

PROBING THE SOLUTION STRUCTURE OF FACTOR H USING HYDROXYL RADICAL PROTEIN FOOTPRINTING AND CROSS-LINKING

**Anna Baud^{1,2}, Florence Gonnet^{1,2}, Isabelle Salard^{1,2}, Maxime Le Mignon^{1,2}, Alexandre Giuliani³,
Pascal Mercère³, Bianca Sclavi⁴, Régis Daniel^{1,2}**

From ¹CNRS, UMR8587, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, 91025 Evry, France, ²Université Evry-Val-d'Essonne, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, 91025 Evry, France, ³Synchrotron Soleil, BP 48 91192 Gif-sur-Yvette CEDEX, France, ⁴CNRS, UMR 8113, LBPA, ENS Cachan, Cachan, 94235, France

*Running title: Solution structure of free and C3b-bound Factor H

Address correspondence to: Dr. Régis Daniel, CNRS, UMR8587, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université Evry-Val-d'Essonne F-91025 Evry, France. Fax: + 33 01 69 47 76 55; E-mail: regis.daniel@univ-evry.fr

Keywords: Factor H, C3b protein, Complement system, hydroxyl radical protein footprinting, mass spectrometry, cross-linking.

ABSTRACT

The control protein Factor H is a crucial regulator of the innate immune complement system, where it is active on host cell membranes and in the fluid phase. Mutations impairing the binding capacity of Factor H lead to severe autoimmune diseases. Here, we studied the solution structure of full-length Factor H, in its free state and bound to the C3b complement protein. To do so, we used two powerful techniques, hydroxyl radical protein footprinting and chemical cross-linking coupled with mass spectrometry, to probe the structural rearrangements and to identify protein interfaces. The footprint of C3b on the Factor H surface matches existing crystal structures of C3b complexed with the N- and C-terminal fragments of Factor H. In addition, we revealed the position of the central portion of Factor H in the protein complex. Moreover, cross-linking studies confirmed the involvement of the C-terminus in the dimerization of Factor H.

INTRODUCTION

Factor H (FH) is a key player in the regulation of the complement system, an essential component of human innate immunity. FH down-regulates the permanently active alternative pathway of the complement system to a basal level, thus preventing deleterious damage to host cells and tissues [1]. FH is an abundant circulating glycoprotein (0.1-0.5 mg/mL) that targets the membrane-bound C3b protein, a hub of the complement cascade. Active in the fluid phase, FH is the most important plasma inhibitor of the complement system. FH is endowed with an extraordinary diversity of modes of action, including competition with Factor B for binding to C3b, involvement as a cofactor in the cleavage of C3b by the protease I, and acceleration of the decay of the C3 convertase C3bBb. Many mutations in FH that impair these activities have been identified and are associated with severe diseases such as age-related macular degeneration, atypical hemolytic uremic syndrome and C3 glomerulopathies [2]. The multiple pathways of action of FH rely on its broad binding capacity that can be seen as a multi-ligand platform for proteins such as C3b and the C-reactive protein, as well as for polyanionic carbohydrates such as heparin, sialic acid clusters and glycosaminoglycans (e.g. heparin sulfate) [3–8]. The structural features of FH promote this diversity of binding partners and mechanisms; numerous studies over the past decade have thus attempted to decipher the structure of FH [9]. However, given its molecular weight (MW 155 kDa) and its flexibility, FH resists conventional biostructural characterization methods, such as NMR and X-ray crystallography. The

flexibility of FH results from its unique structural organization comprising peptide linkers of 3 to 8 amino acids that interconnect 20 homologous globular modules of about 60 residues in length and called short consensus repeats (SCR). This organization leads to a molecule that resembles a string of beads and that can adopt many different conformations. Most of the previous structural studies of FH at the molecular level have been carried out only on individual or a few concatenated SCR modules (Protein Data Bank (PDB): 2RLP, 2RLQ, 1HFH, 2UWN, 2KMS, 4B2R, 4B2S, 3SW0, 4K12) [10–15]. The co-crystal structures of the N-terminal FH SCR modules [1–4] with C3b [16] (PDB: 2WII), and the C-terminal FH SCR modules (19–20) with C3d [17–19] (PDB: 3OXU) have provided important clues on the decay of C3 convertase and the co-factor I activities of FH, as well as its role in the self/non-self discrimination. These two major functional terminal regions of FH have been recently engineered to form a minimal size, functional FH construct [20]. However, the only studies on the full-length molecule employed electron microscopy, X-ray and neutron scattering, and analytical ultracentrifugation, providing low to medium resolution data [21–23]. These studies have revealed multiple fold-back structures in FH and its capacity to self-associate, consistent with its flexible structure [24,25].

Herein, we probed the structure of the full-length, intact FH molecule in solution by chemical cross-linking and X-ray synchrotron-mediated hydroxyl radical footprinting (HRPF) combined with high-resolution mass spectrometry (MS). HRPF combined with MS has recently emerged as a powerful technique that provides single amino-acid resolution on large proteins and macromolecular assemblies that are otherwise inaccessible *via* conventional approaches [26–28]. The hydroxyl radical ($\cdot\text{OH}$) is an efficient labeling reagent that can be generated by the radiolysis of water molecules upon irradiation of the protein sample by a synchrotron X-ray beam. $\cdot\text{OH}$ reacts on a fast timescale with a wide range of amino acids, leading to their oxidation and to a mass increase of +16 Da (or a higher multiple) corresponding to the incorporation of oxygen [29].

In this study, the full size FH was irradiated in solution and the oxidized amino acids were identified. The quantitative determination of the oxidation level using LC-MS/MS analysis led to the definition of the solvent-accessible surfaces on the 20 SCR modules of FH. Furthermore, given that the footprinted regions can vary with conformational rearrangement, ligand binding or biomolecular interactions, we therefore also investigated the solvent-accessible surfaces of the FH molecule in a complex with C3b. We assessed how the different SCR modules are involved in the interaction with C3b, in an attempt to shed light on the FH structural organization in the FH-C3b complex.

EXPERIMENTAL PROCEDURES

Materials and Reagents

All materials were of analytical grade. DL-dithiothreitol (DTT), iodoacetamide, ammonium bicarbonate, phosphate-buffered saline (PBS; 10 mM phosphate buffer, 2.7 mM potassium chloride and 137 mM sodium chloride, pH 7.4) and urea were all purchased from Sigma-Aldrich. Cross-linking reagents BS3-d0 and BS3-d4 were purchased from Thermo Scientific. Acetonitrile was obtained from Fluka, formic acid from Normapur and sequencing-grade modified porcine trypsin from Promega. Complement system proteins C3b and FH purified from human plasma were obtained from Calbiochem (MerckMillipore, Darmstadt, Germany). All buffers and solutions were prepared using ultra-pure 18 M Ω water (MilliQ).

Synchrotron X-ray footprinting of Factor H

Synchrotron irradiation of FH was performed using the Metrology beamline at the SOLEIL synchrotron facility (Saint Aubin, France). The experiment was conducted on the hard X-ray branch of the beamline that emits monochromatic photons in the 100 eV to 40 keV energy range, and also allows access to the synchrotron white radiation. This latter configuration was used for the X-ray footprinting experiments.

First, 2 μM FH was prepared in PBS. The QFM-400 Micro-Volume Quench Flow (Bio-Logic SAS) stopped-flow system was used to deliver the samples in front of the beam and regulate the irradiation time. Then, 14 μL of the 2 μM FH solution was loaded into the instrument and mixed with 0.01 M PBS or C3b solution at a 1:1 molar ratio for 1 s in the mixing chamber. The reaction mixture was delivered to the quartz capillary placed in front of the beam and FH samples in the absence or

presence of C3b were exposed to various radiation times ranging from 0 to 150 ms. Computer-controlled Quench Flow Apparatus enabled immediate quenching of the remaining free hydroxyl radicals with 20 μ L of 100 mM methionine, within 10 ms from X-ray exposure. All experiments were carried out at 21°C and the irradiated samples were stored at -80°C.

Chemical cross-linking of Factor H

Cross-linking of FH was performed with the homobifunctional sulfo-NHS ester cross-linker reagent bis(sulfosuccinimidyl) suberate (BS3-d0 and -d4), characterized by a 11.4 Å spacer arm (Thermo Scientific). BS3 (70 mM in water) was added to FH in a 10:1 molar ratio. After 1 h of incubation on ice, the reaction was stopped with the addition of cold ammonium bicarbonate (50 mM, final concentration). Samples were then immediately mixed with denaturing Laemmli buffer and heated to 95°C for 5 min for subsequent SDS-PAGE analysis. SDS-PAGE was carried out on 6% Tris-glycine polyacrylamide gels following the standard method described by Laemmli [30]. After migration, the gel was stained with Instant Blue™ (Expediton).

In-solution and in-gel proteolytic digestions

Irradiated FH samples in PBS were vacuum-dried and diluted in 14 μ L ultra-pure water. Prior to reduction, 3 μ L of 8 M urea in 400 mM ammonium bicarbonate was added to the sample. Samples were then reduced with the addition of DTT (4.2 mM, final concentration) for 45 min at 37°C, and then alkylated with iodoacetamide (7.6 mM, final concentration) for 30 min in the dark. To dilute the urea, 14 μ L of ultra-pure water was added before adding trypsin (1:25, w/w). Trypsin digestion was carried out overnight at 37°C and was terminated by freezing the samples. Prior to nano LC-MS/MS analysis, protein digests were vacuum-dried and re-suspended in 14 μ L of 98% H₂O, 2% CH₃CN, 0.1% HCOOH (solvent A).

Cross-linked samples separated by SDS-PAGE underwent in-gel digestion. The protein bands were excised and destained with 30% ethanol, reduced with the addition of 50 μ L of 10mM DTT and alkylated with 50 μ L of 55 mM iodoacetamide. In-gel digestion was carried out overnight at 37°C by adding 20 μ L of trypsin at a concentration of 12.5 ng/ μ L. The resulting peptides were extracted by subsequent incubation of the gel pieces in the following extraction solutions: 60% CH₃CN, 0.1% HCOOH; 80% CH₃CN, 0.1% HCOOH; and 100% CH₃CN, 0.1% HCOOH. Supernatants were recovered and vacuum-dried. Pellets were re-suspended in 10 μ L of solvent A prior to nano LC-MS/MS analysis.

Nano liquid chromatography tandem mass spectrometry of irradiated samples

Nano LC-MS/MS analyses were performed on a Dual Gradient Ultimate 3000 chromatographic system (Dionex) coupled to a LTQ-Orbitrap™ XL mass spectrometer (Thermo-Fisher Scientific) equipped with a nanospray source operating in positive ionization mode.

First, 2 μ L of sample was loaded onto a C18 pre-column (Acclaim PepMap C18, 5 mm length x 300 μ m I.D., 5 μ m particle size, 100Å porosity, Dionex) and desalted for 5 min with solvent A. After this step, peptide separation was carried out on a C18 capillary column (Acclaim PepMap C18, 15 cm length x 75 μ m I.D., 3 μ m particle size, 100 Å porosity, Dionex) with a gradient starting at 100% solvent A, ramping to 70% solvent B (80% CH₃CN, 20% H₂O, 0.1% HCOOH) in 30 min, then to 100% solvent B over 5 min (held for 2 min) and finally decreased to 100% solvent A in 10 min.

