Reactivity-Based Analysis of Domain Structures in Native Replication Protein A[†]

Jonathan E. Nuss,[‡] Deacon J. Sweeney, and Gerald M. Alter*

Department of Biochemistry and Molecular Biology and Biomedical Sciences Ph.D. Program, Wright State University, Dayton, Ohio 45435

Received November 23, 2005; Revised Manuscript Received June 9, 2006

ABSTRACT: Replication protein A (RPA) is an essential heterotrimeric ssDNA binding protein that participates in DNA repair, replication, and recombination. Though X-ray and NMR experiments have been used to determine three-dimensional structure models of the protein's domain fragments, a complete RPA structural model has not been reported. To test whether the fragment structures faithfully represent the same portions in the native solution-state protein, we have examined the structure of RPA under biologically relevant conditions. We have probed the location of multiple amino acids within the native RPA three-dimensional structure using reactivity of these amino acids toward proteolytic and chemical modification reagents. In turn, we evaluated different structural models by comparing the observed *native* RPA reactivities with anticipated reactivities based on candidate structural models. Our results show that our reactivity analysis approach is capable of critically assessing structure models and can be a basis for selecting the most relevant from among alternate models of a protein structure. Using this analytical approach, we verified the relevance of RPA fragment models to the native protein structure. Our results further indicate several important features of native RPA's structure in solution, such as flexibility at specific locations in RPA, particularly in the C-terminal region of RPA70. Our findings are consistent with reported DNA-free structural models and support the role of conformational change in the ssDNA binding mechanism of RPA.

Replication protein A (RPA)[1] is a single-stranded DNA binding protein that is essential for the processes of DNA replication, repair, and recombination (*1*, *2*). RPA avidly binds to single-stranded DNA substrates with high affinity and low cooperativity (*3*, *4*). This activity is most likely the reason for RPA's diverse array of in vivo functions such as chromatin unwinding (*5*, *6*) and DNA damage recognition (*7*, *8*). In addition, RPA also participates in a host of protein−protein interactions that are essential for forming the macromolecular complexes that replicate, repair, and change DNA (*9*−*11*).

RPA is composed of a large (70 kDa), a medium (32 kDa), and a small (14 kDa) subunit. It has resisted crystallization. High-resolution structures for seven domains of RPA have been determined by NMR and X-ray crystallographic analysis of RPA fragments (*11*−*15*). In fact, a fragment containing the two most avid ssDNA binding domains, which are located in the central portion of the 70 kDa subunit, has been analyzed by X-ray crystallography in the presence and absence of DNA (*13*, *16*). Models of this region [1FGU and 1JMC (Protein Data Bank identifier)] differ depending on whether ssDNA is bound. Though domains are conformationally similar, their juxtaposition is different (*10*). NMR and X-ray crystallographic analysis of fragments from other regions of the protein reveal that RPA is composed of four additional OB folds, including ones at the N-terminal domain of RPA70, at the C-terminal region of RPA70 (DBDC), spanning RPA14, and in the central region of RPA32 (DBDD) (*11*, *14*). An additional domain at the C-terminal portion of RPA32 does not contain an OB fold (*15*). Crystallography has supplied some quaternary structure information as well. For example, the structural model termed 1L1O contains three domains, one from each RPA subunit. These include RPA14, the central domain from RPA32 (DBDD), and the C-terminal domain from RPA70 (DBDC). The complex is referred to as the minimum trimerization core of RPA (*11*).

Despite all of this structural information, questions remain regarding the protein's structure. Specifically, do the fragments' structures accurately represent the conformations of the same portions of the primary structure in the intact protein?

Investigators have probed protein structures with chemical reagents and proteases that label accessible side chains or nick the peptide backbone after recognizing a subset of side chains on the molecule's surface. In fact, recently these techniques have been used to identify protein interfaces, to detect conformational changes, to identify ligand binding

[†] This study was supported in part by the Research Challenge Program of Wright State University, the Biomedical Sciences Ph.D. Program of Wright State University (fellowship for J.E.N. and D.J.S.), and the Department of Defense, Contract DAMD17-00-C-0020.

* To whom correspondence should be addressed. E-mail: gerald.alter@wright.edu. Phone: 937-775-2200.

[‡] Current address: Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555.

[1] Abbreviations: RPA, human replication protein A; RPA70, the 70 kDa subunit of RPA; RPA32, the 32 kDa subunit of RPA; RPA14, the 14 kDa subunit of RPA; DBDA and DBDB, the second and third OB domain of RPA70; DBDC, the C-terminal domain of RPA70; DBDD, the central domain of RPA32; sulfo-NHS, sulfo-*N*-hydroxysuccinimide acetate; CHCA, α-cyano 4-hydroxycinnaminic acid; SPA, sinnapinic acid; MS, mass spectrometry; MALDI-TOF, matrix-assisted laser desorption−ionization time-of-flight spectrometry.

10.1021/bi052397v CCC: $33.50
Published on Web 07/27/2006

sites, and to inform structure prediction (*9*, *17*−*23*). The increased use of these techniques is linked with advances in mass spectrometric (MS) technology, which makes possible not only simultaneous and accurate identification of residues modified by chemical reagents (or peptides liberated by proteolysis) but also quantification of modification rates (*24*−*26*). Many studies to date report only those sites that are reactive. However, we anticipate that reaction rates, which quantify the kinetics of reactivity, are much more sensitive monitors of reactive group environment than a perfunctory reactive/unreactive classification system. Therefore, we have measured rate constants.

In this study, we have used covalent modification, limited proteolysis, and MS to probe the structures of several domains in native human RPA. We have used reactivity results from these analyses to evaluate the relevance of experimentally derived (from RPA fragments) and theoretically predicted structure models to the native RPA structures.

MATERIALS AND METHODS

Chemicals. Immobilized proteases (trypsin and V8 protease) and sulfo-*N*-hydroxysuccinimide acetate (sulfo-NHS) were purchased from Pierce (Rockford, IL). Soluble sequencing grade proteases (trypsin, chymotrypsin, and V8 protease) were purchased from Roche (Indianapolis, IN). Energy absorbing matrixes, α-cyano 4-hydroxycinnaminic acid (CHCA), and sinnapinic acid (SPA) were purchased from Sigma (St. Louis, MO). All other chemicals were also purchased from Sigma and were of reagent grade or better. Water was deionized using a Millipore (Bedford, MA) Mili Q Plus water purification system.

RPA Purification. Plasmid p11d-tRPA, which contains DNA corresponding to all RPA subunits, was obtained as a generous gift from Dr. Marc Wold (Department of Biochemistry, University of Iowa College of Medicine). It was used to transform the BL21(DE3) strain of *Escherichia coli*. The transformed cells were then induced to overexpress recombinant human RPA, and the protein was purified with a slight modification of a previously described method (*27*). To ensure zinc was maintained in the protein, EDTA was omitted from all buffers and a zinc-containing buffer (termed "Zn buffer"), 20 mM Tris (pH 7.8), 10 mM DTT, 0.01% IGEPAL CA 630, and 10 μM $ZnCl_2$, was used throughout the purification (*28*).

Reactivity of Amino Acids in RPA. (A) Limited Proteolysis. When using proteolytic reactivity to report structural properties of RPA, samples were submitted to limited proteolysis, and rates of individual proteolytic reactions were determined. In these experiments aliquots of purified RPA ([RPA] ≈ 0.5 mg/mL) were treated with trypsin or *Staphylococcus aureus* V8 protease, at pH 7.8 in Zn buffer. Routinely, ∼100 mg of the immobilized protease slurry was added to 250 μL of the buffered RPA solution. When V8 protease was used, conditions were chosen so that cleavage occurred adjacent to glutamic acid residues only (no phosphate ions present). Proteases used in these studies were linked to agarose beads so that reactions were readily stopped at various times (after 2, 5, 10, 20, 40, and 80 min) by centrifuging at low speeds and decanting the supernatant. Hydrolysate from each time point was saved for mass spectrometric (MS) analysis.

(B) Chemical Reactivity. When probing 3-D structural information by means of chemical reactivity, RPA was modified either by the lysine modifying reagent, sulfo-*N*-hydroxysuccinamide acetate (sulfo-NHS), or by free radical hydroxylation (*26*, *29*). In the former case, modifications were routinely carried out in solutions containing 10 μM RPA, 10 μM $ZnCl_2$, and 20 mM HEPES at pH 8.0 and room temperature. The modification reaction was initiated by adding sufficient sulfo-NHS [from a 100 mM stock solution (in doubly deionized H_2O)] to make the RPA mixture 1 mM in sulfo-NHS. At various times (after 2, 5, 10, 20, 40, and 80 min) an aliquot of the reaction mixture was withdrawn and quenched with excess glycine (10 mM).

When modifying with hydroxyl radicals, RPA was dialyzed versus 20 mM HEPES (pH 7.8) containing 10 μM $ZnCl_2$ in preparation for oxidative labeling. Hydroxyl radicals were generated in situ by irradiating samples (0.37 mg/mL in RPA) contained in minicentrifuge tubes with a cesium-137 source. After various periods of irradiation at 0.5 gray/s, the samples were removed from the reaction chamber, and aliquots were withdrawn.

To analyze products, the reacted protein was denatured and exhaustively digested and the modification of the resultant peptides examined. Exhaustive digestion was accomplished by adding acetonitrile to chemically reacted samples to a final concentration of 20% and incubating the samples at 37 °C for 30 min. Samples were then diluted 1:1 with 25 mM ammonium bicarbonate (pH 8.0) to decrease the acetonitrile concentration to 10%. Sequencing grade protease (not on beads), either trypsin, V8 protease, or chymotrypsin, was then added to a concentration of 1/40 (w/w) protease/RPA. Digestions were allowed to proceed overnight at 37 °C.

Mass Spectrometric Characterization of Reaction Products. (A) Collection of MS Spectra. MALDI-TOF mass spectrometry was used to characterize peptides generated during limited proteolysis studies or during exhaustive digestions of the protein adducts from chemical modification studies. MS analyses were performed using a Ciphergen (Fremont, CA) protein chip system, model PBS II. Mass spectra were typically acquired by placing 1 μL of a protein digest (typically 0.25−0.50 mg/mL before digestion) on a gold protein chip spot and mixing 0.5 μL of matrix solution with the spotted sample. Separate laser desorption protocols were optimized to analyze peptides with masses less than 10 kDa and peptides with masses greater than 10 kDa (laser intensity equal to 160 and 273, respectively). At least 100 individual spectra were averaged in each analysis. Masses of each peak were determined with the centroid tool contained in the protein chip analysis package (Ciphergen User's Guide). It should be noted that peptides with masses less than 1 kDa were difficult to observe because of ion suppression caused by matrix peaks.

Matrix solutions used in MALDI experiments were prepared as previously described (*30*). Variation in sample loading and "energy-absorbing-matrix" crystallization on MS chips causes considerable sample-to-sample and peak-to-peak variations in sensitivity. By normalizing MS peak amplitudes to amplitudes of proteolytically resistant peptides and added standards, the effects of sample-to-sample variations were minimized. RPA14 was used as a standard for spectra with mass/charge ranges that were greater than 10000 Thompsons while peptide standards, including angiotensin 1(1296.7 Da), angiotensin 2 (1046.5 Da), and ubiquitin (8565.8 Da), were

added to samples prior to spotting to normalize MS data of less than 10000 Thompsons.

