Published in final edited form as:

Anal Chem. 2017 November 07; 89(21): 11583–11591. doi:10.1021/acs.analchem.7b02915.

Using Covalent Labeling and Mass Spectrometry to Study Protein Binding Sites of Amyloid Inhibiting Molecules

Tianying Liu, Tyler M. Marcinko, Patrick A. Kiefer, and Richard W. Vachet*

Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003, United States

Abstract

Amyloid aggregates are associated with several debilitating diseases, and there are numerous efforts to develop small molecule treatments against these diseases. One challenge associated with these efforts is determining protein binding site information for potential therapeutics because amyloid-forming proteins rapidly form oligomers and aggregates, making traditional protein structural analysis techniques challenging. Using β -2-microglobulin (β 2m) as a model amyloid-forming protein along with two recently identified small molecule amyloid inhibitors (i.e. rifamycin SV and doxycycline), we demonstrate that covalent labeling and mass spectrometry (MS) can be used to map small-molecule binding sites for a rapidly aggregating protein. Specifically, three different covalent labeling reagents, namely diethylpyrocarbonate, 2,3-butanedione, and the reagent pair EDC/GEE, are used together to pinpoint the binding sites of rifamycin SV, doxycycline, and another molecule, suramin, which binds but does not inhibit Cu(II)-induced β 2m amyloid formation. The labeling results reveal binding sites that are consistent with the known effects of these molecules on β 2m amyloid formation and are in general agreement with molecular docking results. We expect that this combined covalent labeling approach will be applicable to other protein/small molecule systems that are difficult to study by traditional means.

Graphical abstract

*Prof. Richard W. Vachet, Department of Chemistry, University of Massachusetts, Amherst, MA 01003. rwvachet@chem.umass.edu. Phone: (413) 545-2733.

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

Supporting Information: Experimental procedures for the ESI-MS titrations; details about the protein thermal stability measurements; calculations of the K_d values for each small molecule binding to β 2m; example SEC-HPLC data; the effect of small molecules on β 2m monomer structure, and additional figures and tables explaining the protein-ligand docking results are found in the Supporting Information.

Amyloid diseases, such as Alzheimer's and Parkinson's, are characterized by the accumulation of insoluble aggregated proteins in cells, tissues, and organs.¹ Dialysis-related amyloidosis (DRA),² which occurs in patients undergoing long-term hemodialysis due to renal dysfunction, involves amyloid deposits of the protein β -2-microglobulin (β 2m) in the musculoskeletal system.³⁻⁵ Currently, there is no treatment for DRA,^{6,7} but recent *in vitro* studies have identified several molecules that can redirect β 2m amyloid formation, suggesting that these molecules might act as prototypes for future drug design.⁸⁻¹⁰ To facilitate drug design efforts, it would be valuable to identify where these molecules bind on β 2m, so that targeted screening of compound libraries could be conducted to find even more potent molecules. A key challenge to experimentally identifying the binding site on β 2m, or any amyloid-forming protein, is how rapidly the monomeric protein is converted to oligomers and aggregates, which makes traditional protein structural analysis techniques, such as X-ray crystallography and NMR, unsuitable. To address this challenge, we are exploring methods based on covalent labeling and mass spectrometry (MS) to rapidly map binding sites before aggregation occurs.

Covalent labeling is a protein surface modification technique that relies on selective¹¹ (e.g. succinimides) or non-selective¹²⁻¹⁸ (e.g. hydroxyl radicals) labeling reagents to covalently modify solvent-exposed amino acid side chains that can then be identified by MS and tandem MS, often in conjunction with bottom-up sequencing. Covalent labeling approaches can be particularly valuable for finding protein-protein or protein-ligand binding sites, as they probe changes in side chain solvent accessibility. Our group has found that diethylpyrocarbonate (DEPC) is a valuable pseudo-selective reagent for studying protein/protein interactions.¹⁹⁻²⁵ DEPC has some advantages over other non-selective reagents, such as hydroxyl radicals, in that it requires no special equipment (e.g. laser or synchrotron), and it results in only a single reaction product, thereby simplifying MS analyses and improving detection sensitivity. DEPC provides good structural coverage as it can react with up to 30% of the residues in the average protein, providing an effective resolution around 8-10 Å.²³ While this level of structural detail can often help define protein-protein interaction sites, this level of resolution may not be sufficient for identifying small molecule binding sites. Consequently, we are exploring the combination of information obtained from DEPC labeling with information from other labeling reagents. In this study, we show that other labeling reagents, namely 2,3-butanedione (BD), which labels Arg residues,²⁶ and the 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)-glycine ethyl ester (GEE) pair, which labels Asp and Glu residues,²⁷ can be used with DEPC to better pinpoint protein-

small molecule binding sites. To demonstrate the effectiveness of this combined labeling approach, we determine the binding sites of three small molecules known to bind to β 2m – doxycycline,⁸ rifamycin SV,⁹ and suramin.²⁸ The former two molecules are known to inhibit Cu(II)-induced β 2m amyloid formation^{29,30} by diverting the reaction toward amorphous aggregates, while suramin has no effect on the amyloid formation reaction.¹⁰ The identified binding sites are consistent with computational modeling and previous biochemical studies, providing validation of the obtained labeling results. We predict that our combined labeling approach should be applicable to other protein-ligand systems that are difficult to study by more traditional methods.

