analytical chemistry

Surface Accessibility and Dynamics of Macromolecular Assemblies Probed by Covalent Labeling Mass Spectrometry and Integrative Modeling

Carla Schmidt,[†] Jamie A. Macpherson,[‡,#] Andy M. Lau,[§,#] Ken Wei Tan,[§] Franca Fraternali,[‡] and Argyris Politis*[,§]

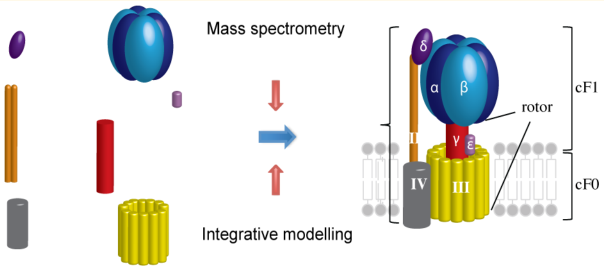
[†]Interdisciplinary Research Center HALOmem, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3, 06120 Halle/Saale, Germany

[‡]Division of Cell & Molecular Biophysics, King's College London, New Hunt's House, SE1 1UL, London, United Kingdom

[§]Department of Chemistry, King's College London, 7 Trinity Street, SE1 1DB, London, United Kingdom

Supporting Information

ABSTRACT: Mass spectrometry (MS) has become an indispensable tool for investigating the architectures and dynamics of macromolecular assemblies. Here we show that covalent labeling of solvent accessible residues followed by their MS-based identification yields modeling restraints that allow mapping the location and orientation of subunits within protein assemblies. Together with complementary restraints derived from cross-linking and native MS, we built native-like models of four heterocomplexes with known subunit structures and compared them with available X-ray crystal structures. The results demonstrated that covalent labeling followed by MS markedly increased the predictive power of the integrative modeling strategy enabling more accurate protein assembly models. We applied this strategy to the F-type ATP synthase from spinach chloroplasts (cATPase) providing a structural basis for its function as a nanomotor. By subjecting the models generated by our restraint-based strategy to molecular dynamics (MD) simulations, we revealed the conformational states of the peripheral stalk and assigned flexible regions in the enzyme. Our strategy can readily incorporate complementary chemical labeling strategies and we anticipate that it will be applicable to many other systems providing new insights into the structure and function of protein complexes.



Mass spectrometry (MS) is an emerging technique in biophysics, and in the last two decades, it has gained in importance when studying the structure and dynamics of macromolecular protein assemblies.[1] Particularly those assemblies which exhibit a certain flexibility and heterogeneity or undergo dynamic interactions with their ligands are the primary targets of structural MS.[2] Various MS techniques each addressing a different question have evolved and are now commonly employed to gain information on composition, stoichiometry, topology, conformation and dynamics.

Most commonly applied is chemical cross-linking,[3−5] a technique which involves covalent linkage of two amino acid side chains in close proximity thus allowing the identification of protein interactions by sequencing the cross-linked dipeptides after enzymatic digestion. MS of intact protein complexes, also called native MS, delivers protein stoichiometries and stable interaction modules enabling the generation of protein interaction networks.[6,7] Together with ion mobility (IM), native MS yields conformation and topology of proteins and their complexes.[8−10] Combining complementary information from chemical cross-linking and native MS delivers valuable insights into the structural arrangements of protein complexes.[11−13]

While cross-linking and native MS identify protein interactions, labeling strategies such as covalent labeling[14] or hydrogen−deuterium exchange (HDX)[15,16] explore solvent accessible surfaces of protein−ligand assemblies. This is of particular interest when studying the dynamics of proteins and their conformational changes,[17,18] for instance upon ligand binding.[19] HDX utilizes the ability of protons to be exchanged with deuterium in solution. The slow exchange rate of protein backbone amide protons causes a mass shift of the protein/peptide, which can be probed by MS. Likewise, chemical labeling approaches introduce modifications to amino acid side chains which can be identified by standard proteomics. Very prominent is hydroxyl radical footprinting involving oxidation of various amino acid side chains.[20] Other labeling strategies employ chemical reagents which are reactive toward specific amino acid side chains.[14]

Received: July 26, 2016
Accepted: January 4, 2017
Published: January 4, 2017

DOI: 10.1021/acs.analchem.6b02875

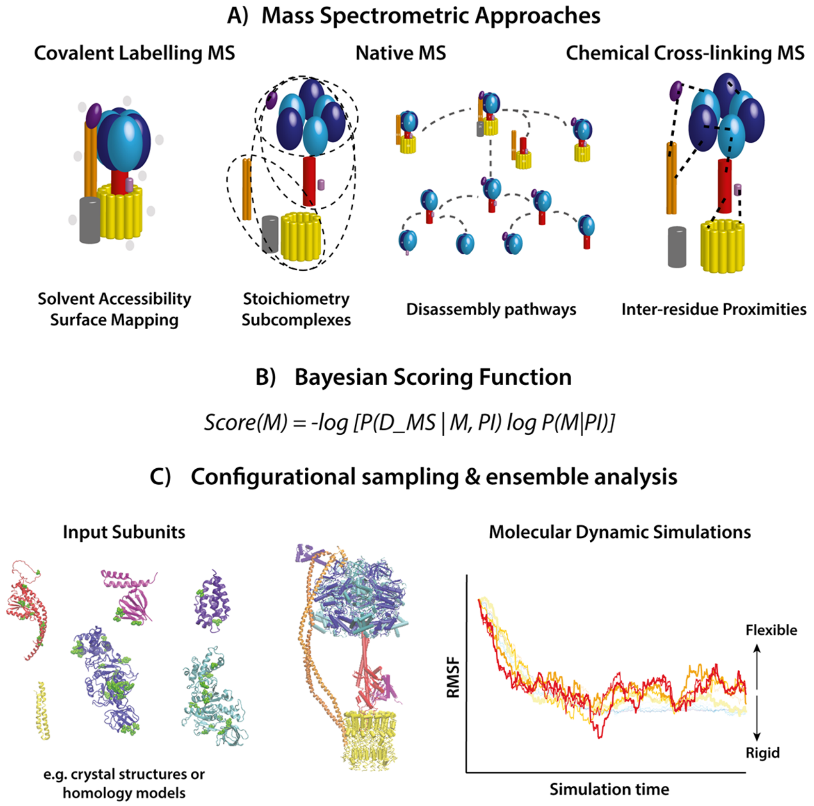

A) Mass Spectrometric Approaches
Covalent Labelling MS
Native MS
Chemical Cross-linking MS
Solvent Accessibility
Surface Mapping
Stoichiometry
Subcomplexes
Disassembly pathways
Inter-residue Proximities
B) Bayesian Scoring Function
Score(M) = -log [P(D_MS | M, PI) log P(M|PI)]
C) Configurational sampling & ensemble analysis
Input Subunits
e.g. crystal structures or
homology models
Molecular Dynamic Simulations
RMSF
Flexible
Rigid
Simulation time

Figure 1. Strategy for protein assembly modeling. (A) Solvent accessibility, inter-residue proximities and disassembly pathways are encoded into modeling restraints. (B) A Bayesian scoring function is employed to build an ensemble of models that match the input data. (C) A representative structure within the top scoring models is subjected to MD simulations enabling to probe the conformational dynamics of the assembly.

