Nucleic Acids Research, 2003, Vol. 31, No. 21 ***e135***
DOI: 10.1093/nar/gng136

Rapid purification of RNA secondary structures

Stacy L. Gelhaus, William R. LaCourse*, Nathan A. Hagan, Gaya K. Amarasinghe and Daniele Fabris

Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Received August 8, 2003; Revised and Accepted September 18, 2003

ABSTRACT

A new method for rapid purification and structural analysis of oligoribonucleotides of 19 and 20 nt is applied to RNA hairpins SL3 and SL2, which are stable secondary structures present on the ψ recognition element of HIV-1. This approach uses ion-pairing reversed-phase liquid chromatography (IP-RPLC) to achieve the separation of the stem–loop from the transcription mix. Evidence is presented that IP-RPLC is sensitive to the different conformers of these secondary structures. The purity of each stem–loop was confirmed by mass spectrometry and PAGE. IP-RPLC purification was found to be superior to PAGE in terms of time, safety and, most importantly, purity.

INTRODUCTION

Highly pure oligonucleotides are of critical importance for *in vitro* studies of nucleic acid–protein interactions. Depending on the desired size and quantity, samples can be extracted from natural sources, synthesized according to phosphoramidite or similar chemistry, or prepared by *in vitro* transcription of DNA templates carried out by the enzyme, T7 polymerase (1–3). The fact that the coupling efficiency of either the chemical or enzymatic procedure is usually <100% ensures the presence of a certain level of sample heterogeneity. Synthetic errors can be caused by decoupling in the synthesizer, adding the wrong base or adding an extra base. Enzymatic synthesis can lead to misincorporation of bases, addition of extra bases at the end of the sequence (most commonly cytosine) and the initiation of synthesis with a mono- or diphosphate nucleotide (4–6). The incidence of errors and the level of heterogeneity are generally increased as a function of the sequence length. For this reason, it is usually necessary to separate the desired sample from the product mixture using a variety of established techniques (7,8). Regardless of the RNA source, accurate structural analysis necessitates the purification of the RNA. Inherent to this process is the repetitive separation of the desired entity from $n-1$ aborts and $n+1$ additions caused by various steps in the synthesis process.

The most common approach to purifying RNA from a transcription mix is based on denaturing polyacrylamide gel electrophoresis (PAGE) followed by electro-elution (8). Unfortunately, this technique comes at a high price in terms of time, labor, cost and toxicity. The initial set-up of the gels is labor intensive, requires precautions to avoid introducing RNase activity into the samples and may lead to toxic exposure to polyacrylamide. The separation itself may take hours and the final resolution may not be adequate. The separated bands are usually excised by hand, increasing the risks of error and contamination. Multiple repetitions of the process may become necessary. In the long run, PAGE can become an expensive method. This is especially true for structural work, which requires large quantities.

An alternative method for the purification of RNA is ion-pairing reversed-phase high-performance liquid chromatography (IP-RPLC). Despite the fact that IP-RPLC is not thought of as one of the standard molecular biology tools, the use of IP-RPLC for nucleotide analysis has steadily increased over the last several decades. IP-RPLC has previously been used for deoxyoligonucleotide purification, analysis of short tandem repeats, small nucleotide polymorphisms and RNA (9,10). Recently, a paper was published showing that footprinting of RNA is also possible by IP-RPLC (11).

Here we report for the first time the differentiation of stem–loop conformations using IP-RPLC. The purification of the stem–loops (i.e. SL3 and SL2, see Fig. 1) located in the ψ-recognition element, necessary for the retroviral packaging of HIV-1 (12–15), is used to highlight this separation effect. For purification, the various stable conformers of the RNA secondary structure are collected as a single fraction and are separated from all components of the transcription mix and abortive sequences. Denaturing IP-RPLC, mass spectrometry and PAGE were used to confirm that the 'packet' of peaks collected in this method all had the same sequence (9).