A parent list was created using MS-Digest (Protein Prospector, UCSF) with the following parameters: trypsin digestion, 1 missed cleavage, peptide mass: 400-1600. Constant modifications: carbamidomethyl; variable modifications: oxidation Met, report multiple charges. The obtained parent list was then manually modified by adding +15.9949 Da to the peptide mass generated by MS-Digest. A parent list including double oxidation (+31.9898 Da) was also created, but only a few multiply oxidized sites were identified (data not shown), probably due to the applied footprinting conditions. For this reason, only single oxidation (+15.9949 Da) modifications were monitored in this study.

One series of MS analyses (ITMS² analyses) consisted of acquisition cycles composed of one MS scan in the Orbitrap analyzer (profile mode; resolution, 60000; m/z range, 400-2000) followed by five MS/MS scans (CID fragmentation and detection in the linear ion trap analyzer; centroid mode; isolation width 3 Da) triggered on the five most intense species from the parent list detected in the preceding MS scan, at a normalized collision energy of 35%. To exclude multiple MS/MS analyses

for the same peptide, dynamic exclusion of 45 s was applied with a repeat count of 1, repeat duration of 30 s and an exclusion mass width of +/- 7ppm. Singly charged ion species were excluded from fragmentation.

Footprinting data analysis

MS data were processed using ProtMapMS software [31] (Neoproteomics, USA), and were also manually validated in raw spectra using Xcalibur version 2.1.0. (Thermo Scientific). The specific sites of modification for each peptide were assigned using a Mascot server and manually validated.

The proportion of unmodified peptides defined as the ratio of unmodified peptides to the total amount of peptides (modified and unmodified) was calculated for each exposure time.

The rate of oxidation of a given peptide was determined from the dose-response plots of the proportion of unmodified peptides as a function of X-ray exposure time (ms). The dose-response plots were fitted to the following equation: $y=e^{-kt}$, where k is the first-order rate constant and t is the exposure time. Then, the dose-response curves generated for each FH peptide were used to compare the degree of oxidation of FH in both states (unbound and bound to C3b) and calculate the degree of protection.

Solvent-accessible surface areas (SASA) of modified FH residues were calculated using GETAREA software [32], based on the available in-solution and crystal structural data of modules and full-length FH (details given in Table 1).

Nano liquid chromatography tandem mass spectrometry of cross-linked samples

Nano LC-MS/MS analyses of the cross-linked samples were performed in the same chromatographic conditions as those used for the irradiated samples, except only 5 μ L of sample was injected.

One series of MS analyses (ITMS² analyses) consisted in acquiring cycles composed of one MS scan in the Orbitrap analyzer (profile mode; resolution, 60000; m/z range, 400-2000) followed by 10 MS/MS scans (CID fragmentation and detection in the linear ion trap analyzer; centroid mode; isolation width, 3 Da) triggered on the 10 most intense species detected in the preceding MS scan, at a normalized collision energy of 35%. To exclude multiple MS/MS analyses for the same peptide, dynamic exclusion of 20 s was applied with a repeat count of 1, repeat duration of 30 s and an exclusion mass width of +/- 7ppm. Singly and doubly charged ion species were excluded from fragmentation.

Cross-linking data analysis

Bioinformatics analysis of cross-linking spectra was conducted as reported by Redeker *et al.* [33]. RAW data acquired on LTQ-Orbitrap mass spectrometer were converted to .MZXML files using the MSConvert software [34]. The resulting .MZXML files were processed with the DeconTools AutoProcessor (<http://omics.pnl.gov/software/decontools-decon2ls>) to de-isotope mass spectra. The resulting .csv files were analyzed using Viper software [35], to identify the d0/d4 peptide pairs corresponding to a mass difference of 4.0247 Da as a signature of peptides that reacted with at least one molecule of BS³-d0/d4. A list of peptide pairs with a maximum mass tolerance of 10 ppm was generated. Mass deviation and peptide elution time were used to manually filter the list of peptide pairs, corresponding to candidate cross-linked peptides or mono-linked or loop-linked peptides. The list of light and heavy precursor masses was used to identify the peptides of BS³-d0/d4 using the xQuest software [36]. MS/MS spectra of the tracked cross-linked peptides were further analyzed manually to confirm their identification.

RESULTS

Hydroxyl radical protein footprinting of the full-size Factor H molecule

The SASA of the full-size FH molecule, either alone or in presence of the C3b protein, was probed using HRPD with the white X-ray beam of the Metrology beamline at the SOLEIL synchrotron facility (France). Irradiated samples underwent tryptic proteolysis and the resulting oxidized FH peptides were identified and quantified by nano LC-nano electrospray ionization (ESI) MS/MS.

The sequence coverage of irradiated FH using MS/MS was on average 73% throughout the entire molecule, providing for a robust analysis of vast portions of the FH molecule (Figure 1). Similar sequence coverage was obtained for non-irradiated FH (data not shown), indicating that X-ray exposure did not damage the FH structure, in agreement with a previous study showing that HRFP does not affect protein structure [37]. The integrity of FH after irradiation was also confirmed by MALDI-TOF analysis (data not shown). Identification of the oxidation sites was based on the detection of peptides with a change in mass due to the hydroxyl radical reaction with an amino acid side chain. Because the MS/MS method targets only the peptides with a mass shift of + 15.9949 Da resulting from the incorporation of oxygen, no other modifications caused by hydroxyl radicals were detected.

The 23 oxidized residues of FH were identified in a reproducible manner (Table 1), based on both automatic and manual investigations of MS/MS spectra. In order of frequency, the affected amino acids were Leu, Ile > Met, Tyr, His > Phe, Val, in agreement with the reactivity towards hydroxyl radicals reported in the literature [38]. They were distributed throughout almost all of the SCR modules, except SCR modules 4, 6, 15 and 19, with the main occurrences in SCR modules (1-2) (each with three oxidized residues) and (7-8 and 14) (each with two oxidized residues). Representative MS/MS spectra of two oxidized FH peptides (92-109) and (110-127) belonging to the SCR2 module are shown in Figure 2.

The obtained experimental footprinting data were compared with the SASA values of the FH amino acid side chains obtained from existing structural data. Note that no crystallographic structure of full-length FH is available to date, and that only the structural data of individual recombinant FH modules obtained either by NMR or crystallographic studies [11,13–15,39,40] could be used for comparison. For most of the oxidized sites identified on the full-length FH molecule in this study, the level of oxidative labelling was in agreement with the solvent accessibility previously reported for isolated FH modules (see below). We therefore produced a map of oxidized residues on the full-length FH molecule by combining the structural data from the different FH modules (Figure 3).

The following eight oxidation sites were identified in SCR modules (1-5) based on MS/MS analysis: Tyr24 (SCR1), Met46 (SCR1), Val54 (SCR1), Leu80 (SCR2), Leu103 (SCR2), Ile120 (SCR2), His173 (SCR3) and Leu262 (SCR5). SASA values of the seven oxidized residues within modules 1-3 were calculated from the NMR structures of SCR(1-2) (PDB : 2RLP) and SCR(2-3) (PDB ID: 2RLQ). These residues, which reacted with the hydroxyl radicals, were moderately to highly exposed to the solvent in SCR modules (1-2) and (2-3) [39] ($SASA \approx 10$ to 30 \AA^2). No solution or crystallographic structures are currently available to assess the solvent accessibility of SCR5; therefore no SASA value could be determined for Leu262 (SCR5). However, its modification by hydroxyl radicals indicates that Leu262 is likely exposed (*i.e.* $SASA \geq 10 \text{ \AA}^2$) to the solvent in free FH.

Four oxidized residues were detected in SCR modules (7-8): His399 (SCR7), Met414 (SCR7), Tyr449 (SCR8) and Ile474 (SCR8). According to the crystal structure of mutated SCR(6-8) in complex with sucrose-octasulfate (SOS) [13] (PDB: 2UWN), His399 and Met414 are highly exposed to the solvent ($SASA$, 65.42 and 85.70 \AA^2 , respectively). This surface exposition is consistent with oxidation of His399 and Met414 by hydroxyl radicals in unbound FH. Oxidation of these residues is very interesting in regard to their proximity to binding sites for polyanions [7] and C3b [41] in SCR7 and confirms the high reactivity of this region of FH.

Residues Tyr449 and Ile474 in SCR8 were modified by hydroxyl radicals, suggesting that these residues are likely exposed to the solvent in free FH. However, these two residues are buried ($SASA=0.31$ and 7.37 \AA^2 , respectively) in the crystallographic structure of SCR(6-8) complexed with SOS [13] (shown in Figure 4).

Six oxidized residues were detected in SCR modules (10-14), including two oxidized residues in SCR14, indicating solvent accessibility of this central part region of FH for which no crystallographic structure is available to date. Nevertheless, solution structures of modules (10-15) and (10-12) have been recently investigated by analytical ultra-centrifugation, small-angle X-ray scattering (SAXS) and NMR [14,15], indicating a relatively rigid and compact structure. These studies led to the deposition of high-resolution structures of FH SCR modules (10-11), (11-12) and (12-13) in the (PDB: 4B2R, 4B2S and 2KMS, respectively), but the structure of SCR14 remains unknown, due to the difficulties in obtaining recombinant pairs of SCR(13-14) and SCR(14-15) [15]. All oxidized residues in

SCR(10-13) exhibited significant solvent accessibilities ($SASA \geq 30 \text{ \AA}^2$), in agreement with the NMR solution structure data of SCR(10-11), (11-12) and (12-13).

Finally, no oxidation was found in SCR19 containing the C3b binding site. Less than half residues (28 out of 58) were solvent accessible. These potential oxidation sites may be shielded by a structural arrangement of SCR19 between SCR18 and SCR20.

Hydroxyl radical footprinting of Factor H in presence of C3b

FH was also exposed to synchrotron X-rays in presence of its binding partner C3b in a 1:1 molar ratio, and oxidation sites were compared between the two states of FH (alone and with C3b). Sequence coverage of FH irradiated with C3b was identical to that of FH alone (74%), providing for a meaningful comparison of the extent of oxidation between the two states of FH.

MS/MS analysis of the C3b-FH mixture revealed significantly decreased oxidation of FH, with only nine residues showing a typical mass shift of +15.9949 Da. Among them, eight oxidized residues were also detected in FH alone (Val54 in SCR1, Met414 in SCR7, Met497 in SCR9, His746 in SCR13, Phe945 in SCR16, Ile981 in SCR17, Ile1041 in SCR18 and Tyr1159 in SCR20), and only one residue (Met927 in SCR16) was differentially exposed to the solvent in FH with C3b. This significant decrease in hydroxyl radical footprinting likely indicates a shielding effect due to C3b, with multiple interaction sites between the two proteins, and a possible conformational change of FH upon C3b binding.

Table 2 summarizes the oxidized residues detected in FH in the presence of C3b. SASA values were compared between the two states of FH for only two oxidized residues Val54 (SCR1) and Tyr1159 (SCR20) because they belong to previously reported X-ray crystallographic structures of FH[1-4] in complex with C3b [16], and FH[19-20] with C3d [17] (PDB: 2WII and 3OXU, respectively). We noticed a significant decrease in solvent exposure for these two residues, especially for Tyr1159 in SCR20 (from 94.6 \AA^2 to 77.5 \AA^2).

