Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins

Diogo B Lima[1,7], John T Melchior[2,7], Jamie Morris[2], Valmir C Barbosa[3], Julia Chamot-Rooke[1], Mariana Fioramonte[4], Tatiana A C B Souza[5], Juliana S G Fischer[5], Fabio C Gozzo[4], Paulo C Carvalho[5,6] & W Sean Davidson[2]

[1]Mass Spectrometry for Biology Unit, CNRS USR 2000, Institut Pasteur, Paris, France. [2]Pathology and Laboratory Medicine, University of Cincinnati, Cincinnati, Ohio, USA. [3]Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. [4]Dalton Mass Spectrometry Laboratory, University of Campinas, São Paulo, Brazil. [5]Group for Computational Mass Spectrometry & Proteomics, Carlos Chagas Institute, Fiocruz Paraná, Brazil. [6]Laboratory of Toxinology, Fiocruz, Rio de Janeiro, Brazil. [7]These authors contributed equally to this work. Correspondence should be addressed to P.C.C. (paulo@pcarvalho.com) or W.S.D. (davidswm@ucmail.uc.edu).

Published online 1 February 2018; doi:10.1038/nprot.2017.113

Cross-linking coupled with mass spectrometry (XL-MS) has emerged as a powerful strategy for the identification of protein–protein interactions, characterization of interaction regions, and obtainment of structural information on proteins and protein complexes. In XL-MS, proteins or complexes are covalently stabilized with cross-linkers and digested, followed by identification of the cross-linked peptides by tandem mass spectrometry (MS/MS). This provides spatial constraints that enable modeling of protein (complex) structures and regions of interaction. However, most XL-MS approaches are not capable of differentiating intramolecular from intermolecular links in multimeric complexes, and therefore they cannot be used to study homodimer interfaces. We have recently developed an approach that overcomes this limitation by stable isotope–labeling of one of the two monomers, thereby creating a homodimer with one 'light' and one 'heavy' monomer. Here, we describe a step-by-step protocol for stable isotope–labeling, followed by controlled denaturation and refolding in the presence of the wild-type protein. The resulting light–heavy dimers are cross-linked, digested, and analyzed by mass spectrometry. We show how to quantitatively analyze the corresponding data with SIM-XL, an XL-MS software with a module tailored toward the MS/MS data from homodimers. In addition, we provide a video tutorial of the data analysis with this protocol. This protocol can be performed in ~14 d, and requires basic biochemical and mass spectrometry skills.

INTRODUCTION

A detailed understanding of the molecular basis of protein–protein interactions is key to our understanding of how protein activity is regulated. Proteins are often regulated by dynamic and highly specific protein–protein contacts. For example, of the ~20,889 enzymes with a known subunit composition, 6,000 + are homo-oligomeric. Among these, most are homodimers (3,775) or homotetramers (1,713). This high incidence of homo-oligomerization probably evolved because of structural and functional advantages such as improved stability, better control over active-site accessibility, increased number of active sites, and support of increasingly complex regulatory strategies[1]. Indeed, the enzymatic activity of several proteins is dependent on dimerization. Some families of proteins, for instance, transition from inactive to active enzymes by an autocleavage process facilitated by homodimer formation[2]. Furthermore, derangements in oligomeric assembly can lead to the formation of pathogenic structures such as amyloid. Thus, we anticipate that this protocol will aid in the study of the structure and function of homodimeric complexes. X-ray crystallography (XRC) and NMR have been tremendously useful in delivering high-resolution protein structures, a key contributor to understanding biological processes at the molecular level. The Protein Data Bank (PDB) currently contains nearly 130,000 structures from ~42,000 unique proteins[3], solved mostly by these two techniques. XRC and NMR have long been the gold standards for the determination of monomeric structures, but both have drawbacks in regard to the determination of the quaternary structure of homodimers (see 'Comparison with other approaches' section). Recent advances in mass spectrometry technology[4] and computational proteomics[5] have led to the development of strategies for defining interaction region(s) between protein complexes by using a combination of chemical cross-linking and mass spectrometry (XL-MS)[6–8]. In these strategies, the two interacting molecules are typically treated with bifunctional (or sometimes trifunctional) reagents that covalently cross-link specific residues, resulting in both intramolecular and intermolecular cross-links. The cross-linked complex is then proteolytically digested, peptides are analyzed by tandem mass spectrometry, and cross-linked peptides (cross-links) are identified. As cross-links have defined reaction sites and spacer-arm lengths, each cross-link provides distance constraints that are used for modeling protein structures and their interactions.

XL-MS has been successfully applied to several heterodimeric complexes. Many important biological processes are mediated by homodimeric or homo-oligomeric interactions. For example, the interaction between HOP2 and MND1, an important complex for DNA repair in plants, was recently worked out by cross-linking and molecular modeling[6]. In another example, the receptor-like protein tyrosine phosphatase-α is thought to exist as a loosely associated dimer in the plasma membrane. This dimeric form has minimal activity because the active site of each molecule is sterically blocked by the partner molecule[9]. Binding of ligands promotes the dissociation to higher-activity monomers. Indeed, there are many other examples of protein-activity regulation by oligomerization[1]. In fact, dimeric species constitute a large number of complexes, and ~35% of PDB structures are functional

biological dimers and two-thirds of them are homodimers. One of the most well-known examples is the dimerization of the insulin receptor A and B isoforms to form the mature receptor[10]. In such a case, conventional cross-linking MS is unable to distinguish between intra- and intermolecular cross-links, as the sequences of the two interacting chains are the same. A solution to this problem is to modify the mass signature of one of the interacting partners, preferably without covalent modification that could affect protein folding or dimerization. We recently developed an approach in which we study cross-linked peptides from homodimers in which one of the monomeric subunits is labeled with a stable isotope incorporated into the peptide chain (using 'heavy' ^{15}N instead of naturally occurring 'light' ^{14}N, for example)[11,12]. In this approach, the heavy and light monomers are mixed at a 1:1 molar ratio under conditions in which the oligomeric interactions can be transiently dissociated and regenerated. Thus, when two differentially labeled monomers are cross-linked, the detectable mass shifts allow for unambiguous distinction between intra- and intermolecular linkages. Intramolecular cross-links will contain cross-linked peptides containing either two heavy or two light peptides, whereas intermolecular cross-links will contain cross-linked heavy–light peptides in addition to the heavy–heavy and light–light peptides. The use of labeled proteins for mapping interaction regions of homodimers, however, is challenging. Although the isotopic-labeling concept is straightforward, there are important experimental and computational issues that must be addressed for a successful workflow. For example, one of the interacting partners must be labeled to high efficiency. In addition, as the light- and heavy-labeled proteins are purified as separate dimers, conditions must be optimized to allow for their efficient unfolding, followed by reassembly in heavy–light homodimers. Finally, software capable of identifying cross-linked peptides with an algorithm capable of dealing with heavy and light labeling, which are specific to each monomer, is required.

Bioinformatics tailored toward cross-linking mass spectrometry

The literature presents several efficient tools for handling XL-MS data; some examples are pLink[13], xQuest/xProphet[14], Hekate[15], XlinkX[16–18] and SIM-XL[5]. We recommend referring to these protocols when addressing nonhomodimeric XL-MS data[14,19]. Dealing with XL-MS data is far more challenging than dealing with data from shotgun proteomics, as the software must handle the combinatorics of peptides that could cross-link; it is well known that an increase in the search space represents a decrease in the sensitivity of the search engine[20]. In homodimer XL-MS analysis, the challenge is increased because, *a priori*, the heavy amino acids would need to be considered as variable modifications, thus expanding the search space exponentially. Finally, the software should be tailored toward constructing 2D visualization maps of the homodimer and only consider heavy–light links to be unambiguous if they are intermolecular. Isotope labeling, together with cross-linking, is a versatile tool that has been previously used in different ways. For example, by using isotope-labeled cross-linkers, Aebersold's group improved the detection and identification confidence of cross-linked peptides by generating spectra with characteristic doublet signals[14,21]. In another approach, isotope labeling was used to probe conformational changes in protein complexes and to quantify changes in subunit interactions in response to different stimuli; to achieve this, the authors compared the ratios of labeled and unlabeled cross-linked peptides[7]. Here, isotopic labeling is used in another innovative approach, i.e., differentiating inter- and intra-protein cross-links and making use of SIM-XL (Spectrum Identification Machine for Cross-Linked Peptides), a software with a module tailored toward the analysis of homodimers, making the analysis straightforward[12]. SIM-XL overcomes the mentioned difficulties by considering the heavy and light labeling as a binary modification (i.e., only heavy or light amino acids can exist in a single peptide). The algorithm has been tuned for accommodating the additional mass complexity imparted by the isotope labeling, and efficiently identifies the site of the interaction. Finally, SIM-XL is the only tool at hand with an integrated dynamic result viewer: besides facilitating analyses of standard cross-linking experiments, it is also capable of seamlessly displaying results from homodimer studies in a 2D map with clickable links that enable viewing of the supporting mass spectra.

Overview of the procedures

We describe a protocol that, assuming a sufficient recombinant expression system is in place, fully covers all steps in analyzing homodimers by XL-MS, using apolipoproteins as an example. As shown in **Figure 1**, this protocol describes how to generate a homodimer (or multimer) with light and heavy monomers; apply the cross-linking reagent; and isolate, enzymatically digest, and identify the cross-linked peptides by mass spectrometry. In what follows, the SIM-XL software is used to identify the cross-linked peptides and to aid in interpretation of results. Finally, SIM-XL can interface with PyMOL to facilitate viewing of the homodimer together with the disposition of the identified cross-links. The data analysis section of this protocol is demonstrated in **Supplementary Video 1**.

Applications of the method

The final product of the protocol described here will be a library of 20–100 cross-links between specific residues within and between proteins participating in a homodimeric interaction in solution. The number of contacts found depends on factors that include the length of the protein, the number of cross-linkable residues, and the solvent accessibility of those sites. The isotopic labeling strategy ensures unambiguous distinctions between intra- and intermolecular cross-links, even if both dimer participants exhibit highly analogous structures. If the Euclidean spacer-arm length of the cross-linking agent used is known, a set of distance constraints can be derived for specific points in the protein sequences. This information can then be used to either test specific structural hypotheses based on data from other methods (i.e., comparing structures predicted from XRC with those in solution) or derive *de novo* structural models using appropriate *in silico* molecular-modeling techniques. This information can then drive new experimentation; for example, a dimeric interaction site identified by XL-MS can be disrupted by site-directed mutagenesis, and the relationship between dimerization and functional activity can be probed. These distance constraints can also be combined with *in silico*-modeling techniques to elucidate the global structure of the dimeric complex and its potential interactions with other molecules.

Comparison with other approaches

For XRC, crystallization artifacts can lead to dimeric species in the crystal that may not have correspondence with the solution structure. For NMR, intra- and interchain nuclear Overhauser

effects (NOEs) cannot be differentiated, thereby hampering the correct modeling of the complex. Thus, alternative experimental approaches are needed to qualify the dimeric structure in solution or in the cell.

More recently, exciting advances in cryo-electron microscopy have made it a powerful additional technique for large biological systems, reaching resolutions comparable with those of the application of XRC to proteins. A number of homodimeric structures have already been solved, such as isocitrate dehydrogenase (PDB ID 5K10), Toll-like receptor 5 (PDB ID 3J0A) and the Xenopus KCNQ1 channel (PDB ID 5VMS), as well as some heterodimeric ones, such as the heterodimeric mammalian eIF3 (PDB ID 5A5U), eRF1/eRF3 pretermination complex (PDB ID 4CRN) and TUBB3 human microtubule assembly (PDB ID 5IJ0). These examples, although not numerous, reveal the potential of cryo-EM to resolve dimeric structures if the right sampling condition is obtained. This technology is particularly promising for advancing the structural understanding of membrane proteins in the coming decade[22]. Many proteins are homo- or heteromultimeric complexes in their active form[23], but as powerful as these standard structural techniques are, establishing the interface between subunits remains difficult. Although theoretical *in silico* modeling studies have proven useful, there is still a substantial need to experimentally define the interaction region(s) between protein complexes, mainly for those with unknown tri-dimensional structures.

Limitations of the approach

Limitations of XL-MS. Although XL-MS is rapidly growing in popularity, the approach has several drawbacks. Chief among these is the relatively limited selectivity of the cross-linking agents. N-hydroxysuccinimide (NHS) esters, for example, are primarily specific for free amines in Lys residues. If a target protein lacks Lys, or has an extensive Lys-poor region, the structural information gleaned from these reagents can be limited. Fortunately, the chemical specificities of available reagents are increasing with cysteine (–SH), carboxyl (–COOH) and even glycoprotein (–CHO)-specific reagents (available in the Thermo Fisher Scientific cross-linker catalog). Another issue is the relatively low yield of the cross-linking reaction for hydrolyzable cross-linkers such as NHS ester–based ones. Hydrolyzable cross-linkers may undergo hydrolysis on one end before anchoring to a protein, thus generating a modification on the reactive site and preventing the cross-linking formation. Thus, of all the peptides analyzed by MS, only a small percentage (<1%) contain useful distance-constraint information. This requires extensive preionization fractionation strategies, and high-resolution and mass-accuracy MS workflows. Even so, ion-suppression effects may result in loss of structural information. For this reason, there is interest in developing trifunctional cross-linking agents containing a purification handle (i.e., biotin) that can be used to enrich the sample for reacted peptides, markedly increasing the chances of identifying a cross-linked peptide[24]. If the protein being studied requires post-translational modifications to dimerize (i.e., phosphorylation, acylation, and so on), it is important that the proteins be cross-linked under conditions in which the modification has been promoted to its maximal extent. Similarly, if a cofactor such as a chaperone is required, this should be included in the cross-linking reaction. As with any exogenous modification of proteins, cross-linking has the potential to cause artifacts. The most likely scenario is that the cross-linker could stabilize a nonnative structure. However, given that the time required to complete a successful cross-link is usually much longer than the time frame of typical molecular dynamics, these effects, according to our experience, have been rare. In addition, it is possible that cross-linking can affect the solubility of a given protein, as the reagents tend to target polar-surface residues. Thus, it is important to verify that the target protein remains soluble for reasonable time periods after cross-linking. Other limitations are that proteins must go through a round of unfolding/refolding, which quite likely will not work for all dimers, and that this approach is limited to proteins *in vitro*, in contrast to 'normal' XL-MS, which can also be done on cells/lysates, for instance. Finally, the complexity of the cross-linked peptides requires high-resolution and mass-accuracy MS instrumentation in order to generate MS2 data of sufficient quality to unambiguously identify the interaction. Fortunately, with the rapid pace of improvement in MS technology, this issue should not be limiting for long.

Limitations of SIM-XL. The current version of SIM-XL has several limitations. First, the seamless integration with raw data from mass spectrometers other than those from Agilent, Thermo Fisher Scientific, or Waters requires exporting data to text-based formats such as MS2, mzML, or MGF. Second, SIM-XL requires a computer with Microsoft Windows 7 or above. Finally, there is an exponentially growing burden on computer resources as the sequence database grows. Although we are working to overcome these limitations, the first of them can already be overcome by referring to the ProteoWizard project[25].

Experimental design

Protein self-association. Homodimer formation is protein specific. It is important to fully characterize the conditions necessary for dimer formation of the protein of interest, including protein concentration, pH and temperature. Moreover, it is necessary to find a way to mix monomers in an effective way, which is also protein specific. This can be accomplished in several ways; the simplest is to mix both heavy and light monomers, and wait for the dimer-formation kinetics to equalize the mixture. This kinetic can be slow, however, and several approaches can be used to speed up the mixing, such as unfolding/refolding, ionic-strength increase, pH changes and dilution/concentration. This protocol was generated based on the conditions required for apolipoprotein (apo)A-I. Unfolding the protein in guanidine is required to achieve 100% monomeric protein for efficient mixing and formation of the dimeric complex containing a single molecule of the wild-type protein and a single molecule of the isotopically labeled protein. For proteins that cannot tolerate denaturation by guanidine, alternative methods for creating a monomeric version of the protein of interest should be determined empirically (e.g., changing buffer ionic strength, pH). In addition, the purity of the wild-type and isotopically labeled protein is critical for efficient mixing. Contamination of the sample can lead to inefficient mixing of the wild-type and isotopically labeled protein, a protein misfolding, and/or ion-signal suppression of the cross-linked peptides of interest during the LC/MS/MS analysis. Thus, we recommend stringent purification of both the wild-type and isotopically labeled protein before mixing and refolding. Finally, both the wild-type and isotopically labeled recombinant protein

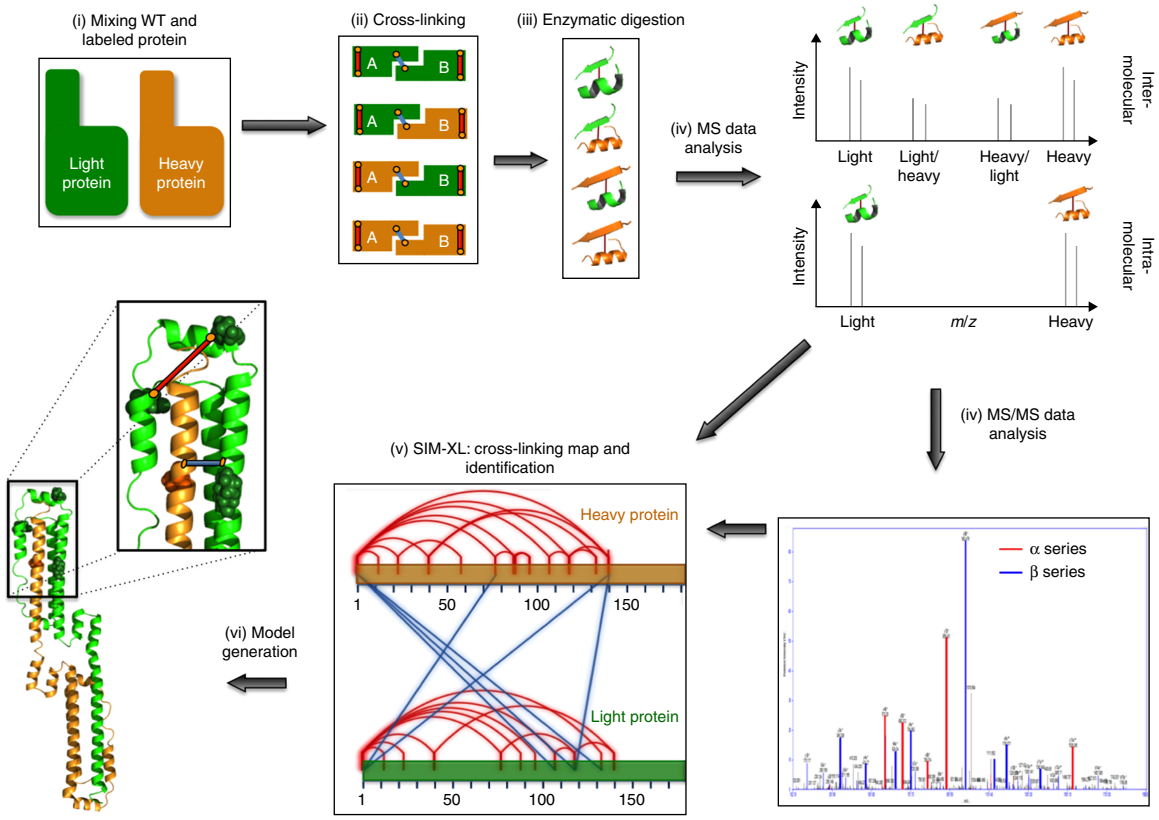

Figure 1 | Overview of the homodimer analysis workflow. A protein is expressed, purified, and its wild-type ('light') and isotopically labeled ('heavy') versions are mixed together (Steps 1–46). Multiple protein complexes are locked by cross-linking to yield mixed (heavy/light and light/heavy) and pure (heavy/heavy and light/light) (Steps 48–50) protein complexes. Homodimers (or multimers of interest) are isolated and enzymatically digested (Steps 51–58), and MS and MS/MS data are acquired (Steps 59–64). The SIM-XL software identifies the cross-linked peptides and generates a protein–protein interaction map. The software capitalizes on the doublets and quadruplets to infer the signature of intra- or interlinks, respectively (Steps 65–113). SIM-XL generates a script that enables viewing the homodimer model with the cross-links on a tool such as PyMOL (Steps 119–123). WT, wild-type.

must have identical amino acid sequences for proper labeling in SIM-XL.

