somalogic™

Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay

SOMAmer reagents bind to their target proteins in a highly specific manner with affinities comparable to and often exceeding those of antibodies. This technical note summarizes some of the existing evidence for the exquisite specificity of SOMAmer reagents.

SOMAmer® reagents: Two elements of specificity in a single protein affinity reagent

Each SOMAmer reagent is characterized by two elements of specificity: strong affinity for its target protein and a slow dissociation rate. These two specificity attributes derive from the unique chemical composition of each reagent.

Since its discovery in 1990, SELEX (Systematic Evolution of Ligands by EXponential enrichment) has proven a powerful tool for identifying nucleic acid-based ligands (aptamers) to a wide range of molecular targets.[1-3]

Nevertheless, identifying aptamers to certain molecular targets, including some protein targets, has remained difficult, in part because the chemical diversity of nucleic acids is more limited than that of proteins.[4]

To overcome this diversity gap, we have developed a versatile method for introducing functional groups that are absent in natural nucleic acid libraries, but found in protein-protein, or small molecule ligand-protein interactions.[5] Such groups, which we add at the 5-position of deoxyuridine, do not disrupt base pairing and are therefore compatible with the enzymatic steps of SELEX. The addition of these modified bases greatly expands the diversity of protein targets for which we can successfully identify a SOMAmer reagent, and by adapting the selection conditions, we can simultaneously identify ligands with slow dissociation rates.

We have solved high-resolution X-ray crystal structures of several SOMAmer:protein complexes, which reveal the major role for the modified side chains of SOMAmer reagents both in SOMAmer folding as well as in creating the binding interface with their target proteins. Three representative structures are shown in Figure 1.

There are several features worth noting. First, there is extensive shape complementarity between the SOMAmer reagents and protein surfaces. The interaction surfaces are relatively large (1097 to 1248 Å^2) in the three examples.[5-8] Superimposed on shape complementarity is functional group complementarity between the interacting surfaces: Polar groups interact with polar groups and hydrophobic side chains of SOMAmer reagents occupy hydrophobic pockets on proteins.

In considering the role of the hydrophobic side chains, it is worth noting that all eight on the PDGF-BB SOMAmer reagent and eight out of ten on the IL-6 SOMAmer reagent make direct contacts with the proteins. It is therefore not surprising that the SOMAmer:protein interface is considerably more hydrophobic than the conventional aptamer:protein interface.

This feature greatly expands the range of epitopes that SOMAmer reagents can bind and allows for identification of ligands that individually bind to a larger fraction of the protein surface.

Affinity for protein targets

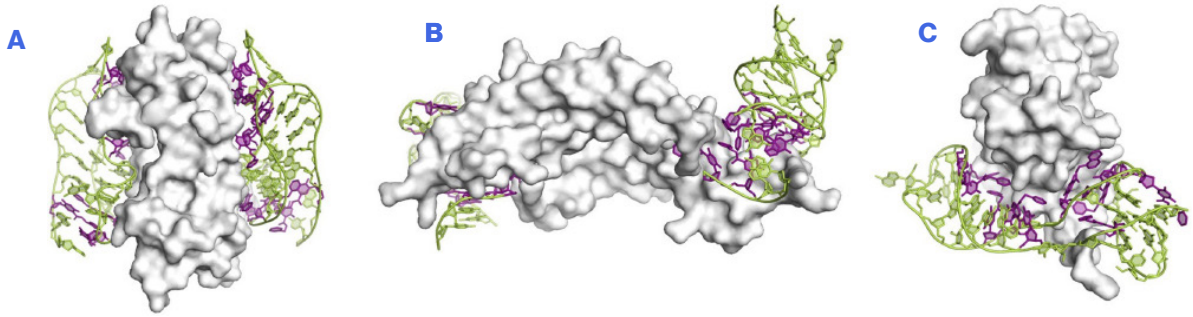


FIGURE 1 X-ray crystal structures of SOMAmer reagents bound to (A) IL-6[6], (B) PDGF-BB[7], and (C) β-NGF[8]. Modified side chains are shown in purple, and the DNA backbone and unmodified bases are in green.

SL00000417 Rev 5: 2022-03
Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay

Slow dissociation rates

Biological samples are highly complex mixtures of proteins and other macromolecules and their metabolites, a subset of which bind nucleic acids or other polyanions such as heparin. Because SOMAmer reagents are comprised of DNA, polyanion-binding proteins will have a tendency to bind all SOMAmer reagents weakly, in a sequence-independent manner.

