REVIEW

Thioesterases: A new perspective based on their primary and tertiary structures

David C. Cantu, Yingfei Chen, and Peter J. Reilly*

Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011

Received 19 April 2010; Accepted 7 May 2010
DOI: 10.1002/pro.417
Published online 17 May 2010 proteinscience.org

Abstract: Thioesterases (TEs) are classified into EC 3.1.2.1 through EC 3.1.2.27 based on their activities on different substrates, with many remaining unclassified (EC 3.1.2.–). Analysis of primary and tertiary structures of known TEs casts a new light on this enzyme group. We used strong primary sequence conservation based on experimentally proved proteins as the main criterion, followed by verification with tertiary structure superpositions, mechanisms, and catalytic residue positions, to accurately define TE families. At present, TEs fall into 23 families almost completely unrelated to each other by primary structure. It is assumed that all members of the same family have essentially the same tertiary structure; however, TEs in different families can have markedly different folds and mechanisms. Conversely, the latter sometimes have very similar tertiary structures and catalytic mechanisms despite being only slightly or not at all related by primary structure, indicating that they have common distant ancestors and can be grouped into clans. At present, four clans encompass 12 TE families. The new constantly updated ThYme (Thioester-active enzYmes) database contains TE primary and tertiary structures, classified into families and clans that are different from those currently found in the literature or in other databases. We review all types of TEs, including those cleaving CoA, ACP, glutathione, and other protein molecules, and we discuss their structures, functions, and mechanisms.

Keywords: clan; primary structure; protein family; tertiary structure; thioesterases; ThYme

Introduction

The thioesterases (TEs), or thioester hydrolases, comprise a large enzyme group whose members hydrolyze the thioester bond between a carbonyl group and a sulfur atom. They are classified by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) into EC (enzyme commission) 3.1.2.1 to EC 3.1.2.27, as well as EC 3.1.2.– for unclassified TEs.[1] Substrates of 15 of these 27 groupings contain coenzyme A (CoA), two contain acyl carrier proteins (ACPs), four have glutathione or its derivatives, one has ubiquitin, and two contain other moieties. In addition, three groupings have been deleted.

The EC classification system is based on enzyme function and substrate identity, and it was first formulated when very few amino acid sequences (primary structures) and three-dimensional (tertiary)

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: U.S. National Science Foundation; Grant number: EEC-0813570.

*Correspondence to: Peter J. Reilly, Department of Chemical and Biological Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011-2230. E-mail: reilly@iastate.edu.

structures of enzymes were available. Another way to classify enzymes is by primary structure into families and by tertiary structure into clans or superfamilies. Some databases are built this way: Pfam[2] has a collection of protein families and domains, and SCOP[3] classifies protein structures into classes, folds, families, and superfamilies. Other databases treat certain enzyme groups more specifically. For instance, MEROPS[4] is a major database for peptidases, and CAZy[5] covers carbohydrate-active enzymes.

It is common to observe that members of more than one EC grouping are found in one enzyme family based on similar amino acid sequences, implying that they have a common ancestor, mechanism, and tertiary structure. Conversely, members of a single EC grouping may be located in more than one enzyme family, being totally or almost totally unrelated in primary structure and potentially in mechanism and tertiary structure.

A further observation is that members of two different enzyme families may have very similar tertiary structures and mechanisms even though their primary structures are very different. This may imply that they are members of the same clan or superfamily, descended from a more distant common ancestor.

In this work, TE primary and tertiary structures will be analyzed to conclude how TEs are divided (and united) into families and clans. Structures, mechanisms, and catalytic residues are compared between families and clans. We compare our findings with existing databases such as Pfam and SCOP. Results also appear in a new continuously updated database, ThYme (Thioester-active enzYmes, http://www.enzyme.cbirc.iastate.edu) that includes families and clans of enzyme groups that are part of the fatty acid synthesis cycle, TEs among them.

Family identification

Family members must have strong (>15%, but typically >30%) sequence similarity and near-identical tertiary structures, and they must share general mechanisms as well as catalytic residues located in the same position.

