UNIVERSITY OF CALIFORNIA,
IRVINE

Instrumentation and Methods Development for NMR of Oriented Biomolecules

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by

John E. Kelly

Dissertation Committee:
Professor Rachel W. Martin, Chair
Professor Athan J. Shaka
Professor John C. Hemminger

2018

DEDICATION

My family for seeing me through this crazy roller-coaster of education.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Rachel Martin, for teaching me everything I know about NMR. I joined the lab knowing only what a 1D spectrum looks like, and she said I could build a probe. She always gave just enough help for me to figure it out on my own, and I am so much the better off for it. I would not have learned anywhere near as much otherwise. I would also like to thank her for always assuring me that I will indeed get out of here one day;

Dr. Kelsey Collier for showing me how to build a probe, and reminding me to always note down all necessary sacrifices to dark gods that make a probe work correctly;

Dr. Suvrajit Sengupta for always knowing how to make the spectrometer work, and you eternal wishes of "Good luuuuccccckkkkkk";

Jessica Kelz for taking all of my problems...I mean projects, and making them yours for the future;

the rest of the Martin lab for helping with fills, odd bio questions, and the occasional necessary coffee excursion...oh and all the heckling during practice talks;

the Game Night crew for stabilizing my insanity, and always being a good time at least once a week to let off steam;

Dr. Nick Vizenor for always being up for a random adventure out, regardless of if it's just groceries, a cool new Yelp place, or Disneyland;

Tory Doolin for accepting me as a required part of being with Nick, and letting me third-wheel all the time;

The research presented herein was supported by grants from NSF, CHE-1308231 and DMS-1361425. This work was also 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.

CURRICULUM VITAE

John E. Kelly

EDUCATION

Doctor of Philosophy in Chemistry **2013-2018**
University of California, Irivne *Irvine, CA*

Master of Science in Chemistry **2013-2016**
University of California, Irivne *Irvine, CA*

Bachelor of Science in Chemistry **2009-2013**
Montana Tech of the University of Montana *Butte, MT*

RESEARCH EXPERIENCE

SAS Probes **2013–2018**
University of California, Irvine *Irvine, California*

Enzyme Discovery **2013–2018**
University of California, Irvine *Irvine, California*

PUBLICATIONS

1. **Advances in Instrumentation and Methodology for Solid-State NMR of Biological Assemblies** R.W. Martin, J.E. Kelly, J.I. Kelz, *J. Struct. Biol.* **2018**, Accepted.

2. **Evolutionary and structural analyses uncover the role of novel gene duplications in pollen feeding butterflies.** G. Smith, J.E. Kelly, A. Macias-Munoz, C.T. Butts, R.W. Martin, A.D. Briscoe, *Proc. R. Soc. B* **2018**, *285*, 20172037.

3. **Design and construction of a quadruple-resonance MAS NMR probe for investigation of extensively deuterated biomolecules.** K.A. Collier, S. Sengupta, C.A. Espinosa, J.E. Kelly, J.I. Kelz, R.W. Martin, *J. Magn. Reson.* **2017**, *285*, 8-17.

4. **Structure prediction and network analysis of chitinases from the Cape sundew, *Drosera capensis*.** M.H. Unhelkar, V.T. Duong, K.N. Enendu, J.E. Kelly, S. Tahir, C.T. Butts, and R.W. Martin, *Biochim. Biophys. Acta, Gen. Subj.* **2016**, *1861*, 636-643.

5. **Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*.** C.T. Butts, X. Zhang, J.E. Kelly, K.W. Roskamp, M.H. Unhelkar, J.A. Freites, S. Tahir and R.W. Martin, *Comput. Struct. Biotechnol. J.* **2016**, *14*, 271-282.

6. **Spatial Reorientation Experiments for NMR of Solids and Partially Oriented Liquids, Progress in Nuclear Magnetic Resonance Spectroscopy.** R.W. Martin, J.E. Kelly, K.A. Collier, *Prog. Nucl. Magn. Reson. Spectrosc.* **2015**, *90-91*, 92-122.

PRESENTATIONS

Oral Presentations:

1. Instrumentation Development for NMR of Oriented Samples. Montana Tech of the University of Montana, December 8, 2017.

Poster Presentations:

1. Instrumentation and Methods Development for NMR of Oriented Biomolecules. Rocky Mountain Conference on Magnetic Resonance, July 23, 2018

2. Design and Construction of a SAS ssNMR Probe for the Investigation of Oriented Samples. 10th Alpine Conference on Solid State NMR, March 27, 2017

3. Design and Construction of a SAS ssNMR Probe for the Investigation of Oriented Samples. 58th Experimental Nuclear Magnetic Resonance Conference, March 27, 2017.

4. Design and Construction of ssNMR Probes for the Investigation of Oriented Solids and Liquids. 58th Rocky Mountain Conference on Magnetic Resonance, July 18, 2016.

ABSTRACT OF THE DISSERTATION

Instrumentation and Methods Development for NMR of Oriented Biomolecules

By

John E. Kelly

Doctor of Philosophy in Chemistry

University of California, Irvine, 2018

Professor Rachel W. Martin, Chair

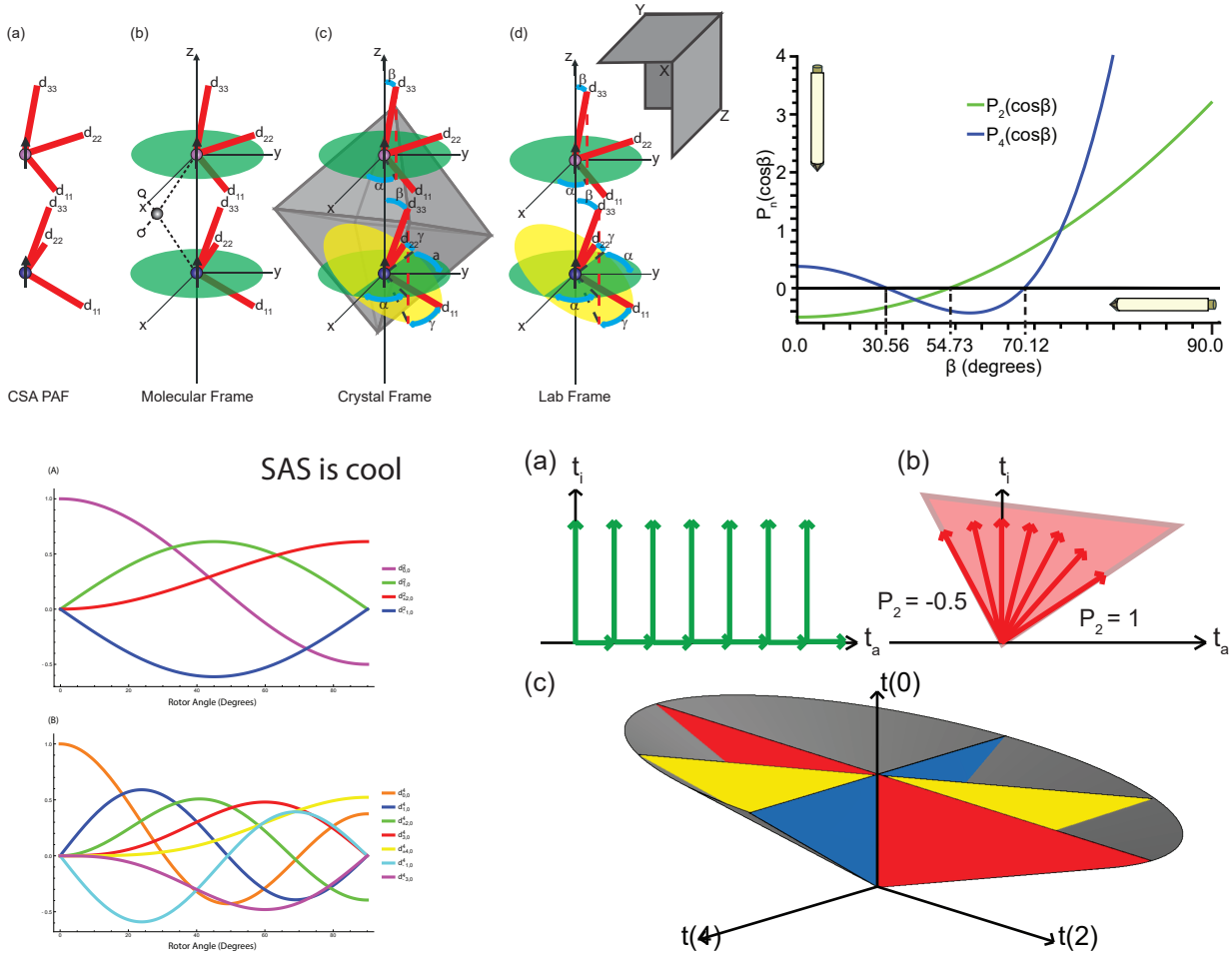
This thesis describes the design and construction of a three-channel ($^{1}H/^{13}C/^{15}N$) switched-angle spinning solid-state NMR probe for a 500 MHz (11.7 T) magnet. The probe is designed for studies of membrane-associated proteins in native-like environments. This probe, which is the next generation of the pneumatic SAS probes built in the Martin Lab, keeps the pneumatic angle switching mechanism from the previous generation, while adding the third channel to enable the triple resonance experiments necessary for protein structure work. The channels utilize transmission line segments that act as tunable reactances, with the matching network for each frequency contained within an outer ground plane. The channels are capacitively coupled to the coil to enable smooth switching without bending the leads repeatedly. In order to study proteins with this probe, we will investigate the angular dependence of decoupling sequences. This is necessary because dipolar couplings are partially averaged out depending on the angle at which the sample is spinning. The angular dependence of popular decoupling sequences will be determined in order to assess how they change with the angle of the sample, enabling us to separately optimize for different angles within a single experiment. With SAS-optimized decoupling sequences, structural studies can be performed on membrane-associated proteins at different angles to extract further distance constraints and orientation information.

Also described is a high-throughput method for re-equilibrating predicted protein structures

and evaluating them to predict their properties in order to search for targets with desired characteristics or other interesting features. This approach allows for the more efficient selection of protein targets for structure determination and biochemical characterization. This method has been used to investigate cocoonases from *Heliconius* butterflies and aspartic proteases from carnivorous plants, as well as other that are described elsewhere.

Chapter 1

Switched-Angle Spinning



The chapter cover figure shows some selected figures illustrating concepts discussed in this chapter; (top left) two P nuclei from a crystal phase in the various frames of reference, (top right) $P_2(cos\beta)$ and $P_4(cos\beta)$ plotted of the range of rotor angles, (bottom left) the reduced Wigner matrix elements as functions of rotor angle, (bottom right) data collection schemes for multidimensional isotropic/anisotropic correlations.

1.1 Motional Reorientation Experiments in Solid-State NMR: History

Motional reorientation experiments in solid-state NMR are extensions of Magic Angle Spinning (MAS) where the rotor axis is changed in order to average out, reintroduce, or scale anisotropic interactions (e.g. dipolar couplings, quadrupolar interactions or chemical shift anisotropies). This chapter focuses on Switched Angle Spinning (SAS), which involves spinning at two different angles sequentially during a multidimensional experiment. In all of these experiments, anisotropic terms in the Hamiltonian are scaled by changing the orientation of the spinning sample relative to the static magnetic field. The majority of this background was previously published in a review paper co-authored with Dr. Rachel Martin and Dr. Kelsey Collier [1].

Solid-state NMR is the structural method of choice for investigating non-crystalline solids, semi-solids, and liquid crystals. Here, the anisotropic interactions that are averaged out in isotropic liquids by molecular tumbling, including magnetic susceptibility differences, chemical shift anisotropy, quadrupolar interactions, and dipolar couplings, are present to a greater or lesser degree depending on the order parameter. Spatially-dependent anisotropic interactions are both the greatest strength and the biggest challenge of solid-state NMR; they provide exquisitely detailed local information about the chemical environment of the nuclear spins at the expense of significantly complicating the spectra. Anisotropic interactions are usually scaled or removed by a combination of Magic Angle Spinning (MAS) and radio frequency (rf) decoupling. Because completely removing these interactions results in the

loss of valuable orientational information that is not available from liquids experiments, the heteronuclear dipolar interactions or chemical shift anisotropies are frequently selectively reintroduced using specialized pulse sequences.

Motional reorientation experiments represent an alternative strategy where the anisotropic interaction terms in the Hamiltonian are modified by changing the orientation of the sample in physical space, in contrast to the more common methods involving modulation of these interactions in spin-space via rotor-synchronized pulse sequences. In VAS, many one-dimensional spectra are collected as a function of spinning angle. In SAS, the structural information from the dipolar couplings and/or chemical shift anisotropies is recovered by switching the sample between two different angles, yielding a two-dimensional NMR spectrum. Thus, SAS measures the isotropic spectrum that is needed to assign the resonances of different chemical sites, while recovering valuable anisotropic information that would be lost in a simple MAS experiment. Dynamic Angle Spinning (DAS) is a special case of SAS for quadrupolar nuclei in which the two angles are chosen to average out the first- and second-order quadrupolar interactions. In the literature, DAS, SAS, and VAS are often used interchangeably, along with other terms such as angle-flipping. For purposes of this thesis, VAS describes a series of 1D experiments collected at different spinning angles, while SAS is a two-dimensional experiment where the indirect and direct dimensions are collected at different angles, often but not necessarily off the magic angle during the evolution period and on the magic angle during the detection period of the pulse sequence. DAS is reserved for the particular case where the angles are chosen to eliminate the first-order and second-order parts of the quadrupolar interaction. The problem posed by the second-order interaction is that its spatial dependence is described by a fourth-rank tensor, such that spinning about any one angle in physical space is insufficient to perform complete spatial averaging, as shown in Figure 1.1. The spins evolve at a rotor angle β_1 during time t_1, the magnetization is stored along the $z-$axis, and then the signal is collected at angle β_2 during time t_2. Given the appropriate choice of spinning angles (usually 30.6° and 70.1°) and evolution times, evolution

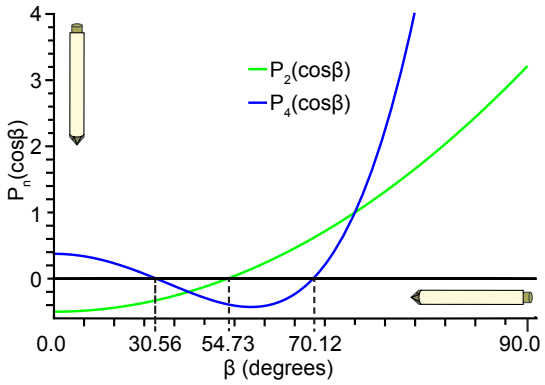


Figure 1.1: The values of $P_2(\cos\beta)$ and $P_4(\cos\beta)$ are plotted over a range of 0 to 90°, relative to B_0. The angles where the respective functions go to zero are indicated with dashed lines. There exists no one sample spinning angle for which $P_2(\cos\beta)$ and $P_4(\cos\beta)$ are both averaged out, which has important implications for spatial reorientation experiments.

under the second- and fourth-rank terms during the second evolution period will cancel that from the first, producing an isotropic spectrum in the indirect dimension, correlated with the anisotropic lineshape in the direct dimension; refinements include pure absorption mode lineshapes [2] and suppression of sidebands [3]. DAS is experimentally quite different from Double Rotaton (DOR), in which the sample, contained within an inner and outer rotor, is spun about two axes, simultaneously averaging out both second-rank and fourth-rank spatial tensors, although in practice similar information is obtained.

These methods are the subject of several excellent review articles and perspective pieces that focus on different aspects of the methodology, including symmetry considerations [4], issues pertaining to DAS specifically [5], the relative merits of DAS vs. DOR and their advantages over prior techniques [6], and their application to solid-state materials [7, 8]. More recent reviews cover modern implementations of DAS and DOR in the context of how they compare to MAS techniques for quadrupolar nuclei [9, 10, 11, 12]. In light of the quality and thoroughness of these previous contributions, DAS is discussed here primarily in the context of instrumentation development. The main focus of this thesis is SAS and VAS experiments for systems composed primarily of spin-1/2 nuclei and deuterium. Although VAS,

SAS, and DAS experiments can be performed with essentially the same instrumentation, the data analysis and scientific objectives of these experiments can be very different, as will be discussed in the following sections on the particular experiments and some applications. Belying their reputation as somewhat exotic techniques, all of the spatial reorientation experiments can be understood in the broader context of experiments that produce either dipolar recoupling or isotropic-anisotropic correlations, both of which are very common objectives in solid-state NMR. These methods have some important advantages and are worth revisiting given some of the recent trends in NMR spectroscopy, including extensive deuteration, fast sample spinning, and the convergence of solids and liquids methodology.

1.1.1 Dipolar recoupling can be accomplished using pulse sequences, manipulation of the sample spinning, or modification of the sample itself.

Dipolar couplings provide orientational restraints and information about local conformational mobility. In rigid solids, which are usually studied under MAS, homonuclear or heteronuclear dipolar couplings are frequently reintroduced using specialized pulse sequences. Recoupling sequences become more challenging to implement in samples where the dipolar interactions are partially averaged by random molecular motion, and more complex symmetry-based pulse sequences become advantageous. For soluble target molecules, residual dipolar couplings (RDCs) and residual CSAs can be induced using a variety of methods, including by mechanically aligning the macromolecule in a stretched gel [13, 14, 15, 16], or magnetically aligning it in a dilute solution of liquid crystals [17]. For membrane-associated peptides and proteins, additional alignment methods are available (reviewed in [18]), since the protein is associated with an orientable lipid bilayer.

This weak-coupling regime can be accessed in solid samples by spinning only slightly off the magic angle (1-3°), heavily scaling the dipolar interactions. Particular spin systems can be

selected using a combination of frequency-selective pulses and multiple-quantum filtering [19]. The use of multiple-quantum filtering, without which it is difficult to separate dephasing due to weak couplings from incoherent signal decay, is an important detail. Even in this weak-coupling limit, dipolar couplings yield valuable orientational restraints. Dipolar couplings scale with internuclear distance as $1/r^3$ for dilute spin systems, whereas NOEs and dipolar couplings in systems subject to spin diffusion scale as $1/r^6$. In the case of many-spin systems such as protonated, uniformly ^{13}C, ^{15}N-labeled samples, magnetization exchange among ^{1}H [20] and ^{13}C [21] nuclei provide long-range distance restraints for structure determination.

1.1.2 Reintroducing the CSA via a pulse sequence or changing the sample orientation provides structural information

The chemical shift anisotropy (CSA) is a consequence of the non-uniform distribution of electron density in chemical bonds. In solid or oriented samples, the chemical shift has a sensitive dependence on molecular orientation, which may be partially or completely averaged by molecular motions. Thus, under conditions where each CSA pattern can be paired with an isotropic chemical shift for assignment purposes, fitting the CSA lineshape can provide beautifully detailed information about the orientation and mobility of particular chemical sites [22]. Despite its obvious utility, this becomes problematic in complex molecules containing many unique sites. When the spinning speed is less than the width of the anisotropic spectrum generated by the chemical shift anisotropy, spinning sidebands appear at multiples of the spinning speed, complicating the spectrum and reducing the effective signal-to-noise ratio, but allowing recovery of the CSA pattern using sideband analysis [23]. When the spinning rate is larger than the width of the pattern (in practice, this means that the sidebands have less than about 10% of the centerband intensity), direct analysis of a 1D spectrum is more difficult; the resulting spectrum has better resolution and reduced overlap between signals from different nuclei, but at the expense of losing the information contained in the sideband manifolds. Several sidebands are needed to determine the principal tensor values,

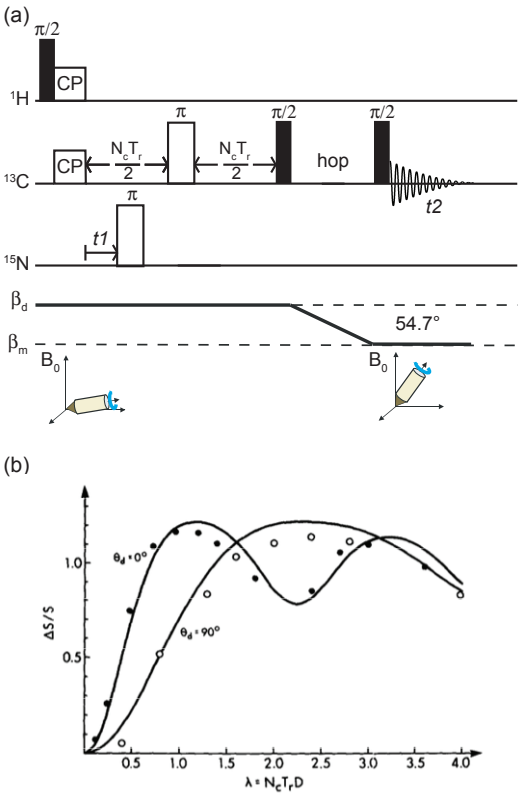
with about 5 sidebands being sufficient to determine the anisotropy, and 6-10 for reliable determination of the asymmetry (η) [24].

Numerous methods have been developed to overcome this limitation including the 1D TOSS (total suppression of sidebands) experiment [25], and refined versions that improve its performance at higher MAS rates and for samples with short T_2 relaxation times [26, 27, 28]. In the PASS (phase-adjusted spinning sidebands) experiment, the phase of the sidebands is adjusted according to the sideband order [29]. Numerous other sequences for observing the CSA patterns under fast MAS have been developed. The chemical shift recoupling experiment of Tycko et al. [30] produces a CSA pattern in the indirect dimension corresponding to that of a well-resolved static spectrum, from which the principal tensor values can be straightforwardly read from the peak and shoulder positions of the CSA pattern. This experiment has given rise to several variants, including a version that is more tolerant of pulse-length errors [31] as well as a constant-time implementation [32]. Another approach correlates the isotropic (fast MAS) spectrum to an "amplified" CSA pattern mimicking that of a sample spinning at a fraction of the actual MAS rate [33]. Many other modern developments have focused on symmetry-based pulse sequences that exploit the different symmetry properties of the CSA and heteronuclear dipolar interaction [34, 35, 36], not only allowing selective reintroduction of these effects, but also enabling determination of their relative tensor orientations [37]. Recoupling the CSA in the context of multidimensional experiments under fast MAS [38, 39] has led to the simultaneous determination of many chemical shift tensors in nanocrystalline proteins [40], and high-quality structures based on the associated orientational restraints [41]. CSA restraints have also been used to improve the quality of structures solved for membrane proteins in phosopholipid bilayers [42]. Introducing residual CSAs can also be accomplished by spinning the sample only slightly off the magic angle [43, 44], as has recently been exploited to investigate the membrane curvature induced by the influenza virus M2 protein [45].