Expression strategy. Each protein poses unique challenges in terms of expression. There are numerous techniques for expression and purification of proteins[26], and a variety of systems are available for isotopic labeling of the protein of interest[27]. Isotope-labeling efficiency is critical for downstream cross-linking identification. Inefficient labeling can result in a 'phantom' Survey Scan (MS1) peak preceding the isotopically labeled peptide peaks during MS1 acquisition, as shown in **Supplementary Figure 1**. Depending on the sensitivity of the mass spectrometer, this peak can be misidentified as the parent ion during generation of the Mascot generic file used by SIM-XL for cross-link identification. Thus, important cross-linked peptides could be overlooked. SIM-XL has a built-in feature to help address this issue, as in Step 92. The wild-type and isotopically labeled protein should be expressed, purified, and stored separately. This protocol is optimized for the expression of wild-type and ^{15}N-labeled complexes, using apoA-I as an example. The recombinant system presented here uses a pET-30a(+) vector that contains a 6×His affinity tag, followed by a TEV cleavage site fused at the N terminus of the mature sequence of apoA-I, which facilitates purification of the protein after its expression in *Escherichia coli* (**Supplementary Fig. 2**). In addition, this system relies on the tobacco etch virus (TEV) protease to cleave the histidine tag from the protein after initial purification. An additional glycine present between the TEV cleavage site and the N terminus was used to enhance TEV cleavage efficiency[28]. Specific details on the vector design and expression of the TEV protease have previously been reported[29].

Cross-linking. There are numerous cross-linking reagents available, and the conditions for cross-linking are specific to the reagent and the protein of interest. A copious number of cross-linking reagents exist that vary in cross-linker reactive groups, spacer-arm length, and membrane permeability. Cross-linkers can be homobifunctional (same reactive ends) or heterobifunctional (different reactive ends), and have been designed to react with primary amines, carboxylic acids, sulfhydryls, and carbonyl groups of a given protein. Spacer-arm length can vary from a 'zero-length' cross-linker (no spacer arm) to, as far as we know, 22 Å. Water-soluble cross-links can be used to study solvent-accessible regions of a given protein; alternatively, lipid-soluble cross-linkers are available for studying membrane-bound proteins. The number of cross-links will depend on whether reactive amino acids are within the upper-limit Euclidean spacer-arm length and accessibility of the amino acids. For NHS esters, the most commonly used cross-linking chemistry, a 100:1 (cross-linker/protein) molar ratio is usually used, but this may be varied depending on the number of reactive residues in the protein. Optimal ratios can be determined for dimeric

species by analyzing the cross-linking products with denaturing SDS-PAGE gels: The optimal amount of cross-linking agent is the minimum amount needed to generate a clear homodimer band by SDS-PAGE without artifactual higher-order multimers. Excessive cross-linking should be avoided to minimize the risk of inducing artificial changes in the protein structure. Cross-linking reactions with NHS-ester-specific cross-linkers are frequently performed at 4–25 °C for 1 h in non-nucleophilic buffers such as phosphate. Tris-based or ammonium buffers must be avoided, as they will compete for the NHS-ester-specific cross-linking reagents.

Dynamic Result Report. SIM-XL provides a Dynamic Result Report that treats intrapeptide and interpeptide link identifications independently. All results can be sorted/searched according to user-specified criteria. They can also be exported as a CSV or PDF file, thus helping with their interpretation. By double-clicking on an entry, the 'Spectrum Viewer' can be accessed, enabling the visualization and/or customization of the spectrum for the identification in that entry. There are several different features, represented by columns that can be displayed or hidden according to the need for details: experimental M + H, theoretical M + H, *m*/*z*, peptide- and protein-type link, primary and secondary scores, precursor charge state, protein, ppm error, file name, peaks matched, alpha and beta chains, peptide sequence, personal assessment, position of the cross-link residue, retention time, and quantitation.

MATERIALS

REAGENTS

- ^{15}N H_4Cl (Sigma-Aldrich, cat. no. 299251)
- 2-Mercaptoethanol (Sigma-Aldrich, cat. no. M6250)
- Agar (Fisher Scientific, cat. no. BP 1423)
- Ammonium bicarbonate (Fisher Scientific, cat. no. A643)
- Basal Medium Eagle (Thermo Fisher Scientific, cat. no. 21010046)
- Biotin (Fisher Scientific, cat. no. BP232)
- Bis(sulfosuccinimidyl)suberate (BS^3; Thermo Fisher Scientific, cat. no. 21580)
- BL21(DE3) *E. coli* cells (Protein Express, cat. no. 961.325)
- Bromophenol Blue (Sigma-Aldrich, cat. no. B0126)
- $CaCl_2 \cdot 2H_2O$ (Fisher Scientific, cat. no. C79)
- Dextrose (D-glucose; Fisher Scientific, cat. no. D16)
- DTT (Fisher Scientific, cat. no. BP172)
- EDTA (VWR, cat. no. 95029-566)
- $FeCl_3$ (Sigma-Aldrich, cat. no. 157740)
- Glacial acetic acid (Fisher Scientific, cat. no. A38) **! CAUTION** Glacial acetic acid is a flammable liquid and vapor. Use personal protective equipment and use it only under an approved chemical fume hood.
- Glycerol, 50% (vol/vol) (Fisher Scientific, cat. no. G33)
- Glycine (Fisher Scientific, cat. no. BP381)
- Guanidine-HCl (Fisher Scientific, cat. no. BP178)
- His-binding resin (Millipore Sigma, cat. no. 69670)
- Imidazole (Acros Organics, cat. no. AC12202)
- Iodoacetamide (Sigma-Aldrich, cat. no. I6125)
- IPTG (Gold Bio, cat. no. I2481C25)
- Kanamycin sulfate (Fisher Scientific, cat. no. BP906-5)
- KCl (Fisher Scientific, cat. no. B366)
- KH_2PO_4 (Fisher Scientific, cat. no. P284)
- LC/MS-grade formic acid (Fisher Scientific, cat. no. A117) **! CAUTION** This reagent is highly caustic. Do not allow its contact with the skin or eyes. Use safety goggles and an 8-inch face shield when handling.
- Leupeptin (Fisher Scientific, cat. no. AAJ18413)
- Luria-Bertani (LB) medium with agar
- MeOH (Fisher Scientific, cat. no. A412)
- $MgSO_4$ (Fisher Scientific, cat. no. M65)
- Na_2HPO_4 (Fisher Scientific, cat. no. BP332)
- NaCl (Fisher Scientific, cat. no. S671-3)
- $NiO4S \cdot 6H_2O$ (Fisher Scientific, cat. no. N73)
- Pepstatin A (Sigma-Aldrich, cat. no. P5318)
- pET-30a(+) vector containing 6×His-tag proximal to the N terminus or distal to the C terminus of the cloned protein (Millipore Sigma, cat. no. 69909)
- PhastGel Blue R (Sigma-Aldrich, cat. no. B4921-20TAB)
- PMSF (Acros Organics, cat. no. AC215740) **! CAUTION** PMSF is dangerous upon inhalation and skin absorption. Wear a mask while working with powder and gloves while working with liquid. If this reagent gets on your skin, wash it immediately.
- Sequencing-grade trypsin (Promega, cat. no. V5111)
- SDS (Fisher, cat. no. BP166)
- Thiamine hydrochloride (Sigma, cat. no. T4625)
- Tobacco etch virus (TEV) protease containing His-tag (Sigma-Aldrich, cat. no. T4455)
- Tris base (Fisher Scientific, cat. no. BP152)
- Tris-HCl (Fisher Scientific, cat. no. BP153)
- Tryptone (Fisher Scientific, cat. no. BP1421) **! CAUTION** This reagent poses an inhalation hazard. Use a surgical mask or a respirator when weighing out the reagent.
- Yeast extract (Fisher Scientific, cat. no. BP1422) **! CAUTION** This reagent poses an inhalation hazard. Use a surgical mask or a respirator when weighing out the reagent.

EQUIPMENT

- 0.2-µm Syringe filters (VWR, cat. no. 28145-501)
- 0.2-µm Membrane filters (Sigma-Aldrich, cat. no. 58060-U)
- 0.45-µm Syringe filters (VWR, cat. no. 28145-505)
- 100 × 15-mm Petri dishes (Fisher, cat. no. FB0875713)
- 1100 Series autosampler/HPLC system (Agilent)
- 20-ml Bond Elut SPE cartridges (Agilent, cat. no. 12131011)
- 20.6-mm Polypropylene frit (Agilent, cat. no. 12131023)
- 250-ml Plastic centrifuge bottles with caps (Fisher Scientific, cat. no. 14-375-353)
- 550 Sonic dismembrator (Fisher Scientific)
- 4–15% Mini-Protean TGX Gels (Bio-Rad, cat. no. 456-1086)
- Acclaim PepMap100 C18 column (Thermo Fisher Scientific)
- ÄKTA pure chromatography system (GE Healthcare)
- Autosampler vials (Agilent, cat. no. 9301-0977)
- Centrifuge (Sorvall, model no. RC-5C Plus Superspeed)
- Falcon 14-ml round-bottom polypropylene test tube (Corning, cat. no. 352059)
- His-binding column (Agilent, cat. no. 12131011)
- Incubator (Isotemp Incubator; Fisher Scientific)
- Incubator and shaker (New Brunswick, model no. Excella E25)
- Incubator and shaker (New Brunswick, model no. Innova 4000)
- IntegraFrit C18 trap column (New Objective, cat. no. IT100-25H002-3P)
- Kimble Kimax 95-ml glass culture tubes (Fisher Scientific, cat. no. 14-923R)
- Kimble Kimax glass culture flasks (Fisher Scientific, cat. no. 10-140)
- Microcentrifuge (Beckman Coulter, model no. Microfuge 20)
- Mini-Protean tetra vertical electrophoresis cell (Bio-Rad, cat. no. 1658005)
- NanoLC-Ultra HPLC system (Eksigent)
- Amicon Stirred Cells (Millipore Sigma, cat. no. UFSC05001)
- QStar XL Hybrid LC/MS/MS system (Sciex)
- SDS-PAGE gels (Bio-Rad)
- Sorvall SA-600 fixed-angle rotor (Thermo Fisher Scientific, cat. no. 28500)
- Fiberlite F14-6 X 250y fixed-angle rotor (Thermo Fisher Scientific, cat. no. 78500)
- Spectrophotometer for A_{280} or Lowry assay readings
- Stainless-steel culture tube closures (Fisher Scientific, cat. no. 05-888)
- 10/300 Filtration column (GE Healthcare, model no. Superdex 200 Increase)
- TripleTOF 5,600+ mass spectrometer (Sciex)
- UV–visible light spectrophotometer (Thermo Fisher Scientific, model no. Evolution 160)
- Vacuum concentrator (SpeedVac or lyophilizer)

REAGENT SETUP

Kanamycin stock Prepare 30 mg/ml kanamycin stock in ultrapure water, and pass it through a 0.2-µm sterile filter into a 14-ml sterile Falcon tube. Prepare aliquots and store them at −20 °C for up to 1 year.

LB medium with agar Dissolve 5 g of NaCl, 5 g of yeast extract, 10 g of tryptone, and 15 g of agar in ultrapure water, and use ultrapure water to bring the volume to 1 liter in a 4-liter Erlenmeyer flask. Autoclave the solution and allow it to cool to ~60 °C, at which temperature the medium is warm but still liquid. Add 1 ml of kanamycin stock per 1 liter of LB medium with agar (final concentration = 30 µg/ml) and mix well. Freshly prepare the medium and use it immediately for the generation of LB agar plates.

LB agar plates Pour 5–10 ml of LB medium with agar onto sterile Petri dishes (just enough to cover the bottom of the plate), allow the medium to solidify, wrap the plates in Parafilm, and store them upside down at 4 °C for up to 2 weeks.

LB medium Dissolve 5 g of NaCl, 5 g of yeast extract, and 10 g of tryptone in ultrapure water, and use ultrapure water to bring the final volume to 1 liter. Prepare 10-ml aliquots in 95-ml culture tubes and 200-ml aliquots in 1-liter culture flasks (use a maximum medium volume of 20% of the total flask volume to ensure good aeration). Seal the containers with steel culture tube closures and autoclave them. The LB medium can be stored at room temperature (21–25 °C) for up to 4 weeks.

5× M9 salts Dissolve 35 g of KH_2PO_4, 35.5 g of Na_2HPO_4, 2.5 g of NaCl and 5 g of $H_4Cl^{15}N$ in ultrapure water, and use ultrapure water to bring the final volume to 1 liter. Autoclave the solution. Freshly prepare the solution, and use it immediately to generate minimal growth medium.

D-Glucose Dissolve 20 g of D-glucose in ultrapure water, and use ultrapure water to bring the final volume to 100 ml. Pass the solution through a 0.2-µm sterile filter. D-glucose should be freshly prepared, and used immediately to generate minimal growth medium.

1 M $MgSO_4$ Dissolve 6.02 g of $MgSO_4$ in ultrapure water, and use ultrapure water to bring the final volume to 50 ml. Autoclave the solution. Freshly prepare the solution, and use it immediately to generate minimal growth medium.

1 M $CaCl_2{\cdot}2H_2O$ Dissolve 7.35 g of $CaCl_2{\cdot}2H_2O$ in ultrapure water, and use ultrapure water to bring the final volume to 50 ml. Autoclave the solution. Freshly prepare the solution, and use it immediately to generate minimal growth medium.

Minimal growth medium Combine 200 ml of 5× M9 salts, 20 ml of D-glucose stock, 10 ml of Basal Vitamins Eagle Media, 2 ml of 1 M $MgSO_4$, 0.1 ml of 1 M $CaCl_2{\cdot}2H_2O$, 10 mg of thiamine hydrochloride, 10 mg of biotin and 16.2 mg of $FeCl_3$, and bring the final volume to 1 liter with ultrapure water. Prepare 10-ml aliquots in culture tubes and 200-ml aliquots in 1-lLliterculture flasks. The medium can be stored at room temperature for at least 48 h.

1 M isopropyl β-D-thiogalactoside Dissolve 2.38 g of isopropyl â-D -thiogalactoside (IPTG) in 10 ml of ultrapure water in a 15-ml plastic conical tube. Using a syringe, pass the solution through a 0.2-µM filter into a fresh 15-ml tube. The solution can be stored at −20 °C for up to 6 months.

1,000× PMSF Dissolve 0.174 g of PMSF in 10 ml of DMSO. Incubate the solution at 50 °C until the particles solubilize. Store the solution at −20 °C for up to 6 months.

1,000× Leupeptin Dissolve 2.378 mg of leupeptin in 5 ml of ultrapure water. Prepare 0.5-ml aliquots. Store the aliquots at −20 °C for up to 6 months.

1,000× Pepstatin A Dissolve 4.801 mg of pepstatin A in 7 ml of DMSO. Prepare 0.5-ml aliquots. Store the aliquots at −20 °C for up to 6 months.

His-binding buffer His-binding buffer is 5 mM imidazole, 500 mM NaCl, and 20 mM Tris-HCl, pH 7.9. Dissolve 0.34 g of imidazole, 29.22 g of NaCl, and 3.15 g of Tris-HCl in ultrapure water, bring the volume to 1 liter, and adjust the pH to 7.9. Freshly prepare the buffer and use it within 48 h. (Optional) Make a 10× stock and store it at 4 °C for up to 1 month.

His-charging buffer His-charging buffer is 50 mM $NiO_4S{\cdot}6H_2O$. Dissolve 13.14 g of $NiO_4S{\cdot}6H_2O$ in ultrapure water, and bring the final volume to 1 liter with ultrapure water. Freshly prepare the buffer and use it within 48 h. (Optional) Make a 10× stock and store it at 4 °C for up to 1 month.

His wash buffer His wash buffer is 60 mM imidazole, 500 mM NaCl, and 20 mM Tris-HCl, pH 7.9. Dissolve 4.08 g of imidazole, 29.22 g of NaCl, and 3.15 g of Tris-HCl in ultrapure water, bring the volume to 1 liter, and adjust the pH to 7.9. Freshly prepare the buffer and use it within 48 h. (Optional) Make a 10× stock and store it at 4 °C for up to 1 month.

His-denaturing buffer His-denaturing buffer is 60 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, and 3 M guanidine-HCl. Dissolve 4.08 g of imidazole, 29.22 g of NaCl, 3.15 g of Tris-HCl, and 286.59 g of guanidine-HCl in ultrapure water, bring the volume to 1 liter with ultrapure water, and adjust the pH to 7.9. Freshly prepare the buffer and use it within 48 h.

