Title
Structure prediction and network analysis of chitinases from the Cape sundew, Drosera capensis

Journal
Biochimica et Biophysica Acta (BBA) - General Subjects, 1861(3)

ISSN
0304-4165

Authors
Unhelkar, Megha H
Duong, Vy T
Enendu, Kaosoluchi N
et al.

Publication Date
2017-03-01

DOI
10.1016/j.bbagen.2016.12.007

Peer reviewed

HHS Public Access

Author manuscript

Biochim Biophys Acta Gen Subj. Author manuscript; available in PMC 2019 August 04.

Published in final edited form as:

Biochim Biophys Acta Gen Subj. 2017 March ; 1861(3): 636–643. doi:10.1016/j.bbagen.2016.12.007.

Structure prediction and network analysis of chitinases from the Cape sundew, Drosera capensis.

Megha H. Unhelkar[†,‡], **Vy T. Duong**[†,¶,‡], **Kaosoluchi N. Enendu**[¶], **John E. Kelly**[†], **Seemal Tahir**[†], **Carter T. Butts**[*,§,‖,⊥], **Rachel W. Martin**[*,†,¶]

[†]Department of Chemistry, University of California, Irvine, Irvine, CA 92697

[‡]These authors contributed equal work.

[¶]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine CA 92697

[§]Department of Sociology, UC Irvine, Irvine, CA, 92697 USA

[‖]Department of Electrical Engineering and Computer Science, UC Irvine, Irvine, CA, 92697 USA

[⊥]Department of Statistics, UC Irvine, Irvine, CA, 92697 USA

Abstract

Background: Carnivorous plants possess diverse sets of enzymes with novel functionalities applicable to biotechnology, proteomics, and bioanalytical research. Chitinases constitute an important class of such enzymes, with future applications including human-safe antifungal agents and pesticides. Here, we compare chitinases from the genome of the carnivorous plant *Drosera capensis* to those from related carnivorous plants and model organisms.

Methods: Using comparative modeling, *in silico* maturation, and molecular dynamics simulation, we produce models of the mature enzymes in aqueous solution. We utilize network analytic techniques to identify similarities and differences in chitinase topology.

Results: Here, we report molecular models and functional predictions from protein structure networks for eleven new chitinases from *D. capensis,* including a novel class IV chitinase with two active domains. This architecture has previously been observed in microorganisms but not in plants. We use a combination of comparative and de novo structure prediction followed by molecular dynamics simulation to produce models of the mature forms of these proteins in aqueous solution. Protein structure network analysis of these and other plant chitinases reveal characteristic features of the two major chitinase families.

General Significance: This work demonstrates how computational techniques can facilitate quickly moving from raw sequence data to refined structural models and comparative analysis, and to select promising candidates for subsequent biochemical characterization. This capability is

[*]To whom correspondence should be addressed buttsc@uci.edu; rwmartin@uci.edu, Phone: 1(949)-824-7959.

Author Contributions

R.W.M. chose the protein set and oversaw the structural biology aspects of the study. C.T.B. performed the cluster analysis, molecular dynamics simulations, and network visualization and analysis. R.W.M., M.H.U., V.T.D., K.E., S.T., and J.E.K. performed sequence annotation and structural analysis. M.H.U., V.T.D, C.T.B. and R.W.M. wrote the manuscript.

increasingly important given the large and growing body of data from high-throughput genome sequencing, which makes experimental characterization of every target impractical.

Keywords

chitinase; protein sequence analysis; protein structure prediction; protein structure network; molecular dynamics; in silico maturation

Introduction

Chitin, a polymer of β-(1,4)-N acetylglucosamine (GlcNAc), is the second-most abundant biopolymer.[1] Chitinases (EC 3.2.1.14) are ubiquitous even among organisms that do not produce chitin, with the latter employing them for purposes of digestion and/or defense. These enzymes cleave chitin at the β-1,4 linkage of N-acetyl glucosamine units, although substantial variation in activity and substrate specificity exists. Some chitinases can also cleave peptidoglycans at β-1,4 linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine, and chitodextrins between N-acetyl-D-glucosamine units. Plant chitinases sometimes have multiple functionalities; some display lysozyme activity,[2] while others have a calcium storage function.[3] In humans, chitinases are produced in response to fungal infections, a feature of the innate immune system that is suppressed in immunocompromised individuals, including AIDS patients, transplant recipients, and burn victims.[4] These enzymes and related chitin-binding proteins are expressed in human lung tissue, where they are dysregulated in cystic fibrosis and asthma.[5]

In plants, these enzymes are expressed in response to environmental stress and pathogen or pest infestation,[6] driving efforts to overexpress particularly effective examples in transgenic crop plants.[7] Carnivorous plants use chitinases as part of the prey capture response: active chitinases have been found in the pitcher fluid of *Nepenthes*,[8,9] and in the digestive fluids of the Venus flytrap.[10] However, the extent to which chitin is used as a nitrogen source remains controversial. *Drosera capensis* plants fed on chitin incorporate its nitrogen into their leaf tissue; however nutrient uptake is less efficient than for plants fed on protein.[11] Examination of insect carcasses after digestion reveals that 40-60% of the total nitrogen is unused,[12,13] consistent with the observation that the remains of insect exoskeletons appear mostly intact.[14] However, chitinase expression is upregulated in the presence of prey in the related species *Nepenthes alata*. In *Drosera rotundifolia,* an increase in both expression of chitinase mRNA and chitinase activity was induced by addition of crustacean chitin with mechanical stimulation of the traps.[15] The prey-induced induction of chitinase activity, despite the low efficiency of chitin use, may indicate that chitinases primarily function to inhibit fungal growth in the traps, just as cytotoxic peptides discourage microbial growth in the fluid of *Nepenthes* pitchers.[16,17]

Here, we compare novel chitinases recently discovered from the genome of the Cape sundew (*Drosera capensis*),[18] to those from other carnivorous plants in order Caryophylalles. The conservation of the overall protein folds and active site architectures suggests that many of the *D. capensis* chitinase sequences form functional enzymes. We use sequence analysis, comparative modeling with all-atom refinement followed by *in silico* maturation,[19] and

investigation of protein structure networks to identify structurally distinct subgroups of proteins for subsequent expression and biochemical characterization.