Differences in oxidized sites between FH irradiated with or without C3b are shown in Figure 3. One of the most important variations was observed in the N-terminal domain which comprised eight oxidized residues in SCR(1-3, 5) in FH irradiated alone and only one in presence of C3b. A significant decrease in modified residues was observed in SCR1 (one oxidized residue in the C3b-FH complex instead of three in unbound FH) as well as in SCR2, SCR3 and SCR5 that lacked any modified residues. A major change in oxidation was also found in the central part of FH, which showed only one oxidized residue in SCR7 (compared with two in unbound FH), and the lack of oxidation in SCR8, SCR10, SCR11, SCR12 and SCR14.

The oxidation rates of modified FH peptides were determined from dose-response plots and compared to those obtained in the presence of C3b (Figure 5). Oxidation rate constants decreased for all FH peptides from the FH-C3b incubation, indicating a lower accessibility of these residues in presence of C3b. This lower reactivity is in line with the lower footprinting of FH in presence of C3b as mentioned above. Together, these results suggest that C3b shields these residues and that FH potentially undergoes a conformational change.

For a given modified peptide, the variation in its oxidation rate between two states of the protein can be expressed as a degree of protection determined by the ratio of the slope values of the dose-response curves. According to previous studies [42], a difference in oxidation rates less than 20% (*i.e.* ratio $\approx 0.8-1.2$) is considered as insignificant, suggesting little or no change in solvent accessibility. On the other hand, a decrease in the oxidation rate greater than 20% (*i.e.* ratio ≥ 1.2) indicates a significant decrease in the solvent accessibility of the peptide, which may result from protection through biomolecular protein-protein interactions. On the contrary, a ratio below 0.8 indicates an increase in the solvent accessibility of the investigated zone.

We observed a ratio of ≥ 4.0 for the 15 peptides located in SCR1, SCR3, SCR5, SCR7, SCR8, SCR10, SCR11, SCR12 and SCR14 that were non-oxidized in presence of C3b, thus confirming a high degree of protection upon C3b binding (Figure 6). These peptides are strong candidates for participation in the interface of the C3b-FH complex.

Ratios ranging between 1.2 and 4.0 fold were observed for the eight peptides that were detected in oxidized form in the presence of C3b, indicating moderate protection of these peptides distributed throughout the FH structure. Only one peptide (915-940, located in SCR16) showed increased solvent accessibility and hence increased solvent exposure upon complex formation.

A decrease in the oxidation of residues located in SCR(7-8) of FH upon C3b binding suggests their implication in a C3b-FH interaction zone. This is consistent with the weak C3b-binding site in SCR(6-8) proposed by Schmidt *et al.* [41].

An dramatic decrease in oxidation was also observed in SCR(10-14): only one residue (His746) was detected as oxidized compared with six modified residues in free FH.

Finally, Tyr1159 located in SCR20 is the only oxidized residue detected in the C-terminal part of FH.

Chemical cross-linking of Factor H

The solvent accessibility study of FH residues provides data on the exposure of the different modules depending on their location in the FH molecule. To obtain further insight into the structural folding and dimerization of FH, the full-length FH molecule was subjected to chemical cross-linking experiments coupled with MS. To do so, FH was incubated in a ratio of 1:10 with the homobifunctional cross-linker NHS-ester BS3 (11.4 Å spacer arm), which targets mainly the solvent accessible lysine residues, and to a lesser extent, serine, threonine and tyrosine side chains [43]. FH and potential cross-linked species were separated by SDS-PAGE. In addition to FH migrating at the expected molecular weight (i.e. 150 kDa), an additional species was observed with an apparent molecular weight of ≈ 300 kDa, corresponding to a dimer of FH. The dimer species and monomeric FH were in-gel digested and the resulting peptides were separated and analyzed using nano LC-MS/MS.

The search for cross-linked peptides was carried out by combining different software, as described by Redeker *et al.* [44]. We found 13 cross-linked peptides belonging to FH. They were mainly detected in the dimer species (10 interpeptides and 2 intrapeptides), whereas only 3 cross-linked species were identified in the FH monomer. Among all these cross-linked peptides, two were identified in both bands. Table 3 groups the cross-linked peptides of FH according to the following nomenclature proposed by Schillig *et al.* [45]: dead-end (type 0), intrapeptide (type 1) and interpeptide (type 2). In accordance with the reactivity of residues towards BS3 NHS-ester [43], the majority of cross-linked amino acids were lysine residues (52% of the cross-linked residues), while serine and threonine accounted for 36% and 12% of cross-linked residues, respectively. No reaction with tyrosine side chains was detected. It is worth noting that a large part of bridged residues belong with regions such as in SCR3, SCR8, SCR11, SCR13, SCR16 and SCR20, where hydroxyl radical-oxidized were detected upon HRP experiment, thus confirming solvent accessibility and reactivity of these regions of FH.

Secondly, cross-linking experiment allowed us to investigate the dimerization state of FH. We assumed that the interpeptides identified by MS/MS in the FH dimer and not in FH monomer originate from two bridged FH molecules. Figure 8A shows a mass spectrum of interpeptide located in SCR20, demonstrating that two FH molecules were proximal during cross-linking reaction. However, the detection of the interpeptide SSQESYAHGYK-WTAK (868-878/1165-1168) from SCR15 and SCR20 not only in the dimer species, but also in FH monomer indicates that this interpeptide may also arise from neighboring peptides within one FH molecule. This result suggests that the SCR15 and SCR20 modules are close enough to each other within one FH molecule, although a previously reported SAXS-derived structural model of FH [25] (PDB: 3GAV-1) showed distant SCR15 and SCR20 modules, a 164 Å distance separating Ser872 and Thr1166 (*i.e.* 14 times longer than the BS3 spacer arm). Given the flexibility of FH, we assume that it can adopt several conformations in solution, among them conformations where SCR15 and SCR20 are much closer than in the SAXS model 3GAV-1 model, and thus allowing cross-linking reactions to occur within the same molecule. In addition, the moderate ionic strength used in this study (114 mM NaCl *vs.* 137 mM for the 3GAV-1 model) may have favored a more folded-back conformation of FH. Furthermore, the high net charge difference between SCR20 (+8) and SCR15 (-4) may also contribute to interactions between these two modules [25]. This result underlines the high flexibility of FH, and, in this respect, it is noteworthy that a major part of the detected cross-linked peptides involved the C-terminal module SCR20, in agreement with a recent study [46] which also reported multiple cross-linking sites in SCR20 of monomeric FH.

Cross-linked peptides within the SCR20 module involved residues Lys1168 and Lys1170. These residues are close enough (8.1 Å) to give rise to an intrapeptide cross-link (WTAKQKLYSR). They were also involved in the WTAKQK- WTAKQK (1165-1170/1165-1170), QKLYSR-QKLYSR

(1169-1174/1169-1174) and QKLYSR- WTAKQK (1169-1174/1165-1170) interpeptides detected in the FH dimer. These C-terminal interpeptides, involving two FH molecules, are consistent with a previous report showing the capacity of FH to self-associate, with a dimerization site in its C-terminus [47].

DISCUSSION

We mapped the FH surface using HRPf and chemical cross-linking coupled with MS. HRPf can provide high-resolution data on whole protein molecules in solution. In this study, we carried out the high-resolution HRPf analysis on the full-length FH molecule (150 kDa) alone and FH in interaction with C3b (330 kDa). HRPf has been used to study protein complexes such as gelsolin (80 kDa) [48], but, to our knowledge, there have been only a few studies on complexes of over 200 kDa [49,50]. We compared the oxidation sites, when possible, with solvent accessibility data available from crystallographic and solution-NMR structures of FH modules in the PDB. The identification of oxidized side chains provided clues on the solvent accessibility of FH peptides, which is particularly interesting in regard to several SCR modules whose structure is not yet known (e.g. SCR5 and SCR14). Moreover, assuming modification of the solvent-accessibility surface of FH upon interaction with C3b, the accurate measurement of the oxidation pattern was used to construct a map of interaction sites on FH.

Twenty-three residues located in 17 SCR modules were oxidized by hydroxyl radicals, showing the presence of reactive sites on the FH surface. FH contains more than 400 hydroxyl radical sensitive residues (Met, Cys, Trp, Tyr, Phe, His, Pro and Leu), yet many of them are protected within the folded structure of the protein and therefore they are not accessible for hydroxyl radical, or are located in the non-covered sequence of FH. Having been selected on their high signal-to-noise ratio and reproducibility, these oxidized residues represent a robust mapping of the solvent accessible area of FH. Roughly 30% of oxidized amino acids detected in unbound FH were located in the SCR(1-4) modules, highlighting this N-terminal domain as a solvent-accessible area of FH. This result confirms the high reactivity of this region of FH already known to contain one C3b binding site [16]. Four oxidized amino acids were identified in the SCR7 and SCR8 modules, which encompass the heparin and C3b binding sites [41], as well as binding site for pathogen components like *Streptococcus pneumoniae* protein PspC [46]. Interestingly, several residues within the SCR(10-16) modules were also oxidized, indicating that this central portion of FH is exposed to the solvent and may contain interaction sites.

Fifteen residues mapped on the surface of free FH were no longer available for hydroxyl radical attack when FH was in a complex with C3b, providing valuable information on interaction zones with C3b as well as on the intra-molecular re-arrangement of FH upon C3b binding.

The decrease in both the number of oxidized residues and the reactivity towards hydroxyl radical observed in SCR[1-3] is consistent with the crystallographic structure of C3b-FH(1-4) [16], in which the interface between C3b and the four N-terminal modules of FH stretches over 100 Å. The oxidized residues identified along the first three N-terminal SCR modules of FH are located in close proximity to the C3b surface identified in the crystallographic structure of C3b-FH(1-4). This can explain the decreased or inhibited oxidation of residues located in this region, as observed with His173 in SCR3. This residue is located within a very short distance to the macroglobulin-2 domain of C3b in the crystallographic structure of C3b-FH(1-4) [16] (about 4 and 3.3 Å away from the C3b residues Ser159 and Gln161, respectively) decreasing the SASA from 69.27 to 18.15 Å² upon C3b binding[16].

Tyr24 showed decreased reactivity towards hydroxyl radicals that could not be attributed to direct protection by C3b binding, as determined from the crystal structure of the C3b-FH(1-4) complex. The decrease in Tyr24 reactivity is likely explained by a conformational adjustment of FH upon C3b binding, as illustrated by the significant decrease in the SASA of Tyr24 from 37 Å² in unbound FH to 5.57 Å² in FH complexed with C3b. Leu103 and Ile120 in SCR2 were highly oxidized in free FH, in agreement with their high SASA values calculated from the NMR-derived structure of SCR(1-2) (PDB: 2RLP) [10]. Interestingly, these residues were only detected in the unoxidized form in the presence of C3b, whereas their SASA values remained high in the C3b-FH(1-4) complex (from 99.83 to 90.37 Å² for Leu103 and from 63.66 to 54.47 Å² for Ile120) as observed on the corresponding crystallographic structure (PDB: 2WII) [16]. Leu103 and Ile120 were not directly involved in the

interaction with C3b although in close proximity to the C3b surface in the C3b-FH(1-4) structure (Figure 7). The observed lack of oxidation indicates decreased accessibility of Leu103 and Ile120 in the full-length FH in the presence of C3b in solution. This change in accessibility suggests a different structural arrangement between the in-solution conformation and the crystal structure, which may arise from different constraints applied on these residues in the full-length FH molecule compared with a partial, four-module FH molecule.