1.1.3 Principles and experimental details of motional reorientation experiments

1.1.3.1 Dipolar couplings

Recoupling of the dipolar coupling between two different rare spins (e.g. ^{13}C and ^{15}N) can be accomplished using rotational echo double resonance (REDOR) [46], rotary resonance recoupling (R^3) [47], symmetry-based sequences [48], and numerous variations for specialized purposes. Historically, all such pulse sequences based on a rotor-synchronized train of π-pulses were very susceptible to π-pulse imperfections, off-resonance effects, and fluctuations in the rotor speed. The spin echo, double resonance with flipped spinning (SEDORFS) was developed to circumvent some of these limitations [49] (shown in Fig. 1.2). SEDORFS is one of a class of experiments that provides simplicity in the pulse sequence at the cost of increased instrumental complexity; here, a single π pulse on each heteronucleus per rotor period is used. The π pulse on ^{13}C, applied halfway through the dipolar evolution period, refocuses the carbon chemical shift. The ^{15}N pulse then dephases the ^{13}C magnetization. In the REDOR-like variant of this method, the carbon and nitrogen π-pulses are coincident. Experiments are performed in the presence and absence of these ^{15}N dephasing pulses, with the ratio of these signals providing a measure of the ^{13}C-^{15}N dipolar couplings. A SEDORFS experiment performed parallel to B_0 is therefore equivalent to a static SEDOR experiment with evolution time 2τ in principle; however in practice the SEDORFS experiment was not performed at 0° because it was done using a modified MAS probe with a solenoid coil. Another variation of this experiment is analogous to dipolar SAS, where the dipolar evolution period is incremented in multiples of the spinning speed, yielding a dipolar Pake pattern in the indirect dimension after a 2D Fourier transform. Interpretation of the SAS-like version is more straightforward, but its implementation requires that the spinning speed be greater than the width of the Pake pattern. Both types of SEDORFS experiment are most straightforwardly applied to systems where there is only a small number of chemically



[t]

Figure 1.2: (a) The SEDORFS pulse sequence starts away from the magic angle. ^{1}H-^{13}C cross-polarization is followed by a dipolar evolution period with a π-pulse on ^{13}C to refocus isotropic chemical shift. The ^{15}N π-pulses induce dipolar dephasing. At the end of the evolution period, the magnetization is stored along the z-axis during the angle switch, and then read out at the magic angle. The experiment is performed with and without the ^{15}N dephasing pulses, and the dipolar coupling value is obtained from the ratio of the signal obtained ΔS/S, similar to the data acquisition scheme in REDOR. Figure adapted from [49] (b) Representative SEDORFS data for the model compound [L-4 ^{13}C-, amide- ^{15}N] asparagine, showing the effect of the spinning angle on the dipolar couplings. Here the experiment was performed with the spinning angle during the evolution period set to 90° (open circles). The SEDOR experiment (filled circles) is equivalent to SEDORFS at 0°, which was not directly performed because it was not possible in the probe used, which had a solenoid coil. The solid lines represent the corresponding theoretical curves. Figure reprinted from [49].

different observed spins, and where the dephasing spins have greatly different chemical shifts.

1.1.3.2 CSA

The chemical shift anisotropy (CSA) contains useful information about chemical bonding, torsion angles, and many other aspects of molecular structure [22]. In relatively simple systems, the CSA lineshape can be resolved in a static spectrum and fitted to obtain chemical information. The Herzfeld-Berger method is a refinement of this technique in which slow MAS is used, and the CSA pattern is recovered from the spinning sideband intensities [23].

Spinning at high frequencies, but off the magic angle also produces a scaled CSA pattern which is especially useful when $90° > \beta > \beta_m$. This approach makes it possible to resolve break points in otherwise overlapping CSA signals [44]. This angular range in particular is useful because of the fact that the scaling factor applied to the CSA pattern ranges from 1 at 0°, parallel to B_0, to 0 at the magic angle, to -1/2 at 90°. Thus, strongly overlapping signals can not only be scaled, but also flipped in sign by spinning closer to 90° with respect to the static field, resulting in overlapping break points being moved to a less crowded region of the spectrum where they can be more easily resolved. This method of finding the principal components of the CSA tensor was demonstrated for polycrystalline samples of the model organic compounds 1,3,5 trimethoxybenzene and p-dimethoxybenzene, both of which were unresolvable by MAS alone at moderate MAS frequencies [43], and for cadmium complexes [50]. The methodology established in this proof-of-principle study was then used to determine the principal components of the CSA tensors in substituted polycylic aromatic hydrocarbons (PAHs), which are difficult to study via NMR because of the long ^{13}C longitudinal relaxation times for unprotonated carbons.

1.1.4 SAS

1.1.4.1 Dipolar couplings

In dipolar SAS, the indirect dimension is obtained off the magic angle, often parallel or perpendicular to B_0, and the direct dimension is detected on the magic angle. This method can be used to measure internuclear distances in solids characterized by strong dipolar couplings [51, 52]. The simplification of these complex spectra enables measurement of even very small dipolar couplings in strongly coupled systems, allowing observation of the fine structure in the powder pattern due to molecular motions and interactions with distant spins.

During t_1, the coherence (usually generated by an initial cross-polarization step, although other methods such as direct polarization or INEPT may also be used) is allowed to evolve under the influence of homonuclear dipolar couplings, often with a π-pulse introduced halfway

through t_1 to refocus chemical shift. The π-pulse is usually not rotor-synchronized, so it introduces sidebands in the ω_1 dimension. After the evolution period, a $\pi/2$ pulse is applied to store the magnetization along the z-axis while the rotor axis is switched to the magic angle. Another $\pi/2$ pulse then reads out the signal, and detection during t_2 occurs while spinning at the magic angle. Proton decoupling is employed throughout, except during the switching time. This experiment results in an isotropic-anisotropic correlation with an isotropic spectrum in ω_1 and a Pake pattern in ω_2. This method was used to determine ^{13}C-^{13}C distances in ^{13}C-labeled zinc acetate and in a 14-residue peptide enriched with ^{13}C at two sites [51]. Distances between 2.4 - 4.2 Å were observed. Although the dipolar Hamiltonians in these systems are truncated at first order (i.e. the relevant dipolar couplings are smaller than the chemical shift differences), complications arise in the peptide spectra because the dipolar couplings are comparable to the homogenous line width governed by the carbon T_2 values. In this case, the dipolar couplings cannot be measured from singularities in the ω_1 spectrum and must be obtained by fitting the Pake patterns and comparing them with the simulated spectra expected from different internuclear distances.

1.1.4.2 CSA

SAS approaches can also be used to correlate the isotropic spectrum to the CSA powder patterns, as first described by Bax et al. [53]. In this experiment, cross-polarization was performed perpendicular to B_0, followed by the evolution period t_1. At the end of t_1, the spinning axis is switched to the magic angle, with the ^{13}C magnetization stored along z during the switching time. Since the evolution period is performed at 90°, the CSA patterns are scaled down by a factor of 2 relative to the static spectrum. This can be an advantage because increasing the width of the f_1 dimension ($\approx$ 6.5 kHz in this example) would require a longer evolution period, which reduces sensitivity and can be problematic for the probe and/or sample if high-power decoupling is needed. Representative data illustrating the correlation between CSA powder patterns and the easily-assigned isotropic spectrum are

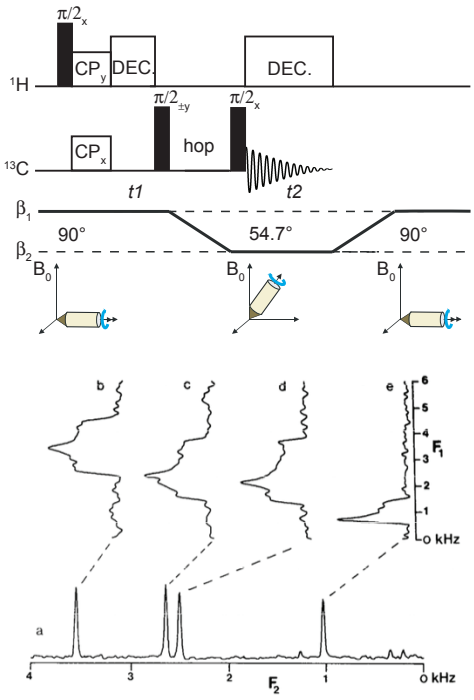


Figure 1.3: Top: SAS pulse sequence used for correlating the CSA powder pattern to the isotropic spectrum. Adapted from [53]. After a CP preparation period, magnetization evolves under the CSA at a spinning angle of 90°. The magnetization is then stored along the $z-$axis while the rotor hops to the magic angle for detection. Bottom: SAS NMR data showing the CSA patterns associated with particular sites in a ^{13}C SAS spectrum of polycrystalline p-methoxybenzene. (a) The isotropic spectrum, collected in F2 at the magic angle. (b)-(e) Slices from the F1 dimension for each isotropic carbon frequency. These powder patterns are half as wide relative to the static spectrum of this molecule because the rotor is spinning at 90°, where the value of $P_2(\cos\beta)$ is $-\frac{1}{2}$. Spectra reprinted from [53].

shown in Figure 1.3.

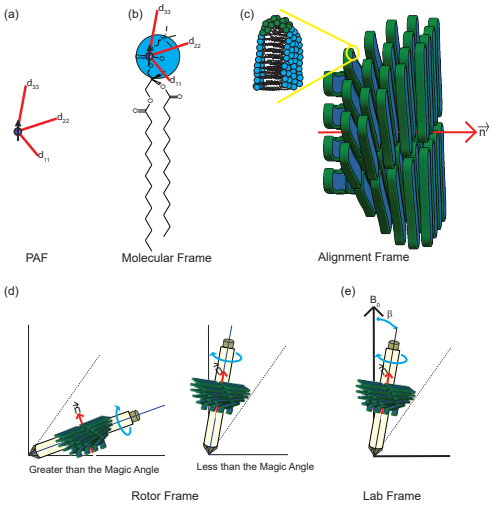
1.2 Goals

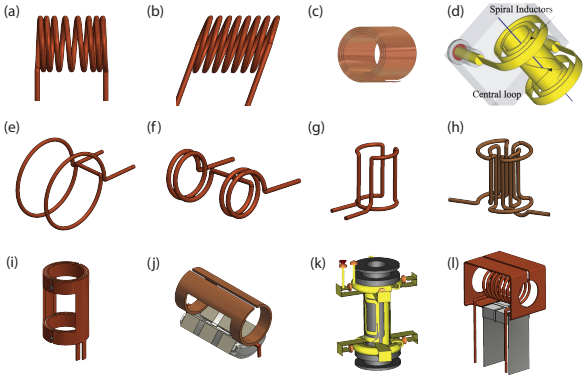
The endgame of this line of research is to solve the protein structures of membrane bound proteins while they are embeded in a membrane mimetic or native membrane. The previous versions of SAS probes in the Martin lab [54] were optimized for investigating liquid crystals and as such have only two channels, ^{1}H and ^{13}C. A triple-resonance SAS probe needs to be designed and built in order to perform the standard experiments used to determine protein structures. These standard experiments can then be built upon to design new experiments to make use of the switching angles, allowing further information (such as CSAs, dipolar

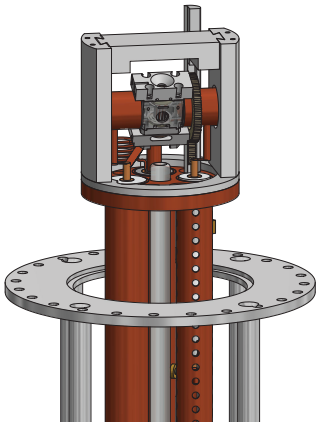
couplings, and orientation information) to be determined more readily.

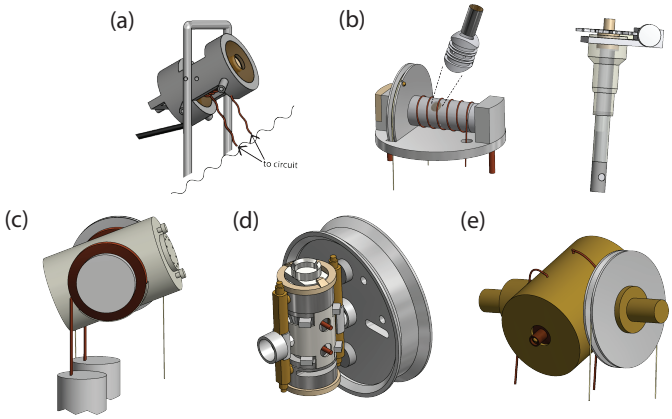
Chapter 2

Evolution of SAS Probes









The chapter cover figure shows some selected figures illustrating concepts discussed in this chapter; (top left) a phosphorus nucleus in a lipid in various frames of reference, (top right) zoo of coils used historically for spatial reorientation experiments, (bottom left) CAD image of a 3 channel SAS probe, (bottom right) zoo of probes used historically for spatial reorientation experiments.

2.1 Instrumentation Development for SAS

In order to obtain the additional anisotropic information (CSAs, dipolar couplings, orientation information, etc.) that Switched-Angle Spinning (SAS) promises to provide, a new design of probe was needed. Probes designs for SAS have evolved over the past few decades, and the current design is based on the developments made for Variable Angle Spinning (VAS), SAS, and Dynamic Angle Spinning (DAS) probes. This chapter discusses these developments before discussing the current design for a triple-resonance SAS probe.

VAS, SAS, and DAS probes are subject to the same design considerations and trade-offs as other solid-state and solution NMR probes; the desiderata for solids and oriented liquids are quite different and result in divergent design decisions. In the case of rigid solids having large quadrupolar interactions or strongly coupled dipolar networks, the most important factor is maximizing the rf field strength delivered for a given input power, either to excite the full bandwidth of quadrupolar nuclei that can have very broad spectra, or in the dipolar case, to provide efficient decoupling of strongly coupled heteronuclei. This is particularly important off the magic angle where dipolar interactions are less attenuated by the sample spinning. In this experimental regime, the samples are usually robust and typical experiment times are relatively short. Minimizing the switching speed is a critical factor because of T_1 relaxation, spin diffusion, and relaxation of magnetization from the centerband to other energy levels in quadrupolar nuclei. For mobile solids or liquid crystals, decoupling is mostly accomplished by random motions of the molecules, and therefore other effects like rf homogeneity and shimming become more relevant.

Regardless of other considerations, the most important design feature of an SAS or DAS probe is the connection between the moving coil, which reorients on a millisecond timescale,

to the stationary part of the circuit that interfaces with the NMR spectrometer. Because SAS and DAS experiments have at least two, and often three, Fourier dimensions, this connection must be able to withstand hundreds of thousands of angle flips without failing or compromising the rf stability. In addition to mechanical robustness of the angle switching mechanism, accuracy and reproducibility of the angle setting is needed in all cases in order to completely cancel the anisotropic interactions. A variety of SAS/DAS designs, both historical and currently used, represent different solutions to the problems presented by these experimental criteria.

Most SAS/VAS/DAS probes are based on established MAS designs, many of which use solenoid coils. Solenoids have many advantages, including their ease of construction, strong rf fields, and reasonably high rf homogeneity (with careful attention to spacing of the turns). They are also generally more efficient than the saddle coils more commonly found in liquids probes in terms of the B_1 strength produced at a given power level, yielding higher sensitivity even though the effective rf field produced at the magic angle is scaled by a factor of $\sin 54.7^\circ = 0.816$ [55]. However, solenoid coils have an important disadvantage in variable angle experiments; because the B_1 field oscillates in the direction along the coil axis, only the field component perpendicular to B_0 interacts with the spins. The 90° pulse length at angle β (p_β) is therefore related to the measured pulse width at 90°, p_{90}, by the expression $p_\beta = \frac{p_{90}}{\sin(\beta)}$.This is not a problem in an MAS probe, where the transverse component is still relatively strong, but in VAS or SAS, the 90° pulse width becomes angle-dependent, such that the 90° pulse with must be calibrated at each angle. Performance is therefore sharply degraded as the coil approaches the B_0 direction (illustrated in Figure 2.1). This situation is frustrating when it is desirable to measure the unscaled values of an anisoptric parameter such as the dipolar couplings or CSA. In such cases, transverse coil designs that provide a constant rf field perpendicular to B_0 regardless of the coil orientation are a better choice, usually with some compromise in rf field strength or homogeneity. The split solenoid, which is illustrated in Figure 2.1, consists of a solenoid with a gap in the middle that is wide

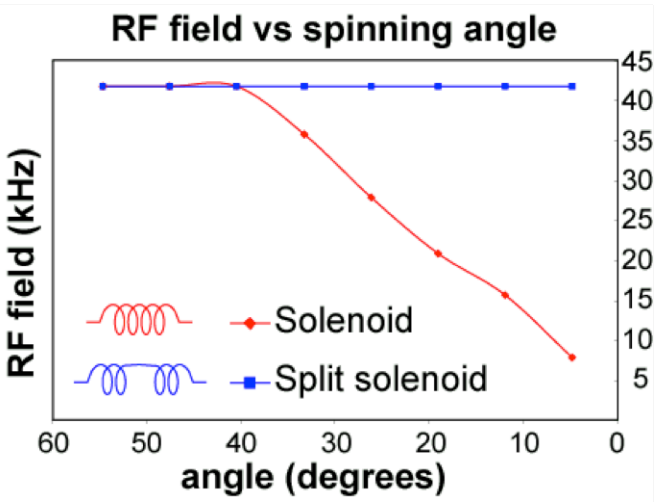


Figure 2.1: Measured B_1 field strengths for a solenoid and a split solenoid (an easily constructed transverse coil) as a function of the angle between the coil axis and B_0 (from [56]). Data were collected on adamantane using a 4 mm rotor in a sliding contact SAS probe similar to the one used in [57].

enough to admit the rotor, which is oriented perpendicular to the coil axis. This coil is easy to fabricate and produces a strong, orientation-independent rf field; however it suffers from reduced rf homogeneity and is therefore not ideal for experiments using many pulses.

Given these design constraints, two broad categories of coils have been used for VAS/SAS/-DAS probes: solenoid-like coils and transverse coils. Schematics of many of these coils are shown in Figure 2.2. Solenoid-like coils have the rf field along the long axis of the coil; however, many variations with different advantages have arisen. A variable pitch solenoid has a much better field homogeneity than that of an evenly-spaced solenoid of the same length and number of turns [58]. The solenoid can also be tilted such that the plane of each loop is not perpendicular to the coil axis [59], which also tilts the rf field it generates. Alonso, et al. have incorporated the Sun and Maciel design into their Double Helix Dipole coil [60], which consists of two coaxial tilted coils that together create an rf field orthogonal to B_0. Other designs seek to reduce the electric field generated by an unbalanced solenoid, especially when the sample behaves as a lossy dielectric, as is the case for most hydrated crystals or precipitates of biological macromolecules. The scroll coil of Stringer et al. is much more homogeneous than a standard solenoid and also has a lower E-field, causing less rf heating

in conductive samples [61]. Another low-E design is the Z-coil of Dillman et al., which is more homogeneous than either an evenly-spaced or a variable pitch solenoid [62]. On the other end of the spectrum are designs where the coil itself is solenoid-like, but the sample axis is perpendicular to that of the turns of the coil, providing a strong rf field at all angles, at the expense of homogeneity and/or sensitivity. A Helmholtz coil is a pair of coils placed at a distance equal to their radii. It is very homogeneous over a limited volume, though it suffers from a low filling factor and thus has limited efficiency. A split solenoid allows the rf field to be transverse to the sample axis, but has low homogeneity due to the shape of the field near the center of the sample.

The other category of coils, more commonly associated with solution-state NMR but finding greater applicability in solid-state probes as well, consists of those with an rf field transverse to the axis of the sample. A single-turn saddle coil has a good filling factor and reasonable homogeneity. A double or triple saddle coil has even better homogeneity than a single turn saddle coil, and also performs better at lower frequencies due to its higher inductance. Kentgens et al. have made use of this approach with a homebuilt DAS probe [63]. The rf circuit in their version uses a three-turn saddle coil with an inner diameter of 4.1 mm and sliding contacts with the leads colinear with the stator rotation axis. The rotor orientation is performed by a pulley system attached to a computer-controlled brushless servo motor.

A double saddle coil has been used in a capacitively coupled double-tuned VAS/ SAS probe for use with mobile solids and liquid crystals [54], with relatively good homogeneity. Some transverse coils are also designed to have a low electric field, such as the early exemplar of the genre, the Alderman-Grant resonator [64]. Its low inductance design greatly reduces rf heating in conductive samples. A similar slotted tube resonator was used in a double-tuned capacitively coupled SAS probe for liquid crystalline samples. The low inductance of this type of coil, however, makes it only effective for detection of high-γ nuclei. The Modified Alderman-Grant Coil [65], with its capacitive bridge, has a low inductance. This makes it particularly well-optimized for higher-γ nuclei.

Solenoid-like coils and transverse coils are frequently combined in static and MAS probes to create a coaxial pair of crossed coils. This type of design has often been used in static probes for oriented samples; a very successful recent version that can be used in both static and MAS probes is Gor'kov's E-free design [66] for a multi-channel probe. This geometry enables the use of a low electric field resonator for the high frequency channel and a solenoid capable of higher rf power for the other channels. Disadvantages include space constraints for the coaxial coils and reduced filling factor of the outer coil. One method for dealing with this, while at the same time increasing the filling factor of the inner coil, is to make the inner coil out of copper foil and then wrap the coil around the stator with capacitors connecting the ends of the foil pattern [67]. This type of design has the potential to be used in SAS probes for deuterated solids and oriented liquids, for the same reasons it has gained popularity in MAS and static solids probes.

Unlike in VAS, where the probe may be retuned at every angle, for all SAS/DAS probes, stability of the circuit tuning upon changing the angle is an important design consideration. This is illustrated in Figure 2.4 for a tuning-tube probe with sliding connections and a stepper motor, based on the design of Medek et al. [57]. A schematic of the angle-switching mechanism for this probe, which uses pulleys connected by stretch-resistant Spectra cord [73] and a stepper motor mounted on a long actuator to minimize its exposure to the strong B_0 field, is shown in Figure 2.4a. The frequency response of the proton channel (500 MHz Larmor frequency) was measured using a Hameg frequency analyzer and is plotted in Figure 2.4b. The probe was initially tuned with the stator positioned at the magic angle, and the angle was then adjusted to 0 ° and 90° using the attached stepper motor, without retuning. It is possible to achieve sufficient tuning stability using sliding contacts, flexible leads, or capacitive coupling; however the moving connection is made, this is an important benchmark for a DAS/SAS probe.