(B) Identification of Peptides. Peptides were routinely identified on the basis of their mass or, when identification was ambiguous, by secondary, exhaustive digestion of fragments. In the former case, samples were evaporated to dryness, digestion buffer was added, and exhaustive proteolytic digests were performed as described above. In the latter instance, secondary digestions were performed on large peptides after separating them either by reversed-phase (C18) HPLC, using a water−acetonitrile gradient containing 0.1% trifluoroacetic acid, or by SDS−PAGE (*30*, *31*). When SDS−PAGE was used, the location of bands were identified using a reversible Coomassie staining reagent, excised, and in-gel digested as described previously. The mass accuracy of the MALDI-TOF (approximately 0.2% of molecular mass) was sufficient so that low molecular mass species (1−4 kDa) could be unambiguously identified on the basis of the measured mass of the peptide.

Reaction Time Courses. The time dependence of the production of peptides and chemically modified peptides that result from proteolytic and chemical reactions was conveniently used to construct rate profiles. Figure 2, left panel, shows representative proteolytic time courses. Each point of the figure represents the average of normalized amplitudes from at least triplicate independent time course determinations. "Averaged" time courses such as these were fit as described in the next paragraph to determine rate constants.

(A) Extraction of Rate Constants from Proteolytic Time Courses. Rate parameters quantifying proteolytic reactivities throughout RPA were deduced in a two-step process: (1) rate constants for peptides produced by digestion were extracted from time courses, and as necessary, (2) these constants were assigned to each end of peptides generated by the proteolysis.

Individual proteolytic reactions were analyzed as first-order/pseudo-first-order processes. This is consistent with concentrations of substrates and proteolytic enzymes being low (<5 and 0.5 μM, respectively) relative to K_m's reported for trypsin-catalyzed (*32*, *33*) and *S. aureus* V8 protease-catalyzed (*34*, *35*) hydrolysis of small substrates. Therefore, peptides including the N- or C-terminal amino acids of RPA subunits were treated as arising from a single first-order reaction. This predicts a hyperbolic time course, which was observed (Figure 2, left panel, filled triangles). Peptides released from the interior of subunit primary structures require a proteolytic reaction at two sites, each first order with respect to substrate. Peptide release was analyzed as two independent (first-order) processes. In this scheme, if rate constants are similar for both processes, the time course is characterized by a lag phase (for example, open triangles, filled circles, Figure 2, left panel). Alternatively, if one constant is much larger than the other, a nearly hyperbolic time course is anticipated and was observed (for example, open circles, Figure 2, left panel). Rate constants were estimated by fitting time courses to eq 1 or 2 for terminal and interior peptides, respectively, using a nonlinear least-squares fitting procedure in the program Slide Write Plus (Advanced Graphics, Encinitas, CA).

Data collected for this publication are very consistent with the kinetic models used here. They display expected rate profile features. Precision of these time courses may not be sufficient to discriminate among alternate kinetic mechanisms predicting similar kinetic profiles. However, we assert that the mechanism modeled here is the simplest one consistent with experimental observations.

$$[\text{peptide}] = A_0(1 - e^{-k_1 t}) \quad (1)$$

$$[\text{peptide}] = A_0(1 - e^{-k_1 t})(1 - e^{-k_2 t}) \quad (2)$$

A_0 was taken as the maximum amplitude for the peptide, while k_1 is the rate constant for one (eq 2) or the only (eq 1) reaction, and k_2 is the constant for the second reaction of the proteolysis. Since we were interested only in primary proteolytic products, data used in the fitting process came from only the time region in which product peptide concentrations were increasing with time. When analyzing two-rate processes (using eq 2), we found the profiles were insensitive to the value of the faster rate constant when it was about 5-fold or greater than the slower one. Yet, the slower rate constant was adequately fit (data not shown). Therefore, when it was too fast to measure, the faster rate constants were assigned a value 5 times the slower rate constant. For this reason, the fastest rates may have been underestimated in this study. When rate constants for the excision of a peptide were very similar or identical, they were not well determined, presumably since there was ambiguity regarding which constant corresponded to a specific rate. Under these conditions, the estimated (reported) rate constants were taken as good approximations of the actual rates.

(B) Assignment of Rate Constants to Peptide Termini. Rates from analyses using eq 2 were correlated with specific ends of peptides generated during proteolysis using several criteria: (A) More reactive sites were initially assigned to regions with the most hydrophilic sequence in the original primary structure. (B) When multiple peptides shared a common cleavage site, rates were assigned to ends so that similarly reactive sites aligned. (C) Assignments were favored when constants for the production of C-termini at each site were matched with similar rate constant values for production of N-termini. (D) Finally, on the very rare occasion when discrepancies became apparent, the exposure of the two alternative locations was assessed using candidate three-dimensional models. Fastest rates were assigned to the most exposed region.

(C) Extraction of Rate Constants from Chemical Modification Reactions. Chemical modification reaction time courses were well fit as a first-order process (eq 3), as expected, since reagent concentrations were in great excess with respect to RPA concentration. In eq 3, *A* and *B* correspond to the concentrations of unmodified and modified peptides at time *t*, while *k* is a pseudo-first-order rate constant for the reaction.

$$A_t/(A_t + B_t) = e^{-kt} \quad (3)$$

The use of ratios of product to reactant concentrations, as in eq 3, has the advantage of accounting for sample to sample variations in MALDI-TOF analysis, thus increasing the accuracy of rate constant determinations. The logarithmic form of eq 3, which we routinely used to analyze time courses, describes a linear plot whose slope is equal to the value of *k*. Figure 2, right panel, shows examples of such

plots for sulfo-NHS and OH modifications. Each point is the average normalized amplitude of at least three determinations. The linearity of these plots validates our first-order analysis and the use of eq 3. Linear regression analysis of these plots typically yielded r^2 values that ranged between 0.7 and 0.95. Modification rate constants are summarized in Table 1 of Supporting Information.

(D) Multiple, Independent Estimates of Rate Constants. In many instances, more than a single peptide had a common N- or C-terminus. Since each peptide resulted in an independent estimation of the rate constant associated with reaction at that site in the native protein, these estimates served to corroborate estimates of the common rate constant. Normally these estimates were averaged. When rate constants were averaged, standard deviations were calculated and are included in Table 1, Supporting Information.

(E) Rate Constant Groups. Regardless of the reagents used, reactive residues were classified according to their reactivity, as reflected by the magnitude of their modification rate constants. Two schemes were used. In the first scheme, each amino acid group contained residues with rate constant values within an approximately 2-fold range. Residues in different groups were color-coded red, orange, yellow, green, cyan, blue, or purple, respectively (from the most reactive group to the least reactive). Unreactive residues were color-coded gray. Molecular models were decorated using this first scheme. In the second scheme, color classes were grouped and given common "reactivity scores". We attempted to populate groups with each reactivity score roughly equally. Groups for each reagent were formed independently of groupings for other reagents. For proteases, reactivity scores of 3 were assigned to residues in red and orange classes, a score of 2 to yellow, green, and cyan classes, and a score of 1 to blue and purple classes. For hydroxylation reactions, a score of 3 was assigned to residues in the red class, 2 to the orange class, and 1 to the yellow class. Finally, for sulfo-NHS modifications, a score of 3 was assigned to residues of the red and orange classes, a score of 2 to the yellow and green classes, and a score of 1 to the cyan class. For all reagents, a score of −3 was assigned to unreactive (gray color class) residues.

(F) Construction of Theoretical Models. Molecular models of several domains were constructed using a remote homology modeling approach (*36*−*38*). Modeled domains included the RPA70 N- and C-terminal domains (as well as the "trimerization core") of RPA (*11*, *14*). In the case of the N-terminal domain, an archetypical OB fold (the DBDA domain of RPA) was used as template. An OB fold domain and a Zn-ribbon domain were combined to model DBDC, the complete RPA70 C-terminal domain. Methods used are described in Supporting Information. Template models and the resultant structure model for DBDC are also shown there (Figure 1, Supporting Information). Finally, the trimerization core was modeled using a successive pairwise domain docking scheme, described in the Results section under the heading Evaluating Different Molecular Docking Topologies. Theoretical models of RPA domain structures are shown in Figures 6, 8, and 9, as well as in Figures 1−4 in the Supporting Information section.

Systematic Scoring of Alternative Models Based on Reactivity Data. To systematically and objectively use enzymatic and chemical reactivity information to select relevant structures, we employed a rudimentary scoring scheme. It is based on the observation that reactivity of residues in a protein is frequently interpreted with respect to their solvent exposure (*30*, *39*, *40*). The intent of our most rudimentary and admittedly simplistic scoring system is to determine for a structure model of a protein a count of the number of times that the minimum requirement, that reactive residues are exposed, is satisfied in a structural model less the number of times it is not. This is accomplished for each monitored site by combining reactivity information (i.e., reactive and unreactive residues are scored as +1 and −1, respectively) with exposure information (exposed and buried residues are scored as +1 and −1, respectively). Multiplying the reactivity score with the exposure score keeps track of whether the reaction rate satisfies expectation based on a particular structure model (as judged by residue exposure). Exposed, reactive residues and buried, unreactive residues meet expectation and have structure scores of 1. Buried, reactive and exposed, unreactive residues do not match expectation and score −1. For the ambiguous case, in which we are not sure if the residue is exposed or buried, i.e., it is on the surface, we do not include the residue in our tally. This is accomplished by giving an exposure score of 0 to residues on the surface, which makes the structure score (exposure X reactivity score) equal 0. The score is summed over all residues analyzed to form a composite score indicating how well reactivity matches expectation based on a structural model, as judged by exposure.

With a small modification, we have extended our scoring system to account for a different level of reactivity. This involves classifying reaction sites we monitor into four groups based on reactivity/rate constants: very reactive, reactive, marginally reactive, and unreactive. These are assigned reactivity scores of 3, 2, 1, and −3, respectively. Accepting for the moment the admittedly simplistic notion that greater reactivity associates with a higher probability of exposure, the calculation of a structure score is directly analogous to the method just described. Exposed residues (exposure score of 1) which are very reactive (reactivity score of 3) have the greatest expectation of being exposed and have the largest score ($1 \times 3 = 3$). Others are proportionally less. Those with moderate activity (reactivity scores of 2) have a structure score of 2 and those with marginal reactivity, a score of 1. For buried residues (exposure scores equal to −1), structure scores are more unfavorable the more reactive the buried group [very reactive residues placed in a model that buries them or unreactive residues that are highly exposed as most unexpected correspond to the most unfavorable scores (−3)] and are most at odds with expectation. We anticipate that residues that are on the surface of a model would not be the most reactive, nor would we anticipate that a partially exposed residue would be completely unreactive. Therefore, we assign a structure score of −1 in both these cases. It is less certain whether moderate reactivity would result from placement on the molecule's surface; therefore, we assign this group a structure score of 0. Finally, we assign a structure score of 1 to marginally reactive groups on the model's surface. Limited exposure is likely to substantially limit reactivity. Again, we consider the effect of solvent exposure here.