Results and Discussion

Two Distinct Families of Carnivorous Plant Chitinases Are Found

Gene sequences annotated as coding for chitinases using the MAKER-P (v2.31.8) pipeline[20] and a BLAST search against SwissProt (downloaded 8/30/15) and InterProScan[21] were clustered by sequence similarity, along with chitinases previously identified from *Dionaea muscipula*[10] and various species of *Drosera* and *Nepenthes*.[22] Annotated sequence alignments of the Family 18 and Family 19 chitinases are shown in Supplementary Figures S1 and S2, respectively. We have identified four fragments ranging from 41%-100% identity to the DcChit1_1 fragment previously found by Renner and Specht in *D. capensis* genomic DNA[22] (Supplementary Figure S3). Several well-characterized reference sequences (e.g chitinases from *Vitis vinifera, Brassica napus,* and *Hordeum vulgare*) are also included for comparison. Using the characterization scheme of the carbohydrate-active enzymes (CAZy) database,[23,24] the chitinases investigated here belong to Family 18 (orange) or Family 19 (green). Overall, the sequence identity among the Family 18 chitinases from Caryophylalles carnivorous plants is much higher than that of Family 19, as illustrated in Figure 1A and B. These two types of chitinases have different folds and are thought to have evolved independently,,[25,26] consistent with their separation into separate clusters (Figure 1C). Family 18 contains types III and V, while types I, II and IV belong to Family 19[10].

D. capensis Chitinases are Predicted to Adopt Folds Consistent with Active Enzymes

Family 18 chitinases, which retain the β-anomeric carbon stereochemistry from the substrate to the product, adopt the $(\alpha\text{-}\beta)_8$ triosephosphateisomerase (TIM)-barrel fold,[28,29] shown for DCAP_0106 in Figure 2A. The *in silico* maturation process, which we have previously described for cysteine proteases,[19] is illustrated in Supplementary Figure S4. The active site (Figure 2B), consists of a characteristic DXXDXDXE motif.[28,29] The “tunnel” containing the active site is shaped by an unusual structural feature, two non-proline cis peptide bonds that are highly conserved, although the particular residues involved are somewhat variable.[3,30] The cis peptide bonds (shown in black in Figure 2C), are captured by the molecular models for all full-length Family 18 chitinases examined here. The shape of the tunnel and the surface formed by the aromatic rings opposite the catalytic D and E residues acts to guide the chitin polymer chains into the active site, leading to processive activity.[31] The ability of Family 18 chitinases to keep the strand that is currently being degraded from re-encountering solid substrate is thought to be a key determinant of their ability to hydrolyze crystalline polysaccharides.[32]

The Family 19 chitinases, all of which are characterized by an anomeric inverting mechanism,[33] have diverse structural features. Much of the structural and functional diversity results from two highly variable regions, the C-rich chitin-binding domain and the P-rich hinge,[34,35] each of which may vary in length or be absent altogether. We have identified two class I chitinases (DCAP_4817 and DCAP_5513) and one class IV chitinase (DCAP_0533) from the *D. capensis* genome. Most of the sequences in this set contain N-

terminal secretion signals, however two *D. spatulata* sequences (Q6IVX2_9CARY and Q6IVX4_9CARY) and the reference sequence CHI2_BRANA contain short C-terminal extensions indicating targeting to the vacuole, consistent with their playing a purely defensive role. One sequence each from *D. capensis* (DCAP_5513), *D. rotundifolia* (Q6IV09_DRORT), and *D. spatulata* (Q6DUK0_9CARY) is missing one or more critical active site residues; in other organisms, enzymatically non-functional chitinase homologs are often present and can serve as chitin-binding proteins.[36] The predicted structure after *in silico* maturation for a representative chitinase, VF-1 from *D. muscipula* (Figure 2) is in good agreement overall with the homology model of Paszota et al.,[10] with the active site residues positioned in a shallow cleft on the surface of the active domain. The two models do differ in the relative orientations of the domains; however examination of the other models in this set suggests that the P-rich hinge is highly flexible (Supplementary Figure S5).

Because sequence identity between our targets and proteins with solved structures is only moderate (in the range of 30-50 %), comparative modeling with all-atom refinement was used. The starting structures are predicted using the Robetta implementation[37] of Rosetta[38]. This approach uses a combination of fragment homology and de novo structure prediction, and is regularly validated via CAMEO[39]. Our modeling approach, in which the starting Rosetta structures are subjected to in silico maturation, was previously validated experimentally when the x-ray structure of a cysteine protease we had previously predicted was solved. The crystal structure of Dionain 1 (PDB ID 5A24),[40] shows excellent agreement with our predicted structure, with the prediction capturing all major secondary structural elements and exhibiting only minor deviations in the flexible loop regions.[19] For the chitinases, fragment homology was the primary method used. Sequence alignments for the target molecule with all of the template sequences used by Rosetta are shown for representative members of Family 18 and Family 19 in Supplementary Figures S6 and S7, respectively. For DCAP_2209 (Family 18), excluding the N-terminal signal sequence, 100% of the sequence aligns with homologous regions in the 11 template sequences (tabulated in Supplementary Table S1). For DCAP_5513, excluding the N-terminal signal sequence, only one 6-residue stretch of the P-rich region is not directly homologous to at least one of the template sequences (tabulated in Supplementary Table S2). As a further validation, a blind structure prediction was performed for the reference sequence HORV2, in which the actual pdb structure of this molecule (PDBID 1CNS, 2BAA)[41] was excluded from the template set. The predicted and experimental structures are shown overlaid in Supplementary Figure S8. After equilibration, the backbone RMSD between these structures was 1.01 Å. All major secondary structure elements are reproduced, with only minor differences in relative orientation as well as some deviation in the loops and termini.

The Class IV Chitinase DCAP_0533 Has Two Functional Domains

We have identified a new class IV chitinase from *Drosera capensis,* DCAP_0533. A class IV chitinase has previously been described as one of the most abundant proteins in the pitcher fluid *N. alata*,[16] where it preferentially hydrolyzes small GlcNAc oligomers over larger polymeric substrates.[42] Unlike other known plant chitinases, DCAP_0533 contains two class IV catalytic domains. The N-terminal domain appears to be fully active, while the C-

terminal domain lacks one of the active residues but containts a full complement of substrate-binding residues(Figure 2E, Supplementary Figures S9-S10). Multidomain chitinases containing dedicated substrate-binding domains have previously been observed in microbes.[43] For example, ChiA from the thermophilic archeon *Pyrococcus kodakaraensis,* has two chitinase domains and three catalytically inactive substrate binding domains, allowing separate optimizaton of substrate binding and catalytic function. [44] AFM data suggests the binding is mostly determined by interaction of the aromatic residues in the binding site (orange in Figure 2E) with the pyranose rings of the substrate.[45] This type of functionality has not been previously observed in plants; we hypothesize that it is an adaptation associated with carnivory, perhaps related to more effective breakdown of small oligosaccharides to components that can be used as a nitrogen source.