Diethylpyrocarbonate (DEPC), employed in this study, was initially used to modify histidine residues. However, DEPC also modifies, with different reactivity, lysine, arginine, tyrosine, threonine and cysteine residues.[21,22] It is an efficient labeling reagent and can probe up to 30% of the protein amino acid sequence. Under acidic and basic conditions or in the presence of nucleophiles, however, DEPC labeling is reversible[23] and experimental conditions have to be carefully optimized.[24]

Structural modeling of proteins and their assemblies includes various computational techniques such as homology modeling, coarse-grained modeling, docking studies or structure prediction.[25−28] In addition, computational simulations can improve our understanding on the dynamic behavior of proteins and their ligands in solution[29] or in the gas phase.[30] The combination of MS approaches and computational methods is increasingly used to study protein complex structures and dynamics. Recent success of hybrid approaches is demonstrated by novel structures of the proteasome,[31,32] the ribosome,[33,34] eukaryotic initiation factors,[35,36] amyloid oligomers,[37] and ATP synthases.[38] A milestone in integrative analysis was the merging of complementary methods[39] and their integration with molecular electron microscopy (EM) maps[35] enabling atomic-level characterization of protein complexes.

We introduce a strategy to study protein complex dynamics by extending the structural toolbox and integrating covalent labeling, cross-linking and native MS with computational modeling. For this, we convert the respective MS data into modeling restraints, which in turn were used to inform a scoring function for generating candidate model structures, while we analyze the prospective models using molecular dynamic simulations (Figure 1). We exemplify this strategy on four well-characterized protein complexes, tryptophan synthase, carbamoyl phosphate synthetase (CPS), the RvB1/RvB2 complex and the catalytic core of cATPase, for which crystal structures are available (Figure S1). We then utilize available information from previous studies together with novel findings on surface accessibility obtained here from covalent labeling and generate a model of the intact F-type ATP synthase purified from spinach chloroplasts. We also subject the top-scoring model to molecular dynamics simulations and identified dynamic and flexible regions within the macromolecular assembly, delivering insights into its function as a nanomotor. The strategy described here is applicable to any protein assembly and provides new opportunities in structural biology linking macromolecular models and their structural dynamics.

■ EXPERIMENTAL SECTION

Protein Purification. Purified tryptophan synthase was a gift of I. Schlichting, Max Planck Institute for Medical Research, Heidelberg, Germany. The RvB1/RvB2 complex was a gift of Karl-Peter Hopfner, Ludwig Maximilian University, Munich, Germany. CPS was provided by F. Raushel, Texas A&M University, College Station. cATPase was purified from spinach leaves and reconstituted in DDM detergent micelles as described previously.[12,40]

DEPC Labeling. Approximately 10 μM of the purified protein complexes were incubated with 8.75, 17.5, 35, or 70 μM DEPC for 1 min at 37 °C. The reaction was quenched by addition of 10 mM imidazole. After quenching the reaction mixture was kept on ice. The proteins were then precipitated with ethanol for 2 h and subsequently digested with trypsin in the presence of RapiGest (Waters) according to manufacturer's protocols.

LC-MS/MS. Dried peptides of cATPase and tryptophan synthase were dissolved in 1% (v/v) formic acid and separated by nanoflow-liquid chromatography on an Dionex UltiMate 3000 RSLC nano System (Thermo Scientific); mobile phase A, 0.1% (v/v) formic acid (FA); mobile phase B, 80% (v/v) acetonitrile 0.1% (v/v) FA. The peptides were loaded onto a precolumn (HPLC column Acclaim PepMap 100, C18, 100 μm I.D. particle size 5 μm; Thermo Scientific) and separated on an analytical column (50 cm, HPLC column Acclaim PepMap 100, C18, 75 μm I.D. particle size 3 μm; Thermo Scientific) at a flow rate of 300 nL/min with a gradient of 5−80% solvent B over 80 min. Peptides were directly eluted into an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Scientific).

MS conditions were: spray voltage of 1.6 kV; capillary temperature of 180 °C; normalized collision energy 35% (q = 0.25, activation time 30 ms). The LTQ-Orbitrap XL was operated in data-dependent mode. MS spectra were acquired in the orbitrap (m/z 300−2000) with a resolution of 30 000 at m/z 400 and an automatic gain control target of 10^6. The five most intense ions were selected for CID fragmentation in the linear ion trap at an automatic gain control target of 30 000. Previously selected ions were dynamically excluded for 30 s. Singly charged ions as well as ions with unrecognized charge state were also excluded. Internal calibration of the orbitrap was performed using the lock mass option.[41]

Peptides and labeled sites were identified using MassMatrix Database Search Engine.[42] Search parameters were as follows: tryptic peptides with a maximum of two missed cleavage sites; carbamidomethylation of cysteine, oxidation of methionine and DEPC-labeled serine, threonine, tyrosine and histidine as variable modifications; mass accuracy filter, 10 ppm for precursor ions, 0.8 Da for fragment ions; minimum pp and pp2 values 5.0, minimum pptag 1.3.

Dried peptides of RvB1/2 and CPS complexes were dissolved in 2% (v/v) ACN, 0.1% FA and separated by nanoflow-liquid chromatography on an Dionex UltiMate 3000 RSLC nano System (Thermo Scientific); mobile phase A, 0.1% (v/v) formic acid (FA); mobile phase B, 80% (v/v) acetonitrile/0.1% (v/v) FA. The peptides were loaded onto a precolumn (HPLC column Acclaim PepMap 100, C18, 100 μm I.D. particle size 5 μm; Thermo Scientific) and separated on an analytical column (50 cm, HPLC column Acclaim PepMap 100, C18, 75 μm I.D. particle size 3 μm; Thermo Scientific) at a flow rate of 300 nL/min with a gradient of 8−90% solvent B over 62 min. Peptides were directly eluted into a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific).

MS conditions were as follows: spray voltage of 1.6 kV; capillary temperature of 250 °C; normalized collision energy 30. The Q Exactive Plus mass spectrometer was operated in data-dependent mode. MS spectra were acquired in the orbitrap (m/z 350−1600) with a resolution of 70 000 and an automatic gain control target of 3×10^6. The 20 most intense ions were selected for HCD fragmentation in HCD at an automatic gain control target of 1×10^5. Previously selected ions were dynamically excluded for 30 s. Singly charged ions, as well as ions with unrecognized charge state, were also excluded. Internal calibration of the orbitrap was performed using the lock mass option.[41]

Peptides and labeled sites were identified using Mascot Search Engine v2.3.02. Search parameters were: Tryptic peptides with a maximum of two missed cleavage sites. Carbamidomethylation of cysteine, oxidation of methionine and DEPC-labeled serine, threonine, tyrosine, and histidine as variable modifications. Mass accuracy filter: 10 ppm for precursor ions, 0.02 Da for fragment ions.

Chemical Cross-Linking of Tryptophan Synthase. Twenty microliters of 20 μM tryptophan synthase were incubated with 20 μL of 2.5 mM bis(sulfosuccinimidyl)suberate (BS3) cross-linker for 1 h at 25 °C at 350 rpm in a thermomixer. After cross-linking, proteins were precipitated with ethanol and digested with trypsin in the presence of RapiGest (Waters) according to manufacturer's protocols. Cross-linked peptides were further separated using SCX Stage Tips (Thermo Scientific) according to the manufacturer's protocol. Peptides were then analyzed by MS and identified as described previously.[12]

Chemical Cross-Linking of CPS and RvB1/B2 Complexes. Ten microliters of 10 μM CPS and 5 μL of 25 μM RvB1/2 were incubated with various concentrations of BS3 cross-linker (final concentrations = 0.5, 0.83, and 1.25 mM) for 1 h at 25 °C at 350 rpm in a thermomixer. Cross-linked proteins were separated by gel electrophoresis (NuPAGE, Invitrogen) and digested in gel as described.[43] Peptides were then analyzed by MS and identified as described previously.[12]

Native Mass Spectrometry. Native MS experiments on tryptophan synthase, CPS, and RvB1/2 were performed on a quadrupole time-of-flight mass spectrometer (Synapt G2Si HDMS, Waters Corp., Manchester, UK). Ten micromolar purified sample was buffer-exchanged in 200 mM ammonium acetate and electrosprayed using gold coated glass capillaries prepared in-house.[44] Typical MS parameters were capillary voltage 1.5−1.7 kV, sampling cone voltage 25−40 V, collision voltage 20 V, bias voltage 20 V, trap collision energy 5 V. MS spectra were processed and analyzed using Masslynx 4.1 (Waters). The spectra were calibrated externally using CsI. Backing pressure: 3.84 mbar. Trap: 0.04 mbar. Helium cell: 3.5 mbar. IMS: 2.6 mbar.