His elution buffer His elution buffer is 1 M imidazole, 500 mM NaCl, and 20 mM Tris-HCl, pH 7.9. Dissolve 68.08 g of imidazole, 29.22 g of NaCl, and 3.15 g of Tris-HCl in ultrapure water, bring the volume to 1 liter and adjust the pH to 7.9. Freshly prepare the buffer and use it within 48 h. (Optional) Make a 5× stock and store it at 4 °C for up to 1 month.

His-stripping buffer His-stripping buffer is 100 mM EDTA, 500 mM NaCl, and 20 mM Tris-HCl, pH 7.0. Dissolve 29.22 g of EDTA, 29.22 g of NaCl, and 3.15 g of Tris-HCl in ultrapure water, bring the volume to 1 liter with ultrapure water, and adjust the pH to 7.0. Freshly prepare the buffer and use it within 48 h. (Optional) Make a 2× stock and store it at 4 °C for up to 1 month. ▲ **CRITICAL** The pH must be 7.0 for the EDTA to solubilize.

His-binding column Apply a frit to the SPE cartridge, and rinse the cartridge thoroughly with ultrapure water. Add ~18 ml of His -binding resin slurry (~9 ml of packed resin). Use gravity flow to drain the storage buffer. Columns can be reused, but we recommend using independent columns for different proteins to avoid the possibility of cross-contamination. Store the cartridge in His-stripping buffer at 4 °C indefinitely.

50 mM NH_4HCO_3 Dissolve 3.95 g of NH_4HCO_3 in ultrapure water, bring the volume to 1 liter with ultrapure water, and adjust the pH to 7.8. Freshly prepare the buffer and use it within 48 h.

PBS PBS is 140 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, and 1.8 mM KH_2PO_4. Dissolve 8.18 g of NaCl, 0.2 g of KCl, 1.43 g of Na_2HPO_4, and 0.24 g of KH_2PO_4 in ultrapure water, bring the volume to 1 liter, and adjust the pH to 7.4. PBS can be stored at room temperature or at 4 °C for at least 2 weeks.

PBS-denaturing buffer Dissolve 286.59 g of guanidine-HCl in PBS, and bring the final volume to 1 liter. Freshly prepare the buffer and use it within 48 h.

BS^3 in PBS Carefully weigh out BS^3 and bring it to a mass concentration of 6.5 mg/ml in PBS. Dissolve the BS^3 by gently pipetting it up and down. Try to add BS^3 to the sample within 60 s after solubilization to avoid hydrolysis of the cross-linker. Freshly prepare the solution and add it to the sample as quickly as possible.

10× Tris-buffered saline Tris-buffered saline (TBS) is 1.50 M NaCl, 100 mM Tris-HCl, 10 mM EDTA, and 0.2% (wt/vol) NaN_3.Dissolve 87.66 g of NaCl, 15.76 g of Tris-HCl, 2.92 g of EDTA, and 2 g of NaN_3 in ultrapure water, bring the volume to 1 liter, and adjust the pH to 8.2. TBS can be stored at room temperature for at least 6 months.

10× Tris-buffered saline − EDTA (TBS−) Tris-buffered saline − EDTA (TBS−) is 1.50 mM NaCl, 100 mM Tris-HCl, and 0.2% (wt/vol) NaN_3. Dissolve 87.66 g of NaCl, 15.76 g of Tris-HCl, and 2 g of NaN_3 in ultrapure water, bring the volume to 1 liter, and adjust the pH to 8.2. TBS− can be stored at room temperature for at least 6 months.

10× SDS running buffer Solubilize 30 g of Tris base, 144 g of glycine, and 10 g of SDS in ultrapure water, bring the volume to 1 liter with ultrapure water, and adjust the pH to 8.3. SDS running buffer can be stored at room temperature for at least 3 months.

SDS sample buffer Solubilize 0.095 g of Tris-HCl, 2.5 ml of 50% (vol/vol) glycerol, 0.2 g of SDS, 0.5 ml of 2-mercaptoethanol, 0.01 g of bromophenol blue, and 7.0 ml of dH_2O. SDS sample buffer can be stored at −20 °C for at least 6 months.

SDS-PAGE destain Combine 300 ml of methanol, 100 ml of glacial acetic acid, and 600 ml of ultrapure water. SDS-PAGE destain can be stored at room temperature for at least 3 months.

Coomassie blue for SDS-PAGE gels Dissolve 5 PhastGel Blue R tablets in 400 ml of dH_2O and stir the solution for 5–10 min. Add 600 ml of MeOH and stir the solution until the MeOH is dissolved. Filter the solution through

a Whatman membrane. The solution can be stored at 4 °C for at least 3 months. For staining gels, mix this solution in a ratio of 1:1 with 20% (vol/vol) glacial acetic acid and use immediately.

200 mM DTT Solubilize 30.85 mg of DTT in 1 ml of 50 mM NH_4HCO_3. Freshly prepare the solution and use it immediately.

800 mM iodoacetimide Solubilize 36.99 mg of iodoacetamide in 250 µl of 50 mM NH_4HCO_3. Freshly prepare the solution and use it immediately.

Trypsin Solubilize Trypsin in 50 mM NH_4HCO_3, pH 7.8. Freshly prepare the solution. Trypsin can be stored at −20 °C for up to 48 h. It is recommended to avoid more than a single freeze–thaw cycle to avoid loss of enzymatic activity.

LC Solvent A LC Solvent A is 0.1% (vol/vol) formic acid in H_2O. Filter the solution using a 0.2-µm membrane. Store the solution at room temperature for up to 1 week.

LC Solvent B LC Solvent B is 0.1% (vol/vol) formic acid in ACN. Filter the solution using a 0.2-µm membrane. Store the solution at room temperature for up to 1 week.

EQUIPMENT SETUP

LC–MS system Direct infusion is typically performed on 5 µg of total protein infused in Solvent A at 6 µl/min for 10 minutes. Both instruments have been discontinued by the original vendors; however, direct injection and deconvolution of intact proteins are readily performed on most mass spectrometers. The parameters should be optimized for the specific instrument. For our QSTAR XL Hybrid LC/MS/MS system, we use the following settings:

Method parameter	Value
Polarity	Positive
HPLC	
Curtain gas	20
Source Gas 1	30
Source Gas 2	0
Spray voltage	5,000
TOFMS	1/cycle
Accumulation time	1 s
Declustering potential	60
Focusing potential	230
Collision energy	5
Scan range	300–40,000 *m/z*

Reliable spectral acquisition of cross-linked peptides will be dependent upon the individual LC/MS setup. The protocol presented here was performed using a TripleTOF 5600+ from Sciex. There are a variety of LC/MS columns and mass spectrometers available that will facilitate identification of cross-linked peptides. Several parameters will require optimization and are unique to our LC/MS system. The primary optimization should be focused on (i) the chromatography elution profile, (ii) the collision energy for peptide fragmentation and (iii) the spectral acquisition parameters. In general, the use of longer C18 columns with a 2- to 3-h gradient and the solvents given above serve as a good starting point for chromatographic separation. Collision energy and spectral acquisition should be optimized to acquire sufficient MS1 counts for initial identification by SIM-XL. It should be noted that some mass spectrometry software is inefficient at assigning charge states to the parent ions, which can result in exclusion of the MS/MS data and ultimately missed identification of cross-linked peptides. We use the following MS settings for cross-link identification:

Method Parameter	Value
Polarity	Positive
Nano-ESI	
Curtain gas	30
Source Gas 1	15
Source Gas 2	0
Interface heater	150 °C
Spray voltage	2,500
TOFMS	1/cycle
Accumulation time	250 ms
Dissociation potential	70
Collision energy	10
Scan range	500–1,800 *m/z*
Information-Dependent Acquisition (MS2)	30/cycle
Accumulation time	75 ms
Dissociation potential	70
Collision energy	43 (+/− 4)
Scan range	100–1,800 *m/z*
Charge (z)	2–5
Intensity minimum	250 counts
Exclusion time	15 s

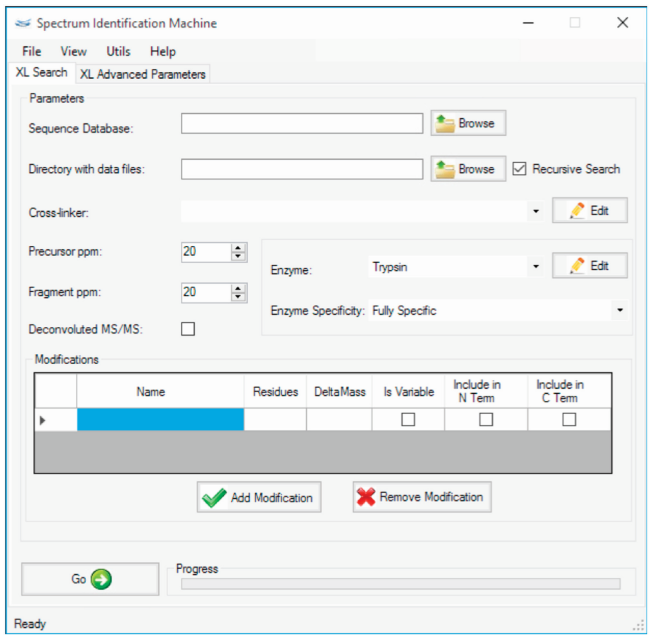

Figure 2 | SIM-XL's main screen. All search parameters, as well as the homodimer-analysis parameters, will be set on this screen. The search can be started through the 'Go' button.

Gel electrophoresis system Remove the gel from its packaging and remove the green tape from the bottom of the gel. Place the gel into the electrode assembly included in the tetra-cell electrophoresis cell package. Place the assembly and the gel into the electrophoresis buffer tank. Dilute 10× SDS running buffer to 1× with ultrapure water, and fill the internal chamber completely and the external chamber to ~1/3 of the total volume. Remove the comb from the gel and load the samples into the wells. Enclose the chamber with the electrophoresis lid and run the samples. Typically, samples are run at 200 V for 35 min, which provides sufficient separation and visualization of apoA-I. Conditions should be optimized for the protein of interest. Remove the gel from the cassette and add Coomassie blue diluted at a 1:1 ratio with 20% (vol/vol) acetic acid. Stain the gel for ~15 min at room temperature (21–25 °C) with gentle rocking. The staining reagent should be removed, and the gel should be incubated with SDS destain until the protein is sufficiently visualized (usually 30–45 min). The destain can be removed, and the gels can be incubated with ultrapure water to rehydrate them, and they can be stored at room temperature for up to 2 weeks.

Hardware requirements A personal computer (PC) with at least 8 GB of RAM (16 GB or more is recommended).

Local storage is required for processing mass spectrometer RAW files. The space occupied by these files can vary substantially, depending on the mass spectrometer used. At least 1 GB of disk space should be available for hosting the mass spectrometer and work files.

Software requirements Microsoft Windows 7 (64-bit version) or later is required. Windows 10 is recommended.

.NET Framework v4.6.1 or later must be installed. A computer with Windows 10 should already have this requirement fulfilled. Nonetheless, if the .NET Framework is not detected during SIM-XL installation, an attempt will be made to automatically install it through Microsoft's website. The latest version, as of the time of this writing, is available from https://www.microsoft.com/en-us/download/details.aspx?id=53345.

Thermo Fisher Scientific MSFileReader should be installed for working directly with the RAW files. Instructions on obtaining this program are available from https://thermo.flexnetoperations.com/control/thmo/download?element=6306677.

SIM-XL setup Go to the SIM-XL home page at http://patternlabforproteomics.org/sim-xl, and click the 'Download' button. After installation, SIM-XL's main screen should pop up(**Fig. 2**). ▲ **CRITICAL** Administrative-access privileges are required for installation. See TROUBLESHOOTING section.

PROCEDURE

Bacterial cell growth and protein expression ● TIMING 3 d

1| Transform pET-30 vector containing His-tagged protein and a TEV cleavage site into BL-21 (DE3)-competent cells. Expression is dependent upon plasmid concentration and cell competency, but typically, for apoA-I, use 1 µl of plasmid (isolated by standard miniprep) per 50 µl of cells.

2| Plate the cells on LB agar plates, and grow colonies by incubation at 37 °C for 12–16 h with the agar upside down.
▲ **CRITICAL STEP** The plates should be incubated and/or stored (see below) with the agar upside down to avoid condensation.
▲ **CRITICAL STEP** Do not overgrow the colonies. This can result in local breakdown of the kanamycin, and formation and survival of satellite colonies that lack the plasmid.
■ **PAUSE POINT** The plates can be removed from the incubator, sealed in Parafilm, and stored upside down at 4 °C for up to 2 weeks.

3| Add 10 µl of kanamycin stock to 95-ml culture tubes containing either 10 ml of LB medium (for wild-type expression) or 10 ml of minimal growth medium (for expression of ^{15}N-labeled protein).

4| Pick single colonies and inoculate them in the culture tubes prepared in Step 3.

5| Cover the culture tubes with steel lids, and incubate them in a tabletop shaker at 225 r.p.m. at 37 °C. Monitor growth using a UV-visible light spectrophotometer measuring absorbance at 600 nm. Grow the colonies until the A_{600} value reaches between 0.7 and 0.8 (typically 12–16 h for apoA-I).
▲ **CRITICAL STEP** Growth rates will be dependent upon the protein of interest. For each protein, it is recommended to experimentally determine the ideal expression time and temperature for optimal expression. Overgrowth will deplete the kanamycin, resulting in the growth of cells lacking the plasmid.

6| Add 200 µl of kanamycin stock to 1-liter flasks containing either 200 ml of LB medium (for wild-type expression) or 200 ml of minimal growth medium (for ^{15}N-labeled protein).

7| Inoculate the medium with 2 ml of confluent 10-ml culture from Step 5 (1 ml per 100 ml of culture).

8| Cover the culture flasks with steel lids and incubate them in a shaker at 225 r.p.m. at 37 °C.
▲ **CRITICAL STEP** Growth rates will be dependent upon the protein of interest. For each protein, it is recommended to experimentally determine the ideal expression time and temperature for optimal expression. Overgrowth will deplete the kanamycin, resulting in the growth of cells lacking the plasmid.

9| Induce the cells with 100 µl of 1 M IPTG in ultrapure water (0.5 mM final concentration).
▲ **CRITICAL STEP** It is important to include a negative-control culture that is not induced with IPTG. This will aid in the identification and optimization of the expression of the protein of interest by comparing noninduced cells with induced cells at Step 11.
? TROUBLESHOOTING

10| Incubate the culture flasks in a shaker for 2 h at 225 r.p.m. at 37 °C.

11| Remove 1 ml of sample and spin it at 15,493*g* for 10 min at room temperature in a tabletop microcentrifuge. Decant the supernatant and solubilize the pellet in 100 µl of SDS sample buffer + 100 µl of ultrapure water. Check the protein expression with 10 µl of sample by running an SDS-PAGE and staining the sample with Coomassie blue. Compare the induced with the noninduced control cultures on the same gel.

12| Transfer the remaining cells to 250-ml centrifuge bottles, and pellet the cells by spinning them in a Sorvall SLA-1500 rotor at 9,268*g* at 4 °C for 10 min in a Sorvall RC-5C centrifuge. Decant the supernatant without disturbing the pellet. Freeze the pellets at −20 °C until ready for protein purification.
■ **PAUSE POINT** The pellets can be stored at −20 °C for at least 4 weeks.

Protein purification ● TIMING 4–5 d

13| Take the cell pellets from the freezer, and allow them to thaw on ice for 10 min. Add 3.0 ml of His-binding buffer per original 200-ml culture. Mix gently by swirling the buffer by hand.

14| If multiple 200-ml cultures were grown, combine the solubilized pellets in a 15-ml plastic conical tube, keeping wild-type and isotopically labeled protein separate. Prepare 6-ml aliquots in separate 15-ml plastic conical tubes for sonication in Step 16.

15| Add 6 µl each of 1,000× PMSF, 1,000× leupeptin, and 1,000× pepstatin to each 6-ml sample.

16| Lyse the cells by sonicating the pellet on ice with the tapered probe for 10 min with a 50% duty cycle. Place the end of the tip at about the point where the tube tapers. Adjust the probe setting so that the probe has a low growl sound, rather than a high-pitched whine (~4 on the Sonic dismembrator).
▲ **CRITICAL STEP** It is important to maintain a low temperature during sonication. Therefore, the samples must be sonicated on ice with pulse intervals short enough to prevent overheating the sample when lysing the cells.

17| Transfer the sample to a 10-ml centrifuge tube, and spin it at 28,384*g* for 20 min at 4 °C in a Sorvall SA-600 fixed-angle rotor in a Sorvall RC-5C centrifuge to pellet the cellular debris.
▲ **CRITICAL STEP** It is possible to obtain ~10% more protein out of the pellets if they are resolubilized in 2 ml of buffer and resonicated. Combine the second supernatant with the first.

18| Pass the supernatant through a 0.45-µm sterile syringe filter. The sample is now ready for purification over the His-binding column.

19| Prepare the His-binding column by sequentially applying 1 column volume (CV) of His-stripping buffer, 2 CVs of ultrapure water, 2 CVs of His-charging buffer, and 2 CVs of His-binding buffer.
▲ **CRITICAL STEP** For all of the following purification steps, it is critical to avoid applying samples containing DTT or EDTA to the His-binding columns. DTT will react with nickel, and turns the column brown. EDTA is a chelator that will remove nickel from the column, reducing binding efficiency.

20| Apply the supernatant from Step 18 to the column, and allow it to pass through the column by gravity flow. His-tagged protein should bind and be retained on the column.
▲ **CRITICAL STEP** We suggest recovering the unbound material (flow-through) to prevent accidental disposal of the protein as a result of any technical problems.

21| Apply 1 CV of His-binding buffer to the column. Drain the buffer.

22| Stop the flow of the column, and apply 1 CV of His-binding buffer. Agitate the resin by gently stirring it with a glass rod. Start the flow of the column and drain the buffer. Repeat this step.

23| Wash the column with 1 CV of His-binding buffer, followed by 1 CV of His wash buffer.

24| Stop the flow and add 1 CV of His-denaturing buffer. Agitate the resin by gently stirring it with a glass rod. Incubate the resin for 15 min, and start the flow and drain the column.

25| Wash the column in 1 CV of His wash buffer. Drain the buffer.