Network Analysis Shows Substantial Topological Differences by Family and within Proteins

When selecting potential targets for biophysical characterization, it is useful to consider general patterns of structural similarity or difference within and between families that may correlate with functional differences. Protein structure networks are useful for this purpose, as they directly encode the potential for direct physical interaction between functional groups (rather than representing detailed structure through properties such as side chain dihedral angles that can often vary substantially and dynamically without impacting protein function). Here we employ the PSN representation of Benson and Daggett,[47] where vertices represent small moieties and edges represent the potential for direct interaction (as determined by moiety-specific proximity constraints). Given two or more such PSNs, we may compare their topology by the structural distance method of,[48] identifying the smallest number of edge changes (i.e. altered inter-moiety interactions) needed to make one PSN isomorphic to the other. Figure 3 depicts respective hierarchical clusterings of the Family 18 (panel A) and Family 19 (panel B) chitinases based on this notion of structural similarity, with distances normalized by the number of vertices to yield a metric with units of average changed interactions per moiety. For Family 18, the pattern of topological similarity is strikingly close to the pattern of sequence similarity, although somewhat more diversity can be seen among structures than among sequences (compare with Figure 1). By contrast, topological clustering of Family 19 chitinases shows substantial differences from the sequence-based clustering. For instance, while DCAP_0533, A9ZMK1_NEPAL, and Q6IV09_DRORT belong to an outlying but internally cohesive cluster with respect to sequence similarity, the three show markedly different topologies (and, indeed, are split between the two large structural clusters characterizing the family). More broadly, we find that the Family 19 chitinases divide structurally into two primary clusters (rather than the four obtained from sequence similarity), both of which are internally heterogeneous and neither of which maps cleanly onto the clusters found by sequence similarity. The relationship between sequence and structure is thus much more tightly coupled for Family 18 than Family 19.

Further insight into the structural differences between the two families can be obtained by considering variation in the properties of their respective PSNs. Here, we examine four basic graph-level indices (GLIs) related to protein network organization. *Transitivity*[49] is defined

as the fraction of *(i,j,k)* two-paths for which there exists an *(i,k)* edge, and is a standard measure of triadic closure; in the PSN context, higher levels of transitivity are associated with structures that are closely and uniformly packed, with few cavities or extended regions. *Degree* is defined as the number of edges incident on a given vertex; for a PSN, this corresponds to the number of other moieties with which a given chemical group is in contact. The standard deviation of the degree distribution within a PSN then provides a measure of the level of heterogeneity in local packing around chemical groups, and we employ it here as a second GLI. At a somewhat less local level, the (degree) *core number* of a given vertex[50] provides a measure of the extent to which that vertex is embedded in a region of high cohesion within the graph. More precisely, the k-th core (or k-core) of a graph is defined as the maximum set of vertices having at least k neighbors within the set. The core number of a vertex is then the number of the highest-order k-core to which it belongs. Although each k-core is not necessarily cohesive as a whole, cores with $k \geq 2$ are composed of *unions* of cohesive subgraphs, such that all vertices with high core numbers necessarily belong to highly cohesive subgroups. In a PSN context, cohesive subgroups of moieties are joined by multiple, redundant paths and cannot be pulled apart without severing large numbers of edges. At the level of the entire PSN, then, the standard deviation of the core number serves as an indicator of the degree of heterogeneity in structural cohesion, and distinguishes between highly organized structures and structures that combine rigidly and loosely bound regions. Finally, we consider an indicator of the global path structure within the PSN, which we call *M-eccentricity*. The *eccentricity* of a vertex is the maximum geodesic distance from that vertex to any other vertex in the graph;[51] we here refer to the corresponding mean geodesic distance as the M-eccentricity. Vertices with high M-eccentricity are on average peripheral to the graph structure, while those with low M-eccentricity are relatively centrally located. At the level of the PSN as a whole, the standard deviation of the M-eccentricity distinguishes between uniformly globular structures and structures with deformations or other elongations, and we employ it as our fourth GLI.

Panel C of Figure 3 shows the distribution of the above GLI values for both chitinase families. All GLIs were calculated using the sna library;[52] to facilitate visualization, each GLI was standardized across the combined set of PSNs by subtracting the mean and dividing by the standard deviation prior to analysis. As is clear from Figure 3, the two families differ markedly on these four characteristics. On average, the Family 18 structures are substantially more homogeneous with respect to extended structure, local packing, and cohesion, while also being less transitive ($p < 0.001$ for all measures, two-tailed t-test). With respect to variation within family, the Family 18 structures show significantly less variability in eccentricity heterogeneity and transitivity (permutation test of logged IQR ratios, respective p values < 1e – 5 and 0.015), but more comparable variability with respect to heterogeneity in local packing and cohesion (respectively $p = 0.073$ and $p = 0.066$, not significant).

To provide an intuition for how these patterns play out in specific cases, Figure 4 shows vertex-level core numbers and M-eccentricity scores for the structures of CF821_NEPMI (Family 18) and DCAP_5513 (Family 19). These structures have low median distance to each other structure in the family, and are hence broadly representative of the classes in question. The core number visualizations of panels (a) and (b) clearly show that

CF821_NEPMI is dominated be a large and uniformly cohesive core region, with few vertices in the outer region (i.e., lower cores). By contrast, the highly irregular structure of DCAP_5513 has numerous areas of low cohesion (including much of the C-rich domain) as well as the highly cohesive region associated with the central helices (compare with Figure 2). Differences in global structure are brought into sharp relief by the M-eccentricity visualizations of panels (c) and (d). The uniform and tightly connected topology of CF821_NEPMI results in a large number of vertices with short path distances to nearly all other chemical groups in the protein, and relatively little overall variation. Moieties in DCAP_5513, on the other hand, may be at an average distance of more than 9 steps from the rest of the protein, with large differences between the relatively central vertices in the helical region and those in the outer portions of the C-rich domain or the P-rich hinge.