Residues His399, Tyr449 and Ile474 located in SCR(7-8) were modified by hydroxyl radicals only in free FH, but remained unmodified in the complex. Interestingly, this loss of accessibility is consistent with a previously reported weak binding site for C3b situated in this region of FH [3,41], but diverges from the proposed models for the C3b-FH complex [13,17], in which FH acquires a bent-back structure and binds with C3b through only the SCR (1-4) and (19-20) modules. Surface plasmon resonance (SPR) studies [41] have demonstrated that SCR(6-8) and possibly SCR9 have low affinity for C3b. Our data confirm these findings, because residues His399, Tyr449 and Ile474 of SCR(7-8) became inaccessible to hydroxyl radicals in the presence of C3b, whereas Met497 located in SCR9 was oxidized in free and bound FH, indicating that this module does not participate in the C3b-FH complex interface.

In the crystallographic structure of SCR(6-8) in complex with SOS (PDB: 2UWN) [13], His399 is exposed to the solvent ($SASA = 65.42 \text{ \AA}^2$) and does not interact directly with this heparin-mimicking sugar. However, the absence of oxidation of His399 in the presence of C3b suggests either that the C3b molecule shields this residue, or that there is a different conformational arrangement of the SCR7 module in the full-length FH molecule compared with a partial FH molecule composed of isolated modules SCR(6-8). The former hypothesis assumes a third C3b-binding site on FH[41]. For example, Met414, which is located on the opposite face of SCR7, was oxidized in the unbound and bound states of FH, suggesting that only one side of SCR7 may interact with C3b, whereas the other face remains exposed to the solvent. This SCR(6-8) region may therefore be involved in the binding of both C3b and glycosaminoglycans.

Neighboring Tyr449 and Ile474, both located in SCR8, were modified by hydroxyl radicals in free FH, indicating a solvent-exposed patch, but they are buried in the crystallographic structure of SOS-SCR(6-8). This discrepancy may again underline a difference in conformation of these modules in the full-size FH in solution and in isolated modules in an X-ray crystal. In addition, artificial constraints on the SCR(6-8) co-crystal due to crystal packing and binding of the highly sulfated carbohydrate ligand SOS cannot be ruled out.

The central SCR(10-14) module exhibited six oxidized residues in free FH, however only His746 (SCR13) remained modified in the presence of C3b, although to a lesser extent. This central region is not known to host a C3b binding site. Nevertheless, the observed oxidation pattern leads to the hypothesis that this central region may be close to C3b in the FH-C3b complex, and may thus be protected from hydroxyl radicals by a neighboring C3b molecule.

Modification of Leu826 and Leu843 (SCR14) by hydroxyl radicals in the unbound FH gave the first insight into the solvent accessibility of these residues, indicating that they are solvent-accessible, because modification of a residue's side chain is unlikely when its SASA is below 10 \AA^2 .

In the C-terminal modules SCR(19-20), which contain C3b and glycosaminoglycan binding sites [17,41], only one residue (Tyr1159, SCR20) was identified as oxidized in both forms of FH. Although Tyr1159 is not located directly in the C3b-FH(19-20) interface zone, it is close to Tyr1190 in SCR20 (about 20 \AA away), which interacts with Asp1093 of C3b. Given that this residue becomes moderately protected in the C3b-FH complex, the neighboring C3b may shield Tyr1159 [17]. The low oxidation level of the C-terminal modules SCR(19-20) may be also a result of the "J-shape" of this C-terminal region [40] showing a 122° angle between SCR18 and SCR19 that can impede the accessibility of SCR19 to hydroxyl radicals. What is more, a difference in conformation, and thus in solvent accessibility, between modules SCR(19-20) within full-length of FH and isolated modules FH(19-20) in complex with C3b cannot be ruled out. Interestingly, the full-length FH bind weakly to C3d (*i.e.* the TED domain of C3b)[41] while the isolated modules SCR(19-20) bind to C3d with a good affinity[17,20], which suggests that the C-terminal region of full-length FH may adopt a conformation masking reactive and binding sites buried in modules SCR(19-20). A recent study confirmed this protected conformation of the C-terminal domain of full-length FH, which becomes exposed upon binding of the N-terminal domain of bacterial protein PspC[46]. The low oxidation level observed

here provides an additional evidence of a locked conformation of the C-terminal modules SCR(19-20) within full-length FH.

The self-association of FH was first reported by Nan *et al.* [47], although it is still under debate as to its occurrence and mechanism *in natura*. Dimerization may mask some binding sites, which can affect the regulation of the alternative pathway of activation by FH. Studies based on SPR [12], non-reducing SDS-PAGE [51] and analytical ultracentrifugation have reported dimerization in FH. Furthermore, X-ray scattering studies have demonstrated two dimerization sites at SCR(6-8) and SCR(15-18) [47]. Our structural study of FH in solution using chemical cross-linking coupled with MS lens further support to the occurrence of dimerization in FH. We detected three interprotein cross-links located in SCR20, confirming the presence of self-association sites within this module. Nevertheless, our study did not confirm the putative dimerization site in SCR(6-8) despite the presence of 14 well-exposed ($SASA > 30 \text{ \AA}^2$) lysine residues within this region of FH.

Finally, our combined HRPF and MS study confirmed the flexibility of the full-size FH molecule in solution. Most importantly, our results demonstrate that several SCR modules may exhibit slightly different structural arrangements in the full-size FH molecule compared with previously described isolated modules (partial molecules). This finding will be useful, for example, for the design of anti-FH antibodies, because antibodies produced from individual or concatenated modules antigens may exhibit differential binding activity on the full-size FH. The footprinting pattern is consistent with the known C3b binding sites such as the N-terminal SCR(1-4) modules that are protected by C3b from hydroxyl radical oxidation. Finally, our study provided new information on the central region of FH that may play a significant role in the folding of FH upon formation of the C3b-FH complex by bringing SCR(10-14) in the proximity to C3b and being involved in a conformational change upon formation of the C3b-FH complex. This study shows that HRPF combined MS has been valuably applied to the high molecular weight and flexible FH molecule. Its application to C3b and C3b in interaction with FH is under progress.

Acknowledgements: We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank P. Da Silva for assistance in using the Metrology beamline, Virginie Redeker for discussion on cross-linking methods, Veronique Legros for help with running the Orbitrap analysis and Cédric Przybylski for MALDI-TOF analyses. CNRS, Evry-Val-d'Essonne University and Genopole supported this work.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

- 1 Ferreira, V. P., Pangburn, M. K. and Cortés, C. (2010) Complement control protein factor H: the good, the bad, and the inadequate. *Mol. Immunol.* **47**, 2187–97.
- 2 Rodriguez, E., Rallapalli, P. M., Osborne, A. J. and Perkins, S. J. (2014) New functional and structural insights from updated mutational databases for complement factor H, Factor I, membrane cofactor protein and C3. *Bioscience Reports* 635–649.
- 3 Sharma, a K. and Pangburn, M. K. (1996) Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10996–1001.
- 4 Okemefuna, A. I., Nan, R., Miller, A., Gor, J. and Perkins, S. J. (2010) Complement factor H binds at two independent sites to C-reactive protein in acute phase concentrations. *J. Biol. Chem.* **285**, 1053–1065.
- 5 Perkins, S. J., Okemefuna, A. I. and Nan, R. (2010) Unravelling protein-protein interactions between complement factor H and C-reactive protein using a multidisciplinary strategy. *Biochem. Soc. Trans.* **38**, 894–900.
- 6 Meri, S. and Pangburn, M. K. (1990) Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 3982–6.
- 7 Blackmore, T. K., Sadlon, T. A., Ward, H. M., Lublin, D. M. and Gordon, D. (1996) Identification of a Heparin Binding Domain in the Seventh Short Consensus Repeat of Complement Factor H. *J. Immunol.*
- 8 Perkins, S. J., Fung, K. W. and Khan, S. (2014) Molecular Interactions between Complement Factor H and Its

- Heparin and Heparan Sulfate Ligands. *Front. Immunol.* **5**, 126.
- 9 Makou, E., Herbert, A. P. and Barlow, P. N. (2013) Functional anatomy of complement factor H. *Biochemistry* **52**, 3949–62.
 - 10 Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrín, D. and Barlow, P. N. (2008) Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem.* **283**, 9475–87.
 - 11 Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B. and Campbell, I. D. (1993) Solution Structure of a Pair of Complement Modules by Nuclear Magnetic Resonance. *J. Mol. Biol.* **232**, 268–284.
 - 12 Oppermann, M., Manuelian, T., Józsi, M., Brandt, E., Jokiranta, T. S., Heinen, S., Meri, S., Skerka, C., Götze, O. and Zipfel, P. F. (2006) The C-terminus of complement regulator Factor H mediates target recognition: evidence for a compact conformation of the native protein. *Clin. Exp. Immunol.* **144**, 342–52.
 - 13 Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Jowitt, T. a, Clark, S. J., Tarelli, E., Uhrín, D., et al. (2007) Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277–83.
 - 14 Schmidt, C. Q., Herbert, A. P., Mertens, H. D. T., Guariento, M., Soares, D. C., Uhrin, D., Rowe, A. J., Svergun, D. I. and Barlow, P. N. (2010) The central portion of factor H (modules 10-15) is compact and contains a structurally deviant CCP module. *J. Mol. Biol., Elsevier Ltd* **395**, 105–22.
 - 15 Makou, E., Mertens, H. D. T., Maciejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., Herbert, A. P., Svergun, D. I. and Barlow, P. N. (2012) Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H. *J. Mol. Biol., Elsevier Ltd* **424**, 295–312.
 - 16 Wu, J., Wu, Y.-Q., Ricklin, D., Janssen, B. J. C., Lambris, J. D. and Gros, P. (2009) Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol., Nature Publishing Group* **10**, 728–33.
 - 17 Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D. T., Svergun, D. I., Johansson, C. M., et al. (2011) Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol., Nature Publishing Group* **18**, 463–70.
 - 18 Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Meri, S., Goldman, A. and Jokiranta, T. S. (2011) Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2897–902.
 - 19 Blaum, B. S., Hannan, J. P., Herbert, A. P., Kavanagh, D., Uhrín, D. and Stehle, T. (2014) Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat. Chem. Biol.* **11**, 77–82.
 - 20 Hebecker, M., Alba-Domínguez, M., Roumenina, L. T., Reuter, S., Hyvärinen, S., Dragon-Durey, M.-A., Jokiranta, T. S., Sánchez-Corral, P. and Józsi, M. (2013) An engineered construct combining complement regulatory and surface-recognition domains represents a minimal-size functional factor H. *J. Immunol.* **191**, 912–21.
 - 21 DiScipio, R. G. (1992) Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149**, 2592–9.
 - 22 Aslam, M. and Perkins, S. J. (2001) Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309**, 1117–38.
 - 23 Perkins, S. J., Gilbert, H. E., Aslam, M., Hannan, J., Holers, V. M. and Goodship, T. H. (2002) Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. *Biochem. Soc. Trans.* **30**, 996–1001.
 - 24 Perkins, S. J., Nealis, A. S. and Sim, R. B. (1991) Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry* **30**, 2847–57.
 - 25 Okemefuna, A. I., Nan, R., Gor, J. and Perkins, S. J. (2009) Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol., Elsevier Ltd* **391**, 98–118.
 - 26 Sclavi, B., Woodson, S., Sullivan, M., Chance, M. R. and Brenowitz, M. (1997) Time-resolved synchrotron X-ray “footprinting”, a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J. Mol. Biol.* **266**, 144–59.
 - 27 Guan, J.-Q. and Chance, M. R. (2005) Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem. Sci.* **30**, 583–92.
 - 28 Huang, W., Ravikumar, K. M., Chance, M. R. and Yang, S. (2015) Quantitative Mapping of Protein Structure by Hydroxyl Radical Footprinting-Mediated Structural Mass Spectrometry: A Protection Factor Analysis. *Biophys. J., Biophysical Society* **108**, 107–115.

