11-25-2020

Proanthocyanidin Block Arrays (PACBAR) for Comprehensive Capture and Delineation of Proanthocyanidin Structures

Shuxi Jing
University of Illinois at Chicago

Wayne E. Zeller
USDA, Agricultural Research Service

Daneel Ferreira
University of Mississippi

Bin Zhou
Chinese Academy of Sciences

Joo-Won Nam
University of Illinois at Chicago

See next page for additional authors

Journal of Agricultural and Food Chemistry, Vol. 68, No. 47 (November 11, 2020): 13541-13549. DOI.

Proanthocyanidin Block Arrays (PACBAR) for Comprehensive Capture and Delineation of Proanthocyanidin Structures

Shuxi Jing
Pharmacognosy Institute and Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois
Wayne E. Zeller
United States Dairy Forage Research Center, Agricultural Research Service (ARS), United States Department of Agriculture (USDA), Madison, Wisconsin
Daneel Ferreira
National Center for Natural Products Research and Department of Biomolecular Sciences, Division of Pharmacognosy, School of Pharmacy, University of Mississippi, University, Mississippi
Bin Zhou
State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, People's Republic of China
Joo-Won Nam

College of Pharmacy, Yeungnam University, Gyeongsan-si, Gyeongsangbuk-do 38541, Republic of Korea

Ana Bedran-Russo

Department of General Dental Sciences, School of Dentistry, Marquette University, Milwaukee, Wisconsin

Shao-Nong Chen

Pharmacognosy Institute and Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois

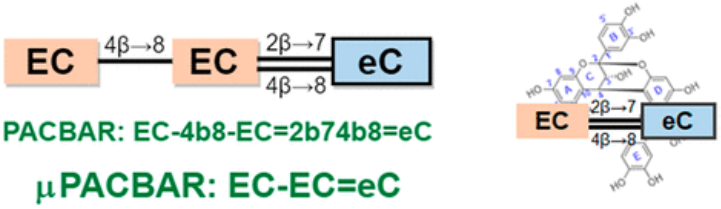
Guido F. Pauli

Pharmacognosy Institute and Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois

Abstract

PACBAR

ProAnthoCyanidin Block ARrays



Proanthocyanidins (PACs) are near-ubiquitous and chemically complex metabolites, prototypical of higher plants. Their roles in food/feed/nutrition and ethnomedicine are widely recognized but poorly understood. With the analysis of evidence that underlies this challenge, this perspective identifies shortcomings in capturing and delineating PAC structures as key factors. While several groups have forwarded new representations, a consensus method that captures PAC structures concisely and offers high integrity for electronic storage is required to reduce confusion in this expansive field. The PAC block arrays (PACBAR) system fills this gap by providing precise and human- and machine-readable structural descriptors that capture PAC metabolomic structural diversity. PACBAR enables communication of PAC structures for the development of precise structure–activity relationships and will assist in advancing PAC research to the next level.

KEYWORDS:

Proanthocyanidins, polyphenols, nomenclature, structure–activity relationships

Background

Abundant contemporary research shows that plant materials rich in oligo- and polymeric proanthocyanidins [PACs, syn. condensed tannins (CTs)] have important roles in food and human nutrition as well as being associated with health benefits when used as dietary supplements.(1,2) A

plethora of reports support this by describing biological activities of PAC-rich extracts and crude natural product mixtures for a host of end points. PACs are oligomeric (defined here as containing 2–9 flavan-3-ol subunits) to polymeric flavan-3-ols that produce anthocyanidins (anthocyanin aglycones) by acid-catalyzed cleavage of the C–C interflavanyl [syn. interflavanoid or IFL bond; not interflavonoid (C-4 carbonyl)] under aerial oxidative conditions. In contrast, leucoanthocyanidins/flavan-3,4-diols generate anthocyanidins by cleavage of the ether or C-4 carbinol (C–O) bond, respectively, upon heating with mineral acid under oxidative conditions.(3)

Recent advances in instrumentation, separation, and structural analysis have made it more possible than ever to characterize PAC materials to the level of single chemical entities and eventually link individual molecules to biological functions. One major impeding factor for establishing such links and advancing the entire field is the realization of the exponential complexity of PACs.(4,5) The diverse set of structurally distinct PAC molecules that nature provides poses unique analytical challenges in structural determination. Importantly, both points reveal shortcomings in the current chemical language, depiction, and nomenclature to communicate PAC structures adequately.

One essential tool for making structure–bioactivity connections is the availability of a public database that makes prior chemistry knowledge, including spectroscopic/spectrometric information [nuclear magnetic resonance (NMR), mass spectrometry (MS), etc.], from PACs accessible to interdisciplinary research. To this end, the U.S. Dairy Forage Research Center (USDFRC) Condensed Tannin NMR database is the most comprehensive tool available to date. It collects basic chemical, sourcing, and reference information on 355 compounds up to tetramers, including their ^{1}H and ^{13}C NMR chemical shift data, and covering reports up to 2015.(6) More importantly, for the first time, this database adopts structural descriptors to represent and search PAC structures. Whereas these “backbone codes” represent a substantial start toward providing a unique tool for cataloguing PACs and rapid electronic searches, the USDFRC database focused on a subset of PACs, failing to capture derivatization, such as galloylation, glycosidation/glycosylation, and methylation, which contribute to the vast, exponential variation of potential isomers.

An indication of the substance of this structural diversity trend can be gleaned from our prior reviews: approximately 500 PACs have been reported from 1992 to 2001,(7−9) with an additional ca. 240 between 2002 and 2010.(1) While fewer reports of new PAC entities have been communicated during the past decade, the ca. 100 reported new PAC structures have grown substantially in structural complexity and notably include many underivatized PACs and the tools for high-accuracy structural and spectroscopic assignments have grown considerably. Collectively, considering the exponential permutational growth of structural possibilities of higher oligomeric PACs,(4,5) the newer reports specifically point to the analytical, structural, and nomenclatural challenges associated with moving this field forward.

