

Chemical and Molecular Composition of the Chrysalis Reveals Common Chitin-rich Structural Framework for Monarchs and Swallowtails

Nicolette F. Goularte,^{1‡} Till Kallem,^{2‡} and Lynette Cegelski^{2*}

¹Department of Structural Biology, 380 Roth Way, Stanford University School of Medicine, Stanford, California 94305

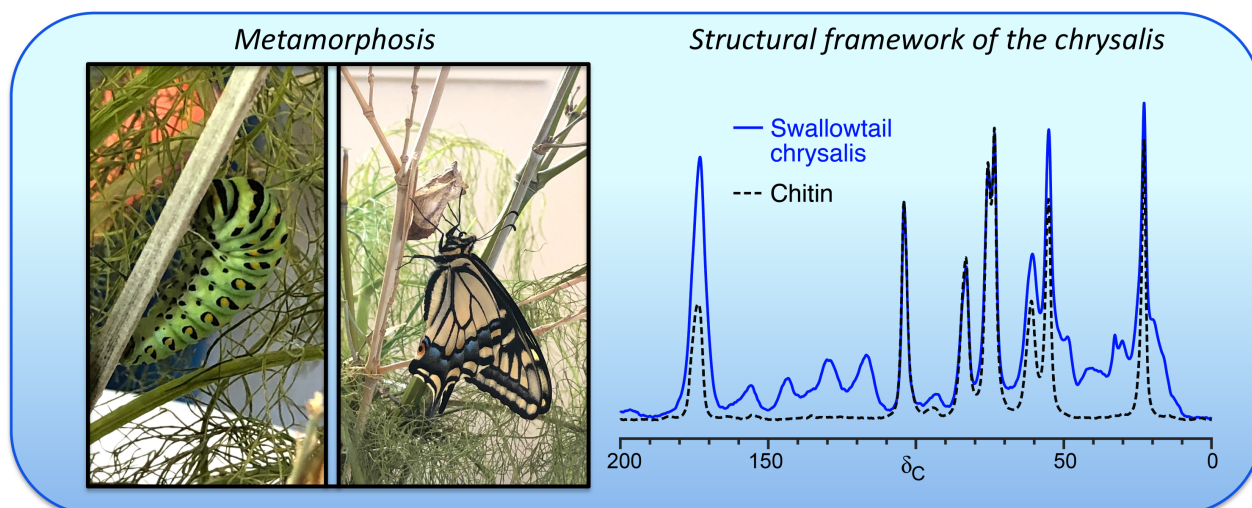
²Department of Chemistry, 380 Roth Way, Stanford University, Stanford, California 94305

*Corresponding author: cegelski@stanford.edu

‡These authors contributed equally to this work.

ABSTRACT

The metamorphosis of a caterpillar into a butterfly is an awe-inspiring example of how extraordinary functions are made possible through specific chemistry in nature's complex systems. The chrysalis exoskeleton is revealed and shed as a caterpillar transitions to butterfly form. We employed solid-state NMR to evaluate the chemical composition and types of biomolecules in the chrysalides from which Monarch and Swallowtail butterflies emerged. The chrysalis composition was remarkably similar between Monarch and Swallowtail. Chitin is the major polysaccharide component, present together with proteins and catechols or catechol-type linkages in each chrysalis. The high chitin content is comparable to the highest chitin-containing insect exoskeletons. Proteomics analysis of associated soluble proteins indicated the presence of chitinases that could be involved in synthesis and remodeling of the chrysalis as well as key cuticular proteins which play a role in the structural integrity of the chrysalis. The nearly identical ¹³C CPMAS NMR spectra of each chrysalis and similar structural proteins supports the presence of underlying design principles integrating chitin and protein partners to elaborate the chrysalis.



INTRODUCTION

Butterflies are admired for their beauty and serve as a spectacular hallmark of insect metamorphosis. The life cycle of insect species that undergo complete metamorphosis includes progression from egg through larva, followed by an intermediate pupal stage through which the larva undergoes a physical transformation into the adult form [1]. The chrysalis is the pupal form of the butterfly. To prepare for metamorphosis, a caterpillar will fasten itself to a surface and shed its skin to reveal the chrysalis exoskeleton which then hardens. After about 10 days, the transformation from chrysalis to butterfly is complete and the butterfly emerges (Figure 1). The empty pupal exoskeleton is also commonly referred to itself as the chrysalis, and we use this interchangeably with chrysalis exoskeleton. The chrysalis performs a crucial role in protecting the pupa during development, while allowing the subsequent emergence of the mature butterfly.

Relatively little is known about the chemical composition of the chrysalis, whereas considerable attention has been given to the moth cocoon, which is the very different shell spun of silk by moth caterpillars, and to the exoskeletons of insects including beetles, blowflies and cicadas [2-5]. Decades of research have also focused on the composition of the hard exoskeletons of crustaceans, including shrimp and lobsters [6, 7]. The notable common structural element found in the exoskeletons of both insects and crustaceans is chitin. Chitin is a homopolymer of β -(1-4)-linked N-acetyl-D-glucosamine and is the second most abundant polysaccharide in nature after cellulose [8]. Chitin is a major product used industrially, with most chitin harvested from shrimp and crab exoskeletons due to their supply and availability [8, 9]. Chitosan can be derived from chitin through deacetylation and is used widely in biomedical applications [8, 9]. Crustaceans generally produce the densely packed crystalline α -chitin [10], while many insect species employ the less crystalline γ -chitin in their exoskeleton [11, 12]. Insect cuticles such as the shed exoskeleton, termed exuviae, can contain up to 45% chitin [4, 13, 14]. In contrast, silkworm (*B. mori*) cocoons contain approximately 20% γ -chitin and, despite this low content, have been investigated as a natural source of γ -chitin [15].

Insect exoskeletons are highly insoluble macromolecular structures which pose a challenge to quantitative analysis using many conventional methods. Solid-state NMR has a rich history in examining exoskeleton of insects and, more recently, in characterizing insect wings and wing membranes [4, 16]. For example, Schaefer *et al.* investigated the Tobacco hornworm species, *M. sexta*, and determined that pupal cuticle contained 35% chitin, 31% protein, and 13% catechols through natural abundance ^{13}C cross-polarization magic-angle spinning (CPMAS) NMR [17]. Gullion *et al.* investigated the composition of cicada wing veins and membranes using ^{13}C CPMAS NMR and determined that veins contained more chitin, whereas membranes that are mechanically flexible contained more protein, including glycine-rich proteins [18]. Interestingly, the cicada wing membrane composition differed from that of the honeybee and butterfly wing membranes in terms of catechol and chitin content. The thinner honeybee and butterfly wing membranes exhibited much higher chitin content than the cicada wing membrane and contained carbon resonances associated with catechols ($\delta_{\text{C}} = 144$ ppm), while the cicada wing membrane lacked catechol-type carbons [19]. Alternative investigations of insect structures, such as the exuviae of blowfly species that did not use CPMAS NMR, required rigorous digestions involving

multiple rounds of acidic, alkaline, oxidative, and/or enzymatic treatments [13, 20]. These methods are valuable, but can underestimate components such as chitin that may not be fully hydrolyzed, wherein the quantification of free glucosamine following acid hydrolysis is often used to estimate chitin in the original material. In addition, some biomolecules are chemically modified and degraded during chemical and enzymatic digestion. Thus, for overall analysis of compositional pools in insoluble systems and desiring to distinguish and quantify protein and polysaccharide components, CPMAS NMR is a uniquely enabling and complementary spectroscopic tool.

