The transition from uni- to multicellular organisms was marked by an increase in the complexity of cell surface proteins that allowed intercellular communication and stable cell adhesion (1). The surfaceome (the collection of proteins residing at the cell surface) acts as the interface between a cell and its environment, mediating crucial cellular behaviours such as cell-cell communication, self and non-self recognition, and cell signalling. Additionally, the surfaceome largely dictates the shape, polarity, differentiation, and motility of cells. Given the crucial role of surface-associated proteins in every aspect of cellular life, they are, unsurprisingly, the molecular targets for roughly 70% of FDA approved drugs (2).

The aforementioned functions of the surfaceome are mediated through interactions of proteins in the cell membrane of the same cell (cis) and through interactions with proteins of neighbouring cells, the extracellular matrix, and circulating ligands (trans). The collection of cis and trans interactions is overall referred to as extracellular protein-protein interactions (ePPIs). For a cell to adapt quickly in a changing microenvironment and respond fast to stimuli while retaining the necessary plasticity, ePPIs are usually transient with low affinities (µM to mM range) to allow for fast dissociation rates. Additionally, surface proteins have a few distinct properties compared to their intracellular counterparts. Firstly, correct folding of these proteins is largely dependent on an oxidising environment for the formation of disulphide bonds. Secondly, given that more than 90% of vertebrate surface proteins are glycosylated, correct folding also requires specialised chaperons and glycosylation machinery (3). Thirdly, membrane-spanning proteins are largely amphipathic due to hydrophilic glycans decorating the extracellular part and stretches of hydrophobic amino acids that span the membrane.

To gain a better understanding of the surfaceome function, elucidating the ePPIs systemically is crucial. However, the physiological characteristics of ePPIs and surface proteins pose biochemical challenges to their study. The production or exogenous expression of surface proteins is cumbersome due to insolubility issues and requirement for post-translational modifications. Additionally, the transient nature of the interactions leads to purification problems in a pull-down experiment. These technical issues may explain why, so far, the characterisation of ePPIs has lagged behind that of intracellular proteins where affinity purification mass spectrometry and yeast two-hybrid methodologies predominate.

**Methodologies for Elucidating ePPIs**

The importance of protein interaction networks is now well-recognised, and a large number of methodologies aiming to build comprehensive protein interactome maps have been developed. The present article will mainly focus on innovative methodologies elucidating trans interactions (see Figure 1, page 18). For a comprehensive review of available methodologies to decipher surfaceome nanoscale organisation (and, thus, cis interactions) see Bausch-Fluck et al (4).

**Surface Plasmon Resonance**

Surface plasmon resonance (SPR) spectroscopy is a biophysical technique for characterising binding partners predominantly developed and performed using BiACore™ technology (5). The SPR-based binding method involves immobilisation of one binding partner on the surface of a sensor chip constructed from a thin gold film over a glass support. To assess equilibrium and binding kinetics, a technology (5). The SPR-based binding method involves immobilisation of one binding partner on the surface of a sensor chip constructed from a thin gold film over a glass support. To assess equilibrium and binding kinetics, a solution containing the second binding partner flows over the surface. Upon protein-protein interaction, the mass of material bound to the surface increases, and this increase is detected as a change in the angle of polarised light reflected from the bottom surface. Notably, the change in refractive index on the surface is linear to the number of molecules bound (5). SPR not only provides rapid qualitative information on protein-protein interaction, but also enables real-time and label-free quantitation of binding kinetics and affinities. Despite these unique advantages, SPR requires prior knowledge of the possible protein pairs, which is why it is ideal as a verification process rather than an unbiased discovery platform.

**Recombinant Protein Libraries**

In the quest to elucidate one-to-many interactions, recombinant protein libraries were developed. In this approach, proteins of interest are recombinantly expressed, purified, immobilised on a solid surface, and then probed with a ligand for assessing interaction by different means. When combined with microarray technology, thousands of proteins can be spotted on a single slide, increasing the throughput of the method. However, recombinant expression of surface proteins (especially membrane proteins) can be challenging. The discovery that the extracellular domain of membrane-anchored proteins...
(ectodomains) expressed as soluble recombinant proteins preserve their ligand binding functions was fundamental in developing protein libraries for elucidating ePPIs (6). Recently, this technique was used to screen a library of 1,300 single-pass transmembrane receptor ectodomains against recombinantly produced envelope proteins of the human cytomegalovirus (7). Nevertheless, protein production and purification is resource intensive, so studies usually focus on a subset of proteins. Nucleic acid programmable protein arrays (NAPPA) are an alternative microarray format whereby printed complementary DNAs (cDNAs) are transcribed and translated in vitro directly onto slides, eliminating the need for protein purification (8). However, it should be noted that NAPPA is a cell-free system where incorrect folding and missing post-translational modifications may affect receptor binding properties.