To address all of these challenges and facilitate comprehension of the complexity of PACs across disciplines, the intent of this perspective is to rationalize and propose expansion of existing systems of PAC structural descriptors and nomenclature.(10) The overarching goal is to capture all current and any potential future PAC chemical entities comprehensively, facilitate communication of PAC structures between researchers of different disciplines, and support efficient electronic searches. This is to be accomplished by (i) achieving a more adequate description of the PAC chemical space (“PACome”) that

has been recognized to exist in plants, (ii) rendering PAC chemical diversity amenable to computational and database (DB) tools, (iii) expanding on the USDFRC database "backbone code" approach, (iv) providing universal communicable language for written and oral communication, (v) continuing the trend of modular PAC depictions that have recently appeared in the literature, and (vi) accommodating predictable growth of the field. Collectively, these points justify the need for a comprehensive yet simple abbreviation scheme that captures PAC structures accurately in a searchable manner. An important goal is to achieve all six points by maintaining full compatibility with existing, traditional conventions, including the somewhat limited but long-used International Union of Pure and Applied Chemistry (IUPAC) nomenclature rules that are indirectly applicable to PACs (numbering system). While structural descriptors and nomenclature may be perceived as rather formal elements of research, the combined experience of the authors predicates the requirement of a strong and comprehensive system that eliminates ambiguity, clarifies scientific meaning, and promotes reporting PAC structures with precision and quality, making them key elements of advancing chemical and interdisciplinary PAC research to the next level.

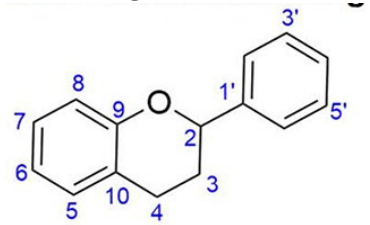
Chemical Diversity

Vast Chemical Space of PACs (PACome)

The structural possibilities of PACs occupy a vast chemical space that appears, to both novice and seasoned scientists, quite chaotic. PACs are presumably biosynthesized from electrophilic aromatic substitution of C-4 of a flavanyl unit (generated from a flavan-3,4-diol or flavan-4-ol) to a nucleophilic flavanyl moiety. PACs are notably distinguished from the related bi- and triflavanoids that are products of phenol oxidative coupling involving flavones, flavanols, etc., possessing a C-4 carbonyl group in every constituent unit.

According to the hydroxylation pattern (Figure S1, Supporting Information), 16 basic PAC units are well-recognized and classified; Table 1 lists their names and natural abundance. All flavans and flavan-3-ols in this list possess (2*S*) and (2*R*,3*S*) absolute configuration, respectively. The most abundant building blocks of PACs, catechin and gallocatechin, are widely distributed in plants, whereas their galloyl esters are characteristic components in green tea (*Camellia sinensis*).(11) PACs containing 5-deoxyflavan-3-ol extension units have only been found in Southern Hemisphere plants: e.g., the profisetinidins are the major constituents of wattle and quebracho tannins, which are important for leather tanning and adhesive manufacturing.(12)

Table 1. Elements of the PACBAR Structural Descriptors and Nomenclature

Monomer	Proanthocyanidin Group and Abundance[a]	Codes[b]			Substituents[c]					
	Flavan Base Structure with Atom Numbering  abundance	Nano Code	Macro Code	*epi* Form	C-3	C-5	C-8	C-3'	C-4'	C-5'
apigeniflavan	proapigeninidins Ø	A	AP	EA	H	OH	H	H	OH	H
afzelechin	propelargonidins +	z	AZ	EZ	OH	OH	H	H	OH	H
butiniflavan	probutinidins Ø	B	BU	EB	H	H	H	OH	OH	H
catechin	procyanidins +++	C	CA	EC	OH	OH	H	OH	OH	H
cassiaflavan	procassinidins Ø	s	cs	ES	H	H	H	H	OH	H
distenin	prodistenidins Ø	D	DI	ED	OH	OH	H	H	H	H
fisetinidol	profisetinidins +	F	FI	EF	OH	H	H	OH	OH	H
gallocatechin	prodelphinidins ++	G	GA	EG	OH	OH	H	OH	OH	OH
guibourtinidol	proguibourtinidins Ø	u	GU	EU	OH	H	H	H	OH	H
luteoliflavan	proluteolindins Ø	L	LU	EL	H	OH	H	OH	OH	H
mesquitol	promelacacinidins Ø	Q	MQ	EQ	OH	H	OH	OH	OH	H
mopanane	promopanidins Ø	M	MO	EM	OCH_2	H	H	OH	OH	H
oritin	proteracacinidins Ø	0	OR	EO	OH	H	OH	H	OH	H
peltogynane	propeltogynidins Ø	p	PE	EP	OCH_2	H	H	H	OH	OH
robinetinidol	prorobinetinidins +	R	RO	ER	OH	H	H	OH	OH	OH
tricetiflavan	protricetinidins Ø	T	TR	ET	H	OH	H	OH	OH	OH

[a]Symbols indicate the abundance of each structural type in nature: "+", high natural occurrence with a substantial ("+") to very large ("+++") number of reported compounds; "⌀", compound class has been discovered but only a few compounds have been reported.

[b]Codes represent the unique, two-letter acronyms for each monomer, to be used in PACBAR naming.

[c]Hydroxylation at C-7 (HO-7 substitution) is considered a default structural element.

PACs are often characterized by the interflavan bond connectivity of their constituent flavan-3-ol units. All PACs contain the single "B-type" linkage consisting of a C–C bond between C-4 of the extender unit and C-6 or C-8 of the contiguous flavan-3-ol unit. The double "A-type" linkages possess an additional ether connectivity between HO-7 or HO-5 (A ring) of the terminal unit and C-2 (C ring) of the contiguous flavan-3-ol unit. PACs can contain only A-type, only B-type, or both A- and B-type linkages, which explains one key element of their structural diversity. According to prior reviews,(1,13) 14 different specific interflavan linkage (IFL) types have been reported (Table S2, Supporting Information). Interestingly, the heterogeneity of IFLs is significantly expanded among 5-deoxy-PACs, such as the profisetinidins, prorobinetinidins, promelacacinidins, proteracacinidins, and proguibourtinidins, where absence of the HO-5 (A ring) substituent allows for a higher proportion of C-4 to C-6 IFLs. This likely arises from the less stable and thus more reactive C-4 carbocations derived from 5-deoxyflavan-3,4-diols and the reduced nucleophilicity of the A ring of 5-deoxyflavan-3-ols that would permit coupling at alternative nucleophilic sites (Figure S1, Supporting Information).(13)