26| Elute the protein with 1.5 CV of His elution buffer and collect the elution. Confirm that the protein of interest is in the elution by running an SDS-PAGE and staining with Coomassie blue. We recommend starting with 5, 10, and 15 µl of sample for electrophoresis to ensure visualization. Leave blank lanes between samples to avoid lane contamination in the case of overloading.
▲ **CRITICAL STEP** We recommend running the eluted sample and the flow-through sample from Step 20 on the same gel for comparison. In the case that the protein of interest is found in the flow-through sample and not in the elution, regenerate the column (Step 27) and re-initialize the His-binding column (Step 19). Repeat Steps 20–26, starting with reapplication of the flow-through that contains the protein of interest. In the case that protein of interest is found in both the flow-through sample and the elution, re-equilibrate the column with 2 CVs of His-binding buffer and repeat Steps 20–26, starting with the reapplication of the flow-through that contains the protein of interest.

27| Regenerate the column by applying 2 CV of His-stripping buffer. Repeat this step until the column is completely white in appearance. Store the column in His-stripping buffer at 4 °C indefinitely.

28| Dialyze the protein into 1× TBS, and determine the protein concentration using the Markwell-modified Lowry assay[30]. Dialysis should be performed at a minimum ratio of 1:100 sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C.
▲ **CRITICAL STEP** Imidazole from the elution buffer will strip some of the nickel off the column into the sample. It is important to dialyze the protein into the buffer containing EDTA to chelate the free nickel in the sample.

29| Remove the His-tag with TEV protease by adding it to a final molar ratio of 20:1 protein/protease. Slowly rock the sample at 25 °C for 2 h. Confirm that the tag is completely cleaved from the protein by SDS-PAGE.
▲ **CRITICAL STEP** We use our own engineered TEV protease[29] that specifically cleaves at the histidine tag on the protein of interest and, itself, contains a histidine tag to facilitate its removal in the following steps. Equivalent TEV proteases are available for purchase; however, careful consideration should be taken to ensure cleavage specificity and that the protease contains a histidine (or equivalent) tag to facilitate its removal from the system.

30| Dialyze the cleaved protein into TBS– to remove the EDTA. Dialysis should be performed at a minimum ratio of 1:100 of sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C.

31| Prepare the His-binding column, as described in Step 19.

32| Apply the cleaved and dialyzed protein to the His-binding column. Collect the flow-through, which should contain the pure protein. The His-tag that was cleaved from the protein, and the His-tagged TEV protease should be retained on the column.

33| Wash the column with 1 CV of His-binding buffer and collect the flow-through.

34| Repeat Step 33.

35| Wash the column with 1 CV of His wash buffer and collect the flow-through.

36| Repeat Step 35
▲ **CRITICAL STEP** Keep the flow-through collected in Steps 32–36 separate. Check the protein purity by SDS-PAGE. Combine the fractions containing pure protein. Repeat the wash steps until the protein is no longer visible by SDS-PAGE.

37| Desalt the sample by dialyzing it against 50 mM NH_4HCO_3, and measure the protein concentration using the Markwell-modified Lowry assay[30]. Dialysis should be performed at a minimum ratio of 1:100 of sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C. Dilute the dialyzed protein to 1 mg/ml using 50 mM NH_4HCO_3.
▲ **CRITICAL STEP** For optimal mass spectrometry results, salts must be removed from the sample to avoid ion suppression. There are a variety of alternative approaches available, such as PD-10 columns, C18 spin columns, or ZipTips, that can be used. In addition, some mass spectrometers contain separate salt traps in tandem with the peptide-separation columns to help circumvent this issue.

38| Evaluate the purity and the molecular weight of both the wild-type and isotopically labeled protein by direct infusion of the protein into a mass spectrometer (see Equipment Setup). Given the protein sequence and the number of nitrogen

molecules present, the efficiency of labeling can be evaluated in the isotopically labeled protein by comparing the experimental molecular weight with the theoretical molecular weight calculated based on the labeling technique. In addition, sample purity can be evaluated; as the fewer the peaks, the cleaner the sample.

39| Prepare aliquots of the samples, freeze them on dry ice, and lyophilize. Store the samples at −80 °C.
■ **PAUSE POINT** Lyophilized protein has been stored for up to 2 years in our laboratory at −80 °C with no observable effects on protein function or structure upon resolubilization.

Preparation and mixing of the wild-type and isotopically labeled protein ● TIMING 1 d

40| Starting with ~5 mg each of lyophilized (or solubilized) wild-type and isotopically labeled protein, unfold the proteins independently by adding PBS-denaturing buffer to a concentration of 0.1 mg/ml.
▲ **CRITICAL STEP** If cysteines are present on the protein, DTT or an equivalent reducing agent must be present in the buffer at a concentration sufficient to reduce and prevent disulfide-bridge formation. We recommend a final concentration of 10 mM of DTT in the denaturing buffer.

41| Solubilize each protein preparation by gentle mixing or rotation for a minimum of 1 h at 4 °C.

42| Determine the protein concentration of the wild-type and isotopically labeled protein using a Lowry the assay.

43| Mix the wild-type and isotopically labeled species at a molar ratio of 1:1. Dilute the mixed protein sample to a concentration of <0.1 mg/ml in PBS-denaturing buffer to ensure that the proteins are in an unfolded, monomeric state.
▲ **CRITICAL STEP** The final volume should result in a yield of 100% monomeric protein to ensure complete dissociation of higher-order oligomers and efficient unfolding. Alternative denaturing reagents can be used, given that the unfolding process is reversible upon removal of the reagent.
? TROUBLESHOOTING

44| Mix the proteins thoroughly by gentle stirring or end-over-end rotation overnight at 4 °C.

Refolding of the wild-type and isotopically labeled protein mixture ● TIMING 1–2 d

45| Refold the mixed protein by slowly dialyzing out the denaturing reagent against PBS. Dialysis should be performed at a minimum ratio of 1:100 of sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h at 4 °C each to remove the denaturing agent from the sample and ensure proper refolding.
▲ **CRITICAL STEP** Careful consideration should be taken for refolding proteins containing cysteines. Some cysteine-containing proteins require disulfide formation for proper refolding and thus, they should be dialyzed into PBS with no reducing agent. Alternatively, other cysteine-containing proteins can form disulfide bridges during the refolding process that result in artifactual structures. In the latter case, DTT or an alternative reducing reagent should be present in the dialysis buffer at a concentration sufficient to prevent disulfide-bridge formation during the refolding process. After refolding has completed, an additional dialysis step can be performed to remove the denaturing reagent.

Cross-linking and isolation of the homodimer complex ● TIMING 48 h

46| Remove the sample from dialysis and slowly concentrate it to a final protein concentration of ~1 mg/ml using a nitrogen stirred cell concentrator.
▲ **CRITICAL STEP** Concentration of the sample for dimer formation should be optimized to avoid aggregation and/or nonspecific protein–protein interactions that may occur using high concentration rates in conjunction with ultracentrifugation. This step is protein specific. We recommend concentration using a nitrogen stirred cell concentrator, which allows variable concentration rates and buffer conditions, and helps avoid oxidative modification of amino acids that could potentially disrupt the native protein complex.

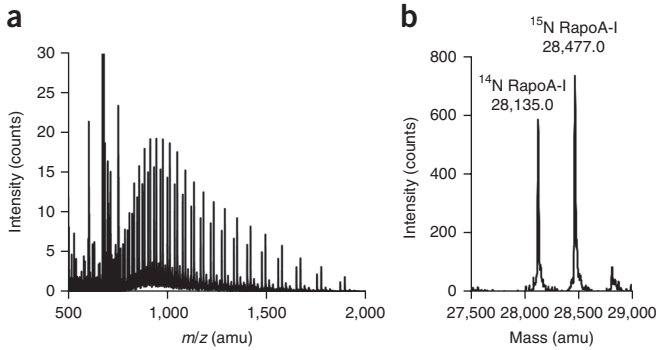

Figure 3 | Mass spectrum and reconstruction of a sample containing wild-type and ^{15}N-labeled apoA-I. 25 µg of sample was removed and dialyzed against 50 mM NH_4HCO_3. Dialysis should be performed at a minimum ratio of 1:100 of sample to dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C. Approximately 5 µg of protein was infused directly onto a QSTAR XL LC/MS/MS system at a flow rate of 6 µl/min. The theoretical mass of ^{15}N-apoA-I is 28,483.69. Each labeled nitrogen results in an ~1.0-Da increase in molecular weight. Thus, protein labeling for this experiment resulted in 6 nonlabeled nitrogens (out of a total of 348 for apoA-I), for an overall efficiency of 98.3%. amu, atomic mass unit; RapoA-I, recombinant apoA-I.

47| (Optional) For optimal mass spectrometry results, remove a 25-µg aliquot of protein, and desalt it by dialyzing against 50 mM NH_4HCO_3. Dialysis should be performed at a minimum ratio of 1:100 of sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C. The experimental ratio of the wild-type to isotopically labeled protein can be evaluated by directly introducing nondigested protein to the mass spectrometer (**Fig. 3**) as described in Step 38. The relative intensity of the peaks between wild-type and isotopically labeled protein should be nearly equivalent if they were mixed and properly refolded. Alternatively, users can digest the sample by proceeding to Step 52 (without cross-linking) and looking at the experimental ratio of the wild-type to isotopically labeled peptides.
▲ **CRITICAL STEP** For optimal mass spectrometry results, salts must be removed from the sample to avoid ion suppression. There is a variety of alternative approaches available, such as PD-10 columns, C18 spin columns, or ZipTips, that can be used downstream after proteolytic digestion of the homodimer to remove excess salt. In addition, some mass spectrometers contain separate salt traps in tandem with the peptide-separation columns to help circumvent this issue.

48| Cross-link the protein by adding BS^3 in PBS to a final molar ratio of 50:1 BS^3/protein. The final concentration of BS^3 should range between 0.25 and 5 mM.
▲ **CRITICAL STEP** BS^3 should be solubilized in PBS immediately before cross-linking the protein. Add BS^3 to the protein within 60 s of solubilization to avoid hydrolysis of the cross-linking reagent.
▲ **CRITICAL STEP** Cross-linking conditions should be optimized for the protein of interest. Cross-linking can be performed at 4 °C, room temperature, or 37 °C for various time intervals. In general, the higher the temperature, the shorter the time frame needed for the cross-linking reaction to come to completion. However, higher temperatures can also result in increased dynamics of the proteins in solution. Typically, for apoA-I, cross-linking is carried out at 4 °C for 12 h to limit the dynamics and ensure completion of the cross-linking reaction.

49| Quench the cross-linking reaction by adding 10× TBS to a final concentration of 20–50 mM Tris, and incubating the mixture at room temperature for 15 min. Alternatively, an excess of nonreacted cross-linker can be removed by dialysis, size-exclusion chromatography, or PD-10 desalting columns.

50| Run 8 µg of non-cross-linked and 16 µg of cross-linked samples on SDS-PAGE, and stain the samples with Coomassie blue to evaluate cross-linking efficiency and ensure that a sufficient amount of the 'locked' homodimer is present for downstream mass spectrometry analysis. An example of efficient cross-linking of apoA-I is shown in lane 1 of **Figure 4b**.
▲ **CRITICAL STEP** If the protein contains cysteines, disulfide bridges can lead to misinterpretation of the cross-linking efficiency of the homodimer complex; i.e., the gel will show abundant homodimers due to disulfide bridges, but it may not contain many cross-links. In this case, the cross-linked sample should be run under reducing conditions by boiling the sample in a sample buffer containing DTT or some equivalent reducing reagent for 10 min to ensure that the homodimer present in the sample has sufficient intermolecular cross-links.
? TROUBLESHOOTING

51| Purify the cross-linked complexes from the monomeric forms of the proteins using size-exclusion chromatography. Separation and purification of multimers should be optimized for the protein of interest. The conditions and purification of apoA-I multimers are shown in **Figure 4**.
▲ **CRITICAL STEP** We recommend size-exclusion chromatography over SDS-PAGE separation when isolating homodimer species from higher-order oligomers or monomeric species. Depending on the efficiency of the cross-linking reaction, it is possible that not all oligomeric species will be locked together with intermolecular cross-links. In SDS-PAGE separation, these 'unlocked' complexes are more likely to dissociate and contaminate lower-order bands. For example, a trimer with inefficient cross-linking could result in intermolecular cross-links forming between two of the three molecules. In that case, the complex could dissociate during electrophoresis into a locked dimer and a monomer. Although minor, these protein conformations are probably not native to the

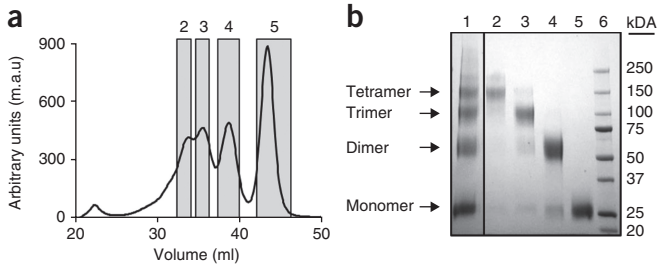

Figure 4 | Cross-linking and purification of apoA-I multimers. Approximately 4 mg of recombinant $^{14/15}$apoA-I was cross-linked with BS^3 at a molar ratio of 50:1 for 12 h at 4 °C. The reaction was quenched, and the sample was concentrated to 250 µl and applied to three Superdex 200 columns in series at a flow rate of 0.3 ml/min. 0.5-ml fractions were collected, and 15 µl of each fraction was evaluated by SDS-PAGE and staining with Coomassie blue. Fractions containing pure multimers were pooled accordingly. (**a**) The fast protein liquid chromatography trace of the cross-linked sample. The shaded areas represent the fractions pooled for each multimer, and the numbers correspond to the lane number in **b**. (**b**) 16 µg of apoA-I cross-linked before separation (lane 1) and 8 µg each of tetrameric apoA-I (lane 2), trimeric apoA-I (lane 3), dimeric apoA-I (lane 4), and monomeric apoA-I (lane 5). Molecular-weight standards are shown in lane 6. m.a.u., milli-arbitrary units.

monomer or dimer and can result in the appearance of 'non-native' cross-links for that particular complex. Dissociation is less likely to occur during the size-exclusion chromatography process. Moreover, it will also improve the efficiency of recovery for cross-linked peptides.
▲ **CRITICAL STEP** Gel-filtration columns and an appropriate resin should be selected based on the protein of interest's molecular weight. If needed, consider using multiple sizing columns in series to optimize species separation or alternative purification strategies such as anion exchange.

52| Collect 0.5-ml fractions and run 15 µl of each fraction on an SDS-PAGE to confirm the purity of the dimer or multimer of interest. Pool the fractions containing pure complex and concentrate them to 1 mg/ml. Confirm the purity of the pooled sample with 8 µg of protein by SDS-PAGE. Run 16 µg of cross-linked protein before purification by gel filtration as a control.

Desalting of the homodimer ● TIMING 1 d
53| Desalt the sample by dialyzing it against 50 mM NH_4HCO_3. Dialysis should be performed at a minimum ratio of 1:100 of sample/dialysis buffer (vol/vol), and the sample should be dialyzed a minimum of 3× for 3–6 h each at 4 °C.
▲ **CRITICAL STEP** For optimal mass spectrometry results, salts must be removed from the sample to avoid ion suppression. There are a variety of alternative approaches available, such as PD-10 columns, C18 spin columns, or ZipTips, that can be used downstream after proteolytic digestion of the homodimer to remove excess salt. In addition, some mass spectrometers contain separate salt traps in tandem with the peptide-separation columns to help circumvent this issue.

54| Remove the desalted sample from the cartridge, and measure the final protein concentration of cross-linked proteins. It is recommended to have a minimum of 10 µg of sample for subsequent preparation for mass spectrometry analysis.
▲ **CRITICAL STEP** Most cross-linking reagents interfere with A_{280} values. We recommend determining the final protein concentration using an alternative assay, such as Markwell Lowry or an equivalent assay immune to cross-linking interference.

Digestion and preparation of the homodimer for mass spectrometer analysis ● TIMING ~16 h
▲ **CRITICAL** If cysteines are present, reduce the disulfide bonds, and alkylate the reduced cysteines according to Steps 55 and 56. If no cysteines are present, proceed directly to Step 57.

55| Reduce the disulfide bonds by adding 200 mM DTT to a final concentration of 10 mM, and incubating the sample at 42 °C for 30 min. If no cysteines are present, proceed directly to Step 57.

56| Alkylate the reduced cysteine residues by adding 800 mM iodoacetamide to a final concentration of 40 mM, and incubating the sample in the dark for 30 min at room temperature.

57| Take the required amount of protein needed for the LC/MS analysis (10 µg allows for several injections), and digest it with sequencing-grade trypsin at a ratio of 1:20 (protein/trypsin) for 16 h at 37 °C.
▲ **CRITICAL STEP** A number of proteolytic enzymes are available for protein digestion. One should carefully consider the cleavage sites of the proteolytic enzyme and how they apply to the sequence of the protein of interest. Depending on the mass spectrometer, larger peptides and cross-linked peptides are more difficult to fragment, which can result in poor fragmentation and mass spectrum of the parent ion. This will result in low scores or missed identifications of cross-linked peptides by SIM-XL.

58| Dry the homodimer digest using a vacuum concentrator.
■ **PAUSE POINT** The dried peptide samples can be stored at −20 °C for at least 1 year.

LC–MS/MS analysis ● TIMING 90 min–3 h, depending on instrument gradient
59| Solubilize the sample in 0.1% (vol/vol) formic acid to a final concentration of 0.2 µg/µl (50 µl for 10 µg of starting material).

60| Vortex the samples and incubate them for 10 min at RT.

61| Centrifuge the sample at 10,000*g* for 10 min at RT to remove any precipitated peptides or contaminants.

62| Transfer the sample to a 250-µl autosampler vial.

63| Inject 5 µl of sample (~ 1 µg of digest) onto a C18 IntegraFrit trap column and run the following gradient:

Time (min:s)	% Solvent B	Flow rate	Step
00:00	0	2 µl/min	Load/desalt on trap
15:00	0		
15:01	5	300 nl/min	Switch trap in-line with column
50:00	40		Gradient elution
55:00	85		
60:00	85		
65:00	5		Re-equilibrate
80:00	5		

▲ **CRITICAL STEP** If you are running multiple samples in series, the column can be cleaned between runs by running a blank sample.