Overall, although recombinant library generation is resource intensive, once established, it can be repeatedly used for different ligands of interest. The major drawback of this methodology is that proteins are immobilised on an artificial surface. This implies that protein orientation/conformation on the slide may differ to the living cell, affecting ligand binding. Additionally, although the amino acid sequence of a protein determines the primary structure, protein function heavily depends on the molecular interactions between the protein and its native microenvironment. Therefore, ligand-binding events that depend on the formation of a protein complex cannot be identified. Last, but not least, certain protein classes, such as multi-pass membrane proteins or proteins containing multiple subunits, cannot be included in the libraries and, therefore, will be missed.

### Table 1: Characteristics of the major methodologies developed to elucidate extracellular protein-protein interactions

<table>
<thead>
<tr>
<th>Methodology</th>
<th>On primary cells</th>
<th>On living cells</th>
<th>With genetic modification of the cells</th>
<th>In the proper cell membrane microenvironment</th>
<th>With membrane proteins being correctly folded and modified</th>
<th>With hypothesis for target interaction needed</th>
<th>With target proteins being immobilised</th>
<th>With targets being artificially overexpressed</th>
<th>With full length functional receptors</th>
<th>Of complexes or cofactors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombinant protein libraries</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Exogenously overexpressed libraries</td>
<td>No</td>
<td>Yes (limited cell types)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ligand-based receptor capture</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Cells Expressing cDNA Libraries of Surface Proteins

Utilising the cell machinery to overexpress cell surface proteins in a living cell exogenously is an attractive alternative compared to spotting recombinant proteins on an artificial surface. In principle, a library of cDNAs encoding membrane proteins is transfected into cultured cell lines, and the overexpressed receptors are probed for extracellular interactions. Similar to NAPPA, a ‘microarray’-based screen has been developed where cDNAs are spotted on slides and are ‘reverse transfected’ into HEK293T cells grown on top, increasing the throughput of pairwise comparisons (9). Using this approach, Turner et al identified the endothelial receptor for a membrane protein of *Plasmodium falciparum* parasites (10).

This cell-based methodology permits performing ligand binding experiments on cells, hence receptors are studied within the context of an intact cell membrane. Additionally, sequence validated open reading frame expression clones are readily available and stable and – given that expression is performed in mammalian cells – proteins are usually properly modified and folded. Nevertheless, this methodology is applicable only to cell lines that can be readily transfected, limiting the number of available model systems. Additionally, cells are genetically modified and technical issues, such as controlling the varying expression levels of transfected proteins, need to be tackled. Furthermore, forced protein overexpression may result in alterations of the surface microenvironment.

### LRC-TriCEPS

To overcome the potential limitations of exogenously overexpressing proteins, an advanced cell-based chemoproteomic approach has been developed namely ligand-based receptor capture (LRC) (11). In this approach, the endogenous receptor repertoire of a given cell serves as an existing bait library that can be probed for ligand interaction. The key component of the LRC methodology is a trifunctional compound (TriCEPS or its latest development named HATRIC) that utilises the extensive glycosylation displayed by the...
majority of cell surface proteins to capture receptor interactions on living cells (12). Experimentally, the first arm is conjugated with the primary amines of a ligand, and the conjugates are added on living cells (mildly oxidised). There, the ligand binds to its target(s), and the second arm of TriCEPS is covalently crosslinked to the glycans of the binding partner. The third arm facilitates target purification for mass spectrometric analysis (11-12). So far, the methodology has been applied in a multitude of ligands, ranging from small molecules to peptides, proteins, antibodies, and even whole pathogens (11-16). The LRC methodology offers several unique advantages. Firstly, it does not require any genetic manipulation and, therefore, can

Figure 1: Schematic representation of the major methodologies developed to elucidate extracellular protein-protein interactions.

A: The binding of molecules on the plasma membrane of a cell can trigger multiple cellular responses, but the binding targets are not always known. Different classes of molecules can act as ligands, such as small molecules, peptides, soluble proteins, antibodies, proteins expressed on other cells, or pathogens. To identify the match between ligand and target(s), multiple methodologies have been developed.