In solution disruption was performed by addition of an organic solvent to the protein complex in ammonium acetate (AA) buffer as described elsewhere.[45] Subcomplexes were generated using 10−40% methanol, dimethyl sulfoxide (DMSO), and acetonitrile (ACN).

Modeling Restraints from Covalent-Labeling MS. Solvent accessibility information from covalent labeling followed by MS was converted into modeling restraints using in-house developed code (https://github.com/apolitis/covalent_labelling_MS). This code iteratively estimates the solvent accessible surface area (SASA) for each residue within all models generated using our sampling algorithm. To calculate the SASA on the surface of each residue we simulated the rolling motion of sphere using a solvent accessible surface function (see Figure S-3). In this function the probe radius of the sphere was 1.8 Å and 5.0 sampling density/ Å^2 for area estimation. The function uses a set of nodal points attributed by xyz coordinates and radius to compute the SASA values. Overall, we report a dimensionless SASA ratio defined as

$$\mathrm{SASA}_x = \frac{\text{accessible surface area of residue } x}{\text{total surface area of residue } x}$$

The returned SASA value per residue is implemented as a structural restraint using a threshold value of 0.25, where if $\mathrm{SASA}_x > 0.25$, then the residue x is exposed, or if $\mathrm{SASA}_x < 0.25$, then the residue x is buried, where x denotes the amino-acid residue.

We iteratively applied this algorithm to all structural models of cATPase, tryptophan synthase, CPS, and RvB1/B2 generated using our Monte Carlo-based strategy. Briefly, we used the list of labeled residues from our covalent labeling mass spectrometry experiments (Tables S4–S7) to interrogate the structural models by satisfaction of modeling restraints. A model was considered if it satisfies the restraint for a specific labeled residue x (histidine, threonine, tyrosine or serine) when the SASA for this residue is greater than 0.25, whereas it violates such restraint for SASA if less than 0.25. For each model structure generated we examined all restraints corresponding to labeled residues and the total score was calculated as

$$S_{\mathrm{SASA},i} = 1 - \frac{\mathrm{RS}}{\mathrm{RT}}$$

where $S_{\mathrm{SASA},i}$ is the score for each model structure i (i = 1, 2, 3, ...) which takes values 0 or 1, RS the number of covalent labeling restraints satisfied in the structure and RT the number of all restraints used, which correspond to the labeled residues from covalent labeling experiments.

The SASA scoring algorithm was implemented within the Integrative Modelling Platform (IMP).[25]

Integrative Modeling. We used an integrative modeling strategy for MS data.[36,39] Structural models of the assemblies were generated using a Monte Carlo search algorithm developed in-house and implemented into IMP.[25] The model building was guided by a scoring function, which estimates the probability of a structural model given existing knowledge of the investigated system and the MS data acquired. The posterior probability $P(M|D_{\mathrm{MS}}, \mathrm{PI})$ for MS Data (D_{MS}) and prior information (PI) is

$$P(M|D_{\mathrm{MS}}, \mathrm{PI}) \propto P(M|\mathrm{PI})P(D_{\mathrm{MS}}|M, \mathrm{PI})$$

where $P(M|\mathrm{PI})$ is the prior, the probability of a model given only existing information on the system and $P(D_{\mathrm{MS}}|M, \mathrm{PI})$ is the likelihood function, expressed as the probability of observing MS data given a structural model and knowledge of the system in question. The score is calculated as the negative logarithm of the likelihood and the existing information (called prior)

$$\mathrm{score}(M) = -\log[P(D_{\mathrm{MS}}|M, \mathrm{PI})\log P(M|\mathrm{PI})]$$

The most likely structural model scores higher according to the posterior distribution. The prior $P(M|\mathrm{PI})$ is the prior probability $P(M)$ accounting for intersubunit connectivities, solvent accessibility, distance restraints and an additional parameter composed of uncertainties; these are the false positives for native MS, cross-linking and covalent labeling MS. The likelihood function $P(D_{\mathrm{MS}}|M, \mathrm{PI})$ for a data point of a data set D of experimentally measured connectivites (native MS, cross-linking MS), distance restraints (cross-linking MS) and solvent accessibilities (CL-MS) is given as

$$P(d_n|Y, \alpha_n, \sigma_n) = \exp\left(-\frac{(d_n - f_n[Y, \omega])}{2\sigma_n^2}\right)$$

where Y is the structure coordinates, σ_n the uncertainty, α_n denotes other parameters, such as ambiguities due to flexibilities, and ω is the weight. The forward function (f_n) predicts the data points, that is, randomly picking a residue that is solvent exposed for a given time point in the experimental measurement (CL-MS) and adopts a conformation consistent with the given connectivities and distance restraints. The uncertainty corresponds to the data points from both measurements that are inconsistent with the structure Y.

We judged the uniqueness of the ensemble of generated models by performing ensemble analysis (e.g., clustering of best-ranking solutions), and the final solution was selected from the major cluster.[45] The Visual Molecular Dynamics (VMD) and the UCSF Chimera packages were used for visualization of the model structures.[46]

Distance Restraints from Cross-Linking MS. Upper bound distance restraints (35 Å) specified from the identified cross-links by applying a cross-linking strategy followed by MS.[36,43] The individual links were implemented into our modeling approach enabling us to guide the search for candidate model structures that fit the input MS data.

Simulations in Explicit Solvent. Explicit solvent MD of the ATPase protein complex were performed and analyzed using the GROMACS 4.6 program[47] using the Amber99sb*-ildn force field parameters.[48] The input structure of the F1 cATPase was assembled from its individual components (crystal structure and homology models) using an MS-restrained strategy as described elsewhere.[39] The initial complex structure, consisting of 56,826 protein heavy atoms, was solvated and minimized in a dodecahedral periodic box of 952 838 TIP3P water molecules[49] with a minimum distance of 1.0 nm between any protein atom and the periodic box. The system charge was neutralized by adding 75 sodium counter-ions to the solvent. The equations of motion were integrated using the leapfrog method[50] with a 2 fs time step. The equilibration protocol hereafter outlined was used: an initial 500 steps of steepest descent energy minimization in solution. This was followed by an equilibration of the system in the canonical ensemble with harmonic positional restraints on the protein heavy atoms using a force constant of 10 000 kJ/mol/nm^2 and gradually reduced to 1000 kJ/mol/nm^2, while increasing the temperature from 50 to 300 K at a constant volume. During this NVT ensemble equilibration, the Berendsen algorithm[50] was employed to regulate the temperature and pressure of the system with coupling constants of 0.2 and 0.5 ps, respectively. A 5 ns NVT equilibration run at 300 K and 1 bar was then performed, following with 2 ns of equilibration in NPT conditions. After successful equilibration of the system, the cATPase complex was then simulated for 40 ns under constant pressure and temperature conditions. Temperature was regulated using the velocity-rescaling algorithm,[51] with a coupling constant (τ) of 0.1. All protein covalent bonds were frozen with the LINCS method,[52] while SETTLE[53] was used for water molecules. Electrostatic interactions were calculated with the particle mesh Ewald method,[54] with a 1.4 nm cutoff for direct space sums, a 0.12 nm FFT grid spacing and a four-order interpolation polynomial for the reciprocal space sums. van der Waals interactions were measured using a 1.4 nm cutoff. The neighbor list for noncovalent interactions were updated every five integration steps.