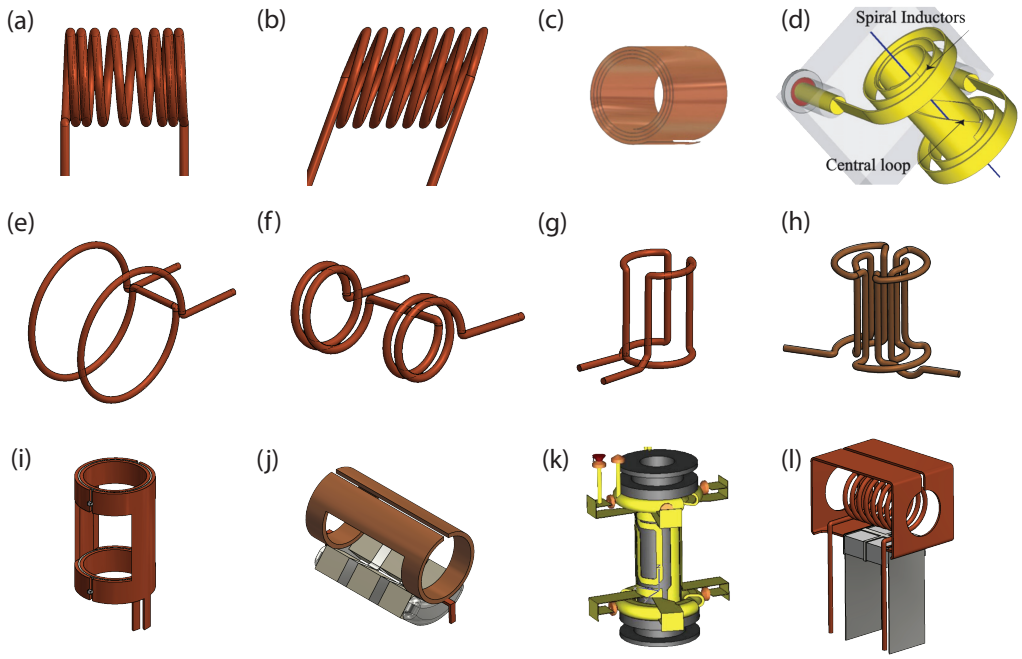


Figure 2.2: An NMR probe coil zoo showing both solenoid-like and transverse coils from the literature, all of which have the potential to be used in SAS/DAS probes depending on the application: (a) Variable pitch solenoid [58] (b) Tilted solenoid [59] (c) Scroll coil [61] (d) Z-coil, figure from [62] (e) Helmholtz coils (f) Split solenoid coil (g) Saddle coil (h) Double saddle coil [63, 54] (i) Alderman-Grant resonator [64] (j) Modified Alderman-Grant resonator [65] (k) Doty's cross-coil design on a ceramic stator, figure from [67] (l) Gor'kov's crossed coil assembly [66]

.

2.1.1 Rigid solids

Many of the various coils used for other solid-state NMR experiments have also been used in VAS/SAS/DAS probes, with different strategies for managing the mobile connection points. A selection of DAS/SAS probes designed for rigid solids is shown in Figure 2.3. The first DAS probe described used a single-tuned solenoid coil wrapped around the spinning assembly, with a sliding contact between the leads emerging from the moving rf coil and the stationary leads attached to the matching network [74]. A refinement of this design used a double-tuned solenoid and beryllium-copper leads that maintain contact with the hopping coil by undergoing a "watch spring" motion [74]. In both of these designs, the 90° pulse width depends on the coil orientation. As is clear from Figure 2.1 the rf field drops off rapidly as β approaches 0°. In the particular case of using a solenoid coil with a relatively high filling

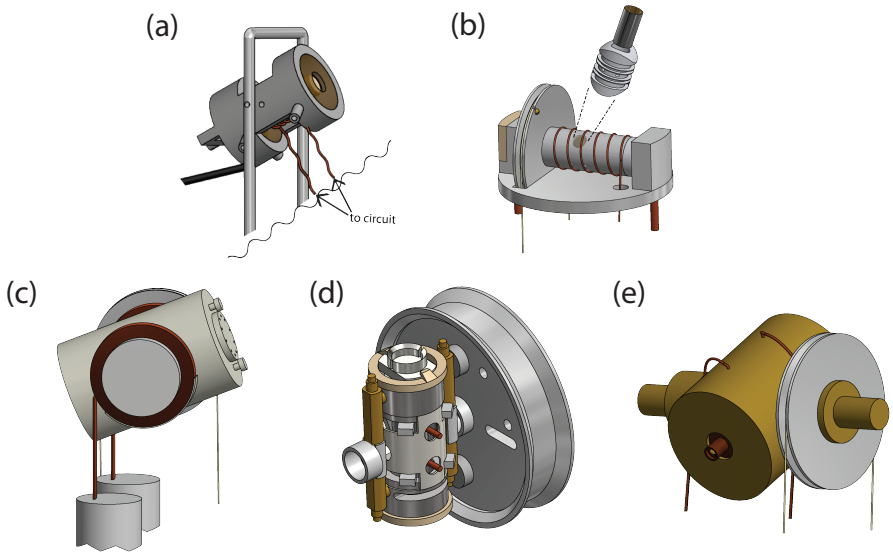


Figure 2.3: SAS/DAS probes for rigid solids: (a) Gerstein's DAS probe enhanced the spinning stability with the addition of another bearing air inlet at the bottom of the standard rotor design [68]. (b) Mueller's DAS probe delivers a consistent B_1 field independently of the spinning angle by using a stationary rf coil wrapped around the entire stator assembly [69]. (c) Medek's DAS probe makes use of robust sliding contacts allowing for use over millions of rotor hops [57]. (d) Doty's HR-VAS probe utilized a cross-coil design to increase the sensitivity of the probe. It is also remarkable for its maximization of homogeneity and elimination of shimming artifacts from the stator assembly [70]. (e) Grandinetti's magic angle flipping (MAF) probe (a descendent of the probe described in [71]) has very precise control of the acceleration during angle switching [72].

factor at the angles required for DAS, this dependence does not present a major problem. For the configuration used in this paper, p_{90} was about 3 μs, and p_β was about 5.5 μs at $\beta_1 = 37.38°$ and 4 μs at $\beta_2 = 79.19°$. These correspond to rf field strengths of 83.3 kHz and 45.5 kHz, respectively, which is sufficient for performing DAS experiments on quadrupolar nuclei with small to moderate quadrupolar linewidths. The leads described in this approach were reported to withstand about three million hops before requiring replacement, corresponding to several DAS experiments.

Further innovations building on this strategy have included more robust flexible leads, e.g. [75] and sliding contacts, e.g. [71]. Two robust sliding contact DAS probes were developed in the 1990s, one based on the "watch spring" connection [71] and used for many pioneering DAS and SAS experiments and the other used by Medek et al. for DACSY [57]. The Medek

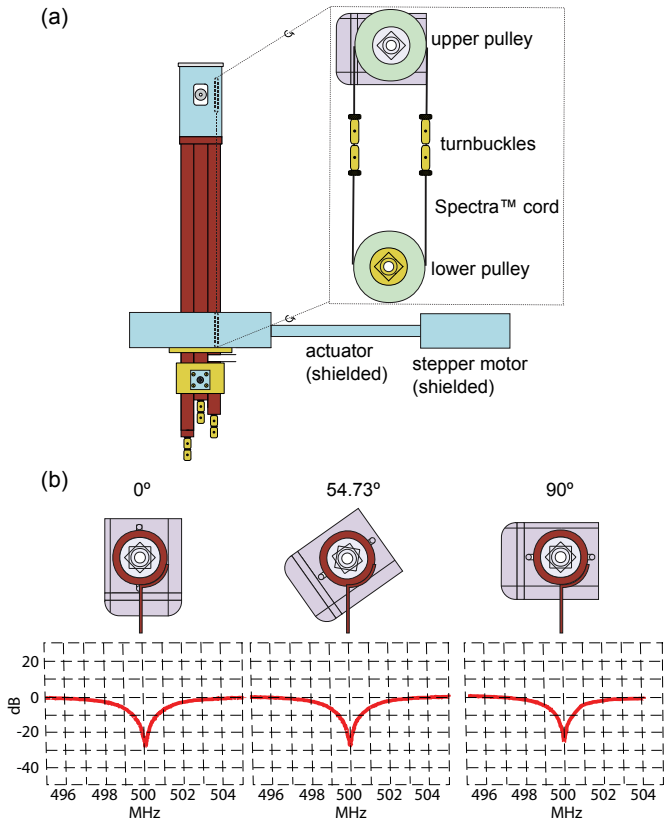


Figure 2.4: (a) A schematic of one strategy for accomplishing angle switching in a DAS/SAS probe. The position of the stator is controlled using pulleys at the top and bottom of the probe assembly, connected to a stepper motor by way of a long brass actuator that keeps the stepper motor from coming too close to the magnet. (b) Tuning stability at a Larmor frequency of 500 MHz for the probe depicted in part (a), measured using a Hameg frequency analyzer after tuning at the magic angle.

probe used a multiply tuned circuit based on a Doty stator and the connection between the rotor and the stationary part of the circuit is made via rings soldered to either side of the coil sliding on beryllium-copper leads. This design was found to be robust over millions of rotor hops, which makes it well suited to perform long multidimensional experiments such as DACSY. In DACSY, several independent 2D DAS experiments, each with a few hundred scans for signal accumulation, are performed over the angle range of $18° \leq \beta_1 \leq 39°$ and $65° \leq \beta_2 \leq 90°$. The probes first used for these experiments were equipped with 5 or 7 mm rotors, spinning at 5 or 10 kHz, respectively. The 90° pulse widths were 1.7-4 μs, depending on the angle between the solenoid coil and B_0. The hopping time required was 45 ms.

The Eastman et al. probe, which was a double-tuned single coil design, was similarly robust.

This probe was used for a variety of applications, including DAS of quadrupolar materials and SAS for measuring internuclear distances [51] as well as investigating chemical exchange [76]. This design used a single solenoid made of copper ribbon wound on the outside of the stator, and connected to the stationary part of the circuit by beryllium copper strips that both bend and slide along the flattened wire of the coil. Because of its solenoid design, the 90° pulse widths of this probe varied from 4-10 μs, depending on the angle used. This probe used a Doty 7 mm MAS rotor system [77], and its mimimum switching time was 30-50 ms, depending on the angles chosen. More recent versions incorporating many of the positive features of this probe have been used for magic angle flipping experiments on solid materials, with smaller rotors, faster hopping times, and better rf efficiency [72, 78]. Commercial implementations are available from both Doty Scientific and Revolution NMR. Another type of single-coil, flexible lead design [68] makes use of high-power transmission lines to provide isolation in multiply resonant circuits [79]. Here the term "transmission line probe" refers to a circuit where the properties of transmission lines are used to provide rf isolation among the different channels [79], as opposed to a "tuning-tube probe," a different strategy in which transmssion line segments are used in lieu of traditional variable capacitors and inductors in order to increase power-handling capability and save space inside the magnet bore [80]. In the McKay transmission line type of design, the sample coil is located at the end of a large diameter transmission line having a length that is a multiple of the high-frequency $\lambda/2$. This enables the tuning and matching elements for this channel (often traditional variable capacitors) to be placed outside the magnet bore. The rf for each lower-frequency channel is fed in through its own transmission line, each of which is connected to the high frequency line at a position corresponding to a node in the high-frequency standing wave. This kind of probe circuit has inherent isolation due to the positioning of this tap on the main transmission line. Multichannel VAS/SAS probes have been constructed using this strategy, taking advantage of its inherently high power handling and good isolation [49, 81]. Another solution to the problem of delivering a consistent B_1 field independent of the spinning

angle is to use a stationary rf coil surrounding the stator assembly, at the expense of filling factor and thus efficiency of the sample coil. A probe of this design was described and its performance demonstrated using ^{2}H VAS and ^{17}O and ^{27}Al DAS [69]. In this design, the stationary coil was a large solenoid oriented perpendicular to B_0 and entirely outside the moving stator assembly. This approach provides a straightforward solution to the problem of connecting the electronics, given that it contains no moving connections, but it is likely impractical for most modern applications, where samples are precious and in short supply, requiring the filling factor to be maximized.

SAS experiments can suffer from instabilities in the rotor spinning speed, particularly in the case of oriented sample NMR, where slow but very stable spinning speeds are desired. Many techniques for stabilizing MAS have been developed and can be applied to SAS as well. For example, buffering the air supply to the spinning controller by using multiple smaller ballast tanks in series rather than large ones can damp out pressure waves generated by the air compressor being turned on and off [82]. Typical spinning speed controllers consist of a feedback circuit that adjusts the flow of drive gas using a mechanical valve, in response to an optical signal [83]. In systems with optical spinning speed detection, painting multiple marks on the rotor such that the apparent spinning speed measured is a multiple of the actual spinning speed enables errors to be found and corrected by the automated spinning controller when they are smaller relative to the actual rotor frequency [84]. Very precise control of the spinning speed can also be achieved by varying the temperature of the drive gas [85], but this approach may be problematic in the case of liquid crystalline samples where small changes in temperature affect the phase behavior and alignment of the sample as well. The paper by Mueller et al. [69] is notable for the careful attention given to minimizing the switching time and stabilizing the spinning. The switching is controlled by a stepper motor, which had precedent in earlier designs, e.g. [53], where a rotor spinning at 2.3 kHz was switched and stabilized within less than one second. Mueller et al. very carefully analyzed the factors limiting the switching speed and the time required to stabilize the rotor. The

authors determined that the minimum reorientation time for a change in angle on the order of $\pi/4$, given optimal control of the switching system, is in the range of 1-4 ms. In practice the limiting factor appears to be achieving precise enough control of the acceleration and deceleration of the stator during the switch. In the Mueller et al. design, switching was accomplished using a pulley system driven by a stepper motor, and switching (including rotor stabilization) required 28 ms to hop a 6.3 mm diameter rotor by 41.74°, the change in spinning angle required by a DAS experiment. Based on analysis of the torque generated by reorienting the rotor, rotor stability was found to be limited by turbulence in the bearing air flow, rather than by failure of the bearing air to support the extra load produced by the rotor during the hop.

The Gerstein SAS probe [68] was also notable for its fast switching with stable spinning: the angle can be changed by 45° in 9 ms with spinning up to 10 kHz, with less than 1% variation in the spinning speed, although the stator assembly "ring" for 25 ms. This probe used a modification of the spinning system designed by Wind et al. [86, 87], where an additional bearing air inlet was added at the bottom of the standard cylindrical rotor design. In the case where slow (≤ 1 kHz), stable spinning is desired, spinning stability can also be enhanced by using drive tips with fewer fins and/or plugging some of the drive air inlet holes.

2.1.2 Liquid crystals and semisolids

VAS experiments on liquid crystals and membrane samples have a different history and purpose from those on rigid solids. Unlike in traditional MAS, the function of the spinning here is to align the director with the rotor axis. Averaging of anisotropic interactions is mostly performed by rotational diffusion, with an assist from rf decoupling. The spinning required for this application is quite slow compared to typical MAS rates, on the order of tens to hundreds of Hz. Previous methods for aligning liquid crystals, including the use of electric fields and spinning the sample perpendicular to B_0 and slower than the critical speed, did not allow orienting the director close to the magic angle. The first VAS experiments on

liquid crystals were performed using an electromagnet with a ^1H Larmor frequency of 80 Hz [88, 89]. A stator assembly was constructed that was capable of spinning 10 mm samples tubes about any axis relative to B_0, allowing the director dynamics to be investigated over the full range of angles. Strong dipolar couplings, and therefore second order spectra, were observed when the sample was spun far from the magic angle. A rotor axis a few degrees off from the magic angle scaled the dipolar couplings, recovering the first-order spectra and enabling the determination of the signs of J and D. This approach takes advantage of the mobility of the sample, making it an asset instead of a liability. Long-term, this method has the potential to enable RDC measurements and structure determination of membrane proteins, which do not assume their biologically relevant conformations in isotropic solution.

In contrast to the DAS/SAS probes built for experiments on solid materials, SAS probes optimized for studies of liquid crystals or semi-solid samples are subject to very different desiderata, other than accuracy and reproducibility of the angle setting, which are equally important in this context. The major considerations are rf homogeneity, spinning stability, temperature control, and robustness of the probe. Because the samples are partially mobile, rotational averaging narrows the line widths considerably, eliminating the need for high-power decoupling. Furthermore, the nuclei involved are usually not quadrupolar, aside from ^{2}H, which has a relatively small quadrupolar interaction. For experiments on oriented membrane samples and other liquid crystals, the filling factor must be maximized to use the small samples available and the possibility of breakage or instability in the leads must be minimized during long experiments with many points in the indirect dimension. The increased stringency of these requirements is somewhat compensated for by the decreased necessity for high excitation or decoupling fields. Some examples of probes built for this purpose are shown in Figure 2.5.

In Variable Angle Spinning (VAS) of liquid crystals, the alignment is dominated by the shear forces produced by spinning; the molecules align parallel or perpendicular to the spinning axis depending on the sign of the magnetic susceptibility anisotropy. Changing the spin-

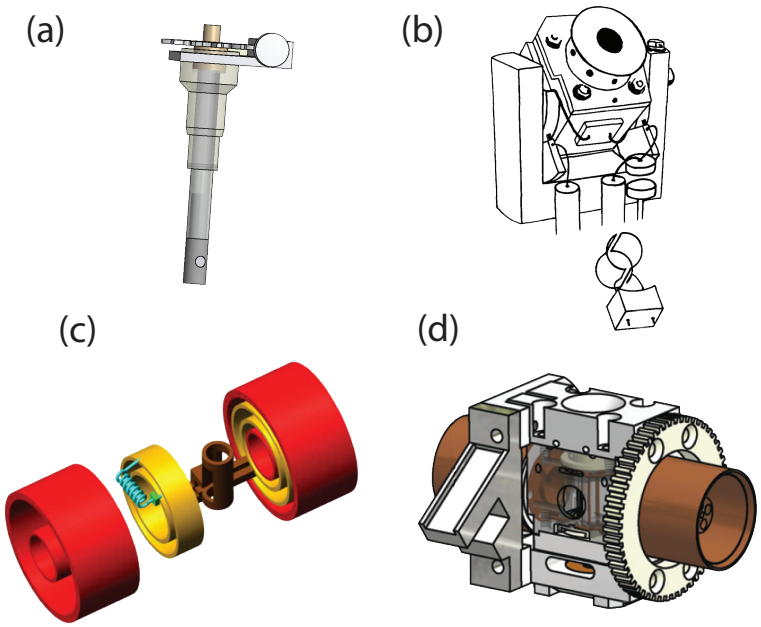


Figure 2.5: VAS/SAS probes for liquid crystals and semisolids: (a) The Väänänen et al. VAS probe provided high resolution and fine control of the spinning angle allowing for very precise measurements of the CH J-couplings [90]. (b) The Tomaselli et al. DAS probe uses sliding copper contacts and a pair of gold pantographs to transfer the rf pulses from the circuit to the transverse coil [91]. (c) The contactless resonator VAS/SAS probe utilizes capacitive coupling, eliminating the need for sliding contacts or flexible leads [92]. (d) Litvak's pneumatic SAS probe also makes use of capacitive coupling, but uses an air-driven switching mechanism instead of a stepper motor, making it possible to place the switching mechanism closer to the spinner module and reducing the hysteresis associated with the actuator [54]. Doty's HR-VAS probe (shown in Figure 2.3d in the solids probe section) can also be used for semisolids with only minor modifications [70].

ning axis scales the dipolar couplings [93, 94]. An early VAS probe designed for studies of liquid crystals was based on a 10 mm liquids probe, with pneumatic spinning at speeds up to 900 Hz (Figure 2.5a). This probe achieved 3 Hz resolution in the ^{1}H spectrum at an operating frequency of 100 MHz, which represented a significant advance in resolution over previous designs, and enabled very precise measurements of $^1J_{CH}$ and the HCH bond angle in methyl iodide. The angle setting was performed using a pinion angle adjustment mechanism controlled by a screw of precisely measured pitch, such that one turn of the screw corresponded to approximately 2°. The angle setting was measured optically, and was found to have excellent linearity over a broad range of angles [90]. Although this method of setting the angle is highly accurate and reasonably convenient, it has not come into widespread use. An updated version of a similar optical system for monitoring the spinning angle, with more details about its implementation and calibration, has been described by Mihaliuk and Gullion [95]. The first VAS and SAS experiments on bicelles established that their behavior is consistent with that of other lyotropic liquid crystals [96]. This work was performed using a homebuilt SAS probe with 4 mm rotors. This probe used a Helmholtz coil, making the 90° pulse length independent of spinning angle. The angle switching was performed using a servo motor controlled by the pulse programmer, which produced a total switching time of 50 ms, including 40 ms for the switching itself and 10 ms for the rotor spinning to stabilize. Another approach to the problem of high-resolution VAS and SAS is the cross-coil HR-VAS probe of Doty and coworkers [70], (Fig. 2.3d) which uses flexible leads and a pulley-based switching mechanism. Cross-coil designs have often been used in NMR probes due to the absence of mutual coupling between orthogonal resonators and the ability to optimize the two coils for different frequency ranges, increasing the sensitivity of the probe [97]. The classic design utilizes a simple solenoid coil for the low-frequency channel(s) and a low-inductance resonator, such as that designed by Alderman and Grant, for proton decoupling [64]. Many variations have been described, with variations in the pitch and cross-section of the solenoid, and different types of slotted-tube resonators or saddle coils for the decoupling coil. Several

modern versions have been built due to increasing interest in biological solid-state NMR, where samples act as lossy dielectrics and reduced electric fields are necessary to limit sample heating. These probes use a low-inductance coil for the proton channel, minimizing the voltage induced across the sample during high-power decoupling [67, 98, 99, 65, 66]. The HR-VAS probe is an early example of this type of probe, notable for its concinnity of design with respect to maximizing static homogeneity and eliminating sources of shimming artifacts due to the stator assembly and circuit elements near the moving coil. Rotor housing components are made cylindrically symmetric to the greatest extent possible, chip capacitors and exhaust holes are oriented at the magic angle with respect to the coil axis, and the effects on the magnetic field from the diamagnetic manifold rings defining the bearing and drive plenums are compensated by insertion of loops of paramagnetic wire. This design represents a major advance in HR-VAS experiments in terms of both rf homogeneity and flexibility of use for both solid and liquid crystalline samples.

Capacitive coupling is another approach to connecting the mobile and stationary parts of an SAS probe. Contactless resonator SAS probes have a mobile assembly built as a pre-tuned double resonant circuit, with the rf feeds for different channels (e.g. proton and carbon) coming in from opposite sides of the coil as shown in Figure 2.5b. The contactless resonator has no mechanical contacts between the coil and the rest of the circuit. The sliding surface is Teflon, which eliminates stability problems due to mechanical aging of sliding contacts and also minimizes stray reactances during reorientation [92]. This design also has the advantage of being balanced on both channels, obviating some of the problems that can occur with unbalanced circuits at higher field strengths, such as mismatch of the rf profiles of different channels [100], rf-induced thermal gradients, and arcing of the coil at the high voltage points at either end [101]. The sample coil itself is a modified version of the slotted tube resonator originally developed by Alderman and Grant to reduce heating due to rf irradiation of conductive samples. It has the added advantage of providing a homogeneous, orientation-independent field. Because it is a single-loop coil with inherently

low inductance, this type of resonator performs best at higher frequencies. This coil showed good ^{1}H-channel performance and excellent RF homogeneity on both channels, but poor ^{13}C sensitivity, indicating that a higher inductance is necessary for carbon-detected experiments [56].