Another important factor driving PAC chemical diversity as well as challenging structural elucidation involves configurational complexity. The stereogenic centers at C-2 and C-3 in the flavan-3-ols lead to the formation of enantiomers and/or diastereoisomers: e.g., catechin possesses the (2*R*,3*S*) absolute configuration, while epicatechin and *ent*-catechin are (2*R*,3*R*)- and (2*S*,3*R*)-configured, respectively. The C-4 configuration at the interflavanyl bond defines the "shape" of the molecule in space. Additionally, rotational hindrance around the IFLs in especially B-type PACs causes the phenomenon of dynamic rotational isomerism (atropisomerism) that significantly complicates NMR spectroscopic investigations(14) and often requires recourse to alternative methods.(15)

Analogous to other chemical classes (peptides, nucleotides, and saccharides), we consider PAC oligomers to have a degree of polymerization (DP) of 2–9 versus polymers with a DP of ≥10. Owing to the structural complexity, low solubility of higher DP PACs, and chromatographic limitations, including atropisomerism, these PACs present higher challenges in purification and structure determination. Reports on the isolation and elucidation of hexamers (DP of 6) are limited to PACs from *Machilus philippinensis*;(16) most recently, the structure of an A-type hexamer from pine bark (*Pinus massoniana*) was fully established by NMR and electronic circular dichroism (ECD) data and supported by phloroglucinolysis.(17) Parallel synthesis, purification, and partial identification of even higher B-type oligo-/polymers up to DP of 11 by ^{1}H NMR and MS data have been reported.(18) Polymers up to DP of 30 have been detected by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF MS),(19) while a DP of 26 was recognized as the electrospray ionization mass spectrometry (ESI–MS) detection limit.(20)

Numbers Game: How Many Individual PACs Are in a Plant?

The theoretical structural possibilities in PAC-rich plants, such as pine (*P. massoniana*) bark, grape (*Vitis vinifera* L.) seed extract, and cacao (*Theobroma cacao* L.), can be calculated on the basis of the constituent monomeric units, IFLs, stereochemistry, and DP that are present in these plants (Table S1, Supporting Information). Using the PAC oligomers from pine as an example and limiting considerations to the current purification and structure elucidation barrier/"wall" with DP of ≤6, all aforementioned factors already make the structural possibilities in excess of 68 000 000 entities. Considering that epiafzelechin was recognized as a new monomeric unit in a trimer,(21) the recognized PAC chemical

space of pine bark continues to expand as new structural features are discovered. Accordingly, PAC structural complexity is increasingly recognized as a factor that challenges the isolation and structural characterization of individual PACs. This also shows how wide the gap between phytochemical and biomedical studies indeed is.

Support for Evolving Structure–Activity Relationships (SARs)

Structures Are Hurdles

PACs are highly distributed in a broad spectrum of foods, forage plants, and agricultural waste (pine bark, peanut skins, etc.). Their well-documented putative effects on mammals, insects, and chemical ecology have drawn attention to agricultural and biomedical research. However, the majority of the bioactivity studies focus on extracts, enriched fractions, or employ only the readily available non-PAC, epigallocatechin gallate (EGCG), and/or PACs like procyanidin B1/2 dimers as "pure" compounds (often without purity analysis).(2,22,23) Dozens of reports have studied the effects of PAC-rich structures on specific proteins or genetic regulations, designating them as "bioactive, natural dietary components" (reviewed in ref (24)).

Informative SARs involving PACs are rare to non-existent. Available information is often confusing as a result of incomplete or missing chemical and/or purity/content characterization of the composition in the tested PAC fractions. Moreover, bioassay interference is prevalent, not only because PACs are prototypical pan-assay interference compounds (PAINS), but also because PACs can act as non-specific aggregators, binders, or precipitation agents in cell-based *in vitro* assays.(25) Importantly, the fact that a given PAC is present in a given plant material or fraction does not indicate its role as a bioactive. Such an assignment requires the rigorous establishment of specificity using pure compounds, demonstration of mechanism of actions, and ideally establishment of SARs. The majority of PAC bioactivity studies lack support by rigorous phytochemical analyses as far as purification and structure elucidation are concerned.

Neither PACs nor PAC Bioactivities Are "All the Same"

Only a few studies on PAC SARs unveiled that different PACs do have specific bioactivities. In the dimeric PACs, dracoflavan B, a pancreatic α-amylase inhibitor from dragon's blood resin (*Daemonorops draco*), its A-ring phenolic group is essential for this activity.(26) Moreover, the interdisciplinary dental research of the authors has recognized specific PACs as promising dentin biomodifiers, with trimers and tetramers exhibiting selective affinity to dentin biomacromolecules (e.g., collagen).(27,28) Studies correlating biomechanical properties with chemical features (constitutional monomers, IFLs, and stereochemistry) are ongoing. Additionally, it has been demonstrated that the addition of galloyl groups to flavan-3-ol monomers and PAC dimers enhance their protein binding affinity toward human parotid salivary proteins(29) bovine serum albumin and human α-amylase(30) compared to the non-galloylated entities. In addition, the presence of A-type linkages in PAC dimers shows a higher affinity toward porcine and bovine trypsin than their B-type linkage counterparts.(31) The presence of A-type linkages was also shown to impact the ability of PACs to inhibit pathogenic *Escherichia coli* infection in epithelial cells.(32) Cases in distinguishing differences in protein affinity of PACs bearing different B-type linkages (i.e., C-4/C-6 versus C-4/C-8) are less clear and appear to depend upon both the protein and PAC structures. For example, with proline-rich saliva proteins, higher tannin-specific activity was

observed with C-4/C-8-linked dimers than their C-4/C-6 isomeric counterparts.(29,33) However, against bovine serum albumin, lysozyme, and trypsin, PACs with a C-4/C-6 terminal IFL appear to be superior protein precipitating agents.(34)

Urgent Need for a SAR-Capable PAC Language

As SAR studies span cross-disciplinary fields, a common communicable language is instrumental in the ability to convey PAC SAR information. Exemplified by our ongoing evaluation of the dental biomodification potency of PACs, aimed at determining pharmacophores, there is an urgent need for a consensus naming system that is rooted in widely accepted rules but still can better communicate the structural subtleties and the chemical complexity of PACs, while also being devoid of the space-consuming complex structural formulas. This may also help to establish a system for PAC bioactivity descriptors beyond the unjustified blanket notion that “all PACs are the same”.