Materials and Methods

Materials

Human full-length β 2m was obtained from Lee Biosolutions (Maryland Heights, MO). Diethylpyrocarbonate (DEPC), doxycycline hyclate, glycine ethyl ester hydrochloride (GEE), imidazole, iodoacetamide, MOPS, MOPS sodium salt, rifamycin SV sodium salt, suramin sodium salt, 2,3-butanedione (BD), tris(2-carboxyethyl)phosphine (TCEP), urea, N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), and L-arginine were obtained from Sigma-Aldrich (St. Louis, MO). Immobilized trypsin and chymotrypsin and triethylamine acetate (pH 8.0, 1 M) were obtained from Princeton Separations (Adelphia, NJ). Acetonitrile, ammonium acetate, CuSO₄, formic acid, potassium acetate, and HPLC grade water were all purchased from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA).

Sample Preparation

Sample preparation for the ESI-MS titration experiments is explained in detail in Supporting Information (SI). Briefly, each sample was prepared with 2.5 μ M β 2m, 20 μ M Cu(II) (CuSO₄), 50 mM ammonium acetate, and variable concentrations of the small molecule of interest over an appropriate concentration range. The samples were incubated for up to 3 h at 37 °C before analyzing the samples by ESI-MS. For covalent labeling and size-exclusion chromatography (SEC) measurements, an 81 μ L solution of 75 μ M β 2m was prepared in 18.5 mM MOPS (pH 7.4), 0.37 M urea and 0.11 M potassium acetate, with 150 μ M Cu(II) (CuSO₄).

Stock solutions of 0.1 M doxycycline, 0.05 M rifamycin SV, and 0.025 M suramin were freshly prepared in distilled and deionized water. Samples containing β 2m and doxycycline or suramin were incubated at 37 °C for 1 h before subjecting β 2m to the measurements of interest. Rifamycin SV was incubated with β 2m at 37 °C for 1 min before conducting the measurements of interest.

ESI-MS Titration Experiments

To identify the appropriate small molecule concentrations for the covalent labeling studies, the dissociation constant (K_d) values for doxycycline, suramin and rifamycin SV bound to β 2m were determined using previously described ESI-MS titration methods.³¹ The values

for suramin and rifamycin SV were determined directly, while for doxycycline a reference ligand method was applied.³⁴ Details of ESI-MS-titration experiments are explained in SI.

Size Exclusion Chromatography (SEC)

Protein samples were separated by a SuperSW2000 SEC column (Tosoh Bioscience, LLC; Tokyo, Japan) with an HP Agilent (Wilmington, DE) 1100 HPLC system. The mobile phase contained 150 mM ammonium acetate (pH 6.8), and a flow rate 0.35 mL/min was used. Before injection, the SEC column was equilibrated by flowing the mobile phase for 30 min. 20 μ L of a prepared sample or calibration standard was injected, and eluting proteins were detected by UV absorbance at 214 nm. A solution containing 5 μ M bovine serum albumin (MW = 66000), 5 μ M ovalbumin (MW = 45000), 5 μ M carbonic anhydrase (MW = 29040), and 5 μ M β 2m (MW = 11731) was used for molecular weight calibration.

Circular Dichroism Spectroscopy

Circular dichroism was performed on a JASCO J-1500 spectrometer (Easton, MD) to measure the impact of the small molecules on the secondary structure of β 2m. Data were collected by scanning from 250 to 195 nm at 25 °C at a data pitch of 0.5 nm with a scan rate of 20 nm/min. Each sample contained 50 μ M β 2m and 100 μ M of the selected small molecule, with 25 mM MOPS and 150 mM potassium acetate at pH 7.4 (urea was omitted to generate higher quality CD spectra). Three scans were averaged per sample. Data processing are found in the SI.

Fluorescence Spectroscopy

To measure the impact of the small molecules on the tertiary structure of β 2m, a PTI Quantmaster 300 was used (Edison, NJ). Intrinsic fluorescence was measured via excitation separately at both 283 and 295 nm. Emission was monitored from 305 to 400 nm. Details of the sample preparation are found in the SI. Data were collected at 1 nm intervals with an integration time of 1 second. Five scans were averaged per sample. Peaks were normalized from 0 to 1 to better compare λ_{max} due to fluorescence quenching from inner filter effects.

For thermal stability experiments, data were collected on a JASCO 1500 spectrometer equipped with a scanning emission monochromator and emission detector for fluorescence detection. Intrinsic tryptophan emission of β 2m was measured every 1 °C from 25 to 80 °C via excitation at 295 nm while monitoring emission from 305-400 nm. Prior to individual scans, samples were equilibrated at the new temperature for 1 minute. λ_{max} of each spectrum was determined by fitting the emission spectra to fourth order polynomial functions in Origin (Northampton, MA). Samples were measured in triplicate and errors bars represent standard error of the mean. Details of the sample preparation are found in the SI.

DEPC Labeling

Stock solutions of DEPC were freshly prepared in acetonitrile. The reaction was initiated by adding a 0.8 μ L aliquot of the DEPC stock solution so that there was a DEPC:protein molar excess of 4:1. In all cases, the final volume of acetonitrile was less than 1%. The reaction was allowed to proceed at 37 °C for 1 min, before being quenched by the addition of imidazole at a final concentration of 10 mM. These modification conditions lead to 0.8 to

1.2 labels per protein on average, which is necessary to maintain the structure integrity of the protein during the modification reaction. These modification conditions have been established by numerous previous studies of DEPC labeling of β 2m by our group.^{19,21-23}

BD Labeling

Stock solutions of BD were freshly prepared in water. The reaction was initiated by adding a 3.3 μ L aliquot of the BD stock solution so that there was a BD:protein molar excess of 1250:1. The reaction was allowed to proceed at 37 °C for 1 min, before being quenched by the addition of an arginine solution to a final concentration of about 140 mM. These modification conditions lead to about 0.9 – 1.1 labels per protein on average, which ensures the structural integrity of the protein during the labeling reaction.