Modeling of the Peripheral Stalk. We performed homology modeling of the peripheral stalks using the MODELLER package.[55] We obtained a reliable homology model (sequence identity >25%) using as templates the *Thermus thermophilus* H-type (PDB ID 3V6I) and bovine mitochondrial (PDB ID 2CLY) ATPases. To compensate for

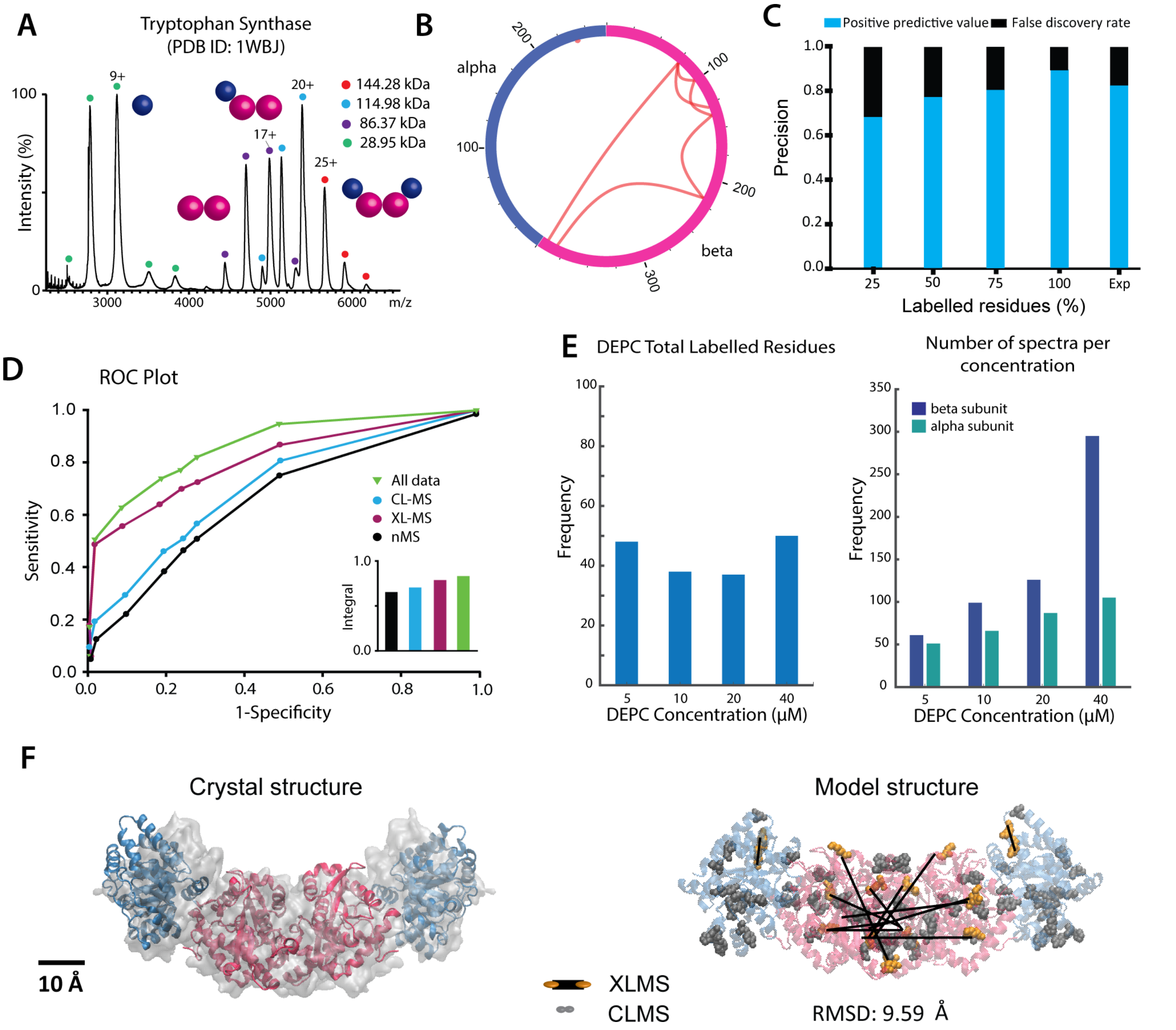


Figure 2. Benchmark analysis on tryptophan synthase. (A) Native MS of the intact complex yielded disassembly pathway. (B) Cross-linking circular plot. (C) The precision of the methodology was estimated by calculating positive predictive values (PPVs) for different amount of theoretical covalent labeling restraints while we use the experimentally available restraints from native and cross-linking MS. (D) ROC curves, plotting the true positive rate (sensitivity) versus false positive rate (1-specificity), to evaluate the confidence level of the restraints. (E) Peptide level analysis plotting the frequency of the DEPC total labeled residues and the number of spectra per concentration shows increase in the labeling residues/spectra with increased concentrations (F) Representative model of the tetrameric tryptophan synthase and its corresponding crystal structure. Inter-residue proximities (XL-MS) and residue solvent accessibilities (CL-MS) are highlighted. The structural similarity of the model to the X-ray structure was assessed using their pairwise r.m.s.d.

the lack of lower part of stalks linking the core with the transmembrane ring, we modeled in the helices using as guide the distance estimated from the missing residues. Homology models for of ε, δ, and γ subunits were also utilized as previously described.[12]

Modeling Scripts, Data, and Results. Our integrative method was implemented in the open source IMP software package (http://integrativemodeling.org). The input data files, modeling scripts, and output models for the tryptophan synthase and cATPase complex are available at https://github.com/apolitis/covalent_labelling_MS. This will allow keen scientists to use our data and/or integrate with their own results for protein assembly modeling.

■ RESULTS AND DISCUSSION

Integrating Covalent Labeling into Computational Modeling. We assessed the predictive power of our integrative method for three-dimensional protein modeling based on structural MS restraints on four protein complexes previously characterized by X-ray crystallography: the 143 kDa tryptophan synthase from *Salmonella typhimurium* (PDB ID 1WBJ),[56] the $\alpha_4\beta_4$ CPS (PDB ID 1BXR, ~640 kDa), the double-heterohexameric ring RvB1/2 (PDB ID 4WVY; 621 kDa) (Figure S1) and the hexameric $\alpha_3\beta_3$-head of cATPase from *Spinacia oleracea* (PDB ID 1FX0; ~328 kDa).[57] Covalent labeling using DEPC, cross-linking with BS3 (Figure S-2) and native MS (Figure S-3) allowed us to label serine, threonine, tyrosine, and histidine residues on the surface of the complex,