In general, TE families were identified in the following way: (1) experimentally confirmed TE sequences were used as queries, (2) a series of successive Basic Local Alignment Search Tool (BLAST)[6] searches and comparison among results reduced query sequences to a few representative ones, (3) the catalytic domains of representative query sequences were subjected to BLAST to populate the families, (4) experimentally confirmed TEs were surveyed to search for missing potential TE families, and (5) the uniqueness of the families was confirmed by multiple sequence alignments (MSAs), by tertiary structure superposition and comparison, and by catalytic residue positions. Methods are detailed in Supporting Information.

Clan identification

Two or more families are grouped into a clan if all the sequences within them show some (<15%) sequence similarity, if their structures are strongly similar (narrowing the search to families with the same fold), and if they share similar active sites and general mechanisms. To consider all aspects of clan classification criteria, several methods are used to combine sequence and structural analysis. In addition, catalytic mechanisms of members of each family were gathered from the literature, and positions of catalytic residues were determined to verify that they coincided. A more detailed description of these methods is found in the Supporting Information.

ThYme database

All the sequences in each family are displayed on the ThYme database website (http://www.enzyme.cbirc.iastate.edu). These sequences are taken, using a series of scripts, from the BLAST results of the catalytic domains of the representative query sequences. Matching accessions, taxonomical data, protein names, and EC numbers are taken from UniProt[7] and GenBank[8] databases. Each TE family is shown on a page where sequences are arranged into archaea, bacteria, and eukaryota, then alphabetically by species. In each row, a single sequence or group of sequences with 100% identical catalytic domains are shown with their protein name and UniProt and/or GenBank accession codes. EC numbers are shown only when they appear in a sequence's UniProt or GenBank annotation. If a crystal structure is known, the Protein Data Bank (PDB, http://www.rcsb.org) accession code also appears. ThYme will be continuously updated: the content of each family will grow as GenBank, UniProt, and PDB do. However, to create a new family, or to merge or delete existing ones, human judgment and manual changes will be necessary.

Thioesterase families

Use of BLAST with TE query sequences followed by construction of MSAs and superposition of tertiary structures yielded 23 families almost completely unrelated by primary structure (Table I).

Enzymes in families TE1–TE13 hydrolyze substrates with various acyl moieties and CoA, those in TE14–TE19 attack bonds between acyl groups and ACP, and those in TE20 and TE21 cleave the bonds between acyl groups and proteins. Members of TE22 and TE23 break bonds between acyl groups and glutathione and its derivatives (Table II). The sulfur-carrying moiety in CoA and ACP is a pantethiene residue, whereas glutathione itself carries

Table I. *Thioesterase Families and Common Names of their Members*

Family	Producing organisms	Genes and/or other names of family members
TE1	A, **B**, E[a]	Ach1
TE2	A, B, **E**	Acot1–Acot6, BAAT thioesterase
TE3	A, **B**	*tesA*, acyl-CoA thioesterase I, protease I, lysophospholipase L1
TE4	**B**, E	*tesB*, acyl-CoA thioesterase II, Acot8
TE5	**B**	*tesC* (*ybaW*), acyl-CoA thioesterase III
TE6	A, **B**, E	Acot7 (BACH), Acot11 (BFIT, Them1), Acot12 (CACH), YciA
TE7	B, **E**	Acot9, Acot10
TE8	A, B, **E**	Acot13 (Them2)
TE9	**B**	YbgC
TE10	**B**	4HBT-I
TE11	**B**	4HBT-II, EntH (YbdB)
TE12	**B**,E	DNHA-CoA hydrolase
TE13	A, **B**	paaI, paaD
TE14	**B**, E	FatA, FatB
TE15	**B**	Thioesterase CalE7
TE16	A, **B**, E	TE domain of FAS (Thioesterase I), TE domain of PKS or NRP (type I thioesterase (TE I))
TE17	**B**	TE domain of PKS
TE18	**B**,E	Thioesterase II, type II thioesterase (TE II)
TE19	**B**	luxD
TE20	**E**	ppt1, ppt2, palmitoyl-protein thioesterase
TE21	A, B, **E**	apt1, apt2, acyl-protein thioesterase, phospholipase, carboxylesterase
TE22	A, **B**, E	*S*-formylglutathione hydrolase, esterase A, esterase D
TE23	A, **B,** E	Hydroxyglutathione hydrolase, glyoxalase II