Polyphenol Confusion

The term polyphenol was initially intended and exclusively used for polymeric (not polyhydroxylated) compounds containing multiple hydroxy-substituted (abbreviation, OL) phenyl (abbreviation, PHEN) constituents, hence the generic PHENOL designation. Typical and valid polyphenol examples include the proanthocyanidins and the hydrolyzable tannins, i.e., gallotannins and ellagitannins. This term has been causing much confusion as has recently been highlighted by a consortium of scientists.(35) In addition, contemporary publications commonly and indiscriminately dub simple phenolic compounds like afzelechin, resveratrol, curcumin, the silybins, and others as “polyphenols”. In these instances, it would be much more appropriate to use the specific type of compound, e.g., isoflavan glycosides. In fact, the term “polyphenol” does not convey any useful meaning but rather introduces confusion and should be avoided altogether. With the facilitation of navigation of all flavan-3-ols, PAC block arrays (PACBAR) contributes to a better understanding of the structural and biological implication of the vast chemical space of both polymeric and polyhydroxylated compounds.

Electronic Storage and Data Mining

Elucidation with Stereochemical Specificity

The decades of progress made in structural determination on PAC research now requires the field to enter into the digital age, thus necessitating digitizing structures into electronically retrievable entities for archiving research data from publications. One of the authors had initiated an online NMR data collection of PACs, the USDFRC CT NMR databasewww.ars.usda.gov/mwa/madison/dfrc/tannin). It provides searchable features like chemical structure, DP, and NMR chemical shifts and particularly leads the way to use structural descriptors to denote PAC structures.(6) This feature makes the database more “user friendly” than general chemistry search tools, which normally need to draw the complex structure or enter the inconsistently used trivial name.

The structural elucidation of PACs can benefit from data collections, for which a significant feature is the repetition of certain flavan units. MS data provide molecular weight information that refer to the DP of PACs; In combination with the NMR data, the configuration of constitutional units and IFLs can be derived via a comparison to well-established cases. The readily accessible NMR database enhances efficiency and accuracy of the structural elucidation/dereplication of PACs (especially higher oligomers) as well as composition analysis of crude materials. Assignment of the absolute configuration of PACs

has recently become accessible via a comparison of ^{13}C NMR chemical shifts to those of PACs with fully established stereochemistry, e.g., tetramers.(28) The diagnostic ^{13}C NMR γ-gauche effect influencing the chemical shifts of C-2 in the extension units is another powerful tool in determining both the relative configurations of C-2 and C-4 and the absolute configuration of monomeric units in oligomers, using reference data with ECD-based absolute configurational assignment of C-4.(21) Because such progress depends upon unambiguity of both the structural assignment and the underlying NMR data, the role of PAC nomenclature cannot be overemphasized.

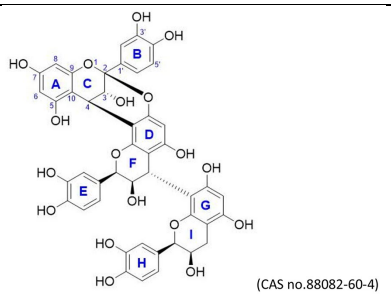
Enhance Database Linguistics To Harness Biological Specificity

While the number and complexity of PACs continue to grow, researchers are striving, often with confusion, to delineate and classify these structures to establish connections with "universal" bioactivities. To encompass the structural diversity of PACs, we, herewith, introduce the development of PACBAR as a tool and inclusive nomenclature that uses modular identifiers and can be used to annotate PACs in this database. A transferable version of the current database will be available for interested researchers or platforms, to support future development in PAC research in the coming "Big(ger) Data" era, such as metabolomics analyses or deep learning in chemical structure annotation. Databases are key tools for understanding chemical space in the literature versus the theoretical permutations emphasized earlier; there is a biosynthetic preference for plants to more commonly produce certain types of PACs.

PACBAR Structural Descriptors

Historically, PACs have been given trivial names, such as procyanidin B1 for epicatechin-(4β→8)-catechin, with the latter name following the now widely used system proposed by Hemingway et al.(10) As newly elucidated PACs became lengthier and more complicated at higher DPs, authors reverted to plant-derived trivial names, such as the trimer, cinnamtannin B-1, from *Cinnamomum* spp. However, because trivial names lack structural information, they are incapable of expressing the structural resemblance or divergence required to communicate, e.g., SAR information or chemical similarity. The PACBAR system incorporates accepted IUPAC nomenclature, works analogous to oligo-/polysaccharide nomenclature,(10) and reconciles all structural variables including flavan monomers (Table 1 and Table S2, Supporting Information). Table 2 shows how PACBAR accommodates all essential chemical identifiers of cinnamtannin B-1 to synthesize three descriptive schemes: the letter- and color-coded graphical PACBAR structure, the plain text macro-PACBAR code, and the minimalist yet fully descriptive micro-PACBAR code.

Table 2. Nomenclature Dilemma and PACBAR Solution Exemplified

CLASSICAL	
Chemical Structure	 (CAS no.88082-60-4)

Trivial name	Cinnamtannin B-1
Prior nomenclature	epicatechin$-(2\beta \rightarrow 7,4\beta \rightarrow 8)-$epicatechin-$(4\beta \rightarrow 8)$-epicatechin
IUPAC name[1]	(1*R*, 5*R*,6*R*, 75,13*S*,21*R*)-5,13-bis(3,4-dihydroxyphenyl)-7-[(2*R*,3*R*)- 2-(3,4-dihydroxyphenyl)-3,5,7 -trihydroxy-3,4-dihydro-2*H*- chromen-8-yl]-4,12,14-trioxapentacyclo[11.7.1.$0^{2,11}$.$0^{3,8}$.$0^{15,20}$]henicosa-2(11),3(8),9,15,17,19-hexaene-6,9,17,19,21-pentol
InChI[2]	InChI=1S/C45H36O18/c46-18-10-27(54)33-31(11-18)62-45(17-3-6-22(49)26(53)9-17)44(59)38(33)36-32(63-45)14-29(56)35-37(39(58)41(61-43(35)36)16-2-5-21(48)25(52)8-16)34-28(55)13-23(50)19-12-30(57)40(60-42(19)34)15-1-4-20(47)24(51)7-15/hI-11,13-14,30,37-41,44,46-59H,12H2/t30-,37+,38-,39-,40-,41-,44-,45+/ml/s1
PACBAR	
macro PACBAR	EC=2b74b8=EC-4b8-EC
micro PACBAR	EC=8EC-8EC
Graphical PACBAR	

[1]Cited from Pubchem.
[2]IUPAC International Chemical Identifier.