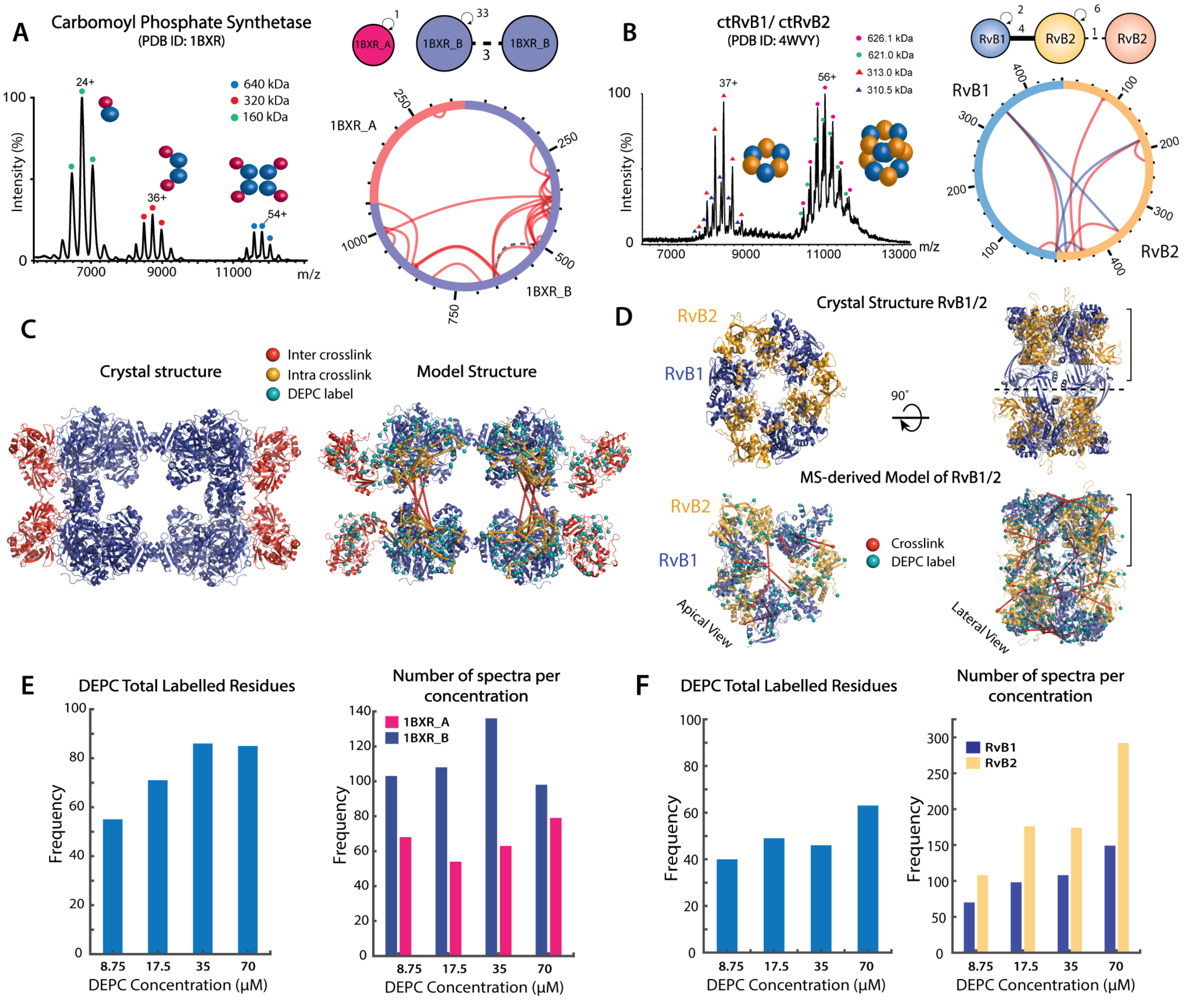


Figure 3. Benchmark analysis on phosphate synthetase (CPS) and RVB1/B2 heterododecamer (A, B) Native and cross-linking MS reveal distinct (sub)complexes and intra- and intersubunit amino-acid level proximities. Identified oligomeric cross-links are shown in the inset small circular (C, D). Integrative modeling results in models in good agreement with the reference crystal structures. (E, F) Peptide level analysis plotting the frequency of the DEPC total labeled residues and the number of spectra per concentration shows increase in the labeling residues/spectra with increased concentrations.

map cross-linked lysines and define stable subcomplexes, respectively. Overall, we identified inter- and intrasubunit cross-links (Tables S1–S3), up to 151 labeled residues (Tables S4, S5, and S7) and several (sub)complexes for tryptophan synthase, CPS and RvB1/2, respectively (Figure 2A, 3A and B, and S4). For the cATPase hexameric head we used previously published cross-linking results and native mass spectrometry[12] and in this study identified 58 solvent-exposed residues (Table S7). With the complementary MS-based data in hand, we applied a computational workflow by first encoding our data into modeling restraints (Figure 1A) and then using a scoring function to guide generation of structural models (Figures 1B and S5 and Experimental Section).

The covalent labeling experiments enabled solvent accessible surface area (SASA) restraints. A SASA restraint is considered to be satisfied if, for each experimentally labeled residue, the theoretically predicted SASA is greater than 25% (Figure S6 and Experimental Section). We plotted the fraction of satisfied residues on the corresponding crystal structure as a function of the percentage of SASA providing justification for its use as a lower bound restraint for modeling (Figure S6). The cutoff is defined as the highest SASA score that gives <10% false positives while the true positives remain over 80% of the total models. The cross-linking experiments allowed upper bound distance restraints (<35 Å).[39] This distance breaks down into 11.4 Å for the linker (BS3), approximately 13 Å for the two lysine side chains and an additional tolerance of 10 Å accounting for flexibility due to protein's motion. The resulting models from application of these restraints were considered to match the data and added to the ensemble that is passed on to the next stage for additional analysis. Clustering analysis[36,45] revealed an ensemble of models with close similarity to the reference crystal structure (r.m.s.d. ranging from 9 to 15 Å) (Figures 2 and 3C and D). Finally, a representative structure from the ensemble was used as a starting model for explicit solvent MD simulations (Figure 1C).

Evaluation of the modeling approach. Having established the validity of using SASA restraints for modeling, we examined the ability of our approach to predict high-resolution models using different levels of theoretically labeled residues ranging from 25% to 100%. Complete theoretical labeling information was extracted from the corresponding crystal structure by assuming as labeled those serines, threonines, tyrosines and histidines with theoretical SASA larger than 25%. The residues with SASA less than 25% were considered buried and therefore were not processed further. A model is defined "good" when it exhibits high structural similarity to the reference crystal structure as calculated by Cα atoms (r.m.s.d. < 12 Å).[39] We estimated ~90% positive predictive value (PPV) within the top-scoring models when all theoretical information was used and a difference of less than 10% PPV when the experimental available data were used (Figure 2C).

To investigate the merit of modeling restraints in predicting the correct structure of the three training complexes (tryptophan synthase, CPS and RvB1/2), we determined receiver-operating characteristic (ROC) plots for the MS techniques employed (Figures 2D and S7). This enabled us to test the ability of our method in generating correct model structures on systems with diverse topological features that include symmetry, ring-like geometries and heteromeric subunits. The area under each curve was determined as a measure of the information content of each restraint, where 0.5 indicates that correct and incorrect structures cannot be discriminated.[45,58] The ROC plots of all three complexes studied here show that inclusion of solvent accessibility restraints from covalent labeling markedly increased (~8–11%) the accuracy of structural prediction (Figures 2D and S7). Increasing the accuracy of the predictions by approximately 10% is an important improvement of the method particularly when building models of multiprotein systems requires a large number of models. For instance, if 10 000 models are generated, a 10% higher accuracy means that the structural prediction leads to 1000 less false positive and false negative models and therefore allows an increased number of "good models" within the top-scoring model structures. This is particularly important for assembling multicomponent systems in a stepwise manner where degeneracy can significantly hinder the accuracy of the resulting predictions.

Concentration Dependence of DEPC Labeling. To assess the effect of concentration on the labeling efficiency, we covalently labeled solvent accessible residues on the three training complexes using a range of labeling concentrations (8.75–70 μM) (Figures 2E, 3E and F, and S8). Using DEPC we targeted histidine, threonine, tyrosine and serine residues, covering ~15–20% of the complex sequence. We plotted the number of experimentally labeled residues over the range of experimental concentrations revealing a significant increase of the labeled residues (10–30%) at higher concentrations (Figures 2E, 3E and F, and S8). A similar trend was found by counting the total number of spectra measured at each concentration used for the experiments (Figures 2E and 3E and F). We overall estimated a 5–10% of the total residues uniquely identified in the two lower concentrations. For modeling purposes we accounted for all labeled residues appearing in at least one concentration.