EDC/GEE Labeling

Stock solutions of EDC and GEE were freshly made in 25 mM MOPS (pH 7.4) at a concentration of 50 mM and 2.0 M respectively. The reaction was initiated by adding an 18 μ L aliquot of the EDC stock solution and an 18 μ L aliquot of the GEE stock solution simultaneously; the EDC:GEE:protein molar excess was 150:6000:1. The reaction was allowed to proceed at room temperature for 30 min, before being quenched by the addition of formic acid to a final concentration of 0.1% (v/v). These modification conditions lead to about 0.4 labels per protein on average, which ensures the structural integrity of the protein during the labeling reaction.

Proteolytic Digestion

Before digestion of the proteins for LC/MS/MS analysis, the labeled protein samples were diluted with HPLC grade water to a volume of 400 μ L and then concentrated using a 10,000 MWCO filter to a final volume of 20 μ L. The samples were then brought to pH 8.0 by adding 70 μ L of 100 mM triethylamine acetate. The disulfide bond was reduced at 37 °C for 10 min using TCEP at a protein:TCEP molar ratio of 1:80, and the resulting reduced cysteines were alkylated in the dark for 30 min using iodoacetamide at a protein:iodoacetamide molar ratio of 1:80. Next, the samples were incubated with 10% (v/v) acetonitrile at 50 °C for 45 min to denature the protein. The protein sample was then applied to the selected enzyme at an enzyme:substrate ratio of 1:10. Digestion of DEPC and BD labeled proteins were performed with immobilized chymotrypsin. EDC/GEE labeled proteins were divided into tubes with equal volumes before either immobilized chymotrypsin or immobilized trypsin was added. After 2 h of digestion at 37 °C, the enzyme was separated from the mixture by centrifugation, and the precipitate was discarded. In the case of the EDC/GEE labeled proteins, the peptide mixture generated by the two enzymes was then mixed at equal volume and analyzed at the same time.

Liquid Chromatography – Mass Spectrometry (LC-MS)

All protein digests were measured immediately after proteolytic digestion by LC-MS. Mass analysis was carried out on a Bruker AmaZon (Billerica, MA) quadrupole ion trap mass spectrometer equipped with an electrospray ionization source. Typically, the electrospray needle voltage was kept at ~4 kV, and the capillary temperature was set at 220 °C. Either

collision-induced dissociation (CID) or electron transfer dissociation (ETD) was applied to identify the position of the labels.

HPLC separation was conducted by a Thermo Scientific (Waltham, MA) Ultimate 3000 HPLC with a Thermo Acclaim PepMap RSLC C₁₈ column (300 µm × 15 cm, 2 µm particle size). Peptide mixtures from the proteolytic digests were eluted using a gradient of acetonitrile containing 20% water and 0.1% formic acid that increased from 5% to 45% for 45 min at a flow rate of 4 µl/min.

Determination of Modification Percentages

The modification percentage of each amino acid was determined in the same manner as described previously²¹⁻²⁵. Briefly, the ion abundance of each peptide of interest (i.e. labeled or unlabeled) was determined from extracted ion chromatograms (e.g. Figure S1). For peptides that contain more than one modified amino acid, it was possible to separate the isomeric peptides (e.g. Figure S1), thereby allowing measurements of individual amino acid modification levels. The % modification was calculated based on eq 1. The modified ion intensities ($I_{unmodified}$) used in equation 1 are the sum of the total ion intensities of all the peptides that contain the modified amino acid of interest. Averaged ion intensity across the selected peak areas of individual peaks were used to determine ion intensity. The percent modification is a relative rather than an absolute value because the modified and unmodified peptides have different ionization efficiencies.

$$\% \text{modification} = \frac{I_{\text{modified}}}{I_{\text{modified}} + I_{\text{unmodified}}} \times 100 \quad (1)$$

Reported percent modifications are averages from three separate experiments. A two-sample unpaired student's t-test at a 95% confidence interval was used to determine if changes in modification levels with and without small molecules were significantly different.

Ion Mobility Mass Spectrometry

A Waters Synapt G2-Si quadrupole time-of-flight (QTOF) equipped with a nanospray source was used to collect native mass spectral data. Electrospray capillaries were prepared in-house by sputter coating gold using a Cressington 108 onto pulled borosilicate thin-wall capillaries (Harvard Apparatus, cat#30-0035, Holliston, MA) in a manner similar to established protocols.³² The electrospray capillary voltage was set to 1.0 kV, while the source temperature was 30 °C. The source offset and sampling cone were set to 20 V. Samples were incubated with the small molecules at 37 °C in solution conditions similar to the labelling experiments. Immediately prior to analysis, the samples were removed from the incubation chamber and rapidly desalted into 100 mM ammonium acetate using a HiTrap column (GE Healthcare, Chicago, Illinois). The desalting process yielded a final sample concentration of approximately 10 µM. Data for each sample was acquired for 10 minutes and averaged over that duration. Collisional cross section (CCS) values were estimated from a linear calibration curve of native-like protein ions using a method described elsewhere.³³

Data analysis was carried out using Waters MassLynx and Driftscope. The extracted arrival time distributions (ATDs) were plotted using Origin.

