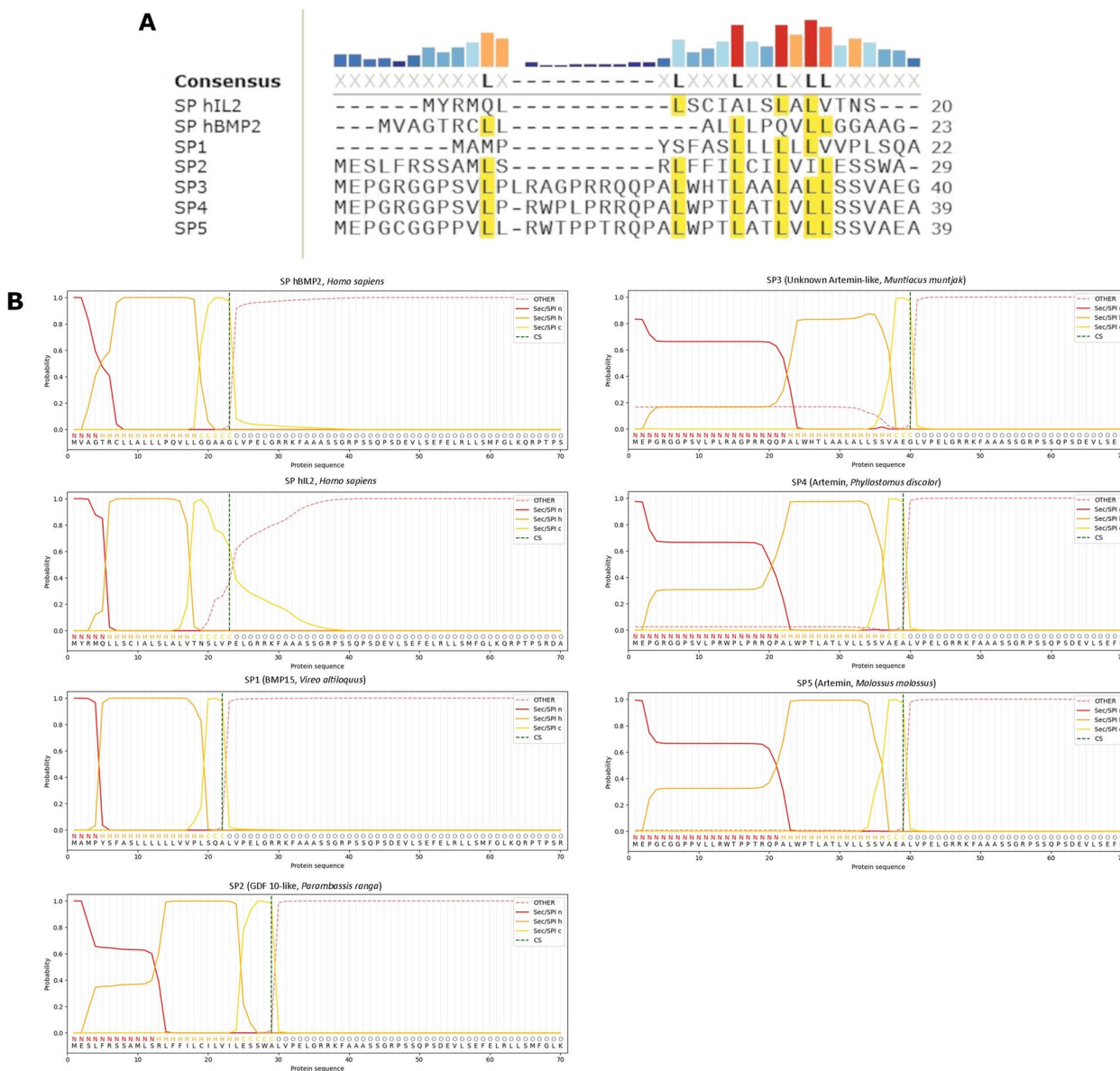
Fig. 3 Further protein sequence analysis (**A**) Protein sequence alignment of the final 7 SP sequence. Note the series of conserved leucine residues, the only conserved element across the final panel. Unsurprisingly the three SPs from Artemin related genes (SPs 3–5) were highly similar. **B** Detailed SP region boundary predictions from the SignalP long model. The Artermin related SPs displayed elongated N-regions in comparison to the rest of the set, corresponding to the position of the insertion in the nucleotide sequences (see Fig. 2A). Also note the leucine motif from the alignment is contained within the SP H-region, thought to be important for maintaining hydrophobicity and ensuring the alpha helical conformation required for transmembrane function