Taken together, these findings suggest substantial structural differences in the basic organization of the Family 18 and Family 19 chitinases, with the former having more internally homogeneous structures, and with structural differences being more closely related to differences in sequence. Family 19 is on the whole more diverse, and contains members that are on average less internally homogeneous. The presence of a higher volume of low-cohesion regions in the Family 19 chitinases suggests that these enzymes may be more prone to thermal denaturation than those in Family 18 (since low-cohesion regions require fewer disrupted edges to pull apart), but may also have functional significance (e.g., by allowing enhanced flexibility). Such structural insights from PSN topology complement those gained by studying specific features, and are more easily extended to analyzing large numbers of sequences.

Materials and Methods

Sequence Alignment and Prediction of Putative Protein Structures

Sequences were aligned with ClustalOmega[53] (gap open penalty = 10.0, gap extension penalty = 0.05, hydrophilic residues = GPSNDQERK, weight matrix = BLOSUM). Secretion signal sequences were predicted using SignalP 4.1.[54] Structure prediction was performed as in.[18] In the first stage, the Robetta[37] implementation of Rosetta[38] was used to produce an initial model for each protein. In the second stage, the model was subjected to "*in silico* maturation." Signal peptides were removed, and disulfide bonds identified by a combination of homology and distance constraints. Protonation states of active site residues were corrected to match literature values where necessary; for Family 18 chitinases, we approximate the sharing of a proton between active site residues D1 and D2 by protonation of D1 (which results in realistic side chain orientations and preserves the attractive interaction between D1 and D2). In the third and final stage, we equilibrated each matured enzyme model in explicit solvent (TIP3P water[55]) under periodic boundary conditions using NAMD.[56] Simulation was performed using the CHARMM36 forcefield,[57] with each model being energy-minimized for 10,000 iterations and then simulated at 293K for 500ps; the final protein conformation was retained for subsequent analysis. For the one reference sequence for which a structure was available (HORV2, PDB ID 2BAA,[58]), this was used as the initial starting model (following removal of heteroatoms and protonation using

REDUCE[59]). PDB files corresponding to the equilibrated structures for all the proteins discussed in this manuscript are available in the supplemental information.

Network Modeling and Analysis

We mapped each equilibrated protein structure to a protein structure network (PSN) as defined by the representation of[47] using software tools from;[18] these in turn make use of VMD[60] and the statnet toolkit[61,62] within the R statistical computing system.[63] To compare PSNs, we use the structural distance approach of,[48] which defines a metric on graph pairs that is in our case equal to the number of edges in one graph that would need to be altered in order to make it isomorphic to the other. (Isolate addition was performed when comparing graphs with differing numbers of vertices.) To remove size effects, the raw distance between each pair of PSNs was normalized by the number of vertices, yielding a metric corresponding to edge changes per vertex. These normalized structural distances were analyzed using hierarchical clustering using R. Additional network analysis and visualization was performed using the network and sna libraries within statnet.[52,62]

Conclusion

Modeling and analysis of Family 18 and 19 chitinases from *D. capensis* and several related species reveal a number of novel enzymes that present promising targets for subsequent expression and biophysical characterization. These include what is to our knowledge the first plant chitinase found with multiple active domains, as well as several proteins that differ in more conventional ways from others in their class. Comparative network analysis of these structures reveals within- and between-family differences in structural properties, with Family 18 chitinases tending to be substantially more homogeneous in internal structure and Family 19 chitinases showing variation in cohesion and packing with possible implications for both function and thermal stability. These results also demonstrate the potential of *in silico* pipelines to move rapidly from genomic DNA to predictions of tertiary structure and comparative analysis thereof. As the “genomic revolution” makes such data available at an ever-increasing rate, such pipelines will become critical to our ability to exploit this scientific resource.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was made possible, in part, through access to the Genomic High Throughput Facility Shared Resource of the Cancer Center Support Grant (CA-62203) at the University of California, Irvine and NIH shared instrumentation grants 1S10RR025496-01 and 1S10OD010794-01; this research was also supported by NSF award DMS-1361425. K.N.E was supported by the UCI UROP program. S.T. and R.W.M. acknowledge the California State Summer School for Math & Science (COSMOS) and NSF grant CHE-1308231.

References

(1). Khoushab F; Yamabhai M Chitin Research Revisited. Marine Drugs 2010, 8, 1988–2012. [PubMed: 20714419]

(2). Rathore A; Gupta R Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives. Enzyme Research 2015, 2015, 1 – 8.

(3). Masuda T; Zhao G; Mikami B Crystal structure of class III chitinase from pomegranate provides the insight into its metal storage capacity. Bioscience, Biotechnology, and Biochemistry 2015, 79, 45–50.

(4). Vega K; Kalkum M Chitin, chitinase responses, and invasive fungal infections. International Journal of Microbiology 2012, 2012, Article ID 920459.

(5). Mack I; Hector A; Ballbach M; Kohlhäufl J; Fuchs KJ; Weber A; Mall MA; Hartl D The role of chitin, chitinases, and chitinase-like proteins in pediatric lung diseases. Molecular and Cellular Pediatrics 2015, 2, DOI 10.1186/s40348-015-0014-6.

(6). Busam G; Kassemeyer H-H; Matern U Differential Expression of Chitinases in Vitis vinifera L. Responding to Systemic Acquired Resistance Activators or Fungal Challenge. Plant Physiol. 1997, 15, 1029–1038.

(7). Karmakar S; Molla K; Chanda P; Sarkar S; Datta S; Datta K Green tissue-specific co-expression of chitinase and oxalate oxidase 4 genes in rice for enhanced resistance against sheath blight. Planta 2016, 243, 115–130. [PubMed: 26350069]

(8). Eilenberg H; Pnini-Cohen S; Schuster S; Movtchan A; Zilberstein A Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. Journal of Experimental Botany 2006, 57, 2775–2784. [PubMed: 16829546]

(9). Rottloff S; Stieber R; Maischak H; Turini FG; Heubl G; Mithöfer A Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. Journal of Experimental Botany 2011, 62, 4639–4647. [PubMed: 21633084]

(10). Paszota P; Escalante-Perez M; Thomsen LR; Risør MW; Dembski A; Sanglas L; Nielsen TA; Karring H; Thøgersen IB; Hedrich R et al. Secreted major Venus flytrap chitinase enables digestion of arthropod prey. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 2014, 1844, 374–383. [PubMed: 24275507]

(11). Pavlovic A; Krausko M; Adamec L A carnivorous sundew plant prefers protein over chitin as a source of nitrogen from its traps. Plant Physiology and Biochemistry 2016, 104, 11–16. [PubMed: 26998942]

(12). Adamec L Leaf absorption of mineral nutrients in carnivorous plants stimulates root nutrient uptake. New Phytolologist 2002, 2155, 89–100.