Chitin-containing exoskeletons are usually associated with insoluble proteins. Considerable attention has been given to possible protein-chitin binding interactions and molecular crosslinks in insect exoskeletons, yielding models for how chitin and the cuticular proteins interact. Noncovalent interactions between chitin and various proteins are possible, wherein cuticular proteins can bind along the surface of chitin microfibrils [14, 21]. Cuticular proteins can be grouped into families by conserved protein sequence motifs, including the Rebers and Riddiford (RR) consensus sequence which is widely conserved in many cuticular proteins and has been shown to bind chitin [22, 23]. Covalent protein-catechol crosslinks have also been detected in insect cuticles [13, 14, 24, 25], and are implicated in insect cuticle sclerotization, the process through which the cuticle hardens and darkens [26]. CPMAS NMR, in particular, has uncovered the extent of sclerotization through detection of increased oxygenated aromatic carbons, such as catechols [17, 24, 27-32]. Catechols found in some insect exuviae have also been isolated after sample digestions and characterized by mass-spectrometry to attempt to identify specific quinone derivatives of catechols that may be covalently linked to cuticular proteins, although many of these remain in the insoluble framework [33]. Indeed, using ^{13}C CPMAS NMR and chemical digestions, Kramer et al. found that only about 1% of the total diphenol/catecholic content could be extracted in acid from sclerotized oothecae and cocoon silks, supporting the intractability and highly crosslinked nature of insect cuticle components [34]. Early isotopic labeling and NMR studies supported the role of histidine and lysine in post-translationally arylated crosslinks in blowfly exuviae [35]. Histidine-catechol crosslinks were confirmed in the Tobacco hornworm species *M. sexta* exuviae by solid-state NMR and mass spectrometry [17, 24, 25]. Some studies have provided evidence for tight chitin-protein associations due to protein-catechol crosslinks in the exuviae of *M. sexta* and sarcophagid species [13, 24].

By analogy to other insect exoskeleton studies, we considered that butterfly chrysalides are likely to contain chitin as well as catechols, but comparable evaluations are currently missing from the literature. Understanding how naturally produced materials function first requires an understanding of their composition and architecture. Our work here was directed to investigating the composition of the empty chrysalides after complete metamorphosis of two butterfly species, Swallowtail and Monarch. We employed ^{13}C CPMAS NMR to define the carbon composition of the chrysalides, integrated with proteomics to identify soluble and insoluble proteins associated with each sample. The chrysalides are highly abundant in chitin and we compare our compositional results with those from other relevant chitin-protein composite materials.

RESULTS AND DISCUSSION

We collected empty chrysalides after the metamorphosis of Monarch and Swallowtail caterpillars to examine the chemical composition in these unique macromolecular structures (Figure 1). A panel of experiments was performed to examine the overall chemical composition of the chrysalis framework and to identify proteins associated with each chrysalis. We first present the natural abundance ^{13}C CPMAS NMR results on the carbon composition and quantify the contribution of chitin to the chrysalides. The chrysalis is an insoluble material and we washed the empty chrysalides once with SDS detergent followed by water washes to remove any adventitiously associated proteins and lipids remaining from internal contact with the larva inside each chrysalis. In subsequent sections, we report on the low-abundance proteins that could be detected through proteomics analysis, and describe additional analysis of trypsin-digested and NaOH-digested samples.

Chrysalis Carbon Analysis by ^{13}C CPMAS NMR. The Monarch and Swallowtail chrysalis ^{13}C CPMAS spectra are exceedingly similar, indicating a high degree of similarity in the carbon composition for each chrysalis (Figure 2A). The major contribution of chitin to the chrysalis is apparent through the overlay of a chrysalis spectrum and the spectrum of commercially available chitin (MP Biomedicals, Solon, OH, USA) isolated from Crustacean shells (Figure 2B). The two spectra in Figure 2B were acquired at a magic-angle spinning (MAS) frequency of 10,056 Hz in order to shift the carbonyl spinning sideband (indicated with an asterisk) from 118 ppm (for 7143 Hz MAS as in Figure 2A) to a region with no other carbon intensity at 94 ppm. The unique peaks attributed to chitin that do not overlap with proteinaceous or lipid biomolecules are the C1 anomeric carbon and the C3, C4, and C5 sugar ring carbons. These are perfectly matched in the chitin and chrysalis spectra (Figure 2B), supporting the presence of chitin as the only major detectable polysaccharide. While the other four chitin carbons have overlapping contributions from other biomolecules, the major contribution of the C2, C6 and methyl carbons are apparent and consistent with the large contribution of chitin in the chrysalis, as further quantified in the subsequent section. Notable carbon regions in the chrysalis spectrum with contributions from proteins include: amino acid carbonyls (centered at 175 ppm); α -carbons between 50-65 and glycine α -carbons at 42 ppm; sidechain methylene and methyl carbons between 10-40 ppm. The unique sidechains (His, Arg, Trp, Tyr, and Phe) contribute intensity to the aromatic sp^2 carbon region between 110-170 ppm. Like the obvious chitin contribution to the chrysalis spectrum, the remarkable intensity in the aromatic region indicates the presence of catechols and catechol-like linkages. The salient chemical shifts associated with chitin as well as catechol and catechol-like linkages, such as those involving protein crosslinks, are provided in Figure 2C, and the detection of these resonances are compatible with studies in other insect systems [17, 18, 19, 24] and as tabulated previously [18, 19]. Specifically, we observed strong peak intensity at 144 ppm (Figure 2B). Carbons at 144 ppm are uniquely attributed to catecholic/diphenolic carbons, such as found in catechols and isodityrosine, whereas the hydroxyl carbon of tyrosine appears at 155 ppm [18, 19, 36]. The chrysalis spectra here reveal an even greater chitin contribution and greater catechol-type carbon intensities than the previously reported honeybee and butterfly wing membranes [19].

CPMAS NMR Analysis of Chitin Contribution to the Chrysalis. We estimated the chrysalis carbon mass attributed to chitin for each chrysalis by determining the percent of the total spectral

area of each chrysalis ^{13}C spectrum attributed to chitin. We also provide the results of quantitative CPMAS experiments that validate this determination. The chitin and chrysalis spectra for Monarch and Swallowtail were normalized to the C1 anomeric peak that is unique to chitin (104 ppm) and lacks any contributions from proteins or catechols. The spectral area of the chitin spectrum was then divided by that of each chrysalis, and the spectra obtained with the magic-angle spinning speed of 7143 Hz was used for this and the additional CPMAS experiments described below. This yielded a percent of chitin carbon mass of 41% for Monarch and 43% for Swallowtail. Additional NMR experiments were performed to validate using the CPMAS spectra to provide this estimate. As shown in Figure S1, all carbons exhibited similar relative intensities whether the spectrum was obtained with a 2 s recycle delay (as in Figure 2) or with a 10 s recycle delay.

Towards evaluation of total carbon areas, a single spectrum can be insufficient to permit quantitative accounting of carbon types. Specifically, there could be significant differences in CP efficiency from ^1H nuclei to ^{13}C nuclei, *i.e.* the buildup of magnetization, and/or in the relaxation dictated by $T_{1\rho}(\text{H})$, the proton spin-lattice relaxation time in the rotating frame [37, 38]. For chrysalides, like many proton-rich biological samples, the polarization build-up is much faster than relaxation ($T_{1s} \ll T_{1\rho}(\text{H})$), as seen in Figure 2D, and quantitative cross-polarization array experiments can be performed in order to account for the possibility of differences in CP dynamics and to determine the absolute carbon intensities [37, 38]. A series of CPMAS spectra were obtained using CP times ranging from 250 μs to 7 ms. Quantitative carbon contributions for selected carbons were obtained by using data from long CP times (4-7 ms) to calculate the magnetization according to $I_t = I_0 [\exp\{-t/T_{1\rho}(\text{H})\}]$ where I_0 is the maximum signal intensity, corresponding to the y-intercepts in Figure 2D. Integrated areas for the three perfectly resolved carbons in chitin and the resolved C1 chitin carbon in the two chrysalides are shown in Figure 1D, with the regions of integration defined in Figure S2. As anticipated, the quantitative carbon accounting for chitin yields a 1:1:1 ratio of carbonyl:C1:methyl carbons (Figure 2D). The CP behavior for chitin within the chrysalides as evidenced by inspection of the resolved and unique C1 peak, was identical to that in pure chitin (Figure 2D). We also present overlays of the full spectra for three of the long CP time acquisitions for each sample that reveal comparable reductions in carbon intensities across each whole spectrum (Figure 2E). Thus, the CPMAS analysis provides a valid quantitative estimate of chitin in chrysalides. The 40-45% chitin content determination is among the highest observed in insect systems and is comparable to higher estimates of chitin in exuviae of the Tobacco hornworm species *M. sexta* (35% chitin) [17], the sheep blowfly (*L. cuprina*) exuviae (40% chitin) [39], and larval cuticle of the blowfly species *C. vicina* (45% chitin) [40].