To study the accuracy and precision of SASA restraint from covalent labeling followed by MS, we projected the labeled serines, threonines, tyrosines and histidines on the crystal structures of tryptophan synthase and the cATPase head and examined their SASA (Figure 4). We revealed high accuracy (>85%) and precision (>80%), confirming the lower bound SASA as a confident restraint for modeling in all benchmark complexes examined in the study.

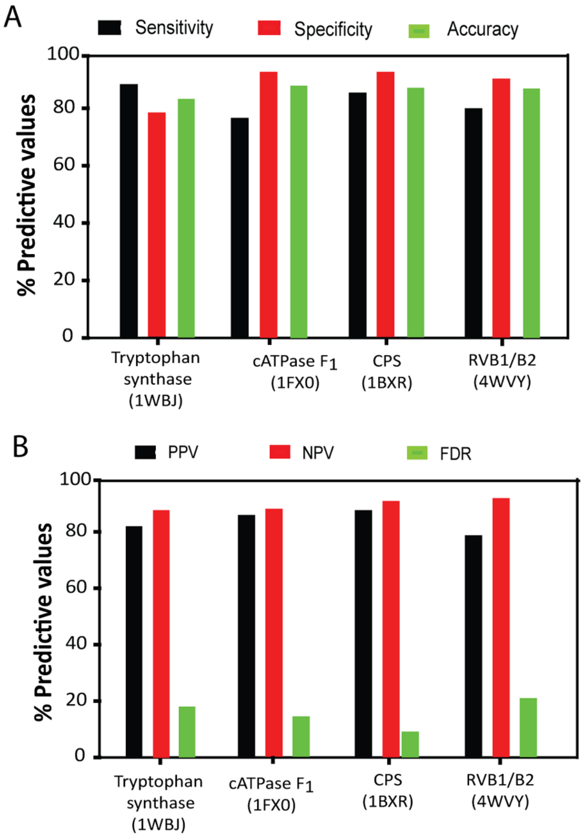


Figure 4. Benchmark analysis of SASA restraint derived from covalent labeling MS experiments. We assessed (A) the sensitivity, specificity, and accuracy and (B) the negative predictive value (NPV), positive predictive value (PPV), and false discovery rate (FDR) using SASA as a restraint through the existing crystal structures of tryptophan synthase, CPS and RVB1/B2 and cATPase (F_1) as references models. SASA area for all residues in the above structures were calculated and compared to the identified labeled sites from covalent labeling MS. A positive or correctly labeled residue is defined as a residue with SASA more than 0.25. False positives or incorrectly labeled residues are identified with calculated SASA below 0.25. Nonexperimentally labeled residues with calculated SASA below 0.25 in the corresponding structure represent true negatives, while false negatives have SASA above 0.25. Sensitivity = TP/(TP + FN), specificity = TN/(TN + FP), accuracy = (TP + TN)/(TP + FP + TN + FN), FDR = FP/(TP + FP), NPV = TN/(TN + FN), and PPV = TP/(TP + FP). TP: True positive. FP: False positive. FN: False negative. TN: True negative.

Solvent Accessibility and Modeling of cATPase. Next, we assembled a model of the intact cATPase from *Spinacia oleracea*. The cATPase generates ATP from ADP and inorganic phosphate using an electrochemical proton gradient across the thylakoid membrane.[57] Its stoichiometry is $\alpha_3\beta_3\gamma\delta\varepsilon$–I–II–$\mathrm{III}_{14}$–IV;[12] however, structural information is limited to crystal structures of the soluble catalytic head $(\alpha_3\beta_3)$[57] and the III_{14} transmembrane ring.[59] Little is known about the structural dynamics of the individual subunits within the assembly. From studies on other ATP synthases, we expect enhanced dynamics for the peripheral stalk, a stator that links soluble and membrane domains and counteracts the torque from "wobbling" of the soluble head during motor rotation[60] of the γ-subunit.[61]

We covalently labeled solvent accessible residues on the surface of cATPase (Figure 5A and B). Different concentrations of DEPC (8.75–70 μM), yielded 75 labeled residues (Table S7) in all protein subunits except ring subunit III. The lack of labeled residues in the membrane ring subunit is attributed to the protective layer of the detergent micelle. However, we identified one labeling site on membrane subunit IV (Tyr 160).

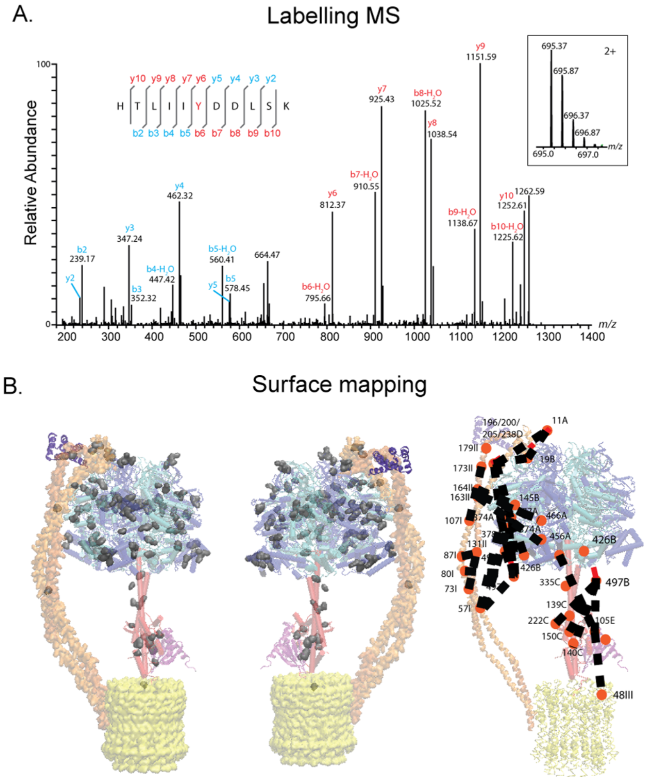


Figure 5. Covalent labeling and cross-linking of cATPase. (A) Example spectrum of a labeled cATPase peptide. *B*- and *y*-ions are assigned. Fragment ions containing the DEPC-modification are shown in red. (B) Covalent labeling analysis reveals solvent accessible residues (gray space fillings) on the surface of the intact enzyme (left and middle panel. Complementary structural information was obtained from chemical cross-linking (right panel).

We used cross-links and dissociation pathways from native MS reported previously[12] providing 11 subcomplexes and a connectivity map (Figures S9 and S10).[36] Covalent labeling data were encoded into modeling restraints and together with distance restraints from cross-linking, enabled us to map the inter-residue proximities and SASA of the cATPase (Figure 5B). By employing our restraint-based modeling approach, we brought together complementary restraints (Experimental Section) allowing us to assemble a structural model of the cATPase (Figure 5B). As input we used the crystal structure of $\alpha_3\beta_3$ and ring III_{14} subcomplexes and homology models of subunits I, II, δ, γ, and ε (Experimental Section). We were unable to position subunit IV as only one residue was labeled and no cross-links or subcomplexes were observed. However, we unambiguously defined the orientation and proximities of the other subunits showing a slight tilting (~4°) of the central axis of the catalytic head with respect to the axis of the membrane ring,[60] consistent with crystal structures of mitochondrial ATPase[62−65] and a model of the V-type ATPase (Figure S11).[60,38,66]

MD Simulations Reveal Flexibility of cATPase. We used the assembled model of cATPase as a starting structure for explicit solvent MD simulations allowing us to examine the architecture and dynamics of the enzyme. Similar to other ATPases, the cATPase γ stalk subunit consists of a globular domain interacting with the α/β-head and δ subunits. Their extended α helices link the F_1 (head) and F_O (transmembrane) domains. To allow for movements of the rotor during the catalytic cycle, the peripheral stalk must exhibit conformational flexibility. We therefore performed simulations for the F_1 domain ($\alpha_3\beta_3\gamma\delta\varepsilon$-I II) (Figure S12 and Experimental Section). Consistent with other ATPases[60,66] we revealed significant flexibility of subunits I, II, γ and δ as calculated by the r.m.s.d. and r.m.s.f. (Figures 6A and B and S13). We projected the r.m.s.f profiles on the surface of the cATPase visualizing dynamic regions in the assembly. Particularly flexible regions were found within the peripheral stalk and γ subunits (Figure 6B). In line with a recent study[67] these regions are connected through a rigid section, which may allow the stalk and the γ subunit to retain their rigidity whereas accommodating the wobbling motion from rotary catalysis. The γ subunit contains an additional loop compared with other ATPases, which is responsible for its deactivation in the absence of light.[61] Interestingly, the fluctuation of the γ subunit predicted by our method includes the ~40 amino acid long loop segment (residues 197–240). It is interesting to speculate that the flexibility of this loop may be related to its role in activation/deactivation of the enzyme suggesting conformational changes during transition from one state to another.