64| Acquire the spectra by operating the mass spectrometer in positive-ion mode, in which each cycle consists of one TOF-MS scan (0.25-s accumulation time in a 350- to 1,500-*m/z* window) followed by 30 information-dependent acquisition-mode MS/MS scans on the most intense candidate ions selected from the initial TOF-MS scan during each cycle. Each product ion scan should have an accumulation time of 0.075 s and a collision energy (CE) of 43 with an 8-unit scan range.

65| Convert the *.wiff file into a Mascot generic file (MGF) using a converter, available in the ProteoWizard Project[25].
▲ **CRITICAL STEP** The parameters for converting a *.wiff file into an MGF file are strictly dependent on the spectral data acquired from the LC/MS system. In general, parameters should be optimized to generate a robust MGF file that contains parent ions and MS/MS data above the 'noise' of the instrument. Generation of MGF files without proper filtering can result in large MGF files with low-quality data and consequent long search times for SIM-XL.

Setup of parameters for analysis of data on SIM-XL ● TIMING not more than 5 min
66| Start the SIM-XL software. Select an input protein-sequence database file in the FASTA format by clicking on the 'Browse' button. SIM-XL enables the reading of FASTA files, but an alternative format containing the target-reverse decoy sequences is also allowed (e.g., T-R, T).

67| Select the directory containing the tandem mass spectra files in any of the following formats: mzML 1.1.0, MS2, MGF, Agilent's RAW, Thermo's RAW, or Waters' RAW. If 'Recursive Search' is checked, all subdirectories will be searched.
▲ **CRITICAL STEP** For reading Thermo Scientific RAW files, the MSFileReader must be installed. Instructions on obtaining this file are available from https://thermo.flexnetoperations.com/control/thmo/download?element=6306677.
▲ **CRITICAL STEP** If any spectrum fails to contain a precursor charge state, SIM-XL will predict it using the YADA algorithm[31] that is internally embedded, and a warning will be issued.

68| Select one of the cross-linkers from the drop-down list. By default, there are five cross-linkers registered: Disuccinimidyl Suberate (DSS), Disuccinimidyl Glutarate (DSG), Disuccinimidyl Glutarate (DSSeb), Disulfide, and Zero-length.

69| (*Optional) Registration of a new cross-linker in the library (Steps 69–72)*. To include a new cross-linker or to edit an existing one, click the 'Edit' button beside the 'Cross-linker' drop-down list. The XL library window will pop up (**Fig. 5**).

70| Fill out the fields 'XL Name', 'XL Mass Shift' (reaction XL mass) in Daltons, 'Reaction Sites', and optionally the 'Modification Mass Shift' (in Daltons as well) and 'Reporter Ion' masses as follows:

'XL Name' is a user-defined identifier for the cross-linker.

'XL Mass Shift' is the net mass of the cross-linker that will be added to the peptide masses.

'Reaction Sites' are all combinations of amino acids that react with the cross-linker. The keyword 'N-TERM' should be included if the cross-linker reacts with the N terminus. For example, the entry for the DSS cross-linker should be 'KK KS SS

KN-TERM SN-TERM'. Similarly, C terminus reactivity can be specified by using the keyword 'C-TERM'. Note that subsequent reaction sites in an entry must be separated by a single space.

'Modification Mass Shift' (optional) defines an artificial modification caused by the cross-linker. For example, DSS/BS^3 can react with a single lysine residue, generating the so-called dead-end modification. If the 'Modification Mass Shift' field is left empty, SIM-XL will always consider all possible combinations of cross-linked peptides.

'Reporter Ions' (optional) defines the *m/z* value of fragments that are specific to cross-linked peptides[13]. If these *m/z* values are given, SIM-XL will search only MS/MS spectra containing at least one of the corresponding fragments, thus speeding up the process; otherwise, all spectra will be searched.

71| To conclude the creation of the new cross-linker, click the 'Update' button. This will make the new cross-linker available within the library and usable for searching.
▲ **CRITICAL STEP** The 'XL Name', 'XL Mass Shift', and 'Reaction Sites' parameters are mandatory.

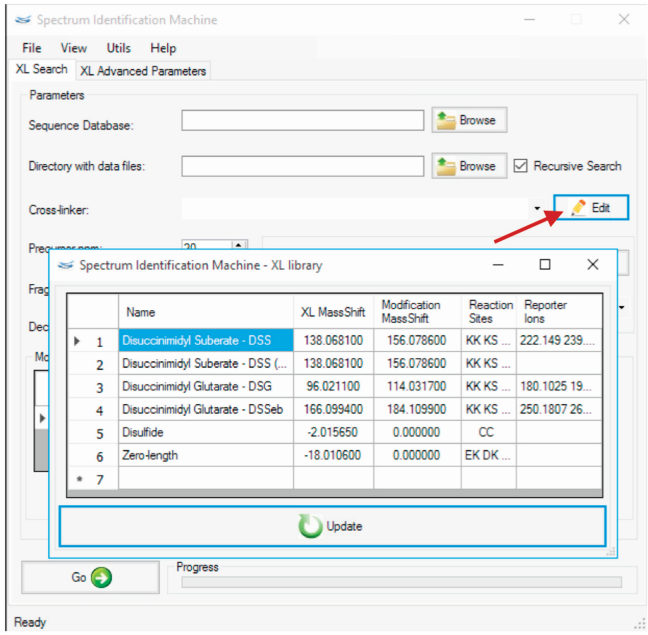

Figure 5 | Cross-linker (XL) library. If the desired cross-linker is not available in the drop-down list, it can be included in the XL library. A window enabling this is accessible by clicking on the 'Edit' button beside the 'Cross-linker' drop-down list (arrow).

72| To delete a cross-linker entry from the library, select the entire line (by clicking on the row's header cell), press the DEL key, click the 'Update' button, and then click the 'OK' button.

73| Select proteolytic enzyme(s) from the drop-down list. By default, there are nine enzymes registered, namely: Asp-N, Arg-C, Chymotrypsin, CNBr, Glu-C, Lys-C, Lys-N, PepsinA, and Trypsin. When selecting more than one enzyme, SIM-XL will perform a sequential digestion on the sequence database.
▲ **CRITICAL STEP** At least one enzyme must be selected.

74| (*Optional) Registration of a new enzyme in the library (Steps 74 and 75)*. To include a new enzyme or to edit an existing one, click the 'Edit' button beside the 'Enzyme' drop-down list; the enzyme library window will pop up (**Fig. 6**). In an empty line, complete the corresponding fields with the enzyme's name, indicate whether cleavage occurs in the C or the N terminus, and provide a regular -expression encoding of the enzymatic cleavage.
▲ **CRITICAL STEP** All fields are mandatory. There is no corrector for regular-expression encodings. For more information on building regular expressions, we refer the reader to https://msdn.microsoft.com/en-us/library/az24scfc(v=vs.110).aspx.

75| To remove an enzyme from the library, select the row containing it, press the DEL key, click the 'Update' button, and then click the 'OK' button.

76| Select an 'Enzyme Specificity' from the drop-down list.

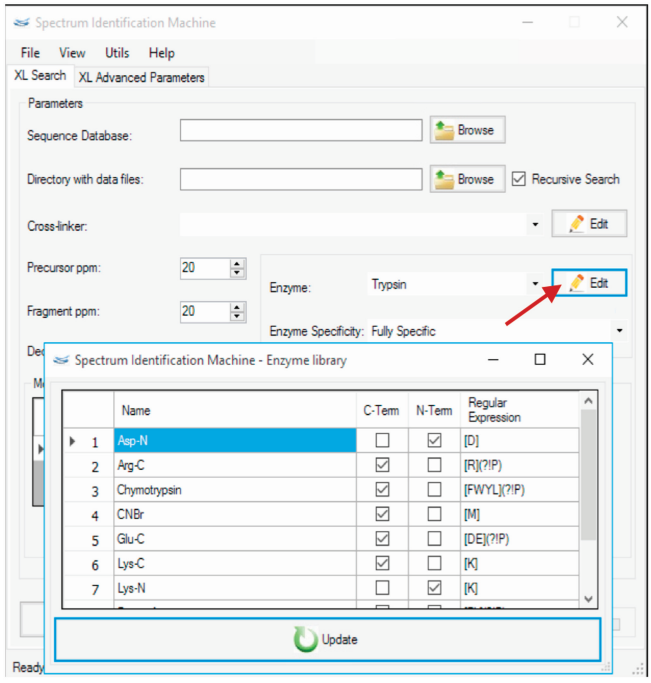

Figure 6 | Inserting enzymes into the enzyme library. Access the enzyme library by clicking on the 'Edit' button indicated by the arrow. All fields are mandatory, and a regular expression is required to specify the cleavage sites of a new enzyme. For example, the typical regular expression for trypsin is [KR](?!P). For more information on building regular expressions, we refer the reader to https://msdn.microsoft.com/en-us/library/az24scfc(v=vs.110).aspx.

77| Specify the 'Precursor' and 'Fragment ppm' tolerances, as seen in **Figure 2**.

78| If the spectra in the data files are deconvoluted, i.e., decharged and deisotoped, then the 'Deconvoluted MS/MS' option can be checked. We refer the reader to YADA as a tool for deconvoluting mass spectra[31].

79| To add a modification from the 'Modification library', select it from the drop-down list, and then click the 'OK' button found in the 'Modifications' group box.

80| (*Optional) Registration of new modifications with the library (Steps 80–84)*. To include a new modification or to edit an existing one, click the 'Edit' button beside the 'Existing modification' drop-down list (**Fig. 7**). A new window will pop up.

81| Fill out the fields 'Name', monoisotopic 'Mass Shift' in Daltons, and 'Amino Acid(s)', and then press the 'Update' button.
▲ **CRITICAL STEP** All fields are mandatory.

82| To delete a modification from the library, select the whole line by clicking on the row's header cell, press the DEL key, click the 'Update' button, and then click the 'OK' button.

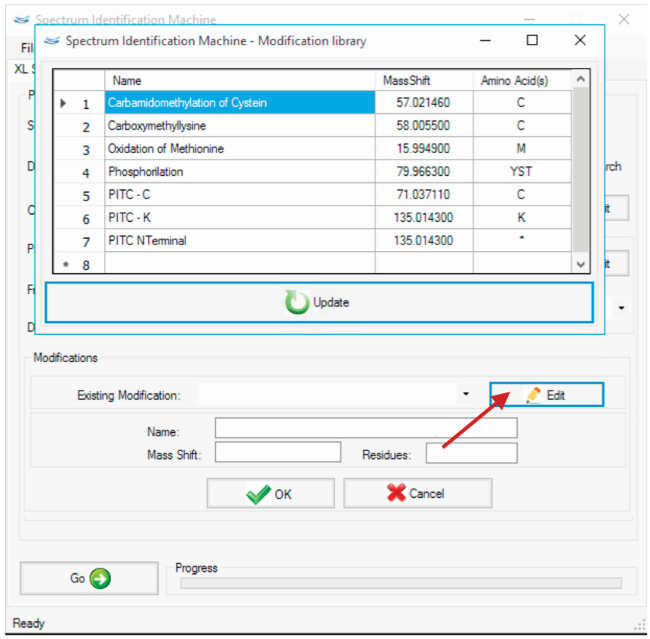

Figure 7 | Adding post-translational modifications to the library. A modification that is not found in the list can be easily inserted into the modification library by first clicking on the 'Edit' button (arrow). A name for the modification is required, as well as a mass shift and the amino acids that will be affected.

83| Indicate whether the modification is a variable one and whether it applies to the C terminus and/or the N terminus by checking the corresponding boxes. For example, if not all methionines in the sample are expected to be oxidized, then the modification should be checked as 'variable'; however, for modifications that are expected in all occurrences of the amino acid, such as, say, carbamidomethylation of cysteine, the 'variable' option should remain unchecked.
▲ **CRITICAL STEP** The search space increases exponentially with the number of variable modifications added, thus increasing the search time.

84| To remove a post-translational modification, select the desired one, and then click the 'Remove Modification' button.

85| Go to the 'XL Advanced Parameters' tab for setting up all parameters to be used in homodimer analyses. Check the 'Homodimer analyses' option, and then click the 'Edit' button. A new window will pop up (**Fig. 8**). There are several parameters on this tab, as detailed in a previous standard XL-MS protocol[19].
▲ **CRITICAL STEP** A protein-sequence database file must be selected before clicking the 'Edit' button.
? TROUBLESHOOTING

86| Select a protein to be labeled in the drop-down 'Protein' list.

87| Select one of the two modes to label a protein isotopically:

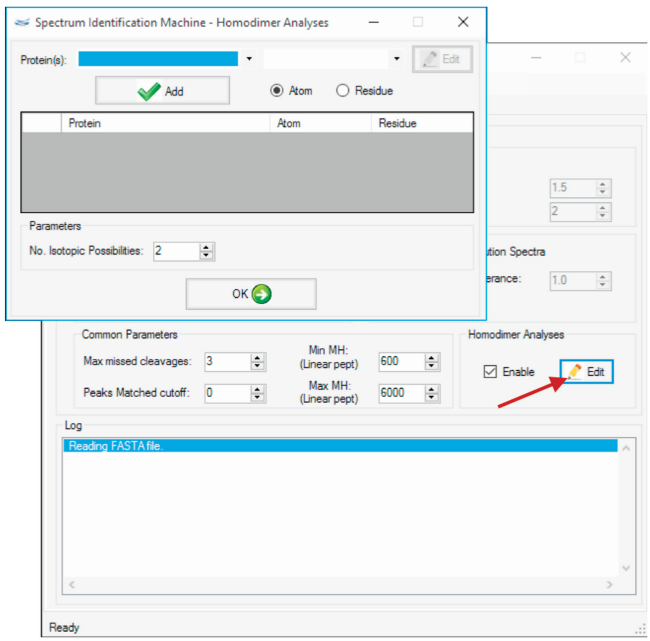

Figure 8 | Setting up homodimer parameters. An analysis tailored toward homodimers can be set up by clicking on the 'Edit' button (arrow) after enabling the corresponding checkbox. A new window will pop up, allowing the configuration of the parameters.

By 'Atom': The protein can be labeled with the 'heavy' version of Carbon (^{13}C) and/or Nitrogen (^{15}N). Select at least one of them from the drop-down list.

By 'Residue': The protein can be labeled with the modified version of a specific residue. By default, there are two modified residues registered: (i) 4–13C Arginine, which represents the residue Arginine with four ^{13}C atoms and (ii) 6–13C Lysine, which represents the residue Lysine with six ^{13}C atoms.
▲ **CRITICAL STEP** At least one option must be selected to label a specific protein.
? TROUBLESHOOTING

88| The 'Residue' option can be modified by the inclusion or removal of a customized residue. For this, click the 'Edit' button beside the drop-down list. A new window will pop up (**Fig. 9**).
▲ **CRITICAL STEP** The 'Edit' button will be enabled only if the 'Residue' option is checked.

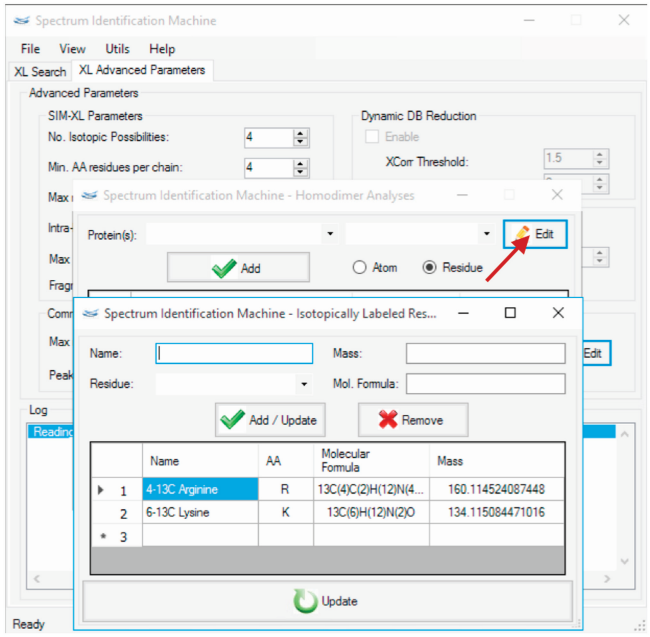

Figure 9 | Adding a new isotopically labeled residue. The 'Edit' button (arrow) in the homodimer analysis window gives way to a secondary window that enables including a new customized residue using a configuration similar to that of SILAC[23].

89| Fill out the fields 'Name', 'Molecular Formula', and 'Mass' in Daltons, and then select a 'Residue' out of the 20 in the drop-down list. Click the 'Add' button, and then click 'Update'. The parameter descriptions are as follows:

'Name' is a user-defined identifier for the cross-linker.

'Molecular Formula' is the chemical representation of the residue.

'Mass' is the isotopic mass calculated according to the molecular formula.

'Residue' is the residue that will be labeled.
▲ **CRITICAL STEP** If the mass provided by the user is inconsistent with the formula, it will be calculated automatically, and a warning will be issued asking for a replacement.

90| To modify a specific residue, click on its row's header cell. All fields will be filled out automatically and can then be changed. Any modifications require clicking on the 'Add' button, and then clicking on 'Update' to conclude.

91| The drop-down list will be updated with all existing residues, once insertions and/or modifications are concluded. Select at least one.

92| The precursor mass stored in the raw data files may not correspond to the monoisotopic peak. Select the 'Number of Isotopic Possibilities' option to allow SIM-XL to infer the monoisotopic peak, required to identify the molecule, at the cost of scaling up the search space. So, for example, for a peptide with a monoisotopic mass of 4,000 Da, the most intense peak in the isotopic envelope is expected to be M+2 (that is, ~4,002 Da), and thus it is the most likely one to be selected for fragmentation. If the number of isotopic possibilities is set to three, SIM-XL will search this MS/MS spectrum considering the precursor masses 4,002, 4,001, and 4,000, plus or minus the given ppm tolerance. In this example, the correct monoisotopic precursor mass is 4,000, which therefore can be correctly identified by SIM-XL. If a high number of isotopic possibilities is set, the search space will increase accordingly and will negatively impact SIM-XL's sensitivity.

93| Conclude the isotopically labeled proteins option by clicking the 'OK' button. If done correctly, a window will pop up saying your protein was successfully labeled.

Performance of the search ● TIMING: from a few minutes up to a couple of hours
94| After setting up all search parameters, click the 'Go' button to start the search. The 'Log' box will display information on search progress.