Chrysalis-associated SDS-soluble Proteins. Beyond the chitin and catechol-type contributions to the chrysalis, we aimed to extract at least partial identification of chrysalis-associated proteins. As described above, the CPMAS spectra in Figure 2 were obtained from chrysalides that were treated with 10% SDS for 1 hour at 95 °C to remove adventitiously associated proteins and lipids. Although the total mass of sample did not change appreciably, the wash did liberate some proteins, as detected by SDS PAGE analysis (Figure 3A) and analyzed by mass spectrometry-based proteomics analysis. The ^{13}C CPMAS spectral comparison for samples before and after the SDS wash is also provided in Figure 3B. These were overall similar with some loss of intensity

in carbonyl and methyl carbons after SDS washing, most likely reflecting the loss of lipids associated with the empty chrysalides after butterfly exit. We hypothesized that we might detect proteins associated with cuticle formation and chitinase enzymes, either for their direct role in chrysalis formation or remaining after formation of the butterfly wing. Indeed, chitin synthesis and chitin degradation are balanced during formation of insect wings (31, 32). For the proteomics analysis (Tables S1-S2), we utilized the Monarch (*Danaus plexippus*) and Old World Swallowtail (*Papilio machaon*) [41] reference genomes, with first sequences reported in 2011 and 2015, respectively. The analysis revealed a number of cuticular proteins as well as chitinases. Specifically, the SDS-soluble proteins associated with the Monarch chrysalis include Putative structural constituent of cuticle Gene KGM_211748 (88 kD), Cuticular protein RR-2 motif Gene KGM_200675 (26.6 kD), Cuticular protein RR-2 motif 76 Gene KGM_208061 (48.9 kD), a chitinase Gene KGM_205594 (50.2 kD), and additional cuticular proteins (Figure 2C and Table S1). The most abundant proteins detected from the Swallowtail chrysalis included Chitinase A Gene RR48_07375 (65.6 kD) and Endochitinase Gene RR48_07756 (63.4 kD); Paternally-expressed gene 3 Gene RR48_11805 (39 kD); and pupal cuticular proteins (Figure 2C and Table S2). Paternally-expressed gene 3 Gene RR48_11805 (39 kD) has 53.6% identity to the major Monarch protein identified in its analysis, Putative structural constituent of cuticle Gene KGM_211748 (88 kD) (BLAST E value $2.5e-91$), and thus is likely a cuticular protein. This analysis provides evidence of cuticular proteins and chitinase that were at least adventitiously bound to the chrysalis and may have been tightly associated. The detection of chitinase is compatible with the role of chitinases in chitin synthesis and remodeling of the chrysalis and/or of the butterfly wings.

Insoluble Proteins Identified through Partial Chrysalis Enzyme Digestion. Most of the proteins in the chrysalis are not SDS soluble, even after the one-hour incubation in 10% SDS at 95 °C, as performed above. Yet, we sought to perform additional digestions on the chrysalides to potentially identify some of the insoluble proteins that are more tightly assembled in the chrysalis. To identify SDS-insoluble proteins, we subjected each solid chrysalis pellet remaining after boiling SDS treatment to trypsin treatment to attempt to liberate peptides associated with the SDS-insoluble proteins that could be accessed by trypsin [20]. Detection by nano LC-MS/MS successfully identified many liberated peptides and we identified candidate associated proteins through a query with Byonic Advanced Peptide and Protein Identification Software [42] (Figure 3D and Tables S3 and S4).

This analysis revealed that half of the proteins liberated and identified above using SDS-treatment (Table S1 and S2) were also found in this insoluble protein analysis using trypsin digestion (Table S3 and S4), all of which were cuticle proteins or hypothesized cuticular proteins. These may be important structural components of the chrysalis with varying levels of cross-linking or association with chitin. In particular, the proteins with the highest probability score for Monarch and Swallowtail in both the SDS-soluble protein analysis and the tryptic digest analysis are the highly similar Putative structural constituent of cuticle Gene KGM_211748 (88 kD) in Monarch and the Paternally-expressed gene 3 Gene RR48_11805 (39 kD) in Swallowtail.

While all proteins identified by tryptic digest of the Monarch chrysalis had a known or hypothesized function, two of the top hits for the Swallowtail chrysalis were annotated as uncharacterized. We investigated these two uncharacterized proteins to determine if they exhibit sequence similarity to other known cuticular proteins. The first uncharacterized protein Gene RR48_13687 (35.5 kD) has high similarity to Cuticular protein hypothetical 7 Gene KGM_204902 (32.9 kD) of *Danaus plexippus* identified in our Monarch soluble and insoluble protein analysis (BLAST search E value $2.9\text{e-}88$). The second uncharacterized protein, Gene RR48_13682, also has similarity to a protein identified in the Monarch tryptic digest analysis, Putative cuticle protein Gene KGM_204899 (22.6 kD) of *Danaus plexippus* (BLAST search E value $6.0\text{e-}37$). Overall, five of the top eight proteins identified in this analysis for Swallowtail have similarity to top hits in the Monarch analysis and the similarity (E value and % identity) are indicated in Table S5.

NaOH Chemical Digestion and Underlying Chitin Framework. Finally, we subjected Monarch chrysalides to additional chemical treatment using boiling NaOH to completely digest all protein and leave only the insoluble chitin to collect by centrifugation [24]. The resulting spectrum is shown in Figure 3E and reveals the underlying insoluble chitin as an overlay with the commercially available chitin used here. The additional small 30-35 ppm peaks observed in the chrysalis chitin remaining from NaOH digestion have been observed in ^{13}C CPMAS spectra of other chemically digested and purified chitin samples, even commercially available chitin from other sources, perhaps due to small molecules tightly trapped in the packing of chitin [11, 43].

Summary. Chitin exists in structural frameworks throughout nature. Moth cocoons, such as Silkworm (*B. mori*), contain approximately 20% chitin [15]. However, moths integrate and spin silk proteins into their protective cocoons and molt inside of them [2], while butterflies do not employ silk for this purpose. We reveal that butterfly chrysalides have a much higher chitin content, similar to that of other insect exoskeletons. Importantly, there is a similar ratio of chitin and protein and catechol type linkages in chrysalides of the Swallowtail and Monarch. There are also common cuticular proteins associated with both chrysalides and chitinases found in association with empty chrysalides. These chemical and molecular compositional comparisons reveal fundamental chemistry and evidence of the structural blueprints that underlie the protective properties that a chrysalis affords the butterfly.

MATERIALS AND METHODS

Sample collection

Monarch butterflies were raised and released by Vanessa Hulan at certified Monarch Waystation 31377 and resulting chrysalides were collected with photographs provided in Figure 1.

Swallowtail caterpillars were donated by Nancy McCarthy and raised and released by Jillian Kanan and the resulting chrysalis was collected with photographs provided in Figure 1.