Non-specific interactions have a strong ionic component, with fast association and dissociation kinetics, so they can be minimized by using competitor polyanions in SOMAmer-based assays. This takes advantage of the slow dissociation rate of specific, high- affinity SOMAmer-protein complexes (generally 60-90 minutes vs. seconds for non-specific interactions).

Examples of SOMAmer specificity

The high degree of specificity that can be achieved by individual SOMAmer reagents in different assays is highlighted in the examples below (other examples available on request).

Growth differentiation factors 8 and 11

Reagents with the ability to distinguish between growth differentiation factor-8 (GDF-8) and GDF-11 are in increasingly high demand. Our original GDF-11-selected SOMAmer reagent binds to both proteins with comparable affinity, which is not surprising given the 90% amino acid identity between the two proteins. To identify specific reagents for GDF-11 and GDF-8, we took advantage of the unique opportunities *in vitro* selection provides to obtain highly selective reagents.

The results include several reagents that bind GDF-11 with very high affinity and show minimal, if any, binding to GDF-8, and vice versa. An example of one of the reagents is shown in Figure 2.

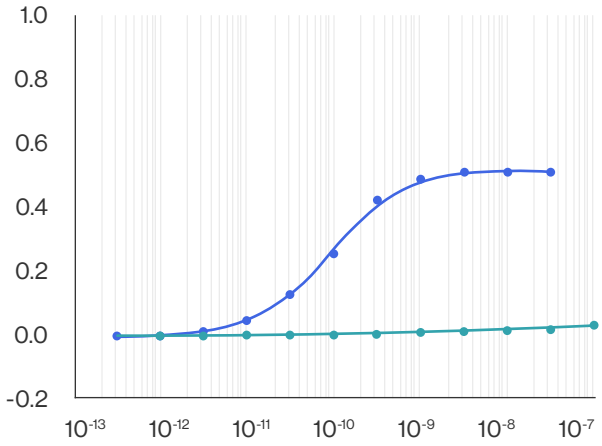


FIGURE 2 Affinity curves for binding of a GDF-11 specific SOMAmer reagent to GDF-11 (blue) and to GDF-8 (green). Protein concentration is shown in units of molarity (M) on the x-axis and the fraction of reagent bound, as measured by radioactivity, is shown on the y-axis.[9]

Single-nucleotide polymorphisms and SOMAmer specificity

Certain single-nucleotide polymorphisms (SNPs) result in a change in the amino acid sequence for the encoded protein. We have observed with the SomaScan® Assay many instances where such a single amino acid change in a protein sequence dramatically affects analyte measurements. The gene for the protein low affinity immunoglobulin gamma Fc region receptor II-a (FCG2A) can contain a common SNP that changes a histidine at position 167 in the protein sequence to an arginine.

In fact, the FCG2A protein used for in vitro selection contained an arginine at position 167. We observed in a SomaScan Assay analysis of healthy individuals that FCG2A signaled ~4,000 – 10,000 Relative Fluorescence Units (RFU) for 2/3 of the individuals, while ~1/3 of the population had a signal ~200 RFU, presumably corresponding to baseline in plasma.

To confirm, purified proteins with either histidine-167 or arginine-167 were used to examine SOMAmer binding. The SOMAmer reagent affinity for the H167R variant is two orders of magnitude higher than that for the wild-type protein, illustrating a striking specificity among two proteins with a single amino acid substitution.

Specificity in cell labeling

SOMAmer reagents can conceivably be used in any assay in which antibodies are traditionally used to detect the presence, quantity, or subcellular localization of a particular protein, including fluorescent cell microscopy. We compared the fluorescence intensity and staining pattern of SOMAmer reagents used on three well-characterized breast adenocarcinoma cell lines known to express very high or very low levels of the receptors ERBB1 or ERBB2, which share 50% amino acid sequence identity (Figure 3).