PACBAR Basics

PACBAR uses monomer codes as follows: (i) a single capital letter code abbreviates the basic flavan unit (Table 1); (ii) prefixes: "e" for "*ent*-" and "E" for "*epi*-"; and (iii) suffixes: "g" for the 3-*O*-galloyl group; e.g., "eECg" is *ent*-epicatechin gallate. The IFLs are represented/drawn as "–" and "═" for single and double linkages, respectively. Configurations and linkages are drawn above and below the bonds/lines using the conventional naming (e.g., 4β→8) in the graphical PACBAR (Figure 1). Structural elements commonly found in PACs are given default status, permitting their exclusion when building minimalist micro-PACBAR code: (a) C-4 as the most frequently linkage site of the extension unit, (b) the ether bond 2[*O*]→7 in A-type PACs, and (c) 4β orientation in IFLs (Figure 1). To simplify textual encoding, macro- and micro-PACBAR use "a/b" instead of "α/β". Table 3 collates more details of the PACBAR nomenclature.

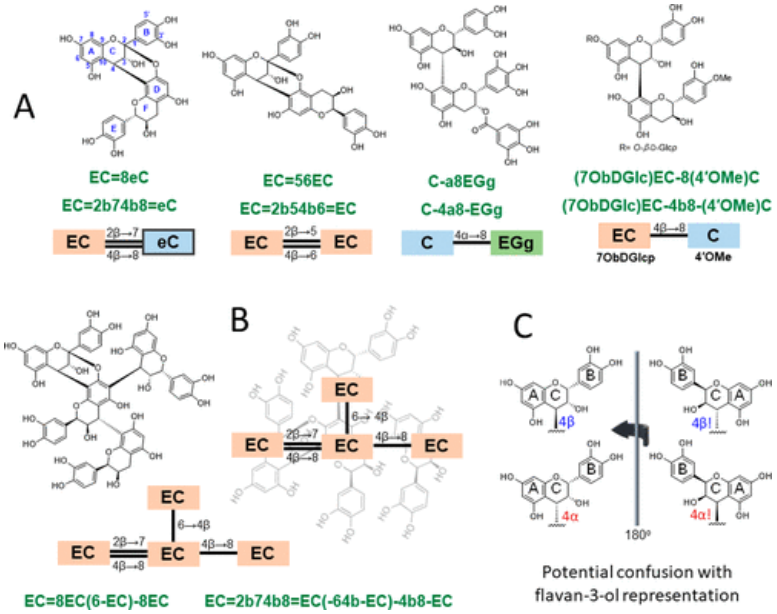


Figure 1. PACBAR nomenclature applied to (A) diverse set of dimers and (B) one branched tetramer. Shown are the classical chemical drawings versus the simplified macro- and micro-PACBAR name pairs (green) versus the graphical PACBAR. (B) Overlay of the PACBAR and the classical structures of the branched tetramer exemplifies how PACBAR avoids the error-prone subtleties of classical drawing while still providing precise structural information and resembling the overall shape of PACs (overlay of PACBAR and classical structure). PACBAR follows the standard method of selecting the longest contiguous chain of flavan-3-ol subunits containing the terminal monomer (i.e., the one possessing a C-4 methylene group). In the case where branching occurs to an equal extent, A-type linkages take precedence over B-type linkages, in the order of more abundant 4→8 having priority over 4→6 linkages. (C) Importantly, PACBAR avoids the confusion potential of 4α/4β designation that can occur in classical drawings when a flavan-3-ol unit is rotated by 180° in the paper plane (not mirrored!) compared to its typical presentation (ring order A[lower left]–C–B[upper right]). In the given example of the tetramer, EC═8EC(6-EC)–8EC, the dashed 4→8 bond still represents a 4β-configured epicatechin unit after the 180° rotation in the paper plane, which is often necessary in the classical drawing format to accommodate certain linkages. Some readers might find it helpful to use 4β to indicate *trans* configuration relative to the C-2 aryl substituent, whereas 4α means *cis* relative configuration. Notably, this situation inverts in the *ent* series of monomers, adding to the potential confusion. Collectively, this highlights another strong rationale for establishing a nomenclature and graphical representation system, such as the PACBAR.

Table 3. Components of the PACBAR Nomenclature

element	graphical PACBAR	macro-PACBAR	micro-PACBAR
abbreviation of basic unit	• use one-letter codea for the basic unit of (2*S*) and (2*R*,3*S*) absolute configuration (e.g., G for gallocatechin)		
	• flavan-3-ols with (2*R*,3*R*) configuration are prefixed with "E" (e.g., EC for epicatechin)		
	• enantiomeric units are prefixed with "e" (e.g., eC for *ent*-catechin)		
	• use blocks of different color for each monomer and bold border for their less common enantiomers		
IFLs	• draw lines that connect blocks to indicate the IFLs	• doubly and singly interflavanyl bonds are symbolized as "═" and "–", respectively	
	• connection sites are denoted above and under the "bond"	• doubly and singly interflavanyl bonds are symbolized as "═" and "–", respectively	
	• keep the arrows and α/β as a means of indicating direction toward the terminal unit as having nucleophile/reactive properties	• use a and b to represesent the α and β configuration of IIFLs	
	• keep the arrows and α/β as a means of indicating direction toward the terminal unit as having nucleophile/reactive properties	• use a and b to represesent the α and β configuration of IIFLs	• consider most common linkage sites (C-4, C-2, and C-7) and configuration (4β) as defaults and drop themb
			• keep one IFL symbol in between the units
substituents		• galloyl group (gallates): add suffix "g"	
		• acetate: add "Ac"	

		• carbohydrates: add their abbreviations, e.g., glc*p* for glucopyranoside	
		• other substituents: use the appropriate IUPAC or ACS abbreviationsc	
branchedand macrocyclic PACs	• longest chain (=contiguous series of monomeric units) takes precedence		
	• determine the longest chain by following C-4 (methylene group) as the default terminal point		
	• add the branched substituents, using brackets for each branching moiety; see Figure 1 for an example of a branched tetramer		
	• branching units or chains are inserted in brackets and listed after the unit of attachment		
	• IFLs are listed in the order in which the atoms are aligned with the main chain; this means that bond directions are annotated from the main chain perspective		
	− e.g., a generic 4β→6 IFL is annotated as 6→4β from the branching monomer point of view; this is in line with the priority of the chain and avoids conflict when the branching unit already has a 4β→6 or 4β→8 bond		
	• in case of a tie in branching points, the following priority rules apply: length of branch > A-type > 4→8 > 4→6 > gallates		
	• IFL numbering proceeds from the main chain toward both the terminal and branched units; accordingly, in the numbering of IFLs at a branching point, the atom numbers of the preceding monomer take priority over the atom numbers of the subsequent unit		

	• to indicate macrocyclic PACs, the chain of flavan-3-ols will be enclosed by pipe (|) universal connector symbols		
applications	• graphical representation, replacing regular structural formulas	• computer language	• plain text in publication
		• database retrieval entry	•pronounceable forms of a PAC name

[a]Monomer abbreviations in Table 1.