Protein-Ligand Docking

Protein-ligand docking was carried out using Glide (Version 2016-1, Schrodinger, LLC, New York). The β 2m monomer NMR structure (PDB ID 1JNJ, first one of the 20 structures) was used as a starting point and was processed and optimized by Protein Preparation Wizard, setting PROPKA at pH=7.0. The 3D structures of doxycycline and rifamycin SV and the 2D structure of suramin (3D structure not available) were downloaded from PubChem (pubchem.ncbi.nlm.nih.gov). All the small molecule structures were then prepared with LigPrep with forcefield OPLS3 applied. Chiralities of 3D structure were maintained and possible states at the target pH 7.4 ± 0.5 were generated. A grid was generated with centroid of specified coordinates (centered on β 2m) that enabled the entire β 2m monomer structure to be covered by adjusting the size of ligand diameter midpoint box, and no other constraint was applied. The docking was performed using Ligand Docking with Extra Precision (XP) mode. The number of poses to be reported was not limited, but post-docking minimization was applied with 20 poses per ligand included with strain correction terms. After the above steps, all the reported results were ranked by Docking Scores, and poses with Docking Score above zero were excluded.

Results and Discussion

K_d values for β 2m with selected molecules

We first determined the K_d values for each of the three molecules when binding to monomeric Cu(II)-bound β 2m. Measuring the K_d values was important for confirming that the molecules bind to β 2m at the relevant concentrations used in the subsequent covalent labeling experiments. K_d values were measured via ESI-MS titration using an approach that is similar to previous work.^{31,34} Figure 1 shows an example of the titration data obtained for suramin. Data for the other molecules, example mass spectra, and details about the method can be found in Supporting Information (Figures S2 to S5). The determined K_d values for doxycycline and suramin bound to Cu(II)- β 2m are $200 \pm 20 \mu\text{M}$ and $40 \pm 2 \mu\text{M}$, respectively, whereas for rifamycin SV, two K_d values are measured: $K_{d1} = 250 \pm 25 \mu\text{M}$ and $K_{d2} = 80 \pm 20 \mu\text{M}$ (Table S2). These results confirm that the molecules bind to the Cu(II)- β 2m monomer, even though their affinities for the protein are relatively low.

Size exclusion chromatography

The measured K_d values allow us to identify concentrations of each small molecule needed to maximize the β 2m-molecule complex concentration so that the covalent labeling experiments could be as sensitive as possible. Because we wanted to measure binding sites under conditions in which β 2m is known to form amyloids and because we previously found that the inhibitors (i.e. rifamycin SV and doxycycline) can influence the formation of off-pathway pre-amyloid oligomers,¹⁰ we needed to optimize the small molecule concentrations used in the covalent labeling experiments to ensure that Cu(II)- β 2m was still monomeric in the presence of the molecules of interest. Thus, we identified the highest concentration of each small molecule that could be used without observing any protein aggregation during

the 1 min covalent labeling reaction. SEC was used to measure the extent of $\beta 2m$ oligomer formation at different concentrations of doxycycline, rifamycin SV, and suramin. From the SEC results (see example chromatograms in Figures S6 and S7 of the SI), we found that up to 1.8 mM of doxycycline and 0.22 mM of suramin could be added without significant $\beta 2m$ aggregation being observed within 1 h. In both cases, these concentrations correspond to over 85% of the protein being bound to the given molecule. In contrast, only 0.25 mM of rifamycin SV could be added without generating a significant amount of $\beta 2m$ oligomers in 1 min. Under these conditions, about 33% of the Cu(II)- $\beta 2m$ complexes are not bound to rifamycin SV.

Rifamycin SV binding

The $\beta 2m$ -Cu(II) complex was covalently labeled in the presence and absence of rifamycin SV as a way to identify the binding site of this molecule. DEPC was first used because it can modify a wide range of amino acids, including His, Lys, Ser, Thr, Tyr, and the N-terminus. Results from the DEPC labeling experiments are shown Figure 2A. Upon binding rifamycin SV, Lys6, Lys91 and Lys94 are found to undergo significant decreases in labeling, while the N-terminus undergoes a significant increase in labeling. It should be noted that the extents of the labeling changes are small in many cases, but the reproducibility of the experiments is usually sufficient to confidently distinguish even small changes. The reason for the small changes in labeling extent is likely attributed to the fact that about 33% of Cu(II)- $\beta 2m$ complex is not bound to rifamycin SV for the conditions under which we performed the covalent labeling experiments.

To assess the potential rifamycin SV binding site, the covalent labeling results were mapped on the NMR structure (PDB ID 1JNJ) of the Cu(II)-free protein. This structure was selected as the model because other work had demonstrated that Cu(II) binding does not dramatically change the native $\beta 2m$ structure.^{19,35} (Note: no Cu(II)- $\beta 2m$ monomer structures are available.) Upon considering the protein model structure, it is clear that Lys6, Lys91 and Lys94 are in relatively close proximity to each other, with about 17 Å between Lys6 and Lys91 and 12 Å between Lys91 and Lys94. Given the size of rifamycin SV, which is approximately 14×7 Å, the flexibility of the lysine side chains, and the fact that Cu(II)-binding causes only minor structural changes to $\beta 2m$,^{19,35} the DEPC labeling results narrow down the binding site to an area near the A and G β strands. The reason for the increased labeling at the N-terminus is not clear at this stage, but it could potentially reflect an allosteric change caused by rifamycin SV binding.