In this analysis, accessibility scores were assigned as follows: A solvent-exposed residue in the model analyzed

[greater than 30% side chain accessible (relative to its accessibility as the central amino acid in a tripeptide containing flanking alanines)] was given a value of +1. Each residue whose side chain was buried in the model (less than 20% side chain solvent accessible) was given a value of −1. Side chains that were partially buried or a surface residue (accessibility between 20% and 30%) were given a value of 0 (*41*). For hydroxylation reactions, individual amino acid targets were not always resolved in our tryptic peptides. For these peptides only, the exposure of potentially reactive residues was calculated by (1) summing their individual solvent-accessible area within the model and (2) dividing by the sum of their individual side chain areas in the standard reference peptide. Potentially reactive residues in these peptides were assigned to relative exposure groups on the basis of this "peptide average" relative exposure.

This scheme takes the rate of reactions into account. It penalizes models that bury reactive residues, expose unreactive residues, and have surface/partially buried residues (20−30%) either unreactive or highly reactive. It awards models with reactive residues exposed, unreactive residues buried, and marginally reactive residues partially exposed.

Molecular Dynamic Simulations. We used the AMBER software suite (Regents of the University of California) to perform molecular dynamic (MD) simulations (*42*). We performed these simulations substantially as previously described (*43*). Structures used were energy-minimized using a combination of steepest descent and conjugate gradient steps. The resulting structures were encased in a water box which was 10 Å larger on each edge than the centered protein molecule. The structure and water were heated in successive 50 K steps, allowing the model to relax at each step, until 300 K was attained. We ran short MD simulations to allow potential and kinetic energies for structures and solvent to reach approximately constant values (after about 10000 1.5 fs steps) prior to performing production MD simulations. We aligned individual structures representing 1 ps steps in the subsequent 100 ps production MD simulations using a root mean squared deviation (RMSD) minimization routine. To identify conformational mobility, we then calculated RMSDs for each α carbon over the simulation. Using these results, we colored PDB models according to RMSDs. Those greater than 3 Å were color-coded magenta, 2.5−3.0 Å red, 2.0−2.5 Å orange, 1.5−2.0 Å yellow, 1.0−1.5 Å green, 0.5−1.0 Å cyan, and <0.5 Å blue (Figure 7).

Computational Tools. Images of structural models were produced using the Swiss Protein Data Bank Viewer (http://www.expasy.org/spdbv/) to generate POV scenes. Scenes were then rendered by Pov-Ray 3.1 (Persistence of Vision, Williamstown, Victoria, Australia; http://www.povray.org). The percent solvent accessibility of side chains from the various structures discussed in this paper was calculated using the GETAREA server (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) (*41*). Default parameters were used in these calculations.

RESULTS

Detection of Reactive Sites in RPA (MS Analysis of RPA Modification Reactions). Representative mass spectra of chemically and proteolytically modified RPA are shown in Figure 1. Together, they demonstrate that, even with the

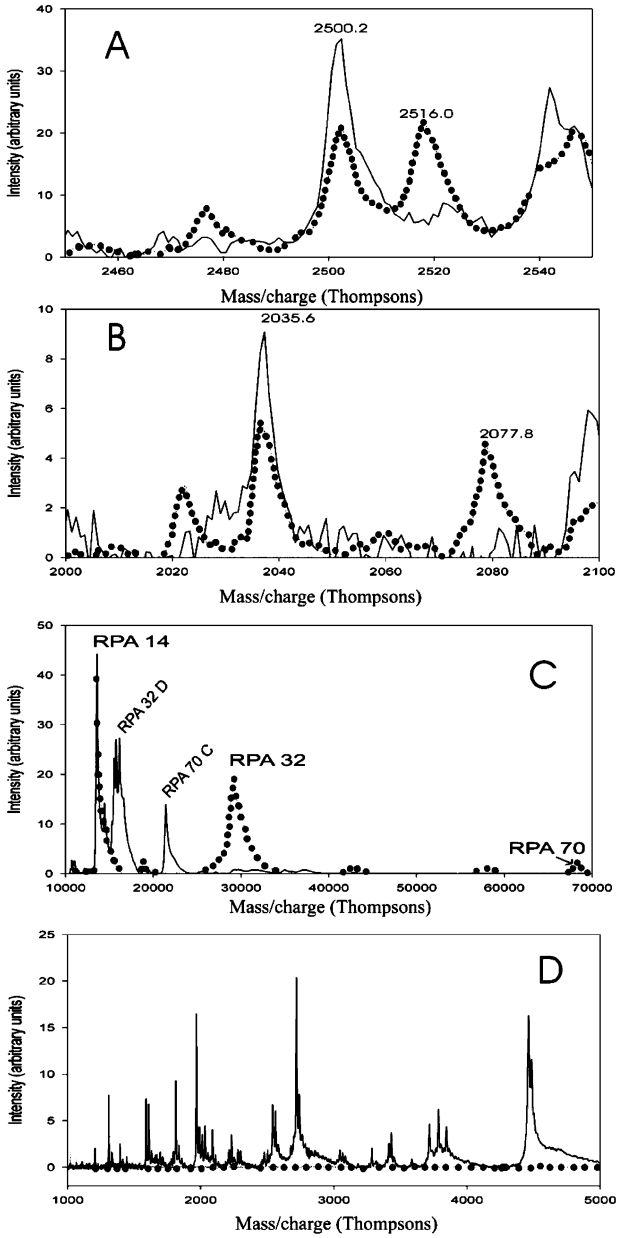

FIGURE 1: Representative mass spectra characterizing the covalent modification and limited proteolysis of RPA. (A) Oxidative modification of tryptic peptide (50−72) from RPA14 by the hydroxyl radical. The black trace shows the MALDI-TOF mass spectrum before modification while the dotted trace shows the same peptide after 30 min of γ radiation. (The data reveal an oxidation product of the peptide that arises approximately 16 Da from the unmodified peak.) (B) Acetylation of V8 protease peptide (253−270) from RPA32 by the reagent sulfo-NHS. The black trace shows the MALDI-TOF mass spectrum before modification while the dotted trace shows the same peptide after acetylation for 40 min. (The data reveal an acetylation product of the peptide that arises approximately 42 Da from the unmodified peak, consistent with the expectant increase in mass upon acetylation.) (C) MALDI-TOF mass spectrum, showing the mass/charge range between 10 and 70 Thompsons, of limited trypsin digestion of RPA before proteolysis (dotted trace) and after 90 min of digestion. RPA14 is resistant to trypsin digestion while the other two subunits are degraded during the time course. Two other domains, RPA70 C (432−616) and RPA32 D (43−171, 41−171, 43−180, and 41−180) are particularly resistant to digestion. (D) MALDI-TOF mass spectrum, showing mass/charge range between 1000 and 5000 Thompsons, before digestion (dotted trace) and after 90 min of trypsin digestion. (The data indicate that though no peptides in this mass range were present in the sample before proteolysis, many small peptides are generated by the end of the time course.)

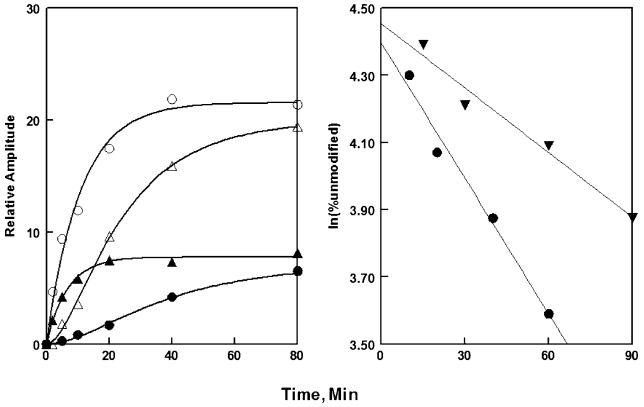

FIGURE 2: Time course for free radical hydroxylation of tryptic peptide 50−72 of RPA14 (triangles) and sulfo-NHS acetylation of 253−270 of RPA32 (circles), right panel, and production of tryptic peptides corresponding to 41−180 (RPA32) (○), 43−180 (RPA32) (△), and 576−616 (RPA70) (▲) as well as a V8 proteolytic peptide corresponding to 561−579 (RPA70), left panel. In the right panel, plots were constructed according to eq 3 in Materials and Methods. Lines correspond to best linear regression lines through experimentally measured time points. In the left panel, points correspond to experimental measurements (normalized using an internal standard as mentioned in Materials and Methods), while associated curves are nonlinear least-squares best fits to eq 2 (Materials and Methods) except for (▲), which was fit using eq 1.

small amount of sample protein routinely used in analyses (1 μL of a 0.5−1 mg/mL solution), sufficiently intense spectra are obtained to detect modification of the protein at multiple sites. Peptides modified by free radical hydroxylation (panel A) and sulfo-NHS acetylation (panel B) during time courses have anticipated molecular mass increases, that is, 16 Da for hydroxyl radical attack at the Met side chain sulfur (panel A) and 42 Da for acetylation of a lysine side chain in panel B (*44*). As mentioned in the Materials and Methods section, molecular masses of peptides and, when necessary, secondary digests were used to identify the peptide modified and thus the amino acid(s) modified. For some hydroxylation reactions, individual amino acid targets in tryptic peptides were not resolved since the amount of hydroxylation product was too small for analyses using secondary hydrolyses. These are noted in results reported here (Table 1, Supporting Information). Spectra such as those shown in panels A and B of Figure 1 were used to construct reaction time courses. Panels C and D in Figure 1 show representative tryptic proteolysis results for native RPA at time zero (dotted lines) and after substantial digestion (solid lines). They show RPA has multiple proteolysis sites. Domains of RPA32 and RPA70 offer some protection against proteolysis. The 14 kDa subunit of RPA is resistant to tryptic attack under the conditions used here (*31*). V8 protease digestion of RPA produced a similar pattern of digestions (data not shown). Normalized MS peaks, like those shown in Figure 1, were used to construct time courses such as those shown in Figure 2 for chemical and proteolytic reactions.

Rate Constant Determinations. As discussed in the Materials and Methods section, time courses for tryptic and V8 protease digestion of RPA are well described by eqs 1 and 2 (Figure 2). Further, hydroxylation and acetylation reactions are well fit as first-order (pseudo-first-order) processes (eq 3) (Figure 2). These observations are consistent with the mechanisms leading to eqs 1−3 and support their validity in our analysis.

Statistical parameters (r) describing the quality of the fit of our data had values uniformly greater that 0.7 and mostly greater than 0.9 (Table 1, Supporting Information). Further, constants derived by this fitting process usually had standard error of estimate values less than 25%. Large standard deviations frequently occurred when averaged rate constants included one whose value was estimated, that is, one that was more than 5-fold faster than the peptide's other rate constant. In these situations the estimated value could be very low and so increase the standard deviation of the average (Supporting Information, Table 1). Rate constants are summarized in Table 1 (Supporting Information). Amino acids were grouped according to rate constant values as described in the Materials and Methods section.