The second-generation contactless resonator probe addressed two outstanding issues: the need for better performance on the carbon channel and a faster switching mechanism [54] (Figure 2.5d). The required higher inductance is provided by a double saddle coil, which is pictured in Figure 2.2(h). This coil was chosen as a compromise between low frequency and high frequency performance. This probe was designed to deliver a constant rf field strength independent of spinning angle to allow for maximum flexibility in testing sample preparations and experiments for SAS in oriented systems. Because a self-supporting wire gives poor field quality and is prone to misalignment that impairs spinning, the top and bottom of the coil are supported by machined Kel-F tablets. The inner diameter of the ceramic piece must provide enough room for free rotor spinning. In MAS probes, the Hall sensor is usually mounted directly on the stator [102] and calibrated by recording the voltage when the stator is at the magic angle, measured by maximizing the sidebands of KBr [103]. In an SAS probe, the sensor cannot be mounted on the spinner module during an experiment. Thus, the sensor is attached to the stator by a removable bracket and calibrated, with the Hall voltage recorded for a range of angles. Angle switching in most SAS probes has been performed using a stepper motor or servo motor. This provides great flexibility of control over the motion of the spinner; however, the necessity of locating the magnetic motor far enough from the superconducting magnet leads to slower switching and possible failure of the motor. Using pneumatics [91] it is possible to place the switching mechanism close to the spinner module. Solenoid valves replace the motor, and the motion is actuated using pressurized air.

In the Litvak SAS probe, the angle switching is performed by a steam engine-type pneumatic mechanism with a geared connection between the piston and the spinner module. The

cylinder is mounted between the base plate and the middle plate of the probe body. The vertical position of the cylinder is adjusted through a worm drive. The piston travel is set between the bottom of the cylinder and a separate brake disk, which moves inside the cylinder. It is connected to a threaded rod outside the cylinder through a clip. To set β_1 (smaller) and β_2 (larger), one adjusts the vertical position of the pneumatic cylinder and of the brake disk, respectively. Pneumatic switching is achieved by alternately applying pressure and vacuum. Linear motion of the piston is converted into the angle switching of the spinner through a rack-and-pinion gear. The angle switch time depends on the pressure applied, with higher pressure resulting in faster switching. Figure 2.6 shows the time-domain signal of KBr for different values of the switching delay. With $t = 0$ set when the stator begins to move (12 ms from the closing of the valve), it takes 14-16 ms for the stator to reach the magic angle and an additional 1 ms for the spinning to fully stabilize. The minimum time required for the angle switching with 75 psi input pressure is thus 17 ms, including the time for the spinning to stabilize. Given that the ^{1}H longitudinal relaxation time T_1 is on the order of a second for a typical hydrated solid protein sample, this is sufficient for performing the desired experiments. Furthermore, this is the time required for switching from 0° to the magic angle; in order to obtain manageable coupling values, real experiments are likely to require a smaller change in angle. Regardless of how the angle setting and swtiching are accomplished, the hopping time, including the time it takes for the spinning speed to stabilize, is a key DAS/SAS probe benchmark that should be optimized.

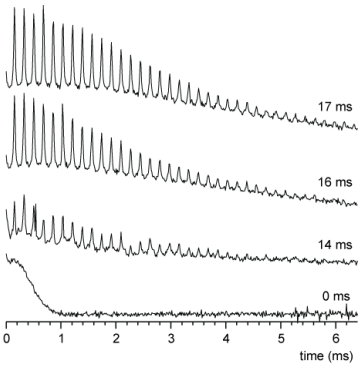
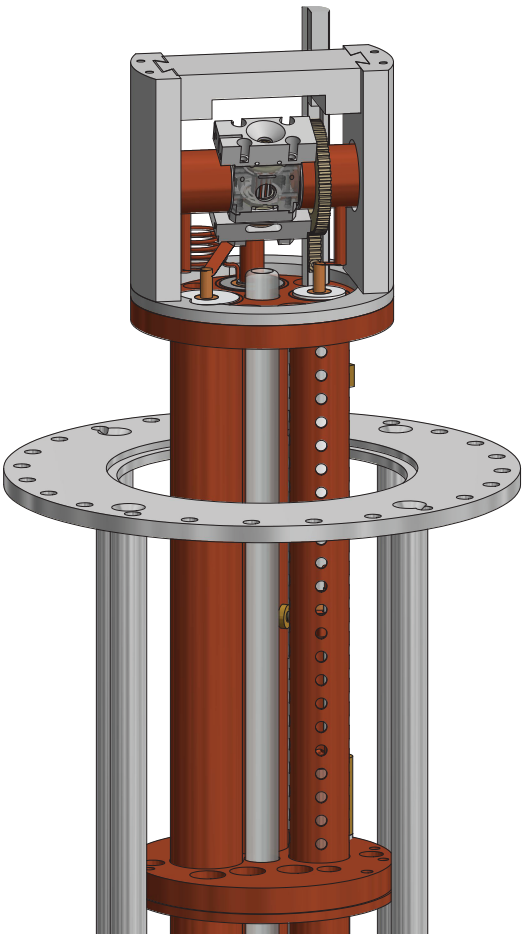
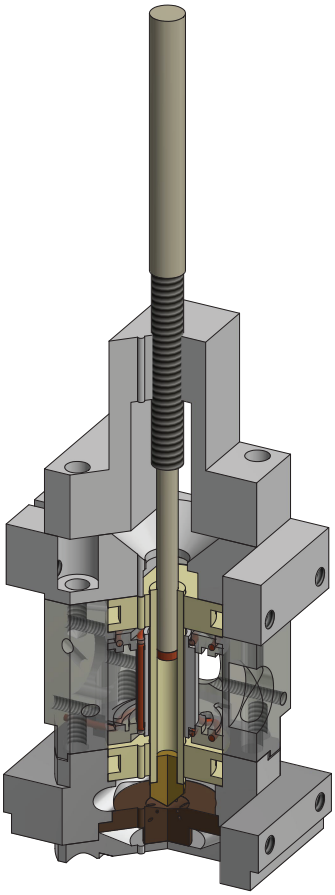


Figure 2.6: Time-domain ^{79}Br signals of KBr as a function of switching time. Here $t = 0$ is counted as the time when the stator begins to move, 12 ms after the TTL pulse signaling the closure of the valve. No rotational echoes can be seen at $t = 0$, as the sample is spinning at 0 °. 14 ms later, the sample is moving toward the magic angle. At 16 ms, the sample has reached the magic angle, but takes an additional ms to fully stabilize, as can be seen by comparing the small rotational echoes near the end of the acquisition time. 16 scans were taken to confirm that the measurement is reproducible. [54]

Chapter 3

Design of a Triple Resonance SAS Probe





The chapter cover figure shows some selected figures illustrating concepts discussed in this chapter; (left) CAD image of an assembled triple resonance SAS probe, (right) ball shift assembly used to test the homogeneity of the magnetic field produced by the probe.

3.1 Designing a Triply-Resonant Probe for SAS

Building on the previous generations of SAS probes built in the Martin lab, this triple-resonance SAS probe will utilize the pneumatic switching mechanism of the Litvak probe [54] and the capacitive coupling to the coil of the Qian probe [92], modified to allow 3 separate channels. Using these previous probes as a template, the most important remaining design considerations are how to configure the capacitive coupling for 3 channels and, closely related, which type of coil(s) will be necessary for that design.

3.1.1 Coils

There are a wide variety of coils that have been used for solid-state NMR probes (see Figure 2.2), but for this design only a small selection of them are useful, shown in Figure 3.1. A variable-pitch solenoid has a reasonably homogeneous B_1 field [104] and can transmit relatively high power rf pulses. This makes it one of the most common coil choices for solid-state NMR probes. Solenoids are axial coils (meaning that the magnetic field produced by the coil is along the axis of the coil), therefore the effective B_1 field is just the x-component of the field vector since NMR is set up to detect in the x-y plane. This is something of a complication for switched-angle spinning because the angle of the coil, and thus the x-component of the B_1 field, will change during an experiment. The change in the effective B_1 field can be accounted for by calibrating the field at each angle and setting the pulse sequence to use different power levels for the pulses at each angle. While somewhat tedious during the initial setup, this does make solenoids a valid choice for use in SAS, but there are also other options.

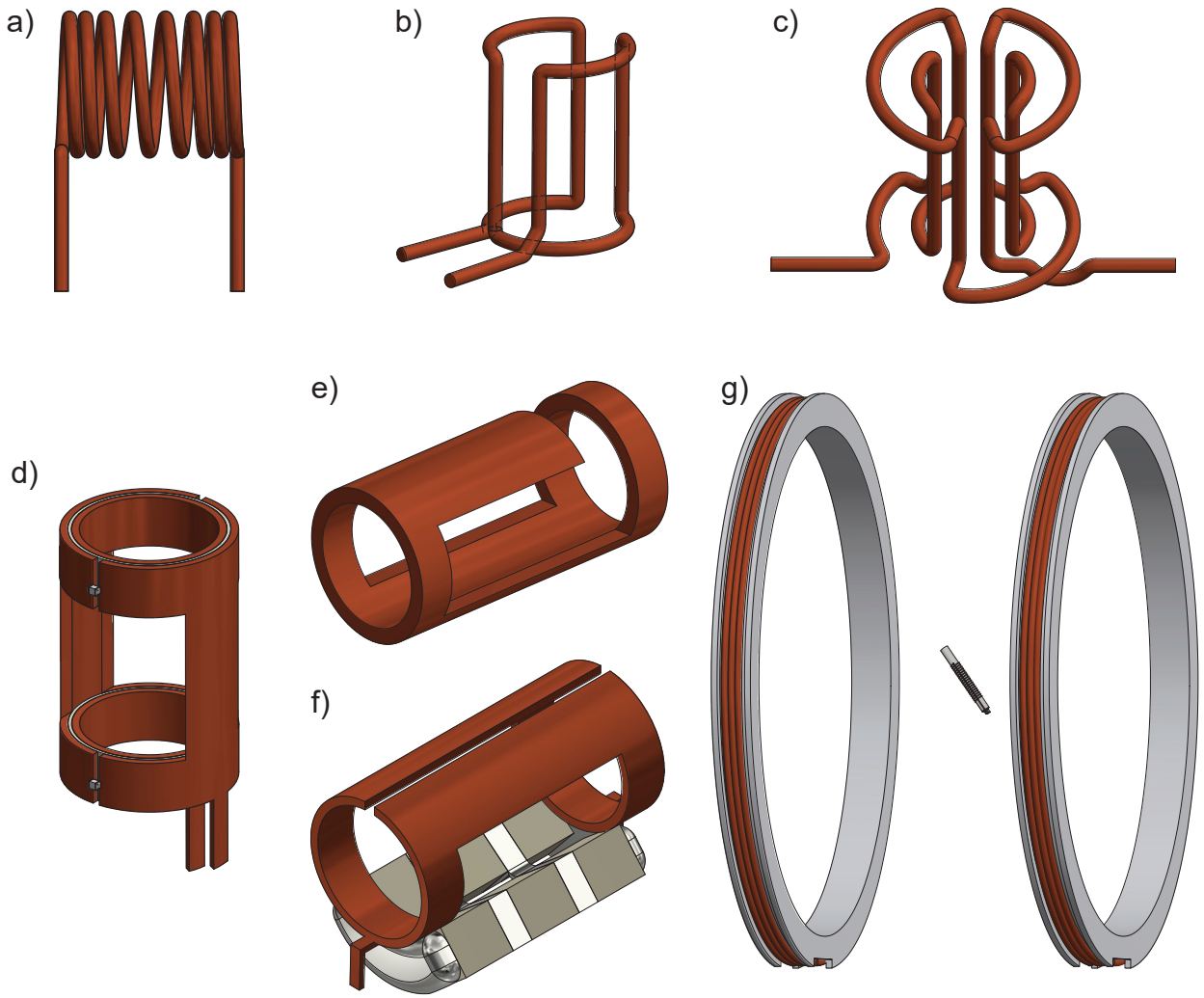


Figure 3.1: A selection of rf coils that are of use with SAS (a) variable pitch solenoid [104], (b) saddle coil, (c) double-saddle coil [54], (d) Alderman-Grant resonator [64], (e) Modified Alderman-Grant resonator [105], (f) Modified Alderman-Grant resonator [65], (g) microcoil inductively coupled to a Helmholtz array [106].

Transverse coils are ideally suited to SAS since the B_1 field produced by them is orthogonal to the coil axis. This means that the B_1 field can be made coaxial with the switching axis, resulting in a constant B_1 field regardless of angle. The caveat for transverse coils is that they have lower inductance than solenoids and therefore cannot produce such strong B_1 fields. This can be helpful in some instances (reducing sample heating) and a hindrance in others (producing reasonable B_1 fields for lower-γ nuclei); furthermore, some transverse coils are designed to have moderate inductance making them capable of producing reasonable

B_1 fields for lower-γ nuclei. The Litvak probe uses a double-saddle coil [54], which is a transverse coil designed to have a higher inductance than that of a traditional saddle coil or an Alderman-Grant type resonator. Alderman-Grant type resonators are designed to be low-E coils, meaning that the E-field produced when running a current through it is much lower [64]. Low-E coils greatly reduce the sample heating experienced by lossy dielectric samples (such as biological samples in ionic buffers) [64, 65], though this also makes them poorly suited for low-γ nuclei.

Given the past designs and the considerations for adding a third channel to the capacitive coupling (discussed in the next section), two coil configurations have been chosen for further testing. For a single coil option, the double-saddle coil was chosen due to its higher inductance compared to other transverse coils, this arrangement will put the ^{1}H circuit on one side of the coil while the ^{13}C and ^{15}N circuits are on the other side (see Fig. 3.2). This design will have reasonable homogeneity across all 3 channels, a better filling factor for ^{1}H, and allow for higher power pulses; however, the tuning network will be more complex and require more isolation elements, as all 3 channels are on the same coil. As a crossed-coil configuration, a solenoid that is coaxial with a Zhang MAG resonator [105], in this case the MAG, will be tuned to the ^{1}H frequency, while the solenoid will have the ^{13}C and ^{15}N circuits (see Fig. 3.3). This design will have less sample heating from ^{1}H decoupling because the intended sample system is membrane proteins within bicelles, a lossy biological sample. It will also be easier to tune, since the circuits are more isolated; however, alignment of the two coils can be tricky, and by its nature would be capable of lower power levels than the single coil system. Also, because of the changing x-component of the B_1 field of the solenoid at each angle, the power levels for those channels will need to be calibrated for each angle so that the pulse sequence can use the proper power level to give an equivalent pulse at each angle.

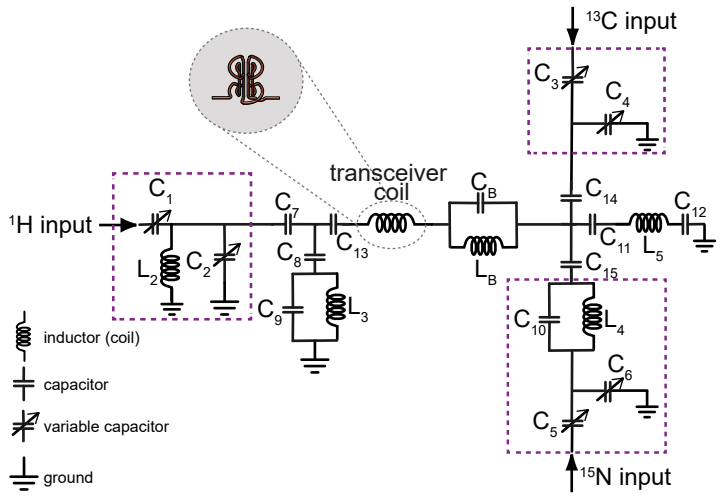


Figure 3.2: Circuit diagram for a triple-resonance SAS probe using a double-saddle coil.

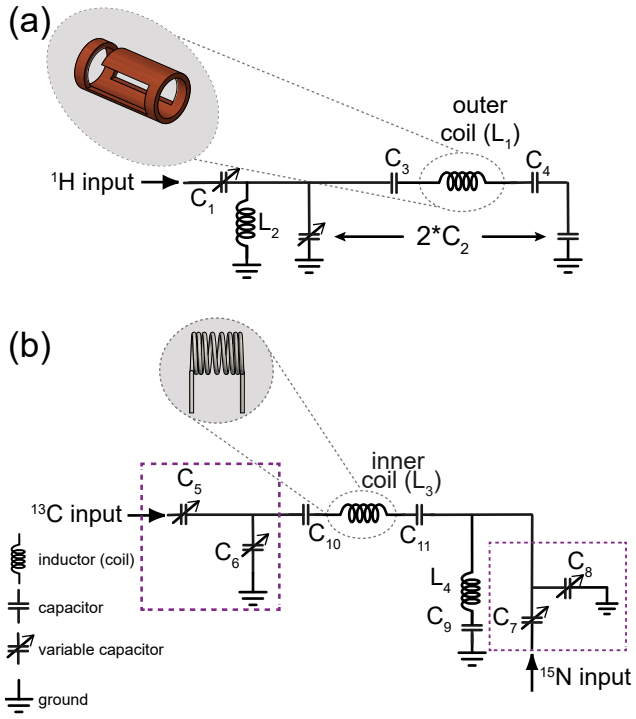


Figure 3.3: Circuit diagrams for a triple-resonance SAS probe using crossed-coils. (a) ^{1}H circuit diagram for the Zhang MAG. (b) ^{13}C/^{15}N circuit using a variable-pitch solenoid.

3.1.2 Capacitive Coupling

The two potential capacitive coupling arrangements corresponding to each of the chosen coil designs will slightly different from each other, and each has potential advantages and

disadvantages. These arrangements are illustrated in Figure 3.4. Since each channel needs to have its own capacitance connecting it to the coil, both designs will have at least one side of the capacitive coupling stacked.

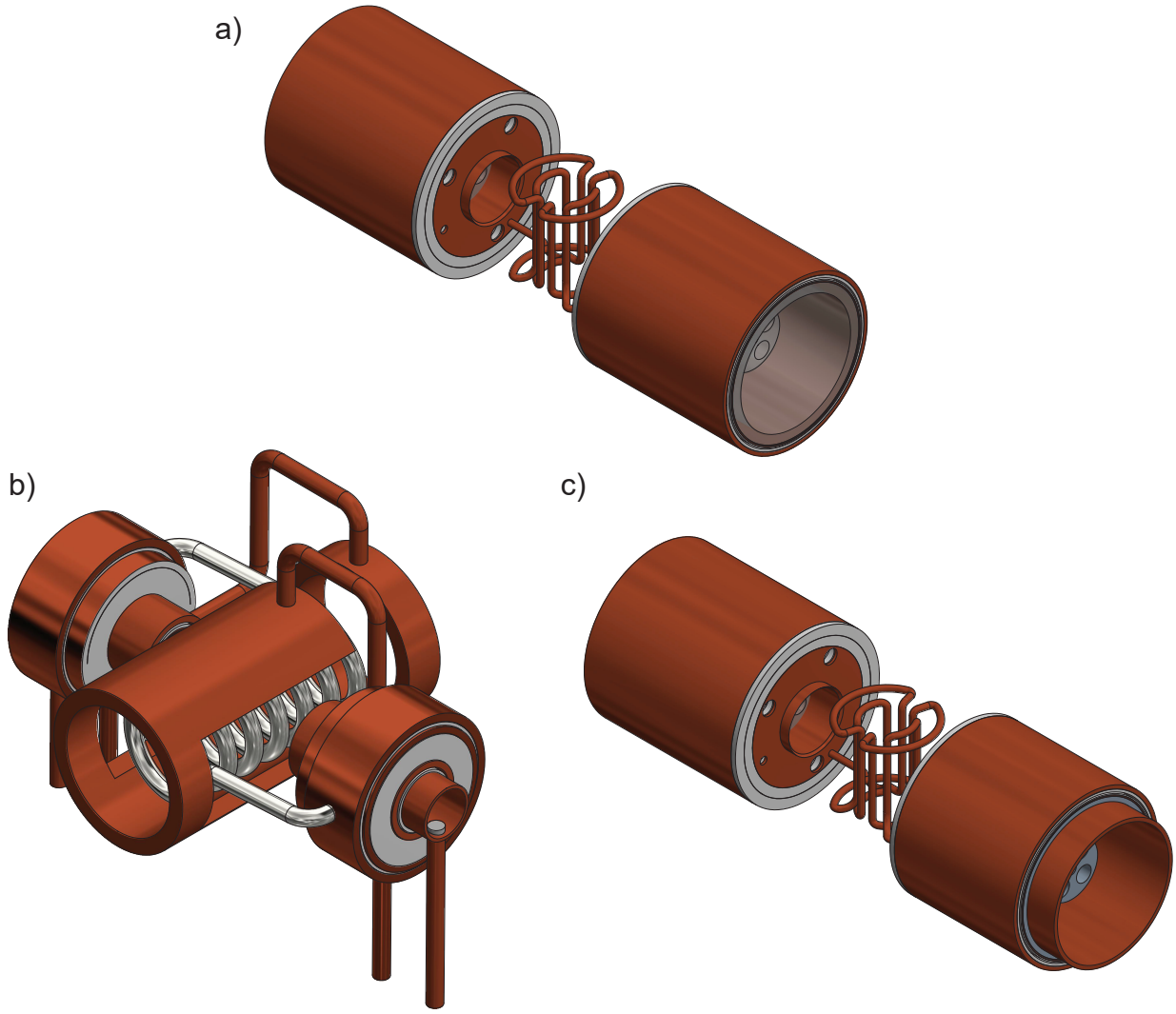


Figure 3.4: (a) The capacitive coupling used in the Litvak SAS probe [54], (b) cross-coil design for capacitive coupling for the new 3 channel SAS probe, (c) single-coil stacked capactive coupling design for the new 3 channel SAS probe.

As shown in Fig. 3.4c, with the single coil arrangment, using a double-saddle coil, only one side of the circuit has stacked capacitors. This could have been arranged a few ways, including two separate pairs of cylinders (as in Fig. 3.4b) or having the outer cylinder split in half with each half being one channel's capacitance. Here, the design that both is, in my

opinion, functional and easiest to machine was chosen. In this design there are 3 concentric cylinders on one side of the coil, separated by PTFE dielectrics. The ^{15}N capacitance will be from the outer cylinder (which is static and connected to the channel) to the middle cylinder (connected to the coil, and therefore mobile). The ^{13}C capacitance will be from the inner cylinder (again static and connected to the channel) to the middle cylinder. The other side of the coil has the same design as in the Litvak probe [54]. This design will have fewer pieces to machine, while still being electrically equivalent to the other options, with a distinct capacitance for each channel.

Fig. 3.4b shows the capacitive coupling arrangement for a crossed-coil design. This design has stacked capacitors on both sides of the coil. The stacked capacitors in this design consist of 4 concentric cylinders. The inner two cylinders on either side serve as the capacitors for the ^{1}H channel and its balance capacitor, and the outer two cylinders on each side are the capacitors for the ^{13}C and ^{15}N channels. The outer of the two cylinders in each pair is the static one connected to the channels, the inner cylinders of the pairs are attache to the coils and mobile. My main concern about this arrangement (though not directly a result of this capacitive coupling design) is that the alignment of the crossed-coils will suffer from switching, potentially interfering with the sample spinning during the experiment.

