BENAROYA RES INST AT VIRGINIA). A similar approach uses immobilised antibodies which can bind to a target protein. A library of peptides in solution is then added to the immobilised antibodies, and once the peptides which bind to the immobilized antibodies are identified, sequences are computationally aligned with that of the whole target protein (WO 2004092741, UNIV MONTANA STATE). Other experimental methods commonly used include phage display and proteolysis of protein-antibody complexes combined with mass spectrometry [2, 7]. A problem with all existing experimental approaches for potential epitope mapping is that they require antibodies, which might not be available, or their quality and performance might not be sufficient for mapping.

Other methods that allow predicting the interactions between antibody-antigen and hence determination of potential epitopes are based on molecular modelling and computer simulation, like those described in US2014100834 (MACROMOLTEK). A large variety of *in silico* products for epitope screening has recently emerged [2], including software to screen antigens based on their genetic profile [8] and protein-protein docking methods based on B-cell epitope databases [9, 10], together with one-structure-based prediction tools [11]. Additional approaches include web-based tools which scan a whole antigen and can predict the most probable epitope regions [12].

However, these methods are not always accurate and require detailed information about protein structure. As an alternative, molecular imprinting methods have been used to assess if a small library of peptides were conformational epitopes of a protein (creatine kinase), and the method still required information about the protein and previous knowledge of the candidate epitope sequences [13].

There is therefore a need to provide improved methods and materials that may be used to determine, on a protein target, a binding site or epitope for a binding partner, such as an antibody or antigen binding fragment thereof.

SUMMARY

In one aspect, the present disclosure refers to a method for determining, on a protein, a binding site for a binding partner, comprising: (i) contacting the protein with a plurality of monomers, and polymerising the monomers to create a protein:polymer complex; (ii) digesting the protein in the complex to produce a peptide:polymer

complex; and (iii) isolating the peptide:polymer complex; and optionally(iv) sequencing the peptide; wherein the peptide corresponds to a binding site for a binding partner.

In another aspect, the present disclosure refers to the use of a molecularly imprinted polymer (MIP) for producing a composition suitable for determining, on a protein, a novel binding site for a binding partner, optionally wherein the binding site is an epitope to which an antibody binds.

In yet another aspect, the present disclosure refers to a kit for determining, on a protein, a binding site for a binding partner, the kit comprising: (i) a plurality of monomers suitable for contacting a target protein and polymerising the monomers to create a protein:polymer complex; and (ii) one or more protein cleaving agent.

DETAILED DESCRIPTION

The invention provides methods, compositions, and kits for the characterisation and analysis of proteins.

Hence, in a first aspect of the invention, there is provided a method for determining, on a protein, a binding site for a binding partner, comprising:

- contacting the protein with a plurality of monomers, and polymerising the monomers to create a protein:polymer complex;
- digesting the protein in the complex to produce a peptide:polymer complex; and
- isolating the peptide:polymer complex; and optionally sequencing the peptide, wherein the peptide corresponds to a binding site for a binding partner.

Advantageously, the method of the first aspect does not require information about protein structure or the availability of antibodies for experimental identification of peptide sequences which are binding site candidates. The characterisation and analysis may be for obtaining information related to amino acid sequences exposed on a protein surface. The information derived is useful for the identification of binding sites upon a protein, which may be suitable for the binding of binding partners, such as antibodies, aptamers, affimers, small drug molecules, MIPs or other natural or synthetic receptors. The information about these sequences may be obtained using molecular imprinting and sequencing techniques. Knowledge about these amino acid sequences may be used for generating binding partners, such as high affinity antibodies or other natural or synthetic receptors, for research, diagnostic, or therapeutic applications.

The protein upon which the method of the invention is performed, which may be referred to as the target protein, can be any molecule or composition comprising a component which comprises a chain of amino acid monomers or a peptide. The protein may be a single or multi-domain protein. The protein may be in a complex with other components, such as nucleic acids (DNA and/or RNA), cofactors, prosthetic groups, co-substrates, inorganic ions, vitamins, or sugars. The protein may include naturally occurring standard amino acids, naturally occurring non-standard amino acids, or non-naturally occurring amino acids. The protein may include modifications, including post-translational modifications or synthetic modifications. The protein may be a naturally occurring protein, or may be a protein not found in nature. The protein may be selected from the group consisting of: protein exposed on a cell surface, extracellular protein, intracellular protein, antibody, non-immunoglobulin protein, glycoprotein, phosphorylated protein, nucleoprotein, lipoprotein, glucosaminoglycan, mucoprotein, synthetic protein, or protein-DNA and protein-saccharide complexes. The protein may be a biomarker of a pathological state or a disease state. In some embodiments, two or more proteins may be used in a mixture or separately. Other embodiments may include an array of two or more proteins. In another embodiment, the protein is a complex with an enzyme inhibitor, drug molecular or other protein.