and non-significant decreases. Data were highly variable across all groups, with the exception of the pVax_GFP negative control. There appeared to be a minor blanking issue as small negative hBMP2 concentrations were measured for the negative control across all replicates. This was thought to be a minor issue as these values were only a small fraction of that detected in the other samples. It was suspected this was caused by slight differences in

blocking behaviour between the fresh complete media used for the standard curve and the conditioned media that made up the samples. C2C12s conditioned media ELISA results showed significant decreases versus the pVax_BMP2 positive control for SPs 1 & 2, SP hBMP2 TISIGNER and SP hIL2, while in silico SPs 3,4 & 5 showed non-significant decreases (see Fig. 4B). After 7 days ALP activity was again at background levels in

Wilkinson *et al. BMC Biotechnology*   *(2024) 24:34*

Page 11 of 16

Table 2 Overview of species, protein, sequence data and calculated RNA properties for the manually selected sequences

| Species	Protein	Abbreviated name	UniProt accession	Nucleotide database & accession	Signal peptide nucleotide sequence (5'–3')	Signal peptide amino acid sequence (N–C)	SignalP score	mRNA MFE (kcal/mol)	mRNA opening energy (kcal/mol)	SP score	SP Z-score
Homo sapiens	BMP2	SP hBMP2	P12643	NCBI nucleotide: NM_001200.4	ATGGTGGCC GGGACCCGC TGTCTTCTAGCG TTGCTGCTTCCC CAGGTCCTCCTG GGCGGCGCG GCTGGC	MVAGTRCL-LALLLPQVLLG-GAAG	0.9988	-462.3	-205	2.26	1.0
-	hBMP2 TISIGNER codon opti-mised	SP hBMP2 TIS	-	-	ATGGTTGCCGGT ACGGCGTGTCTA CTAGCATTGCTG CTTCCCCAGGTC CTCGGGC GGCGCGGCT GGC	MVAGTRCL-LALLLPQVLLG-GAAG	0.9988	-459.8	-219	2.1	0.8
Homo sapiens	IL2	SP hIL2	P60568	NCBI nucleotide: NM_000586.4	ATGTACAGGATG CAACTCCTGTCT TGCATTGCACTA AGTCTTGCACTT GTCACAAACAGT	MYRMQLLSCIAL-SLALVTNS	0.9994	-435.3	-417	1.04	-0.6
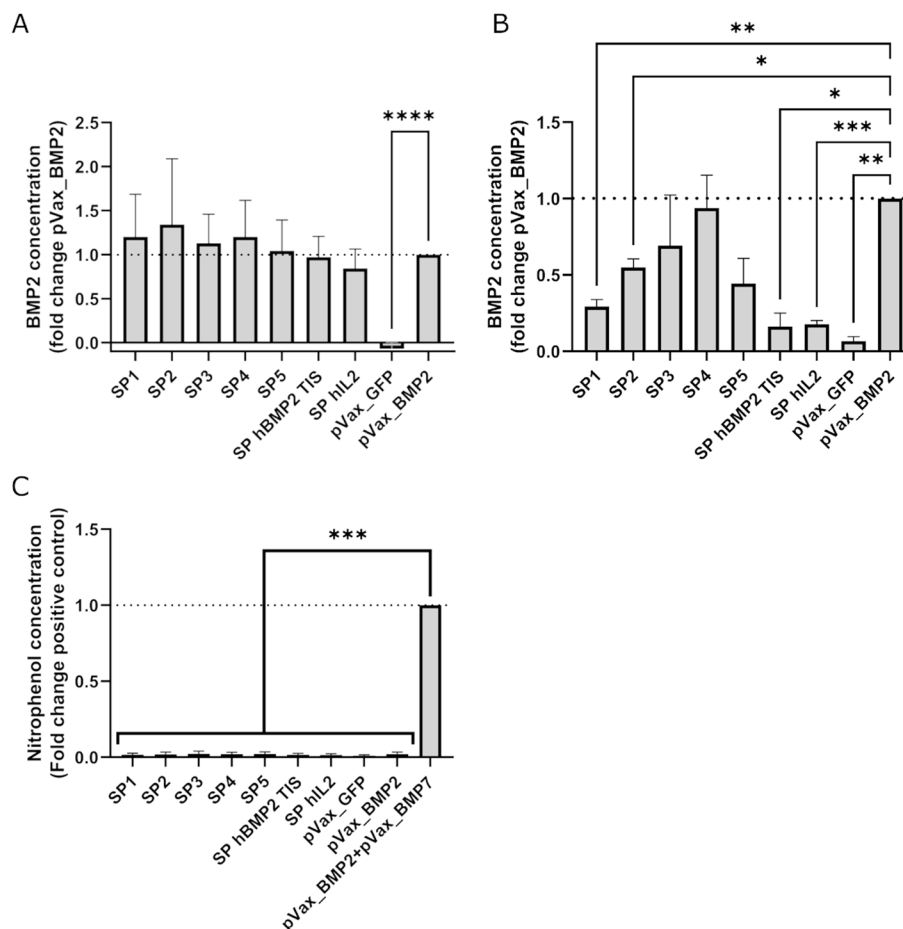
Fig. 4 In vitro validation data (A) hBMP2 ELISA data from HEK293T cells. All computationally selected sequences showed non-significant increases in secretion. $N = 7 \pm SD$, **** = $p < 0.0001$ (B) hBMP2 ELISA data from C2C12 cells. All computationally selected sequences showed decreased secretion versus the positive control, with several showing significant decreases. Particularly notable were the sequences that showed major decreases in the C2C12 results compared to the HEK293T experiment. These were SP1, SP2, SP hIL2 and SP hBMP2 TIS. $N = 3 \pm SD$. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ C) ALP assay data from C2C12 cells. All groups except the positive control showed background levels of ALP activity, indicating the novel SPs were not sufficient to induce osteogenesis in this context. $N = 3 \pm SD$, *** = $p < 0.001$