MATERIALS AND METHODS

Reagents

All reagents and solvents were HPLC grade (Fisher Scientific, Springfield, NJ). Triethylammonium acetate (TEAA) was obtained as a 20× concentrate (Transgenomic, San Jose, CA) and diluted with either water and/or acetonitrile (ACN). Water was purified using a reverse osmosis system coupled with a

*To whom correspondence should be addressed. Tel: +1 410 455 2105; Fax: +1 410 455 2608; Email: lacourse@umbc.edu
Present address:
Gaya Amarasinghe, Department of Biochemistry, University of Texas, Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA

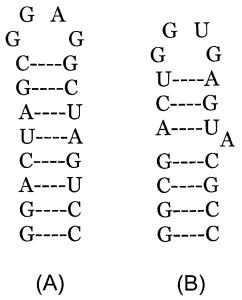

Figure 1. Sequence and structure of SL2 (**A**) and SL3 (**B**).

multi-tank/ultraviolet/ultrafiltration station (US Filter/IONPURE, Lowell, MA). The SL2 and SL3 stem–loops were prepared as in Milligan and Uhlenbeck (3) and used without further treatment. Samples were stored in a freezer at –20°C or lyophilized and then frozen for long-term storage.

Liquid chromatography

Analytical separations were carried out on the WAVE Nucleic Acid Analysis System (Transgenomic) using Hitachi System Manager (HSM) software. The HSM program is interfaced with the pump (model L7100), autosampler (model L7200), oven (model L7300) and UV detector (model L7400) through the D-700 interface module. A semi-preparative system consisted of the same elements except it was plumbed to accept larger flow rates. SL2 and SL3 were purified on the analytical and semi-preparative systems, respectively.

The column (part no. NUC-99-3860) used was a non-porous, C-18 modified polystyrene-divinylbenzene (PS-DVB), semi-preparative column (6.0 × 79 mm; 3 µM particles) from Transgenomic. The oven temperature was set at 30°C. A two-solvent gradient was used with a mobile phase consisting of an ion-pairing agent and organic modifier. Solvent A consisted of 0.1 M TEAA and solvent B consisted of 0.1 M TEAA and 25% ACN. The solvents were made using RNA-grade water. The flow rates were 0.75 and 1.2 ml·min^{-1} for the analytical and semi-preparative systems, respectively. The gradient used for the stem–loop purification is shown in Table 1. Injection volumes were 80 and 500 µl for the analytical and semi-preparative systems, respectively. Several water injections were run before injecting the RNA in order to wash the system through with the RNA-grade solvents. Also, the samples were pipetted into autosampler tubes after wiping down the pipette and tips with RNaseZap from Sigma®. Collection was done at 260 nm on the analytical system, and off the maximum absorbance wavelength, at 300 nm, on the semi-preparative system. This was done to prevent saturation of the detector.

Separation fractions were collected using a model FCW-180 Fraction Collector (Transgenomic). The fraction collector was set to collect the main stem–loop peaks by setting a specific time window and threshold value. Collection would not take place unless the peaks exceeded the threshold value within that specified time window. New Falcon® tubes were used for collecting the RNA product.

Mass spectrometry

Mass spectrometry (MS) analyses were performed on a JEOL (Tokyo, Japan) HX110/HX110 four-sector mass spectrometer equipped with an Analytica of Brandford (Brandford, CT) thermally assisted electrospray ionization (ESI) source. Lyophilized samples from long-term storage were re-dissolved in 10 mM ammonium acetate (pH adjusted to 7.0) to a final concentration of ~10 µM. Each determination was carried out by injecting 10 µl aliquots through a loop injector at a constant flow rate of 1 µl/min. Negative-ion mode spectra were the averaged profile of up to 20 scans with a duty cycle of ~20 s. Resolution was set to 500 by adjusting the slit width, accuracy was determined to be ~410 p.p.m. or better.

PAGE

RNA samples before and after IP-RPLC purification were analyzed on an analytical 20% denaturing PAGE. Gels were prepared according to the standard protocol (8) and run with 1× Tris-borate–EDTA buffer at 200 V for 60 min. Samples were brought up in 50% glycerol before loading. Staining was carried out with Stainzall® for 20 min and the gel was destained overnight.