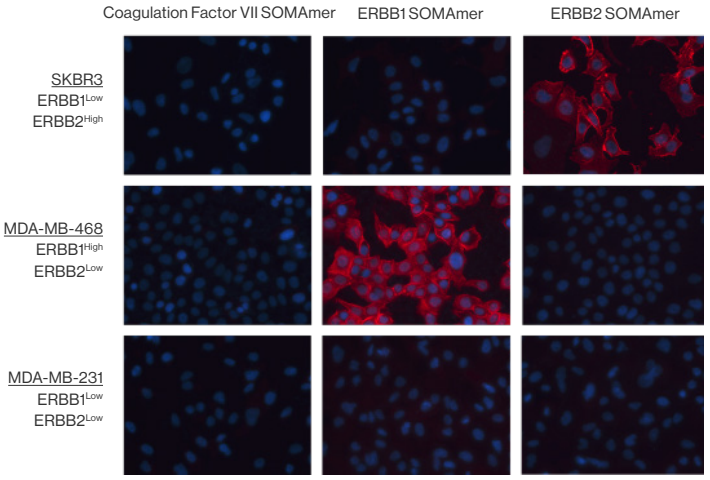


FIGURE 3 Fluorescent cell microscopy using SOMAmer reagents as primary detection reagents. ERBB1-specific staining is limited to the ERBB1^{High} cell line MDA-MB-468. ERBB2-specific staining is limited to the ERBB2^{High} cell line SKBR-3. The ERBB1^{Low}, ERBB2^{Low} cell line MDA-MB-231 is not stained by either SOMAmer reagent. The control reagent against coagulation factor VII does not stain SKBR-3, MDA-MB-468, or MDA-MB-231 cell lines.

All SOMAmer reagents contained a 5′ fluorophore for detection purposes. A significant fluorescent staining pattern was only observed with the ERBB1-specific reagent on the ERBB1^{High} cell line and with the ERBB2-specific reagent on the ERBB2^{High} cell line. A control reagent against coagulation factor VII did not label any of the three tested cell lines. The results indicate that the ERBB1 and ERBB2-specific SOMAmer reagents are indeed specific for their target receptor, even in the presence of a highly expressed, related receptor, and that SOMAmer reagents, in general, do not non-specifically label cells in the presence of polyanionic competitor.

Direct assays to assess specificity

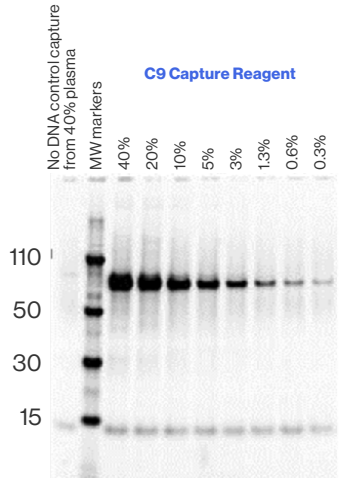
Ultimately, the best way to assess specificity of an analyte-binding reagent within a biological sample is to capture the analyte using the reagent and then use peptide mass-fingerprinting to determine the identity of the protein(s) captured. However, this approach is only possible for the highest abundance analytes.

Figure 4 shows an example of a high abundance protein for which specific target capture by its SOMAmer reagent has been confirmed with peptide mass-fingerprinting. In the figure, total protein captured out of plasma diluted between 40% and 0.3% is shown for the complement component 9 (C9). There was no depletion or other pretreatment of the plasma. Complement C9 is expected to be approximately 1 µM in 100% plasma and the SOMAmer reagent affinity is 60 pM.

Specificity in the SomaScan Assay

The highly multiplexed SomaScan Assay takes advantage of both of the specificity attributes of SOMAmer reagents described above (high affinity and slow dissociation rates for cognate proteins) and includes assay steps to maximize that specificity. For more detailed information on the SomaScan Assay, please consult the SomaScan Assay Technical Note (available at www.somalogic.com).

Briefly described, the biological sample is incubated with a mixture of SOMAmer reagents pre-immobilized on beads through a 5′ biotin group (each SOMAmer reagent also contains a photocleavable group and a fluorescent tag at its 5′ end).



IPI00022395 (100%), 63, 174.9 Da
C9 Complement component C9
37 exclusive unique peptides, 54 exclusive unique spectra,
78 total spectra, 322/559 amino acids (58% coverage)

FIGURE 4 (Left) SDS-PAGE with Alexa-647-labeled total proteins captured by the complement component C9 SOMAmer reagent at various plasma dilutions. The small band just below 15 kDa in each capture lane is streptavidin monomer that dissociated from the beads used in the assay. **(Right)** Sequence coverage from peptide mass fingerprinting verifying that C9 is captured from plasma after enrichment using the SOMAmer reagent selected using C9. Yellow/bold indicates peptides found in the mass spectrometry data and green are potential modifications in the identified spectra (typical in MS analyses).

somalogic

SL00000417 Rev 5: 2022-03
Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay

SOMAmer-protein complexes are formed during an incubation step that establishes conditions that promote complex formation approaching equilibrium, which we refer to as "catch-1." The beads are then washed to remove unbound proteins followed by biotin labeling of all bead-associated proteins.