all groups except the osteogenesis positive control (see Fig. 4C).

Regression analysis of SP scores and ELISA data indicated that there was a significant correlation between SP score and BMP2 secretion in HEK293T (see Fig. 5A; gradient = 0.0509, $R^2 = 0.096$, $F_{[1, 44]} = 4.976$, $p = 0.0305$). No correlation was found between SP score and the heparin supplemented C2C12 ELISA data (see Fig. 5B).

Discussion

Here we have created an alternative approach to SP prediction, hoping to take advantage of the new generation of computational tools to create quantitative *in silico* predictions of SP effectiveness. We chose to first test the approach with human BMP2, an important osteogenic protein long used in regenerative approach in bone both as a recombinant protein and a gene

therapy. Notably BMP2 is known to be a difficult-to-express protein [34, 35], and it was hoped SP engineering might help to alleviate this issue.

While many tools now exist to predict the presence of SPs, they are not capable of making quantitative predictions of how they will perform. To add more utility to the *in silico* segment of the screen than simply filtering out likely negatives (still leaving an overwhelming number of sequences), an additional approach was required. mRNA structure prediction was used to attempt to provide this additional predictive power. Briefly, the signal peptide coding region for SRP SPs contains the start codon and lies only slightly downstream of the ribosomal attachment point. Consequently secondary structure in the region can have a pronounced influence on translation. There is evidence that less stable secondary structure in a ± 15 bp window around the translational start site
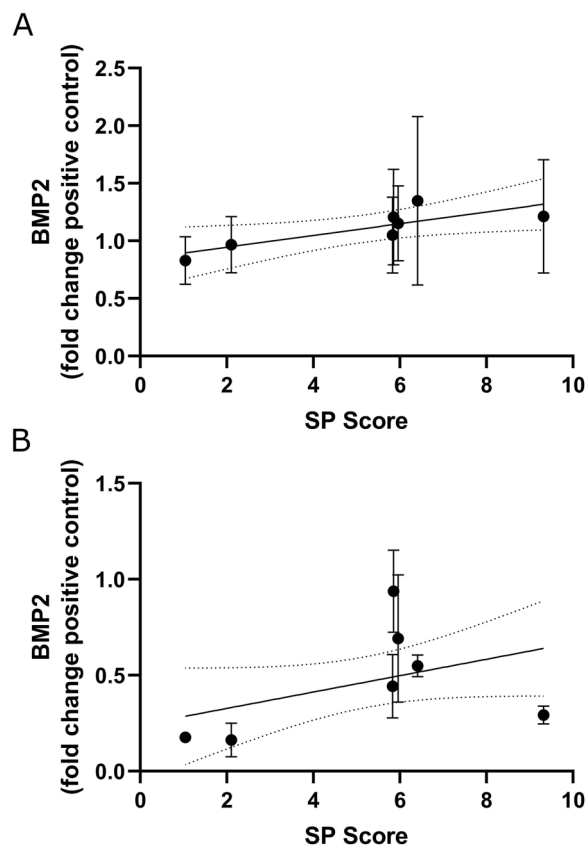
Fig. 5 Linear regression analyses of the SP score values and ELISA data (**A**) ELISA data from the HEK239T experiment. Correlation was found to be significant (gradient = 0.0509, R2 = 0.096, F [1, 44] = 4.976, *p* = 0.0305) (**B**) ELISA data from the C2C12 experiment. No significant correlation was found (gradient = 0.0427, R2 = 0.138, F [1, 19] = 3.038, *p* = 0.0975)

(which includes the beginning of the SP coding sequence) can improve translation initiation and early elongation [3–5]. Outside the translational start site mRNAs with a higher global stability appear to show improved translation, thought to be due to a combination of improved half-life and the generally higher stability of sequences containing many codons with abundant tRNAs [36, 37]. Consequently, it was though that heterologous SP-BMP2 fusions with mRNAs predicted to display less stable secondary structure at the translation start site and higher global stability might show improved translation and thus more protein export. An mRNA structure and stability based metric, SP score, was devised to assess these two properties simultaneously. It was hypothesised that SP score would correlate with levels of secreted protein.