Whether proteolytic rate constants were determined using all peptides or only those greater than 20 AA's long, similar rates were usually found (Table 1, Supporting Information). The largest consequence of excluding small peptides from the analysis was that reactivity information from locations uniquely sampled in reactions leading to small peptides was not used in the analysis of models.

Distribution of Reactive Sites in the RPA Primary Structure. Figure 3 shows the location of reactive sites within RPA's reported structural domains, which are marked by the bars in each panel. Reactive sites are distributed throughout domains. Regions with residues reactive to one protease are frequently reactive to others. This is consistent with the reactivity of both proteases being modulated by similar features (exclusive of specificity), i.e., exposure. Chemical reactions also tend to occur in proteolytically reactive areas (Figure 3). Again, this is consistent with reactivity being dependent on one or more common (though not necessarily sole) features, such as exposure.

Reactivity-Based Analysis of RPA Structure Models. Alternate models of the N-terminal, middle (DBDA and DBDB), and the C-terminal (DBDC) domain three-dimensional structure of RPA70 are shown in Figures 4−6 and in Supporting Information. Using the criterion of reactivity, these models were tested for their accuracy in describing the structures of the respective portions of RPA within the context of the complete protein and under conditions in which modifications were performed. Together, they show that most reactive residues are located on the surface regardless of the specific model. However, close inspection indicates that environmental properties of specific residues vary between alternate models of the same domain's structure. In a number of instances, particular models place residues in impossible locations, based on reactivity data. We interpret these as "fatal flaws" of the respective models. This is one method we used to eliminate models that are not consistent with reactivity results. We present the results of this type of analysis in stick and water-accessible surface models highlighting specific residues.

Fatal Flaw Analysis. (A) DNA Binding Domains A and B from the Central Region of RPA70. We examined alternate structure models of the DBDA and DBDB domains deduced using X-ray crystallographic analysis of this fragment crystallized in the presence [1JMC (PDB designation), lower panels of Figures 4 and 5] (*13*) and absence of an eight-base oligo(dC) (*16*) [1FGU (PDB designation), upper panels of Figures 4 and 5]. Reactivities of amino acids measured in the intact protein were used to decorate each model. As

Table 1: Fatal Flaw Residues Important for Selecting between Alternate Models

				residue environment				
				exposure		secondary		
	domain	reactive class[a]	model (type)	%	class	structure[b]	charge[c]	best model
RPA70								
N-terminal	E67	Yel/Yel	NTM (T)	1.5	i[d]	H		
			1EWI (N)	42	o	C		X
DBDA, DBDB	K379	Cyan/Cyan; Yel[e]	1JMC (X)	3.5	i	S	−1	
			1FGU (X)	38	o	S	+2	X
	E225	Red/Red	1JMC (X)	10.2	i	S		
			1FGU (X)	30.7	o	S		X
	W414, F415	Yel	1JMC (X)	12.3, 3.2	i	H_{end}		
			1FGU (X)	18.1, 24.4		H_{end}		X
DBDC	K473	Grn[e]	HYB (T)	0.5	i	S	+1	
			1L1Oc (X)	60.7	o	β-bulge	+1	X
	E474	Blu/Blu	HYB (T)	24.8		S		
			1L1Oc (X)	51.5	o	β-bulge		X
	K489	Or/Or; Or[e]	HYB (T)	20.8		S_{end}	+3	
			1L1Oc (X)	66.6	o	S	+2	X
	E501	Blu/Blu	HYB (T)	25.6		S_{end}		X?
			1L1Oc (X)	46.8	o	C		
	K551	Cyn[e]	HYB (T)	71.4	o	C	−2	X
			1L1Oc (X)	0.1	i	C	−1	
trimerization core								
DBDC	R568	Grn/Grn	1 (T)	12.6	i	S_{end}		
			3 (T)	37.1	O	S_{end}		X
	R600	Yel/Yel	2 (T)	12.3	i	C		
			3 (T)	47.1	o	C		X
RPA14[f]	F109, F112, F113	NR	1L1O (X)	14.4, 18.4, 3.4	i	C, C, C		X
			1L1Oa (X)	51, 73, 57.8	o	C, C, C		
RPA32 MD[f]	M97, M102	NR	1L1O (X)	2.0, 1.5	i, i	C, S_{end}		X
			1L1Oa (X)	57.1, 1.5	o, i	C, S_{end}		

[a] The reactivity class of amino acids based on rates determined using all peptides is separated by a slash from the reactivity class for the same amino acid using edited peptides. V8 protease is the reagent reactive with E, trypsin with K and K (unless marked by footnote *e*), and free radical hydroxylation with W, F, and M. [b] The reactive amino acid is characterized as in footnote *a*: helix, H; strand of β-sheet, S; coil, C; or a β-bulge. The subscript "end" denotes the location at the terminus of the indicated structure. [c] Charge indicates the sum of charges on side chains within 6 Å of the indicated amino acid. [d] Classification of exposure according to relative solvent accessibility with >30% being classified as o and <20% as i. [e] Reactivity toward sulfo-NHS. [f] Models compared are the subunit exported from the trimerization core (1L1Oa) versus the subunit within the context of the trimerization core.

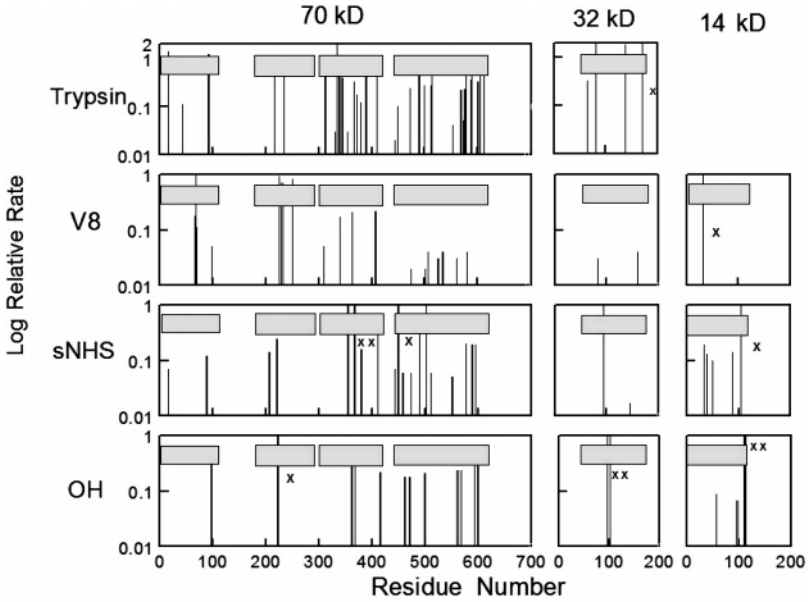

FIGURE 3: Distribution of trypsin, V8 protease, sulfo-NHS, and free radical hydoxylation reaction sites (respectively labeled row of panels) in domains (indicated by horizontal gray bars) of 70, 32, and 14 kDa subunits of RPA (respectively labeled columns of panels). (Note the logarithmic amplitude axis.) Magnitudes of rate constants used in this plot are summarized in Table 1, Supporting Information. Monitored sites that were unreactive are also shown. They are represented by vertical lines from top to bottom of the graph panels with X's to their right. Reaction conditions and analytical methods are those described in the Materials and Methods section.

anticipated, reactive residues are clearly located on the surface of both models, in most cases (Figure 4). However, closer examination indicates that significant differences between models occur at several locations.

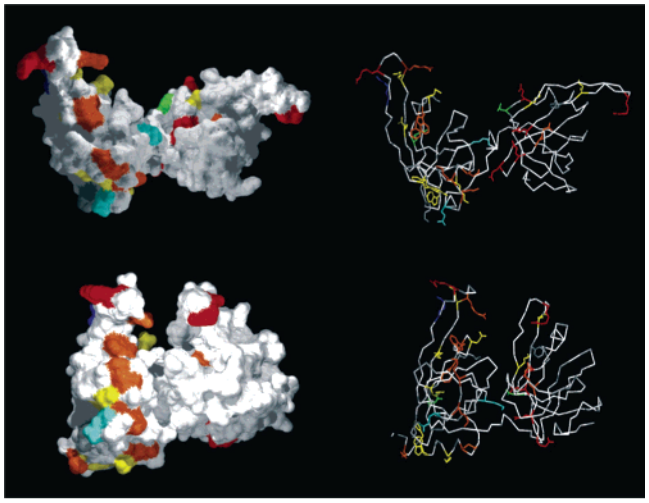

FIGURE 4: Stick (right panels) and Connelly surface (left panels) structure models of RPA's DBDA and DBDB crystallized in the absence (upper panels) (1FGU model) and presence (lower panels) (1JMC model) of a ssDNA dT octanucleotide. The latter model is presented after removing the bound oligo(dT). Residues are colored according to their reactivity as described in the Materials and Methods section.

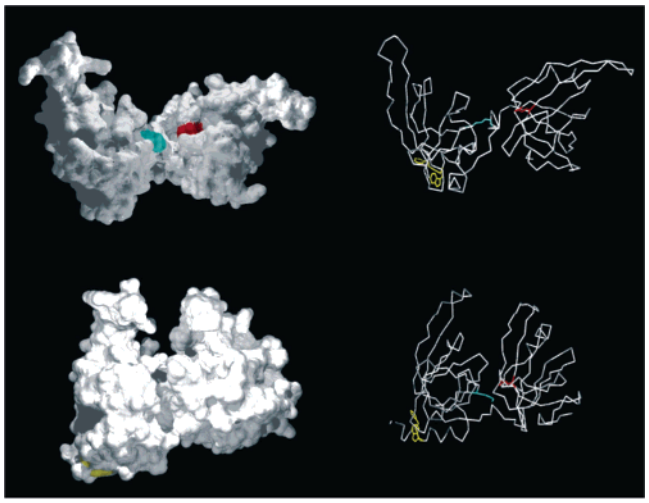

FIGURE 5: Stick (right panels) and Connelly surface (left panels) structural models of DBDA and DBDB as described in Figure 4. Side chains of selected amino acids are shown in the right panels. Corresponding solvent surfaces for these residues are indicated in the left panels. Surface and side chains of Glu 225 are shown in red, of Lys 379 are shown in cyan, and of Trp 414 and Phe 415 are shown in yellow.