- 29 Xu, G. and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–43.
- 30 Laemmli, U. K. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685.
- 31 Kaur, P., Kiselar, J. G. and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Anal. Chem.* **81**, 8141–9.
- 32 Fraczkiewicz, R. and Braun, W. (1997) Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.* **19**, 319–333.
- 33 Redeker, V., Bonnefoy, J., Le Caer, J.-P., Pemberton, S., Laprévotte, O. and Melki, R. (2010) A region within the C-terminal domain of Ure2p is shown to interact with the molecular chaperone Ssa1p by the use of cross-linkers and mass spectrometry. *FEBS J.* **277**, 5112–23.
- 34 Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egerton, J., et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920.
- 35 Monroe, M. E., Tolić, N., Jaitly, N., Shaw, J. L., Adkins, J. N. and Smith, R. D. (2007) VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **23**, 2021–3.
- 36 Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L. N., Beck, M., Schmidt, A., Mueller, M. and Aebersold, R. (2008) Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318.
- 37 Downard, K. M., Maleknia, S. D. and Akashi, S. (2012) Impact of limited oxidation on protein ion mobility and structure of importance to footprinting by radical probe mass spectrometry. *Rapid Commun. Mass Spectrom.* **26**, 226–30.
- 38 Xu, G. and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Anal. Chem.* **77**, 4549–55.
- 39 Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrin, D. and Barlow, P. N. (2008) Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem.* **283**, 9475–87.
- 40 Morgan, H. P., Mertens, H. D. T., Guariento, M., Schmidt, C. Q., Soares, D. C., Svergun, D. I., Herbert, A. P., Barlow, P. N. and Hannan, J. P. (2012) Structural analysis of the C-terminal region (modules 18-20) of complement regulator factor H (FH). *PLoS One* **7**, e32187.
- 41 Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrin, D., Barlow, P. N. and Blaum, S. (2008) A New Map of Glycosaminoglycan and C3b Binding Sites on Factor H. *J. Immunol.* **181**, 2610–2619.
- 42 Kamal, J. K. A., Benchaar, S. a, Takamoto, K., Reisler, E. and Chance, M. R. (2007) Three-dimensional structure of cofilin bound to monomeric actin derived by structural mass spectrometry data. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7910–5.
- 43 Stefanie, M., Bich, C., Touboul, D. and Zenobi, R. (2009) Chemical cross-linking with NHS esters : a systematic study on amino acid reactivities. *J. Mass Spectrom.* **2009**, 694–706.
- 44 Redeker, V., Pemberton, S., Bienvenut, W., Bousset, L. and Melki, R. (2012) Identification of Protein Interfaces between α -Synuclein, the Principal Component of Lewy Bodies in Parkinson Disease, and the Molecular Chaperones Human Hsc70 and the Yeast Ssa1p. *J. Biol. Chem.* **287**, 32630–9.
- 45 Schilling, B., Row, R. H., Gibson, B. W., Guo, X. and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14**, 834–50.
- 46 Herbert, A. P., Makou, E., Chen, Z. A., Kerr, H., Richards, A., Rappsilber, J. and Barlow, P. N. (2015) Complement Evasion Mediated by Enhancement of Captured Factor H: Implications for Protection of Self-Surfaces from Complement. *J. Immunol.* **195**, 4986–98.
- 47 Nan, R., Gor, J. and Perkins, S. J. (2008) Implications of the progressive self-association of wild-type human factor H for complement regulation and disease. *J. Mol. Biol.* **375**, 891–900.
- 48 Kiselar, J. G., Janmey, P. a, Almo, S. C. and Chance, M. R. (2003) Structural analysis of gelsolin using synchrotron protein footprinting. *Mol. Cell. Proteomics* **2**, 1120–32.
- 49 Xu, G., Liu, R., Zak, O., Aisen, P. and Chance, M. R. (2005) Structural allostery and binding of the transferrin*receptor complex. *Mol. Cell. Proteomics* **4**, 1959–67.
- 50 Kiselar, J. G., Mahaffy, R., Pollard, T. D., Almo, S. C. and Chance, M. R. (2007) Visualizing Arp2/3 complex

activation mediated by binding of ATP and WASp using structural mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. **104**, 1552–7.

- 51 Jokiranta, T. S., Hellwage, J., Koistinen, V., Zipfel, P. F. and Meri, S. (2000) Each of the three binding sites on complement factor H interacts with a distinct site on C3b. J. Biol. Chem. **275**, 27657–62.

FIGURE LEGENDS

Figure 1: Sequence coverage of irradiated Factor H.

The Factor H sequence is shown in single-letter amino acid code. The individual short consensus repeats (SCRs) are indicated in brackets. The color of the amino-acid sequence indicates identification of the protein using tandem mass spectrometry (MS/MS): black, sequence not identified in MS/MS analysis; blue, sequence determined using MS (73%); red, oxidized amino acid; pink, glycosylation site; blue arrow: peptide identified as oxidized in unbound Factor H; green arrow, peptide identified as oxidized in the C3b-FH complex.

Figure 2: Representative tandem mass spectra of peptides 92-109 (A) and 110-127 (B) obtained after 30 ms irradiation and trypsin digestion of Factor H.

Mass shifts +16 Da (shown in green) indicate that oxidation occurred in positions Leu103 (A) and Ile120 (B). Peptide γ -fragments are shown in red and b-fragments in blue.

Figure 3: Comparison of oxidized sites in Factor H irradiated as an unbound protein and in a complex with C3b.

A: Unbound Factor H (FH); B: FH irradiated in presence of C3b. FH is represented as a gray surface. The red spheres indicate oxidized residues. Orange circles highlight the FH regions showing differences in oxidative labeling between unbound FH and FH in complex with C3b. PDB ID: 3GAV-1.

Figure 4: Representation of oxidized residues H399, M414, Y449, and I474, located in short consensus repeat [7-8] of Factor H, on the crystallographic structure of C3b-FH[6-8] complexed with sucrose octasulfate [13].

Factor H is shown in blue, oxidized residues in yellow, and SNPs Y384H and Q382 in red. Sucrose octasulfate (SOS) is represented by black sticks.

Figure 5: Representative dose-response plots showing the extent of unmodified FH peptides as a function of X-ray exposure time (ms). Dose-reponse plots were constructed for Factor H (FH) peptides identified in both unbound (black) and complexed (red) states of FH.

Figure 6: Degree of protection of oxidized Factor H peptides defined as the ratio of oxidation of a given peptide in unbound FH to oxidation of the same peptide in the C3b-FH complex.

Differences in oxidation rates less than 20% (corresponding to 0.8-1.2 fold change) are considered negligible. A decrease in the oxidation rate greater than 20% (≥ 1.2 fold) is considered as a significant decrease in solvent accessibility of the peptide, and thus its protection through the interaction between the protein of interest and its ligand. On the contrary, "the negative protection" degree (≤ 0.8 fold) indicates an increase in solvent accessibility of the investigated region.

Figure 7: Surface representation of protected Factor H residues on the crystal structure of C3b-FH[1-4] complex.

The C3b surface is shown in gray, Factor H short consensus repeat modules [1-4] in blue, highly protected residues in red and moderately protected in orange. PDB ID: 2WII.

Figure 8: Representative tandem mass spectra of BS3 cross-linked peptides 1165-1170/1165-1170 (A) and 1169-1174/1165-1170 (B) derived from trypsin digestion of Factor H dimer.

Table 1: Oxidized residues identified by nano liquid chromatography-tandem mass spectrometry in unbound Factor H.

Solvent-accessible surface area (SASA) values of residues belonging to short consensus repeats (SCR) [1-3] were calculated from the NMR solution structure (PDB ID: 2RLP for SCR[1-2] and 2RLQ for SCR[2-3]). SASA values of residues located in SCR[7-8] were calculated from the crystal structure of SCR[6-8] (His402 risk variant) in complex with sucrose octasulfate (SOS) (PDB ID: 2UWN) and SASA values of residues located in SCR9 were calculated from the crystal structure of SCR9 with choline binding protein (PDB ID: 4K12). SASA values of oxidized residues located in the central region of Factor H (FH) were calculated from the NMR solution structures of SCR[10-11] (PDB ID: 4B2R), SCR[12-13] (PDB ID: 2KMS) and SCR[15-16] (PDB ID: 1HFH). The crystallographic structure of the C-terminus of FH (PDB ID: 3SW0) was used to determine SASA values for residues located in SCR18 and SCR20.

The SAS values of residues Leu262, Leu826, Ile843 and Ile981 cannot be determined because there are no structural models for them.

Please note that the residue numbering does not include the 18-residue long signal peptide.

SCR	Peptide range	Sequence of oxidized peptide	Oxidized amino acid	SASA [Å ²]	PDB structure	Structure type
1	11-33	NTEILTGSWSDQTYPEGTQAIYK	Tyr24	37.06	2RLP	NMR Solution
1	40-49	SLGNVIMVCR	Met46	13.20	2RLP	
1	50-60	KGEWVALNPLR	Val54	45.15	2RLP	
2	65-91	RPCGHPGDTPFGTFTLTGGNVFEYGVK	Leu80	15.34	2RLP	
2	92-109	AVYTCNEGYQLLGEINR	Leu103	99.83	2RLP	
2	110-127	ECDTDGWTNDIPICEVVK	Ile120	63.66	2RLP	
3	166-182	IEGDEEMHCSDDGFWSK	His173	69.27	2RLQ	
5	248-259	SCDNPYIPNGDYSPLR	Leu262	-	-	-
7	393-406	SIDVACHPGYALPK	His399	65.42	2UWN	X-ray crystallography
7	407-423	AQTTVTCMENGWSPTPR	Met414	85.70	2UWN	
8	433-452	SSIDIENGFISESQYTYALK	Tyr449	0.31	2UWN	
8	461-478	LGYYTADGETSGSITCGK	Ile474	7.37	2UWN	
9	490-500	SCDIPVFMNAR	Met497	59.53	4K12	
10	577-606	FSCKPGFTIVGPNVQCYHFGLSPLDLPICK	Phe583	31.17	4B2R	NMR Solution
11	607-624	EQVQSCGPPPELLNGNVK	Leu619	72.49	4B2R	
12	712-731	SITCIHGVWWTQLPQCVAIDK	His717	77.76	2KMS	
13	737-748	SSNLILEEHLK	His746	120.95	2KMS	
14	817-836	VSVLCQENYLIQEGEITCK	Leu826	-	-	-
14	840-849	WQSIPLCVEK	Ile843	-	-	-
16	941-954	CFEGFGIDGPAIAK	Phe945	56.84	1HFH	NMR Solution
17	969-987	TDCLSLPSFENAI ₁ PMGEKK	Ile981	-	-	-
18	1027-1044	DTSCVNPPTVQNA ₁ Y ₁ VS ₁ R	Ile1041	20.52	3SW0	X-ray
20	1154-1164	EIMENY ₁ NIALR	Tyr1159	94.58	3SW0	crystallography

Table 2: Oxidized residues identified by nano liquid chromatography-tandem mass spectrometry of Factor H in the presence of C3b.