Due to these factors, the arrangement chosen to use for the probe is the double-saddle coil design (Fig. 3.4c). CAD images of all the pieces were made (available in Appendix A: CAD Drawings), and the pieces were machined.

3.2 Construction

The CAD images for the parts of this probe are shown in Appendix A: CAD Drawings. The dimensions shown in the CAD images are in inches. All of the pieces, except for the stator, were machined in the UCI Student Machine Shop (a small number of the more intricate pieces were made by the Professional Shop staff).

3.2.1 Probe Frame

The probe is designed to fit inside the 500MHz wide-bore Oxford magnet. The diameter of the bore with the shims inserted is 2.89 inches, therefore the largest diameter of the probe frame (the outer diameter of the can) will be 2.85 inches to ensure everything fits smoothly inside. For the purposes of this discussion, the support frame will refer to the rods and plates that hold up the probe frame as well as screw to the bottom of the magnet (Fig. 3.5a) and the probe frame will refer to the connected pieces that hold the probe together and serve as the outer ground plane of the circuits (Fig. 3.5b).

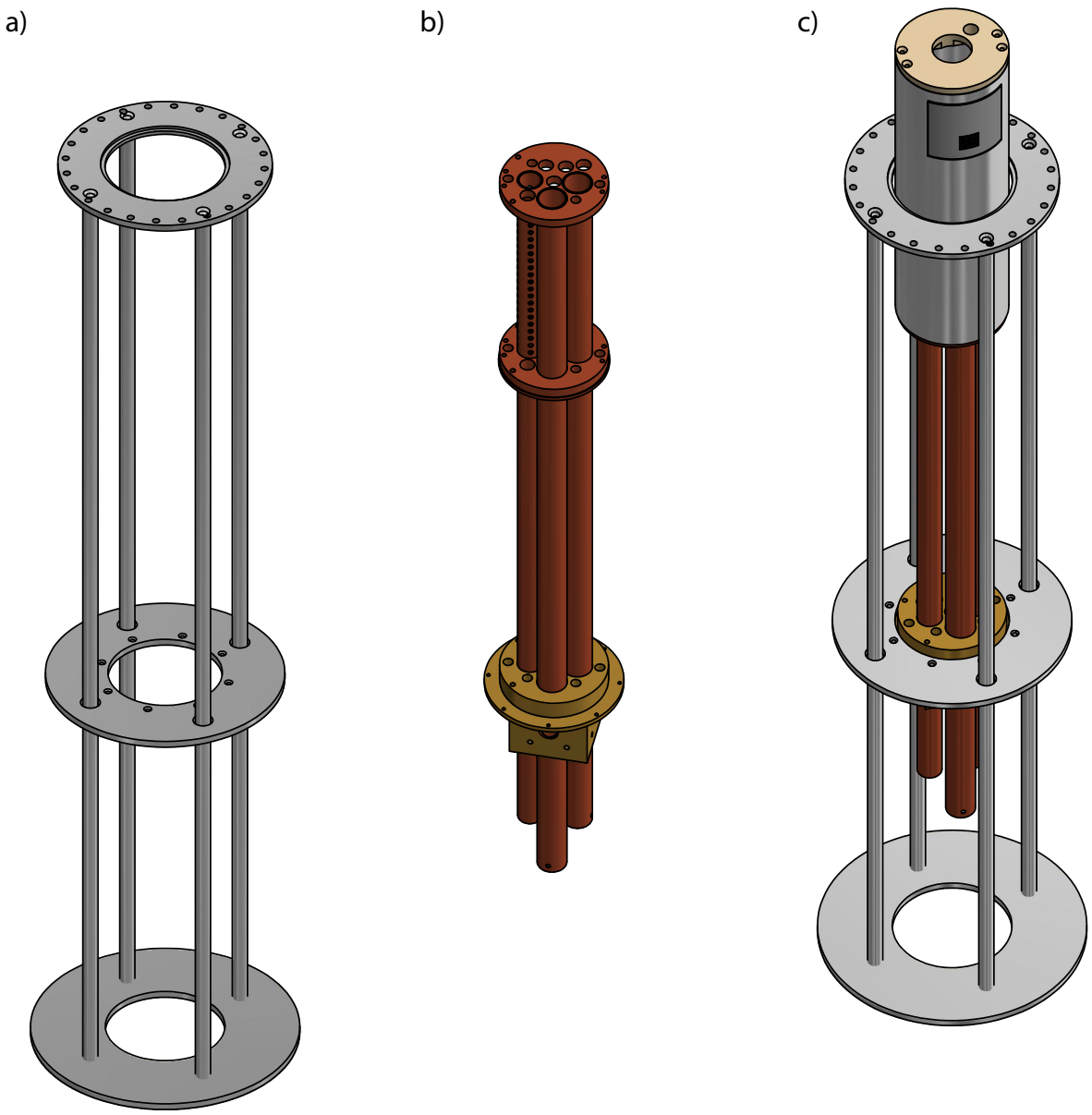


Figure 3.5: a) The support frame that holds up the probe exoskeleton and is attached to the magnet itself. b) The probe frame that serves as the ground plane and contains the tuning networks (is inserted into the magnet). c) How the probe frame and support frame attach together.

The support frame consists of 4 rods that connect 3 plates that hold the probe frame. The top plate has 4 holes that are counter bored to insert the rods and countersunk so that the screws that attach to the rods are flush with the plate. The top plate also has a series of holes around the outer edge that are for the screws that attach the frame to the magnet, the

large number of these holes allows for a large number of orientations of the probe within the magnet so that the probe can be better aligned within the shims. The middle plate has 4 clearance holes that the rods go through, and 8 holes around the center that attach to the brass ring of the probe frame and serve as the primary attachment of the probe frame to the support frame. The bottom plate only has the 4 holes to attach the rods, which are counter bored and countersunk the same as the top plate. All of these pieces are made of aluminum.

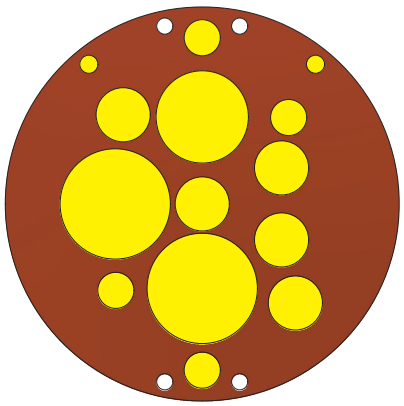


Figure 3.6: The yellow shaded holes in this ring are the primary holes. These are the ones that are common to all three rings as well as the brass block.

The probe frame consists of 2 OFHC copper rings (the top and middle), 1 brass ring, 1 brass support block, and the 3 OFHC copper channels. The 3 rings all have the same set of holes to contain the channels, airlines, dewar, isolation elements, and SAS components (hereon referred to as the primary holes, shown in Figure 3.6). The top ring just has the primary holes. The middle ring has the primary holes as well as a small lip to support the bottom of the can. The brass ring, besides the primary holes, has a larger lip with 8 threaded holes to attach the to the middle plate. In addition to the primary holes (the same as the rings), the brass block also has holes on the sides that line up with the holes in the channels and are where the connections from the spectrometer will insert into the probe, there are also threaded holes on each side to attach N-type connectors to the block. The heteronuclear

channels are of different lengths based on their frequency, but are the same diameters and design otherwise. They have one hole near the bottom for the spectrometer connections, and one small hole at the bottom to secure the knob assembly to the channel. The ^{1}H channel is a smaller diameter due to its higher frequency, and in addition to the connection and securing holes in the heteronuclear channels, also has a series of holes near the top that are used to secure the inductive stub to the ground plane.

In addition to the probe frame and support frame pieces, the final frame assembly also has an aluminum can that serves to further isolate the tune networks, as well as the probehead, from stray frequencies. There is also a cap to to secure the can to the rest of the probe (these are shown in the whole assembly of the frames in Fig. 3.5c). All the pieces of the probe and support frames are available as CAD images in Appendix A: CAD Drawings as Figures A.1-A.10.

3.2.2 Tuning Network

The design used for the tuning networks has been adapted from that of Martin, et al. [80]. This tuning tube design utilized transmission line segments as the discrete elements of the circuit. This enables each tuning channel to be shielded from the others within its outer ground plane, which also saves space within the probe and enables the use of higher power levels during experiments.

The ^{13}C and ^{15}N channels are extremely similar to the original design [80]. The transmission line segments are concentric cylinders that have been machine from OFHC copper (McMaster-Carr), and the dielectrics are machined from Teflon PTFE (DuPont). The detailed CAD images of these pieces are available in Appendix A: CAD Drawings as Figures A.22 & A.23.

Both the heteronuclear channels use the same design, differing only in the dimensions required to tune to the proper frequencies. As can be seen in the exploded view in Fig. 3.7, all of the pieces are concentric and the match and tune capacitors share a plane. A pulse from

the spectrometer enters at the bottom of the rf feed and travels up to the inner conductor of the match capacitor. The match capacitance is between the match conductor and the inner diameter of the tune conductor (and is adjusted by a knob at the bottom of the channel that moves the match conductor in and out of the tune conductor), the pulse would then travel from the tune conductor to the static cylinder of the capacitive coupling and then to the mobile coil. The tune conductance is that from the tune conductor to the outer ground plane of the channel and is adjusted by turning a knob at the bottom of the channel that moves the tune dielectric up and down between the tune conductor and the outer channel.

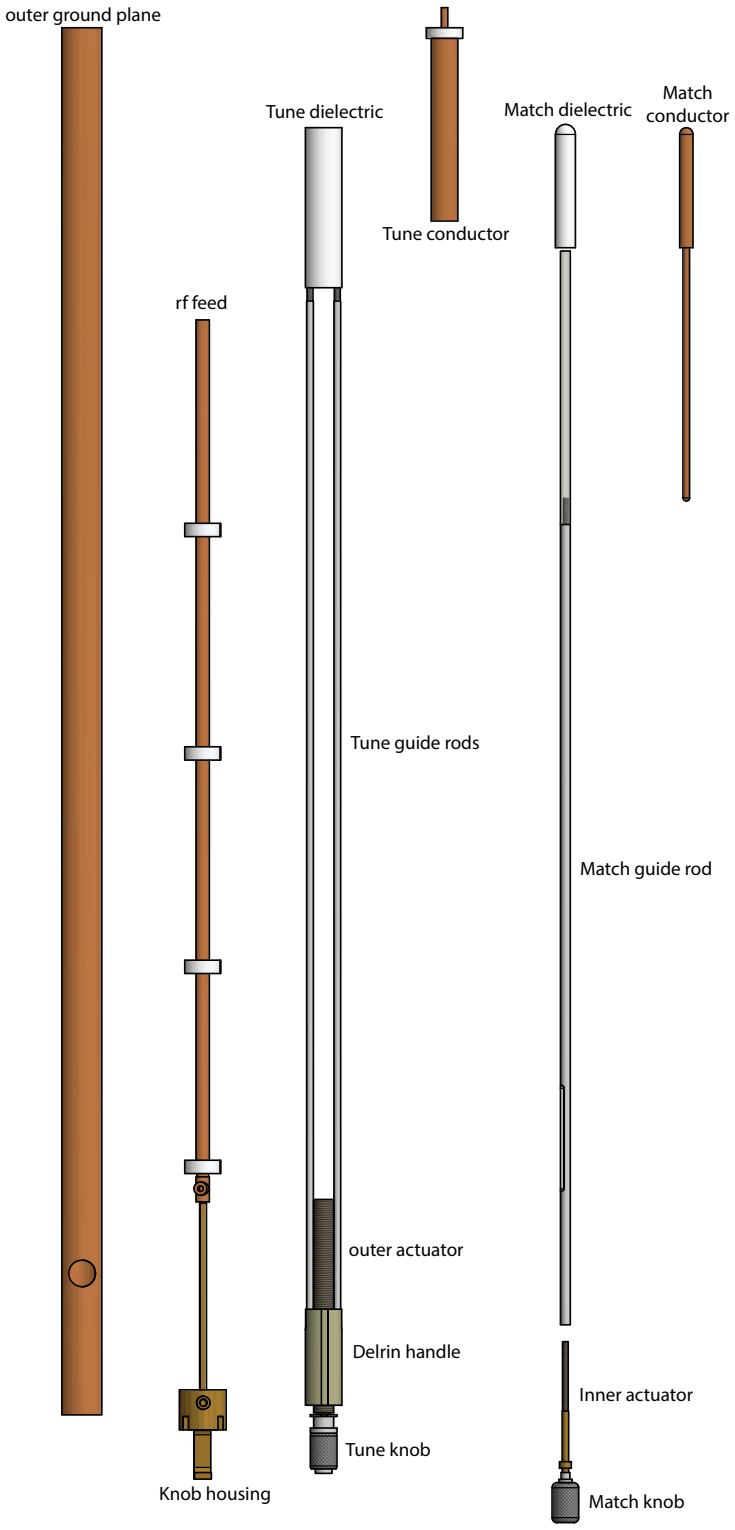


Figure 3.7: Exploded view of the heteronuclear channels. The pieces are all concentric, with the pieces to the right fitting inside the pieces to the left of them.

The knob assemblies at the bottom of the channels are the same for each of the channels (including ^{1}H). The bottom knob adjusts the match by turning an inner actuator (concentric inside the outer actuator) that moves the match guide rod up and down inside of the rf feed. The top knob adjusts the tune by turning the outer actuator which moves a Delrin acetal resin (DuPont) handle up and down. This handle is in turn attached to two guide rods that that attach to, and move, the tune dielectric at the top of the channel. An exploded view of the knob assembly is shown in Fig. 3.8.

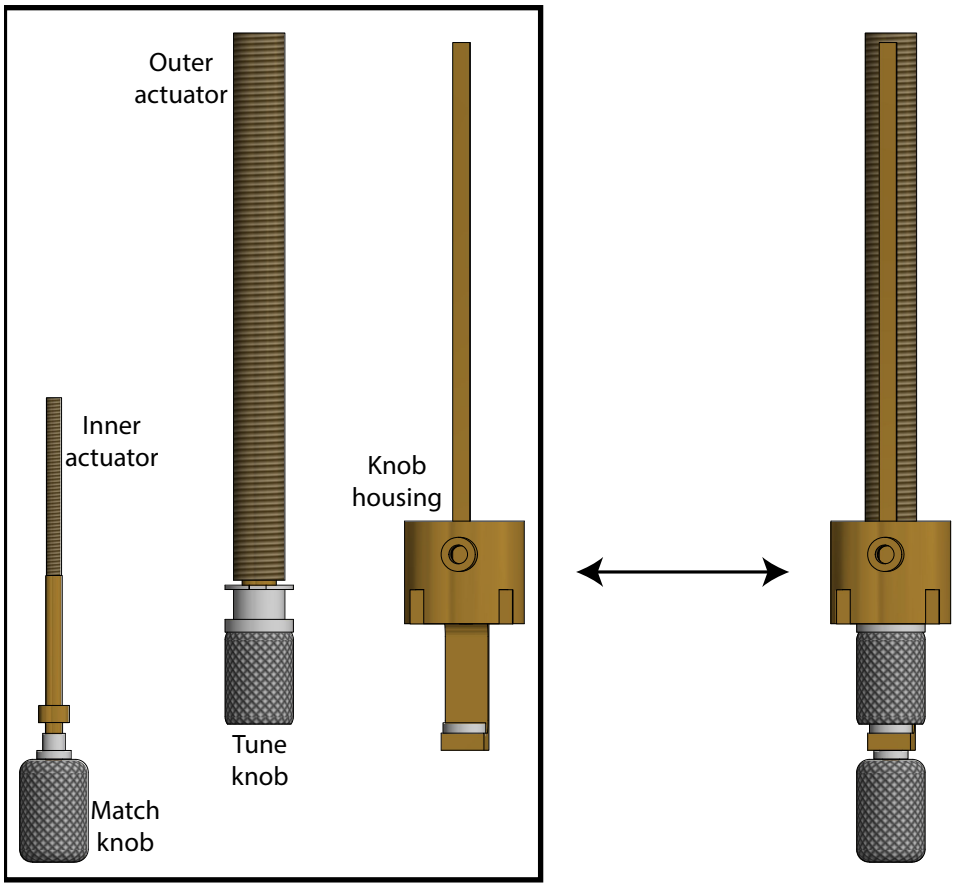


Figure 3.8: Exploded view of the knob assembly that controls tune and match.

The ^{1}H channel differs from the heteronuclear channels because there is an extra inductive element added to the circuit to aid in bringing the resonant frequency higher while still having the capacitive pieces be easily machinable dimensions. The design of this inductive 'stub'

has been modified slightly from the original design by Martin, et al. [80]. The detailed CAD images of these pieces are available in Appendix A: CAD Drawings as Figures A.24-A.27. The electrical path of a pulse from the spectrometer through the ^{1}H channel is very similar to that of the heteronuclear channels in the beginning. A pulse would enter at the bottom of the rf feed and travel up towards the match conductor, however, in this case the match conductor is a thing copper sleeve that is concentric outside the tune conductor. The match capacitance is that from this sleeve to the outer diameter of the tune conductor. From here the circuit experiences an inductance to ground (through the inductive stub screwed to the outer ground plane), the tune capacitance (from the tune conductor to the outer channel), and to the coil through the capacitve coupling as in the heteronuclear channels. The adjustment of the match and tune capacitances functions as described previously. A zoomed in view showing the difference in the upper portions of the heteronuclear vs. ^{1}H channels is shown in Fig. 3.9.

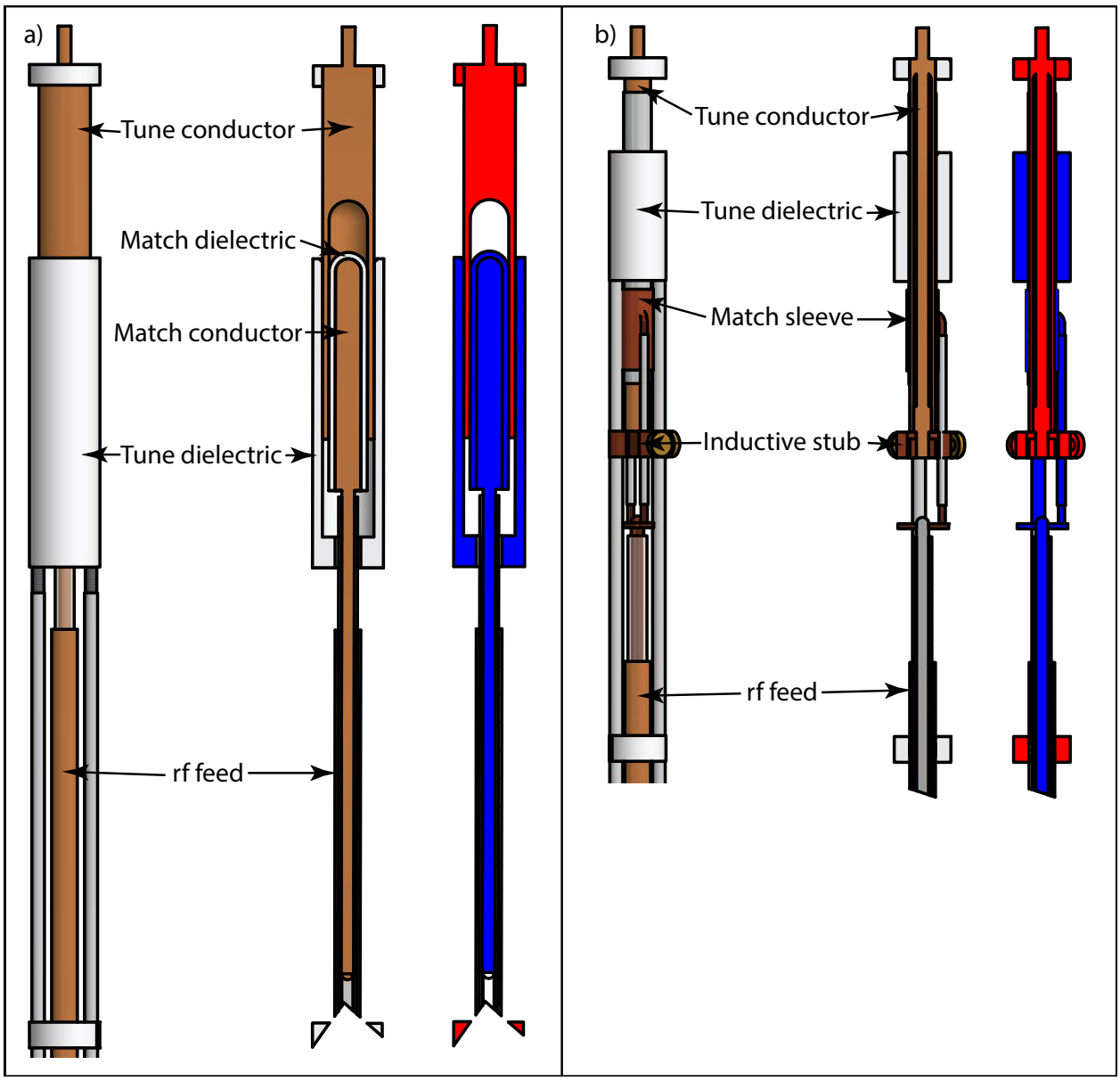


Figure 3.9: Comparison of the tune and match elements of the heteronuclear channels (a) and the ^{1}H channel (b). The channels are very similar in design, with the major differences being the presence of the inductive stub in the ^{1}H channel as well as the match being an outer sleeve instead of an inner conductor. In both panels, the sections that have been shaded red are static within the channel, and those shaded blue are the components that move to adjust the tune and match capacitances.

3.2.3 Capacitive Coupling

The capacitive coupling is one of the key points of the chosen SAS design. Given the large number of angle switches that the probe will experience over its lifetime, there is a need for a

robust connection between the mobile and stationary parts of the probe that also reduces the friction caused by the switching. This is done by capacitively coupling the tuning networks to the signal coil of the probe. One plane of the cylindrical capacitor connects to the coil and will move with the stator during the angle switches. The other plane is connected to the channels of the tuning networks. To stabilize this connection, a small section of the outer planes is made thicker to accommodate the attachment of a $\frac{1}{8}$ in. copper rod as the connection to the channels. These raised sections can be seen in Fig. 3.10, which illustrates the chosen capacitive coupling design for this probe.