To obtain additional confirmation of the binding site and further narrow down the $\beta 2m$ residues that interact with rifamycin SV, we employed two other types of labels. The arginine- selective reagent 2,3-butanedione (BD) was selected because Arg3 and Arg12 flank the residues that are part of the A β strand, Arg97 is near the G β strand, and Arg45 is on the opposite side of the protein (see Figure S8). We predicted that changes in the extent of labeling of the nearby Arg residues would further help narrow potential rifamycin SV binding sites. From the BD labeling results (Figure 2A), only Arg97 is found to undergo a significant change in labeling, as its labeling percentage decreases from 6.4 ± 0.3 to 5.4 ± 0.2 in the presence of rifamycin SV. Insignificant labeling changes to Arg3, Arg12, and

Arg45 along with a decrease in Arg97 labeling suggest that rifamycin SV binds predominantly along the G β strand.

We also used the reagent pair EDC/GEE to monitor changes in the solvent accessibility of Asp and Glu residues. Seven Asp and Glu residues are modified by EDC/GEE, including Glu16, Asp34, Glu36, Asp38, Glu50, Asp53, and Asp59, but none of these residues undergo a significant change in labeling, which is not surprising considering that all but one of these residues are located on the opposite face of the protein (see Figure S9) from the residues identified by DEPC and BD labeling. The fact that there is no change in labeling on Glu16, which is located on the loop that is between the A and B β strands, further confirms that rifamycin SV is not binding in this region of the protein but is binding closer to the G β strand.

Protein-ligand docking was also performed using the program Glide from the Schrödinger software suite to predict rifamycin SV's binding site on β 2m. The Cu(II)-free β 2m monomer (PDB ID 1JNJ) was used as the protein model, even though β 2m's fluorescence λ_{max} , CD spectroscopy, its thermal stability, and ion mobility data reveal rifamycin SV binding slightly perturbs the native β 2m structure (Figures S10 and S11). It should be noted that while it is known that Cu(II) binding does not have a dramatic effect on β 2m structure,^{19,35} it does have a measurable effect on β 2m thermal stability (Figure S11). In comparison, rifamycin binding to the Cu(II)-bound protein stabilizes the structure to be more similar to the native. Thus, in the absence of any other β 2m structural model, we feel that the chosen model structure is a reasonable starting point for docking experiments. Upon using this structural model for docking, we find that the predicted binding site is the pocket between Lys75 and Arg97 (Figure 2D). In light of our covalent labeling, we conclude that rifamycin SV likely binds near Lys94 and Arg97 and perhaps Lys91, which is along the G β strand of the protein. A perfect match between the docking and labeling results is not expected given the fact that rifamycin SV mildly perturbs the native structure of β 2m (Figure S11), but the similarity is encouraging. The decreased labeling of Lys6 may be due to its proximity to the other residues, subtle structural changes that the protein undergoes upon binding to rifamycin SV, or could perhaps be due to burial in the small fraction of protein molecules that have a second rifamycin SV bound to them.

More importantly, the binding site that is identified by covalent labeling is consistent with what is known about pre-amyloid oligomeric β 2m structures^{21,22} and our recent work on β 2m amyloid inhibition by rifamycin SV.¹⁰ Our group recently showed that rifamycin SV inhibits β 2m amyloid formation by preventing the β 2m dimer from forming an amyloid-competent tetramer. We had previously found that the β 2m tetramer is formed by two dimers interacting between the G β strands and the D β strands.²² Considering the binding site identified in the current work, rifamycin SV's proposed mode of action for preventing β 2m amyloid formation is to bind near the G β strands (i.e. Figure 2) and prevent two dimers from forming an amyloid-competent tetramer.

Doxycycline binding

The binding site of doxycycline was determined in a similar manner to rifamycin SV. DEPC labeling was first applied and numerous residues undergo significant changes in labeling

(Figure 3A). More sites undergo significant changes than with rifamycin SV, and this increased number of residues is probably because more than 90% of Cu-protein complex is bound to doxycycline. Again, the Cu(II)-free β 2m NMR structure (PDB ID 1JNJ) is used to map the labeling changes for reference. The sites that undergo decreased labeling can be clustered into two primary groups: 1) Lys6, Lys91, and Lys94, and 2) Lys41, Lys48 and His51/Ser52. This clustering of residues is based on the distribution of these sites on the protein surface and the size of one doxycycline molecule (about $13 \times 7 \text{ \AA}$). Lys75 and Ser57/Lys58 are not included in either group because they are somewhat distant from each cluster of residues. It is difficult to exclude either group based on the extent of the labeling decrease because allosteric changes or changes in the local microenvironment can alter the reactivity of residues for reasons other than decreased solvent accessibility. These possible changes are particularly true for Lys residues, as disruption of salt bridges or changes in pK_a , due to microenvironment changes, can influence their reactivity with DEPC.³⁶

Interestingly, the N-terminus, His13, Ser28, His31 and Ser33 all undergo significant increases in labeling. Among these five residues, the N-terminus, Ser28, His31 and Ser33 are close to or part of the known Cu(II) binding site.^{19,35} Increased labeling at these residues suggests that Cu(II) binding is affected by doxycycline binding. Indeed, as we reported in recent work, β 2m's affinity for Cu(II) decreases by approximately 25-fold in the presence of doxycycline.¹⁰ The increased labeling extent of His13 is more difficult to explain but may be due to an allosteric change.