RESULTS

Initial separation conditions were selected via the Wavemaker software®. Since the software was originally developed for DNA, thymine was replaced by uracil when entering the

Table 1. Gradient program for the isolation of SL2 and SL3 in IP-RPLC

SL2 (flow rate: 0.75 ml·min^{-1})			SL3 (flow rate: 1.20 ml·min^{-1})		
Time (min)	Solvent A (%)	Solvent B (%)	Time (min)	Solvent A (%)	Solvent B (%)
0.0	90	10	0.0	90	10
5.0	80	20	5.0	90	10
8.0	75	25	8.0	80	20
10.0	70	30	10.0	60	40
12.0	62	38	12.0	42	58
15.0	60	40	14.0	40	60
16.0	53	47	16.0	10	90
18.0	50	50	18.0	0	100
20.0	0	100	18.5	0	100
22.0	0	100	19.0	90	10
22.1	90	10	25.0	90	10
24.0	90	10			

sequence information. The effect of this base exchange on the initial separation parameters was determined to be minimal because the structures of the two bases are sufficiently similar. Under these conditions, each of the stem–loops eluted earlier than predicted for a linear sequence of the same number of bases. The conditions for separation were further optimized from Wavemaker® to increase resolution as well as equilibration time for each injection. Table 1 gives the optimized gradient programs used for the SL2 and SL3 separations. The temperature was set at 30°C to more closely match the conditions used in the nucleic acid–protein interaction studies and to maximize the efficiency of the separation of the stem–loops from aborts and other impurities.

Figure 2 shows the chromatograms of the crude transcription mix for SL3 (Fig. 2A) and SL2 (Fig. 2B) using the conditions outlined above. Typical reaction mixtures include a DNA template (single stranded), a primer (17 nt, single stranded), the triphosphate ribonucleotides (NTPs), the enzyme, T7 polymerase, magnesium, Tris buffer and a mixture of RNA products, including the target, a number of abortive sequences, and possibly the target plus an additional base. In each chromatogram, peak I is the solvent front, which is mainly comprised of the unused NTPs and T7 polymerase. The group II peaks are comprised of aborts of the desired sequence. Some of the aborts, which do not form a secondary structure, elute later than the stem–loops even though their sequence may be shorter (see Fig. 2B). Both SL3 (III) and SL2 (III) eluted as packets of four main peaks and three main peaks, respectively. Chromatograms of the purified SL3 (Fig. 3A) and SL2 (Fig. 3B) show only the major RNA species of interest. Since the ratio of individual peaks of the different conformers is concentration dependent, the purified SL2 (Fig. 3B) appears to elute as a single peak due to an over 10-fold decrease in its injected concentration. When the peaks of the major species (injected at higher concentrations) were collected and re-injected individually the chromatograms showed the same profiles of multiple peaks. The return of one peak to the same profile of multiple peaks is an indication of the presence of conformer equilibrium. As mentioned previously, the retention time for these species was shorter than the calculated retention time for a single-stranded species with the same number of nucleotides. Previous studies have determined that the mechanism of IP-RPLC separation is dependent upon not only the number of charges, but also on charge accessibility (16,17). As a consequence, IP-RPLC is sensitive to the shape corresponding to the hairpin conformation. It is plausible that the difference in the predicted retention time between the stem–loops and their hypothetical linear analogs is due to restricted access of the charges of the stem–loops.