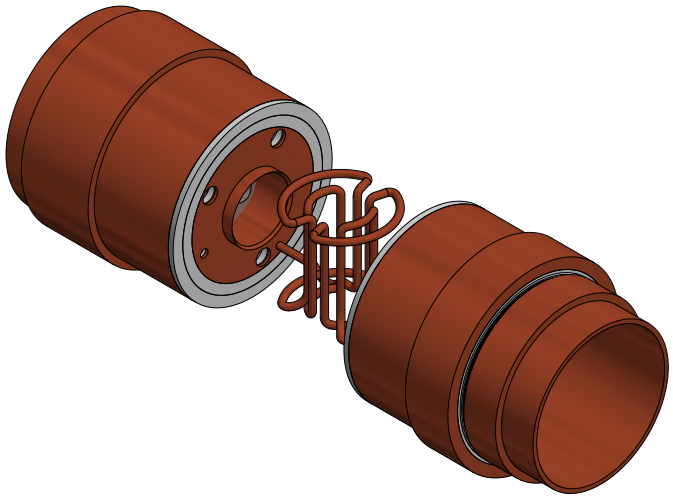


Figure 3.10: The selected cacacitive coupling design for this probe. A single pair of cylinders makes up the ^{1}H capactive coupling and the heteronuclear side has 3 concentric cylinders making up the 2 coupling capacitances.

3.2.3.1 ^{1}H

Due to the single coil design of this probe, the two sides of the coil have different numbers of channels on them. Therefore, the capacitive coupling design for the ^{1}H side of the coil is different from the heteronuclear side. The ^{1}H capacitance is adapted straight from Litvak's work [54], and a cross-section view is shown in Fig. 3.11.

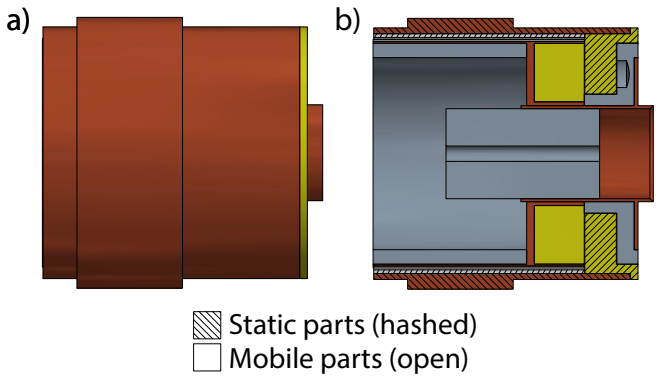


Figure 3.11: The ^{1}H cacacitive coupling a) and a cross-section view b) illustrating how it fits together. The hash marked areas denote the stationary portions of the capacitive coupling.

3.2.3.2 ^{13}C and ^{15}N

The design of the heteronuclear capacitve coupling is a little more complex, since there needs to be a separate capacitance for each channel on this side. This is done by adding a third concentric cylinder to the design used for the ^{1}H capactive coupling. The resulting design has the central cylinder of the 3 concentric cylinders as the one connected to the coil. The capacitance from the inner cylinder to the center cylinder has calculated to match that of the outer cylinder to the central cylinder. The lengths of the cylinders have been shortened to allow for the total length of the stacked capacitors to match that of the ^{1}H side. A cross-section view of this new design is shown in Fig. 3.12.

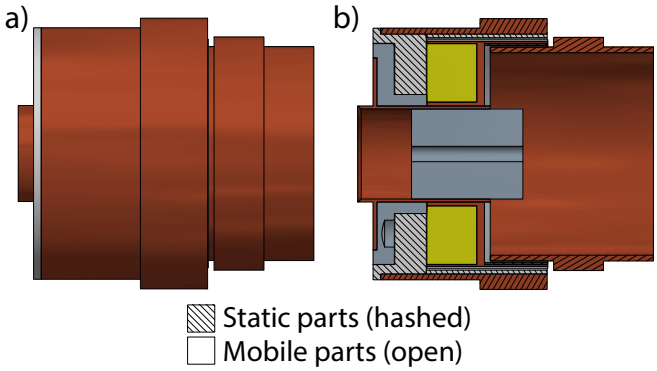


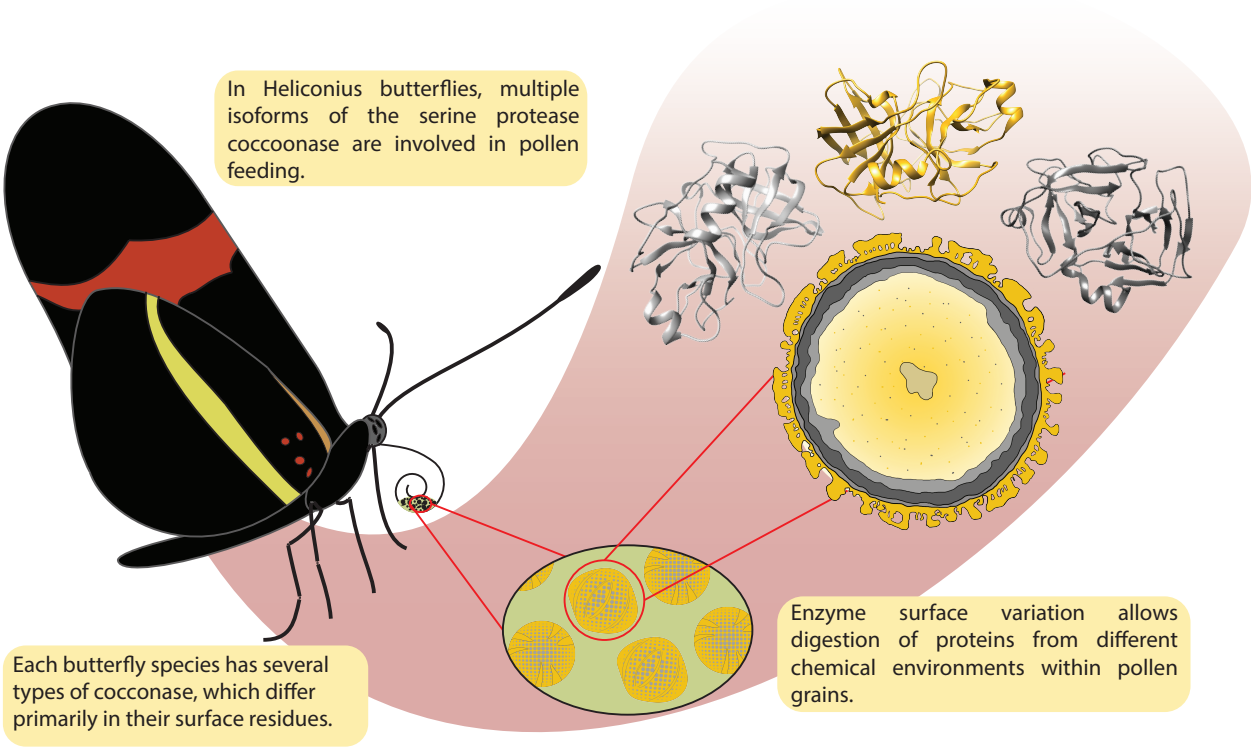
Figure 3.12: The ^{13}C and ^{15}N cacacitive couplings a) and a cross-section view b) illustrating how it fits together. The hash marked areas denote the stationary portions of the capacitive coupling.

3.3 Outlook

At this point, the design of the probe is complete. Also, a large portion of the probe frame and supports have been machined and assembled. What remains is the machining of the capacitive coupling components, and the final assembly of the probe. From there, the testing and tuning of the probe can commence.

Chapter 4

Enzyme Discovery



The chapter cover figure shows some selected figures illustrating concepts discussed in this chapter; a *Heliconius* butterfly with representations of the pollen it feeds on and the predicted protein structures.

4.1 Motivation

In the age of inexpensive genome sequencing, it is impractical to experimentally solve the structure of each newly discovered protein due to requisite monetary and time resources. In fact, only about 1% of protein sequences listed in the UniProtKB database have structures in the Protein Data Bank (PDB), which means there are as many as five million known protein sequences with no currently solved structures [107]. It is for these reasons that the Martin lab has begun using computational methods to predict protein structures and bioinformatics tools to predict their properties as described for the aspartic proteases investigated in Chapter 5. Using the target selection pipeline shown in Figure 4.1, proteins can be sorted and filtered to search for targets with desired characteristics or other interesting features. This approach allows for the more efficient selection of protein targets for structure determination and biochemical characterization. Interesting targets to study are selected by analysis of Protein Structure Networks (PSN), which is performed in collaboration with Prof. Carter T. Butts research group. The analysis uses Benson and Daggett's PSN representation [108], where vertices represent chemical moieties and edges represent the potential for direct interaction (as determined by moiety-specific proximity constraints). PSNs reveal the presence or absence of unique structural features within the group of enzymes being studied. Through an encoding of the potential for direct interaction between functional groups, PSNs are useful tools for quickly and inexpensively analyzing the structural properties of proteins. Additionally, PSNs can help predict properties related to structural differences such as thermal stability, aggregation propensity and overall activity of the enzymes. PSN analysis helps in selecting interesting proteins for expression and characterization by maximizing the probability of selecting proteins that display useful functionality [109, 110]. Before PSN analysis can be performed, the selected protein sequences need to be aligned and annotated to extract

the information needed by the maturation process. Then PSN analysis can be run on the mature structures. In addition to the PSN analysis, we also perform principal component analysis , using the fraction of polar residues and charged residues, mean residue charge, and mean hydrophobicity as the components. This provides us with further information about the surface residues of proteins and how they interact.

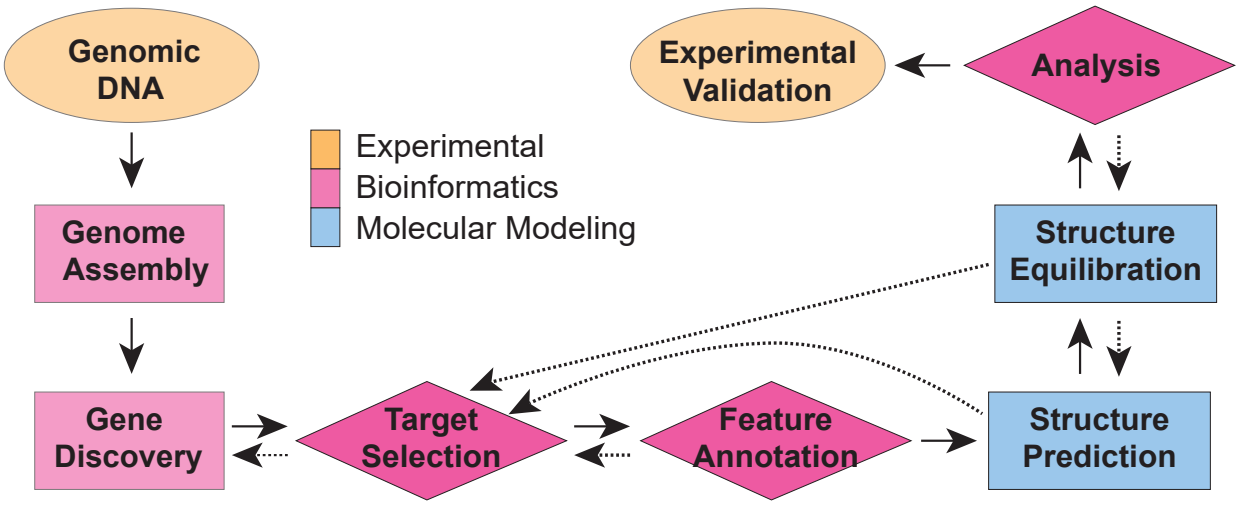


Figure 4.1: The target selection pipeline illustrates the process of discovering new enzymes to their experimental validation. The color key represents whether the step is experimental (orange), bioinformatics (pink) or molecular modeling (blue). The direction of arrows shows the flow of information to the next step. Dotted arrows indicates going back to an earlier step to modify conditions or fix errors. Using standard flow chart symbols, the square steps shows processing step while the diamond steps show decision making. Figure credit: Megha Unhelka [111].

4.2 Methods

A brief technical description of our methods follows. More detail is available in both the "how-to" section of this chapter (Section 4.5) and in the methods section corresponding to the investigation of Aspartic Proteases from *D. capensis* in Chapter 5.

Sequences were aligned using ClustalOmega [112] (gap open penalty = 10.0, gap extension penalty = 0.05, hydrophilic residues = GPSNDQERK, weight matrix = BLOSUM). The

presence and cleavage sites for N-terminal secretion signal sequences were predicted using SignalP 4.1 [113]. Structure prediction was performed in three stages as described in Butts et al. [114, 109]. Simulation was performed using the CHARMM36 forcefield [115], 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 reference sequences for which an experimentally determined structure was available, this was used as the initial starting model (following removal of heteroatoms and protonation using REDUCE [116]). Relative solvent accessibility (RSA) values are calculated for all equilibrated structures using DSSP 2.2.1 [117]; residues with RSA values less than 0.2 were regarded as buried, with other residues classified as solvent exposed. For the set of solvent exposed residues within each structure, the fraction of polar residues and charged residues, mean residue charge, and mean hydrophobicity (using the scale of Kyte and Doolittle [118]) were calculated. All data analysis and visualizations were performed using R [119]. Projections of the original variables into the principal component analysis (PCA) space were also calculated to assist with interpretation.

One of the benefits of this high-throughput method for investigating and comparing proteins without solved structures is that it can be applied to many different classes of proteins. This has allowed us to have several people, including high school students and undergraduates with minimal training, working on this project, each working on a different class of protein. However, our initial work showed a need to standardize the method, in order to keep the processes running smoothly and efficiently, while minimizing errors. Since we have a need for this process to be easily understood by inexperienced users, I have detailed the standardized alignment and annotation processes in a how-to format as a manual for new users, included at the end of the chapter, Section 4.5. It is important for methods like this to be easily accessible to new users, because it is something that an inexperience student can do with very minimal training. This allows them to get research experience that they feel is making progress, while allowing them time to be trained in the rest of the lab skills they will need to

do other, more wet lab, research. This approach has been our idea for implementing more citizen science, and in addition to having undergraduate students working on this project, we have also had several high school students assisting. This has enabled some of them to even have a published paper before they graduate high school.

4.3 Cocoonases

4.3.1 Background

The *cocoonase* gene is a single-copy gene in several butterfly and moth genomes (the silkmoth *Bombyx mori*, diamond backed moth *Plutella xylostella*, and monarch butterfly *Danaus plexippus*, and the Glanville fritillary *Melitaea cinxia*; [120]). However, heliconiine butterflies harbor at least five duplicate copies, indicating recent gene duplication events unique to *Heliconius* butterflies. Smith et al. [120] discovered that multiple paralogs of *cocoonase* mRNA are upregulated in the proboscis of heliconiines when compared to two other tissues (antennae and legs). Further, they saw that these *cocoonases* are not expressed in the salivary glands of *Heliconius melpomene*, suggesting that, like moths, butterflies directly secrete this digestive enzyme from the proboscis. This is further supported by the presence of cocoonase in the saliva of *H. melpomene* adults [121]. Unlike moths, adult heliconiine butterflies express multiple, sequence-divergent versions of cocoonase in their mouthparts. This, and the fact that butterflies do not pupate within a silk cocoon but escape with relative ease from a chrysalis, suggests that cocoonase may have a different function in butterfly adults post-emergence. One potential function is the pre-digestion of pollen granules during feeding. *Heliconius* butterflies directly feed on pollen by collecting and digesting pollen on their proboscides, a behavior that is not seen in other butterflies [122, 123]. Nectar also contains small amounts of amino acids from dissolved pollen, and thus heliconiine butterflies may have coopted cocoonase for the digestion of peptides found in their natural diet, and *Heliconius* butterflies may use these specifically for feeding on pollen.

New cocoonase enzymes have potential commercial value across numerous applications. In the silk industry, degumming of silk by sericin removal is carried out in order to improve the quality of the fibroin silk fibers [124]. Current industrial methods for removing sericin involve chemical incubation (e.g., with alkaline solutions), which also degrades fibroin, reducing silk quality [125]. Because cocoonase hydrolyzes sericin but leaves the fibroin fibers untouched the enzyme is of great interest to the silk producing industry. The discovery of duplicate copies of *cocoonase* with divergent amino acid sequences might lead to the discovery of a protease that is more effective at degumming. Proteases are also useful for medical applications; recent work on the Chinese silkworm cocoonase has shown that it cleaves fibrin and fibrinogen both in vitro and in an animal model of thrombosis [126], demonstrating its utility for studying thrombosis and potentially treating blood clots. As we show below, the heliconiine cocoonases exhibit substantial variation in surface properties (particularly hydrophobicity), potentially facilitating their diffusion into diverse chemical environments. Such a mixture of enzymes with differing surface characteristics could prove useful in degrading complex proteinaceous material in the presence of chemical detergents, a common requirement of enzymatic cleaning agents.

Our collaborators in the Briscoe lab have investigated the duplications of these cocoonases from genome and transcriptome data [120, 127]. Our initial hypothesis involved their enzyme products acting on different substrates. We use comparative modeling and protein structure analysis to infer functional (and perhaps adaptive) differences when compared to the single-copy moth *cocoonase*. Our modeling data of 30 individual cocoonases indicate that, contrary to our hypothesis, all the cocoonase enzymes have trypsin-like specificity, while significant differences are found among the surface residues of different cocoonase types, suggesting enzyme adaptation to different chemical environments.

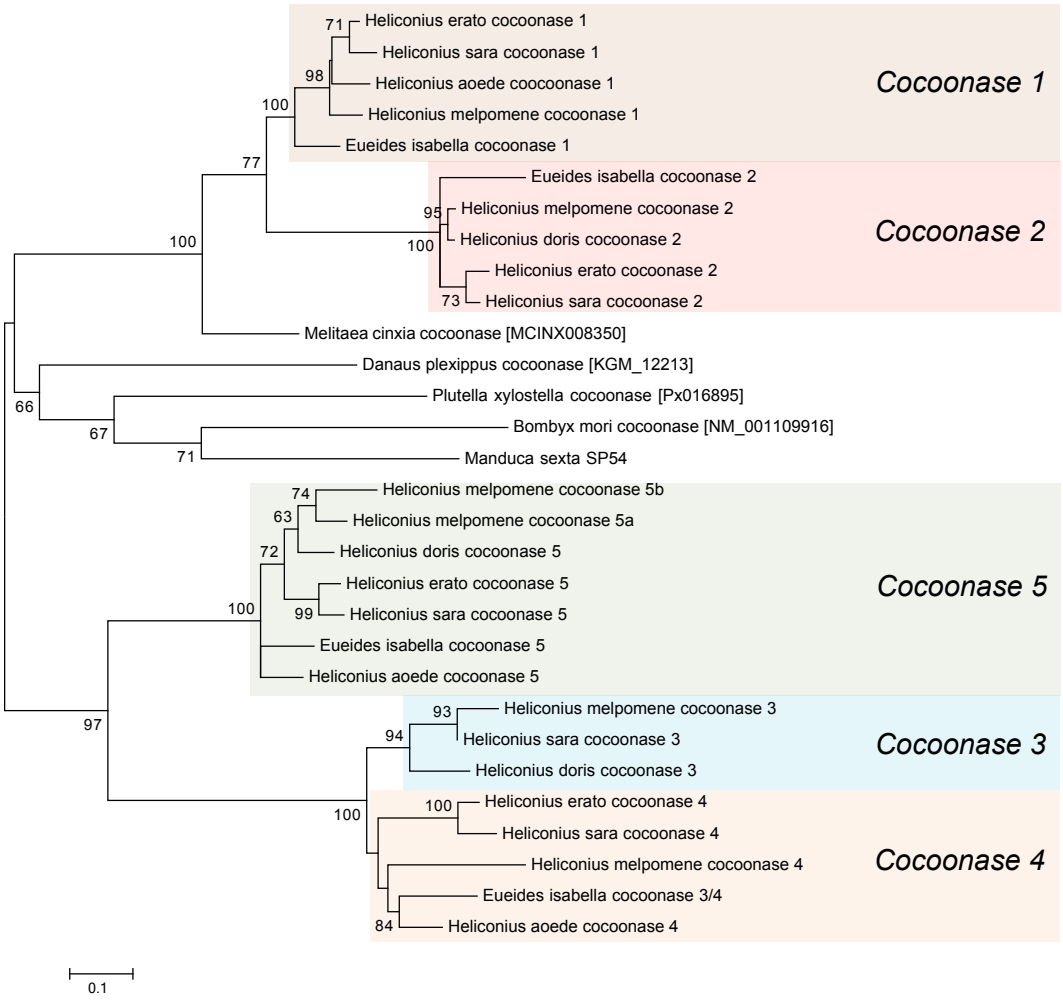


Figure 4.2: Maximum likelihood tree of cocoonase amino acid sequences. Numbers above branches are bootstrap support values (percent out of 1000 replicates) and scale bar is the branch length. Highlighted are the cocoonase clusters: cocoonase 1 genes are highlighted in brown, cocoonase 2 in red, cocoonase 3 blue, cocoonase 4 orange, and cocoonase 5 (a/b) green. Tip labels include the species name followed by, where applicable, the gene name or gene ID.

4.3.2 Molecular modeling predicts high structural homology to trypsin

The molecular models of cocoonase are characterized by a trypsin-like fold, as illustrated for a representative example (*D. plexippus* cocoonase; Figure 4.3a). Functional features examined include the length, sequence and positioning of the surface loops [128, 129], the

conformation, and dynamics [130] of the backbone around a glycine residue near the entrance to the specificity pocket (Gly 216 in trypsin) [131]. Figures 4.3b and 4.3c show the surface loops for a representative cocoonase (*H. melpomene* cocoonase 1) in comparison to trypsin and chymotrypsin. Although there are minor differences in backbone position, for both loops the conformation adopted is clearly that of trypsin. This is true for all of the full-length cocoonase models examined here.

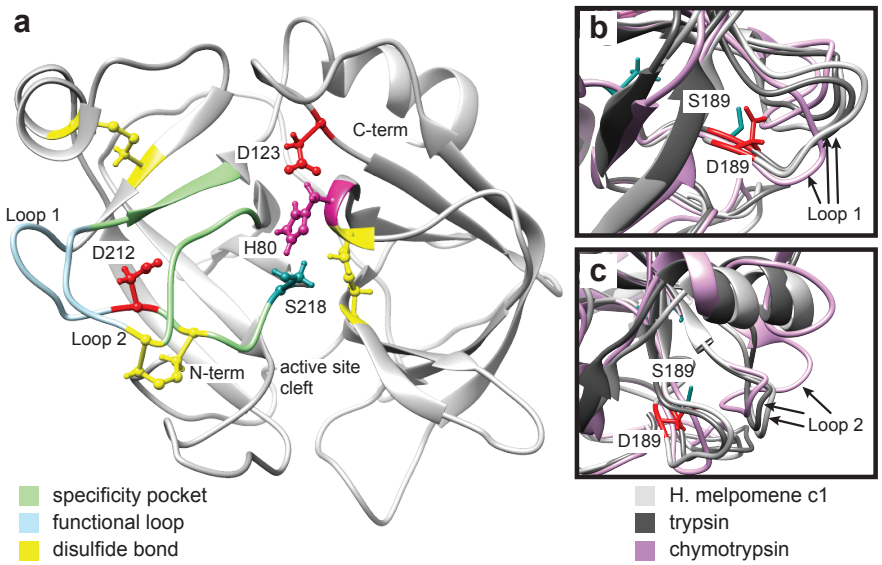


Figure 4.3: (a) Equilibrated molecular model of the active form of *D. plexippus* cocoonase, labeled to show the essential features of a trypsin-like serine protease. The catalytic triad residues H80, D123, and S218 (zymogen numbering) are shown as ball and stick models. The presence of a negatively charged residue (D212) in the bottom of the specificity pocket (green) indicates that this enzyme, like the other cocoonases, has a preference for a positively Arg or Lys in the position immediately prior to the cut site. The loops surrounding the pocket (blue) are also functionally important. The three conserved disulfide bonds are shown in yellow. (b) and (c) Comparison of *H. melpomene* cocoonase 1 (c1; light grey) to trypsin (dark gray) and chymotrypsin (lavender).