There are three areas within the protein where reactive residues appear more exposed in the 1FGU model than in the 1JMC model (Table 1, Figure 5). All of these indicate that the 1FGU model is the more relevant one. The first area is lysine 379 (cyan side chain), which is a site of both trypsin digestion and chemical modification by sulfo-NHS. The moderate reactivity of this residue is more consistent with its 38% solvent accessibility in the 1FGU model than its virtual burial in the 1JMC model. [Notice lysine 379's contribution to 1FGU's solvent-accessible surface in Figure 5, upper left panel, compared with no contribution to the surface in 1JMC (Figure 5, lower left panel).] Further, the sulfo-NHS reactivity of this residue is more consistent with the positive charge in this region of the 1FGU model

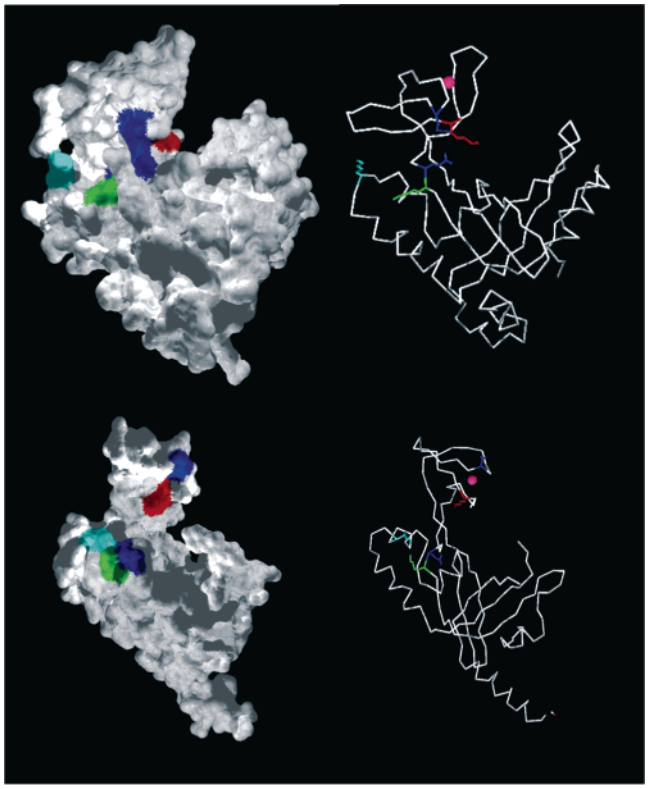

FIGURE 6: Stick (right panels) and Connelly surface (left panels) of the HYB (upper panels) and 1L1Oc (lower panels) models of RPA70's DBDC. In these models, the side chain and surface of Lys 473 are shown in green, the Lys 489 side chain and surface are shown in orange, and the Lys 551 side chain and surface are shown in cyan. The upper blue side chain and associated surface correspond to Glu 501 and the lower ones to Glu 474 in each panel.

("charge" column of Table 1), as opposed to negative charge in the region in the 1JMC model. [Sulfo-NHS, being negatively charged, should be attracted to the positive region, enhancing its reactivity. Similarly, we anticipate it will tend to be excluded from the negative regions as in 1JMC (Table 1), decreasing its reactivity at that site (*39*).] Glutamic acid 225, which is a site of rapid V8 protease digestion, is the second area (red side chain, Figure 5). The amino acid's reactivity is more consistent with its greater than 30% exposure in the 1FGU model than its virtual burial (10% exposure) in the 1JMC model. [The residue is in similar secondary structures in both models (Table 1).] The third area is the 411−431 peptide of RPA70, which is a very reactive hydroxylation target (Table 1, Supporting Information). This peptide contains two aromatic amino acids which are potential targets for hydroxylation (yellow side chains, Figure 5). Yet Phe 415 and Trp 414 are only about 3% and 10%, respectively, solvent accessible (an accessibility frequently classified as buried) (Table 1) in the 1JMC model. In contrast, both residues are about 20% exposed in the 1FGU model, which is more consistent with the reactivity we observe (Table 1). [Both residues are on the "back" of the model in Figure 5 (lower left panel) and are only partially visible at the bottom in this view of the model.] On the basis of these results we conclude that the 1FGU structure is more representative of this portion of RPA in the native apoprotein under the conditions that we modified it. This is consistent with our expectation since DNA was not present in the RPA we analyzed (see Discussion).

(B) C-Terminal Domain of RPA70. Similar analysis of the C-terminal domain of RPA70 (DBDC) indicates that the model deduced by X-ray crystallography, 1L1Oc (PDB designation), represents the native protein's structure better than our theoretical model of the same domain (HYB, Supporting Information).

The experimentally derived and theoretical models have striking similarities. Both are composed of an OB and a Zn-ribbon fold (Supporting Information). Excluding the Zn-ribbon region and the OB fold's C-terminal helix, the α carbons of the two models can be superimposed with a RMSD of 2.1 Å. Alternatively, the Zn-binding region of the two models can be superimposed with an RMSD of 2.4 Å when the remainder of the model is ignored (Magic Fit function of SPDBV; http://us.expasy.org/). Primary differences between models include the juxtaposition of the OB and Zn-ribbon folds, which are connected via an apparently flexible neck region (Figure 6).

As with DBDA and DBDB, reactive residues generally contribute significantly to the protein's solvent-accessible surface (Figure 1, Supporting Information). However, differences between models are apparent. For analysis of alternative structural models for DBDC, the following three areas are particularly relevant (Figure 6): (1) a β-bulge in the 1L1Oc structure model, (2) an interface region between the Zn-ribbon and the OB fold, and (3) a portion of the protein structure that molecular dynamic simulations indicate is very flexible. In the first region, neighboring residues, K473 (green side chain, Figure 6) and E474 (blue side chain, Figure 6), have moderate and low reactivates, respectively, which are most consistent with their location in the 1L1Oc model rather than HYB (Table 1). The primary difference between models in this region is that the 1L1Oc model, but not HYB, contains a β-bulge including both K473 and E474 (Figure 6, green and lower blue side chain). The bulge interrupts a β-strand and likely relieves overwhelming inhibition of V8 protease activity at E474. This accounts for the modest V8 reactivity observed at that site. [Researchers have argued that residues in β-sheet structures have severely limited reactivity or are unreactive toward proteases owing to the rigid and extended conformation of peptide backbones in this secondary structure (*46*−*48*).]

The β-bulge irregularity causes the K473 ϵ-amino group to be much more exposed in the 1L1Oc model (Figure 6, lower left panel) than it would be in a regular sheet structure as in HYB (concave green, Figure 6, upper left panel) (Table 1). This accounts for K473's moderate reactivity with sulfo-NHS. Reactivity at both E474 and K473, then, argues for the 1L1Oc model.

A second area contains (1) K489 of DBDC (orange side chain, Figure 6), which is quite reactive toward enzymatic and chemical (NHS) reagents (Table 1), and (2) E501 (upper blue side chain, Figure 6), which is modestly reactive toward V8 protease. These residues are in a portion of the 1L1Oc and HYB models whose exposure varies with the juxtaposition of the domain's OB and Zn-ribbon folds. As a result, solvent accessibility of these residues is greater in the 1L1Oc model than in the HYB model. The substantial reactivity of trypsin at K489, then, is most consistent with the 1L1Oc model (Table 1). Further, the combination of exposure and positive charge in the region is consistent with the strong reactivity of this lysine toward sulfo-NHS. However, the reactivity of E501 is surprisingly low given the residue's level of exposure in the 1L1Oc model (∼47%) [relative to other V8-reactive locations such as E225 (∼31% solvent accessible)] (Table 1). It is perhaps more consistent with the limited exposure in the HYB model (∼25% solvent accessible). Further, the residue's location in secondary structure varies between models. It is at the end of a β-strand in HYB rather than within a coil (as in 1L1Oc), consistent with limited rather than substantial proteolytic reactivity at this position. Then, as a whole, reactivity at this location is more consistent with the 1L1Oc model but has features quite compatible with the HYB model.

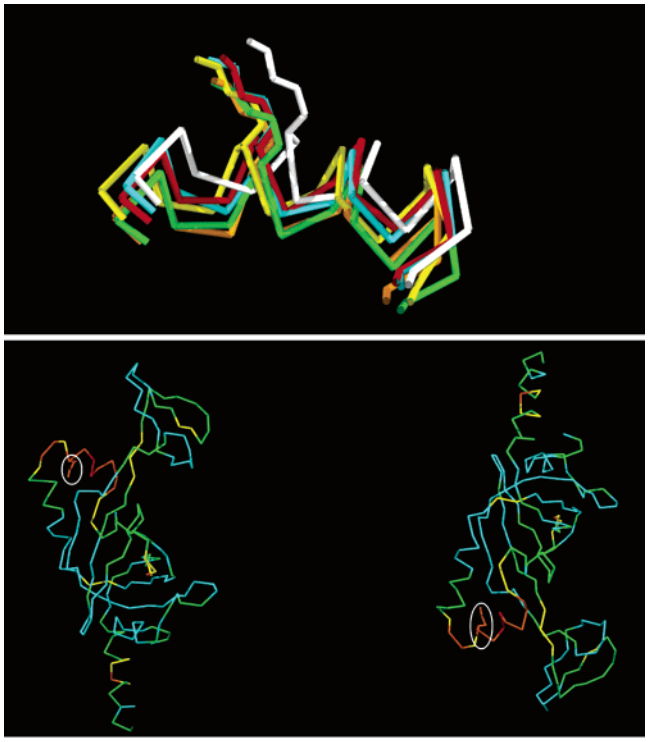

FIGURE 7: Models of the RPA70 C-terminal domain structure in the region of Lys 551 during molecular dynamic simulations (upper panel) and flexibility of this domain as judged by RMSD during molecular dynamic simulations (lower panel). The position of the K551 side chain is shown before (white) and after 20, 40, 60, 80, and 100 ps simulation (cyan, green, yellow, orange, and red, respectively) (upper panel). In the lower panel the structure (1L1Oc) is colored according to RMSD movement during simulations with red being the greatest movement and orange, yellow, green, and cyan being progressively less.

We conclude that the 1L1Oc model is also preferred at the third site. However, this conclusion is based on the flexibility of this region predicted by molecular dynamic simulation results described in the next paragraph. At K551, which is modestly reactive toward sulfo-NHS, alternative models (1L1Oc and HYB) differ (Table 1). Low reactivity is consistent with net negative charge in the vicinity of K551, a feature of both models. However, even low reactivity is inconsistent with a static 1L1Oc model since it buries the K551 ϵ-amino group. However, the activity also seems surprisingly low for the large K551 exposure predicted by the HYB model (Table 1).

Interestingly, molecular dynamic simulations conducted on 1L1Oc show that lysine 551 moves early in MD simulations to a more exposed position (Figure 7). In fact, after 50 ps of MD the ϵ-amino group of this residue is nearly

Table 2: HEX Docking Parameters

receptor	ligand	energy			rank[a]
		total	shape	force	
RPA32D[b]	RPA14[c]	−754.0	−899.2	145.2	1
RPA14−RPA32D[d]	RPA70C[e]	−523.0	−448.6	−74.4	1
RPA14−RPA32D	RPA70C	−489.6	−412.4	−77.2	2
RPA14−RPA32D	RPA70C	−473.1	−397.6	−75.5	3

[a] Docking simulations are ranked on the basis of the total calculated energy. The docking results labeled "1" are the lowest energy results for a particular simulation. [b] The central domain from RPA32 (43−171) was used as the receptor for this simulation. [c] RPA14 was used as the ligand for this docking simulation. [d] The complex of RPA32D and RPA14 was used as the receptor for these simulations. [e] The C-terminal domain from RPA70 (449−616) was used as the ligand for these simulations.