[a] A, archaea; B, bacteria; E, eukaryota. Most prevalent producers bolded.

the sulfur moiety, and in non-ACP proteins, the sulfur-carrying moiety is built up mainly from a cysteine residue.

All tertiary structures within each family have almost identical cores and very strong overall resemblance (Table III) shown by $RMSD_{ave}$ values of <1.8

Table II. *Thioesterase Functions and Substrate Specificities*

Family	General function	EC number	Preferred substrate specificity (if known)
TE1	Acyl-CoA hydrolase	3.1.2.1, 2.8.3.–	Acetyl-CoA
TE2	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.2, 2.3.1.65	Palmitoyl-CoA, bile-acid-CoA
TE3	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.20, 3.1.1.2, 3.1.1.5	Medium- to long-chain acyl-CoA
TE4	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.2, 3.1.2.27	Short- to long-chain acyl-CoA, palmitoyl-CoA, choloyl-CoA
TE5	Acyl-CoA hydrolase	3.1.2.–	Long-chain acyl-CoA, 3,5-tetradecadienoyl-CoA
TE6	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.1, 3.1.2.2, 3.1.2.18, 3.1.2.19, 3.1.2.20	Short- to long-chain acyl-CoA, C_4–C_{18}
TE7	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.1, 3.1.2.2, 3.1.2.20	Short- to long-chain acyl-CoA
TE8	Acyl-CoA hydrolase	3.1.2.–	Short- to long-chain acyl-CoA, C_6–C_{18}
TE9	Acyl-CoA hydrolase	3.1.2.–, 3.1.2.18	Short- to long-chain acyl-CoA, 4-hydroxybenzoyl-CoA
TE10	Acyl-CoA hydrolase	3.1.2.23	4-Hydroxybenzoyl-CoA
TE11	Acyl-CoA hydrolase	3.1.2.–	4-Hydroxybenzoyl-CoA
TE12	Acyl-CoA hydrolase	3.1.2.–	1,4-Dihydroxy-2-napthoyl-CoA
TE13	Acyl-CoA hydrolase	3.1.2.–	Short and medium-chain acyl-CoA, several hydroxyphenylacetyl-CoA substrates
TE14	Acyl-ACP hydrolase	3.1.2.–, 3.1.2.14	Short- to long-chain acyl-ACP, C_8–C_{18}
TE15	Acyl-ACP hydrolase	—	—
TE16	Acyl-ACP hydrolase	3.1.2.14[a]	Long-chain acyl-ACP, various polyketides and non-ribosomal peptides
TE17	Acyl-ACP hydrolase	3.1.2.14[b]	Several polyketides
TE18	Acyl-ACP hydrolase	3.1.2.–, 3.1.2.14	Medium-chain acyl-ACP, various polyketides and nonribosomal peptides
TE19	Acyl-ACP hydrolase	2.3.1.–	Myristoyl-ACP
TE20	Protein-palmitoyl hydrolase	3.1.2.–, 3.1.2.22	Palmitoyl-protein
TE21	Protein-acyl hydrolase	3.1.2.–, 3.1.1.1	—
TE22	Glutathione hydrolase	3.1.2.12, 3.1.1.1, 3.1.1.6	*S*-Formylglutathione
TE23	Glutathione hydrolase	3.1.2.6	D-Lactoylglutathione

[a] TE domain. FASs, PKSs, and NRPs can have several EC numbers such as 2.3.1.85, 2.3.1.94, 2.3.1.–, 2.7.7.–, and 5.1.1.–.
[b] TE domain of PKSs.

