Multimerization Determinants Reside in Both the Catalytic Core and C Terminus of Avian Sarcoma Virus Integrase*

Mark D. Andrake‡ and Anna Marie Skalka§

From the Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania 19111

We have shown previously that the active form of avian sarcoma virus integrase (ASV IN) is a multimer. In this report we investigate IN multimerization properties by a variety of methods that include size exclusion chromatography, chemical cross-linking, and protein overlay assays. We show that removal of the nonconserved C-terminal region of IN results in a reduced capacity for multimerization, whereas deletion of the first 38 amino acids has little effect on the oligomeric state. Binding of a full-length IN fusion protein to various IN fragments indicates that sequences in both the catalytic core (residues 50–207) and a C-terminal region (residues 201–240) contribute to IN self-association. We also observe that the isolated C-terminal fragment (residues 201–286) is capable of self-association. Finally, a single amino acid substitution in the core domain (S85G) produces a severe defect in multimerization. We conclude from these analyses that both the catalytic core and a region in the nonconserved C terminus are involved in ASV integrase multimerization. These results enhance our understanding of integrase self-association determinants and define a major role of the C-terminal region of ASV integrase in this process.

The integration of viral DNA into the genome of the host cell is a unique and vital step in the normal life cycle of retroviruses (for recent reviews, see Refs. 1 and 2). The retroviral integrase (IN) is both necessary and sufficient for the integration of a linear DNA with viral ends into a target DNA in vitro (3, 4). Our knowledge of retroviral integrase structure and function continues to be refined by identification of domains of the enzyme that contribute to the various functions required during integration (Fig. 1). These include DNA binding, catalytic activities, and the multimerization required for the coordinated joining of processed viral ends to the site of integration in host DNA. Several investigations have mapped a nonspecific DNA binding domain to C-terminal portions of both ASV and HIV-1 IN (5–9). Catalytic functions have been localized to the central core domain, which is resistant to limited proteolysis (10), and contains a conserved triad of acidic residues [D,D,D(35E)] that is presumed to bind the divalent cations that are required for catalytic activity (5, 11). The isolated catalytic domain is competent to perform a concerted cleavage-ligation activity (12–14) and contributes to the recognition of conserved CA residues present at the 3' ends of retroviral terminal repeats and other transposable elements (15). However, it cannot perform the viral DNA end processing and joining reactions required for insertion of viral DNA into target DNA sequences. The crystal structures of the catalytic cores of both HIV-1 IN and ASV IN have been solved recently (16, 17).

Many enzymes that catalyze DNA recombination require the formation of multimeric protein-DNA complexes. Detailed structural information is available for some of these (18–22). However, important questions concerning the stoichiometry of IN protein and DNA substrates in the nucleoprotein complex that is competent for integration remain unanswered. In addition, the role of protein multimerization in substrate binding and catalysis is yet to be clearly defined.

Both biochemical and genetic studies indicate that multimerization is a functionally important property of retroviral integrases. ASV IN, purified from viral particles, and bacterially expressed HIV-1 IN were observed to migrate in glycerol gradients at a position consistent with a dimer molecular weight (23, 24). Gel filtration studies also suggest that purified HIV-1 IN forms dimers (25). We have used sedimentation and kinetic studies to demonstrate that purified ASV IN, which exhibits a reversible mass action between monomer ↔ dimer ↔ tetramer, must multimerize to perform its catalytic function (26). Similar sedimentation analyses have been performed with HIV-1 IN (27). The coordinated action of Moloney murine leukemia virus IN on both ends of viral DNA was demonstrated by mutagenesis studies and analysis of intermediates produced in vivo (28). Our laboratory has recently obtained similar results with ASV IN in vitro, using DNA substrates that link two viral DNA end sequences (29). Multimerization has also been inferred from the enzymatic complementation of two defective HIV-1 IN mutants when incubated together (30, 31). Finally, a yeast transcriptional reporter system was used to analyze HIV-1 IN homomeric interactions and to determine a minimal domain required for self-association (32).

The work presented here includes physical analyses of the multimeric state of purified ASV IN and various IN fragments using size exclusion chromatography (SEC) and chemical cross-linking techniques. To further delineate the regions of this enzyme that contribute to self-association, we have also employed a modification of a protein overlay blot technique which tests directly for binding between two potential associating proteins. In addition to demonstrating a role for the conserved catalytic core domain in ASV IN dimerization, our results un-

---

* The work was supported in part by National Institutes of Health Grants CA-47486 and CA-06927, a grant for infectious disease research from Bristol-Myers Squibb Foundation, and also by an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported in part by United States Public Health Service Fellowship 5F32 AI08642 and T32 CA09035-17.

§ To whom correspondence should be addressed: Fox Chase Cancer Center, Institute for Cancer Research, 7701 Burholme Ave., Philadelphia, PA 19111. Tel.: 215-728-2490; Fax: 215-728-2778.

1 The abbreviations used are: IN, integrase; PAGE, polyacrylamide gel electrophoresis; ASV, avian sarcoma virus; GST, glutathione S-transferase; DSP, dithiobis(succinimidylpropionate); EDC, 1-ethyl-3(3-dimethylaminopropyl)carbodiimide; BS5, bis(sulfosuccinimidyl)suberate; MES, 2-(N-morpholino)ethanesulfonic acid; SEC, size exclusion chromatography; HIV-1, human immunodeficiency virus type 1; CHAPS, 3-[3-cholamidopropyl]dimethylammonio-1-propanesulfonic acid.
cover an important determinant for multimerization located in the less conserved C-terminal domain.