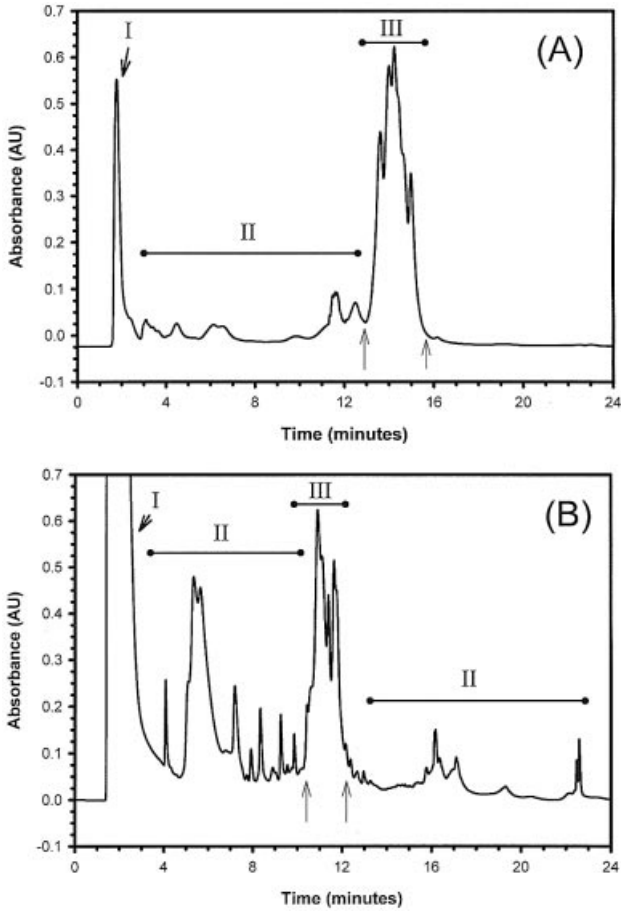

Figure 2. Chromatograms of the crude transcription mix for: (**A**) SL3 showing the solvent front (I), abort sequences (II) and main product peaks (III); (**B**) SL2 showing the solvent front (I), abort sequences (II) and main product peaks (III). The vertical dashed line indicates where the cuts were taken for purification.

DISCUSSION

The presence of multiple peaks, which may represent the equilibrium of conformers, can be explained by the stem–loops' ability to maintain a certain degree of secondary structure during the separation. In agreement with IP-RPLC theory, retention times decreased for the 'packet' of peaks, which is indicative of stable secondary structure over the range of organic modifier concentrations used. If an increase in organic modifier had denatured the secondary structure, phase, the retention would have increased. Further and more definitively, varying the temperature from 35 to 50°C in 2°C increments, starting at 36°C under the on-line oven control, made the three peaks of SL2 coalesce into a single peak (Fig. 4). This melting curve was done under initial Wavemaker® gradient conditions before the gradient was further optimized. The fact that the melting temperature of SL2 calculated by Wavemaker software® is 58°C suggests that the peaks coalescence may be due to the opening of the stem's double-stranded structure. The presence of multiple conformers is also consistent with the observation that the ion-pairing reagent is a singly charged monoamine, which may not offer the same secondary structure stabilization produced by coordination of the divalent cation Mg^{2+}.

By comparison, the pooled fractions containing multiple peaks in the SL2 chromatography provided a single band on a 20% denaturing PAGE, which showed the same migration time as the SL2 marker (Fig. 5). Similarly, the 'packet' of peaks, designated as the major stem–loop product, purified by IP-RPLC and by the established PAGE-base protocol were analyzed by ESI-MS (Fig. 6A, PAGE and B, IP-RPLC). In

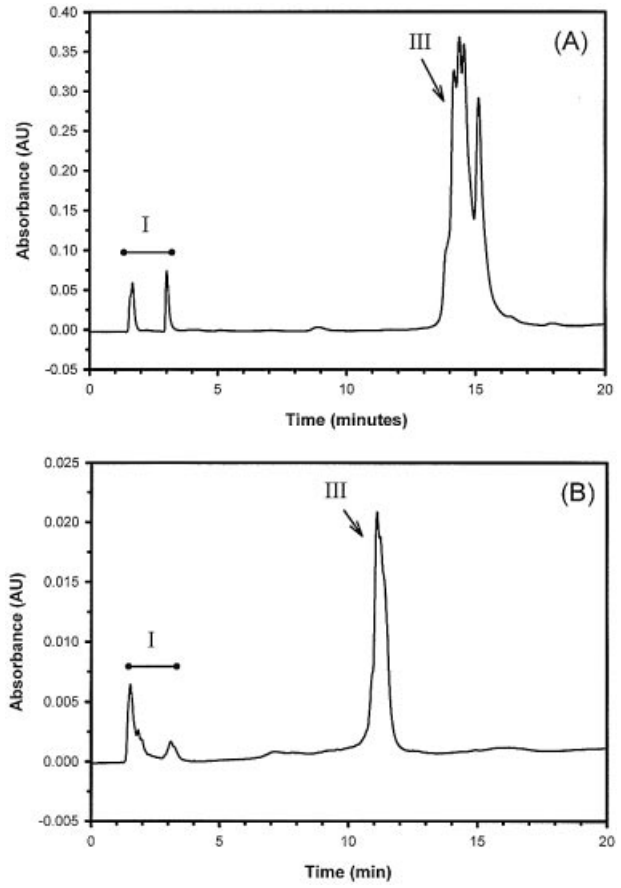

Figure 3. Chromatograms of the purified (**A**) SL3 and (**B**) SL2. The grouping of peaks (III) indicates the main product that was collected from the crude transcription mix of each template synthesis.