[b]The descriptors for these default features are left out to keep micro-PACBAR names concise.

[c]Abbreviations and names of common substituents in PACs are listed in Table S3, Supporting Information.

Additional Considerations

PACBAR adopts American Chemical Society (ACS) terminology for Me, Ac, Bu, and Bn substituents. Table S3 of the Supporting Information collates acronyms for common functional groups. For example, 7-*O*-β-d-Glc*p*-epicatechin-(4β→8)-4′-*O*-methylcatechin could be encoded as (7ObDglcp)EC-8(4′OMe)C (Figure 2). Because flavan-3-ols with (2*R*,3*TBFS*) versus (2*S*,3*R*) absolute configuration are intrinsically dextro- versus levorotatory, the usage of the optical rotation signs, (+) versus (−), is superfluous; instead, names such as catechin (C) versus *ent*-catechin (eC) are recommended. PACBAR does not cover non-PAC flavan or flavan-3-ol constituent units.

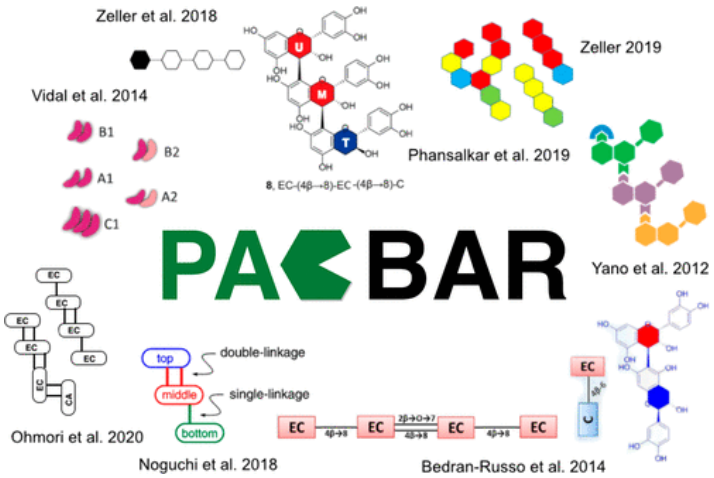


Figure 2. The PACBAR scheme and nomenclature consolidate the formats in numerous recent PAC publications(4,27,36−40) that seek to capture the PAC building patterns and the three-dimensional (3D) shapes of the molecules in a variety of ways.

Practical Application Scenario

The color-coded graphical PACBAR is for visual purposes intended to replace the chemical formula and can function as a precise but “graphical abstract” for PAC structures. The plain text macro- and micro-PACBARs are not only compatible with current nomenclature but also computer/database-readable and communicable. The macro-PACBAR contains all elements of a PAC name, is fully descriptive without knowledge of the default elements (see above), and fully amenable to computational and database tools. Meanwhile, micro-PACBAR uses default structural features to reduce the code length for enhanced communication purposes and intended to replace trivial or systematic names.

Advancing Interdisciplinary PAC Research

PAC research spans multiple disciplines, including human and ruminant health, productivity and sustainability, material sciences, and chemical ecology. This perspective is not intended to detract from or substitute the informative and often wonderful cartoon representations forwarded by many authors of PAC structures in their papers. At the same time, the thrust of new and expanded analytical methods providing detailed structural analysis of purified PACs makes it necessary to establish consensus development of a universal PAC nomenclature scheme. The proposed PACBAR system

accurately captures PAC structures, allows for rapid visualization, and can be readily reduced to an electronic searchable entry to foster interdisciplinary research.

Terms & Conditions

Funding

The University of Illinois at Chicago and Marquette University authors acknowledge the long-term funding of their research on PACs as biomodifiers of dentin through Grants R01 DE021040, R56 DE021040, and R01 DE028194 from the National Institute of Dental and Craniofacial Research (NIDCR)/National Institutes of Health (NIH).

Notes

The authors declare no competing financial interest.

Abbreviations Used	
DP	degree of polymerization
ECD	electronic circular dichroism
ESI–MS	electrospray ionization mass spectrometry
IFL	interflavan linkage
MALDI–TOF MS	matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
NMR	nuclear magnetic resonance
PAC	proanthocyanidin
PACBAR	proanthocyanidin block arrays
SAR	structure–activity relationship

References

1 Ferreira, D.; Marais, J. P. J.; Coleman, C. M.; Slade, D. Proanthocyanidins: Chemistry and Biology. In *Comprehensive Natural Products II*; Liu, H.-W., Mander, L., Eds.; Elsevier: Amsterdam, Netherlands, 2010; Vol. 6, Chapter 18, pp 605– 661, DOI: 10.1016/B978-008045382-8.00648-1 .

2 Neilson, A. P.; O'Keefe, S. F.; Bolling, B. W. High-Molecular-Weight Proanthocyanidins in Foods: Overcoming Analytical Challenges in Pursuit of Novel Dietary Bioactive Components. *Annu. Rev. Food Sci. Technol.* **2016**, *7* (1), 43– 64, DOI: 10.1146/annurev-food-022814-015604

3 Xie, D.-Y.; Sharma, S. B.; Paiva, N. L.; Ferreira, D.; Dixon, R. A. Role of Anthocyanidin Reductase, Encoded by BANYULS in Plant Flavonoid Biosynthesis. *Science* **2003**, *299* (5605), 396– 399, DOI: 10.1126/science.1078540

4 Bedran-Russo, A. K.; Pauli, G. F.; Chen, S.-N.; McAlpine, J.; Castellan, C. S.; Phansalkar, R. S.; Aguiar, T. R.; Vidal, C. M. P.; Napotilano, J. G.; Nam, J.-W.; Leme, A. A. Dentin Biomodification: Strategies,

Renewable Resources and Clinical Applications. *Dent. Mater.* **2014**, *30* (1), 62– 76, DOI: 10.1016/j.dental.2013.10.012

5 Naumann, H. D.; Tedeschi, L. O.; Zeller, W. E.; Huntley, N. F. The Role of Condensed Tannins in Ruminant Animal Production: Advances, Limitations and Future Directions. *Rev. Brasil. Zootec.* **2017**, *46* (12), 929– 949, DOI: 10.1590/s1806-92902017001200009