MULTIPLIZATION OF ASV INTEGRASE

Plasmids and Cloning

The cloning of nonfused ASV IN and IN fragments has been described (15). To produce plasmid DNA clones encoding IN fragments with the indicated boundaries fused to glutathione-S-transferase (GST), polymerase chain reaction was performed with Vent polymerase (Boehringer) for 25 cycles from the template pRC23p32 (33), using primers with flanking BamHI and EcoRI sites suitable for insertion in the expression vector, pGEX-2TK (Pharmacia Biotech Inc.). This expression vector encodes a kinase labeling site at the junction between GST and the cloned insert of the fusion protein. Polymerase chain reaction fragments were purified prior to digestion and ligation; cloning and screening procedures followed standard practice (34).

The IN(201–286)st clone, which expressed the IN fragment 201–286, included a streptavidin epitope tag at the C terminus. It was constructed in three steps as follows. 1) A DNA duplex fragment encoding a kinase/strep-tag sequence (35) of 15 amino acids (Arg-Arg-Ala-Ser-Ala-Trp-Arg-His-Pro-Gln-Phe-Gly-Gly) with compatible ends was inserted into BamHI/HindIII-digested pSE380 (36). 2) A polymerase chain reaction fragment encoding IN amino acids 201–286 (using primers with flanking NcoI and BamHI sites) was inserted into the pSE380 strep-tag derivative. 3) The IN-strep-tag encoding DNA fragment was purified and ligated to NcoI/HindIII-digested expression vector pET-20b (Novagen). The polypeptide expressed from this construct includes a leader sequence (pelB) from the pET-20b vector, which is intended for periplasmic targeting of overexpressed proteins. However, the most effective purification of IN(201–286)st was from whole cell lysates. The apparent molecular weight from SDS gels indicate that this leader is cleaved by a signal peptidase. The IN sequences of all clones were confirmed by sequencing, and fusion proteins of expected molecular weight were synthesized in all cases.

FIG. 1. Domain structure of retroviral integrases. The figure shows the relative positions of the major domains along the linear sequence of retroviral integrases. The scale of amino acid numbering is indicated at the top (ASV IN = 286 amino acids; HIV-1 IN = 288 amino acids). The catalytic core is an evolutionarily conserved region among retroviral INs and certain transposases; the acidic residues (D, D, E) presumed to bind the divalent cations required for activity have been positioned to scale. Another conserved motif, the HHCC region, is located in the N-terminal region. It contains appropriately spaced histidines and cysteines characteristic of several Zn²⁺-binding domains. The amino acid sequence in the C-terminal region (amino acids beyond 200) is not highly conserved among retroviral INs. It has been shown to contain determinants for nonspecific DNA binding and, in addition to the core region, likely plays a role in substrate binding during catalysis. The location of the two self-association regions for ASV IN identified in this report are also indicated. The five ASV IN truncated proteins used in this report are diagrammed below, with the shaded box indicating the portion of IN encoded. The construct name includes the end points of amino acids contained in each polypeptide.
min at 4°C. After quenching, the glutathione resin was washed 4 times with NETN buffer prior to elution of the labeled fusion protein as described above.

Strap-tag Protein Expression and Purification—Induction of expression and cell lysis were performed essentially as described for GST-fusion proteins above. Streptavidin-agarose (Pierce) was added to cleavage affinity purification buffer, and the batch procedure. Fusion proteins were exposed to the resin for longer than 1 h at 5°C. The resin was then washed 5 times with 20 ml HEPES, pH 8, 0.5 mM NaCl, 1% (v/v) glycerol, 1 mM EDTA. Strap-tag-fusion proteins were eluted from the resin with 1 mM D-biotin in elution buffer. Eluted fusion protein was then dialyzed overnight at 5°C against 50 mM HEPES, pH 8, 0.5 mM NaCl, 1% (v/v) glycerol, 1% (v/v) D-threonine. This procedure allowed the quantitation of an identically loaded Coomassie-stained gel of GST-fusion proteins for both probe and targets. GST has been shown to be a dimer in the active state (37).

Chemical Cross-linking
Cross-linking of IN and IN fragments was carried out using a variety of chemical reagents (Pierce) with buffer and protein concentrations noted in the figures. Reactions using the reagents DSP and BS3 were performed in 100 mM HEPES buffer, pH 8.0. Reactions with the reagent EDC were performed in 100 mM MES, pH 5.6, and 5 mM N-hydroxysuccinimidyld ester. Unless otherwise noted, all reactions contained 600 mM NaCl and were quenched with a molar excess of either glycerine or lysis prior to the addition of an equal volume of SDS sample buffer containing 280 mM 2-mercaptoethanol. The 2-mercaptoethanol was omitted if the cross-linker contained a disulfide bond for cleavage, i.e. DSP. Samples were subsequently heated at 95°C for 10 min. Covalently linked multimers were detected by separation in 12% or 15% SDS-polyacrylamide gels and silver staining.