▲ CRITICAL STEP If any required parameter is left unfilled, the search will not be started.
? TROUBLESHOOTING

95| Once all parameters have been set, save all of them for future searches. This is accomplished by selecting 'Save SIM-XL params' from the 'File' menu or pressing CTRL + S, as can be seen in **Figure 10**.

96| (Optional) If you want to load a previously saved file of search parameters, select 'Load SIM-XL params' from the 'File' menu or press ALT + L (**Fig. 10**).

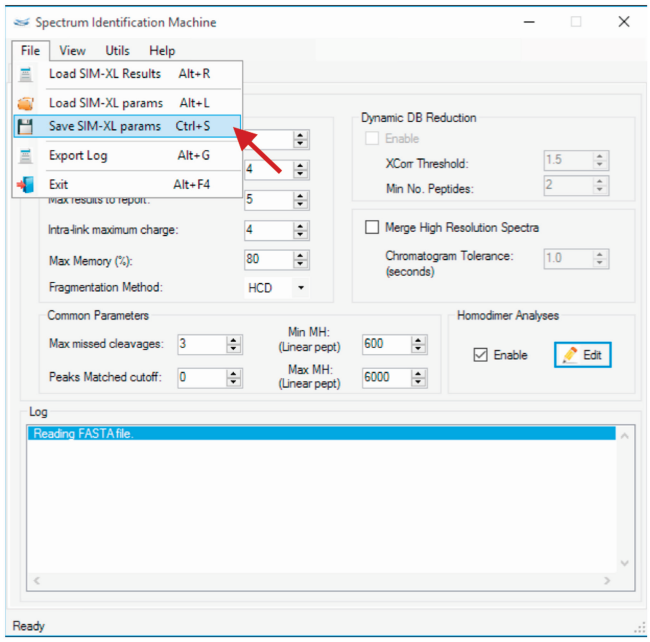

Figure 10 | Saving SIM-XL parameters. All parameters can be saved/reloaded to avoid setting up the same configuration more than once by clicking on the menu option (arrow).

Evaluation of the results ● TIMING variable, up to 2 h

97| Once the search is finished, a 2D map of protein -interaction results (2D-Map) will be shown. This feature is an interactive map showing all the cross-links identified with scores above the cutoff values given in the 'Score (intra-link)' and 'Score (inter-link)' fields. In the 2D-Map (**Fig. 11**), there are two ways tor assess protein–protein interactions. The default, referred to as 'Protein-Bar', represents each protein as a rectangle of size directly proportional to its sequence length, with ticks marking the residue numbers at the bottom. The protein's ID is shown outside the rectangle. Each inter-protein cross-link is represented as a blue line, and each intra-protein cross-link is represented as a red arc. Customize the viewer by left-clicking on the rectangles and dragging them around, as well as zooming in or out with the zoom bar. Hover the mouse pointer over a cross-linker representation to open a pop-up window showing linker details such as the linking amino acids and their positions.
▲ CRITICAL STEP The 2D-Map is displayed only if no more than 5,500 cross-links are identified.
? TROUBLESHOOTING

98| Zoom in and out of the 2D-Map to facilitate interpretation.

99| Display or hide proteins on the 2D-Map by checking/unchecking them on the right side of the map.

100| In the case of four or more proteins, the Circular Viewer can be enabled to simplify visualization. To do this, check the 'Circular Viewer' option, located between the 'Min peaks matched per chain' and 'Zoom' fields. This viewing mode allows evaluation of all identified cross-links in a way similar to the previous one, as shown in **Figure 12**.

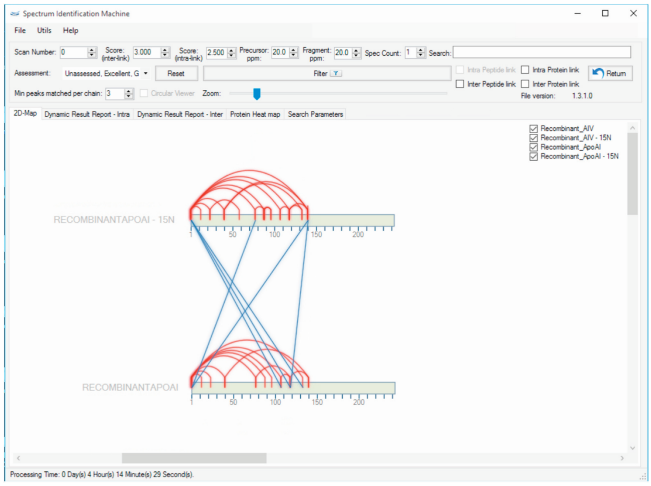

Figure 11 | 2D-Map of protein-interaction results. SIM-XL is the first XL tool to generate a 2D interactive map showing the protein–protein interactions with clickable links for browsing the supporting mass spectra. The display can be customized by selecting the proteins of interest from the list of proteins available on the right.

101| The Circular Viewer also enables switching to Protein-Bar mode just by right-clicking on the desired protein. Use the checkboxes displayed on the 2D-Map's right side to select which proteins to display.
▲ CRITICAL STEP At least one protein must be checked.

102| Highlight the interactions between two proteins by clicking on one of them and then clicking on the other. All identifications between them will turn darker.

103| Right-click a cross-linker representation to open a pop-up window showing all the identified cross-links, with corresponding scan numbers, scores, and charge

states. Each identification can be evaluated by selecting one of the available options: 'Excellent', 'Good', 'Medium', 'Fair', and 'Poor'.

104| Access the spectrum representing each specific identification by left-clicking on the desired identification to open the 'Spectrum Viewer' (this feature is discussed below in this protocol).

105| Export all cross-links visualized on the 2D-Map as a PNG, TIFF, or JPG image, or as a PDF file, just by clicking on the 'File' menu, selecting '2D-Map', then 'Save Image' (ALT + I), or 'Export 2D-Map' (ALT + R).

106| Use the fields below to filter the results in order to improve the display:

'Scan number': only spectra whose scan numbers match that of this field will be shown.

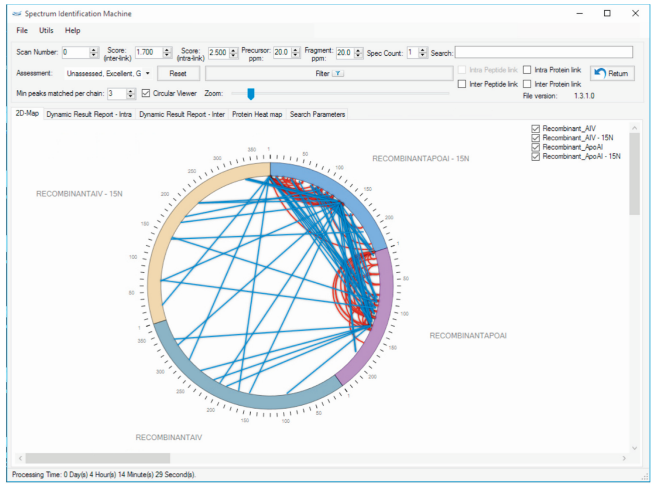

Figure 12 | Using the Circular Viewer mode to visualize cross-links from three or more proteins. When many proteins are identified, this viewing mode makes it easier to see which proteins are interacting with one another.

'Intra- and inter-link score': only identifications containing scores greater than or equal to this value will be shown.

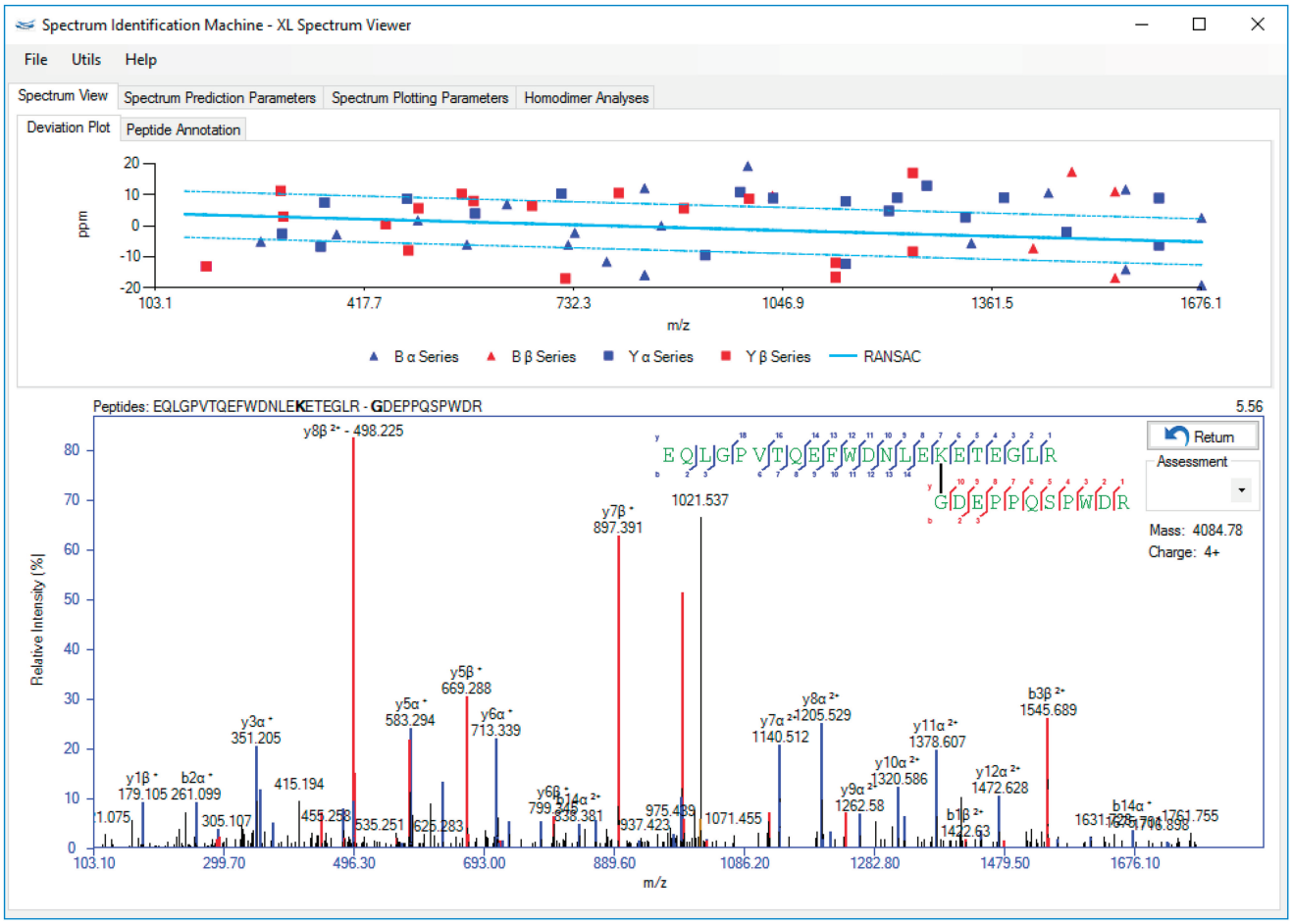

Figure 13 | The Spectrum Viewer allows seamless evaluation of identifications. The corresponding result can be tentatively altered while one checks whether the changes lead to improved peak matching. Three types of evaluation are offered: Spectrum and peptide annotation, and the Random Sample Consensus (RANSAC), which shows how well matched the experimental and theoretical peaks are. A personal assessment can also be done to be used for filtering in a downstream analysis; each spectrum can be graded from excellent to poor by the drop-down menu displayed in the upper right-hand corner of the mass-spectrum viewer. This figure is not a screenshot; we copied the contents of the peptide annotation tab and pasted them into the spectrum view to enable a broader understanding of the functionalities.

'Precursor and fragment ppm error': only identifications with a ppm error less than or equal to this value will be shown.

'Spectrum counter': only results with a minimum identification number will be shown.

'Search a specific peptide sequence': all identifications matching the desired sequence will be shown.

'Personal assessment': only results with a specific assessment will be shown.

'Searching only for intra- and/or inter-peptide links, intra- and/or inter-protein links': only results conforming to the specific criteria given in these fields will be shown.

'Minimum peaks matched per chain': only identifications with a minimum number of peaks matched in each peptide chain will be shown.

▲ **CRITICAL STEP** All results to be exported will be selected according to the contents of the Dynamic Result Report.

107| SIM-XL provides the Spectrum Viewer (**Fig. 13**), which displays an annotated XL mass spectrum. Navigate to the 'Spectrum View' tab to browse the spectrum, zoom in on it and out, and easily view which peak was attributed to the corresponding fragment. To zoom in, click and drag the mouse over the desired *m/z* range. To zoom out, double left-click the mouse. If some peak was annotated wrongly, undo the annotation by simply right-clicking on the desired peak and confirming the action.

▲ **CRITICAL STEP** There are many parameters on the 'Spectrum Prediction Parameters' and 'Spectrum Plotting' tabs. They were all described in a previous protocol[19]; so here, we focus only on discussing homodimer analyses.

108| The 'Homodimer Analyses' tab enables customization of the isotopically labeled proteins to check the identification. To accomplish this, first check the 'Enable' option (**Fig. 14**).

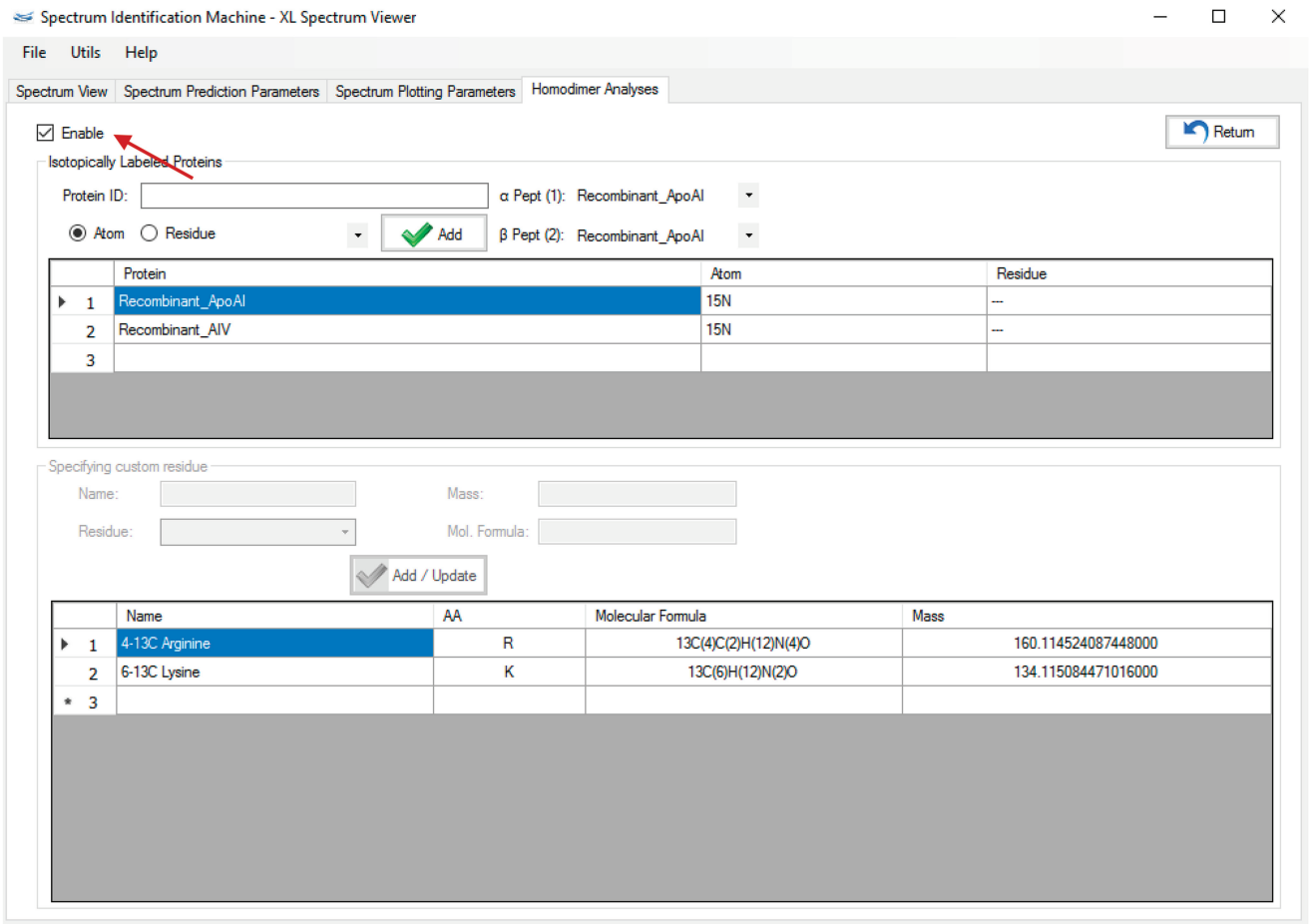

Figure 14 | Customizing the homodimer parameters on XL Spectrum Viewer. To obtain a better homodimer analysis, the previously discussed parameters can be changed, and others can be similarly added, to check the confidence of the identification. To customize it, click on the 'Enable' option (arrow).

109| Enter a protein ID for identifying the modified protein.

110| As described in Step 87, check 'Atom' if the protein sequence will be modified by atoms, indicating which ones will label the protein (^{13}C and/or ^{15}N). Otherwise, check 'Residue' and select the desired modified residue.

111| *Addition of a new modified residue (Steps 111–113)*. Insert a new modified residue as a specific modification (explained in Steps 88 and 89)

112| Select for the alpha and/or beta peptide of the protein sequence that will be isotopically labeled.

113| Annotate a new spectrum using the new parameters through the 'Spectrum Prediction Parameters' tab and by clicking on the 'Plot' button.

? TROUBLESHOOTING

114| SIM-XL can load results in the SIM-XL (*.simxlr) or mzIdentML file (v1.2) formats. The first of these allows evaluation of each identified cross-link dynamically through the 2D-Map or the Dynamic Result Report. Although the mzIdentML format does not contain all parameters required by SIM-XL to evaluate all features, with the original tandem mass spectra file, it is possible to load the identified cross-links. Load the results in the 'Result Browser' window by selecting 'Load Results' from the 'File' menu or pressing ALT + S. Otherwise, if the main window is open, select 'Load SIM-XL Results' from the 'File' menu or press ALT + R. The mzIdentML format is required by PRIDE to perform a complete submission to its repository, and SIM-XL is the first tool to upload this format in the v1.2, the latest format defined by the HUPO Proteomics Standard Initiative (http://www.psidev.info/tools-implementing-mzidentml).