Strong conservation of sequence and structural properties in the cocoonase specificity pockets is not limited to the stabilizing Asp residue; examination of the sequence alignments (Figures 4.4 and 4.5, detail in Figure 4.6) reveals significant variation in only three sites in the specificity pocket itself, one of which is the Ser/Ala in position 188 (trypsinogen numbering). The side chains of the other residues involved (W215 and D217 in trypsinogen) point outward,

away from the substrate-binding pocket. Furthermore, the conformation about G216, a critical structural residue involved in regulating substrate binding [131], is similar to that of trypsin in the cocoonases. Based on the sequence and structural evidence, the evolution of these enzymes does not seem to be the result of pressure to produce proteases with different substrate specificity. Therefore, understanding the functional origin of cocoonase gene duplication in pollen-feeding butterflies must focus on other structural and sequence features.

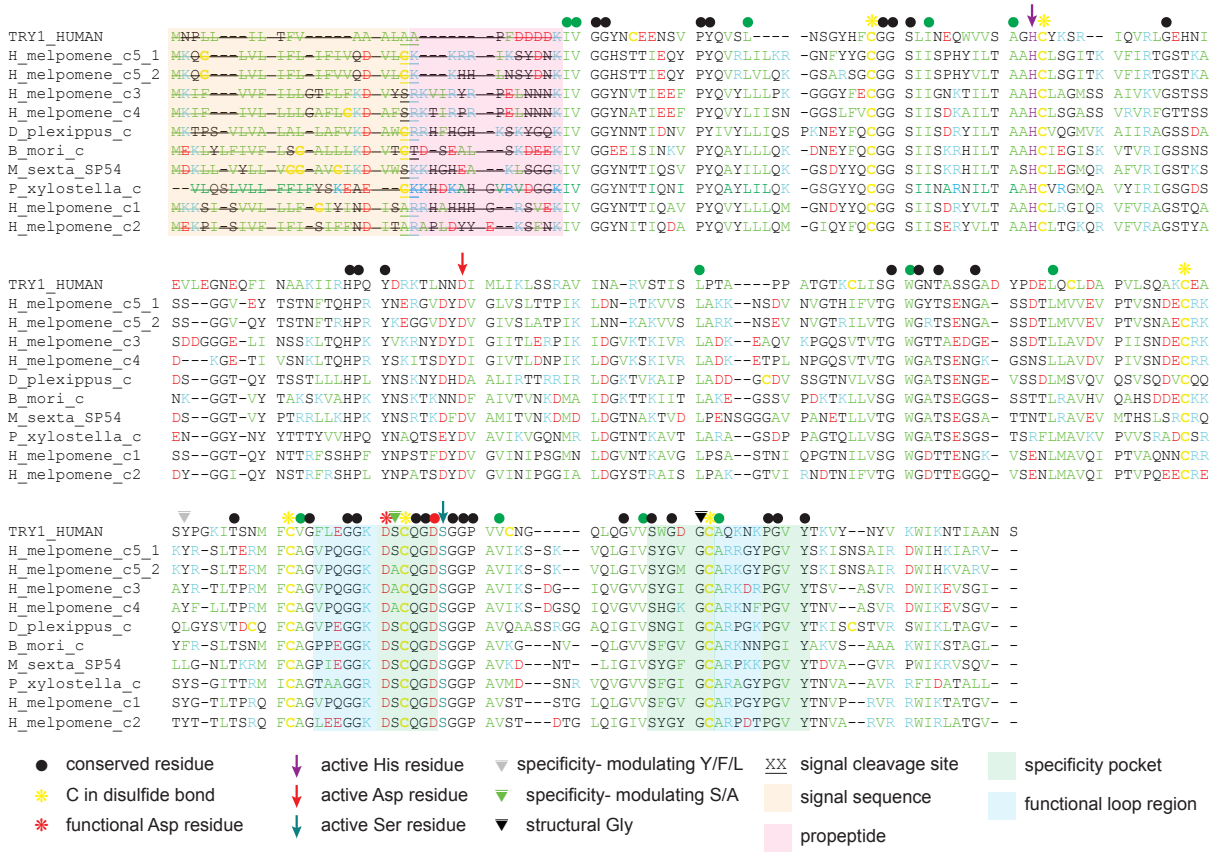


Figure 4.4: Cocoonases from *H. melpomene* compared with those from other lepidopteran species and with human trypsin. (H_melpomene_c5_1 is *H. melpomene cocoonase 5a*, and H_melpomene_c5_2 is *H. melpomene cocoonase 5b*.)

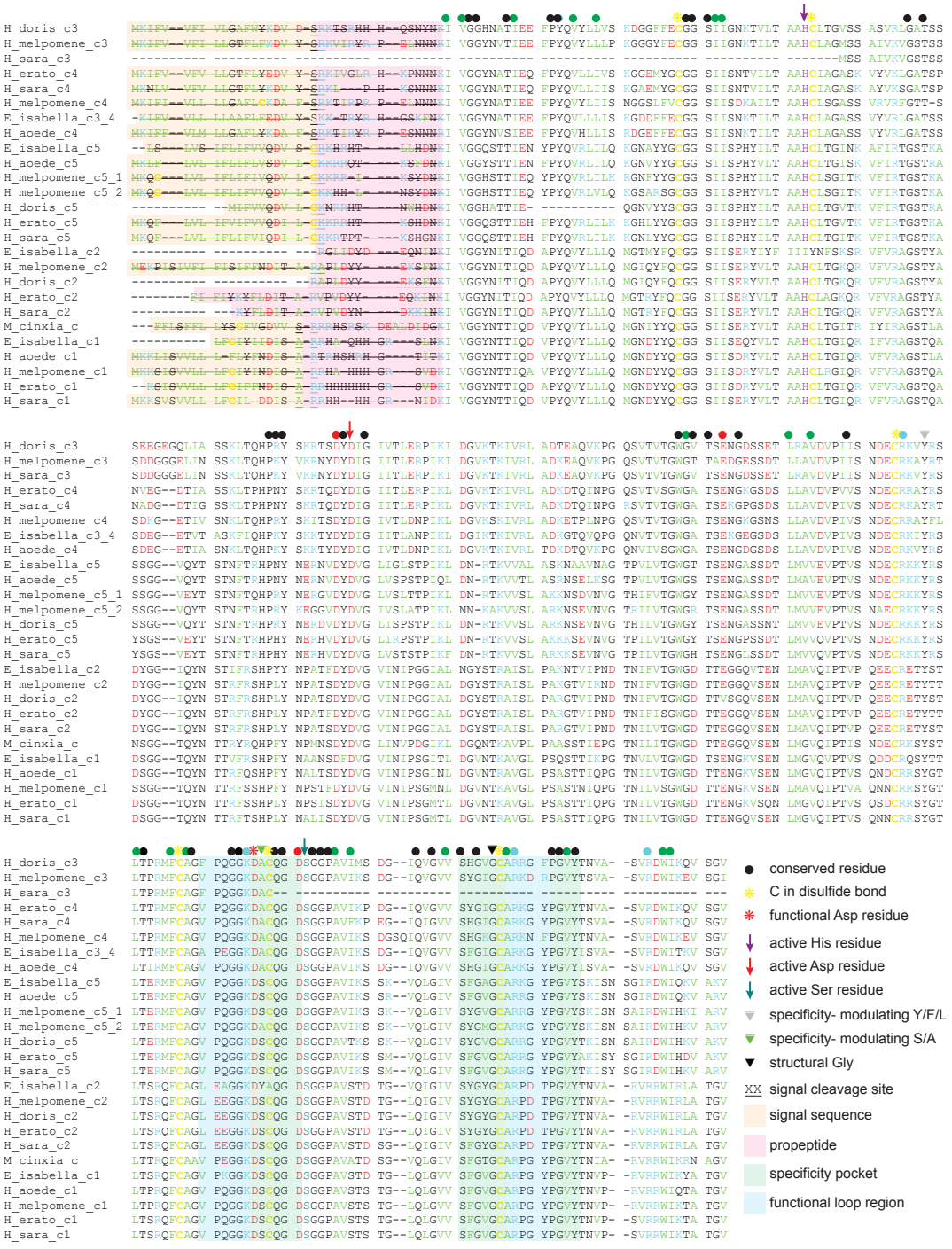


Figure 4.5: Cocoonases from *Heliconius* and *Eueides* butterflies. See Figure 4.4 for legend.

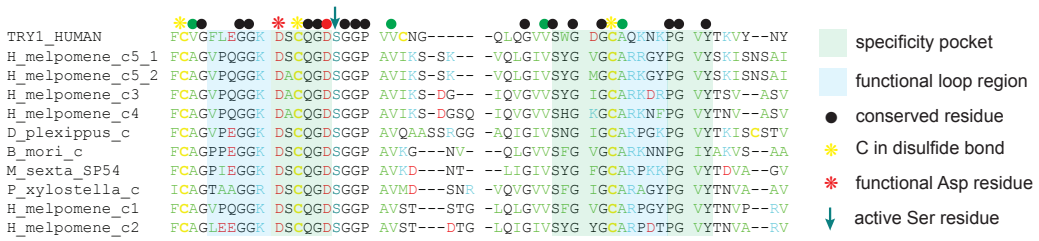


Figure 4.6: Amino acid sequence alignment comparing a representative of each type of heliconiine cocoonase with those of other lepidopterans. The sequence region shown contains the active Ser, the specificity pocket, and the functional loops. Full alignments are given in Figures 4.4 and 4.5. (H_melpomene_c5_1 is *H. melpomene cocoonase 5a*, and H_melpomene_c5_2 is *H. melpomene cocoonase 5b*.)

4.3.3 Diversity of surface residues suggests specialization for different environments

Within each cocoonase type, sequences from different species demonstrate a high level of conservation. Figure 4.7 illustrates the sequence conservation for all cocoonases, plotted on the structure of *H. melpomene* cocoonase 1. Strikingly, the side chains of most of the highly conserved residues are directed toward the interior of the protein, while many of the variable residues project into the solvent. This pattern suggests that the evolution of different cocoonase types may have been driven by the need to diffuse into and remain solvated in different environments, rather than by differences in substrate specificity.

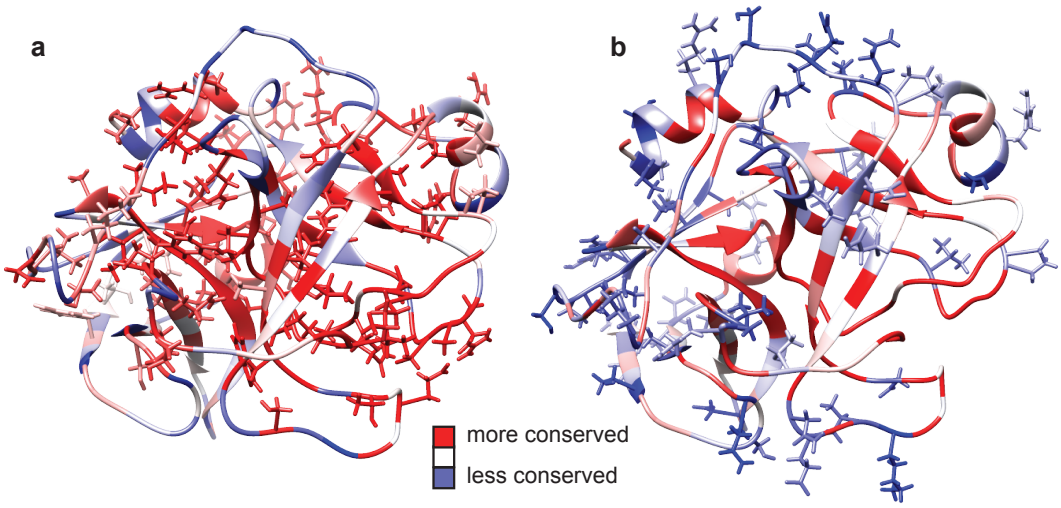


Figure 4.7: Conservation map of all cocoonase sequences plotted on the molecular model of *H. melpomene* cocoonase 1. The percent conservation for each position ranges from red (most conserved) to blue (least conserved). (a) Side chains for the most conserved residues. (b) Side chains for the least conserved residues.

The observation that cocoonases differ primarily in their outward-facing residues suggests that functional differences may alter their surface properties; such an adaptation could arise e.g. in response to the need to efficiently diffuse through the chemically heterogeneous materials that comprise pollen grains. The properties most likely to vary in such a scenario are residue charge, overall hydrophobicity, and whether or not the residue side chain is polar, all of which influence solvation and ability to diffuse within relatively polar vs. non-polar environments. Principal Component Analysis (PCA) of solvation-relevant surface properties for the protein set (Figure 4.8) shows that the five sequence-based clusters clearly follow a well-ordered, circumplex pattern from those whose surfaces are relatively non-polar and hydrophobic (cocoonase 1), to those with more moderate levels of hydrophobicity and relatively negative mean charge (cocoonases 2 and 3), to those that are relatively polar, hydrophilic, and neutral (cocoonase 4), to those that are polar and hydrophilic with a relatively positive mean charge (cocoonase 5). With the exception of cluster 3, the five clusters are well-separated in PCA space, and there is no indication of clustering by species (as might be expected if surface characteristics were primarily a response to species-specific

feeding requirements). Taken together, these results are consistent with the hypothesis that differences in sequence among cocoonases are primarily driven by a requirement for diversity in solvation-related properties, and that this requirement is broadly shared among pollen-feeding lepidopterans.

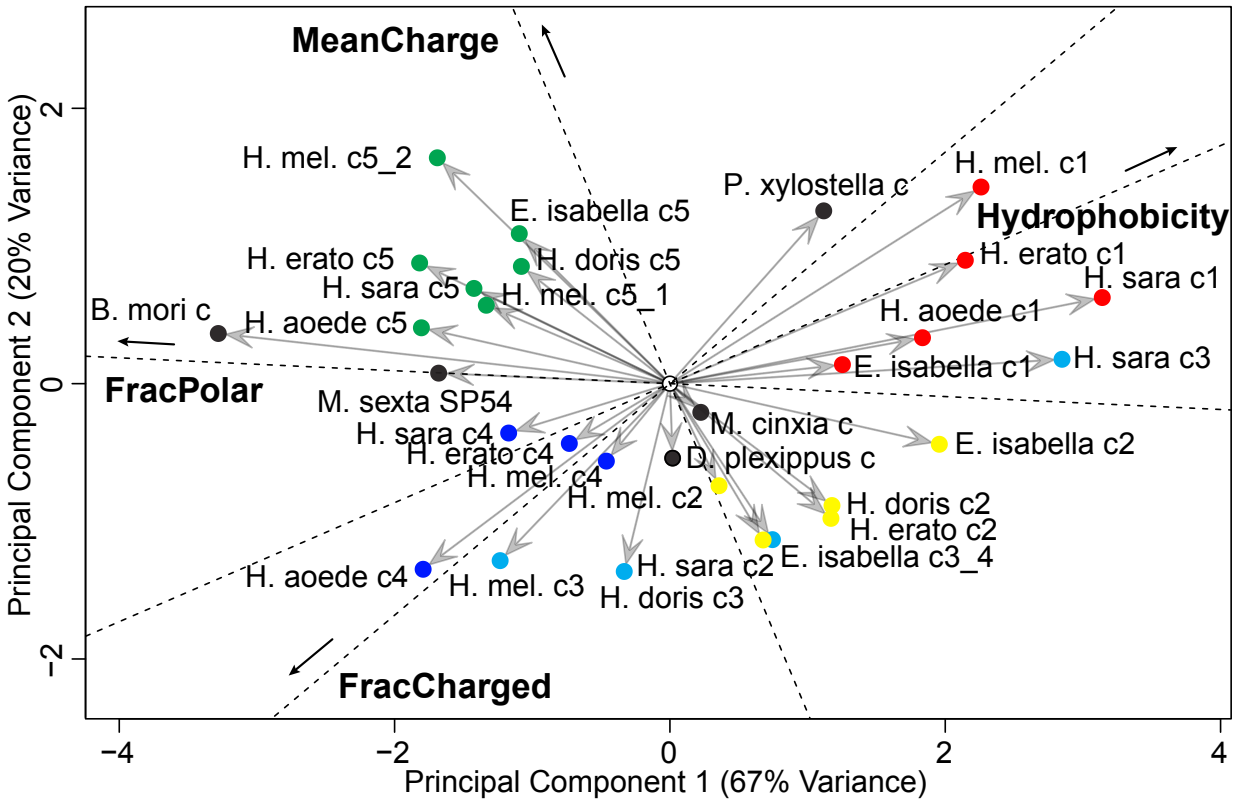


Figure 4.8: PCA of solvation-relevant surface residue properties. Projections of original variables into the PCA space are shown by dotted lines; arrows indicate positive direction. Original variables include: overall mean residue charge (MeanCharge), fraction of residues that are charged (FracCharged), fraction of residues that are polar (FracPolar), and overall hydrophobicity (Hydrophobicity). Colors indicate protein sequence-based clustering (see Figure 4.2 for clusters).

In summary, *cocoonase* is highly expressed in the mouthparts of butterflies and the function of this protease in *Heliconius* is likely to involve feeding, specifically pollen feeding. Solvent exposed surface residues tend to have a higher mutation rate than slower evolving interior residues, which are constrained by their important roles in structures such as the active site [132]. While surface residues may play a role in substrate and modulatory ligand recognition

when associated with the active site and/or specialized regions such as binding pockets [133], we see no evidence of systematic variation in such residues among the heliconiine cocoonases. Variation is instead seen in residues comprising the bulk of the protein surface, which are more typically implicated in stability and solvation within particular chemical environments [134, 135]. Cocoonase surface residue diversity may therefore be important for solvation and stability in the heterogeneous mix of chemical microenvironments comprising pollen grains [136], which is further increased by biochemical differences in the plants on which they feed.

4.4 Summary

With these methods, I have done a lot of sequences alignments and annotations, as well as the analysis of the results. I learned that the surface residues are the defining characteristic of the cocoonases from the *Heliconius* butterflies. All of this work as helped me to streamline the process that we use to do these studies, and make this type of work more accessible to an inexperience scientific audience. The how-to manual that I have written to explain our methods immediately follows this section.

4.5 Alignment and Annotation Manual

4.5.1 Sequence Alignment

Before beginning the alignment process, first the proteins in the chosen class must be clustered according to sequence (dis)similarity. The sequence dissimilarity used to make these clusters is the e-distance metric of Székely and Rizzo [137] (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, which we define as the raw sequence dissimilarity (1 - (%identity)/100). Each cluster can then be separately aligned with appropriate reference sequences. Reference sequences should be as similar as possible to those in the cluster and

have been functionally characterized; a solved structure is ideal, if there is none available, we need at least the placements of the active site residues, disulfide bonds, and any other functionally relevant residues (such as metal binding sites, pro-sequences, etc.). These references are found using the BLAST search tool on the UniProtPK website.

The next step is to generate the initial alignment. If the cluster only has 2 sequences (one reference and one of interest), then we use Needle [138] to generate the alignment. For all other clusters, we use ClustalOmega [112]. Both of these tools are part of the EMBL-EBI bioinformatics web and programmatic tools framework [139]. On the ClustalOmega website (`https://www.ebi.ac.uk/Tools/msa/clustalo/`), paste the list of FASTA sequences for the cluster into the box. For most cases, running with the default options is best, however, the "output format" should be set to "ClustalW" (not "ClustalW with character counts"). When it is done running, click on the "results summary" tab. From there right click (or control click on Mac) on "Alignment in CLUSTAL format" and save it as a .clustal file. These must be .clustal files (not .txt), since that is what the alignment script is set to accept as an input. Also make sure to save the file names with no spaces, use underscores instead (test_cluster.clustal). At this point, the .clustal file has the alignment of the protein sequences in rows of 60 residues with no spaces in each row, it also has the names of the proteins as the full name from the FASTA header which are often long and uninformative.

The script to reformat the .clustal file into the alignment used for the rest of the annotation process uses Python. For Windows users, Python must first be installed before the script can run (Python comes pre-installed on Mac computers). Install the current version of Python (3.x), at this point the alignment script may be used. There are two versions of the alignment script, '60_to_100.py' and '60_to_100_w_color.py'. '60_to_100.py' takes the .clustal alignment and reformats it to be lines of 100 residues with a space every 10 residues to make it easier to read and locate particular residues. The script also renames the proteins to match the UniProt or PDB identifiers, and the sequences from the *Drosera capensis* to have the form DCAP_#### instead of their MakerP names. The '60_to_100_w_color.py'

does everything that the previous script does, as well as coloring the residues according to the scheme described in Table 4.1. There are PC and Mac versions of each of these scripts (the names end in _pc or _mac), but the only difference between them is the file paths associated with each operating system. These scripts are shows in Appendix B: Alignment Script as Listings B.1 and B.2 respectively.

Table 4.1: Residue Coloring Scheme

Residues	Type	Formatting
A,F,M,L,I,V,W	hydrophobic	Green
D,E	negatively charged	Red
R,K	positively charged	Cyan
C	cysteine	Yellow & Bold
All others		Black

After choosing the appropriate script, make a copy of it to the Desktop. Open the script in a text editor and change the username to match the computer being used. Also change the name of the input and output files (default is test.clustal) to match the current alignment being formatted. After saving, the script can be run from the command terminal, outputting a formatted .rtf file. This file should be checked to make sure all the protein names changed correctly, occasionally there are odd FASTA headers that the script doesn't know how to deal with and they will remain in their long form. These can be either changed manually now in the .rtf file, or they can be changed in the .clustal file and rerun if the formatting is disturbed too much. After checking the protein names, this is now the final alignment file that will be imported into Adobe Illustrator and annotated.

4.5.2 Annotation

There are two parts to the annotation process, the annotation figure (which is the alignment figure with annotations added to show conserved residues, disulfide bonds, active sites, etc.) and a spreadsheet (.csv) detailing the information from the figure as well as other information important to the *in silico* maturation process.

To import the alignment into Adobe Illustrator, first open up the 'alignment_template.ai' file. "Save As" the appropriate name to match the cluster (if using the lab copy of Adobe Illustrator CS6 for Windows, make sure to save as the CS5 version to keep compatibility with the lab Mac version of Adobe Illustrator). Then go to "File" and "Place" and select the .rtf file generated from the alignment. This make take a few minutes, as it sometimes hangs up on the Place process depending on the size of the cluster. Once this is done, the alignment will be in a text box that can be expanded to fill the artboard in the template. Now the annotating can begin.