To reduce the high dimensionality of the MD trajectories and to identify the dominant molecular motions of the peripheral stalks, we performed principal component analysis (PCA). We showed that both peripheral stalks undergo a "bending" motion with particular flexible regions located at the initial and terminal ends of the stalks (Figure 6C). The flexibility of stalks is likely to be an intrinsic property enabling them to adjust during the catalytic rotation of the molecular motor. This is consistent with the twisting motion of the catalytic head of an A-ATPase proposed previously[60] and may be related to the intermediate states of rotary ATPases during ATP synthesis.[68,69]

■ CONCLUSIONS

We presented here a strategy for interrogating the structure and dynamics of multiprotein assemblies. These assemblies are difficult to study by traditional tools, which limits our knowledge of their function. In our strategy, we incorporated modeling restraints derived from covalent labeling MS in the form of SASA. We integrated, using a scoring function, the SASA restraint with the connectivity and distance restraints from native and chemical cross-linking MS, respectively.

We assessed the predictive power of the method by reconstructing the 3D assembly structure of tryptophan synthase, CPS and RvB1/2 with high accuracy and precision. The integration of a novel combination of MS-based methods markedly increased the predictability of the method as shown by ROC plots and enabled us to suggest a confident model of cATPase, a particularly challenging target in structural biology. We observed a ~10% increase in the overall predictability of the integrative methodology when covalent labeling was added to native and cross-linking MS. Such an increase may have a significant effect in differentiating between closely related states and in cases where ambiguous or incomplete data sets exist. In

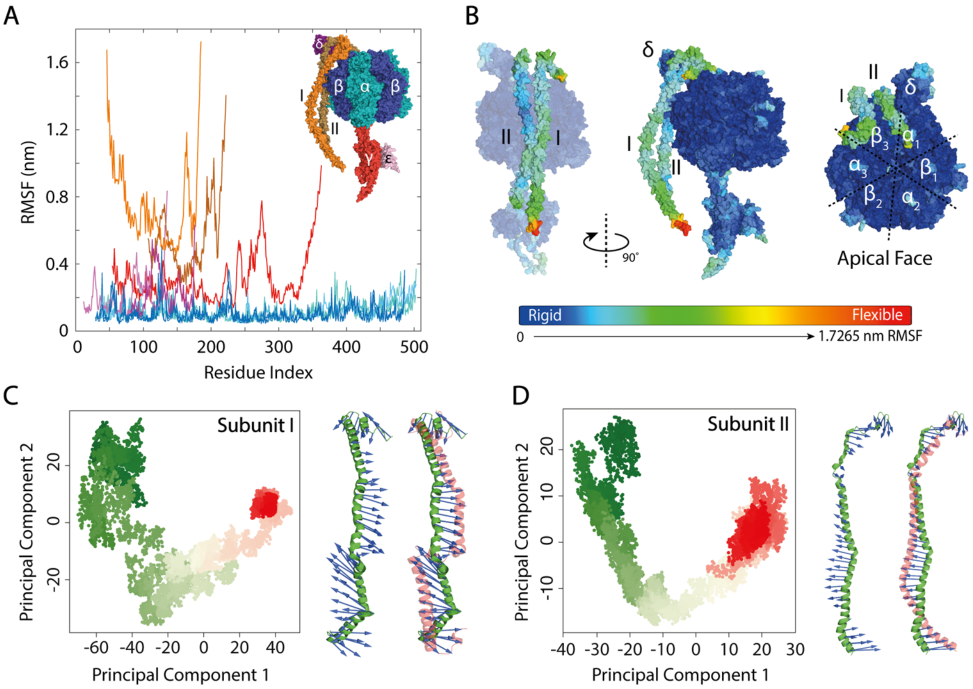


Figure 6. Explicit solvent MD simulations of cATPase (A) Plot of the residue r.m.s.f. within each subunit over time reveals regions of enhanced flexibility (B) Mapping the conformational fluctuation predicted on the surface of cATPase. (C, D) Principal component analysis revealed the local dynamics and directionality of motion within the peripheral stalk subunits. The scatterplot colored with a gradient from green to red indicates "start" and "end" of the simulations, respectively.

principle, our workflow allows the incorporation of every labeling strategy and we expect that the application of complementary techniques targeting different amino acid side chains will improve the predictability even further. Such an increase in predictability can lead to high confident models of multiprotein complexes and is primarily important for those systems where limited information are attained by other biophysical methods, such as the cATPase.

We provided an additional dimension by subjecting the cATPase model structure to solution phase simulations, allowing us to assign flexible regions within the complex. Performing such simulations were possible by the assembly of a confident model of cATPase from our restraint-based strategy, thus demonstrating how static structural predictions and dynamic simulations can be integrated for understanding complex biological systems. Even though the main strength of our strategy is its ability to simultaneously incorporate various labeling methods, it becomes more powerful when combined with high-resolution information on the individual assembly subunits.[70] We envision that the combination of labeling MS with accurate modeling and simulations may be used in future to study many other multiprotein complexes currently eluding structure determination.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b02875.

> X-ray crystal structures of benchmark complexes, tryptophan synthase gel, SDS-PAGE of the cross-linking experiments for CPS and RvB1/B2 complexes, mass spectra of tryptophan synthase, plot of model structure scores for the tryptophan synthase, description of the solvent accessibility algorithm, ROC plots, DEPC labels, benchmark analysis on the hetero-hexameric α/β cATPase, connectivity map of cATPase, model structures of ATPases, molecular dynamic simulations analysis, solution-phase MD simulation of cATPase, identified cross-links for tryptophan synthase, carbamoyl phosphate synthetase, and RvB1/RvB2 complex, labeled amino-acid residues of tryptophan synthase, labeled carbamoyl phosphate synthetase amino-acid residues, labeled RvB1/RvB2 amino acid residues from covalent labeling MS, and labeled cATPase amino-acid residues (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: argyris.politis@kcl.ac.uk.

ORCID

Argyris Politis: 0000-0002-6658-3224

Author Contributions

[#]J.A.M. and A.M.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.S. acknowledges funding from the Federal Ministry for Education and Research (BMBF, ZIK programme, 03Z22HN22), the European Regional Development Funds (EFRE, ZS/2016/04/78115), and the MLU Halle-Wittenberg. A.P. acknowledges funding from the Wellcome Trust (109854/Z/15/Z). We thank Prof. Dame Carol V. Robinson for generous support. We thank Uwe Plessmann for help with Mascot searches.

REFERENCES

(1) Benesch, J. L.; Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011**, *21*, 641−649.

(2) Schmidt, C.; Robinson, C. V. *FEBS J.* **2014**, *281*, 1950−1964.

(3) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell. Proteomics* **2010**, *9*, 1634−1649.

(4) Rappsilber, J. *J. Struct. Biol.* **2011**, *173*, 530−540.