Solvent-accessible surface area (SASA) values of residues belonging to short consensus repeat (SCR)1 were calculated from the NMR solution structure (PDB ID: 2RLP). SASA values of residues located in SCR[7-8] were calculated from the crystal structure of SCR[6-8] (His402 risk variant) in a complex with sucrose octasulfate (SOS) (PDB ID: 2UWN) and SASA values of residues located in SCR9 were calculated from the crystal structure of SCR9 complexed with choline binding protein (PDB ID: 4K12). SASA values of oxidized residues located in the central region of Factor H (FH) were calculated from the NMR solution structures of SCR[12-13] (PDB ID: 2KMS) and SCR[15-16]

(PDB ID: 1HFH). The crystallographic structure of the C-terminus of FH (PDB ID: 3SW0) was used to determine SASA values for residues located in SCR18 and SCR20. SASA values of FH residues in complex with C3b were calculated from the crystal structures of SCR[1-4] in complex with C3b (PDB ID: 2WII) and SCR[19-20] in complex with C3d (PDB ID: 3OXU).

The SASA values of residues Met497 and Ile981 cannot be determined because there are no structural models for them.

SCR	Peptide range	Sequence of oxidized peptides	Oxidized residue	SASA [\AA^2] unbound	SASA [\AA^2] complex	PDB structure
1	50-60	KGEW <u>V</u> ALNPLR	Val54	45.15	28.28	2RLP/2WII
7	407-423	AQTTVTC <u>M</u> ENGWSPTPR	Met414	85.70	-	2UWN
9	490-500	SCDIPV <u>F</u> MNAR	Met497	-	-	
13	737-748	SSNLILEE <u>H</u> LK	His746	120.95	-	2KMS
16	915-940	SPPEISHGVVAH <u>M</u> SDSYQYGEEVTYK	Met927	78.04	-	1HFH
16	941-954	CFEG <u>F</u> GIDGPAIAK	Phe945	56.84	-	1HFH
17	969-987	TDCLSLPSFENAI <u>P</u> MGEKK	Ile981	-	-	
18	1027-1044	DTSCVNPPTVQNA <u>I</u> VSR	Ile1041	20.52	-	3SW0
20	1154-1164	EIMEN <u>Y</u> NIALR	Tyr1159	94.58	77.52	3SW0/3OXU

Table 3: Cross-linked peptides of Factor H detected by nano liquid chromatography-tandem mass spectrometry.

Factor H (FH) treated by the cross-linking reagent BS3 was separated by SDS-PAGE, and the monomer and dimer bands were in-gel digested. Resulting peptides were analyzed by nano liquid chromatography-tandem mass spectrometry.

FH	Modification	Peptide sequence	SCR	Amino-acid sequence of peptides
Dimer	Interpeptide	187-193/ 240-247	3-4	CVEI <u>S</u> CK-PLP <u>S</u> CEEK
Dimer	Interpeptide	229-239/ 1189-1192	4-20	GDAVCTE <u>S</u> GWR-L <u>S</u> SR
Dimer	Interpeptide	453-456/ 61-64	8-1	E <u>K</u> AK- <u>K</u> CQK
Dimer	Interpeptide	625-628/ 868-678	11-15	E <u>K</u> TK-SSQE <u>S</u> YAHGYK
Dimer	Interpeptide	732-736/ 737-748	13-13	L <u>K</u> <u>K</u> CK-SSNLILEE <u>H</u> LK
Dimer & monomer	Interpeptide	868-878/ 1165-1168	15-20	SSQE <u>S</u> YAHGYK-W <u>T</u> AK
Dimer & monomer	Intrapeptide	955-968	16	CLGE <u>K</u> W <u>S</u> HPPSCIK
Dimer	Interpeptide	1023-1026/ 1175-1184	18-20	P <u>T</u> CR-TGE <u>S</u> VEFVCK
Dimer	Interpeptide	1165-1170/ 1165-1170	20-20	W <u>T</u> AKQK- W <u>T</u> AKQK
Dimer	Intrapeptide	1165-1174	20	W <u>T</u> AK <u>Q</u> KLYSR
Dimer	Interpeptide	1169-1174/ 1165-1170	20-20	Q <u>K</u> L <u>Y</u> SR- W <u>T</u> AK <u>Q</u> K
Dimer	Interpeptide	1169-1174/ 1169-1174	20-20	Q <u>K</u> L <u>Y</u> SR- Q <u>K</u> L <u>Y</u> SR
Monomer	Dead-end	737-751	13	SSNLILEE <u>H</u> L <u>K</u> NKK