(13). Pavlovic A; Krausko M; Libiakova M; Adamec L Feeding on prey increases photosynthetic efficiency in the carnivorous sundew *Drosera capensis*. Ann. Bot. 2014, 113, 69–78. [PubMed: 24201141]

(14). Juniper B; Robins R; Joel D The Carnivorous Plants; Academic Press: London, UK, 1989.

(15). Matusíková I; Salaj J; Moravcíková J; Mlynárová L; Nap J; Libantová J Tentacles of in vitro-grown round-leaf sundew (Drosera rotundifolia L.) show induction of chitinase activity upon mimicking the presence of prey. Planta 2005, 222, 1020–1027. [PubMed: 16049675]

(16). Hatano N; Hamada T Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. Journal of Proteomics 2012, 75, 4844–4852. [PubMed: 22705321]

(17). Buch F; Rott M; Rottloff S; Paetz C; Hilke I; Raessler M; Mithöfer A Secreted pitfall-trap fluid of carnivorous Nepenthes plants is unsuitable for microbial growth. Annals of Botany 2013, 111.

(18). Butts CT; Bierma JC; Martin RW Novel proteases from the genome of the carnivorous plant *Drosera capensis*: structural prediction and comparative analysis. Proteins: Structure, Function, and Bioinformatics 2016, *in press,* doi:10.1002/prot.25095.

(19). Butts CT; Zhang X; Kelly JE; Roskamp KW; Unhelkar MH; Freites JA; Tahir S; Martin RW Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, Drosera capensis. Computational and Structural Biotechnology Journal 2016, *in press*.

(20). Campbell M; Law M; Holt C; Stein J; Moghe G; Hufnagel D; Lei J; Achawanantakun R; Jiao D; Lawrence CJ et al. MAKER-P: A Tool-kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiology 2013, 164, 513–524. [PubMed: 24306534]

(21). Quevillon E; Silventoinen V; Pillai S; Harte N; Mulder N; Apweiler R; Lopez R InterProScan: Protein Domains Identifier. Nucleic Acids Research 2005, 33, W116–W120. [PubMed: 15980438]

(22). Renner T; Specht CD Molecular and functional evolution of Class I chitinases for plant carnivory in the caryophyllales. Molecular Biology and Evolution 2012, 29, 2971–2985. [PubMed: 22490823]

(23). Cantarel BL; Coutinho PM; Rancurel C; Bernard T; Lombard V; Henrissat B The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Research 2009, 37, D233–D238. [PubMed: 18838391]

(24). Lombard V; Ramulu HG; Drula E; Coutinho PM; Henrissat B The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 2014, 42, D490–D495. [PubMed: 24270786]

(25). Merzendorfer H The cellular basis of chitin synthesis in fungi and insects: Common principles and differences. European Journal of Cell Biology 2011, 90, 759 – 769. [PubMed: 21700357]

(26). Lukas Hartl SZ; Seidl-Seiboth V Fungal chitinases: diversity, mechanistic properties and biotechnological potential. Appl Microbiol Biotechnol 2011, 533 – 543. [PubMed: 22134638]

(27). Székely GJ; Rizzo ML Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. Journal of Classification 2005, 22, 151–183.

(28). Kesari P; Patil DN; Kumar P; Tomar S; Sharma AK; Kumar P Structural and functional evolution of chitinase-like proteins from plants. PROTEOMICS 2015, 15, 1693–1705. [PubMed: 25728311]

(29). Van Aalten D; Komander D; Synstad B; Gåseidnes S; Peter M; Eijsink V Structural insights into the catalytic mechanism of a family 18 exo-chitinase. Proceedings of the National Academy of Sciences 2001, 98, 8979–8984.

(30). Terwisscha van Scheltinga A; Kalk K; Beintema J; B.W. D Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure 1994, 2, 1181–1189. [PubMed: 7704528]

(31). Horn SJ; Sikorski P; Cederkvist JB; Vaaje-Kolstad G; Sórlie M; Synstad B; Vriend G; Vårum KM; Eijsink VGH Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. Proceedings of the National Academy of Sciences of the United States of America 2006, 103, 18089–18094. [PubMed: 17116887]

(32). von Ossowski I; J. S; Koivula A; Piens K; Becker D; Boer H; Harle R; Harris M; Divne C; Mahdi S et al. Engineering the exo-loop of Trichoderma reesei cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. Journal of Molecular Biology 2003, 333, 817–829. [PubMed: 14568538]

(33). T. F Chitinolytic enzymes: catalysis, substrate binding, and their application. Current Protein and Peptide Science 2000, 1, 105–124. [PubMed: 12369923]

(34). Meins F; Fritig B; Linthorst H; Mikkelsen J; Neuhaus J; Ryals J Plant chitinase genes. Plant Molecular Biology Reporter 1994, 12, S22–S28.

(35). Neuhaus J; Fritig B; Linthorst H; Meins F; Mikkelsen J; Ryals J A revised nomenclature for chitinase genes. Plant Mol Biol Rep 1996, 14, 102–104.

(36). Purushotham P; Arun PVPS; Prakash JSS; Podile AR Chitin binding proteins act synergistically with chitinases in *Serratia proteamaculans* 568. PLoS ONE 2012, 7, e36714. [PubMed: 22590591]

(37). D.E. K; Chivian D; Baker D Protein Structure Prediction and Analysis Using the Robetta Server. Nucleic Acids Research 2004, 32, W526–31. [PubMed: 15215442]

(38). Raman S; Vernon R; Thompson J; Tyka M; Sadreyev R; Pei J; Kim D; Kellogg E; DiMaio F; Lange O et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins 2009, 77, 89–99. [PubMed: 19701941]

(39). Haas J; Roth S; Arnold K; Kiefer F; Schmidt T; Bordoli L; Schwede T The Protein Model Portal - a comprehensive resource for protein structure and model information. Database (PMID: 23624946), 2013.