Size Exclusion Chromatography of Full-length and Truncated IN Proteins
Size exclusion chromatography was performed using a Superdex 75/HR 10/30 column (Pharmacia) on a Rainin HPLC system with native IN or nonfused IN fragments. A flow rate of 0.5 ml/min with a mobile phase of 50 mM HEPES, pH 7.5, 0.5 mM NaCl, 1% (v/v) glycerol was used for all experiments. Typically, 25–50 µl of purified IN polypeptides at a concentration of 30 µM were injected. Samples were subjected to centrifugation at 10,000 × g prior to injection on the column. Absorbance of the column eluate was monitored at both 280 and 220 nm. Samples from peak fractions were monitored by SDS-PAGE for the presence of the expected protein species. The column was calibrated using seven different globular proteins as molecular weight standards, and the apparent molecular weight of each sample peak was determined using linear regression of the log of known molecular weight versus the elution behavior (Kow or elution time).

The SEC analyses presented do not attempt to account for the dissociation kinetics of IN multimers. The observation of discrete monomer and dimer peaks in some chromatograms (for example, Fig. 4, traces 4 and 5) indicates that the dissociation rate of dimer to monomer should be slow relative to the column run time. The intent of the experiments presented here was to compare the behavior of different polypeptides and not to establish absolute quantitative constants for IN self-association. Sedimentation equilibrium analysis is more suited to the quantitative determination of association constants, and the results of such studies (in progress) will be reported separately.

Protein Overlay Binding Assay
To test the ability of a labeled protein probe to bind to a defined set of target IN fragments, standard SDS-PAGE was performed on the target polypeptides prior to electrophoretic transfer to either a polyvinylidene difluoride or nitrocellulose membrane using a 3-buffer semidry technique (according to Millipore recommended procedures). Subsequent steps of blocking, renaturing, and probing were carried out at 4°C. For buffers requiring dithiothreitol, this reagent was added just prior to use. The transfer membrane was blocked in binding buffer (25 mM HEPES-KOH, pH 7.7, 25 mM NaCl, 5 mM MgCl2, 1 mM dithiothreitol) containing 5% (w/v) powdered milk and 0.05% (v/v) Nonidet P-40 (blocking solution) for at least 1 h (typically overnight). Target polypeptides were denatured on the membrane by two successive incubations in binding buffer containing 6 M guanidine HCl for 10 min. Slow renaturation was accomplished by 6 successive dilutions with an equal volume of binding buffer, each with a 10–15 min incubation. After two washes with binding buffer, the membrane was treated once more with blocking solution for 1 h and for 30 min in blocking solution with only 1% powdered milk. A molar excess of labeled probe protein was diluted into probe buffer (20 mM HEPES, pH 7.7, 75 mM KCl, 0.1 mM EDTA, 2.5 mM MgCl2, 1% powdered milk, 0.05% Nonidet P-40, 1 mM dithiothreitol). This probe mixture also included at least a 2-fold molar excess of unlabelled GST protein to block nonspecific binding to the GST-fusion targets. The probe was incubated with the renatured blot for greater than 6 h, followed by three successive 10-min washes in probe buffer alone. The membrane was then dried and the radioactivity quantitated on a Fuji BAS1000 phosphorimaging system. The amount of radioactive probe bound to target bands of interest was normalized to the quantity of the target polypeptide loaded on the gel, as assessed by densitometric quantitation of an identically loaded Coomassie-stained gel. In Fig. 6, the level of binding to GST alone is indicated and should be considered a background for each probe. Note that the binding of the gN(1–286) probe to the full-length target is greater than 20-fold above binding to GST alone, demonstrating an adequate signal-to-noise ratio under our assay conditions. Experiments were repeated with several of the constructs using nonfused protein targets (data not shown), and similar results were obtained.

RESULTS
Integrase Multimers Revealed by Chemical Cross-linking—Chemical cross-linking has been employed successfully to examine the protein-protein associations of many multisubunit enzymes. We have used a variety of chemical cross-linking reagents to identify the multimeric forms of ASV IN that exist in solution. As shown in Fig. 2A, titration of the cross-linker DSP revealed dose-dependent formation of covalently linked forms of IN (lanes 2–6). As the concentration of cross-linker was increased, the amount of dimers, tetramers, and higher order multimers increased while the amount of free monomer decreased. The dominant multimer observed was a dimer. Optimal concentrations of DSP were between 1 and 2 mM; higher concentrations of the cross-linker produced aggregates which failed to enter the gel (Fig. 2A, lane 7). The DSP cross-linker contains a disulfide bond which can be reduced, thereby breaking the covalent link between subunits. As expected, treatment with β-mercaptoethanol led to the disruption of cross-linked multimers (Fig. 2A, compare lanes 6 and 8). Fig. 2B shows results of treatment with increasing concentrations of the “zeolength” cross-linker EDC. Due to the short length of the covalent bond formed (the length of a typical amide bond), detection of cross-linked multimers using this reagent (lanes 2–5) provided strong evidence that the subunit association is structurally significant (38, 39).

Other reagents with different chemistries (e.g. glutaraldehyde and dimethyl 3,3′-dithiodipropionimidate) were also examined for their ability to covalently cross-link IN multimers. Multimeric complexes similar in composition and amount to those observed with DSP and EDC were detected with these reagents (data not shown), suggesting that a variety of reactive residues (basic, acidic, and others) must be present at or near the interface between monomer subunits.

Experiments in which the IN concentration was decreased while the cross-linker concentration was held constant showed that dimers are present at IN concentrations as low as 70 nM (Fig. 2C). These data also reveal a concentration dependence for tetramer formation (Fig. 2C, compare lanes 2 and 3). These results are consistent with previous estimates of a Kd (monomer-dimer) in the 1–5 μM range (26).