B: In recombinant protein libraries, secreted factors, glycosylphosphatidylinositol, and the ectodomains of single-pass transmembrane receptors can be expressed in heterologous expression systems. The produced proteins carry purification tags to be isolated and conjugated on a solid surface. Then, they are probed with a ligand to assess interaction by different means (e.g., enzymatic reaction or fluorescence). A major limitation of this approach is that proteins are immobilised on an artificial surface, and protein orientation/conformation on the slide may differ to the living cell. Additionally, expressed proteins may not carry post-translational modifications. In cells exogenously overexpressing membrane proteins, a library of cDNAs encoding membrane proteins is transfected into cultured cell lines. The overexpressed proteins are presented in high amounts at the cell surface, where they can be probed with a ligand to assess interaction. Given that this is a cell-based system, proteins are usually properly folded and modified. Additionally, target proteins are located in the plasma membrane, thus, in their natural microenvironment where they can form cis interactions. Also, due to the overexpression of target proteins, detection of interactions may be more sensitive. However, only a handful of cell lines can be readily transfected that do not necessarily reflect the biology of the ligand-target interaction. Therefore, biologically relevant interactions may be missed because target proteins are not expressed in the selected cell line. In the LRC methodology, the endogenous receptor reporters of a given cell serves as an existing bait library that can be probed for ligand interaction. In an LRC experiment, no genetic manipulation is required, and, therefore, any cell-based model (including primary cells and tissues) can be used. Also, targets are located within the context of the natural cell-specific surface microenvironment, so they are fully functional and modified. Finally, the methodology allows the discovery of interactions between a ligand and formed complexes or interactions that require the presence of cofactors.

C: Explanation of used graphics (see page 19)
be applied on a multitude of cell lines – including primary cells – and even tissues. Secondly, receptors are located within the context of the natural cell-specific surface microenvironment, so they are fully functional. Thirdly, it is hypothesis-free, meaning that no previous knowledge or speculation about the target is required. Finally, the methodology allows the discovery of interactions between a ligand and formed complexes or interactions that require the presence of cofactors. Albeit the unique advantages, an interaction may be missed, such as, for example, in the case of the target receptor not being expressed in the selected model system or TriCEPS coupling hindering the ligand. For this reason, previous binding experiments using the ligand TriCEPS conjugates are recommended (17). Additionally, the LRC methodology enables the identification of glycosylated targets and, therefore, a small percentage of receptors (less than 10%) may be missed.

Final Thoughts

The cell surfaceome and its interactions play a pivotal role in virtually all processes of a living cell. Given the limited compatibility of widely used protein-protein interaction technologies to detect extracellular interactions, alternative methodologies have been developed. In the present review, an overview has been provided of the latest methodologies used to study extracellular protein-protein interactions and their strengths and limitations discussed (see Table 1, page 17). Driven by the characteristics of each technology, researchers should independently evaluate what is the most relevant platform for their specific project.

References

Dr Maria P Pavlou received her PhD in translational proteomics from the Department of Laboratory Medicine and Pathobiology at University of Toronto, Canada. Upon PhD completion, Maria moved to Switzerland to pursue a postdoctoral fellowship in the Institute of Molecular Systems Biology at the Swiss Federal Institute of Technology (ETH) in Zürich, focussing on host-pathogen interactions. In 2017, she joined Dualsystems Biotech as a senior scientist, and, a year later, she was promoted to Chief Scientific Officer, leading the research team to develop further the LRC methodology and establish new services. Email: maria.pavlou@dualsystems.com

Dr Paul Helbling did his Master’s degree in biology at the University of Zürich, Switzerland, and obtained his PhD in developmental biology at ETH Zürich. After his PhD, he worked at the Foundation for Cardiovascular Research and organised clinical studies according to Good Clinical Practice. Paul then moved on to MDS Pharma Services, where he became responsible for the ligand binding group that specialised in the analysis of biologics and antidrug antibodies. His next position was in business development at Solvias, responsible for southern Europe and Germany and opening new market territories in Turkey. Paul then became Head of Business Development at IBR, before joining the team at Dualsystems Biotech as CEO, focussing on identifying targets and off-targets on the living cells using liquid chromatography-mass spectrometry. Email: paul.helbling@dualsystems.com

About the authors

13. Li Z et al, Staphylococcal superantigens use LAMA2 as a coreceptor to activate T cells, J Immunol 200(4): pp1,471-9, 2018
15. Galoian K et al, Toll like receptors TLR1/2, TLR6 and MUC5B as binding interaction partners with cytotnsic profile rich polypeptide 1 in human chondrosarcoma, Int J Oncol 52(1): pp139-54, 2018

Covering every stage of the supply chain

SAMEDAN LTD
Suite E, 11 Bell Yard Mews, 175 Bermondsey Street
London SE1 3TN
+44 (0)20 7724 3456
www.samedanltd.com

Subscribe to any of the titles in our portfolio by visiting www.samedanltd.com