The protein:SOMAmer complexes are then released from the beads back into solution by photocleavage with UV light, diluted and challenged with a polyanionic competitor. It is in this step that a second dimension of specificity is achieved through what we term the "kinetic challenge."

Because dissociation rates of cognate SOMAmer-protein interactions are much slower than those of non-specific interactions, a polyanionic competitor, present in excess, rapidly occupies binding sites freed by the dissociated non-cognate complexes and prevents their rebinding.

One of the reasons a high degree of multiplexing is achievable in the SomaScan Assay is the fact that, since all SOMAmer reagents are polyanions, there exists a common non-denaturing competitor capable of competing for virtually all non-specific protein interactions.

Following the kinetic challenge, the complexes are re-captured on a second set of beads via biotin on the proteins ("catch-2"), followed by additional washing to remove unbound SOMAmer reagents.

Finally, the SOMAmer reagents that remain associated with beads are eluted by denaturation and hybridized to complementary probes printed on a DNA microarray. In this manner, protein concentrations in the original sample are quantitatively converted into addressable fluorescence signals on the array.

Based on pull-down experiments that mimic conditions of our assay, we have observed a high degree of specificity with fluorescence signals reflecting cognate SOMAmer-protein interactions, even at a high protein multiplicity. Aside from the intrinsic specificity of SOMAmer reagents and the off-rate differential between specific and non-specific interactions, the use of two separate bead capture steps contributes to overall specificity. For more details on the SomaScan Assay two-catch approach, please see the SomaScan Assay Technical Note (available at www.somalogic.com).

Even with this level of specificity, individual analyte measurements in any given biological sample can be affected in an idiosyncratic manner by various components of that sample, usually referred to as "matrix effects." To monitor such effects, we have developed a group of nucleic acid sequences that have similar composition to SOMAmer reagents but have no known specific molecular target. We include these sequences (called "spuriomers") in the assay as an index of non-specific matrix effects.

Figure 5 shows the matrix effects for different kinds of biological samples. Matrix effects appear to be sequence dependent; some matrices increase signals on certain control sequences more than others. Most importantly, detection of signals in the SomaScan Assay over and above those typical of non-specific interactions revealed by the spuriomers allows the discovery of true biomarkers in a given matrix.

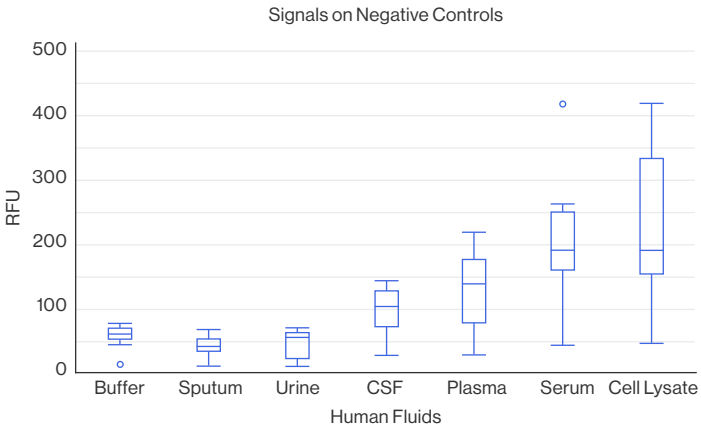


FIGURE 5 Signals on negative controls for a variety of human matrices. The signals are presented as box and whisker plots where the median is designated by the horizontal line in the box that represents the center half of the data. Specific signals are typically well above 1,000 RFUs.

somalogic

SL00000417 Rev 5: 2022-03

Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay

Specificity verification from genetic associations

The SomaScan Assay enables broad assessment of the effects that natural sequence variation in the human genome has on circulating protein levels. Linking proteins to risk-associated genetic variants can expose causal factors in disease and reveal new drug targets.

A genetic variant that modulates protein abundance is called a protein quantitative trait locus (pQTL), and a variant located within or near the gene that encodes the measured protein is called a cis-pQTL. Discovery of a cis-pQTL provides strong orthogonal evidence that the SOMAmer reagent in the SomaScan Assay is accurately and robustly measuring its target protein.

In 2017, Suhre et al. published the first proteogenomic study using the SomaScan Assay to quantify 1,124 plasma proteins in 1,000 individuals and identified 539 associations between 284 proteins, including 384 cis-pQTLs.[10] In 2018, Sun et al. conducted a larger study using the SomaScan Assay to quantify 3,622 plasma proteins in 3,301 participants and identified 1,927 associations between 1,478 proteins and 764 genomic regions, including 549 cis-pQTLs.[11] Emilsson et al. later published a study using the SomaScan Assay to quantify 4,137 serum proteins in 5,457 individuals.[12] They identified 1,046 cis-pQTLs and confirmed 80% of all previously reported cis-pQTLs.