Cross-linking Analysis of Truncated Derivatives of IN—We have previously described a series of N- and C-terminal deletions in ASV IN (15). Two of these, IN(1–207) and IN(52–207), were shown to have lost normal processing and joining activities, but retained endonuclease and cleavage-ligation activity.
Multimerization of ASV Integrase

Chemical cross-linking of IN subunits. A, titration of the cross-linking reagent DSP. Reactions containing 7 μM IN (wild type, nonfused) and various concentrations of DSP were incubated for 30 min at 22 °C. They were then quenched by addition of a molar excess of glycine and an equal volume of SDS sample buffer without a reducing agent (except in lane 8 where 2-mercaptoethanol was included). Samples were analyzed by 12% SDS-PAGE and silver staining. The positions of the molecular mass markers are noted to the left; positions of IN multimers are indicated to the right (1-IN = IN monomer, 2-IN = IN dimer, 4-IN = IN tetramer). The concentrations of DSP in lanes 2, 3, 4, 5, 6, and 7 were 2 μM, 20 μM, 0.2 mM, 1 mM, 2 mM, and 20 mM, respectively. Lane 8 contained 2 mM DSP. The presence of trace amounts of dimer in the lane with no cross-linker is due to fortuitous disulfide bonds formed between native IN monomers that are not reduced, since the gel sample buffer could not contain reducing reagents directed against ASV IN (data not shown). Cross-linking of this fragment can account for some of the broadness of the dimer band seen in this figure and Fig. 3, B, cross-linking with the reagent EDC. Reactions containing 10 μM IN, 650 mM NaCl, and increasing concentrations of EDC were incubated for 30 min at 22 °C. Labeling is as in A. The concentrations of EDC in lanes 1, 2, 3, 4, and 5 were 8 μM, 40 μM, 0.2 mM, 1 mM, and 5 mM, respectively. C, chemical cross-linking with lower concentrations of IN. Concentrations of IN in lanes 2, 3, 4, and 5 were 7 μM, 3.5 μM, 70 μM, and 350 μM, respectively. The assignment of the band labeled 4-IN (tetramer) is more apparent in lower percentage gels. Reaction mixtures contained 0.5 mM NaCl, 0.2 mM BS3, and were incubated for 15 min at 22 °C. No cross-linker was added in lane 1. Labeling is as in A.

The cross-linker BS3 was used to determine the ability of these truncated forms of IN to form covalently linked dimers (see Fig. 3). The results showed significantly less cross-linked dimer with IN(1-207) (Fig. 3A, lanes 7-9), compared to full-length IN (Fig. 3A, lanes 3-5). At a concentration of 2 mM BS3 in which all full-length IN was covalently linked in multimeric forms, most of IN(1-207) remained monomeric. We conclude that IN(1-207) is deficient in multimerization. Analysis of IN(52-207), which also lacks amino acids from the N terminus, showed a similar deficiency (Fig. 3B, compare lane 2 with lane 8). In contrast, deletion of residues 5-38 from the N-terminal segment alone did not cause reduction in multimerization (Fig. 3B, compare lane 2 and lane 4).

No heterodimerization was observed in mixtures of full-length and IN(1-207) polyptides in this assay (data not shown). It is possible that under these experimental conditions, exchange of monomer subunits proceeds too slowly for heterodimers to form. However, this could also reflect the fact that the affinity of an IN(1-207) subunit for the full-length IN is significantly lower than two full-length IN monomers for each other.

Multimerization Detected by Size Exclusion Chromatography—The oligomeric composition of IN and IN truncated proteins in solution was also investigated by size exclusion chromatography. IN these experiments, the properties of full-length and truncated proteins were compared at similar molar concentrations. Under the conditions described for Fig. 4, full-length IN eluted at a position consistent with a dimer molecular size (trace 1). The peak was asymmetric, with a shoulder at the position of the monomer. At lower loading concentrations of IN (not shown), a greater percentage eluted in the monomer shoulder, as expected for a concentration-dependent monomer-dimer equilibrium. No dimer was detected with IN(52-207) under the same conditions; this polyptide comprises the catalytic core of ASV IN (trace 6). However, much higher loading concentrations of IN(52-207) did reveal a minor peak consistent with a dimer molecular size (data not shown). An additional

1 R. A. Katz, G. Merkel, and A. M. Skalka, submitted for publication.

2 R. A. Katz, G. Merkel, and A. M. Skalka, submitted for publication.
Multimerization of ASV Integrase

Size exclusion chromatography was performed as described under “Materials and Methods.” The chromatograms display the absorbance at 220 nm as a function of elution time. The trace numbers noted in the text are found to the left of each chromatogram. For reference, the elution positions of 2 (of a total of 7) globular molecular mass standards are indicated with dotted vertical lines. The loading concentration of each polypeptide was approximately 30 μM. Lightly shaded traces indicate proteins that are competent for multimerization; hashed traces indicate proteins partially defective in multimerization, and darkly shaded traces indicate proteins most defective in multimerization. We note that full-length IN elutes with an apparent molecular mass that can be best assigned as a dimer (~57.8 kDa), yet at a time slightly later than expected for an ideal globular dimer. This behavior is similar to that observed with HIV-1 IN (25). The slightly retarded elution time most likely reflects nonspecific interaction of IN with the column matrix. This has been observed with other SEC media used for ASV IN (data not shown) and was observed for the catalytic core of HIV-1 IN, where inclusion of CHAPS in the running buffer decreased elution time (43).