EDC/GEE labeling was again applied to narrow down potential binding sites. As shown in Figure S12, Glu50 and Asp53 are near His51, offering a potential means of distinguishing the second group of residues identified by DEPC labeling (i.e. Lys41, Lys48 and His51/Ser52) from the first group (i.e. Lys6, Lys91, and Lys94). We find that only Glu50, of all the Asp and Glu residues, undergoes a significant decrease in labeling, which is consistent with doxycycline binding near the cluster defined by Lys41, Lys48, and His51/Ser52 but not the cluster defined by Lys6, Lys91, and Lys94. Moreover, Asp34, Glu36, and Asp38 are on the same C β strand as Lys41, but these residues do not undergo a decrease in labeling, indicating that doxycycline is not binding to residues on the C β strand. The fact that both Glu50 and Asp53 are on the same D β strand as Lys48 and His51/Ser52, but only Glu50, not Asp53, undergoes a significant labeling decrease suggests that doxycycline only binds to the beginning of D β strand.

Because no Asp or Glu residues are near the first cluster of residues identified by DEPC labeling, we further investigated the binding site by BD labeling. Arg3, Arg45, and Arg97 surround cluster 1 (Figure S13). Arg3 and Arg97 undergo an increase in labeling, suggesting that doxycycline is not binding near Lys6, Lys91, or Lys94. The changes in the labeling of these Lys residues may be caused by DEPC reactivity decreases caused by changes in the nearby microenvironment. Interestingly, the labeling extent of Arg45, which is near Lys41, does not change significantly. Alongside the lack of labeling changes associated with Asp34, Glu36, and Asp38, no apparent change in the labeling extent at Arg45 further suggests that doxycycline does not bind to the C β strand. Considering the labeling results as a whole, we conclude that doxycycline binds near Lys48, Glu50 and His51/Ser52. These residues all

undergo significant decreases in labeling ratio and all are located on the beginning of D β strand.

We also used protein-ligand docking with Glide to predict the doxycycline binding site on β 2m. The Cu(II)-free β 2m monomer (PDB ID 1JNJ) was again used as the protein model because the fluorescence λ_{max} , CD spectrum, thermal stability, and ion mobility data indicate that doxycycline binding only slightly perturbs the protein's structure (Figures S10 and S11). Indeed, doxycycline appears to perturb the structure even less than rifamycin SV, suggesting that the chosen model is even more appropriate for docking. Using this structural model, docking predicts a binding site that is in the pocket between the C and D β strands (Figure 3C). This docking position is very consistent with the labeling data that shows several residues on the D β strand, including Lys48, Glu50, and His51/Ser52, undergoing significant changes in labeling (Figure S14). Given the fact that several residues on C β strand have no significant changes in the labeling degree and decreases in labeling are observed along the D β strand, it is reasonable to conclude that doxycycline binds near the D β strand.

Binding at the D β strand is consistent with our group's recent work that shows that doxycycline can redirect β 2m amyloid formation by perturbing the β 2m dimer structure such that it cannot form an amyloid-competent tetramer.¹⁰ As indicated above and as previously shown, the β 2m tetramer is formed by two dimers interacting between the G β strands and D β strands.²² Binding of doxycycline to the D β strand, as indicated by our labeling results, reveals how the amyloid formation pathway can be redirected to form non-amyloid aggregates. It should be noted that Giorgetti et al. studied doxycycline/ β 2m binding by NMR and identified a different binding site that involves β strands A and B.⁸ These previous measurements, however, were done under non-amyloid forming conditions in which the protein was present in pure water at a pH between 5.5 and 6. The difference in the binding site that we identify as compared to this previous study likely reflects the fact that the protein likely has a significantly different structure under amyloid-forming conditions at pH 7.4, which are the conditions we have used in our studies here.

Suramin binding

Suramin binds with greater affinity to β 2m than doxycycline and rifamycin SV, but as described previously it does not inhibit amyloid formation.¹⁰ Thus, we hypothesized that the identity of the suramin binding site would be different from the doxycycline and rifamycin SV binding sites such that it would not interfere with oligomer formation. DEPC labeling results in four sites that decrease in labeling (Figure 4A), and these sites are clustered into two groups based on proximity and suramin's size: 1) Lys41 and Lys48 and 2) the N-terminus and Ser57/Lys58. Suramin is a larger molecule than doxycycline and rifamycin SV, and it is less conformationally constrained than the other two, which suggests that it could potentially cover a larger area of β 2m's surface.

With relatively little distinguishing information obtained from DEPC labeling alone, we also used EDC/GEE labeling to narrow further the binding site. Several Asp and Glu residues along the C and D β strands can be probed (Figure S15), and strikingly Asp34, Glu36, and Glu50 all undergo significant decreases in labeling (Figure 4A). These three residues are

positioned near Lys41 and Lys48, suggesting that suramin binds near the C and/or D β strand, rather than near the N-terminus. Curiously, Asp38 and Asp53 undergo increased labeling despite being positioned on the C and D β strands.