Solid-state NMR measurements

Whole chrysalis samples were ground by mortar and pestle. SDS and NaOH treated samples were washed three times with water after treatment and lyophilized. All samples were packed into 3.2-mm zirconia rotors. All CPMAS NMR (34) experiments were performed in an 89-mm-bore 11.7 T magnet using an Agilent BioMAS NMR probe with a DD2 console (Agilent Technologies). Spectrometer ^{13}C chemical shift referencing was performed by setting the high-frequency adamantane peak to 38.5 ppm (35). Samples were spun at 7143 Hz and 10,056 Hz and experiments were performed at room temperature. The field strength for ^{13}C cross-polarization was 50 kHz with a 10% ^1H linear ramp centered at 57 kHz for MAS experiments at 7143 Hz and at 60kHz for 10,056 Hz MAS experiments. ^1H decoupling was performed with TPPM at 83kHz. The recycle time was 2 s for all experiments except for the 10 s comparison in Figure S1. NMR spectra were processed with 80-Hz line broadening. CP array experiments were performed at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7 ms.

SDS treatment and proteomics analysis

Chrysalis samples from both Monarch and Swallowtail were incubated in sample buffer with 10% SDS (250mM Tris pH 8.8, 10% SDS, 20% glycerol, 5% BME, 0.5mg/mL bromophenol blue) and boiled for 1 hour. After boiling, samples were centrifuged at 12,000 rpm for 10 minutes. The pellet with remaining solid chrysalis material was washed in water 3 times and either lyophilized for NMR analysis or trypsin treated for analysis of the more insoluble proteins associated with the chrysalides. The supernatant was examined by SDS-PAGE using NuPAGE 4-12% Bis-Tris gel in 1X MES, run for 1 hr at 150V, stained with Expedeon Instant Blue, and photographed. For inclusive proteomic analysis of the SDS soluble and the insoluble trypsin-treated samples, a limited gel was run in NuPAGE 4-12% Bis-Tris gel in 1X MES for 8 min at 150V and excised for by nano LC-MS/MS proteomic analysis on the Thermo Scientific Orbitrap QE HF-X at the Vincent Coates Foundation Mass Spectrometry Laboratory at Stanford University. Analysis was performed using Byonic software by Protein Metrics.

NaOH digestion

Whole Monarch chrysalis samples were ground by mortar and pestle. 20 mg of ground chrysalis was divided into two glass tubes each treated with 1.5 mL of 1 M NaOH for 2 hours at 98°C. NaOH was removed and samples were combined, washed three times with water, and lyophilized for NMR analysis.

Commercial chitin

Pure chitin (isolated from Crustacean shells) was purchased from MP Biomedicals (Solon, OH, USA) Catalog # 101334.

ACKNOWLEDGEMENTS

We acknowledge Vanessa Hulan, Jillian Kanan, and Nancy McCarthy for donating, raising, and releasing the butterflies. We also acknowledge Kratika Singhal, Fang Liu, and the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry for assistance with proteomics analysis. This work was supported in part by NIH P30 CA124435 utilizing the Stanford Cancer Institute Proteomics/Mass Spectrometry Shared Resource. N.F.G acknowledges support from the Stanford Graduate Fellowship. This work was supported by the National Science Foundation Award 2001189 (L.C.).

REFERENCES

- [1] Snodgrass, R. E. (1954). Insect metamorphosis. Washington, D.C: Smithsonian Institution.
- [2] Yonemura, N., Sehna, F. (2006). The design of silk fiber composition in moths has been conserved for more than 150 million years, *J. Mol. Biol.* **63**, 42-53.
- [3] Kramer, K. J., Koga, D. (1986). Insect chitin: physical state, synthesis, degradation and metabolic regulation, *Insect Biochem.* **16**, 851-877.
- [4] Kramer, K. J., Hopkins, T. L., Schaefer, J. (1995). Applications of solids NMR to the analysis of insect sclerotized structures, *Insect Biochem. Mol. Biol.* **25**, 1067-1080.
- [5] Chen, F., Porter, D., Vollrath, F. (2012). Morphology and structure of silkworm cocoons, *Mater. Sci. Eng. C* . **32**, 772-778.
- [6] Boßelmann, F., Romano, P., Fabritius, H., Raabe, D., Epple, M. (2007). The composition of the exoskeleton of two crustacea: The American lobster *Homarus americanus* and the edible crab *Cancer pagurus*, *Thermochim. Acta.* **463**, 65-68.
- [7] Nagasawa, H. (2012). The crustacean cuticle: structure, composition and mineralization, *Front. Biosci.* **4**, 711-720.
- [8] Rinaudo, M. (2006). Chitin and chitosan: properties and applications, *Prog. Polym. Sci.* **31**, 603-632.
- [9] Ravi Kumar, M. N. V. (2000). A review of chitin and chitosan applications, *React. Funct. Polym.* **46**, 1-27.

- [10] Minke, R., Blackwell, J. (1978). The structure of α -chitin, *J. Mol. Biol.* **120**, 167-181.
- [11] Kaya, M., Mujtaba, M., Ehrlich, H., Salaberria, A. M., Baran, T., Amemiya, C. T., et al. (2017). On chemistry of γ -chitin, *Carbohydr. Polym.* **176**, 177-186.
- [12] Jang, M. K., Kong, B. G., Jeong, Y. I., Lee, C. H., Nah, J. W. (2004). Physicochemical characterization of α -chitin, β -chitin, and γ -chitin separated from natural resources, *J. Polym. Sci. A Polym. Chem.* **42**, 3423-3432.
- [13] Lipke, H., Geoghegan, T. (1971). The composition of peptidochitodextrins from sarcophagid puparial cases, *Biochem. J.* **125**, 703-716.
- [14] Hackman, R. H., Goldberg, M. (1978). The non-covalent binding of two insect cuticular proteins by a chitin, *Insect Biochem.* **8**, 353-357.
- [15] Paulino, A. T., Simionato, J. I., Garcia, J. C., Nozaki, J. (2006). Characterization of chitosan and chitin produced from silkworm crysalides, *Carbohydr. Polym.* **64**, 98-103.
- [16] Peter, M. G., Grun, L., Forster, H. (1984). CP/MAS- ^{13}C -NMR spectra of sclerotized insect cuticle and of chitin, *Angew. Chem. Int. Ed.* **23**, 638-639.
- [17] Christensen, A. M., Schaefer, J., Kramer, K. J., Morgan, T. D., Hopkins, T. L. (1991). Detection of cross-links in insect cuticle by REDOR NMR spectroscopy, *J. Am. Chem. Soc.* **113**, 6799-6802.
- [18] Gullion, J. D., Gullion, T. (2017). Solid-state NMR study of the cicada wing, *J. Phys. Chem. B.* **121**, 7646-7651.
- [19] Eddy, S., Gullion, T. (2021). Characterization of insect wing membranes by ^{13}C CPMAS NMR, *J. Phys. Chem.* **125**, 931-936.
- [20] Zhou, Y., Badgett, M. J., Billard, L., Bowen, J. H., Orlando, R., Willis, J. H. (2017). Properties of the cuticular proteins of *Anopheles gambiae* as revealed by serial extraction of adults, *PLoS ONE*. **12**, 1-21.
- [21] Hackman, R. H., Goldberg, M. (1977). Molecular crosslinks in cuticles, *Insect Biochem.* **7**, 175-184.
- [22] Rebers, J. E., Riddiford, L. M. (1988). Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes, *J. Mol. Biol.* **203**, 411-423.
- [23] Rebers, J. E., Willis, J. H. (2001). A conserved domain in arthropod cuticular proteins binds chitin, *Insect Biochem. Mol. Biol.* **31**, 1083-1093.
- [24] Schaefer, J., Kramer, K. J., Garbow, J. R., Jacob, G. S., Stejskal, E. O., Hopkins, T. L., et al. (1987). Aromatic cross-links in insect cuticle-detection by solid-state ^{13}C and ^{15}N NMR, *Science*. **235**, 1200-1206.
- [25] Kerwin, J. L., Turecek, F., Xu, R., Kramer, K. J., Hopkins, T. L., Gatlin, C. L., et al. (1999). Mass spectrometric analysis of catechol-histidine adducts from insect cuticle, *Anal. Biochem.* **268**, 229-237.