13 amino acids at the N terminus of the catalytic core, IN(39–207), enhanced dimerization to an intermediate level (trace 5). IN(1–207) eluted with a profile similar to that observed for IN(39–207). Most of the IN(1–207) protein eluted as expected for a monomer (apparent molecular mass ~23 kDa), with a smaller peak containing 15–20% of the eluting protein at the position of a dimer (trace 4). In contrast to C-terminal truncations, the N-terminal deletion protein IN(39–286) exhibited an elution pattern like that of the full-length IN (trace 3).

In the course of site-directed mutagenesis studies, our laboratory has prepared a number of altered ASV IN proteins that contain single amino acid substitutions in residues that are highly conserved in retroviruses and certain other transposable elements (11). A number of these proteins (D64E, T66A, F126A, D121E, L163A, H9N, K206A, R227A) were examined by SEC (data not shown). Only one of these, S85G, exhibited a significant difference when compared with wild type protein. As illustrated in Fig. 4 (trace 2), S85G eluted exclusively as a monomer.

The results of these SEC analyses are in general agreement with those of the chemical cross-linking studies. Both sets of data indicate that the catalytic core domain of ASV IN can dimerize, but with reduced efficiency compared to the full-length protein. Addition of the C-terminal region appears to restore full multimerization capability. We conclude that self-association determinants are located in both the core and C-terminal regions of ASV IN. The inability of the S85G mutant to dimerize suggests that substitutions in this residue alter the catalytic core structure, or the way in which the core interacts with the C-terminal domain. Analysis of the crystal structure of ASV IN(52–207) reveals that the side chain of this residue participates in a network of hydrogen bonds in a tight turn between two β-strands (17).

Localization of Sdf-associating IN Domains—In order to map the self-association domains of ASV IN and to gain a better understanding of their relationship to one another, we used deletion mutagenesis coupled with a protein overlay technique. This technique (40, 41) employs a labeled protein to probe a Western blot of target proteins that are first denatured and then renatured. We modified this procedure to investigate the self-association potential of individual regions of ASV IN. For the first experiment, a labeled GST-IN fusion protein, gIN(1–286), was the probe for a series of renatured targets which included full-length IN, IN(1–207), and other non-integrase protein controls. The data showed efficient binding of the probe to full-length IN (Fig. 5B, lane 3) and no detectable binding to the bovine serum albumin and molecular mass standards (Fig. 5B, lanes 1 and 2). Probe binding to full-length IN could be competed by incubation of the blot with unlabeled, full-length IN fusion, but not with the GST portion of the fusion protein alone (data not shown). Therefore, we conclude that the binding reaction is specific for IN-derived polypeptides.

The C-terminal truncation protein showed reduced binding (Fig. 5B, lane 4), even though equal molar amounts of this
Multimerization of ASV Integrase

Fig. 6. Mapping determinants of IN self-association. The left part of the figure shows the set of truncated proteins generated to test for probe binding. The dark bars indicate the portions of IN included in each GST-fusion, and the corresponding name is listed in the first column of the table to the right. Dotted lines delineate the catalytic core and C-terminal regions proposed to be involved in ASV IN self-association. The variable shading of the C-terminal self-association region (here and in Fig. 1) indicates the relative contribution to multimerization conferred by the sequences within this region. The last two columns of the table present the normalized data for binding of two different probes, full-length IN fusion, gIN(1–286), and gIN(201–236), to each target protein. ND indicates not done.

| Target | % Binding to full-length target | Probe gIN1-286 | Probe gIN201-236 |
|--------|---------------------------------|---------------|-----------------|
| glN1-286 | 100                             | 100            |                 |
| glN60-286 | 93                              | 75             |                 |
| glN120-286 | 40                              | 71             |                 |
| glN156-286 | 47                              | 83             |                 |
| glN178-286 | 59                              | 74             |                 |
| glN201-286 | 37                              | 67             |                 |
| glN223-286 | 12                              | 73             |                 |
| glN240-286 | 5                               | ND             |                 |
| gst     | 4                               | 14             |                 |
| glN1-236 | 69                              | 136            |                 |
| 1-207   | 26                              | 0              |                 |

polypeptide were used in the assay (Fig. 5A, compare lanes 3 and 4). The relative amount of probe bound was quantitated by radioanalytic imaging and normalized to the amount of protein present on the filter. These calculations indicated that the C-terminal truncation protein bound to the probe with 5- to 10-fold lower efficiency than the full-length IN. Since these results were consistent with those from our physical assays of IN multimerization, we used this method to screen a series of nested N-terminal truncated IN proteins. For ease and uniformity of purification, these truncated proteins were constructed as GST-fusion proteins (see Fig. 6 and "Materials and Methods"). The fusion proteins were expressed in E. coli, affinity-purified, and used as targets in the protein overlay assay. The relative binding capacity of these proteins was tested with two probes, full-length gIN(1–286) and gIN(201–236), and quantitated as described above. Fig. 6 provides a map of the IN deletion proteins tested and a summary of the results of this quantitation expressed as percent of probe bound to a full-length IN target protein.