To test the approach a small subset of the top results from the pipeline and two manually selected sequences were subjected to more detailed *in silico* analysis and tested in vitro. The top five candidates from the pipeline (termed SPs 1–5) were derived from avian, piscine

and mammalian species, with SPs drawn from BMP15, GDF10 and three ARTN related genes (see Table 1). The two manually selected sequences were a codon optimised variant of the native hBMP2 SP, and the human IL2 SP. A codon optimised hBMP2 SP allowed examination of the influence of nucleotide sequence changes divorced from changes to the protein sequence. In this case the tool TISIGNER was used to optimise the sequence for reduced opening energy [28], rather than a traditional codon adaption index (cAI) or tRNA adaption index (tAI) based approach. The hIL2 SP was chosen as it is a well-established choice in SP engineering, having been demonstrated to improve secretion for a variety of other proteins [38–42].

In vitro testing in HEK293T showed none of the tested sequences were significantly different from the positive control, and that the data were highly variable (see Fig. 4A). It was suspected that the HEK293T data variability was due to their tendency to form aggregates, leading to increased error during cell seeding which in turn influenced transfection unpredictably [43]. Despite none of the SPs inducing a significant increase in secretion, a correlation between SP score and secreted protein was found (see Fig. 5A), in agreement with the initial hypothesis. This was a promising result and suggested that the mRNA structure predictions had been accurate and had some influence on SP efficacy. Further work with a larger dataset would be required to confirm this fully.

C2C12 ELISA data showed that SPs 1 & 2 and the manually selected sequences showed a significant decreased in secretion versus the positive control, while SPs 3,4 & 5 were not significantly different (see Fig. 4B). There was no correlation between SP score and BMP2 secretion in C2C12 (see Fig. 5B). While the overall results in C2C12 did not agree with the hypothesis, they raised several interesting additional questions. The massive decrease in secretion seen in the two manually selected sequences was particularly interesting. In the case of SP hBMP2 TIS this indicated that the nucleotide sequence was vitally important in this context, as a small number of synonymous mutations almost completely obviated secretion. SP hBMP2 TIS was predicted to show very similar opening energy to the native SP sequence, indicating that either the predictions had been inaccurate or additional factors were at play. Change in codon availability from HEK293T to C2C12 was considered as a possible cause, as the lines are human and murine respectively. A cAI based analysis with ATGme [44] and the murine codon usage table (http://www.kazusa.or.jp/codon/cgi-bin/showc odon.cgi?species=10090), indicating an addition of a single additional low availability codon (< 10%) in the SP hBMP2 TIS versus the native sequence. This was thought to be unlikely to be the sole cause of the massive drop in