(40). Ris/or MW; Thomsen LR; Sanggaard KW; Nielsen TA; Th/ogersen IB; Lukassen MV; Rossen L; Garcia-Ferrer I; Guevara T; Scavenius C et al. Enzymatic and structural characterization of the

major endopeptidase in the Venus flytrap digestion fluid. The Journal of Biological Chemistry 2016, 291, 2271–2287. [PubMed: 26627834]

(41). Song HK; Suh SW Refined structure of the chitinase from barley seeds at 2.0 Åresolution. Acta Crystallographica D: Biological Crystallography 1996, 52, 289–298. [PubMed: 15299702]

(42). Ishisaki K; Honda Y; Taniguchi H; Hatano N; Hamada T Heterogonous expression and characterization of a plant class IV chitinase from the pitcher of the carnivorous plant *Nepenthes alata*. Glycobiology 2012, 22, 345–351. [PubMed: 21930651]

(43). Talamantes D; Biabini N; Dang H; Abdoun K; Berlemont R Natural diversity of cellulases, xylanases, and chitinases in bacteria. Biotechnology for Biofuels 2016, 9, DOI: 10.1186/s13068-016-0538-6.

(44). Tanaka T; Fujiwara S; Nishikori S; Fukui T; Takagi M; Imanaka T A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. Applied and Environmental Microbiology 1999, 65, 5338–5344. [PubMed: 10583986]

(45). Kikkawa Y; Fukuda M; Kashiwada A; Matsuda K; Kanesato M; Wada M; Imanaka T; Tanaka T Binding ability of chitinase onto cellulose: an atomic force microscopy study. Polymer Journal 2011, 43, 742–744.

(46). Si J; Yan R; Wang C; Zhang Z; ; Su X TIM-Finder: A new method for identifying TIM-barrel proteins. BMC Structural Biology 2009, 9, doi:10.1186/1472-6807-9-73.

(47). Benson NC; Daggett V A chemical group graph representation for efficient high-throughput analysis of atomistic protein simulations. Journal of Bioinformatics and Computational Biology 2012, 10, 1250008. [PubMed: 22809421]

(48). Butts CT; Carley KM Some Simple Algorithms for Structural Comparison. Computational and Mathematical Organization Theory 2005, 11, 291–305.

(49). Wasserman S; Faust K Social Network Analysis: Methods and Applications; Cambridge University Press: Cambridge, 1994.

(50). Seidman SB Network Structure and Minimum Degree. Social Networks 1983, 5, 269–287.

(51). West DB Introduction to Graph Theory; Prentice Hall: Upper Saddle River, NJ, 1996.

(52). Butts CT Social Network Analysis with sna. Journal of Statistical Software 2008, 24.

(53). Sievers F; Wilm A; Dineen D; Gibson TJ; Karplus K; Li W; Lopez R; McWilliam H; Remmert M; Söding J et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011, 7, 539–539, 21988835[pmid]. [PubMed: 21988835]

(54). Petersen T; Brunak S; von Heijne G; Henrik Nielsen H SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 2011, 8, 785–786. [PubMed: 21959131]

(55). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics 1983, 79, 926–935.

(56). Phillips JC; Braun R; Wang W; Gumbart J; Tajkhorshid E; Villa E; Chipot C; Skeel RD; Kalé L; Schulten K Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 2005, 26, 1781–1802. [PubMed: 16222654]

(57). Best RB; Zhu X; Shim J; Lopes PEM; Mittal J; Feig M; MacKerell AD Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ, and side-chain χ_1 and χ_2 dihedral angles. Journal of Chemical Theory and Computation 2012, 8, 3257–3273. [PubMed: 23341755]

(58). Hart PJ; Pfluger HD; Monzingo AF; Hollis T; Robertus JD The refined crystal structure of an endochitinase from *Hordeum vulgare L.* seeds at 1.8Åresolutiontructure of an endochitinase from *Hordeum vulgare L.* seeds at 1.8Åresolution. Journal of Molecular Biology 1995, 248, 402. [PubMed: 7739049]

(59). Word JM; Lovell SC; Richardson JS; Richardson DC Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. Journal of Molecular Biology 1999, 285, 1735–1747. [PubMed: 9917408]

(60). Humphrey W; Dalke A; Schulten K VMD: visual molecular dynamics. Journal of Molecular Graphics 1996, 14, 33–38, 27–28. [PubMed: 8744570]

(61). Handcock MS; Hunter DR; Butts CT; Goodreau SM; Morris M statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. Journal of Statistical Software 2008, 24, 1–11. [PubMed: 18612375]

(62). Butts CT network: a Package for Managing Relational Data in R. Journal of Statistical Software 2008, 24.

(63). R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2015.

Highlights:

We report eleven new chitinases from the carnivorous plant *Drosera capensis*.

A novel two domain class IV chitinase similar to those found in microbes was found.

Protein structure prediction and comparison to other carnivorous plant chitinases reveals commonalities.

Sequence and structural motifs are conserved among carnivorous plant chitinases.

Protein structure networks reveal structural differences and predict functionality.