Deletion of the N-terminal HHCC domain did not significantly change the binding with full-length probe (compare the gIN(1–286) and gIN(60–286) targets), consistent with results from SEC and cross-linking with the IN(39–286) protein. Further N-terminal truncation, which removed part (gIN(120–286)) or all (gIN(156–286)) of the D,D(35)E region, reduced binding to an intermediate level, 40–60%. Continued truncation did not change binding significantly until the deletion extended into the C-terminal domain which includes amino acids downstream of residue 201. After that, binding of the probe continued to decrease [gIN(223–286)] until the deletion included amino acid 239 [gIN(240–286)], when it reached a background level, equivalent to binding to the GST alone control.

These results delineate two regions critical to IN self-association. The first corresponds to the central catalytic D,D(35)E domain, and the second lies in a C-terminal domain including, but perhaps not limited to, amino acids 201–240 of ASV IN. Results with targets which are truncated from the C terminus were consistent with this designation. The target protein gIN(1–236) bound the full-length probe with approximately 70% efficiency, whereas the efficiency with IN(1–207) was only 26%.

The association properties of the C-terminal domain were further investigated by probing the same set of target proteins with the gIN(201–236) fusion protein. The results with this probe revealed a pattern distinct from that of the full-length probe. All of the N-terminal truncation proteins up to and including gIN(223–286) bound the gIN(201–236) probe with approximately 70–80% the efficiency of the full-length target (Fig. 6, last column). The C-terminal truncation IN(1–236) bound this probe as well as or better than the full-length IN, but no binding above background was observed to the IN(1–207) target.

It is possible that this method could detect both quaternary interactions between IN monomers and tertiary interactions that reflect the folding of domains within an IN monomer polypeptide. However, results with the gIN(201–236) probe, which showed equivalent binding to all targets that contained residues 201–240, make it unlikely that this reaction is simply mimicking tertiary interactions. They suggest, instead, that this C-terminal peptide is capable of specific association with the homologous region in another target polypeptide.

An IN(201–286) Fragment Can Self-associate—As a final test of the self-associating capability of the C-terminal domain, we expressed and partially purified IN(201–286) as a streptagged protein (35) and analyzed this polypeptide by SEC and chemical cross-linking (Fig. 7). Under the SEC conditions used, a monomer of this fragment is expected to elute at 24.5 to 25 min. The results showed that this C-terminal fragment eluted at 20.5 min, consistent with a formation of a multimer (trimmer or tetramer). Material in the void volume contained other protein and nucleic acid contaminants, whereas the multimer peak contained the bulk of the IN(201–286) fragment (confirmed by SDS-PAGE analysis of fractions, data not shown). This C-terminal fragment was also tested in chemical cross-linking experiments, performed as in Fig. 3B. These results showed that the C-terminal polypeptide is able to form dimers, trimers, and tetramers in solution (Fig. 7 inset at top). The self-association properties of this fragment are not as limited as those observed with other fragments analyzed. It could be that this isolated C-terminal fragment is less sterically restrained, allowing it to form multimers (e.g. trimers) not found with other larger fragments. We conclude that the C-terminal fragment, IN(201–286), can self-associate as an independently expressed polypeptide.

**DISCUSSION**

Multimerization Detected by Cross-linking and SEC—We have investigated the self-association properties of ASV IN and IN fragments using a variety of methods, each of which offers...
Multimerization of ASV Integrase

Fig. 7. Self-association of the C-terminal IN fragment IN(201–286). SEC was performed using the C-terminal fragment IN(201–286) as in Fig. 4. The chromatogram shows two peaks: the first eluted in the void volume of the column, and the second (labeled IN(201–286)) eluted at a position consistent with a multimer of the IN(201–286) fragment. The multimer peak was confirmed to contain the IN(201–286) fragment by SDS-PAGE analysis of the fractions. The 260:280 nm absorbance ratio of the void peak was consistent with the presence of both nucleic acid and protein components in this fraction, which could account for the observed aggregation. The inset shows results from a chemical cross-linking experiment with the IN(201–286) fragment and cross-linker BS3 under the conditions described in Fig. 3B. The positions of covalently linked dimers, trimers, and tetramers are noted.

distinct advantages and limitations. Chemical cross-linking is the least stringent of the tests for multimerization because protein concentration can be controlled to favor association. SEC is the most stringent assay presented here because dissociative forces due to dilution are prominent during migration of multimeric proteins through a column, and the detection of multimers depends on the rate of dissociation relative to column run times (42). Accordingly, chemical cross-linking of the ASV IN catalytic core (IN(52–207)) clearly reveals dimer formation (Fig. 3), whereas in SEC, this fragment runs predominantly as a monomer.

Results obtained with these nonequilibrium methods are consistent with independent estimates of dissociation constants determined from sedimentation equilibrium experiments. Whereas full-length ASV IN has a $K_d$ (monomer-dimer) of 1–5 $\mu$M (26), ASV IN(52–207) has a $K_d$ (monomer-dimer) in excess of 500 $\mu$M. The HIV-1 catalytic core (residues 50–212) has been reported to have a stronger association than that of the analogous ASV fragment, and dimers have been observed with chemical cross-linking, SEC, and sedimentation analysis of the HIV-1 IN core fragment (27, 30, 43).