(5) Sinz, A. *Expert Rev. Proteomics* **2014**, *11*, 733−743.

(6) Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem.* **2007**, *76*, 167−193.

(7) Heck, A. J. *Nat. Methods* **2008**, *5*, 927−933.

(8) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. *Chem. Soc. Rev.* **2010**, *39*, 1633−1655.

(9) Jurneczko, E.; Barran, P. E. *Analyst* **2011**, *136*, 20−28.

(10) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. *Anal. Chem.* **2009**, *81*, 248−254.

(11) Morgner, N.; Schmidt, C.; Beilsten-Edmands, V.; Ebong, I. O.; Patel, N. A.; Clerico, E. M.; Kirschke, E.; Daturpalli, S.; Jackson, S. E.; Agard, D.; Robinson, C. V. *Cell Rep.* **2015**, *11*, 759−769.

(12) Schmidt, C.; Zhou, M.; Marriott, H.; Morgner, N.; Politis, A.; Robinson, C. V. *Nat. Commun.* **2013**, *4*, 1985.

(13) Sinz, A.; Arlt, C.; Chorev, D.; Sharon, M. *Protein Sci.* **2015**, *24*, 1193−1209.

(14) Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785−815.

(15) Engen, J. R. *Analyst* **2003**, *128*, 623−628.

(16) Beveridge, R.; Migas, L. G.; Payne, K. A.; Scrutton, N. S.; Leys, D.; Barran, P. E. *Nat. Commun.* **2016**, *7*, 12163.

(17) Katta, V.; Chait, B. T.; et al. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214−217.

(18) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158−170.

(19) Konermann, L.; Rodriguez, A. D.; Sowole, M. A. *Analyst* **2014**, *139*, 6078−6087.

(20) Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514−3543.

(21) Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895−2904.

(22) Miles, E. W. *Methods Enzymol.* **1977**, *47*, 431−442.

(23) Melchior, W. B., Jr.; Fahrney, D. *Biochemistry* **1970**, *9*, 251−258.

(24) Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 899−907.

(25) Russel, D.; Lasker, K.; Webb, B.; Velazquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. *PLoS Biol.* **2012**, *10*, e1001244.

(26) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779−815.

(27) Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *PLoS One* **2010**, *5*, e12080.

(28) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *Nucleic Acids Res.* **2005**, *33*, W363−367.

(29) Hospital, A.; Goni, J. R.; Orozco, M.; Gelpi, J. L. *Adv. Applic. Bioinf. Chem.* **2015**, *8*, 37−47.

(30) van der Spoel, D.; Marklund, E. G.; Larsson, D. S.; Caleman, C. *Macromol. Biosci.* **2011**, *11*, 50−59.

(31) Lasker, K.; Forster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1380−1387.

(32) Tomko, R. J., Jr.; Taylor, D. W.; Chen, Z. A.; Wang, H. W.; Rappsilber, J.; Hochstrasser, M. *Cell* **2015**, *163*, 432−444.

(33) Greber, B. J.; Bieri, P.; Leibundgut, M.; Leitner, A.; Aebersold, R.; Boehringer, D.; Ban, N. *Science (Washington, DC, U. S.)* **2015**, *348*, 303−308.

(34) Greber, B. J.; Boehringer, D.; Leibundgut, M.; Bieri, P.; Leitner, A.; Schmitz, N.; Aebersold, R.; Ban, N. *Nature* **2014**, *515*, 283−286.

(35) Erzberger, J. P.; Stengel, F.; Pellarin, R.; Zhang, S.; Schaefer, T.; Aylett, C. H.; Cimermancic, P.; Boehringer, D.; Sali, A.; Aebersold, R.; Ban, N. *Cell* **2014**, *158*, 1123−1135.

(36) Politis, A.; Schmidt, C.; Tjioe, E.; Sandercock, A. M.; Lasker, K.; Gordiyenko, Y.; Russel, D.; Sali, A.; Robinson, C. V. *Chem. Biol.* **2015**, *22*, 117−128.

(37) Hall, Z.; Schmidt, C.; Politis, A. *J. Biol. Chem.* **2016**, *291*, 4626−4637.

(38) Zhou, M.; Morgner, N.; Barrera, N. P.; Politis, A.; Isaacson, S. C.; Matak-Vinkovic, D.; Murata, T.; Bernal, R. A.; Stock, D.; Robinson, C. V. *Science* **2011**, *334*, 380−385.

(39) Politis, A.; Stengel, F.; Hall, Z.; Hernandez, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R. *Nat. Methods* **2014**, *11*, 403−406.

(40) Varco-Merth, B.; Fromme, R.; Wang, M.; Fromme, P. *Biochim. Biophys. Acta, Bioenerg.* **2008**, *1777*, 605−612.

(41) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4*, 2010−2021.

(42) Xu, H.; Hsu, P. H.; Zhang, L.; Tsai, M. D.; Freitas, M. A. *J. Proteome Res.* **2010**, *9*, 3384−3393.

(43) Schmidt, C.; Robinson, C. V. *Nat. Protoc.* **2014**, *9*, 2224−2236.

(44) Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2*, 715−726.

(45) Hall, Z.; Politis, A.; Robinson, C. V. *Structure* **2012**, *20*, 1596−1609.

(46) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(47) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(48) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.

(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(51) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(52) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(53) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.

(54) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(55) Webb, B.; Sali, A. *Curr. Protoc. Bioinf* **2014**, *47*, 5.6.1−5.6.32.

(56) Kulik, V.; Hartmann, E.; Weyand, M.; Frey, M.; Gierl, A.; Niks, D.; Dunn, M. F.; Schlichting, I. *J. Mol. Biol.* **2005**, *352*, 608−620.

(57) Groth, G.; Pohl, E. *J. Biol. Chem.* **2001**, *276*, 1345−1352.

(58) Alber, F.; Kim, M. F.; Sali, A. *Structure* **2005**, *13*, 435−445.

(59) Vollmar, M.; Schlieper, D.; Winn, M.; Buchner, C.; Groth, G. *J. Biol. Chem.* **2009**, *284*, 18228−18235.

(60) Stewart, A. G.; Lee, L. K.; Donohoe, M.; Chaston, J. J.; Stock, D. *Nat. Commun.* **2012**, *3*, 687.

(61) Samra, H. S.; Gao, F.; He, F.; Hoang, E.; Chen, Z.; Gegenheimer, P. A.; Berrie, C. L.; Richter, M. L. *J. Biol. Chem.* **2006**, *281*, 31041−31049.

(62) Giraud, M. F.; Paumard, P.; Sanchez, C.; Brethes, D.; Velours, J.; Dautant, A. *J. Struct. Biol.* **2012**, *177*, 490−497.

(63) Stock, D.; Leslie, A. G.; Walker, J. E. *Science* **1999**, *286*, 1700−1705.

(64) Watt, I. N.; Montgomery, M. G.; Runswick, M. J.; Leslie, A. G.; Walker, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16823−16827.

(65) Zhou, A.; Rohou, A.; Schep, D. G.; Bason, J. V.; Montgomery, M. G.; Walker, J. E.; Grigorieff, N.; Rubinstein, J. L. *eLife* **2015**, *4*, e10180.

(66) Zhou, M.; Politis, A.; Davies, R.; Liko, I.; Stewart, A.; Stock, D.; Robinson, C. V.; et al. *Nat. Chem.* **2014**, *6*, 208−215.

(67) Papachristos, K.; Muench, S.; Paci, E. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1203−1212.

(68) Abrahams, J. P.; Leslie, A. G.; Lutter, R.; Walker, J. E. *Nature* **1994**, *370*, 621−628.

(69) Walker, J. E. *Angew. Chem., Int. Ed.* **1998**, *37*, 2308−2319.

(70) Schneider, M.; Belsom, A.; Rappsilber, J.; Brock, O. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 152−163.