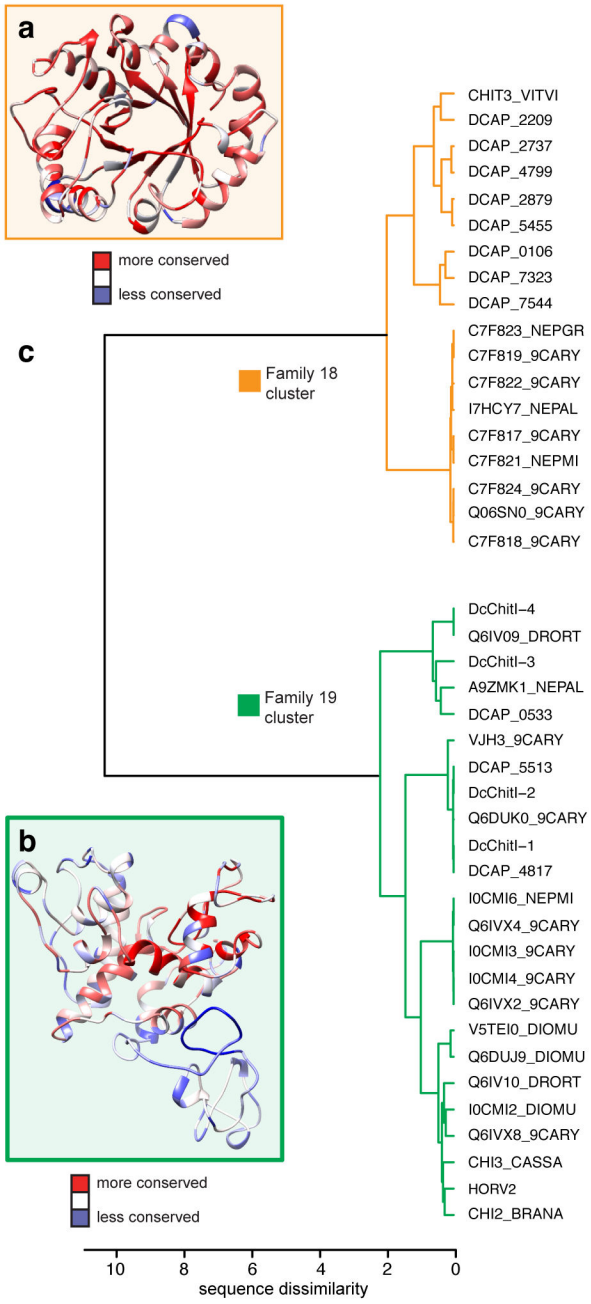

Figure 1:

Clustering of chitinases identified from the *D. capensis* genome, compared with those from other Caryophylalles carnivorous plants and well-characterized reference sequences. All of the sequences examined belong to GH Families 18 or 19. The sequence dissimilarity used here is the e-distance metric of Székely and Rizzo[27] (with $\alpha = 1$). This parameter is a weighted function of within-cluster similarities and between-cluster differences with respect to a user-specified reference metric, defined here as the raw sequence dissimilarity (1 - (%identity)/100).

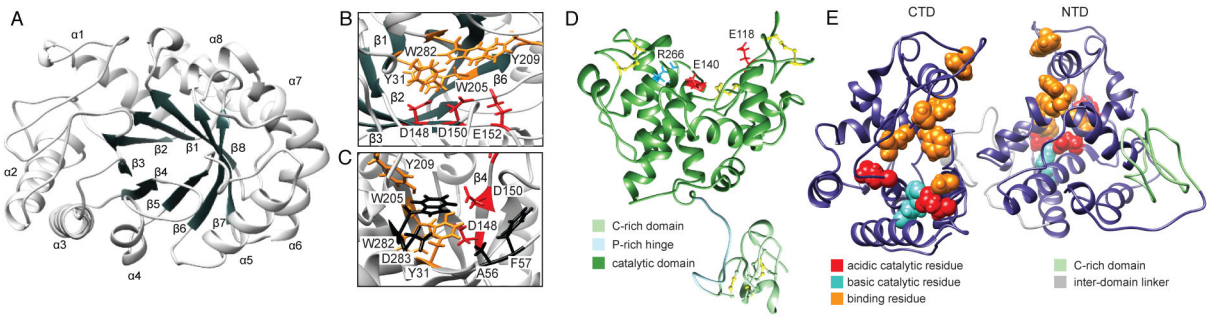

Figure 2:

Equilibrated structures of the mature sequences of chitinases from carnivorous plants. A. DCAP_0106, a representative Family 18 chitinase, after *in silico* maturation. Numbering of secondary structure elements follows the convention of Si et al.[46] B. Notably, the tunnel containing the active site has two surfaces with different chemical properties; the aromatic rings (orange) hold the more hydrophobic face of the chitin polymer in place, while the acidic residues (red) perform hydrolysis of the glycosidic linkages. C. Two conserved non-proline cis peptide bonds (black) are critical to shaping the active site tunnel in Family 18 chitinases. D. Chitinase VF-1 from *Dionaea muscipula* V5TEI0_DIOMU,[10] with important sequence features and active site residues labeled (red: acidic active residue. blue: basic active residue. yellow: disulfide bond). E. The two-domain chitinase DCAP_0533. Color coding is as in D, with the addition of substrate-binding residues in orange.