Table III. *Thioesterase Folds*

Family	Fold	$RMSD_{ave}$ (Å)	P_{ave} (%)	PDB files
TE1	NagB	1.25	96.4	2G39, 2NVV
TE2	α/β-Hydrolase	1.00	96.6	3HLK, 3K2I
TE3	Flavodoxin-like	0.58	96.6	1IVN, 1J00, 1JRL, 1U8U, 1V2G, 3HP4
TE4	HotDog	0.90	33.3	1C8U, 1TBU
TE5	HotDog	—	—	1NJK
TE6	HotDog	1.39	75.9	3B7K, 2Q2B, 2V1O, 2QQ2, 1YLI, 3BJK, 3D6L
TE7	HotDog	—	—	
TE8	HotDog	0.58	88.3	2H4U, 3F5O, 2F0X, 2CY9
TE9	HotDog	1.19	88.8	2PZH, 1S5U, 3HM0, 1Z54
TE10	HotDog	0.67	97.1	1BVQ, 1LO7, 1LO8, 1LO9
TE11	HotDog	0.87	93.9	1Q4S, 1Q4T, 1Q4U, 1VH9, 2B6E, 1SC0, 3LZ7
TE12	HotDog	—	—	2VEU
TE13	HotDog	0.43	94.6	2FS2, 1PSU, 2DSL0, 1J1Y, 1WLU, 1WLV, 1WM6, 1WN3
TE14	HotDog	1.65	87.7	2OWN, 2ESS
TE15	HotDog	—	—	2W3X
TE16	α/β-Hydrolase	1.51	66.9	2VZ8,[a] 2VZ9,[a] 2PX6, 1XKT, 2ROQ,[b] 2CB9, 2CBG, 2VSQ, 1JMK
TE17	α/β-Hydrolase	1.67	82.4	1MO2, 1KEZ, 1MN6, 2H7X, 2H7Y, 2HFK, 2HFJ, 1MNA, 1MNQ
TE18	α/β-Hydrolase	0.83	97.2	3FLA, 3FLB, 2RON,[b] 2K2Q[b]
TE19	α/β-Hydrolase	—	—	1THT
TE20	α/β-Hydrolase	1.41	91.2	1EH5, 3GRO, 1EI9, 1EXW, 1PJA
TE21	α/β-Hydrolase	0.82	96.7	1FJ2, 1AUO, 1AUR, 3CN7, 3CN9
TE22	α/β-Hydrolase	1.69	78.9	3FCX, 3C6B, 2UZ0, 1PV1, 3I6Y, 3E4D, 3LS2
TE23	Lactamase	1.67	78.5	2QED, 1XM8, 2P18, 2GCU, 2Q42, 1QH3, 1QH5, 2P1E

[a]2VZ8 and 2VZ9 have TE domains in their FASTA format. Therefore, these were picked up by BLAST, but their PDB files do not include the TE domain, and they were not included in the RMSD calculation.
[b]NMR-resolved structures not included in RMSD calculation.

Å and P_{ave} values of >75% (see Supporting Information for definitions), with two exceptions. TE4 has a P_{ave} value of 33.3% because it has only two crystal structures, of which one monomer (1C8U) is a double HotDog, whereas another monomer (1TBU) is incomplete with only a single HotDog. Similarly, in TE16 the P_{ave} value is 65.8% because the TE domain of one structure (2VSQ) is smaller than the rest.

Of the families whose members hydrolyze acyl-CoAs, all have HotDog[9,10] folds (Table III, Figs. 1 and 2) except for TE1, TE2, and TE3. TE1 enzymes have NagB folds, and they have acetyl-CoA hydrolase (EC 3.1.2.1) activity as well as acetate or succinate-CoA transferase (EC 2.8.3.–) activity. They are found mainly in bacteria and fungi, although they are also present in archaea. Enzymes coded by the acetyl-CoA hydrolase *ACH1* gene from *Saccharomyces cerevisiae* are present in TE1.[11] Fungal enzymes in this family are involved with acetate levels and CoA transfer in mitochondria.[12]