In 2021, Pietzner et al. published a large study using the SomaScan Assay that measured 4,775 unique proteins in plasma samples from 10,708 individuals and identified 10,674 pQTLs for 3,892 distinct proteins, of which 1,569 (60%) were cis-pQTLs.[13] They linked the cis-pQTLs to a large range of traits and diseases revealing 1,859 gene-protein-disease connections for 412 unique proteins. Connecting cis-pQTLs to disease endpoints provides the strongest possible evidence that the protein target is causal in the disease.

Also in 2021, Ferkingstad et al. published the largest proteogenomic study to date.[14] They used the SomaScan Assay to measure the levels of 4,719 proteins in the plasma of 35,559 Icelanders and mapped them to 27.2 million sequence variants. They discovered 18,084 pQTLs involving 4,631 proteins (94% of the proteins measured), including 1,881 cis-pQTLs. This study replicated 83% of the pQTLs reported in Sun et al. The levels of proteins in plasma were tested for associations with 373 diseases and other traits and yielded 257,490 associations. Twelve percent of roughly 50,000 DNA variants associated with diseases also associated with protein levels.

Table 1 summarizes the pQTL associations identified by these five studies. Overall, significant cis-pQTL associations have been identified for 2,365 SOMAmer reagents.

Study	No. of samples in study	No. of SOMAmer reagents in SomaScan Assay	No. of unique proteins measured	No. of pQTLs reported	No. of cis-pQTLs reported
Suhre et al.[10]	1,000	1,124	1,092	539	384
Sun et al.[11]	3,301	3,622	2,994	1,927	549
Emilsson et al.[12]	5,457	4,783	4,137	3,134	1,046
Pietzner et al.[13]	10,708	4,979	4,775	10,674	1,569
Ferkingstad et al.[14]	35,559	4,907	4,719	18,084	1,881

TABLE 1 Summary of pQTL associations for five proteogenomic studies using the SomaScan Assay.

somalogic™

SL00000417 Rev 5: 2022-03

Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay

Conclusion

We have developed a new class of analyte-specific reagents that we have named SOMAmer reagents, or Slow Off-rate Modified Aptamers, a name that reflects both the chemical structure of these reagents as well as the kinetic stability of the complexes they form with proteins. SOMAmer reagents have a wide range of potential applications across the life sciences, from basic discovery through clinical diagnostics, and overcome the inherent limitations of both antibodies (multiplex size limits) and traditional aptamers (reduced range of protein targets and generally short off-rates) to create a "next generation" of affinity reagents. For more information, visit www.somalogic.com.

References

1. Tuerk C & Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505-510.
2. Ellington AD & Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287):818-822.
3. Ozer A, Pagano JM, & Lis JT (2014) New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization. *Mol. Ther. Nucleic Acids* 3:e183.
4. Gold L, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 5(12):e15004.
5. Vaught JD, et al. (2010) Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc*. 132: 4141-4151.
6. Gelinas AD, et al. (2014) Crystal structure of interleukin-6 in complex with a modified nucleic acid ligand. *J. Biol. Chem.* 289(12):8720-8734.
7. Davies DR, et al. (2012) Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. USA* 109(49):19971-19976.
8. Jarvis TC, et al. (2015) Non-helical DNA Triplex Forms a Unique Aptamer Scaffold for High Affinity Recognition of Nerve Growth Factor. *Structure* 23(7):1293-1304.
9. Ochsner UA, et al. (2019) Targeting unique epitopes on highly similar proteins GDF-11 and GDF-8 with modified DNA aptamers. *Biochemistry* 58(46):4632-4640.
10. Suhre, K et al. (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun*. 8:14357.
11. Sun, BB et al. (2018) Genomic atlas of the human plasma proteome. *Nature* 558:73-79.
12. Emilsson, V et al. (2018) Co-regulatory networks of human serum proteins link genetics to disease. *Science* 361(6404):769-773.
13. Pietzner, M et al. (2021) Mapping the proteo-genomic convergence of human diseases. *Science*:eabj1541.
14. Ferkingstad, E et al. (2021) Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet*. 53(12):1712-1721.

somalogic

SL00000417 Rev 5: 2022-03

Specificity Elements of SOMAmer® Reagents and the SomaScan® Assay