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 14 of 16

secretion. Of course, the possibility that the tRNA balance of C2C12 has significantly diverged from the mouse reference data cannot be ignored. An additional possible cause was interaction with antisense oligonucleotides or proteins that are found in C2C12 and not HEK293T. Further investigation would be required to identify these factors.

The SP hIL2 result was also notable as this marks a rare occurrence of the SP not offering improved performance versus a native SP, and indeed reducing it massively [38, 40–42, 39]. It was suspected that this might be related to the predicted change in SP cleavage site for the SP hIL2 fusion. It is well established that multiple SP cleavage sites can be observed using one protein and cell type [45–47], therefore perhaps cleavage at both sites was seen in both lines but the balance between the two sites (or potentially even more than two sites) was shifted. Alternatively, cleavage may have occurred at the correct site but the SP introduced a more subtle alteration to early folding. Why this might occur in C2C12 and apparently not in HEK293T was unknown. Further work using LC-MS/MS would be able to empirically establish the cleavage site/s [45–47].

ALP assay results at day 7 again showed background levels of ALP activity in all groups except the osteogenesis positive control (see Fig. 4C). This indicated that the novel SPs alone were not sufficient to induce an osteogenic effect in C2C12s. This observation was in line with previous data, where BMP2 is known to require either a high concentration or co-expression with other factors such as BMP7 to strongly induce osteogenesis [31, 48–50]. Combined with the ELISA data this was a clear indication that the SPs tested here were not sufficient to improve vector performance. Due to the lack of significant improvement in ELISA and ALP data and the small size of the in vitro data set it was decided to not perform specificity and sensitivity calculations.

There were several clear weaknesses with the approach that were imposed by current limitations in computational methods. While 5' end mRNA structure likely does play a role in how an SP influences protein secretion, it is certainly not the only factor. Interactions between the SP, SRP machinery and propeptide are known to be important [51–54], however making relevant predictions related to these interactions is difficult with existing computational tools and no established methods exist. While recent advances in multimer tertiary structure prediction by tools such as AlphaFold2 are impressive and have seen some use in modelling SP/translocon interactions [55, 56], modelling of the kinetics of the interactions of multiple multi-subunit complexes is not within current capabilities. If tools able to produce reliable predictions of these interactions are developed, they could perhaps be

combined with mRNA structure predictions to improve efficacy. The initial limitation of the sequence panel to the TGF-β superfamily dataset was an attempt to address this problem. It was hoped that SPs from these proteins would be more likely to retain their function in a new but similar context, though it is possible that this was not a reasonable assumption and the dataset was too diverse.

It should also be noted that while RNA structure prediction has seen continuous improvement and a similar burst in progress with machine learning based approaches, predictions are still frequently inaccurate in many contexts [10]. This is thought to be due to biases in the annotated RNA structure data sets used for model training, which are frequently dominated by short ncRNAs and are unable to account for all possible biomacromolecule interactions [57]. Consequently, the mRNA structure predictions used in this approach must be taken with a reasonable degree of scepticism. Advances in high through RNA structure determination techniques promise to address these problems [10, 58, 59], and the ever growing volume of training data from wider contexts promises to allow continued improvement of RNA structure modelling in future. Despite this, the significant correlation observed between SP score and secreted protein in the HEK293T data suggests the predictions were somewhat accurate, though further work would be required to conclusively establish this was not due to chance. Future work could employ a library of synonymous variants to separate the influence of nucleotide sequence from protein structure.

Conclusions

The mRNA structure-based method for SP effectiveness prediction described here was capable of identifying previously untested SPs capable of functioning comparably to the native sequence in HEK293T. Additionally there was a significant correlation between model predictions and secreted protein in HEK293T, though none of the SPs showed a significant improvement in secretion versus the native SP. The approach was not effective in C2C12 cells with several SPs inducing a significant decrease in secretion. Particularly poor performance from the codon optimised SP hBMP2 TIS indicated the importance of synonymous mutations in the SP and merits further study. These results suggest the mRNA structure prediction approach requires further improvement before it can produce significant improvements in this context, and ideally should be combined with protein sequence based predictions of SP activity in future.