Both the cross-linking and SEC methods record the behavior of the majority of molecules in the protein preparations, whereas multimerization inferred from enzymatic complementation (30, 31) or a transcriptional reporter system (32) could reflect the activity of a small fraction of protein present. In addition, the latter assays include DNA substrates which could facilitate the formation of multimers of IN. We have performed cross-linking experiments in the presence of various DNA substrates and have failed to detect enhancement of ASV IN multimerization (data not shown). However, this could reflect the adverse affects of the high salt conditions required to keep the protein soluble. We note also that enzymatic complementation cannot identify regions of IN that contribute to self-association if they do not include the minimal region necessary for catalytic activity. The analyses reported here do not require overlap with catalytic regions and represent the first evidence that important determinants of self-association reside in the C-terminal region of a retroviral integrase.

Multimerization Detected by Protein Overlay—The protein overlay technique provides an intermediate level of stringency for detection of protein-protein associations. It differs from the first two solution methods in several ways: protein probes are tested for binding to immobilized renatured targets, binding conditions are easily manipulated, and the specificity and sensitivity are high. This method offers an opportunity to investigate binding to partially purified target proteins and permits rapid screening of many potential partners for protein-protein associations. However, unlike the first two methods, it does not allow direct determination of the multimeric state and stoichiometry of the interacting partners. From these overlay experiments, we identified both the catalytic core and C-terminal domains as contributing to multimerization. We also observed that the C-terminal domain can specifically associate with itself. This latter property was confirmed by SEC and chemical cross-linking experiments with an isolated C-terminal fragment. Considered together, results from our analyses indicate that the two self-association domains of ASV IN can act independently, predominantly through interactions between homologous domains in each monomer. However, our data do not rule out cooperativity of the two self-association domains in the native full-length protein, nor can we exclude the possibility of other interactions not identified by these methods that could also contribute to the stability of a multimer.

Since multimerization is required for IN function, inhibitors of self-association may be of potential use in antiviral therapy. The protein overlay method may be particularly suited for the identification of peptide inhibitors that interfere with this property and presumably viral integration. There is precedent for such an inhibitor strategy with the retroviral protease (44, 45).

Relationship between IN Structure, Multimerization, and Function—Several aspects of the domain structure of IN proteins revealed in these studies are consistent with recent information obtained from x-ray crystallographic analysis of the HIV-1 IN (50–212) (16) and ASV IN (52–207) (17) catalytic cores. In both analyses, an extensive interface between two monomers in the unit cell was observed, with a large solvent-excluded surface area. The presence of an extensive dimer interface in both structures is consistent with the contention that the catalytic core of one monomer interacts with the core of another. Despite the similar size of the two core domains, the ASV IN structure lacks the sixth C-terminal helix observed in HIV-1. In the HIV-1 dimer, this helix from one monomer extends out from the structure and interacts across the dimer interface with the analogous helix from the second monomer. It is possible that the added stability conferred by this interaction accounts for the tighter association of the HIV-1 core dimer relative to the ASV core noted above. The topology of both dimers suggests that the C-terminal domains of the full-length proteins would be in close proximity and available for interaction with each other across a multimeric interface.

It is still unclear whether retroviral integrase functions as a dimer or a tetramer. Formation of a tetramer might require interactions across two separate protease interfaces, one for dimerization and a second for the association of two dimers into
a tetramer. Currently, it is not possible to conclude that the ASV IN C-terminal domain is involved in either of the postulated interfaces. However, it is noteworthy that C-terminal truncated proteins, IN(52–207) and IN(1–207), do not form tetramers in chemical cross-linking experiments as do full-length IN, IN(39–286), and IN(201–286) (Figs. 2 and 6). Whether this is due to the absence of the C-terminal domain remains to be investigated.

The topology of their folding places the retroviral integrases in a family of nucleases that includes RNase H, the RuvC resolvase, and the MuA transposase (46). Despite sharing a similar fold, and probably similar reaction chemistries, these diverse nucleases differ in substrate specificity, multimeric structure, and the requirement for coordination of cleavages performed. For example, RNase H is known to act as a monomer, whether as an independent domain or in the context of HIV-1 reverse transcriptase (47), and, correspondingly, its function does not require coordination of multiple cleavages. In contrast, RuvC is known to act as a dimer and performs two DNA cleavages to resolve a Holliday junction, but is not involved in joining of DNA strands (48). It is apparent from comparison of their crystal structures that the RuvC dimeric interface (49) differs from that of the IN core structures. A primary challenge for the future will be to determine how aspects of protein sequence and structure give rise to the specific quaternary interactions that allow each of these proteins to perform their specialized functions. Further study of ASV IN self-association should help to identify the relevant distinguishing features of integrase structure.

Acknowledgments—We are grateful to Richard Katz, Barbara Müller, George Kukolj, and Tom Laue for encouragement and helpful advice during the course of this work. We are also indebted to George Merkel for purification of several of the nonfused proteins used in this work.