6 Zeller, W. E.; Schatz, P. F. The U.S. Dairy Forage Research Center (USDFRC) Condensed Tannin NMR Database. *J. Agric. Food Chem.* **2017**, *65* (25), 5104– 5106, DOI: 10.1021/acs.jafc.7b02314

7 Ferreira, D.; Bekker, R. Oligomeric Proanthocyanidins: Naturally Occurring *O*-Heterocycles. *Nat. Prod. Rep.* **1996**, *13* (5), 411– 433, DOI: 10.1039/np9961300411

8 Ferreira, D.; Li, X. C. Oligomeric Proanthocyanidins: Naturally Occurring *O*-Heterocycles. *Nat. Prod. Rep.* **2000**, *17* (2), 193– 212, DOI: 10.1039/a705728h

9 Ferreira, D.; Slade, D. Oligomeric Proanthocyanidins: Naturally Occurring *O*-Heterocycles. *Nat. Prod. Rep.* **2002**, *19* (5), 517– 541, DOI: 10.1039/b008741f

10 Hemingway, R. W.; Foo, L. Y.; Porter, L. J. Linkage Isomerism in Trimeric and Polymeric 2,3-*cis*-Procyanidins. *J. Chem. Soc., Perkin Trans. 1* **1982**, (5), 1209– 1216, DOI: 10.1039/p19820001209

11 Hemingway, R. W. Biflavonoids and Proanthocyanidins. In *Natural Products of Woody Plants*; Rowe, J. W., Ed.; Springer: Berlin, Germany, 1989; Springer Series in Wood Science, Vol. 12, pp 571– 651, DOI: 10.1007/978-3-642-74075-6_17 .

12 Venter, P. B.; Sisa, M.; van der Merwe, M. J.; Bonnet, S. L.; van der Westhuizen, J. H. Analysis of Commercial Proanthocyanidins. Part 1: The Chemical Composition of Quebracho (*Schinopsis lorentzii* and *Schinopsis balansae)* Heartwood Extract. *Phytochemistry* **2012**, *73* (1), 95– 105, DOI: 10.1016/j.phytochem.2011.10.006

13 *Flavonoids: Chemistry, Biochemistry and Applications*; Andersen, Ø. M., Markham, K. R., Eds.; CRC Press (Taylor & Francis Group): Boca Raton, FL, 2006; pp 1256.

14 Fletcher, A. C.; Porter, L. J.; Haslam, E.; Gupta, R. K. Plant Proanthocyanidins. Part 3. Conformational and Configurational Studies of Natural Procyanidins. *J. Chem. Soc., Perkin Trans. 1* **1977**, (14), 1628, DOI: 10.1039/p19770001628

15 Esatbeyoglu, T.; Jaschok-Kentner, B.; Wray, V.; Winterhalter, P. Structure Elucidation of Procyanidin Oligomers by Low-Temperature ^{1}H NMR Spectroscopy. *J. Agric. Food Chem.* **2011**, *59* (1), 62– 69, DOI: 10.1021/jf1032334

16 Lin, H.-C.; Lee, S.-S. Proanthocyanidins from the Leaves of *Machilus philippinensis*. *J. Nat. Prod.* **2010**, *73* (8), 1375– 1380, DOI: 10.1021/np1002274

17 Zhou, B.; Alania, Y.; Reis, M. C.; McAlpine, J.; Bedran-Russo, A. K.; Pauli, G.; Chen, S.-N. Rare A-Type, New Spiro-Type, and Highly Oligomeric Proanthocyanidins from *Pinus massoniana*. *Org. Lett.* **2020**, *22* (14), 5304– 5308, DOI: 10.1021/acs.orglett.0c01439

18 Kozikowski, A. P.; Tückmantel, W.; Böttcher, G.; Romanczyk, L. J., Jr. Studies in Polyphenol Chemistry and Bioactivity. 4.(1) Synthesis of Trimeric, Tetrameric, Pentameric, and Higher Oligomeric Epicatechin-Derived Procyanidins Having All-4,8-Interflavan Connectivity and Their Inhibition of Cancer Cell Growth through Cell Cycle Arrest. *J. Org. Chem.* **2003**, *68* (5), 1641– 1658, DOI: 10.1021/jo020393f

19 Takahata, Y.; Ohnishi-Kameyama, M.; Furuta, S.; Takahashi, M.; Suda, I. Highly Polymerized Procyanidins in Brown Soybean Seed Coat with a High Radical-Scavenging Activity. *J. Agric. Food Chem.* **2001**, *49* (12), 5843– 5847, DOI: 10.1021/jf010307x

20 Mouls, L.; Mazauric, J.-P.; Sommerer, N.; Fulcrand, H.; Mazerolles, G. Comprehensive Study of Condensed Tannins by ESI Mass Spectrometry: Average Degree of Polymerisation and Polymer

Distribution Determination from Mass Spectra. *Anal. Bioanal. Chem.* **2011**, *400* (2), 613– 623, DOI: 10.1007/s00216-011-4751-7

21 Zhou, B.; Alania, Y.; Reis, M.; Phansalkar, R.; Nam, J.-W.; McAlpine, J.; Chen, S.-N.; Bedran-Russo, A. K.; Pauli, G. Tri- and Tetrameric Proanthocyanidins with Dentin Bioactivities from *Pinus massoniana*. *J. Org. Chem.* **2020**, *85* (13), 8462– 8479, DOI: 10.1021/acs.joc.0c00783

22 Smeriglio, A.; Barreca, D.; Bellocco, E.; Trombetta, D. Proanthocyanidins and Hydrolysable Tannins: Occurrence, Dietary Intake and Pharmacological Effects. *Br. J. Pharmacol.* **2017**, *174* (11), 1244– 1262, DOI: 10.1111/bph.13630

23 Rauf, A.; Imran, M.; Abu-Izneid, T.; Iahtisham-Ul-Haq; Patel, S.; Pan, X.; Naz, S.; Sanches Silva, A.; Saeed, F.; Rasul Suleria, H. A. Proanthocyanidins: A Comprehensive Review. *Biomed. Pharmacother.* **2019**, *116*, 108999, DOI: 10.1016/j.biopha.2019.108999

24 Bladé, C.; Arola-Arnal, A.; Crescenti, A.; Suárez, M.; Bravo, F. I.; Aragonès, G.; Muguerza, B.; Arola, L. Proanthocyanidins and Epigenetics. In *Handbook of Nutrition, Diet, and Epigenetics*; Patel, V., Preedy, V., Eds.; Springer: Cham, Switzerland, 2019; pp 1933– 1956, DOI: 10.1007/978-3-319-55530-0_16 .