Abbreviations
4NP   4-Nitrophenol
ALP   Alkaline Phosphetase
BMP2   Bone Morphogenetic Protein 2

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 15 of 16

MFE	Minimum Free Energy
NPP	Nitrophenol Phosphate
rhBMP2	Recombinant human Bone Morphogenetic Protein 2
Sec/SPI	Sec substrates cleaved by SPase I
SP	Signal Peptide
SRP	Signal Recognition Particle
TISIGNER	Translation initiation coding region designer

Availability of data and materials
The datasets used and analysed during the study were publicly available and access links are provided in the main text. Full results from the in silico screen can be provided on request.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Consent for publication
Not applicable.

Author details
[1]Department of Mechanical Engineering, Institute of Medical and Biological Engineering, University of Leeds, Leeds, UK. [2]Faculty of Biological Sciences, University of Leeds, Leeds, UK. [3]School of Biomedical Sciences, Faculty of Biological Sciences, University of Leeds, Leeds, UK. [4]Oral Biology, Faculty of Medicine and Health, University of Leeds, Leeds, UK.

References
1. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: structure, roles, and applications. Eur J Cell Biol. 2018;97(6):422–41.
2. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 2014;505(7485):696–700.
3. Corley M, Solem A, Phillips G, Lackey L, Ziehr B, Vincent HA, et al. An RNA structure-mediated, posttranscriptional model of human α-1-antitrypsin expression. Proc Natl Acad Sci. 2017;114(47):E10244–53.
4. Mustoe AM, Busan S, Rice GM, Hajdin CE, Peterson BK, Ruda VM, et al. Pervasive Regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. Cell. 2018;173(1):181–e19518.
5. Mustoe AM, Corley M, Laederach A, Weeks KM. Messenger RNA structure regulates translation initiation: a mechanism exploited from Bacteria to humans. Biochemistry. 2018;57(26):3537–9.
6. Xiang Y, Huang W, Tan L, Chen T, He Y, Irving PS, et al. Pervasive downstream RNA hairpins dynamically dictate start-codon selection. Nature. 2023;621(7978):423–30.
7. Wang Y, Mao Y, Xu X, Tao S, Chen H. Codon Usage in Signal Sequences affects protein expression and secretion using Baculovirus/Insect cell expression system. PLoS ONE. 2015;10(12):e0145887–0145887.
8. Cheng Y, Liu S, Lu C, Wu Q, Li S, Fu H, et al. Missense mutations in the signal peptide of the porcine GH gene affect cellular synthesis and secretion. Pituitary. 2016;19(4):362–9.
9. Xu Y, Verma D, Sheridan RP, Liaw A, Ma J, Marshall NM, et al. Deep Dive into Machine Learning models for Protein Engineering. J Chem Inf Model. 2020;60(6):2773–90.
10. Zhang J, Fei Y, Sun L, Zhang QC. Advances and opportunities in RNA structure experimental determination and computational modeling. Nat Methods. 2022;19(10):1193–207.
11. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5.
12. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 2017;33(21):3387–95.
13. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12(1):941.
14. Katagiri T, Watabe T. Bone morphogenetic proteins. Cold Spring Harb Perspect Biol. 2016;8(6).
15. Bostrom MPG. Expression of Bone Morphogenetic Proteins in Fracture Healing. Clin Orthop Relat Res [Internet]. 1998;355. https://journals.lww.com/clinorthop/fulltext/1998/10001/expression_of_bone_morphogenetic_proteins_in.13.aspx.
16. Huntley R, Jensen E, Gopalakrishnan R, Mansky KC. Bone morphogenetic proteins: their role in regulating osteoclast differentiation. Bone Rep. 2019;10:100207.
17. Sánchez-Duffhues G, Hiepen C, Knaus P, ten Dijke P. Bone morphogenetic protein signaling in bone homeostasis. Muscle Bone Interact. 2015;80:43–59.
18. Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, et al. Bone morphogenetic protein (BMP) signaling in development and human diseases. Genes Dis. 2014;1(1):87–105.
19. Marsell R, Einhorn TA. The role of endogenous bone morphogenetic proteins in normal skeletal repair. Injury. 2009;40(Suppl 3):S4–7.
20. Khan SN, Lane JM. The use of recombinant human bone morphogenetic protein-2 (rhBMP-2) in orthopaedic applications. Expert Opin Biol Ther. 2004;4(5):741–8.
21. Schlundt C, Bucher CH, Tsitsilonis S, Schell H, Duda GN, Schmidt-Bleek K. Clinical and Research approaches to treat non-union fracture. Curr Osteoporos Rep. 2018;16(2):155–68.
22. Hacobian AR, Posa-Markaryan K, Sperger S, Stainer M, Hercher D, Feichtinger GA, et al. Improved osteogenic vector for non-viral gene therapy. Eur Cell Mater. 2016;31:191–204.
23. UniProt. The universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480–9.
24. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2007;35(suppl1):D5–12.
25. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988–95.
26. Cummins C, Ahamed A, Aslam R, Burgin J, Devraj R, Edbali O, et al. The European Nucleotide Archive in 2021. Nucleic Acids Res. 2022;50(D1):D106–10.
27. Davis P, Zarowiecki M, Arnaboldi V, Becerra A, Cain S, Chan J, et al. WormBase in 2022—data, processes, and tools for analyzing Caenorhabditis elegans. Genetics. 2022;220(4):iyac003.
28. Bhandari BK, Lim CS, Gardner PP. TISIGNER.com: web services for improving recombinant protein production. Nucleic Acids Res. 2021;49(W1):W654–61.
29. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2 0 Algorithms Mol Biol. 2011;6(1):26–26.
30. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. Bioinforma Oxf Engl. 2015;31(20):3377–9.