- [26] Hopkins, T. L. (1992). Insect cuticle sclerotization, *Annu. Rev. Entomol.* **37**, 273-302.
- [27] Merritt, M. E., Christensen, A. M., Kramer, K. J., Hopkins, T. L., Schaefer, J. (1996). Detection of intercatechol cross-links in insect cuticle by solid-state carbon-13 and nitrogen-15 NMR, *J. Am. Chem. Soc.* **118**, 11278-11282.
- [28] Williams, H. J., Scott, A. I., Woolfenden, W. R., Grant, D. M., Vinison, S. B., Elzen, G. W., et al. (1988). In vivo and solid state ^{13}C nuclear magnetic resonance studies of tyrosine metabolism during insect cuticle formation, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **89**, 317-321.
- [29] Goldberg, W. M., Hopkins, T. L., Holl, S. M., Schaefer, J., Kramer, K. J., Morgan, T. D., et al. (1994). Chemical composition of the sclerotized black coral skeleton (Coelenterata: Antipatharia): a comparison of two species, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **107**, 633-643.
- [30] Kramer, K. J., Christensen, A. M., Morgan, T. D., Schaefer, J., Czapla, T. H., Hopkins, T. L. (1991). Analysis of cockroach oothecae and exuviae by solid-state ^{13}C -NMR spectroscopy, *Insect Biochem.* **21**, 149-156.
- [31] Holl, S. M., Schaefer, J., Goldberg, W. M., Kramer, K. J., Morgan, T. D., Hopkins, T. L. (1992). Comparison of black coral skeleton and insect cuticle by a combination of carbon-13 NMR and chemical analyses, *Arch. Biochem. Biophys.* **292**, 107-111.
- [32] Kramer, K. J., Morgan, T. D., Hopkins, T. L., Christensen, A. M., Schaefer, J. (1989). Solid-state ^{13}C -NMR and diphenol analyses of sclerotized cuticles from stored product Coleoptera, *Insect Biochem.* **19**, 753-757.
- [33] Kramer, K. J. (2001). Oxidative conjugation of catechols with proteins in insect skeletal systems, *Tetrahedron.* **57**, 385-392.
- [34] Kramer, K. J., Bork, V., Schaefer, J., Morgan, T. D., Hopkins, T. L. (1989). Solid state ^{13}C nuclear magnetic resonance and chemical analyses of insect noncuticular sclerotized support structures: Mantid oothecae and cocoon silks, *Insect Biochem.* **19**, 69-77.
- [35] Sugumaran, M., Lipke, H. (1982). Crosslink precursors for the dipteran puparium, *Proc. Natl Acad. Sci. USA.* **79**, 2480-2484.
- [36] Cegelski, L., O'Connor, R. D., Stueber, D., Singh, M., Poliks, B., Schaefer, J. (2010) Plant cell-wall cross-links by REDOR NMR spectroscopy. *J. Amer. Chem. Soc.* **132**, 16052-16057.
- [37] Stejskal, E., Schaefer, J., Steger, T. R. (1978). High-resolution ^{13}C nuclear magnetic resonance in solids, *Faraday Symp. Chem. Soc.*, 56-62.
- [38] Kolodziejewski, W., Klinowski, J. (2002). Kinetics of cross-polarization in solid-state NMR: a guide for chemists, *Chem. Rev.* **102**, 613-628.
- [39] Gilby, A. R., McKellar, J. W. (1970). The composition of the empty puparia of a blowfly *J. Insect Physiol.* **16**, 1517-1529.

- [40] Hackman, R. H., Goldberg, M. (1981). A method for determinations of microgram amounts of chitin in arthropod cuticles, *Anal. Biochem.* **110**, 277-280.
- [41] Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L., et al. (2015). Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies, *Nat. Commun.* **6**, 1-10.
- [42] Bern, M., Kil, Y. J., Becker, C. (2012). Byonic: advanced peptide and protein identification software, *Curr. Protoc. Bioinformatics.* **40**, 13.20.01-13.20.14.
- [43] Majtan, J., Bilikova, K., Markovic, O., Grof, J., Kogan, G., Simuth, J. (2007). Isolation and characterization of chitin from bumblebee (*Bombus terrestris*), *Int. J. Biol. Macromol.* **40**, 237-241.
- [44] Schaefer, J., Stejskal, E. O. (1976) Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *J. Am. Chem. Soc.* **98**, 1031-1032.
- [45] Morcombe, C. R., Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson.* **162**, 479-486.

FIGURES

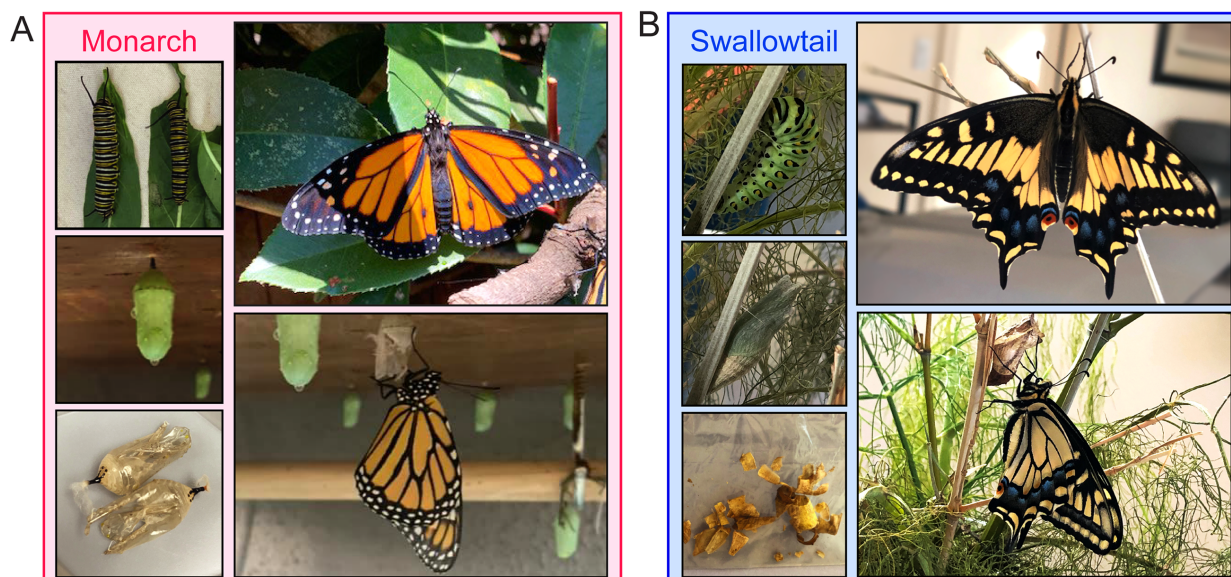


Figure 1. Chrysalides from Butterfly Metamorphosis. (A) Monarch and (B) Swallowtail butterflies were raised from caterpillars. Photographs present caterpillars preparing to form a chrysalis, formed chrysalides, butterfly emerged near to chrysalis, and empty chrysalides collected for compositional analysis.