30% accessible (colored side chains, Figure 7). This is consistent with the observed reactivity at that site arising from a dynamically relaxed 1L1Oc crystal structure. In fact, results of MD simulation involving the whole RPA70 N-terminal fragment (Figure 7, bottom panels) demonstrate that K551 is located in a region that this method predicts is the most mobile segment of the RPA70 N-terminal domain. A similar dynamic change was not found for the NHS-reactive residues in the HYB model. Therefore, we conclude that the flexible 1L1Oc model is most consistent with reactivity results in this region.

(C) N-Terminal Domain of RPA70. Reactive residues of the RPA70 N-terminal domain are generally exposed in alternative models derived either from NMR analysis (1EWI model) or by homology modeling methods (NTM model; Supporting Information, Figure 2). However, inspection of reactivity results as just described indicates that the NMR-derived structure for the N-terminus of RPA70 (which was obtained from a fragment of RPA) more accurately reflects this portion of the native protein than our predicted model (Supporting Information). A key residue in this analysis is E67 which is V8 protease reactive. This is inconsistent with its virtually buried position and location in an α-helix in our predicted structure (NTM) (Table 1). However, the reactivity is quite consistent with its exposed position and location in a coil region in the 1EWI model (Table 1, Supporting Information, Figure 3).

Evaluating Different Molecular Docking Topologies. A structural model for the "minimum trimerization core" of RPA (*11*) was probed using the methods describe above. Models evaluated included one deduced from X-ray crystallographic analysis of a RPA fragment containing the trimerization core (1L1O) and three models deduced by docking component pieces of the trimerization core using the molecular docking software, HEX (*49*−*51*).

For docking experiments, individual domains from 1L1O were docked in a pairwise fashion. The juxtaposition of a subunit, RPA14 (structure model 1L1Oa), and the central domain of another subunit, RPA32 (structure model 1L1Ob), was determined first since this interaction was characterized by the largest decrease in free energy of any pairwise docking combination we performed (Table 2). This energetically most favorable docking result yielded a model nearly identical to the model of these domains in the context of the trimerization core model deduced from X-ray crystallographic analysis (1L1O). In this two-domain structure derived from docking

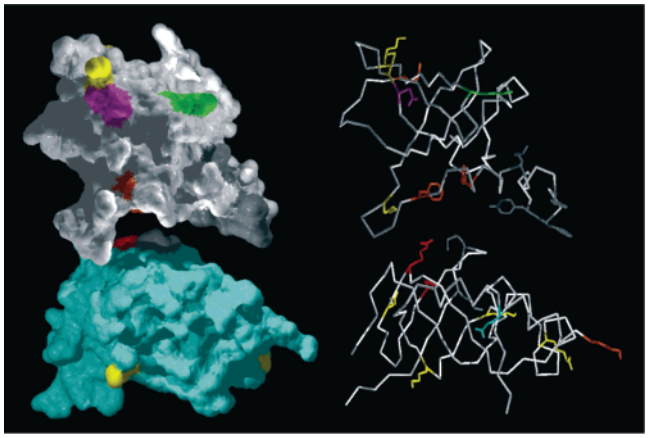

FIGURE 8: Stick (right) and Connelly surface (left) models of the best scoring HEX docking results involving RPA14 (upper domain) and RPA32 (lower domain). The domains of the complex have been separated along the vertical axis to aid viewing of subunit interfaces. Side chains of reactive residues are shown in multiple colors according to their reactivities. Gray side chains (Phe 106, Phe 112, and Phe 113 of RPA14 and Met of RPA32) are unreactive and are both exposed on individual domain surfaces and occluded from solvent in the two-domain complex.

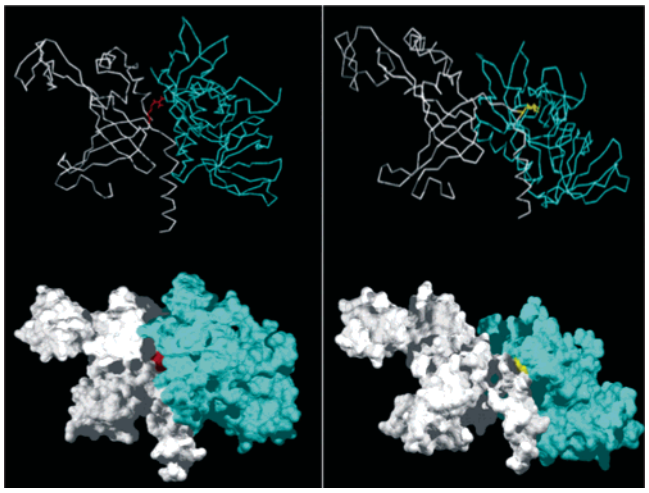

FIGURE 9: Stick (upper panel) and Connelly surface (lower panel) models of the lowest energy docking (left panels) and the second lowest energy docking (right panels) solutions for combining 1L1Oc (white) with the RPA14 (1L1Oa)−RPA32 fragment (1L1Ob) complex (cyan). HEX was used in all docking procedures. R568 (red) is substantially buried in the lowest energy solution (left panels) and R600 is substantially buried in the second lowest energy solution (right panels).

experiments, no reactive residues were buried (Figure 8). Further, several residues on the surface of these domains that are NOT reactive in the native protein are buried by domain contacts (F106, F112, and F113 of RPA14 and C219 of RPA32) (Table 1).

The RPA14−RPA32 fragment just described was docked with the C-terminal domain from RPA70 (1L1Oc). Free energies for the three most energetically favorable docking results are summarized in Table 2. The lowest and next to lowest energy models (models 1 and 2, Table 1), however, buried different residues (12% solvent accessible) that were reactive in native RPA (Arg 568 and Arg 600, Table 1, Figure 9). In contrast, all of the reactive residues in the third lowest energy docking result are at least 30% accessible after docking (model 3, Table 1). We conclude the latter is the

Table 3: Reactivity Scores of RPA Domain Models[a]

		reagent[b]						total score	
method	model	trypsin	edited[c]	V8	edited	hydroxyl radical	sulfo-NHS		edited
2-class[d]	EWI	4		4		1	4	13	
	NTM	4		2		1	4	11	
	FGU	11		7		1	3	22	
	JMC	9		5		0	1	15	
	1L1Oc	12		3		2	9	26	
	HYB	12		1		3	7	23	
4-class[e]	EWI	11	11	9	9	3	4	27	27
	NTM	8	8	5	5	3	4	20	20
	FGU	25	25	15	15	4	−3	41	41
	JMC	13	13	9	9	2	1	25	25
	1L1Oc	26	25	12	7	3	19	60	54
	HYB	28	25	12	7	3	13	56	48

[a] Reactivity scoring is based on the accessibility of amino acid side chains with the exception of sulfo-NHS reactivity, which is scored using the accessibility of the ϵ-amine (see Results section for detailed description of scoring). [b] Score resulting from action of indicated reagent. [c] Rate from peptides less than 20 amino acids long not included in analysis. [d] Residues classified only as reactive or unreactive (see Results section for detailed description of scoring). [e] Residues classified on the basis of rate constant values as in one of three activity levels or as unreactive.

Table 4: Reactivity Scores of Trimerization Core Docking Solutions[a]

		reagent[b]						total	
method	model	trypsin	edited[c]	V8	edited	hydroxyl radical	sulfo-NHS		edited
2-class[d]	1[f]	14		7		5	8	34	
	2[g]	14		11		5	8	38	
	3[h]	16		10		4	9	39	
4-class[e]	1	27	24	10	9	3	4	44	40
	2	27	24	18	13	5	4	54	46
	3	31	28	16	11	5	4	56	48

[a] Reactivity scoring is based on the accessibility of amino acid side chains with the exception of sulfo-NHS reactivity, which is scored using the accessibility of the ϵ-amine (see Results section for detailed description of scoring). [b] Score resulting from action of indicated reagent. [c] Rate from peptides less than 20 amino acids long not included in analysis. [d] Residues classified only as reactive or unreactive (see Results section for detailed description of scoring). [e] Residues classified on the basis of rate constant values as in one of three activity levels or as unreactive. [f] Scoring of the highest ranking docking result (1 in Table 2). [g] Scoring of the second highest ranking docking result (2 in Table 2). [h] Scoring of the third highest ranking docking result (3 in Table 2).

best model of this region in the intact protein. Interestingly, the latter result is virtually identical to the X-ray crystallographically determined 1L1O trimerization core model (α carbon RMSD = 0.62 Å).

Systematic Scoring. We used a second general approach to evaluate candidate structural models and docking results. Using the systematic scoring scheme described in the Materials and Methods section, which incorporate observed reaction rates and residue exposure in a candidate structure, scores for each structure were assigned. Different variations of the scheme were employed (see Materials and Methods). These included or ignored reactions resulting in small peptides and considered or disregarded magnitudes of rate constants. Results for different models and methods are listed in Table 3 (domain structures) and Table 4 (the trimerization core).

DISCUSSION

A number of structural models characterizing RPA have been reported (*11*−*16*). These anchor our understanding of the mechanisms of action of this important protein. However, these studies have been conducted on fragments of the native protein in nonphysiological environments. For RPA, the reasonable, but unverified, assumption is that these models represent the structure of respective areas in the native protein. Here we use proteolytic and chemical reactivity analysis (p-RAN and c-RAN) for evaluating the relevance of three-dimensional models to the protein structures. Specifically, we used this methodology to test the validity of fragment-derived structural models for describing the three-dimensional structure of the native protein.

Reactivity-Based Analysis of RPA Domain Models. The rationale for our approach is that different models may place given amino acids in different environments and so suggest different reactivities. By comparing observed reactivities with anticipated ones, a model that is most consistent with the structure of the protein (under the conditions in which it is modified) can be selected from several possibilities. For the best measure of reactivity, we have used reaction rate constants at individual amino acids rather than simply cataloguing which residues which do or do not react. Since, in this study, we measured reactivities of native RPA under solution-state conditions similar to those used when monitoring RPA activity, we assume the reactivities reflect features of RPA relevant to its biological activity.

(A) Models. Candidate structural models used in this study include those reported from X-ray crystallographic (*11*−*13*, *16*) and NMR (*14*, *15*) studies of RPA fragments (in either the presence or absence of ssDNA). Other models were derived in modeling studies (Materials and Methods and Supporting Information) that relied heavily on homology modeling (*36*). They were constructed prior to the publication

of reports describing the structural determination of the same domains by NMR or X-ray methods. These structural predictions represent reasonable alternative theoretical models of pertinent domains.

(B) Reaction Rates. Reactivities toward free radical hydroxylation and sulfo-NHS (acetylation) as well as trypsin and V8 proteolysis were determined. In aggregate, approximately 110 site-specific reactions in domains throughout the protein structure were measured. Rate constants, determined from reaction time courses, were used to place residues with similar rate constants in groups that were color-coded red to blue as described in Materials and Methods. Residues in models of candidate structures were decorated on the basis of this color scheme (see below). For systematic scoring (see below), rate constants describing reactive residues were used to reclassify residues into four larger reactivity-based groups. (See Materials and Methods section for details.)

Analysis of Candidate RPA Models. We have used two general methods to analyze models based on chemical and proteolytic reactivity. In one, we examined alternate models to determine whether they clearly included one or more fatal flaws, that is, whether a reactive residue was clearly in an environment that prohibits reaction or, conversely, whether one or more unreactive residues were in environments conducive to reaction. In a second method, we have used systematic, arithmetic schemes based on solvent exposure of all probed residues to assess how well reactivity measurements were represented in specific models.