▲ **CRITICAL STEP** The mzIdentML results can be loaded within SIM-XL only by accessing the 'Load Results' option. The original data file (e.g., an mzML, MGF, MS2, .d (Agilent) or RAW file) is required to be loaded along with the mzIdentML file.

115| SIM-XL allows the association of extracted ion chromatograms (XICs) with identifications. It also allows viewing of the intensity versus time plot. The area under the curve (i.e., XIC) can be used as a surrogate for an identification's relative abundance, as can be seen in **Figure 15**. A normalization factor, given by the total ion current of the full run, is provided to

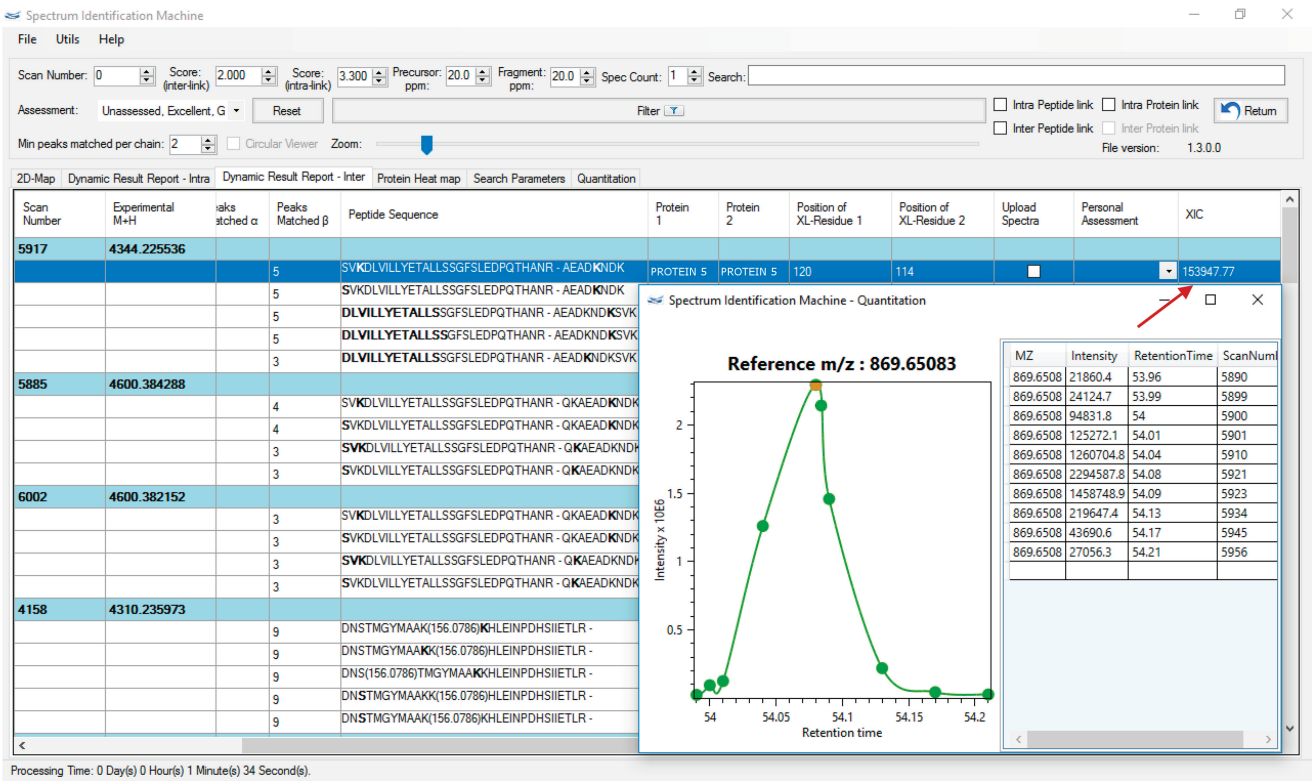

Figure 15 | Relative XL quantitation. SIM-XL provides a relative quantitation value for each identification. The Quantitation window displays a plot in which each green point represents the intensity of the precursor peak extracted from the respective MS1 event. The orange point indicates when the MS/MS event that enabled the identification occurred. The XIC value (i.e., area under the curve) is indicated in the SIM-XL interface by a red arrow.

enable comparison of XICs of the same identification across different runs. The normalization factor is simply the XIC of the entire run.

116| Enable the 'Quantitation' module (by pressing ALT + Q or by clicking the 'Utils' menu and 'Quantitation' to calculate the XICs). A new window will pop up (**Fig. 16**).

117| Select a directory containing all tandem mass spectra files and then click the 'OK' button (**Fig. 17**). Progress in all steps will appear in the 'Log' box. When the process is finished, the Dynamic Result Report will be updated, and a new column, titled XIC, will be displayed (**Fig. 15**).
▲ **CRITICAL STEP** Currently, only Agilent and Thermo RAW files are supported for calculating XICs.

118| Display the XIC plot by double-clicking on a specific result.
▲ **CRITICAL STEP** The value '−1' in a cell of the XIC column indicates that the XIC operation failed. This may occur when not enough MS1 events are found.

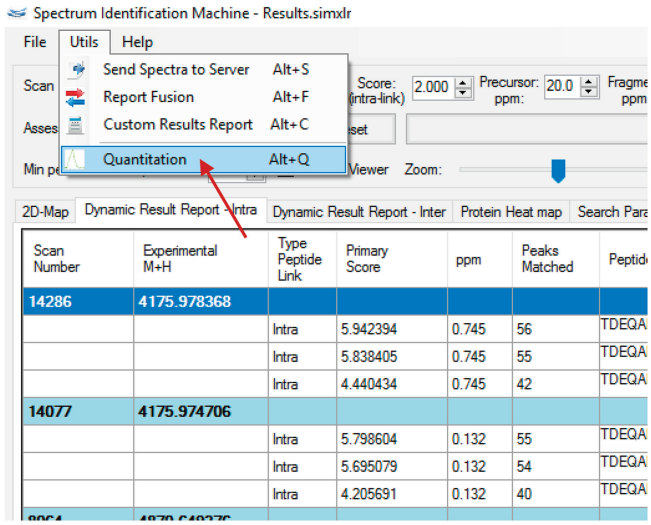

Figure 16 | Enabling the Quantitation module. The Quantitation module can be enabled through the 'Utils' menu on the Result Browser by selecting 'Quantitation' (arrow) (or pressing ALT + Q).

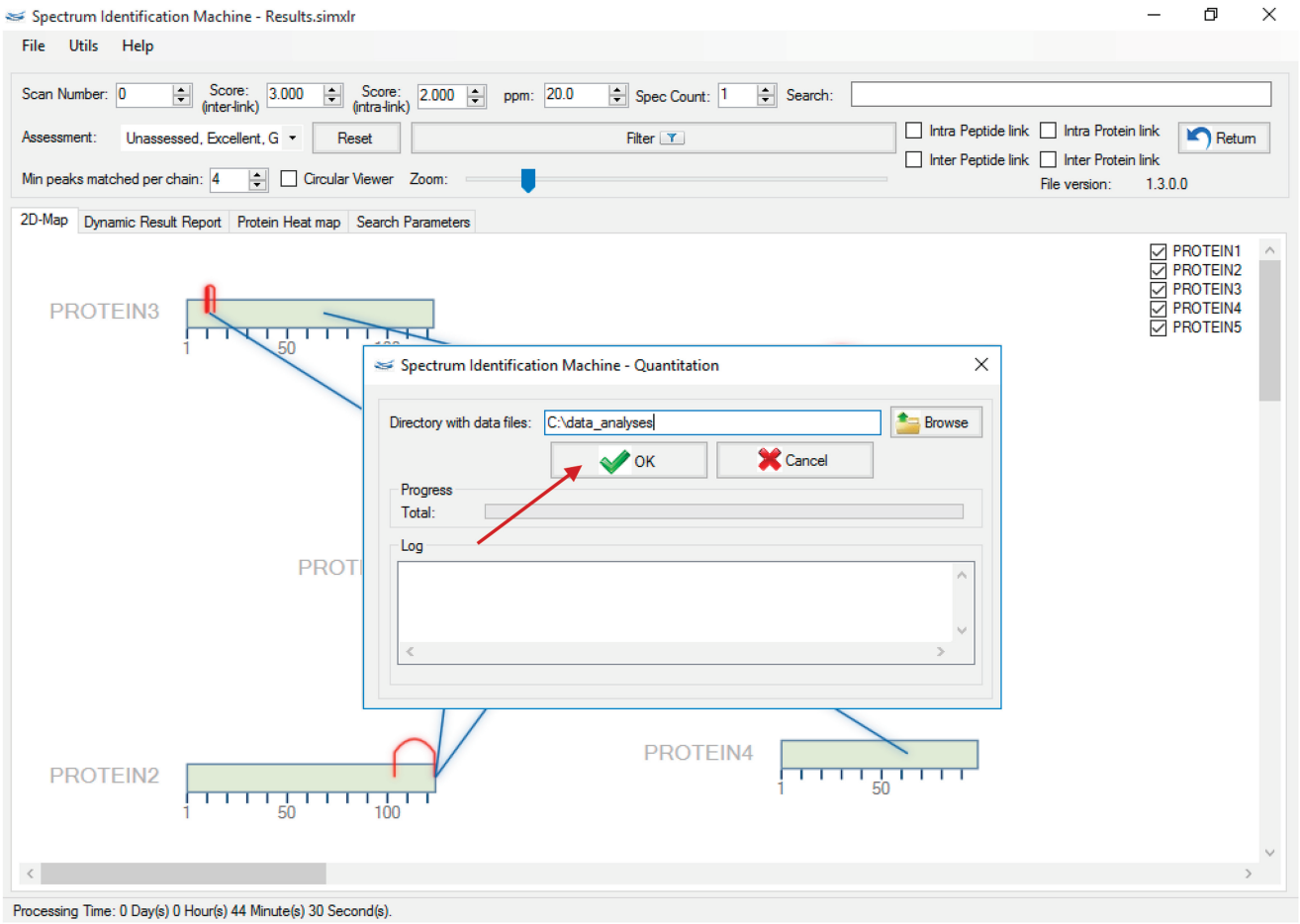

Figure 17 | Calculating the XICs. A directory containing the RAW files is required to calculate the XICs.

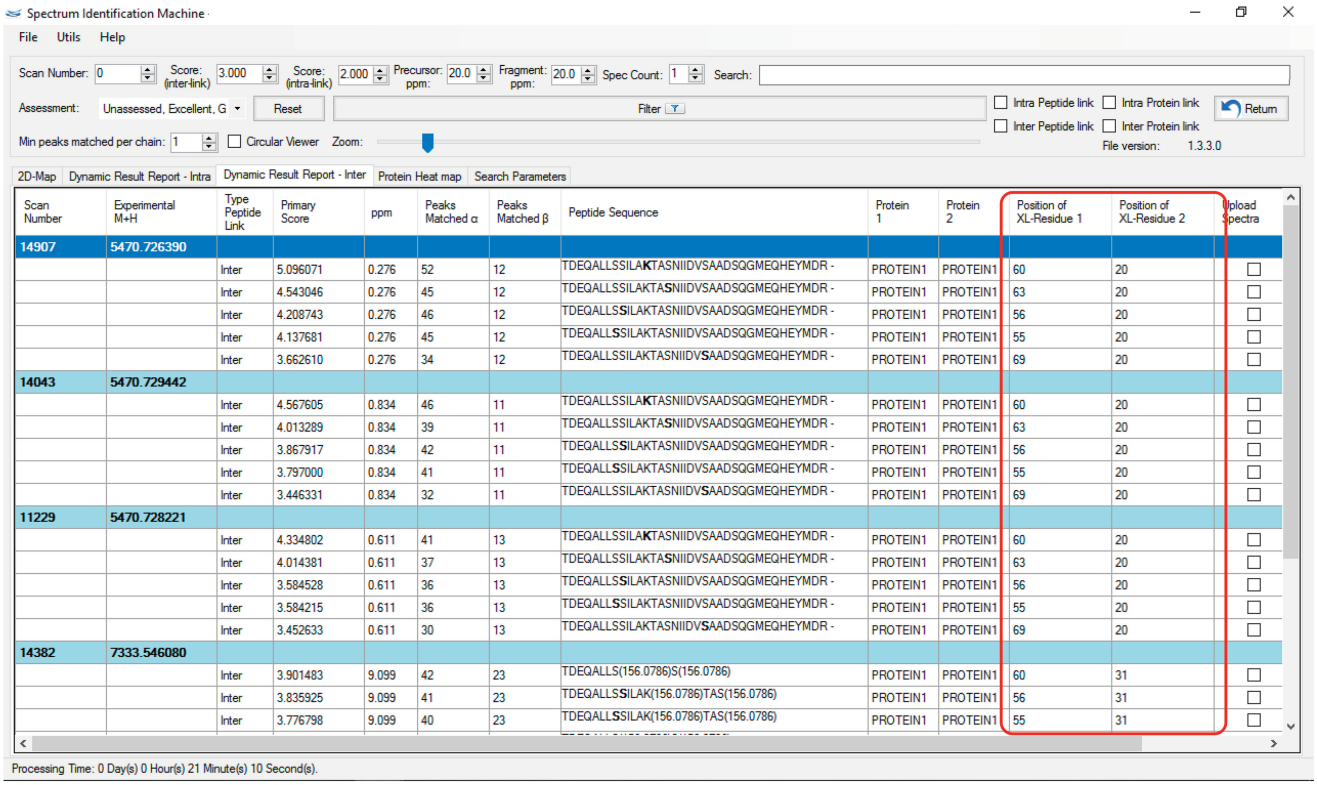

Figure 18 | Spatial constraints from identified residues. SIM-XL can generate a PyMOL script from spatial constraints obtained from the Dynamic Result Report.

Generation of the PyMOL script according to identified spatial constraints ● **TIMING not more than 10 min**

119| SIM-XL enables generation of a PyMOL script from spatial constraints obtained from the Dynamic Result Report (**Fig. 18**). Use these data for obtaining structural information such as the folding of each protein (**Fig. 19**), the topology of the protein complex, and the protein–protein interactions map.

120| Export the PyMOL script by going to the 'File' menu on Results Browser, pressing 'Export Results', and pressing the 'PyMOL Script' option (or pressing ALT + P). A new window will be opened (**Fig. 20**).

121| Set the parameters (font size and color, dash width, dash color and dash length, and the type of target atom: carbon alpha or carbon beta).

122| Select the PDB file for each identified protein, and set, optionally, a sequence offset.

▲ **CRITICAL STEP** The PDB file is mandatory.

123| Click the 'Go' button to generate the script. All data from the Dynamic Result Report will be considered.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**. If the question is not covered in this protocol, please refer to the discussion forum on our SIM-XL Google group, made available through the project's website at http://patternlabforproteomics.org/sim-xl, or by selecting 'Help', and then 'Discussion forum', in the SIM-XL graphical user interface.

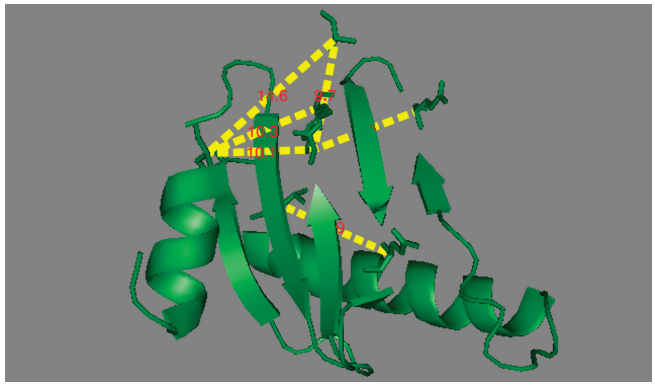

Figure 19 | Protein folding obtained from the identified spatial constraints. The results from the Dynamic Result Report can be used to obtain important structural information, such as the folding of each protein.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
SIM-XL setup	SIM-XL fails to install	.NET Framework may not be properly installed	It is necessary to update to .NET 4.6 or later. The latest version of the .NET Framework can be downloaded from Microsoft's website
9	The protein fails to express itself as compared with the noninduced control	Expression plasmid lacking expression machinery IPTG was either not added or not effective	Check the sequence of the expression plasmid Prepare a fresh stock of IPTG
43	Only intramolecular cross-links are identified; i.e., cross-links are identified only within the wild-type protein and within the isotopically labeled protein	The wild-type and isotopically labeled species must be completely unfolded when mixed together for equivalent participation of each species in a homodimeric complex. If the homodimers were not completely unfolded and dissociated before mixing, the wild-type and isotopically labeled monomers were not free to participate in an intermixed homodimer complex	Dilute the wild-type and isotopically labeled protein to a lower concentration during the unfolding step
50	No cross-linked dimer appears on an SDS gel	The protein or cross-linking reagent concentration is too low The protein is misfolded The protein is in an incompatible buffer	First, try increasing the final concentration of the cross-linking reagent, and then try increasing the protein concentration Refolding the protein is critical for the formation of the physiologic dimer complex. One strategy is to dialyze out the denaturing buffer into lower concentrations of the denaturing agent before final removal; i.e., first dialyze the denaturing buffer into 2 M guanidine followed by 1 M guanidine followed by PBS We recommend running a positive control in the same buffer using a protein known to easily form oligomeric complexes, such as BSA
85	The message 'Please specify a database before editing the protein labeling' is displayed	No protein database has been selected in the 'Sequence database' field	First, select a protein database in the 'Sequence database' field, and then click on the 'Edit' button in the 'Homodimer Analyses' group box
87	The markup is not inserted into the table	The same markup already exists in the table or a different one exists for the same protein. Only one label mode is allowed for a protein: Atom or Residue	Remove the protein label with the conflicting markup
94	The comet tries to read Thermo RAW files and displays the message, 'Retrieving the COM class factory for component with CLSID failed due to the following error: 80040154 Class not registered.'	MSFileReader is not installed	Install MSFileReader, available on Thermo's website

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
97	The message 'There are too many spectra to be visualized. Increase the score filter to obtain a 2D-Map!' is displayed	The 2D-Map does not support plotting too many cross-links	It is necessary to apply more stringent-quality filters. In addition, if there are many identified proteins, we recommend enabling the 'Circular Viewer' mode
	Scores of cross-linked peptides are low	MS/MS spectra are poor because of insufficient mass amounts of the proteolytic enzyme MS/MS spectra are poor because of a cross-linker inhibiting the digestion MS/MS spectra are poor because of insufficient fragmentation of the parent ion of the cross-linked peptides	Cross-link ID scores depend on the quality of the MS/MS spectra. First, try increasing the amount of the proteolytic enzyme Enzymatic digestion can be hindered by cross-linkers. For instance, Trypsin cleaves after lysines except when a cross-linker is present. Try decreasing the cross-linker/protein ratio to free up some lysines for digestion while still capturing the dimer Cross-linked peptides may require optimization of the MS collision energy. The mass-to-charge ratios are higher than for free peptides, so try increasing the CE to get better fragmentation of the parent ion
113	The message 'No data to visualize' is displayed	No peaks overlap between the experimental and the theoretical mass spectra	Change the parameters on the 'Spectrum Prediction Parameters' and/or 'Homodimer Analyses' tabs, or change the experimental ions on the 'Spectrum Plotting Parameters' tab

● TIMING

Steps 1–12, bacterial cell growth and protein expression: 3 d
Steps 13–39, protein purification: 4–5 d
Steps 40–44, preparation and mixing of the wild-type and isotopically labeled protein: 1 d
Step 45, refolding of the wild-type and isotopically labeled protein mixture: 1–2 d
Steps 46–52, cross-linking and isolation of the homodimer complex: 48 h
Steps 53 and 54, desalting of the homodimer: 1 d
Steps 55–58, digestion and preparation of the homodimer for mass spectrometer analysis: ~16 h
Steps 59–65, LC–MS/MS analysis: 90 min–3 h, depending on instrument gradient
Steps 66–93, setup of parameters for analysis of data on SIM-XL: <5 min
Steps 94–96, performance of the search: from a few minutes up to a couple of hours
Steps 97–118, evaluation of the results: variable, up to2 h
Steps 119–123, generation of the PyMOL script according to the identified spatial constraints: <10 min
Note: For Steps 94–96), search time can range from a few minutes up to a couple of hours, varying mostly with sample complexity, the number of variable post-translational modifications, the number of spectra, resolution, PPM, mass spectrometer, LC gradient length, and so on, as well as with the computer's configuration. We use as an example the computational burden of an experiment performed by us that evaluated a crystal structure of C-terminal-truncated apolipoprotein A-I in solution that reveals a structural dynamics related to lipid binding. In this case, the database file had two protein sequences, and SIM-XL analyzed 11,670 tandem mass spectra. It short-listed 24 intramolecular links with scores greater than 2.0 and 56 intermolecular links with scores greater than 3.2.