BD labeling was used further to identify the suramin binding site. Consistent with the EDC/GEE labeling decreases found for residues on the C β strand, Arg45, which is also on the C β strand (Figure S16), undergoes a significant labeling decrease (Figure 4A). The three other Arg residues in the protein, including Arg3, which is near the N-terminus, do not undergo significant changes in labeling. Considering the labeling data as a whole, the suramin binding site appears to be located along the C β strand and the C-D loop, based on the decreased labeling of Asp34, Glu 36, Lys41, Arg 45, and Lys48 (Figure 4B). The distance from Asp34 to Lys41 is approximately 23 Å, which can be easily spanned given the molecular dimensions of suramin. Moreover, suramin has negatively charged sulfonate groups that could likely bind to Lys41, Arg45, and Lys48. These sulfonate groups might also explain why Asp38, and perhaps Asp53, undergo increased labeling as they are electrostatically repelled by the sulfonate groups and made more solvent accessible. As expected from the outset, this binding site along the C β strand is in contrast to the doxycycline and rifamycin SV binding sites.

Once again, protein-ligand docking was considered for predicting the suramin binding site on β 2m. Because fluorescence, CD spectroscopy, thermal stability, and ion mobility measurements suggest β 2m's structure is significantly perturbed upon suramin binding (Figures S10 and S11), we did not expect good agreement between the docking and covalent labeling results. Indeed, all 7 poses with docking scores below zero place suramin near the N- terminus or near the A or D β strands (Figure S17 and Table S3). These poses do not agree well with the labeling data. In fact, these docking results position suramin near the known β 2m dimer interface,^{21,22} which is not consistent with suramin's inability to inhibit β 2m amyloid formation or the progression of the oligomers that precede amyloid formation.¹⁰ Thus, we conclude that the docking results are inaccurate, highlighting the value of experimental labeling data. This inconsistency between theory and experiment likely arises because the β 2m model used for docking is probably not a good one.

Based on the labeling results, we conclude that suramin binds along the C β strand and near the C-D loop (Figure 4C), which is consistent with the observation that suramin is not capable of inhibiting β 2m amyloid formation.¹⁰ From our previous studies of β 2m oligomer structures,^{21,22} the C β strand has not been implicated as part of the interface in either dimer or tetramer. Binding of suramin to this region of the protein then likely explains why it does not substantially influence β 2m amyloid formation.

Conclusions

Our results indicate that covalent labeling together with MS is a useful tool for studying protein-ligand binding sites. Changes in labeling extents in clusters of adjacent amino acid residues, with and without ligand molecules present, allow ligand binding site information to be obtained. Confidence in the binding site determination is notably improved when multiple covalent labeling reagents are used together. Using DEPC, BD, and the EDC/GEE pair as

labeling reagents, we identify the binding sites of two small-molecule inhibitors of β 2m amyloid formation. The identified binding sites are consistent with the known effects of these molecules and with molecular docking results, thereby providing validation of the covalent labeling method. Rifamycin SV is found to bind to β 2m on the G β strand, while doxycycline binds on the D β strand of β 2m. These binding sites are near the known interfaces of the pre-amyloid tetramer of β 2m²² and are consistent with previous measurements that indicate that these molecules inhibit amyloid formation by preventing the amyloid-competent tetramer from forming.¹⁰ Moreover, the binding site of suramin, which does not inhibit β 2m amyloid formation, occurs near the C β strand and C-D loop of the protein, a region not involved in any known oligomer interface, further validating the accuracy of the covalent labeling method. The fact that three molecules bind to different sites on β 2m may seem surprising, given the small size of the protein, but its β -sandwich structure presents distinct surface regions and inter-sheet binding pockets that evidently can be exploited by the three chemically different molecules. The covalent labeling/MS method described here appears to be suitable for providing protein-ligand binding information for systems that are difficult to study by other methods. In addition, the method might be used to facilitate drug design efforts by improving the targeted virtual screening of compound libraries by narrowing the protein sites most likely to influence a protein's reactivity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant R01 GM015092.

References

1. Hazenberg BP. Rheum Dis Clin North Am. 2013; 39:323–345. [PubMed: 23597967]
2. Floege J, Ketteler M. Kidney Int. 2001; 59:S164–S171.
3. Drücke, Tilman B, Ziad A. Massy Semin Dial. 2009; 22:378–380. [PubMed: 19708985]
4. Heegaard NH. Amyloid. 2009; 16:151–173. [PubMed: 19657763]
5. Gejyo F, Odani S, Yamada T, Honma N, Saito H, Suzuki Y, Nakagawa Y, Kobayashi H, Maruyama Y, Hirasawa Y, Suzuki M. Kidney Int. 1986; 30:385–390. [PubMed: 3537446]
6. Yamamoto S, Kazama JJ, Maruyama H, Nishi S, Narita I, Gejyo F. Clin Nephrol. 2008; 70:496–502. [PubMed: 19049706]
7. Gejyo F, Kawaguchi Y, Hara S, Nakazawa R, Azuma N, Ogawa H, Koda Y, Suzuki M, Kaneda H, Kishimoto H, Oda M. Artif Organs. 2004; 28:371–380. [PubMed: 15084199]
8. Giorgetti S, Raimondi S, Pagano K, Relini A, Bucciantini M, Corazza A, Fogolari F, Codutti L, Salmona M, Mangione P, Colombo L. J Biol Chem. 2011; 286:2121–2131. [PubMed: 21068391]
9. Woods LA, Platt GW, Hellewell AL, Hewitt EW, Homans SW, Ashcroft AE, Radford SE. Nat Chem Biol. 2011; 7:730–739. [PubMed: 21873994]
10. Marcinko TM, Dong J, LeBlanc R, Daborowski KV, Vachet RW. J Biol Chem. 2017; 292:10630–10638. [PubMed: 28468825]
11. Mendoza VL, Vachet RW. Mass Spectrom Rev. 2009; 28:785–815. [PubMed: 19016300]
12. Xu G, Chance MR. Chem Rev. 2007; 107:3514–3543. [PubMed: 17683160]
13. Wang L, Chance MR. Anal Chem. 2011; 83:7234–7241. [PubMed: 21770468]