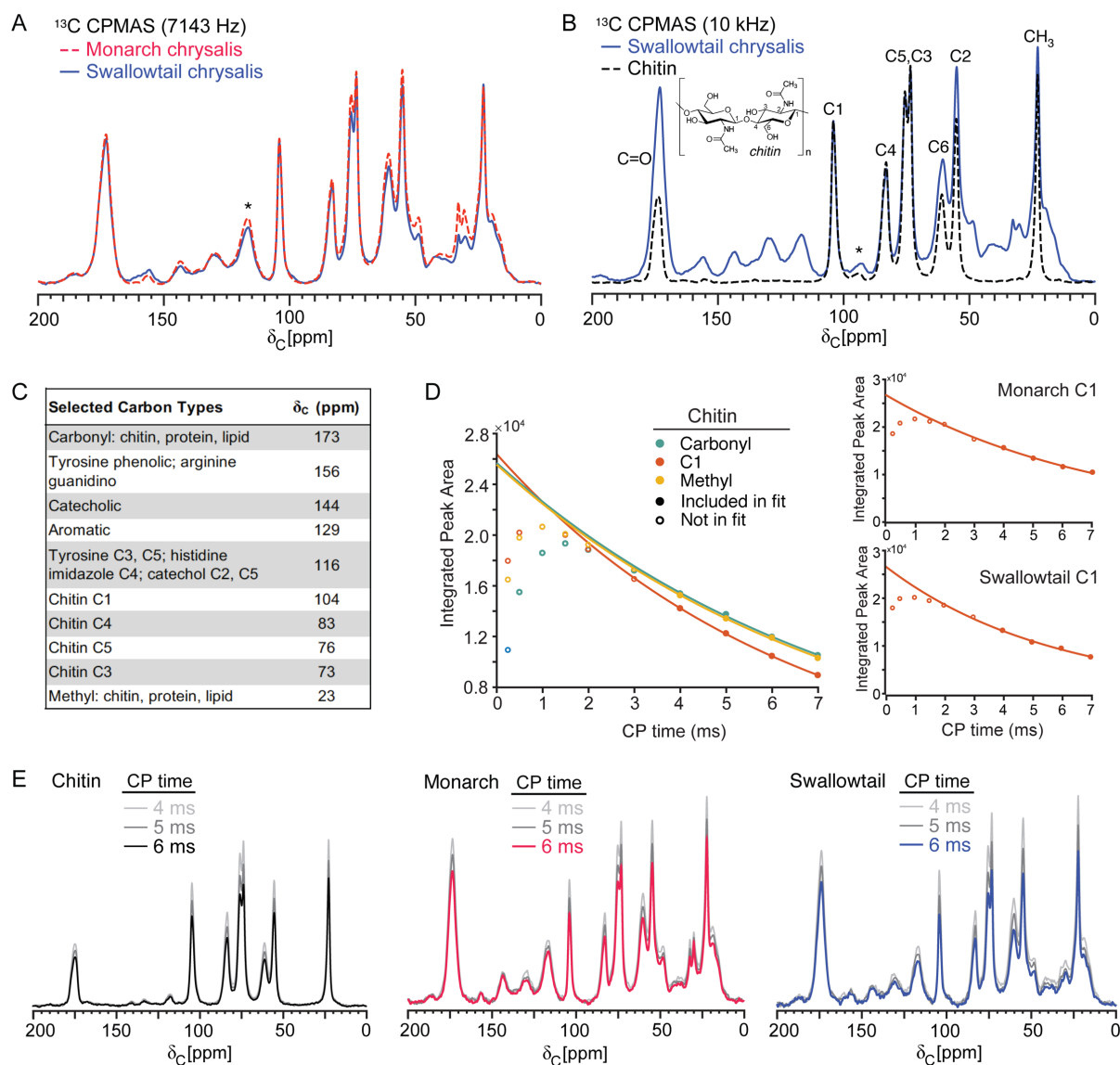


Figure 2. Chrysalis Compositional Analysis. (A) ^{13}C CPMAS spectral comparison of Monarch (dashed red spectrum) and Swallowtail (solid blue spectrum) chrysalides, with spectra resulting from 32,768 scans, with MAS at 7143 Hz, and scaled by sample mass (20 mg Monarch; 14 mg Swallowtail). Carbonyl spinning sideband is noted by an asterisk. (B) ^{13}C CPMAS spectral comparison of the Swallowtail chrysalis and commercially available chitin obtained with 10,056 Hz MAS. Carbonyl spinning sideband at 94 ppm is indicated by an asterisk. (C) ^{13}C CPMAS spectral assignments of chitin and catechol species. (D) CP array analysis for chitin showing integrated areas for carbonyl (teal), methyl (yellow), and C1 (orange) carbons as well as that of the C1 anomeric carbon in Monarch and Swallowtail chrysalides (orange). Exponential fits are shown, using the 4, 5, 6, and 7 ms CP time acquisitions, indicated by closed circles, for extrapolation back to time zero. (E) ^{13}C CPMAS spectral comparison of commercially available chitin (2048 scans), Monarch chrysalis (4096 scans) and Swallowtail chrysalis (4096 scans), acquired with varied CP times and MAS at 7143 Hz.