25 Bisson, J.; McAlpine, J. B.; Friesen, J. B.; Chen, S. N.; Graham, J.; Pauli, G. F. Can Invalid Bioactives Undermine Natural Product-Based Drug Discovery?. *J. Med. Chem.* **2016**, *59* (5), 1671– 1690, DOI: 10.1021/acs.jmedchem.5b01009

26 Toh, Z. S.; Wang, H.; Yip, Y. M.; Lu, Y.; Lim, B. J. A.; Zhang, D.; Huang, D. Phenolic Group on A-Ring Is Key for Dracoflavan B as a Selective Noncompetitive Inhibitor of α-Amylase. *Bioorg. Med. Chem.* **2015**, *23* (24), 7641– 7649, DOI: 10.1016/j.bmc.2015.11.008

27 Vidal, C. M. P.; Leme, A. A.; Aguiar, T. R.; Phansalkar, R.; Nam, J.-W.; Bisson, J.; McAlpine, J. B.; Chen, S.-N.; Pauli, G. F.; Bedran-Russo, A. Mimicking the Hierarchical Functions of Dentin Collagen Cross-Links with Plant Derived Phenols and Phenolic Acids. *Langmuir* **2014**, *30* (49), 14887– 14893, DOI: 10.1021/la5034383

28 Nam, J.-W.; Phansalkar, R. S.; Lankin, D. C.; McAlpine, J. B.; Leme-Kraus, A. A.; Vidal, C. M. P.; Gan, L.-S.; Bedran-Russo, A.; Chen, S.-N.; Pauli, G. F. Absolute Configuration of Native Oligomeric Proanthocyanidins with Dentin Biomodification Potency. *J. Org. Chem.* **2017**, *82* (3), 1316– 1329, DOI: 10.1021/acs.joc.6b02161

29 de Freitas, V.; Mateus, N. Structural Features of Procyanidin Interactions with Salivary Proteins. *J. Agric. Food Chem.* **2001**, *49* (2), 940– 945, DOI: 10.1021/jf000981z

30 Soares, S.; Mateus, N.; de Freitas, V. Interaction of Different Polyphenols with Bovine Serum Albumin (BSA) and Human Salivary α-Amylase (HSA) by Fluorescence Quenching. *J. Agric. Food Chem.* **2007**, *55* (16), 6726– 6735, DOI: 10.1021/jf070905x

31 Helsper, J. P. F. G.; Hoogendijk, J. M.; van Norel, A.; Kolodziej, H. Characterization and Trypsin Inhibitor Activity of Proanthocyanidins from *Vicia faba*. *Phytochemistry* **1993**, *34* (5), 1255– 1260, DOI: 10.1016/0031-9422(91)80011-O

32 Feliciano, R. P.; Meudt, J. J.; Shanmuganayagam, D.; Krueger, C. G.; Reed, J. D. Ratio of "A-Type" to "B-Type" Proanthocyanidin Interflavan Bonds Affects Extra-Intestinal Pathogenic *Escherichia coli* Invasion of Gut Epithelial Cells. *J. Agric. Food Chem.* **2014**, *62* (18), 3919– 3925, DOI: 10.1021/jf403839a

33 Bacon, J. R.; Rhodes, M. J. C. Development of a Competition Assay for the Evaluation of the Binding of Human Parotid Salivary Proteins to Dietary Complex Phenols and Tannins Using a Peroxidase-Labeled Tannin. *J. Agric. Food Chem.* **1998**, *46* (12), 5083– 5088, DOI: 10.1021/jf980791c

34 Ezaki-Furuichi, E.; Nonaka, G.-i.; Nishioka, I.; Hayashi, K. Affinity of Procyanidins (Condensed Tannins) from the Bark of *Rhaphiolepis umbellata* for Proteins. *Agric. Biol. Chem.* **1987**, *51* (1), 115– 120, DOI: 10.1080/00021369.1987.10867976

35 Frank, J.; Fukagawa, N. K.; Bilia, A. R.; Johnson, E. J.; Kwon, O.; Prakash, V.; Miyazawa, T.; Clifford, M. N.; Kay, C. D.; Crozier, A.; Erdman, J. W., Jr.; Shao, A.; Williamson, G. Terms and nomenclature used for plant-derived components in nutrition and related research: efforts toward harmonization. *Nutr. Rev.* **2020**, *78* (6), 451– 458, DOI: 10.1093/nutrit/nuz081

36 Noguchi, Y.; Takeda, R.; Suzuki, K.; Ohmori, K. Total Synthesis of Selligueain A, a Sweet Flavan Trimer. *Org. Lett.* **2018**, *20* (10), 2857– 2861, DOI: 10.1021/acs.orglett.8b00873

37 Zeller, W. E. Activity, Purification, and Analysis of Condensed Tannins: Current State of Affairs and Future Endeavors. *Crop Sci.* **2019**, *59* (3), 886– 904, DOI: 10.2135/cropsci2018.05.0323

38 Phansalkar, R. S.; Nam, J.-W.; Leme-Kraus, A. A.; Gan, L.-S.; Zhou, B.; McAlpine, J. B.; Chen, S.-N.; Bedran-Russo, A. K.; Pauli, G. F. Proanthocyanidin Dimers and Trimers from *Vitis vinifera* Provide Diverse Structural Motifs for the Evaluation of Dentin Biomodification. *J. Nat. Prod.* **2019**, *82* (9), 2387– 2399, DOI: 10.1021/acs.jnatprod.8b00953

39 Yano, T.; Ohmori, K.; Takahashi, H.; Kusumi, T.; Suzuki, K. Unified Approach to Catechin Hetero-Oligomers: First Total Synthesis of Trimer EZ-EG-CA Isolated from *Ziziphus jujuba*. *Org. Biomol. Chem.* **2012**, *10* (38), 7685– 7688, DOI: 10.1039/c2ob26337h

40 Ohmori, K.; Suzuki, K. Recent Advances in Polyphenol Research. In *Recent Advances in Polyphenol Research*; Reed, J., de Freitas, V., Quideau, S., Eds.; John Wiley & Sons: Hoboken, NJ, 2020; Vol. 7.