Wilkinson *et al. BMC Biotechnology*   (2024) 24:34

Page 16 of 16

31. Feichtinger GA, Hacobian A, Hofmann AT, Wassermann K, Zimmermann A, van Griensven M, et al. Constitutive and inducible co-expression systems for non-viral osteoinductive gene therapy. Eur Cell Mater. 2014;27:166–84. discussion 184.

32. Zhao B, Katagiri T, Toyoda H, Takada T, Yanai T, Fukuda T, et al. Heparin potentiates the in vivo ectopic bone formation induced by bone morphogenetic protein-2. J Biol Chem. 2006;281(32):23246–53.

33. Kim MG, Kim CL, Kim YS, Jang JW, Lee GM. Selective endocytosis of recombinant human BMPs through cell surface heparan sulfate proteoglycans in CHO cells: BMP-2 and BMP-7. Sci Rep. 2021;11(1):3378.

34. Jérôme V, Thoring L, Salzig D, Kubick S, Freitag R. Comparison of cell-based versus cell-free mammalian systems for the production of a recombinant human bone morphogenic growth factor. Eng Life Sci. 2017;17(10):1097–107.

35. Riedl SAB, Jérôme V, Freitag R. Repeated transient transfection: an alternative for the recombinant production of difficult-to-express proteins like BMP2. Processes. 2022;10(6).

36. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015;160(6):1111–24.

37. Mauger DM, Cabral BJ, Presnyak V, Su SV, Reid DW, Goodman B et al. mRNA structure regulates protein expression through changes in functional half-life. Proc Natl Acad Sci. 2019;116(48):24075 LP – 24083.

38. Bamford RN, DeFilippis AP, Azimi N, Kurys G, Waldmann TA. The 5´ untranslated region, Signal Peptide, and the Coding sequence of the Carboxyl Terminus of IL-15 participate in its multifaceted translational Control1. J Immunol. 1998;160(9):4418–26.

39. Zhang L, Leng Q, Mixson AJ. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J Gene Med. 2005;7(3):354–65.

40. Loomis RJ, DiPiazza AT, Falcone S, Ruckwardt TJ, Morabito KM, Abiona OM et al. Chimeric Fusion (F) and Attachment (G) Glycoprotein Antigen Delivery by mRNA as a Candidate Nipah Vaccine. Front Immunol [Internet]. 2021;12. https://www.frontiersin.org/articles/https://doi.org/10.3389/fimmu.2021.772864.