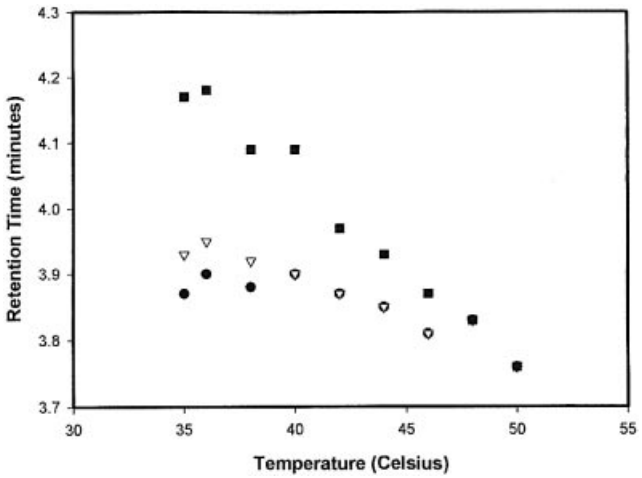

Figure 4. Plot of peak retention verses column temperature for SL2. Each of the three main peaks comprising the major species collected for SL2 is identified by a different symbol, either a square, triangle or a circle. As the temperature is increased, the three main peaks coalesce into one.

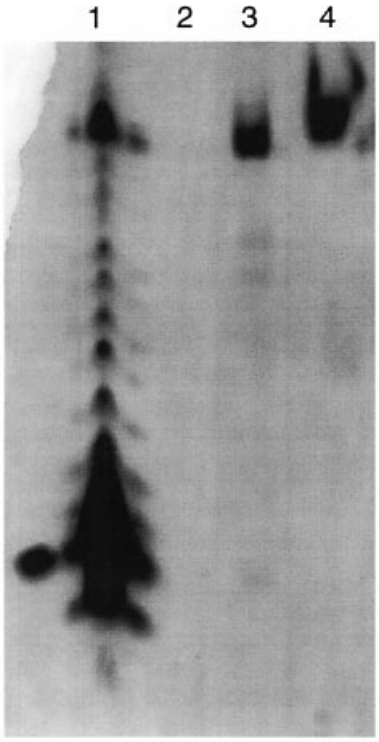

Figure 5. PAGE verification of IP-RPLC-purified SL2. Crude transcription mix (lane 1), IP-RPLC-collected solvent front (lane 2), IP-RPLC-purified SL2 (lane 3) and an SL2 molecular weight marker (lane 4).

provides an experimental mass of 6374 ± 2 Da for the PAGE sample and 6371 ± 2 Da for the IP-RPLC sample (18). Both determinations were in close correlation with the molecular weight calculated from the sequence (6371.6 Da). Both spectra indicated the presence of SL2-PO_3^{3-}, which can be attributed to a small percentage of diphosphate (NDP) in the NTPs used in the synthesis mixture that has been incorporated by the T7 polymerase (6). The presence of an NDP or an NTP at the 5′ of the construct does not change the overall charging of the analyte and thus is inconsequential for the separation. It should also be noted that the PAGE-purified sample revealed the presence of a minor species, which was absent in the IP-RPLC sample, and corresponds to the extra-template addition of cytosine (an additional experimental mass of 304 Da) by the T7 polymerase during the synthesis of the oligonucleotide (5). It is clear that the resolution afforded by PAGE was insufficient to remove these extended products from the reaction mixtures, while IP-RPLC showed clearly superior results. Similar results were obtained for the SL3 construct (data not shown). Hence, analysis by IP-RPLC under denaturing conditions, mass spectrometry and PAGE all showed that the packet of peaks representing individual conformers of a particular stem–loop consists of a single

Other considerations (Table 2) should be made concerning the advantages of IP-RPLC over the PAGE method. While the latter involves long electrophoresis and electro-elution times (hours time scale), the IP-RPLC method allows one to obtain suitable separation and collection in a much shorter time frame