SCR, short consensus repeat

Figure 1

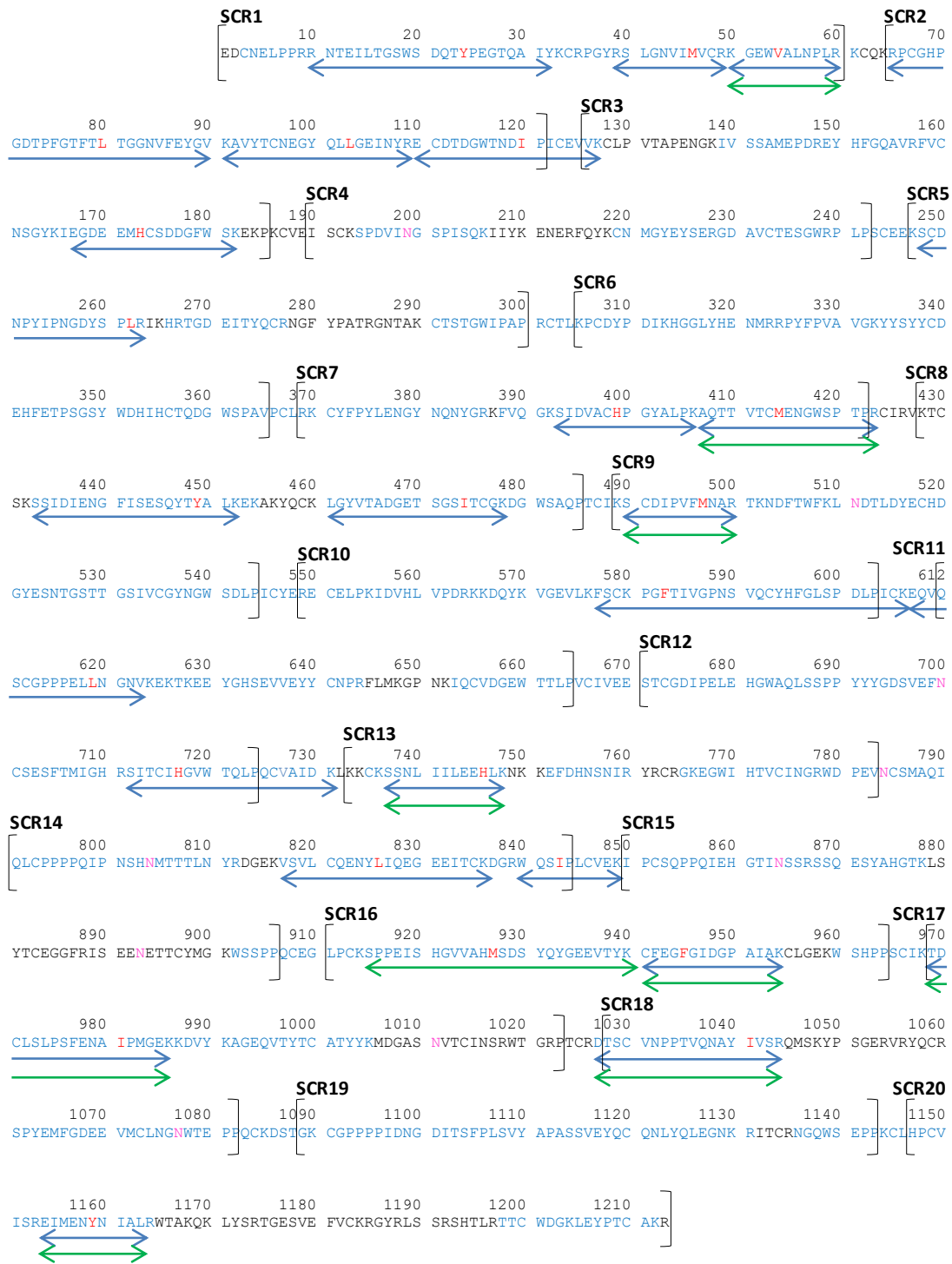


Figure 2A

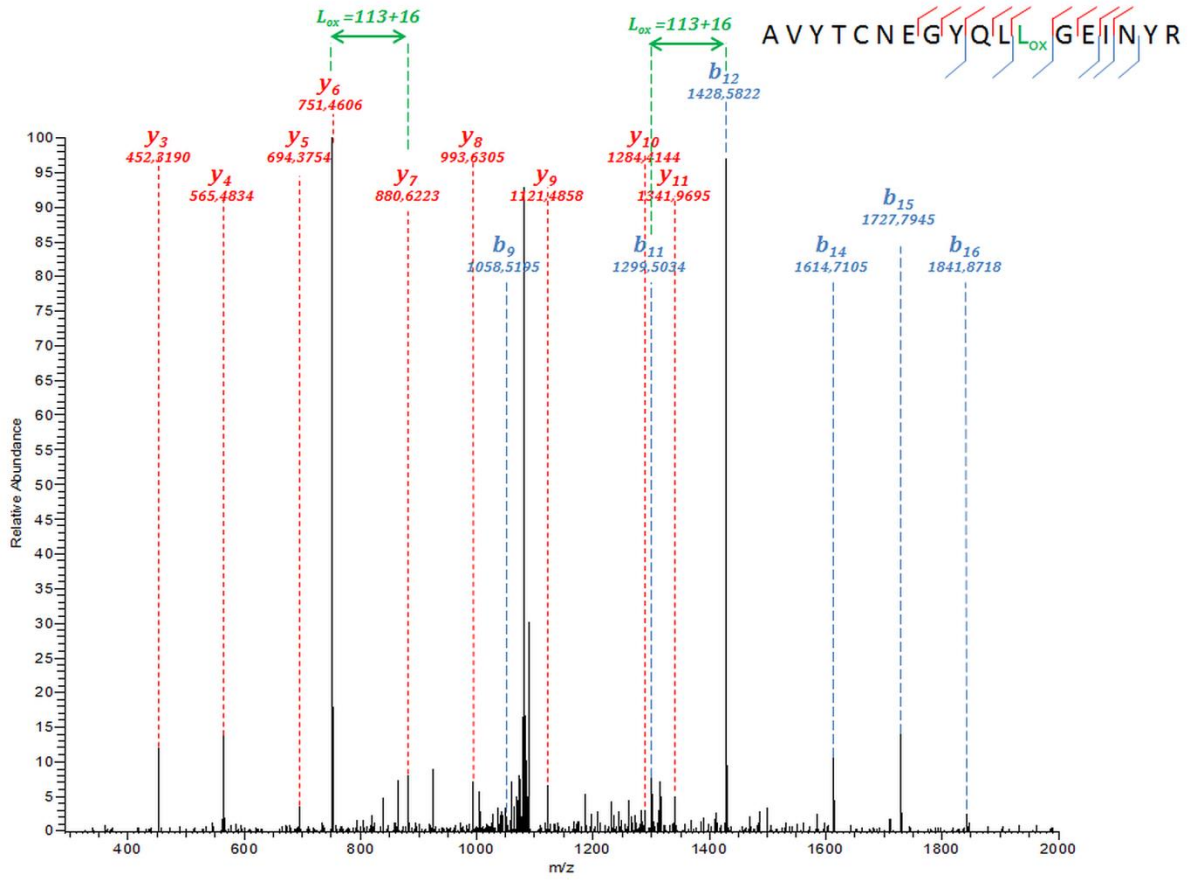


Figure 2B

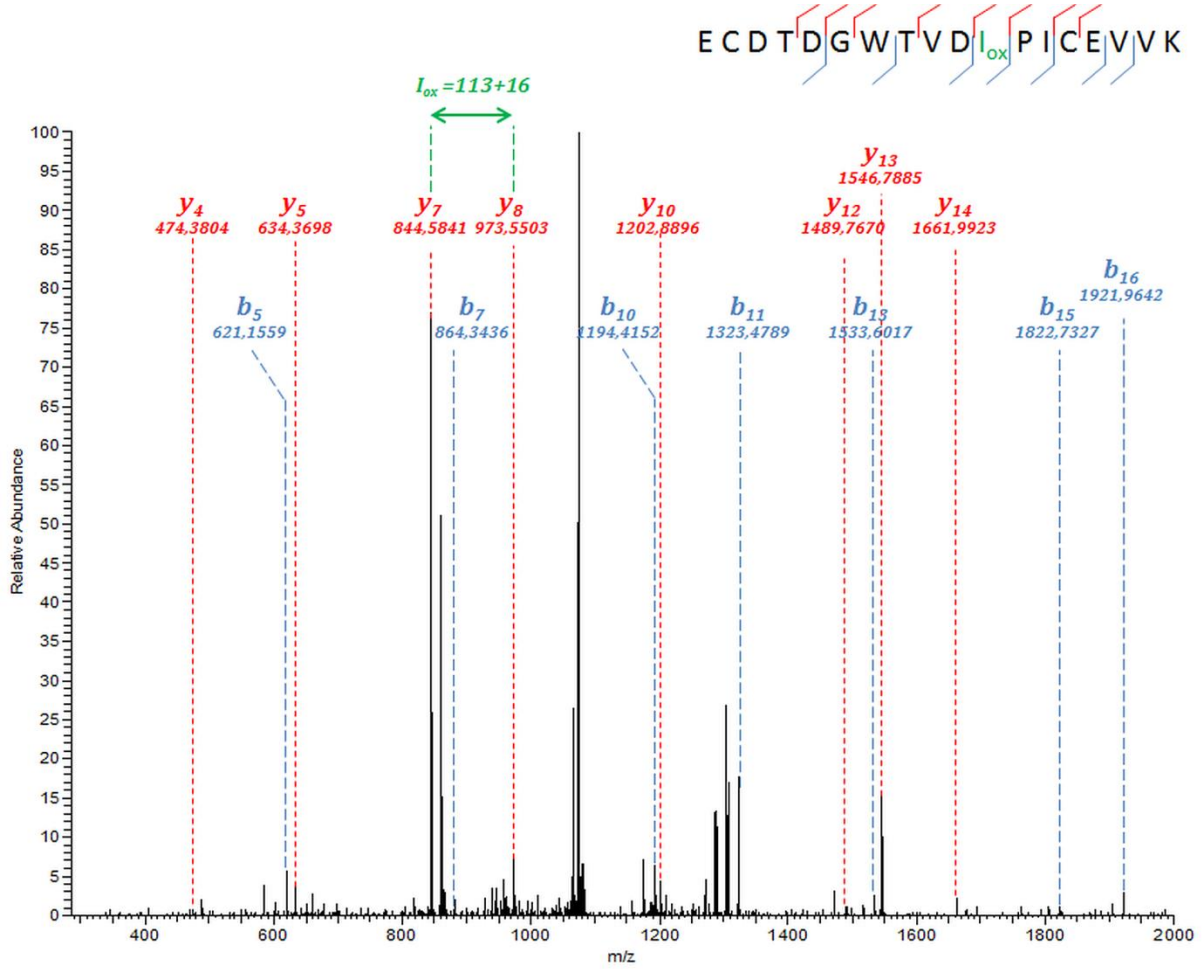


Figure 3

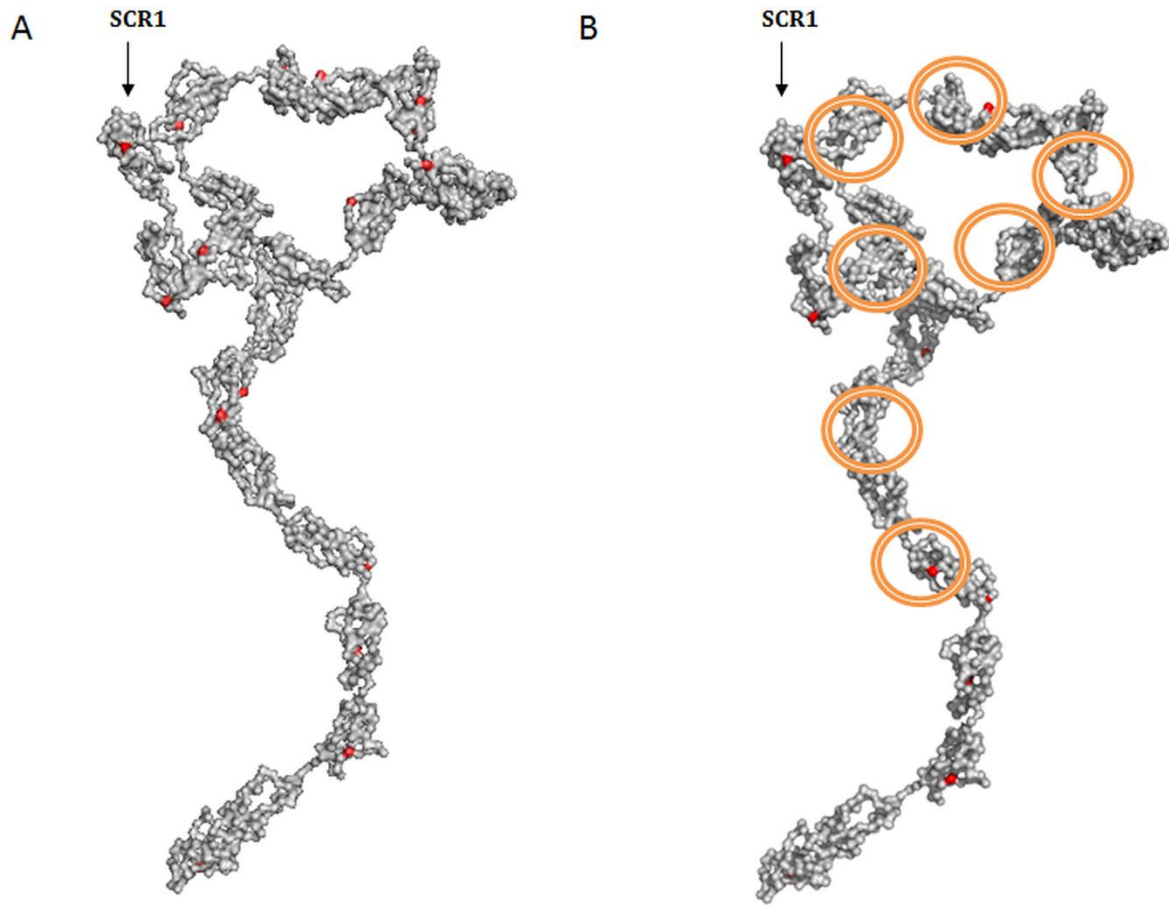


Figure 4