ANTICIPATED RESULTS

This protocol stems from our previous work that used XL-MS, isotopic labeling, SIM-XL, and small-angle X-ray scattering (SAXS) to define structural differences between crystallized and in-solution forms of dimeric apoA-I[12], the most abundant protein among high-density lipoproteins. Our work demonstrated the power of the isotopic-labeling strategy for probing the structure of a homodimer, and allowed us to show that the structure of apoA-I in solution can vary from that predicted by

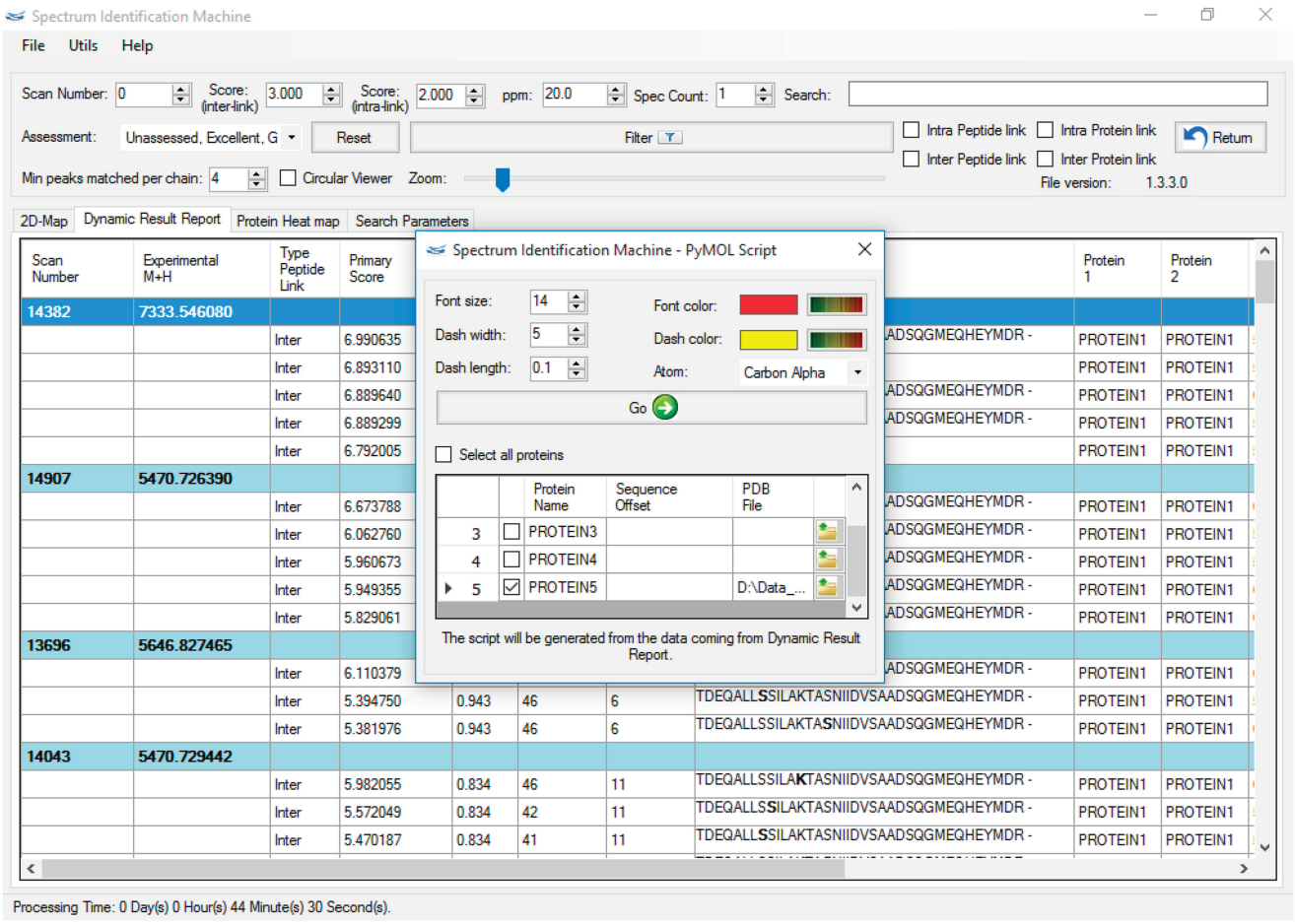

Figure 20 | Exporting the PyMOL script. A script can be generated from the Dynamic Result Report and can be customized according to some parameters, e.g., the font size and color, and the dash's style. The Protein Data Bank (PDB) file for each identified protein is required as well.

XRC. This allowed us to propose new hypotheses for apoA-I structural dynamics and interaction with lipid. Herein lies the main advantage of the XL-MS approach, i.e., the data reflect the solution state of a target protein with functionally relevant structural dynamics included. With appropriate modification, we believe that this protocol can be applied to virtually any protein that can be isotopically labeled and mixed with its unlabeled counterpart in a physiologically relevant way. The most obvious applications are to soluble interacting proteins that can be directly combined in the test tube (similar to our example), but one can envision extending the approach to membrane proteins residing in vesicular or micellar systems that preserve their function. Using these approaches, investigators can ask important biological questions such as (i) how many molecules of a given protein are involved in a homo-oligomeric complex? (ii) What are the conformational consequences/changes invoked by the interaction?, and perhaps most importantly, (iii) can this interaction be promoted or inhibited to impact a biological or pathological process?

As good as SIM-XL and other software tools are becoming, we believe that it is still important to do critical and manual assessment of the results. **Figure 11** demonstrates the 2D interaction map between the heavy and light monomers of apoA-I; the intra-links (red lines) from each monomer should corroborate with its heavy/light counterpart. Even though the intra-link pattern is similar when comparing both monomers, it is common for each monomer to have unique intra-links, most likely because of undersampling. In addition, note that the interlinks occur within the same regions (domains) as intra-links. Moreover, different thresholds were applied when accepting inter- and intra-links, as interlinks usually have higher similarity scores, as their numbers of amino acids are usually higher; we recommend using 3.0 and 2.5 for inter- and intra-links, respectively. We argue that, different from proteomics, in which a threshold of false discovery is tolerable, the incorporation of a single false-positive cross-link is unacceptable because it may have major impacts on subsequent protein-modeling analyses. The spectrum information is accessible by right-clicking on a link (red or blue lines in **Fig. 11**), leading to the contents of **Figure 13**. As previously described, each spectrum can be assigned a user assessment ranging from excellent to poor; this enables the retrieval of high-quality user-assessed spectra when building a model. The number of intra-links in homodimer experiments is usually low, enabling them to be evaluated within <1 h. Assessing the alpha and

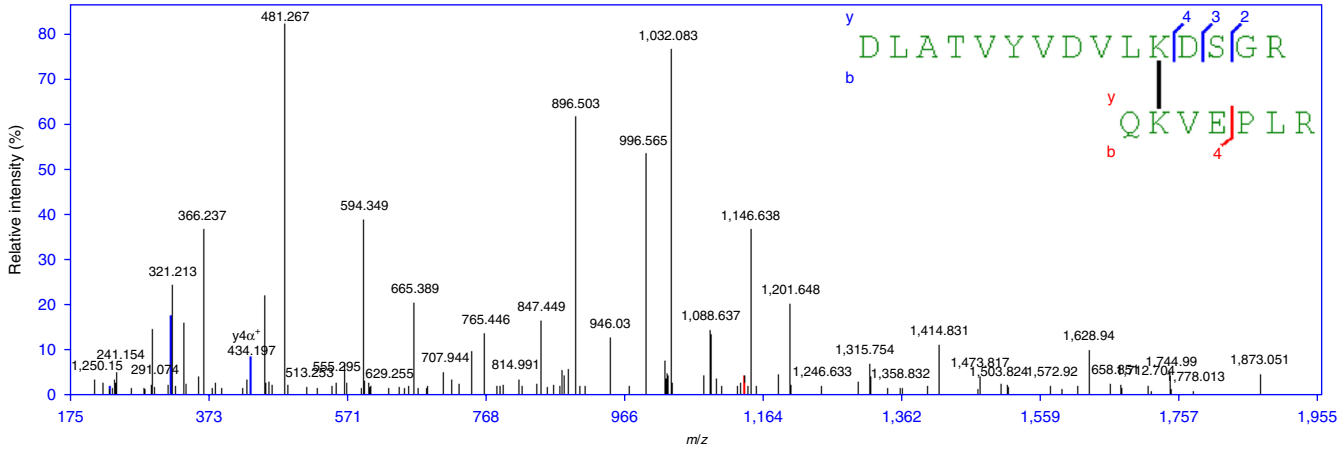

Figure 21 | A poor identification. The most intense peaks are unassigned, and there are no *y* or *b* ions to the left and right of the cross-linker sites of the alpha chain, characterizing an unconfident identification.

beta chain coverage is critical; this can be achieved by clicking on the 'Peptide Annotation' tab. We replicated the contents of this tab on the spectrum view for demonstration purposes (**Fig. 13**). A confident identification should have a reasonable coverage (>50%) on both chains. The SIM-XL algorithm, by default, will discard identifications containing chains with <4 amino acids. **Figure 13** exemplifies what we consider a confident identification, as the most intense peaks are assigned within <20 ppm. On the other hand, **Figure 21** demonstrates an identification with a low score (0.44) that was discarded in our analysis; most intense peaks remain unassigned, and there are no *y* or *b* ions to the left and right of the cross-linker site of the alpha chain. In addition, MS1 can be accessed within the raw data file to manually assess the intensity and accuracy of the parent ion picked by the MS for fragmentation. SIM-XL provides the *m/z* value, precursor charge, and retention time in the 'Results' tab, which aids in quick location and identification of the mass of interest. The appearance and intensity of the 'phantom' peak on an isotopically labeled cross-linked peptide (**Supplemental Fig. 1**) can also be monitored. If the 'phantom' peak is ≥25% of the fully labeled parent ion, the number of isotopic possibilities should be increased to five or more when searching the data file with SIM-XL (Step 92).

The raw data described in this protocol, together with the SIM-XL results were deposited in PRIDE[32] under identifier PXD006574. The spectra that were considered to be high-quality identifications are marked as 'Excellent' in SIM-XL's user assessment and thus can be quickly retrieved. This should enable readers to replicate our results and serve as a standard for applying this protocol to their proteins of interest. Finally, we make a user discussion forum available at the software's website http://patternlabforproteomics.org/sim-xl.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS V.C.B. acknowledges a FAPERJ BBP grant, as well as support from CNPq and CAPES. P.C.C. acknowledges the Fundação Araucária Universal/Jovem Pesquisador Grant and support from PAPES VII and Universal CNPq. D.B.L. and J.C.-R. acknowledge the Centre National de la Recherche Scientifique (CNRS) for financial support. F.C.G. and M.F. acknowledge support from FAPESP (grants 2014/17264-3 and 2012/10862-7) and CNPq. The mass spectrometry methods were established in the UC Proteomics Laboratory on the Sciex 5600 + TripleTOF system, funded in part through an NIH-shared instrumentation grant (S10 RR027015-01; K.D. Greis, principal investigator). This work was also supported by an American Heart Association postdoctoral fellowship grant (16POST27710016 to J.T.M.) and by funding from the National Institutes of Health Heart, Lung and Blood Institute (R01 GM098458 and P01 HL128203 to W.S.D.).

AUTHOR CONTRIBUTIONS D.B.L., P.C.C., F.C.G., and V.C.B. have participated in the development of the SIM-XL project since its beginning in 2015. D.B.L., J.T.M., W.S.D., P.C.C., and F.C.G. participated in developing the SIM-XL module tailored for homodimers. J.T.M., W.S.D., J.C.-R., J.S.G.F., J.M., and T.A.C.B.S. participated in describing the experimental methodology. D.B.L., P.C.C., and V.C.B. participated in writing the data-analysis methodology. M.F. participated in developing the SIM-XL software and in describing the computational methodology. All authors participated in writing the Introduction and Anticipated Results sections. D.B.L., J.C.-R., and P.C.C. created the supplementary video. All authors read and approved the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Marianayagam, N.J., Sunde, M. & Matthews, J.M. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–625 (2004).
2. Brannigan, J.A. *et al.* A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**, 416–419 (1995).
3. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
4. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **25**, 663–682 (2006).
5. Lima, D.B. *et al.* SIM-XL: a powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **129**, 51–55 (2015).
6. Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540 (2011).
7. Schmidt, C. & Robinson, C.V. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **9**, 2224–2236 (2014).

8. Sinz, A., Arlt, C., Chorev, D. & Sharon, M. Chemical cross-linking and native mass spectrometry: a fruitful combination for structural biology. *Protein Sci.* **24**, 1193–1209 (2015).
9. Jiang, G., den Hertog, J. & Hunter, T. Receptor-like protein tyrosine phosphatase alpha homodimerizes on the cell surface. *Mol. Cell. Biol.* **20**, 5917–5929 (2000).
10. Vashisth, H. & Abrams, C.F. Docking of insulin to a structurally equilibrated insulin receptor ectodomain. *Proteins* **78**, 1531–1543 (2010).
11. Walker, R.G. *et al.* The structure of human apolipoprotein A-IV as revealed by stable isotope-assisted cross-linking, molecular dynamics, and small angle X-ray scattering. *J. Biol. Chem.* **289**, 5596–5608 (2014).
12. Melchior, J.T. *et al.* An evaluation of the crystal structure of C-terminal truncated apolipoprotein A-I in solution reveals structural dynamics related to lipid binding. *J. Biol. Chem.* **291**, 5439–5451 (2016).
13. Iglesias, A.H., Santos, L.F.A. & Gozzo, F.C. Identification of cross-linked peptides by high-resolution precursor ion scan. *Anal. Chem.* **82**, 909–916 (2010).
14. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.* **9**, 120–137 (2014).
15. Holding, A.N., Lamers, M.H., Stephens, E. & Skehel, J.M. Hekate: software suite for the mass spectrometric analysis and three-dimensional visualization of cross-linked protein samples. *J. Proteome Res.* **12**, 5923–5933 (2013).
16. Du, X. *et al.* Xlink-Identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. *J. Proteome Res.* **10**, 923–931 (2011).
17. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A.J.R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8**, 15473 (2017).
18. Liu, F., Rijkers, D.T.S., Post, H. & Heck, A.J.R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
19. Borges, D. *et al.* Using SIM-XL to identify and annotate cross-linked peptides analyzed by mass spectrometry. *Protoc. Exch.* http://dx.doi.org/10.1038/protex.2015.015 (2015).
20. Borges, D. *et al.* Effectively addressing complex proteomic search spaces with peptide spectrum matching. *Bioinforma. Oxf. Engl.* **29**, 1343–1344 (2013).
21. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318 (2008).
22. Bai, X. *et al. Nature* **525**, 212–217 (2015).
23. Keskin, O., Tuncbag, N. & Gursoy, A. Predicting protein-protein interactions from the molecular to the proteome level. *Chem. Rev.* **116**, 4884–4909 (2016).
24. Kang, S. *et al.* Synthesis of biotin-tagged chemical cross-linkers and their applications for mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 1719–1726 (2009).
25. Chambers, M.C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
26. Rosano, G.L. & Ceccarelli, E.A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
27. Becker, G.W. Stable isotopic labeling of proteins for quantitative proteomic applications. *Brief. Funct. Genomic Proteomic* **7**, 371–382 (2008).
28. Kapust, R.B., Tözsér, J., Copeland, T.D. & Waugh, D.S. The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294**, 949–955 (2002).
29. Tubb, M.R., Smith, L.E. & Davidson, W.S. Purification of recombinant apolipoproteins A-I and A-IV and efficient affinity tag cleavage by tobacco etch virus protease. *J. Lipid Res.* **50**, 1497–1504 (2009).
30. Markwell, M.A.K., Haas, S.M., Bieber, L.L. & Tolbert, N.E. A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* **87**, 206–210 (1978).
31. Carvalho, P.C. *et al.* YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics* **25**, 2734–2736 (2009).
32. Vizcaíno, J.A. *et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).