(A) Fatal Flaw Domain Structures. The results of the fatal flaw approach are summarized in Table 1, "best model" column. Key fatal flaw residues for different models are listed with pertinent characteristics of their environments in this table. These results demonstrate that frequently, when using the criterion of enzymatic/chemical reactivity, a model was preferred at a particular site for several reasons. Relevant factors included solvent exposure (necessary for reaction with polar reagents), participation of the amino acid in secondary structure (which generally inhibits proteolytic cleavage), and local charge (*39*, *48*). For example, in the vicinity of K379, 1FGU was more consistent with reactivity results than 1JMC, whether exposure or local charge was considered (Table 1, Figure 5). Similarly, in the vicinity of E67 in the RPA70 N-terminal domain, 1EWI was a better model than NTM whether solvent exposure or placement in secondary structure was considered. (See Materials and Methods section for detailed arguments.)

As Table 1 indicates, a clear choice of models is apparent for the RPA N-terminal domain (1EWI), the DBDA and DBDB domains (1FGU), and the trimerization core (1L1O). Analysis of alternative models (1L1Oc and HYB) for the structure of the C-terminal region of RPA70 is complex. We conclude that the 1L1Oc model is more consistent with reactivity data than the HYB model, with the caveat that its structure is slightly different in the native RPA structure (similar to the MD-relaxed structure) (Figure 7) relative to the crystal structure (to explain reactivities of K551, K489, and E501) (Figure 6, Table 1).

Interestingly, the analysis of models for the C-terminal end of the 70 kDa subunit demonstrates the ability of chemical reactivity methodology to suggest the presence of a β-bulge (*52*). This is a feature which is very difficult to predict during model construction. Therefore, reactivity data are quite useful for improving structure models.

(B) Fatal Flaw Docking Structure. It is noteworthy that, in the case of the trimerization core, potentially reactive residues, based on their location on the surface of individual domain structures, were buried by domain−domain contacts and so were unreactive (Table 1, Figure 8). Though this did occur with a low-energy model, it was not the minimum energy result indicated by HEX (Table 3). This clearly demonstrates the great potential for using unreactive residues located on the surfaces of substructure (domain structures here) models together with reactive residues to indicate the topology of structures. In combination, reactive and unreactive residues facilitate identifying interaction sites that lead to higher levels of structural organization.

(C) Systematic Scoring of Structures. Using a simple exposure-based approach, we have classified residues as exposed (>30% solvent accessible), surface (30−20% solvent accessible), or buried (<20% solvent accessible) according to their solvent accessibility (*41*). As the above discussion indicates, we do not wish to imply that exposure is the only relevant factor for reactivity. However, we have assumed that reactive residues must be exposed, buried residues cannot be reactive, and surface residues may or may not be reactive.

Experimentally observed reaction rates were grouped differently for different analyses. In one approach, we considered a residue as either reactive (no matter how slowly) or not reactive. [This method is equivalent to evaluating reactivity on the basis of a single (frequently assumed) end point, rather than taking into account rate constants.] In the second method we classified residues as unreactive or in three levels of reactivity. The latter allowed about 4−8-fold variation in rate constants within each level of reactivity. Reactivity measures of the native protein and accessibility features of models were combined and summed over each model to give an overall structure score for that model, as described in the Materials and Methods section. Results are summarized in Tables 3 and 4. For alternate models of the same structure, the ones with higher scores are most consistent with reactivity results. We examined the same two alternative models for each domain using the "systematic scoring" approach that we analyzed using the fatal flaw approach. Our results indicate that regardless of whether we classified reactivity using a 2-class system (reactive or not reactive) or a 4-class system (very reactive, reactive, marginally reactive, or unreactive) the same models for each pair scored highest, EWI, FGU, and LIOc (Table 3). Similarly, analysis of alternate docking models identified the same model as best regardless of the scoring method used (Table 4). Lastly, the same model was best whether the systematic scoring or the fatal flaw method was used (Tables 1 and 4).

Scores resulting from reactions at different regions of RPA with each reagent used are shown in Tables 3 and 4. They show reagents contributed variably to the summed reactivity results and occasionally favored a model contrary to the consensus score model (for example, Table 3, sulfo-NHS column, FGU and JMC models; Table 4, V8 results, models 2 and 3). We assert this is the result of the rather simplistic analysis of reactivity used in the systematic scoring method. It considers only solvent accessibility. In the case of at least some sulfo-NHS reactions with lysine side chains, exposure

is a necessary but not a sufficient condition. The negative charge on the reagent might interfere with its approach to an otherwise reactive lysine. This likely occurs in the region of the trypsin-reactive residues, K311 and K313 of RPA70. These residues are in a region of net negative electrostatic potential and are unreactive with respect to sulfo-NHS. Similarly, an unreactive ionization state of lysine side chains might be stabilized by the residue's environment. For example, trypsin-reactive K499 is solvent exposed but unreactive with respect to sulfo-NHS. It is significant that the primary amine of this lysine side chain is within 3 Å of the carboxlytic acid side chain of D527. These side chains likely form a salt bridge which stabilizes the protonated, *unreactive*, ionization state of the K449 side chain. Hubbard and co-workers have enumerated a number of properties important for reactivity of proteases (*48*). It is clear that many properties may modulate activity. In fact, we are currently developing more realistic, multifaceted analyses of reactivity. For the present, however, the method's ability to select the same consensus models regardless of the systematic scoring system used, as well as the method's ability to recapitulate choices made using the more usual fatal flaw approach, attests to its surprising accuracy and robustness.

Since release of an internal peptide of the protein requires cleavage at two sites, it occurred to us that it is possible that cleavage at the first site might affect reactivity at the second. However, correlations between regions where residues are susceptible to attack by protease and chemical modification reagents (Figure 3) are gratifying and suggest that this is not the case. In a similar vein, models showing all reactive residues (Figure 4 and Figures 2 and 4 of Supporting Information) indicate that many of the same hot spots on the model surface are identified by multiple reagents. This suggests the data from release of internal peptides is not suspect. Finally, we considered that cut sites close to each other provide one, but not the only, scenario by which a second proteolytic cut could be influenced by the first. One way this could occur is if the cut sites are close in the primary structure. Therefore, we determined rate constants at each site following two protocols: one including data from all peptides released and a second excluding consideration of rate constants from peptides smaller than 20 amino acids (Tables 3 and 4). The results were the same whether all peptides cleaved from the protein or only larger (>20 amino acids) peptides (Table 3, edited columns) were used to quantify reactivity. All indications are that, for RPA, the influence of a first reaction on the second reaction to release a peptide is not large.

RPA Structure. (A) Chemical and Enzymatic Modification of RPA. Others have reported chemical modification of RPA. One group recently used the acetylating reagent biotin-NHS, and another involved cross-linking RPA with DNA (*53*). Both reagents and conditions used by these investigators differ from those used here. In one instance, an uncharged NHS reagent was used instead of the negatively charged sulfo-NHS used here. In another, DNA was used as a coreagent, while it was not used here. Further, in both reported instances, Zn-deficient, EDTA-containing buffers were used while EDTA-deficient Zn-containing buffers were used here. We chose to use Zn-containing buffer since RPA contains a Zn-binding site, and other investigators have reported that Zn influences RPA function (*28*, *54*−*57*). Despite these differences, previous modification results are generally consistent with structural selections we have made and tend to corroborate our conclusions, even as ours tend to corroborate theirs.

Similarly, other researchers have reported the use of proteolysis to probe RPA structure and RPA structural changes (*5*, *31*, *58*). These investigators have monitored the protein's structure, concentrating on proteolytic cleavage that generates large fragments of RPA. Using mass spectroscopic analysis, we have been able to examine a wider range of fragments and found substantially more reactive sites (Figure 3 and Table 1 in Supporting Information). Results reported here and in Nuss and Alter (*31*) substantially identify cleavage sites proposed by previous investigators. Therefore, our results are in substantial agreement with those reports. We have examined the reactivity of portions of the RPA that are not in domain structures. These data will be presented in a future publication which deals more completely with these latter reactive sites.

(B) Structure of Domains within RPA. Our results presented here do address the relevance of RPA fragment structures to the native protein's structure. In general, they confirm the applicability of fragment structures to the corresponding structure in the native protein. On the basis of native RPA reactivities, we conclude that the 1EWI and the 1FGU models faithfully represent these portions of the native protein structure. Similarly, our results indicate that trimming approximately 70% of the amino acids in RPA from the protein, to form the protein's trimerization core, does not change the juxtaposition of the remaining domains. Our studies also indicate that the C-terminal domain of RPA, which contains the Zn-binding domain, is substantially like the C-terminal domain of the 1L1O model of RPA, particularly if the model is allowed to be flexible (Figure 7).

These observations are consistent with the characterization of RPA as a "string" of autonomously folding domain structures within the protein. It is consistent with a recent report describing the denaturation of RPA (*31*). In that study, substantial changes in the RPA structure, as monitored by fluorescence, were reported, though small changes in ssDNA binding properties and only modest changes in the CD properties of the polypeptide backbone were noted. The latter measurement probably reports the state of the protein's domains. Our results confirm the assertion that these domains are quite stable, being very similar in crystallized fragments and in the native protein. This, in turn, supports the contention that the RPA domains of this structure may rearrange substantially without unraveling.

The two models for the DBDA and DBDB, 1FGU (Apo) and 1JMC (crystallized with bound ssDNA), clearly predict different reactivities for RPA residues. These differences do not simply reflect the portion of the structure that is occluded by the binding of a molecule of ssDNA. (Bound oligonucleotide was stripped from the 1JMC model prior to using it for reactivity analyses.) Rather, they reflect a reorientation of the two domains in these models (*13*, *16*). Our results clearly show that the apo model (1FGU) more accurately represents the orientation in the native protein structure than the holo (1JMC) model. If a "plus DNA" conformation predominated in native RPA even in the absence of the oligonucleotide, our methodology would have detected it. In sum, our results add weight to the view that DBDA and

DBDB conformation/orientation accompanies binding of ssDNA to RPA (*16*). An intriguing question remains: Is this only a portion of a larger ligand-associated perturbation of the RPA conformation (*5*, *59*)?

ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. Vernon Anderson for providing us access to the Cs137 irradiator used in this study, Dr. David Cool for providing us access to the Cyphergen MS instrument used here, and the BMS Ph.D. as well as the Research Incentive Program of Wright State University for grant support of the studies reported here.