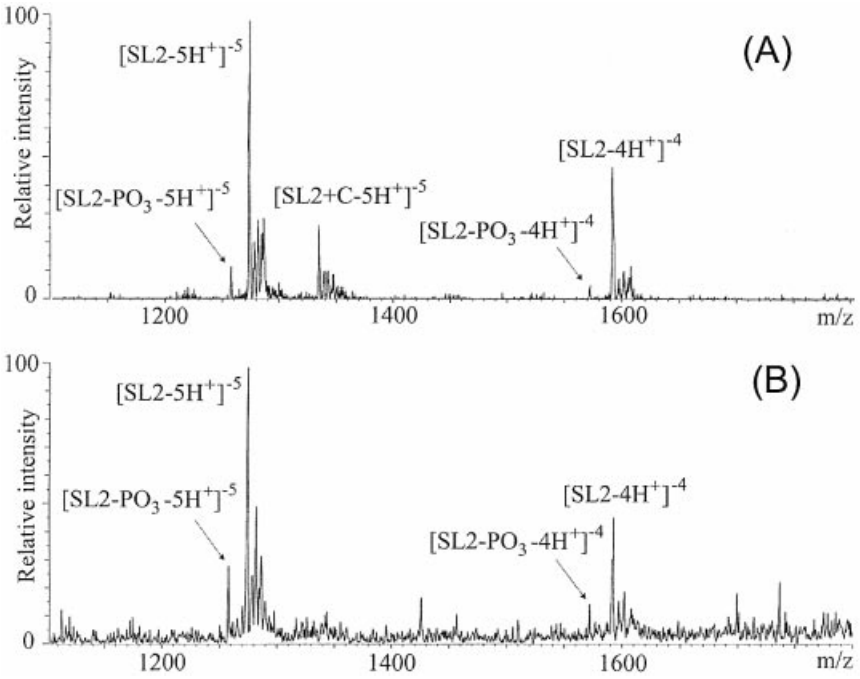

Figure 6. ESI-MS spectra of (**A**) PAGE-purified and (**B**) IP-RPLC-purified SL2. The PAGE purification shows the presence of SL2 + 1 cytosine.

Table 2. Comparison of purification methods

Item	SDS–PAGE	Analytical IP-RPLC	Semi-preparative IP-RPLC
Steps in purification	4	2	2
Overall time	2 days	5.5 h	30 min
Maximum sample size	2 ml	2 mg/80 µl	2 mg/1.8 ml
Cost per run	$$	$	$

of resolution due to the frequent practice of overloading. IP-RPLC offers the same resolution obtained in the analytical scale for the preparative scale up to 2 mg. In addition, there is no manual handling of the RNA samples, as no gel excision is required and injections can be automated. This results in a much lower probability of RNase contamination. Finally, the initial cost of the gel apparatus is cheaper than the HPLC system; however, cost per run for the chromatography system is less costly than PAGE, making IP-RPLC more cost efficient in the long run.

The IP-RPLC method has proven to be a robust, convenient and efficient alternative to PAGE for the purification of RNA products synthesized *in vitro* by T7 polymerase transcription. IP-RPLC was shown to be superior to PAGE in several areas. By cutting the purification process down from several days to several hours, more RNA can be processed and more experimentation done with the possibility of simultaneously cutting the costs of each synthesis. Single-base resolution can be maintained for analytical and semi-preparative scale necessary to purify the large quantities of products required for structural and biochemical investigations.

This technique does not preclude its application to isolation of secondary structures synthesized by various methods. It was shown that the separation mechanism is sensitive to the analyte conformation of the stem–loops chosen. It is this characteristic of conformer sensitivity that is exploited by the chromatography for isolation of the major species from the surrounding matrix. This feature could pave the way for intriguing biophysical applications to other systems with secondary structure and conformer equilibrium where kinetics allows.

ACKNOWLEDGEMENTS

We would like to thank Dr Michael Summers, HHMI, for donating the stem–loop, SL2, for the project. We are deeply grateful for the financial support from Transgenomic Inc., San Jose, CA. The SL3 synthesis and MS analysis was supported

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.