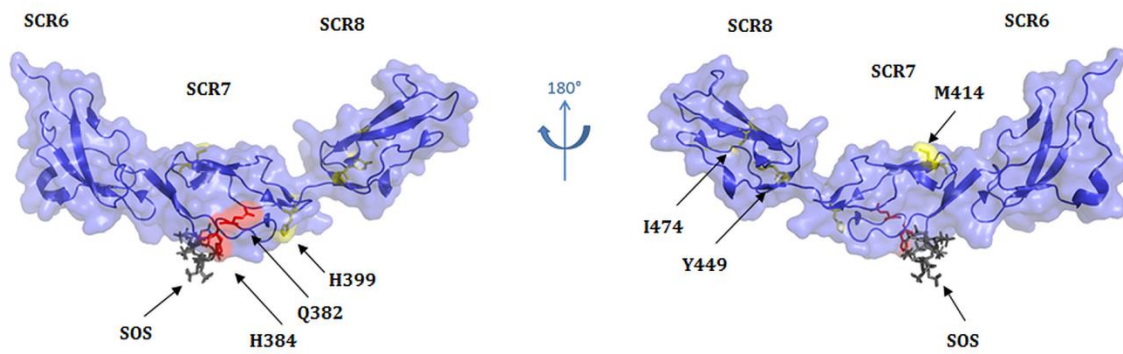


Figure 5

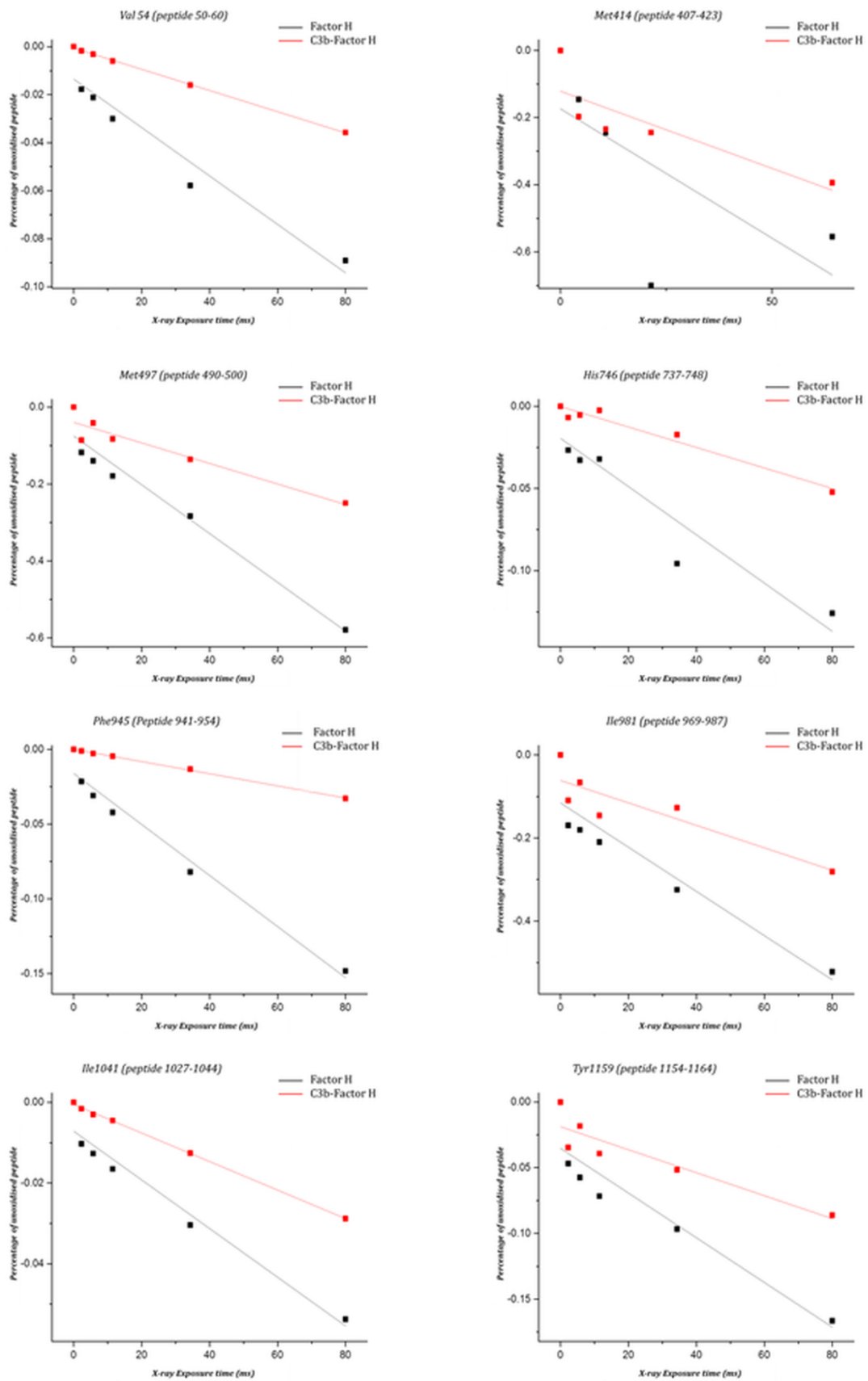


Figure 6

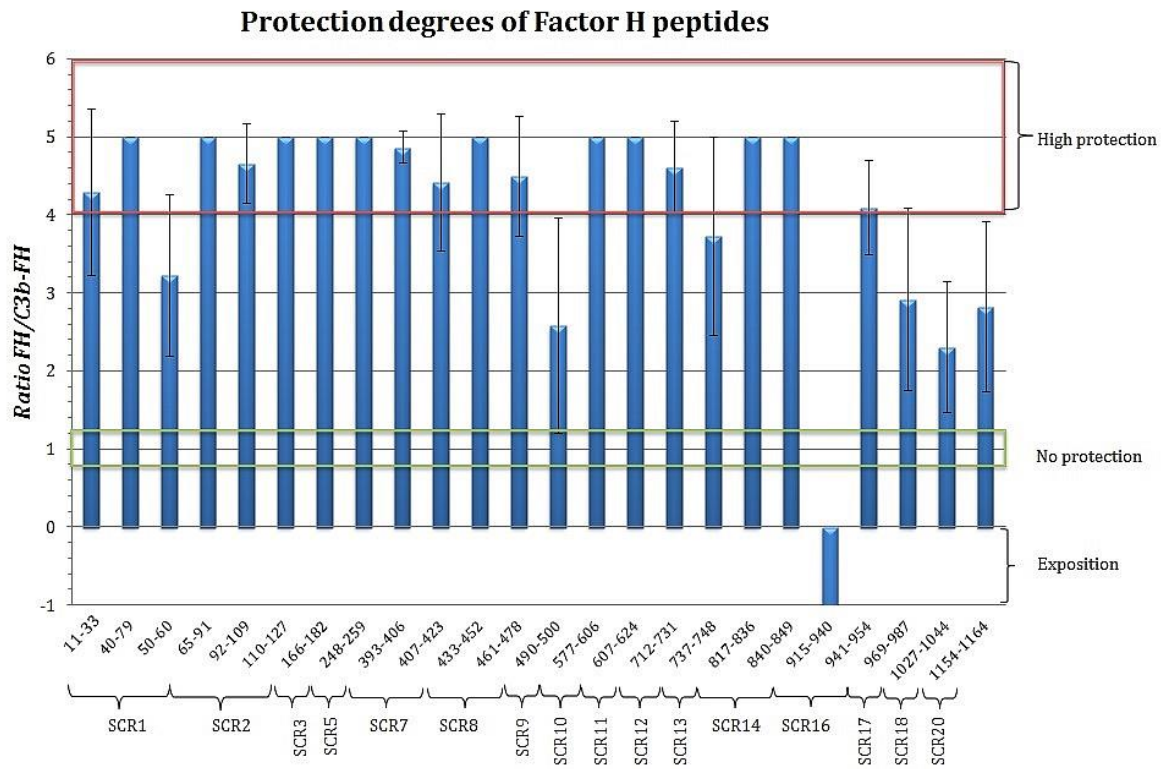


Figure 7

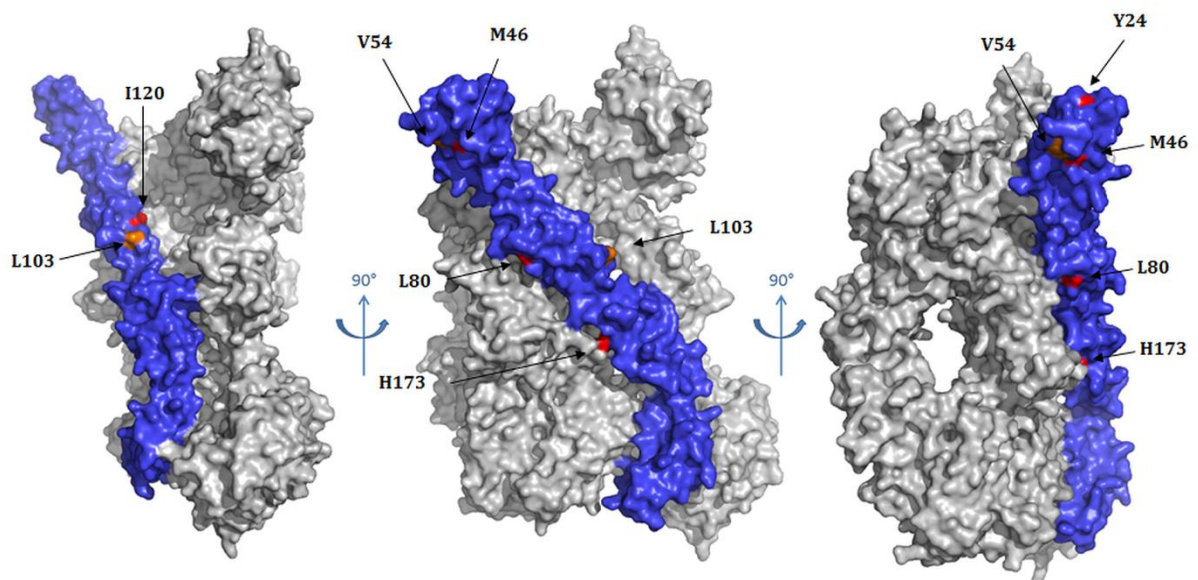


Figure 8A

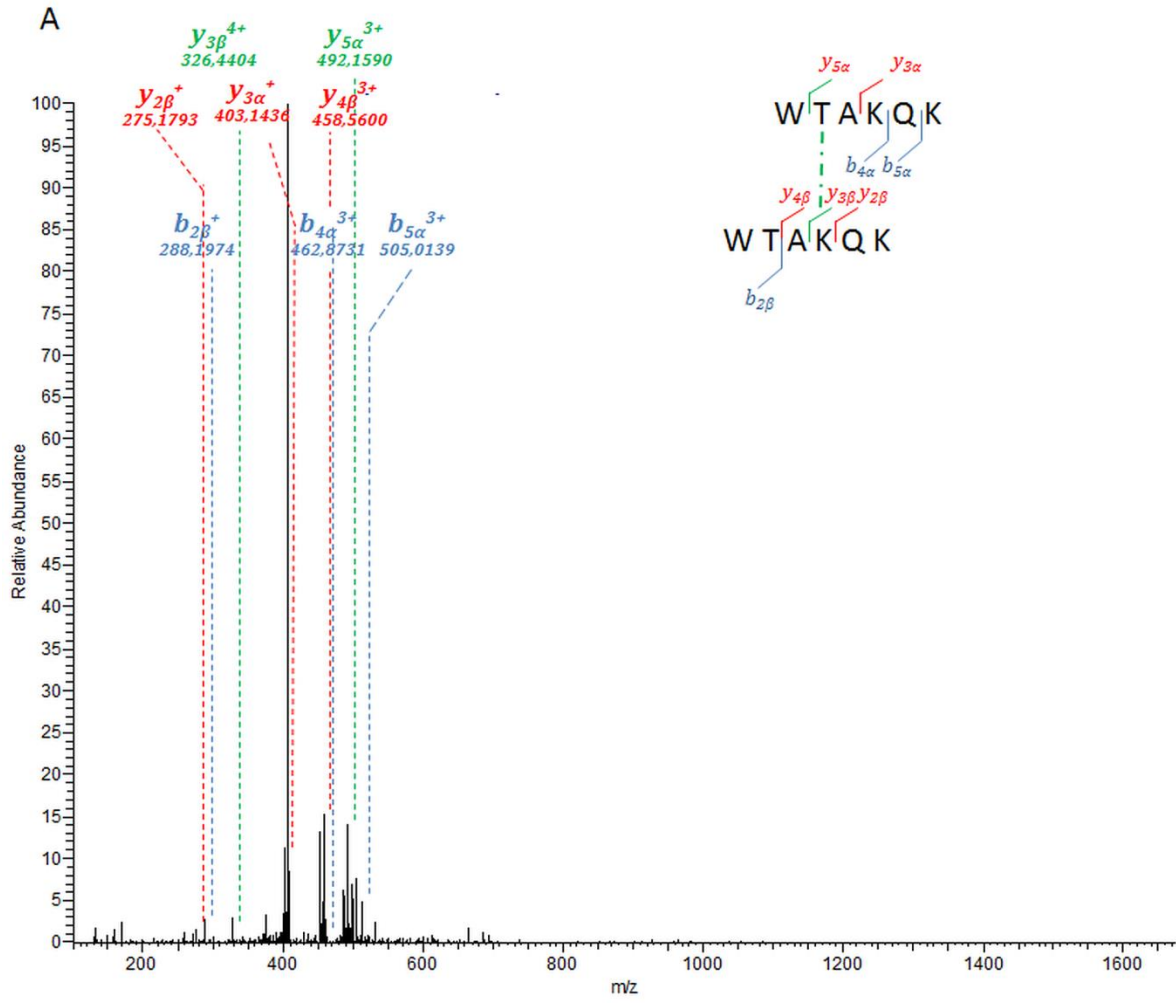


Figure 8B