14. Chen J, Rempel DL, Gau BC, Gross ML. *J Am Chem Soc.* 2012; 134:18724–18731. [PubMed: 23075429]
15. Zhang Y, Wecksler AT, Molina P, Deperalta G, Gross ML. *J Am Soc Mass Spectrom.* 2017; 28:850–858. [PubMed: 28255747]
16. Jumper CC, Schriemer DC. *Anal Chem.* 2011; 83:2913–2920. [PubMed: 21425771]
17. Jumper CC, Bomgarden R, Rogers J, Etienne C, Schriemer DC. *Anal Chem.* 2012; 84:4411–4418. [PubMed: 22480364]
18. Manzi L, Barrow AS, Scott D, Layfield R, Wright TG, Moses JE, Oldham NJ. *Nat Commun.* 2016; 7:13288. [PubMed: 27848959]
19. Mendoza VL, Vachet RW. *Anal Chem.* 2008; 80:2895–2904. [PubMed: 18338903]
20. Srikanth R, Mendoza VL, Bridgewater JD, Zhang G, Vachet RW. *Biochemistry.* 2009; 48:9871–9881. [PubMed: 19754160]
21. Mendoza VL, Antwi K, Barón-Rodríguez MA, Blanco C, Vachet R. *W Biochemistry.* 2010; 49:1522–1532.
22. Mendoza VL, Barón-Rodríguez MA, Blanco C, Vachet R. *W Biochemistry.* 2011; 50:6711–6722.
23. Zhou Y, Vachet RW. *J Am Soc Mass Spectrom.* 2012; 23:708–717. [PubMed: 22298289]
24. Zhou Y, Vachet RW. *J Am Soc Mass Spectrom.* 2012; 23:899–907. [PubMed: 22351293]
25. Borotto NB, Zhou Y, Hollingsworth SR, Hale JE, Graban EM, Vaughan RC, Vachet RW. *Anal Chem.* 2015; 87:10627–10634. [PubMed: 26399599]
26. Fliss H, Viswanatha T. *Can J Biochem.* 1979; 57:1267–1272. [PubMed: 540238]
27. Kaur P, Tomechko SE, Kiselar J, Shi W, Deperalta G, Wecksler AT, Gokulrangan G, Ling V, Chance MR. *MAbs.* 2015; 7:540–552. [PubMed: 25933350]
28. Regazzoni L, Colombo R, Bertolotti L, Vistoli G, Aldini G, Serra M, Carini M, Facino RM, Giorgetti S, Stoppini M, Caccialanza G. *Anal Chim Acta.* 2011; 685:153–161. [PubMed: 21168564]
29. Eakin CM, Knight JD, Morgan CJ, Gelfand MA, Miranker AD. *Biochemistry.* 2002; 34:10646–10656.
30. Eakin CM, Attenello FJ, Morgan CJ, Miranker AD. *Biochemistry.* 2004; 24:7808–7815.
31. El-Hawiet A, Kitova EN, Klassen JS. *Biochemistry.* 2012; 51:4244–4253. [PubMed: 22564104]
32. Hernández H, Robinson CV. *Nat Protoc.* 2007; 2:715–726. [PubMed: 17406634]
33. Bush MF, Hall Z, Giles K, Hoyes J, Robinson CV, Ruotolo BT. *Anal Chem.* 2010; 82:9557–9565. [PubMed: 20979392]
34. Liu L, Kitova EN, Klassen JS. *J Am Soc Mass Spectrom.* 2011; 22:310–318. [PubMed: 21472590]
35. Eakin CM, Knight JD, Morgan CJ, Gelfand MA, Miranker AD. *Biochemistry.* 2002; 41:10646–10656. [PubMed: 12186550]
36. Isom DG, Castañeda CA, Cannon BR. *PNAS.* 2011; 108:5260–5265. [PubMed: 21389271]

Figure 1. ESI-MS titration curve, indicating the binding affinity of suramin to Cu(II)- β 2m. The meaning of $R/(R+1)$ and how this information is used to determine K_d is shown in the Supporting Information.

Figure 2.

Labeling results with rifamycin SV. (a) Changes in covalent labeling modification percentages with rifamycin SV bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in β 2m. (b) Ribbon structure of β 2m, showing the seven β strands labeled A through G. (c) β 2m surface structure illustrating the sites of significant changes in covalent labeling induced by rifamycin SV. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (d) Protein-ligand docking result.

Figure 3.

Labeling results with doxycycline. (a) Changes in covalent labeling modification percentages with doxycycline bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in $\beta 2m$. (b) $\beta 2m$ surface structure illustrating the sites of significant covalent labeling changes induced by doxycycline. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (c) Protein-ligand docking result.

Figure 4. Labeling results with suramin. (a) Changes in covalent labeling modification percentages with suramin bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in $\beta 2m$. (b) $\beta 2m$ surface structure illustrating the sites of significant covalent labeling changes induced by suramin. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (c) Proposed suramin binding site based on covalent labeling data.