41. Billerhart M, Schönhofer M, Schueffl H, Polzer W, Pichler J, Decker S, et al. CD47-targeted cancer immunogene therapy: secreted SIRPα-Fc fusion protein eradicates tumors by macrophage and NK cell activation. Mol Ther Oncolytics. 2021;23:192–204.

42. Egan KP, Awasthi S, Tebaldi G, Hook LM, Naughton AM, Fowler BT et al. A trivalent HSV-2 gC2, gD2, gE2 nucleoside-modified mRNA-LNP vaccine provides outstanding Protection in mice against genital and non-genital HSV-1 infection, comparable to the same antigens derived from HSV-1. Viruses. 2023;15(7).

43. Sticky Issues with 293 Cells [Internet]. Culture Collections; [cited 2023 Oct 25]. https://www.culturecollections.org.uk/news/ecacc-news/sticky-issues-with-293-cells.aspx.

44. Daniel E, Onwukwe GU, Wierenga RK, Quaggin SE, Vainio SJ, Krause M, ATGme. Open-source web application for rare codon identification and custom DNA sequence optimization. BMC Bioinformatics. 2015;16(1):303.

45. Park JH, Lee HM, Jin EJ, Lee EJ, Kang YJ, Kim S, et al. Development of an in vitro screening system for synthetic signal peptide in mammalian cell-based protein production. Appl Microbiol Biotechnol. 2022;106(9–10):3571–82.

46. Huang Y, Fu J, Ludwig R, Tao L, Bongers J, Ma L, et al. Identification and quantification of signal peptide variants in an IgG1 monoclonal antibody produced in mammalian cell lines. J Chromatogr B Analyt Technol Biomed Life Sci. 2017;1068–1069:193–200.

47. Haryadi R, Ho S, Kok YJ, Pu HX, Zheng L, Pereira NA, et al. Optimization of Heavy Chain and Light Chain Signal Peptides for high level expression of therapeutic antibodies in CHO cells. PLoS ONE. 2015;10(2):e0116878–0116878.

48. Kawai M, Bessho K, Maruyama H, Miyazaki J ichi, Yamamoto T. Simultaneous gene transfer of bone morphogenetic protein (BMP) -2 and BMP-7 by in vivo electroporation induces rapid bone formation and BMP-4 expression. BMC Musculoskelet Disord. 2006;7:62.

49. Feichtinger GA, Hofmann AT, Slezak P, Schuetzenberger S, Kaipel M, Schwartz E et al. Sonoporation increases therapeutic efficacy of inducible and constitutive BMP2/7 in vivo gene delivery. Hum Gene Ther Methods. 2013/11/27 ed. 2014;25(1):57–71.

50. Kaito T, Morimoto T, Mori Y, Kanayama S, Makino T, Takenaka S, et al. BMP-2/7 heterodimer strongly induces bone regeneration in the absence of increased soft tissue inflammation. Spine J off J North Am Spine Soc. 2018;18(1):139–46.

51. Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W. Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nat Struct Mol Biol. 2008;15(5):494–9.

52. Zhang X, Schaffitzel C, Ban N, Shan S. ou. Multiple conformational switches in a GTPase complex control co-translational protein targeting. Proc Natl Acad Sci. 2009;106(6):1754–9.

53. Jungnickel B, Rapoport TA. A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. Cell. 1995;82(2):261–70.

54. Zhang X, Rashid R, Wang K, Shan S. Ou. Sequential checkpoints govern substrate selection during cotranslational protein targeting. Science. 2010;328(5979):757–60.

55. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun. 2022;13(1):1265.

56. Gutierrez Guarnizo SA, Kellogg MK, Miller SC, Tikhonova EB, Karamysheva ZN, Karamyshev AL. Pathogenic signal peptide variants in the human genome. NAR Genomics Bioinforma. 2023;5(4):lqad093.

57. Danaee P, Rouches M, Wiley M, Deng D, Huang L, Hendrix D. bpRNA: large-scale automated annotation and analysis of RNA secondary structure. Nucleic Acids Res. 2018;46(11):5381–94.

58. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods. 2014;11(9):959–65.

59. Smola MJ, Weeks KM. In-cell RNA structure probing with SHAPE-MaP. Nat Protoc. 2018;13(6):1181–95.

Publisher's Note