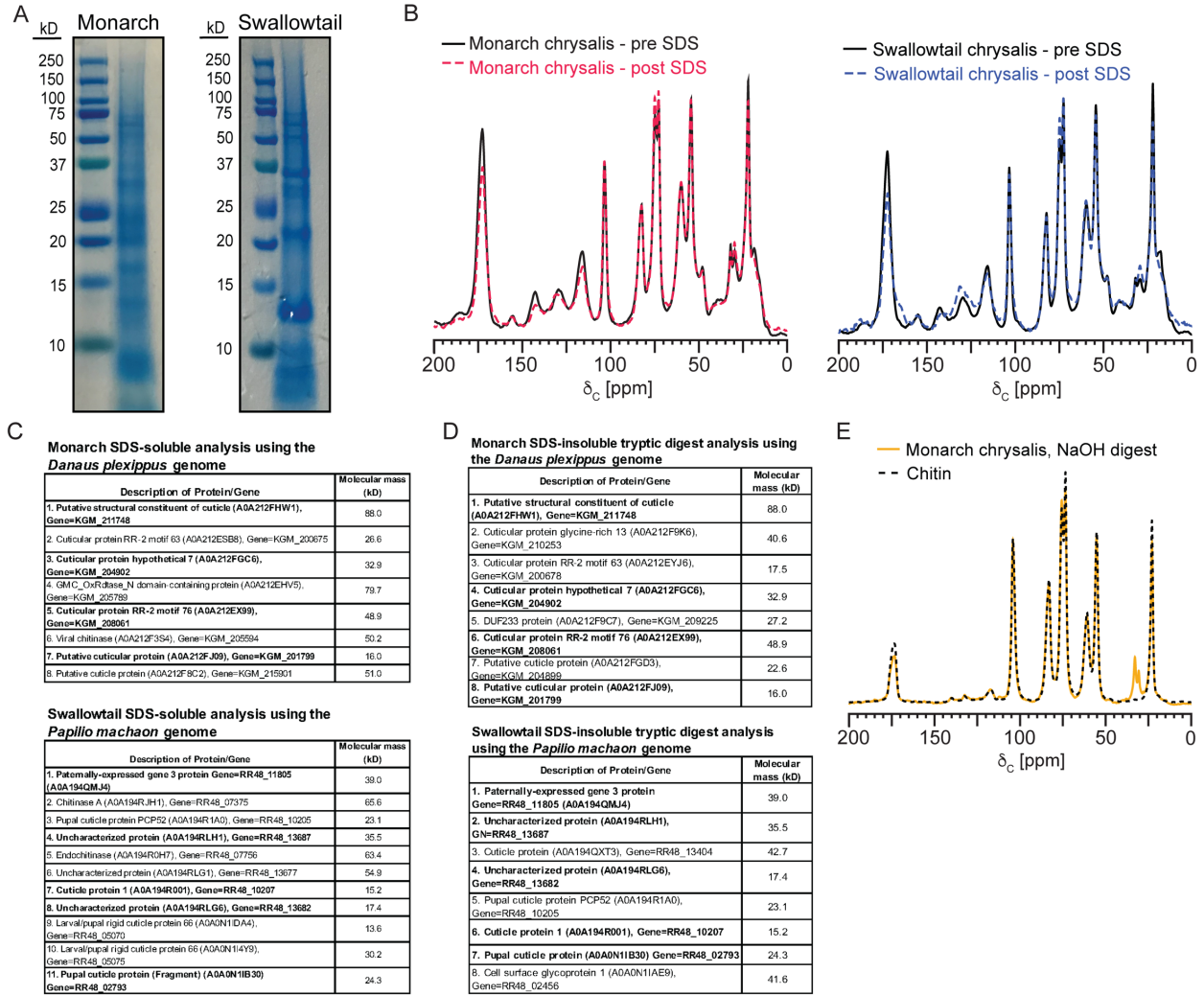


Figure 3. Chrysalis Digestion and Protein Analysis. (A) SDS-PAGE gel showing SDS-soluble proteins associated with the Monarch and Swallowtail chrysalides. (B) ^{13}C CPMAS spectra of the Monarch chrysalis and Swallowtail chrysalides before and after SDS treatment. The post-SDS treated sample spectra are the same as presented in Figure 2A and obtained with 7143 Hz MAS. (C) SDS-soluble proteins identified from nano LC-MS/MS with a Log Probability, *i.e.* Log base 10 of the protein p-value, at or greater than 100. Entries in bold are also found in the insoluble tryptic digest analysis. Additional details provided in Tables S1 and S2. (D) Top proteins identified from nano LC-MS/MS of SDS-insoluble pellets treated with trypsin to liberate accessible peptides from chrysalis-integrated proteins. Additional details provided in Tables S3 and S4. (E) ^{13}C CPMAS spectral overlay of the Monarch chrysalis after NaOH digestion (gold) with that of commercial chitin (dashed black).

Chemical and Molecular Composition of the Chrysalis Reveals Common Chitin-rich Structural Framework for Monarchs and Swallowtails

Nicolette F. Goularte,^{1‡} Till Kallem,^{2‡} and Lynette Cegelski^{2}*

¹Department of Structural Biology, 380 Roth Way, Stanford University School of Medicine, Stanford, California 94305

²Department of Chemistry, 380 Roth Way, Stanford University, Stanford, California 94305

*Corresponding author: cegelski@stanford.edu

‡These authors contributed equally to this work.

SUPPLEMENTARY INFORMATION

Contents:

Supplementary Figures (Figure S1 and S2)

Supplementary Tables (Tables S1-S5)

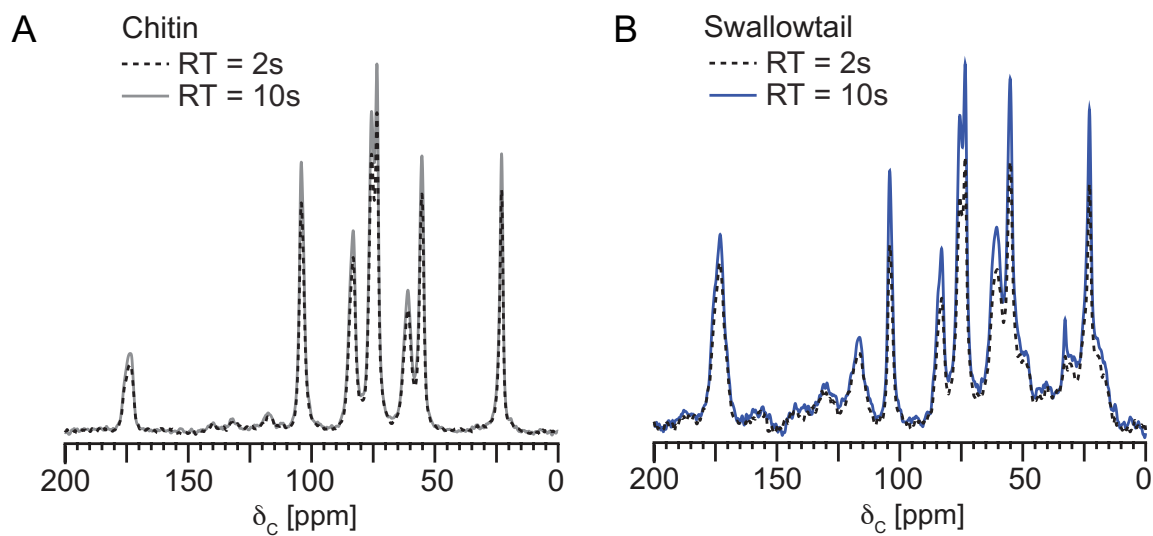


Figure S1. (A) ^{13}C CPMAS spectral comparison of commercially available chitin acquired with 2s (dashed black) and 10s (grey) recycle delay at 7143Hz, each the result of 1024 scans. (B) ^{13}C CPMAS spectral comparison of the Swallowtail chrysalis acquired with 2s (dashed black) and 10s (blue) recycle delay at 7143Hz, each the result of 2048 scans.

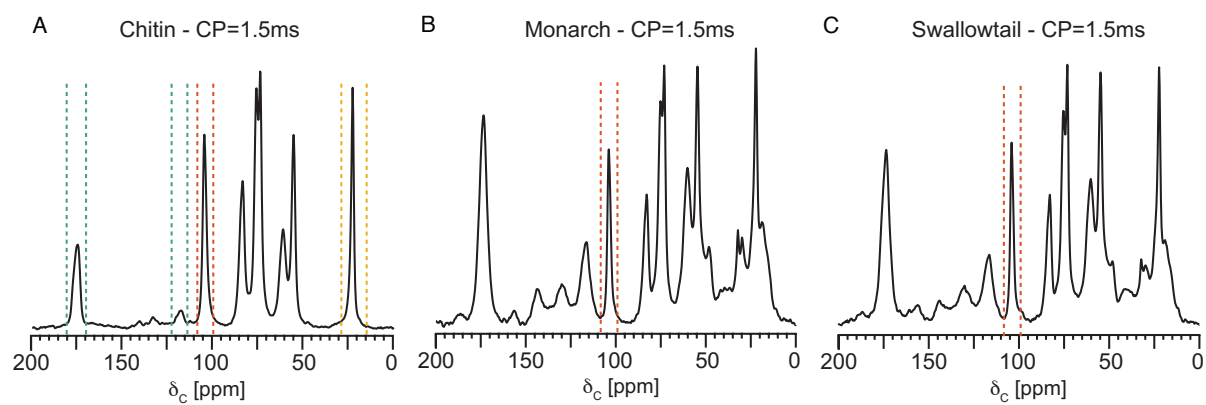


Figure S2. ^{13}C CPMAS spectrum at 7143Hz MAS and cross-polarization of 1.5 ms for (A) Chitin (B) Monarch chrysalis and (C) Swallowtail chrysalis marking the bounds of each peak integrated in Figure 2D.

Table S1. Soluble Monarch chrysalis proteomics data. Top proteins identified post SDS treatment from nano LC-MS/MS with a Log Prob (or Log base 10 of the protein p-value) at or greater than 100. Entries in bold are also found in the Monarch insoluble protein analysis (Table S3).