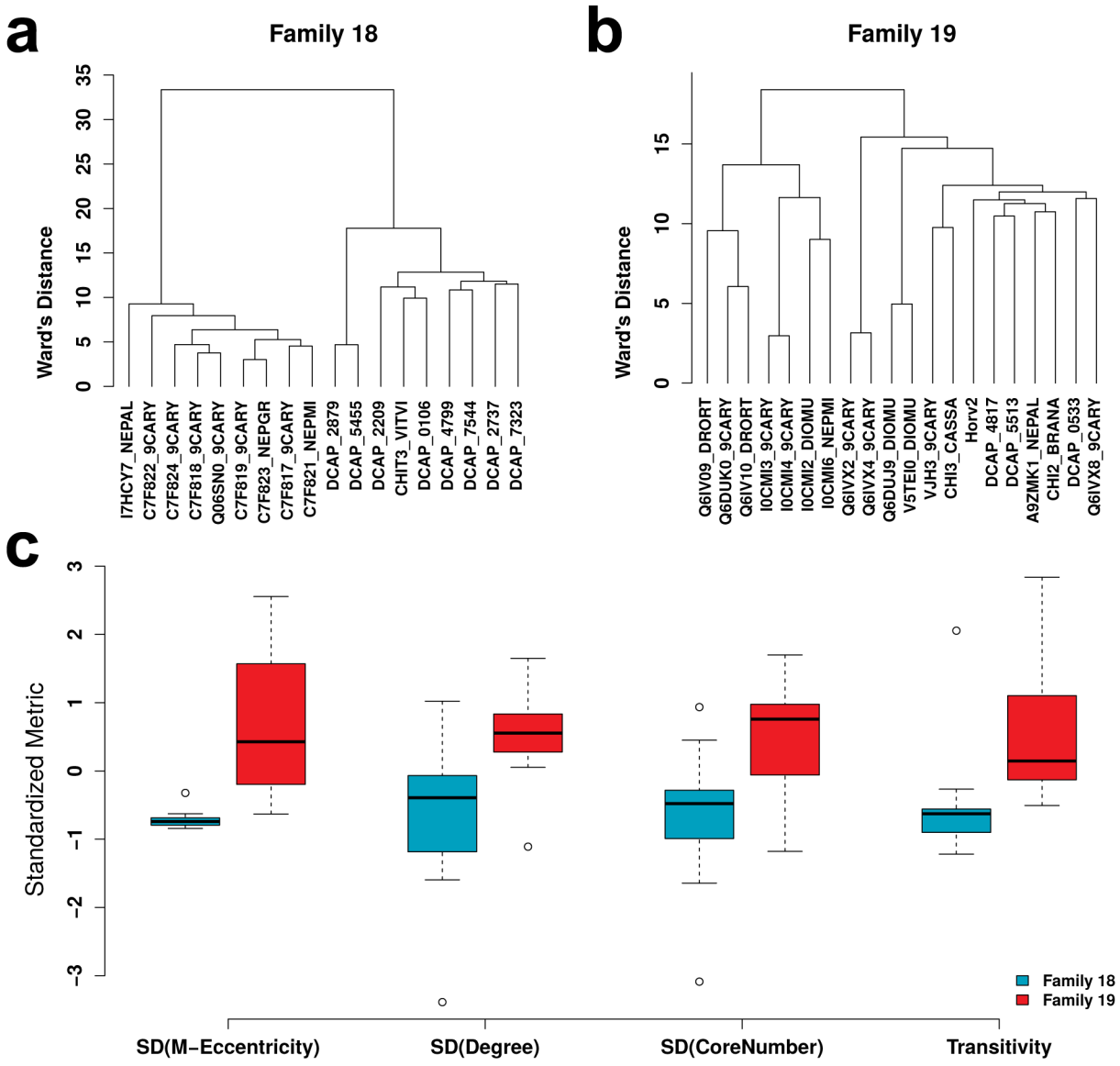

Figure 3:

(a)-(b) Within-family clustering of chitinases by normalized structural distances. Ward's method (in the generalization of[27]) was employed to construct a hierarchical clustering of Family 18 (a) and Family 19 (b) chitinases based on topological dissimilarity. Sequence similarity is broadly recapitulated by the structural distances in Family 18, while Family 19 shows distinct patterns of variation. Differences between families are large, as illustrated in (c), which shows distributions of M-eccentricity variation, degree variation, core number variation, and transitivity by family. Family 19 chitinases tend to be markedly more internally heterogeneous, with chemical groups whose local structural environments vary far more than their counterparts in Family 18. Family 19 chitinases also show a higher overall level of triadic closure, as captured by transitivity.

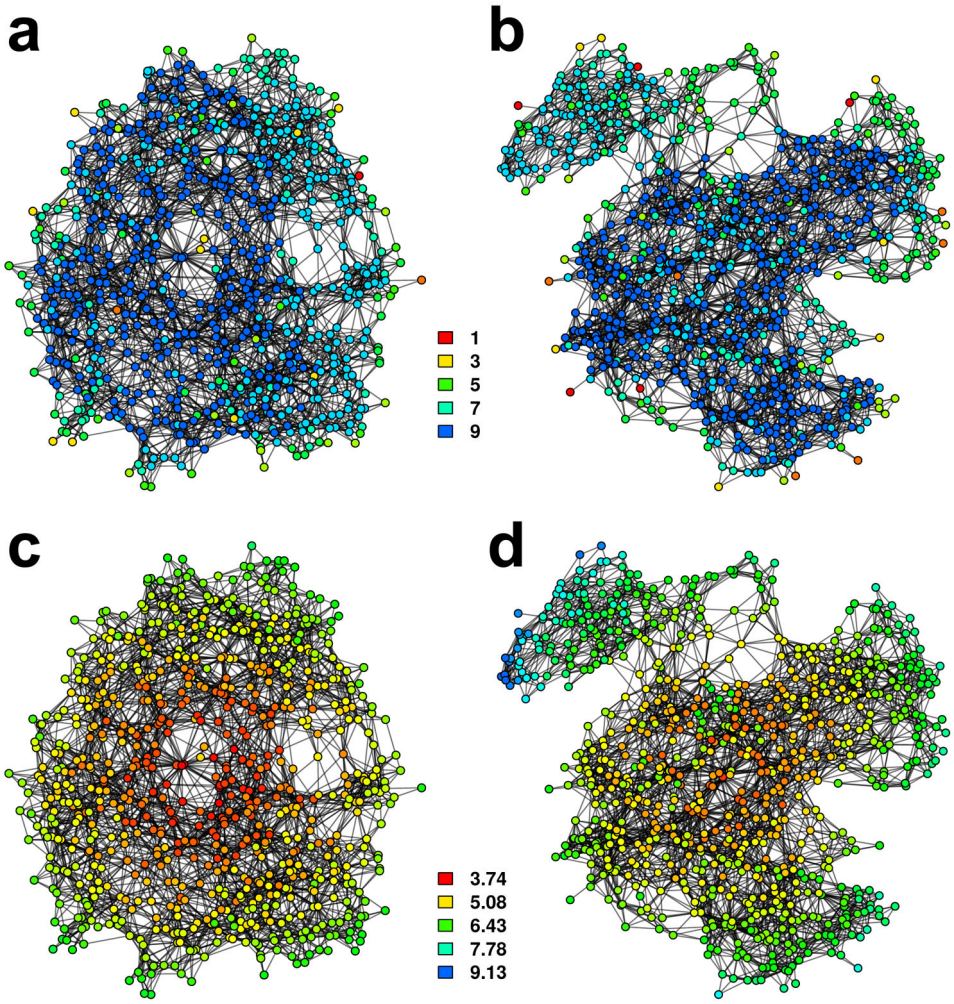

Figure 4:

PSN Visualizations for family-representative structures C7F821_NEPMI (Family 18, (a) and (c)) and DCAP_5513 (Family 19, (b) and (d)). In panels (a) and (b), vertices are colored by *k*-core number; vertices with higher core numbers are embedded in more strongly cohesive local structures. Panels (c) and (d) show vertices by M-eccentricity (with higher values indicating a higher mean distance to other vertices in the network). The much higher level of internal heterogeneity in DCAP_5513 versus C7F821_NEPMI is immediately evident, with the former containing complex and irregular structure that subjects some vertices to higher levels of both cohesion and proximity than others.