TE2 enzymes have α/β-hydrolase[13] folds (Figs. 3 and 4). They are mainly found in eukaryotes (animals), but they are also present in bacteria. They have mostly palmitoyl (EC 3.1.2.2) and bile acid-CoA:amino acid *N*-acyl transferase (BAT) (EC 2.3.1.65) activities. The acyl-CoA TE (Acot) enzymes ACOT1, ACOT2, ACOT4, and ACOT6 from *Homo sapiens* are present in this family, as well as the Acot1 through Acot6 enzymes from *Mus musculus*, *Rattus norvegicus*, and similar species.[14] Also in TE2 are the BAAT TEs that transfer bile acid from bile acid-CoA to amino acids in the liver; these conjugates later solvate fatty acids in the gastrointestinal tract.[15]

Enzymes in TE3 are part of the SGNH hydrolase superfamily with a flavodoxin-like fold. They are mainly found in bacteria and have acyl-CoA hydrolase (EC 3.1.2.20), arylesterase (EC 3.1.1.2), and lysophospholipase (EC 3.1.1.5) activities. Some TE3 enzymes come from the *tesA* gene, and they are located in the periplasm and are involved in fatty acid synthesis.[16] TE3 enzymes are also called acyl-CoA thioesterase I, protease I, and lysophospholipase L_1, and the genes that code for them, *tesA*, *apeA*, and *pldC*, respectively, are nearly identical.[17]

The rest of the acyl-CoA hydrolase families have HotDog folds. TE4 enzymes, present in bacteria and eukaryotes, are acyl-CoA hydrolases as well as palmitoyl-CoA (EC 3.1.2.2) and choloyl-CoA (EC 3.1.2.27) hydrolases. The Acot8 gene encodes for peroxisomal TEs,[18] which are found in TE4. Also in this family are acyl-CoA thioesterase II enzymes, encoded by the *tesB* gene, that can hydrolyze a broad range of medium- to long-chain acyl-CoA thioesters, but whose physiological function is not known.[19]

TE5 acyl-CoA enzymes, also known as thioesterase IIIs, are present in bacteria. They are encoded by the *tesC* (or *ybaW*) gene and are long-chain acyl-CoA TEs preferring 3,5-tetradecadienoyl-CoA as a substrate.[20]

TE6 members, present in eukaryotes, bacteria, and archaea, have acyl-CoA hydrolase activities