Monarch soluble analysis using *Danaus plexippus* genome

Description of Protein/Gene	Molecular mass (kD)	Log Prob	Total Intensity	Number of spectra	Unique peptides	Coverage Percent
1. Putative structural constituent of cuticle (A0A212FWH1), Gene=KGM_211748	88.0	563.15	6.34E+07	356	114	90
2. Cuticular protein RR-2 motif 63 (A0A212ESB8), Gene=KGM_200675	26.6	194.66	1.75E+07	108	44	81
3. Cuticular protein hypothetical 7 (A0A212FGC6), Gene=KGM_204902	32.9	124.78	6.94E+06	47	27	48
4. GMC_OxRdtase_N domain-containing protein (A0A212EHV5), Gene=KGM_205789	79.7	118.69	3.60E+06	51	30	40
5. Cuticular protein RR-2 motif 76 (A0A212EX99), Gene=KGM_208061	48.9	112.38	1.62E+07	100	27	34
6. Viral chitinase (A0A212F3S4), Gene=KGM_205594	50.2	107.28	4.74E+06	56	27	38
7. Putative cuticular protein (A0A212FJ09), Gene=KGM_201799	16.0	106.16	6.29E+06	95	28	75
8. Putative cuticle protein (A0A212F8C2), Gene=KGM_215901	51.0	98.25	2.85E+06	52	23	41

Table S2. Soluble Swallowtail chrysalis proteomics data. Top proteins identified post SDS treatment from nano LC-MS/MS with a Log Prob (or Log base 10 of the protein p-value) at or greater than 100. Entries in bold are also found in the Swallowtail insoluble protein analysis (Table S4).

Swallowtail soluble analysis using *Papilio machaon* genome

Description of Protein/Gene	Molecular mass (kD)	Log Prob	Total Intensity	Number of spectra	Unique peptides	Coverage Percent
1. Paternally-expressed gene 3 protein Gene=RR48_11805 (A0A194QMJ4)	39.0	681.92	1.57E+10	635	112	74
2. Chitinase A (A0A194RJH1), Gene=RR48_07375	65.6	300.42	5.05E+08	266	58	52
3. Pupal cuticle protein PCP52 (A0A194R1A0), Gene=RR48_10205	23.1	272.99	1.11E+09	168	46	79
4. Uncharacterized protein (A0A194RLH1), Gene=RR48_13687	35.5	259.56	6.25E+08	111	29	58
5. Endochitinase (A0A194R0H7), Gene=RR48_07756	63.4	215.12	2.63E+08	133	40	47
6. Uncharacterized protein (A0A194RLG1), Gene=RR48_13677	54.9	214.89	1.16E+09	153	39	64
7. Cuticle protein 1 (A0A194R001), Gene=RR48_10207	15.2	174.61	8.61E+08	244	37	78
8. Uncharacterized protein (A0A194RLG6), Gene=RR48_13682	17.4	167.65	1.03E+09	94	33	68
9. Larval/pupal rigid cuticle protein 66 (A0A0N1IDA4), Gene=RR48_05070	13.6	157.06	8.72E+08	162	26	81
10. Larval/pupal rigid cuticle protein 66 (A0A0N1I4Y9), Gene=RR48_05075	30.2	116.40	4.11E+08	76	18	35
11. Pupal cuticle protein (Fragment) (A0A0N1IB30) Gene=RR48_02793	24.3	113.76	5.63E+08	98	33	88

Table S3. Insoluble Monarch chrysalis proteomics data. Top proteins identified from nano LC-MS/MS of the SDS-insoluble pellet treated with trypsin. Entries in bold are also found in the Monarch soluble protein analysis (Table S1).

Monarch insoluble analysis using *Danaus plexippus* genome

Description of Protein/Gene	Molecular mass (kD)	Log Prob	Total Intensity	Number of spectra	Unique peptides	Coverage Percent
1. Putative structural constituent of cuticle (A0A212FHW1), Gene=KGM_211748	88.0	139.32	1.71E+08	137	36	62
2. Cuticular protein glycine-rich 13 (A0A212F9K6), Gene=KGM_210253	40.6	76.12	3.67E+07	69	20	63
3. Cuticular protein RR-2 motif 63 (A0A212EYJ6), Gene=KGM_200678	17.5	63.48	5.19E+07	70	21	79
4. Cuticular protein hypothetical 7 (A0A212FGC6), Gene=KGM_204902	32.9	61.77	4.26E+07	51	14	46
5. DUF233 protein (A0A212F9C7), Gene=KGM_209225	27.2	58.14	6.63E+07	69	19	63
6. Cuticular protein RR-2 motif 76 (A0A212EX99), Gene=KGM_208061	48.9	54.94	1.29E+08	127	19	30
7. Putative cuticle protein (A0A212FGD3), Gene=KGM_204899	22.6	51.10	1.19E+08	38	11	51
8. Putative cuticular protein (A0A212FJ09), Gene=KGM_201799	16.0	42.98	2.89E+07	34	10	65

Table S4. Insoluble Swallowtail chrysalis proteomics data. Top proteins identified from nano LC-MS/MS of the SDS-insoluble pellet treated with trypsin. Entries in bold are also found in the Swallowtail soluble protein analysis (Table S2).

Swallowtail insoluble analysis using *Papilio machaon* genome

Description of Protein/Gene	Molecular mass (kD)	Log Prob	Total Intensity	Number of spectra	Unique peptides	Coverage Percent
1. Paternally-expressed gene 3 protein Gene=RR48_11805 (A0A194QMJ4)	39.0	93.83	9.52E+07	85	20	39
2. Uncharacterized protein (A0A194RLH1), GN=RR48_13687	35.5	56.46	5.42E+07	44	14	34
3. Cuticle protein (A0A194QXT3), Gene=RR48_13404	42.7	38.26	4.57E+07	55	12	28
4. Uncharacterized protein (A0A194RLG6), Gene=RR48_13682	17.4	30.14	2.25E+07	17	5	19
5. Pupal cuticle protein PCP52 (A0A194R1A0), Gene=RR48_10205	23.1	26.08	1.09E+07	19	7	32
6. Cuticle protein 1 (A0A194R001), Gene=RR48_10207	15.2	25.54	1.33E+07	22	12	54
7. Pupal cuticle protein (A0A0N1IB30) Gene=RR48_02793	24.3	24.06	2.62E+07	32	9	42
8. Cell surface glycoprotein 1 (A0A0N1IAE9), Gene=RR48_02456	41.6	22.81	4.93E+06	12	8	30

Table S5. Similarity data from BLAST search of top insoluble Swallowtail *Papilio Machaon* proteins (from Table S4) with the resulting Monarch *Danaus plexippus* genome matches.

BLAST search: Swallowtail	BLAST result: Monarch	E-value	% Identity
1. Paternally-expressed gene 3 protein Gene=RR48_11805 (A0A194QMJ4)	1. Putative structural constituent of cuticle (A0A212FHW1), Gene=KGM_211748	2.50E-91	53.6%
2. Uncharacterized protein (A0A194RLH1), GN=RR48_13687	4. Cuticular protein hypothetical 7 (A0A212FGC6), Gene=KGM_204902	2.90E-88	51.3%
3. Cuticle protein (A0A194QXT3), Gene=RR48_13404			
4. Uncharacterized protein (A0A194RLG6), Gene=RR48_13682	7. Putative cuticle protein (A0A212FGD3), Gene=KGM_204899	6.00E-37	60.7%
5. Pupal cuticle protein PCP52 (A0A194R1A0), Gene=RR48_10205			
6. Cuticle protein 1 (A0A194R001), Gene=RR48_10207	8. Putative cuticular protein (A0A212FJ09), Gene=KGM_201799	1.70E-52	66.4%
7. Pupal cuticle protein (A0A0N11B30) Gene=RR48_02793			
8. Cell surface glycoprotein 1 (A0A0N11AE9), Gene=RR48_02456	2. Cuticular protein glycine-rich 13 (A0A212F9K6), Gene=KGM_210253	0.0	84.4%