REFERENCES

1. Katz, R. A., and Skalka, A. M. (1994) Annu. Rev. Biochem. 63, 133–173
2. Kulkosky, J., and Skalka, A. M. (1994) Pharmacol. Ther. 61, 185–203
3. Bushman, F. D., Fujiwara, T., and Craige, R. (1990) Science 249, 1555–1558
4. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) Cell 63, 87–95
5. Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J., and Skalka, A. M. (1991) Nucleic Acids Res. 19, 851–860
6. Mumm, S. R., and Grandgenett, D. P. (1991) J. Virol. 65, 1160–1167
7. Woerner, A. M., and Marcus, S. C. (1993) Nucleic Acids Res. 21, 3507–3511
8. Peñas-Lutzke, R. A., Vink, C., and Plasterk, R. H. (1994) Nucleic Acids Res. 22, 4125–4131
9. Engelman, A., Hickman, A. B., and Craige, R. (1994) J. Virol. 68, 5911–5917
10. Engelman, A., and Craige, R. (1992) J. Virol. 66, 6361–6369
11. Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) Mol. Cell. Biol. 12, 2331–2338
12. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) Science 255, 763–766
13. Donzella, G. A., Jonsson, C. B., and Rothe, M. J. (1993) J. Virol. 67, 7077–7087
14. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craige, R. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3428–3432
15. Kulkosky, J., Katz, R. A., Merkel, G., and Skalka, A. M. (1995) Virology 206, 448–456
16. Dyda, F., Hickman, A. B., Jenkms, T. M., Engelman, A., Craige, R., and Davies, D. R. (1994) Science 266, 1891–1896
17. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. and Skalka, A. M. (1995) J. Mol. Biol. 253, 333–346
18. Baker, T. A., and Mizuzuki, K. (1992) Genes Dev. 6, 2221–2232
19. Grindley, N. (1993) Science 262, 738–740
20. Hughes, R. E., Rice, P. A., Stetzel, T. A., and Grindley, N. (1993) EMBO J. 12, 1447–1458
21. Segal, A. M., and Nash, H. A. (1993) EMBO J. 12, 4567–4576
22. Rice, P. A., and Stetzel, T. A. (1994) EMBO J. 13, 1514–1524
23. Grandgenett, D. P., Vora, A. C. and Schift, R. D. (1978) Virology 89, 119–132
24. Sherman, P. A., and Fyle, J. A. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5119–5123
25. Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1993) J. Virol. 67, 425–437
26. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M. (1992) J. Biol. Chem. 267, 16037–16040
27. Hickman, A. B., Palmer, I., Engelman, A., Craige, R., and Wingfield, P. (1994) J. Biol. Chem. 269, 29279–29287
28. Murphy, J. S., and Eoff, S. P. (1992) J. Virol. 66, 5092–5095
29. Kukolj, G., and Skalka, A. M. (1995) Genes & Dev. 9, 2556–2567
30. Engelman, A., Bushman, F. D., and Craige, R. (1993) EMBO J. 12, 3269–3275
31. van Gent, D. C., Vink, C., and Oude Goomer, A., and Plasterk, R. H. A. (1993) EMBO J. 12, 3261–3267
32. Kalpana, G. V., and Goff, S. P. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 10923–10927
33. Terry, R., Sadt, D. A., Katzmann, M., Cobi, D., and Laue, T. M., and Skalka, A. M. (1988) J. Virol. 62, 2358–2365
34. Abate, F. M., Broxton, A., Randolph, E. R., Moore, D. D., Seidman J. G., Smith, J. A., and Struhl, K. (1992) Short Protocols in Molecular Biology 2nd Ed, Green Publishing Associates and John Wiley and Sons, New York
35. Schmidt, T. G., and Skerra, A. (1994) J. Chromatogr. A 676, 337–345
36. Broeys, J. (1990) DNA (NY) 8, 759–777
37. Rushmore, T. H., and Pickett, C. B. (1993) J. Biol. Chem. 268, 11475–11478
38. Grabak, Z., and Gergely, J. (1993) Anal. Biochem. 185, 131–135
39. Koole, G. R., Mehravan, M., and Martinson, H. G. (1981) Mol. Cell. Biochem. 34, 3–13
40. Carr, D. W., and Scott, J. D. (1992) Trends Biochem. Sci. 17, 246–249
41. Kiel, N. J., Krek, W., and Lilkens, W. R., DeCaprio, J. A., Aihsenbaum, F., Fuchs, C. S., Chittenden, T., Li Y., Farnham, P. J., Bhan, M. A., Livingston, D. M., and Flemington, E. K. (1992) Cell 70, 351–364
42. Stevens, F. J. (1989) Biophys. J. 55, 1155–1167
43. Jenkins, T. M., Mack, A. B., Dyda, F., Ghirlando, R., Davies, D. R., and Craige, R. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 6057–6061
44. Yang, Z. Y., Poorman, R. A., Magnol, L. L., Heinrikson, R. L., and Keedy, F. J. (1991) J. Biol. Chem. 266, 15951–15954
45. Baba, L. M., Rose, J., and Craik, C. S. (1992) Protein Sci. 1, 1244–1253
46. Yang, W., and Stetzel, T. A. (1995) Structure 3, 131–134
47. Smith, J. S., and Roth, M. J. (1993) J. Virol. 67, 4037–4049
48. West, S. C. (1994) Cell 76, 9–15
49. Arisio, M., Vassylev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994) Cell 76, 1063–1072
Multimerization Determinants Reside in Both the Catalytic Core and C Terminus of Avian Sarcoma Virus Integrase
Mark D. Andrake and Anna Marie Skalka

J. Biol. Chem. 1995, 270:29299-29306.
doi: 10.1074/jbc.270.49.29299

Access the most updated version of this article at http://www.jbc.org/content/270/49/29299

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 48 references, 21 of which can be accessed free at http://www.jbc.org/content/270/49/29299.full.html#ref-list-1