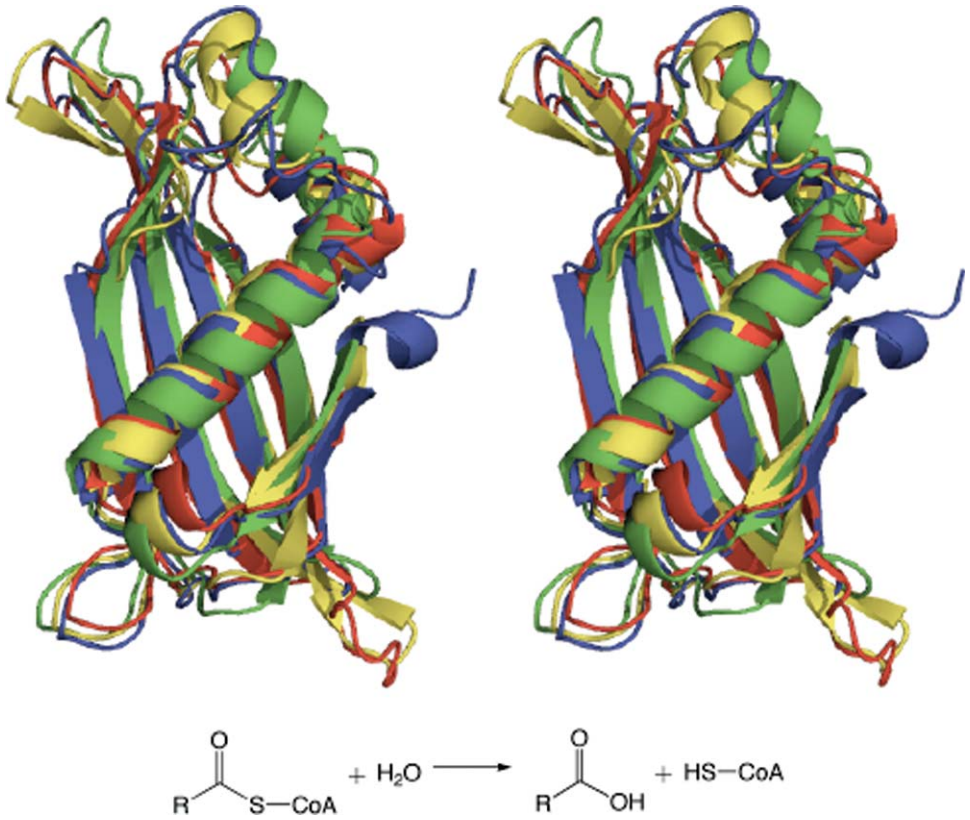

Figure 1. Superimposed tertiary structures of single representatives of each TE family in a clan: TE-A acyl-CoA hydrolases from *Escherichia coli* (TE5) (green), *Helicobacter pylori* (TE9) (red), *Pseudomonas sp*. (TE10) (yellow), and *Prochlorococcus marinus* (TE12) (blue).

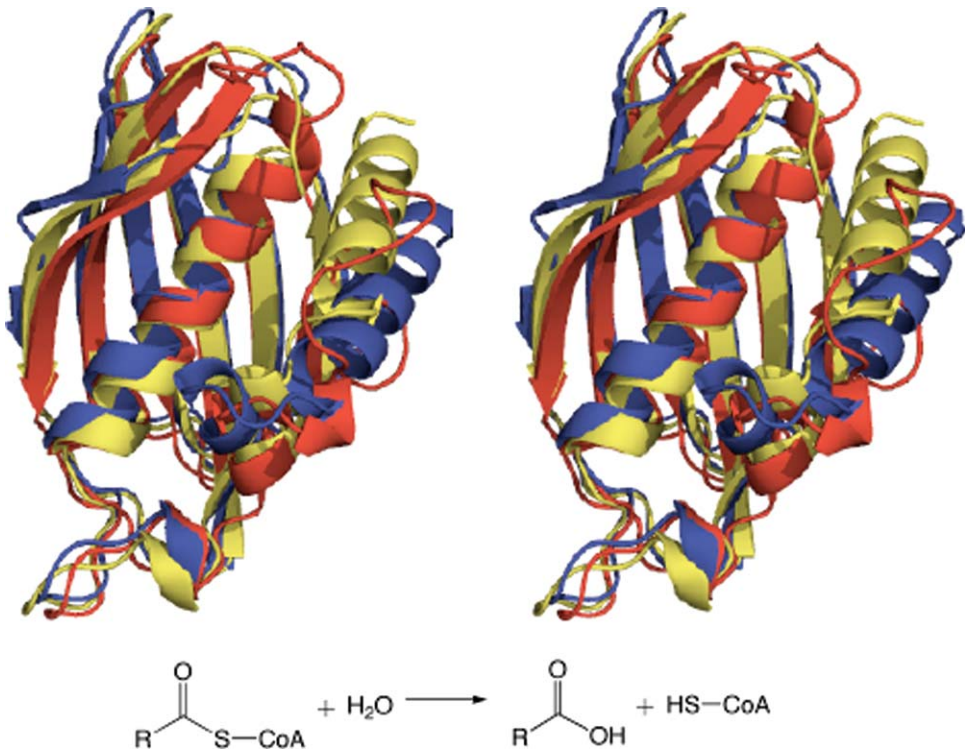

Figure 2. Superimposed tertiary structures of single representatives of each TE family in a clan: TE-B acyl-CoA hydrolases from *Homo sapiens* (TE8) (blue), *Arthrobacter* sp. (TE11) (red), and *E. coli* (TE13) (yellow).

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.