SUPPORTING INFORMATION AVAILABLE

Descriptions of theoretical modeling procedures, rate constant values for all chemical and proteolytic reactions, and graphics showing both modifications and fatal flaw residues. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Iftode, C., Daniely, Y., and Borowiec, J. A. (1999) Replication protein A (RPA): the eukaryotic SSB, *Crit. Rev. Biochem. Mol. Biol. 34*, 141−180.
2. Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism, *Annu. Rev. Biochem. 66*, 61−92.
3. Kim, C., Paulus, B. F., and Wold, M. S. (1994) Interactions of human replication protein A with oligonucleotides, *Biochemistry 33*, 14197−14206.
4. Kim, C., and Wold, M. S. (1995) Recombinant human replication protein A binds to polynucleotides with low cooperativity, *Biochemistry 34*, 2058−2064.
5. Gomes, X. V., Henricksen, L. A., and Wold, M. S. (1996) Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA, *Biochemistry 35*, 5586−5595.
6. Treuner, K., Ramsperger, U., and Knippers, R. (1996) Replication protein A induces the unwinding of long double-stranded DNA regions, *J. Mol. Biol. 259*, 104−112.
7. Thoma, B. S., and Vasquez, K. M. (2003) Critical DNA damage recognition functions of XPC-hHR23B and XPA-RPA in nucleotide excision repair, *Mol. Carcinog. 38*, 1−13.
8. Robison, J. G., Elliott, J., Dixon, K., and Oakley, G. G. (2004) Replication protein A and the Mre11.Rad50.Nbs1 complex co-localize and interact at sites of stalled replication forks, *J. Biol. Chem. 279*, 34802−34810.
9. Kiselar, J. G., Janmey, P. A., Almo, S. C., and Chance, M. R. (2003) Visualizing the Ca^{2+}-dependent activation of gelsolin by using synchrotron footprinting, *Proc. Natl. Acad. Scie. U.S.A. 100*, 3942−3947.
10. Murzin, A. G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for nonhomologous sequences, *EMBO J. 12*, 861−867.
11. Bochkareva, E., Korolev, S., Lees-Miller, S. P., and Bochkarev, A. (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA, *EMBO J. 21*, 1855−1863.
12. Bochkarev, A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding, *EMBO J. 18*, 4498−4504.
13. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA, *Nature 385*, 176−181.
14. Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X., and Wold, M. S. (1999) Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker, *J. Biomol. NMR 14*, 321−331.
15. Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J., Edwards, A. M., and Chazin, W. J. (2000) Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA, *Cell 103*, 449−456.
16. Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding, *EMBO J. 20*, 612−618.
17. Peterson, J. J., Young, M. M., and Takemoto, L. J. (2004) Probing alpha-crystallin structure using chemical cross-linkers and mass spectrometry, *Mol. Vision 10*, 857−866.
18. Novak, P., Kruppa, G. H., Young, M. M., and Schoeniger, J. (2004) A top-down method for the determination of residue-specific solvent accessibility in proteins, *J. Mass Spectrom. 39*, 322−328.
19. Kruppa, G. H., Schoeniger, J., and Young, M. M. (2003) A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry, *Rapid Commun. Mass Spectrom. 17*, 155−162.
20. Guan, J. Q., Vorobiev, S., Almo, S. C., and Chance, M. R. (2002) Mapping the G-actin binding surface of cofilin using synchrotron protein footprinting, *Biochemistry 41*, 5765−5775.
21. Guan, J. Q., Almo, S. C., Reisler, E., and Chance, M. R. (2003) Structural reorganization of proteins revealed by radiolysis and mass spectrometry: G-actin solution structure is divalent cation dependent, *Biochemistry 42*, 11992−12000.
22. D'Ambrosio, C., Talamo, F., Vitale, R. M., Amodeo, P., Tell, G., Ferrara, L., and Scaloni, A. (2003) Probing the dimeric structure of porcine aminoacylase 1 by mass spectrometric and modeling procedures, *Biochemistry 42*, 4430−4443.
23. Goldsmith, S. C., Guan, J. Q., Almo, S., and Chance, M. (2001) Synchrotron protein footprinting: a technique to investigate protein−protein interactions, *J. Biomol. Struct. Dyn. 19*, 405−418.
24. Liu, R., Guan, J. Q., Zak, O., Aisen, P., and Chance, M. R. (2003) Structural reorganization of the transferrin C-lobe and transferrin receptor upon complex formation: the C-lobe binds to the receptor helical domain, *Biochemistry 42*, 12447−12454.
25. Maleknia, S. D., Ralston, C. Y., Brenowitz, M. D., Downard, K. M., and Chance, M. R. (2001) Determination of macromolecular folding and structure by synchrotron X-ray radiolysis techniques, *Anal. Biochem. 289*, 103−115.
26. Nukuna, B. N., Sun, G., and Anderson, V. E. (2004) Hydroxyl radical oxidation of cytochrome *c* by aerobic radiolysis, *Free Radical Biol. Med. 37*, 1203−1213.
27. Henricksen, L. A., Umbricht, C. B., and Wold, M. S. (1994) Recombinant replication protein A: expression, complex formation, and functional characterization, *J. Biol. Chem. 269*, 11121−11132 [erratum: (1994) *J. Biol. Chem. 269*, 16519].
28. Bochkareva, E., Korolev, S., and Bochkarev, A. (2000) The role for zinc in replication protein A, *J. Biol. Chem. 275*, 27332−27338.
29. Salvucci, M. E. (1993) Covalent modification of a highly reactive and essential lysine residue of ribulose-1,5-bisphosphate carboxylase/oxygenase activase, *Plant Physiol. 103*, 501−508.
30. Nuss, J. E., Patrick, S. M., Oakley, G. G., Alter, G. M., Robison, J. G., Dixon, K., and Turchi, J. J. (2005) DNA damage induced hyperphosphorylation of replication protein A. 1. Identification of novel sites of phosphorylation in response to DNA damage, *Biochemistry 44*, 8428−8437.
31. Nuss, J. E., and Alter, G. M. (2004) Denaturation of replication protein A reveals an alternative conformation with intact domain structure and oligonucleotide binding activity, *Protein Sci. 13*, 1365−1378.
32. Tsunematsu, H., Imamura, T., and Makisumi, S. (1983) Kinetics of hydrolysis of Nα-benzoyl-*p*-guanidino-L-phenylalanine *p*-nitroanilide by trypsin, *J. Biochem. 94*, 123−128.
33. Pozsgay, M., Szabo, G., Bajusz, S., Simonsson, R., Gaspar, R., and Elodi, P. (1981) Investigation of the substrate-binding site of trypsin by the aid of tripeptidyl-*p*-nitroanilide substrates, *Eur. J. Biochem. 115*, 497−502.
34. Nagata, K., Yoshida, N., Ogata, F., Araki, M., and Noda, K. (1991) Subsite mapping of an acidic amino acid-specific endopeptidase from *Streptomyces griseus*, GluSGP, and protease V8, *J. Biochem. 110*, 859−862.
35. Houmard, J. (1976) Kinetic investigation of the staphylococcal protease-catalyzed hydrolysis of synthetic substrates, *Eur. J. Biochem. 68*, 621−627.
36. Tang, C. L., Xie, L., Koh, I. Y., Posy, S., Alexov, E., and Honig, B. (2003) On the role of structural information in remote homology

detection and sequence alignment: new methods using hybrid sequence profiles, *J. Mol. Biol. 334*, 1043−1062.
37. Zhao, L., Cox, A. G., Ruzicka, J. A., Bhat, A. A., Zhang, W., and Taylor, E. W. (2000) Molecular modeling and in vitro activity of an HIV-1-encoded glutathione peroxidase, *Proc. Natl. Acad. Sci. U.S.A. 97*, 6356−6361.
38. Bates, P. A., Kelley, L. A., MacCallum, R. M., Sternberg, M. J. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM, *Proteins, Suppl. 5*, 39−46.
39. Means, G., and Feenery, R. (1971) *Chemical Modification of Proteins*, Holden-Day, San Francisco, CA.
40. Gomes, X., Henricksen, L., and Wold, M. (1996) Proteolytic mapping of human replication protein A: evidence fro multiple structureal domains and a contormational change upon interaction with single-stranded DNA, *Biochemistry 35*, 5586−5595.
41. Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules, *J. Comput. Chem. 19*, 319−333.
42. Pearlman, D. A. C., Caldwell, J. W., Ross, W. R., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a computer program for applying molecular mechanics, normal-mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, *Comput. Phys. Commun. 91*, 1−41.
43. Letner, C., and Alter, G. (1994) The evolution of lactate dehydrogenase conformation during molecular dynamics simulation, in *Proceedings of the First Electronic Computational Chemistry Conference*, ARInternet.
44. Zhang, K., Tang, H., Huang, L., Blankenship, J. W., Jones, P. R., Xiang, F., Yau, P. M., and Burlingame, A. L. (2002) Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry, *Anal. Biochem. 306*, 259−269.
45. Deleted in proof.
46. Ellison, D., Hinton, J., Hubbard, S. J., and Beynon, R. J. (1995) Limited proteolysis of native proteins: the interaction between avidin and proteinase K, *Protein Sci. 4*, 1337−1345.
47. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors, *J. Mol. Biol. 220*, 507−530.
48. Hubbard, S. J., Beynon, R. J., and Thornton, J. M. (1998) Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures, *Protein Eng. 11*, 349−359.
49. Mustard, D., and Ritchie, D. W. (2005) Docking essential dynamics eigenstructures, *Proteins: Struct., Funct., Bioinf. 60*, 269−274.
50. Ritchie, D. W. (2003) Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2, *Proteins 52*, 98−106.
51. Ritchie, D. W., and Kemp, G. J. (2000) Protein docking using spherical polar Fourier correlations, *Proteins 39*, 178−194.
52. Creighton, T. (1993) *Proteins: Structure and Molecular Properties*, W. H. Freeman, New York.
53. Doneanu, C. E., et al. (2004) Mass spectrometry of UV-cross-linked protein-nucleic acid complexes: Identification of amino acid residues in the single-stranded DNA-binding domain of human replication protein A, *Anal. Chem. 76*, 5667−5676.
54. Park, J. S., Wang, M., Park, S. J., and Lee, S. H. (1999) Zinc finger of replication protein A, a non-DNA binding element, regulates its DNA binding activity through redox, *J. Biol. Chem. 274*, 29075−29080.
55. Walther, A. P., Gomes, X. V., Lao, Y., Lee, C. G., and Wold, M. S. (1999) Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit, *Biochemistry 38*, 3963−3973.
56. Lao, Y., Lee, C. G., and Wold, M. S. (1999) Replication protein A interactions with DNA. 2. Characterization of double-stranded DNA-binding/helix-destabilization activities and the role of the zinc-finger domain in DNA interactions, *Biochemistry 38*, 3974−3984.
57. Eckerich, C., Fackelmayer, F. O., and Knippers, R. (2001) Zinc affects the conformation of nucleoprotein filaments formed by replication protein A (RPA) and long natural DNA molecules, *Biochim. Biophys. Acta 1538*, 67−75.
58. Pestryakov, R., Weisshart, K., Schlott, B., Khodyreva, S., Kremmer, E., Grossie, F., Larvrik, O., and Nasheuer, H. (2003) The C-terminal RPA70 and the central RPA32 domains are involved in the interactions with the 3′-end of a primer-template DNA, *J. Biol. Chem. 278*, 17515−17524.
59. Blackwell, L. J., Borowiec, J. A., and Masrangelo, I. A. (1996) Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase, *Mol. Cell. Biol. 16*, 4798−4807.

BI052397V