Start the annotation by marking the conserved residues with the • symbol that matches the color of the residue, place the symbol above the column of the conserved residues. Conserved residues must be identical in every protein in the cluster to be marked, though they should also be marked if they are present at least 75% of the sequences (the remainder just having a - in that location). Mark the cysteines that participate in disulfide bonds with a large yellow $*$ symbol, determine which cysteines are participating in disulfide bonds by matching to the reference sequence in UniProt. Mark important active site residues with a $\downarrow$ in the color corresponding with that site. The signal sequences (if present) should be highlighted with a pale orange color. The pro-sequences (if present) should be highlighted with a pale pink color. Residues that are cut off in the mature protein need to be ~~strikethrough~~ (i.e. pro-sequence, signal sequence, etc). The cleavage sites residues (the last one cut off and the first one remaining) should be underlined, both for where the anything is cut off the mature sequence [~~X~~X] and where any other cleavage occurs (such as the signal sequence from the pro-sequence [~~X~~X], etc.). Other important binding sites (such as substrate or metal binding) should be marked with appropriately colored $*$ symbols as necessary for that particular cluster or class of enzymes. Examples of these symbols and colors are available in the template file and are also shown in Figure 4.9. These annotations create a straightforward way of visually representing the differences and similarities withing a cluster of sequences. They also clearly illustrate which portions of the sequence remain in the active enzyme, and

which portions are initially present to serve other purposes, such as directing the enzyme to its final location in the organism or protecting the active site.

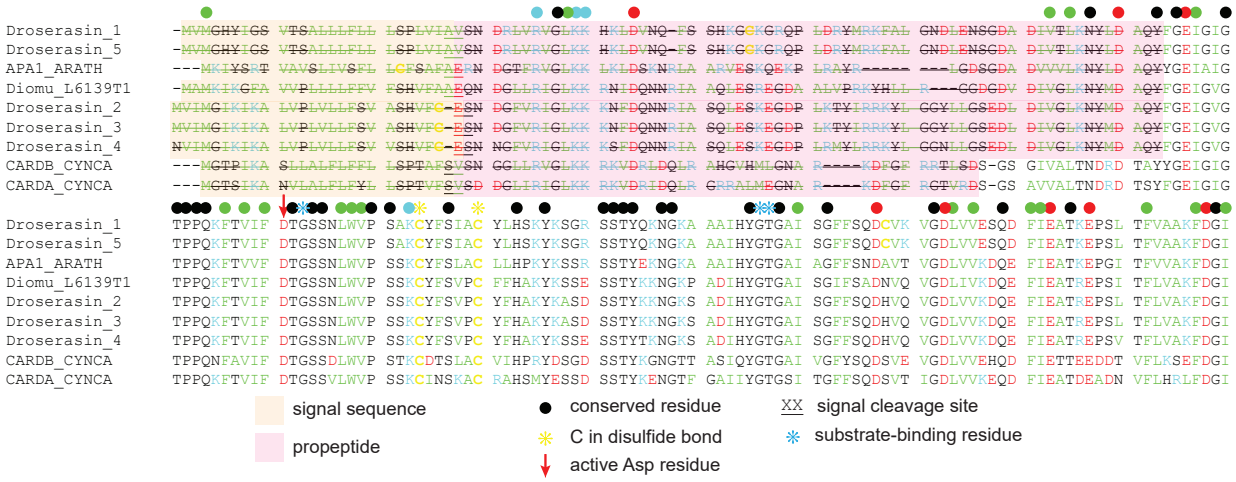


Figure 4.9: Excerpt of the sequence annotation for Droserasin cluster of the aspartic proteases discovered from D. capensis [109]. Active site residues are shown in colored arrows (red marking active aspartic acids). Conserved residues are marked with colored dots corresponding to the type of residue (see Table 4.1 for coloring scheme), cysteines involved disulfide bonds are marked in yellow asterisk, the signal sequences are highlighted in the light orange color and the pro-sequence in pink. Cleavage sites are underlined, and any part of the sequence not present in the mature protein is struckthrough.

The next step is to fill out a spreadsheet with information needed by the *in silico* maturation script to run. The basic information needed is listed in Table 4.2, with extra columns (such as SBR#Pre - the 4 residues before Signal Binding Residue #) being added as necessary for each class of enzyme; there should be a Prefix and Suffix column for each special residue of interest or the start and end of each special protein domain. An example spreadsheet is available to start from as 'annotation_spreadsheet_template.csv'. The spreadsheet should include all proteins and references for all clusters within a class of enzymes. Most of the columns are self-explanatory from the table, but a couple of them require additional files. DSRefFASTA is the file name for .fasta file containing only the FASTA sequence of the reference protein for the applicable cluster. DSRefBonds is a list of the disulfide bonds locations for the reference protein; these can be found in the UniProtKB entry for the reference protein, and should be a .disulfides file containing pairs of numbers corresponding to the residues that make up

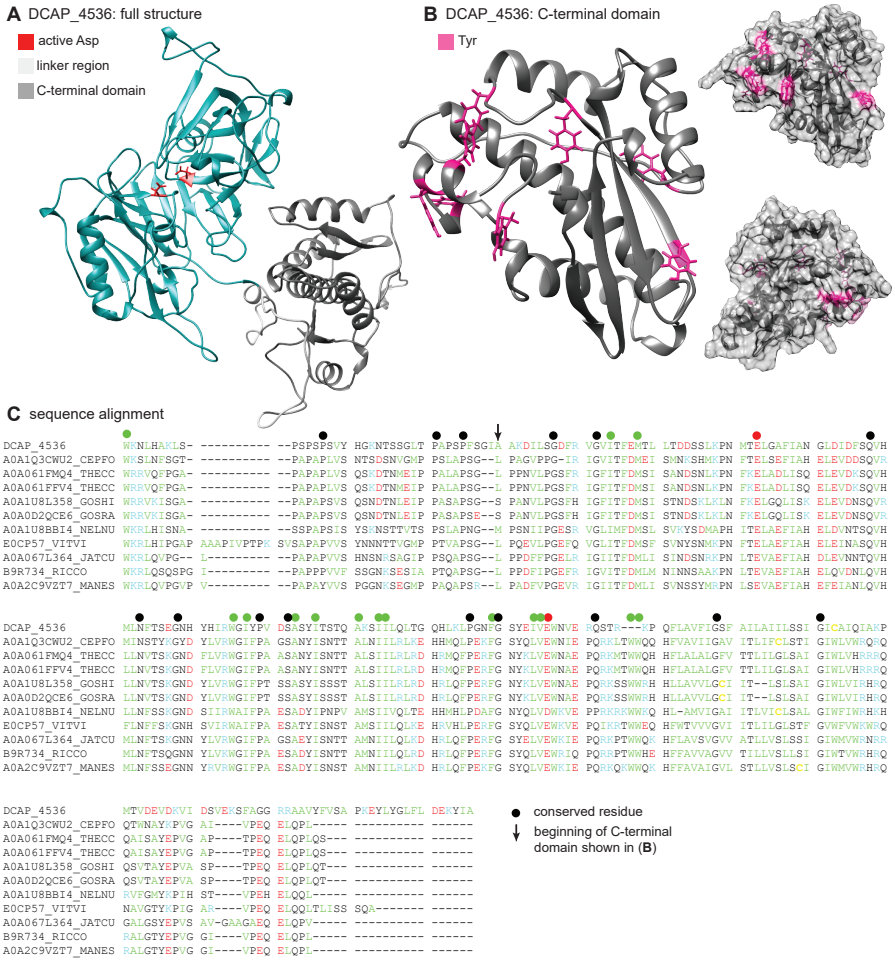
each disulfide bond.

Table 4.2: Annotation Spreadsheet Headings

Column Heading	Description
PDB	full file name of structure (i.e. DCAP_7323_full_m1.pdb)
Class	type of enzyme
DSRefFASTA	FASTA sequence for reference comparing disulfide bonds to (.fasta file)
DSRefBonds	listing of locations of DS bonds (.disulfides file)
ProNTPre	4 residues before the N-terminal pro-sequence cut off site
ProNTSuf	4 residues after the N-terminal pro-sequence cut off site
ASPreX#	4 residues before Active Site residue # (one for each AS residue in the protein)
ASSufX#	4 residues after Active Site residue # (one for each AS residue in the protein)
ProCTPre	4 residues before the C-terminal pro-sequence cut off site
ProCTSuf	4 residues after the C-terminal pro-sequence cut off site
pH	varies depending on final environment of the protein (i.e. vacuole vs cytoplasm, etc)
Temp	varies depending on final environment of the protein (i.e. psychrophiles vs mesophiles, etc)
Preprocess	(TRUE/FALSE)
Box	
EqTime	
BoxAdjTime	
Equilibrate	(TRUE/FALSE)
Note	anything important (i.e. no pro-sequence)

Chapter 5

Aspartic Proteases



The chapter cover figure shows some selected figures illustrating concepts discussed in this chapter; (a) Rosetta model of the pre-equilibration structure of DCAP_4536, (b) Detail of the C-terminal domain of DCAP_4536, (c) sequence alignment for the C-terminal domain of DCAP_4536 with those of selected other plant aspartic proteases.

5.1 Introduction

Aspartic proteases are endopeptidases that function best in acidic environments. In plants, they are used in a variety of contexts including protein processing, [140] resistance to pathogens [141], drought response, [142] and programmed cell death during senecence [143]. In addition to these basic functions, carnivorous plants require a variety of proteases with different substrate affinities and cleavage sites to effectively digest the proteins from their prey. However, with the exception of the nepethesins and the cysteine protease dionain 1 [144], carnivorous plant digestive enzymes have yet to be investigated in detail. Therefore, digestive proteases from carnivorous plants represent an underutilized resource for chemical biology and biotechnology. These enzymes have the ability to digest prey items down to small molecule components, even without the benefit of mastication or other tissue processing. They also must withstand onslaughts from bacterial and fungal growth. Depending on the species, these enzymes are optimally active over a pH range of 2-6 [145, 146]. Enzymes that are active in mildly acidic solutions are particularly attractive for bottom-up proteomics applications: target proteins are less susceptible to disulfide scrambling and other modifications that cause problems at the higher pH values preferred by mammalian trypsin. The pH of *D. capensis* digestive mucilage is approximately 5-5.5 [147], while the pH of plant vacuoles ranges from 4.5-6 [148]. The aspartic proteases discussed in this work are classified as family A1 by the MEROPS database [149]. Family A1 is further divided into subfamilies A1A (pepsin-like proteases) and A1B (nepenthesin-like proteases), both of which are found in *D. capensis* [114].

The cardosins (MEROPS A1A) are among the best-characterised plant aspartic proteases

due to their role in the traditional preparation of sheep's milk cheese in the Mediterranean region [150, 151]. In this application, cardosins A and B cleave ovine caseins with a different activity profile from mammalian enzymes such as chymosin, resulting in a unique set of peptide products generated, resulting in the characteristic flavor and texture of the cheese [152, 153].

Aspartic proteases from the related genus *Nepenthes* are already being used in mass spectrometry applications where protein samples are digested and the fragments analyzed [154, 155, 156]. Extension of these techniques to ever more complex mixtures requires an extensive toolbox of enzymes with different substrate cleavage sites. In general, the nepenthesins are characterized by high stability over a wide range of pHs and temperatures, with maximum activity at pH 2.5 [157]. Nepenthesins 1 and 2 are being used in hydrogen/deuterium exchange mass spectrometry (HDX-MS) in place of the more common pepsin. The nepenthesins are well-suited to this application due to their activity profiles: unlike pepsin, they provide rapid and efficient cleavage C-terminal to K, R, and H residues [158]. Like other aspartic proteases, nepenthesins are inhibited by pepstatin A [159].

5.2 Methods

5.2.1 Sequence Alignment

Sequence alignments were performed using ClustalOmega [112], with settings for gap open penalty = 10.0 and gap extension penalty = 0.05, hydrophilic residues = GPSNDQERK, and the BLOSUM weight matrix. The presence and position of a signal sequence flagging the protein for secretion was predicted using the program SignalP 4.1 [113]. Quality control was performed using both sequence alignment and inspection of the Rosetta structures; proteins where one of the critical active residues (D 100 or D 287, APA1_ARATH, full-length sequence numbering) is missing or that contained truncations in the active region were not selected for modeling or further analysis. Removal of sequences that are thus

unlikely to produce active proteases left seven aspartic proteases from *D.capensis* and three from the Venus flytrap *Dionaea muscipula*.

5.2.2 Sequence Alignment and Prediction of Putative Protein Structures

Structures were predicted using a three-stage process. First, an initial model was created for each complete sequence using the Robetta server [160]; the Robetta implementation of the Rosetta [161] system generates predictions from sequence information using a combination of comparative modeling and all-atom refinement based on a simplified forcefield. For each enzyme in this set, two structures were predicted using Robetta. The full-length structures include the N-terminal signal peptides and pro-sequences, and in the case of the droserasins, the PSIs, all of which are cleaved during maturation *in vivo*. Because the emphasis here is on the potential use of these enzymes in biotechnology or chemical biology applications, we also modeled the forms that would likely be produced for use in the laboratory, consisting of the catalytic domains only, with one continuous chain excluding the PSI. These preliminary structures were then subject to *in silico maturation*, in which disulfide bonds identified by homology to known homologs were added and the protonation states of active site residues were fixed to their literature values. Finally, in the third phase, each corrected structure was equilibrated in explicit solvent under periodic boundary conditions in NAMD [162] using the CHARMM22 forcefield [115] with the CMAP correction [163] and the TIP3P model for water [164]; following minimization, each structure was simulated at 293K for 500ps, with the final conformation retained for subsequent analysis. This process was performed for the seven protease sequences from *D. capensis*, four from related carnivorous plants, and three well-characterized reference sequences from other organisms (see below); where published structures were available, these were used as the initial starting model (following removal of heteroatoms and protonation using REDUCE [116] as required). The PDB files corresponding to the equilibrated structures for all the proteins discussed in this manuscript are

available in the Supplementary Information. Figures showing selected predicted structures were generated using Chimera [165].

5.2.3 Network Modeling and Analysis

In collaboration with Prof. Carter Butts research group, each equilibrated protein structure was mapped to the network representation of [108] using custom scripts employing both VMD [166] and the `statnet` toolkit [167, 168] within the R statistical computing system [119]. Each vertex within the resulting protein structure network (PSN) represents a chemical group, with edges representing potential interaction as determined by proximity within the protein structure. PSNs were then compared using the structural distance technique of [169], which provides a uniform way to compare the underlying structures of networks (i.e. graphs) with different vertex sets; this involves mapping both graphs onto a common vertex set (adding isolated vertices to the smaller graph as needed) such that the differences between the two mapped networks are minimized with respect to an underlying metric. The value of this metric after mapping is the structural distance. Here, distances were computed between unlabeled graphs based on an underlying Hamming metric, and can be interpreted as the minimum number of edge changes required to transform a member of the isomorphism class of the first graph (i.e., the set of all graphs having the same underlying typology) into a member of the isomorphism class of the second (or vice versa). The raw structural distance between each pair of PSNs was then normalized by graph order, yielding a metric corresponding to edge changes per vertex. Normalized structural distances between PSNs were analyzed via metric multidimensional scaling and hierarchical clustering using R. Additional network visualization and analysis was performed using the `sna` library [170] within `statnet`.

5.2.4 PCA

Relative solvent accessibility (RSA) values were calculated for all equilibrated structures using DSSP 2.2.1 [117]; residues with RSA values less than 0.2 were regarded as buried, with

other residues classified as solvent exposed. For the set of solvent exposed residues within each structure, the fraction of polar residues and charged residues, mean residue charge, and mean hydrophobicity (using the scale of Kyte and Doolittle [118]) were calculated. All data analysis and visualization was performed using R [119]. The first two principal components jointly accounted for 82% of the standardized variance. Projections of the original variables into the PCA space were also calculated to assist with interpretation.

5.3 Results and Discussion

5.3.1 *D. capensis* Aspartic Proteases Cluster Into A1A and A1B Subfamilies Based on Protein Sequence

Several aspartic proteases with moderate sequence identity to mammalian pepsin have been identified from the genome of *D. capensis*. In carnivorous plants, MEROPS A1 proteases are among those involved in digestion of prey. All *D. capensis* sequences previously annotated as coding for MEROPS A1 aspartic proteases using the MAKER-P (v2.31.8) pipeline [171] and a BLAST search against SwissProt (downloaded 8/30/15) and InterProScan [172] were clustered by sequence similarity. Several previously-characterized aspartic proteases from other plants are also included as reference sequences. The *D. capensis* aspartic protease sequences, clustered in Figure 5.1 divide into two types: A1A (the droserasins) and A1B (nepenthesins). Protein sequence alignments comparing the sequences of these enzymes' active regions with well-characterized references can be found in Figures 5.2 (full-length sequences, including signal sequences, pro-sequences, and PSIs) and 5.3 (mature constructs) for the droserasins and Figure 5.4 for the nepenthesins.

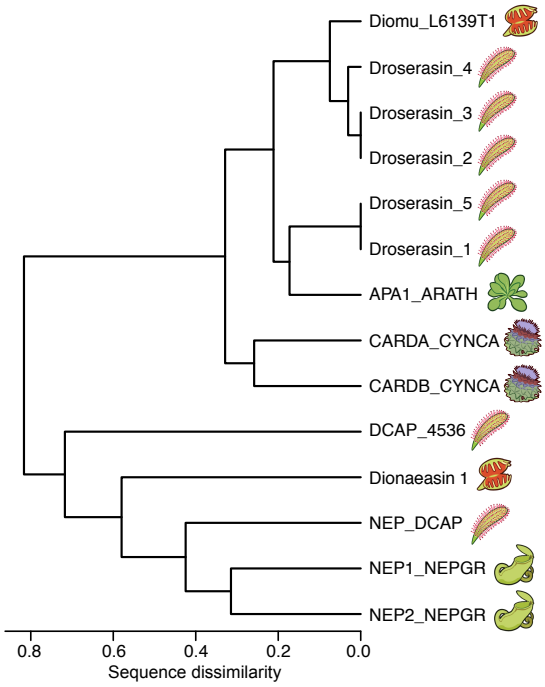


Figure 5.1: Clustering of aspartic protease sequences identified from the *D. capensis* genome along with related sequences from other plants and well-characterized reference sequences.

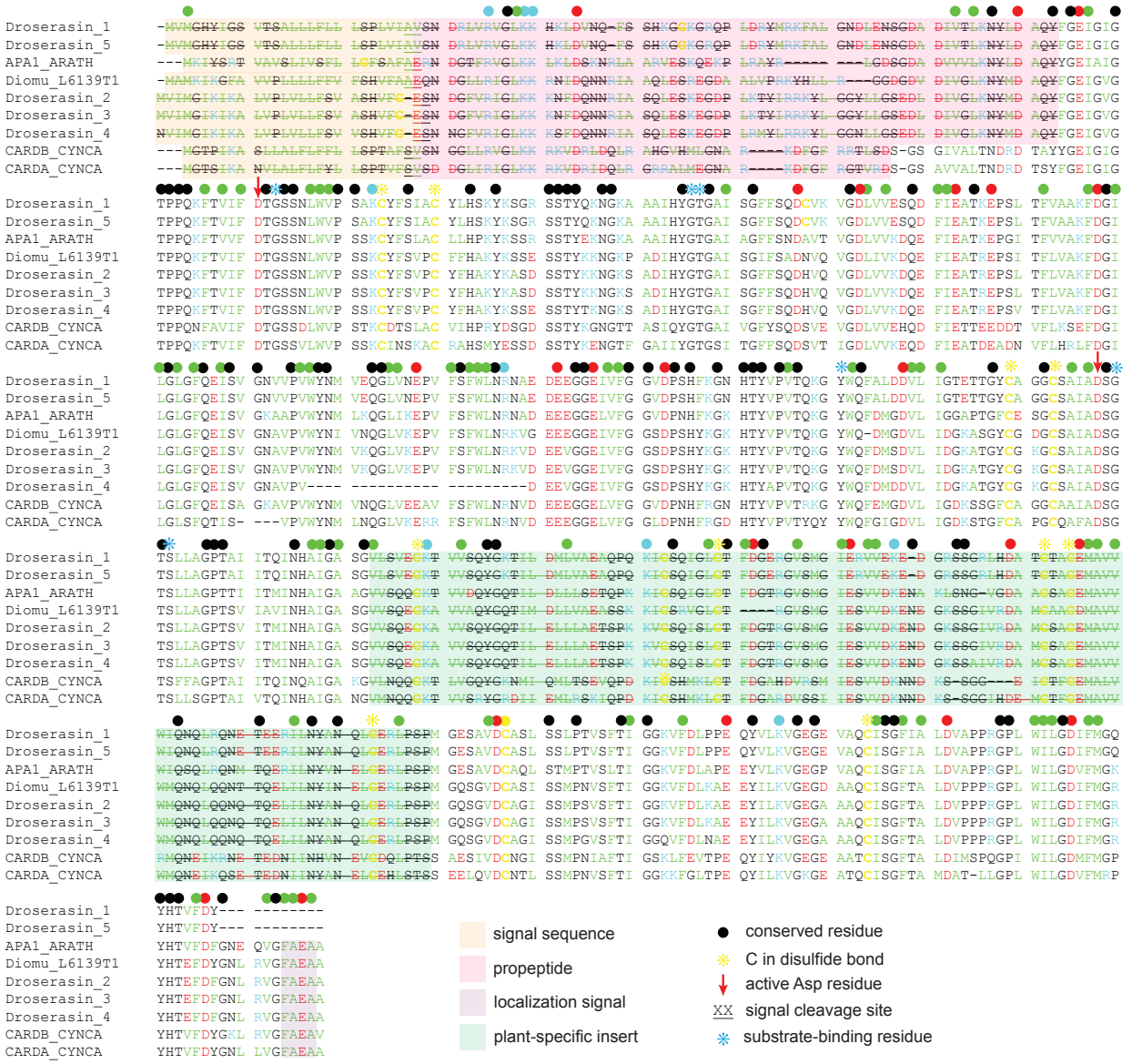


Figure 5.2: Sequence comparison for the droserasins. Hydrophobic residues are shown in green, positively charged residues in blue, negatively charged residues in red, and cysteines in yellow. Conserved Cys residues involved in structure-stabilizing disulfide bonds are indicated with yellow asterisks, while other residues conserved across all the sequences considered are indicated with solid dots. The active aspartic acid residues are indicated with red arrows. The predicted signal and pro-sequences, which are strongly conserved, are highlighted using colored boxes. Strikethrough text indicates parts of the sequence that are expressed but removed during post-translational processing; for most of these proteins, this includes both an N-terminal region comprising of the signal peptide and the pro-sequence and the internal PSI region. The localization signal, present in all but Droserasins 1 and 5, is the vacuole targeting signal.

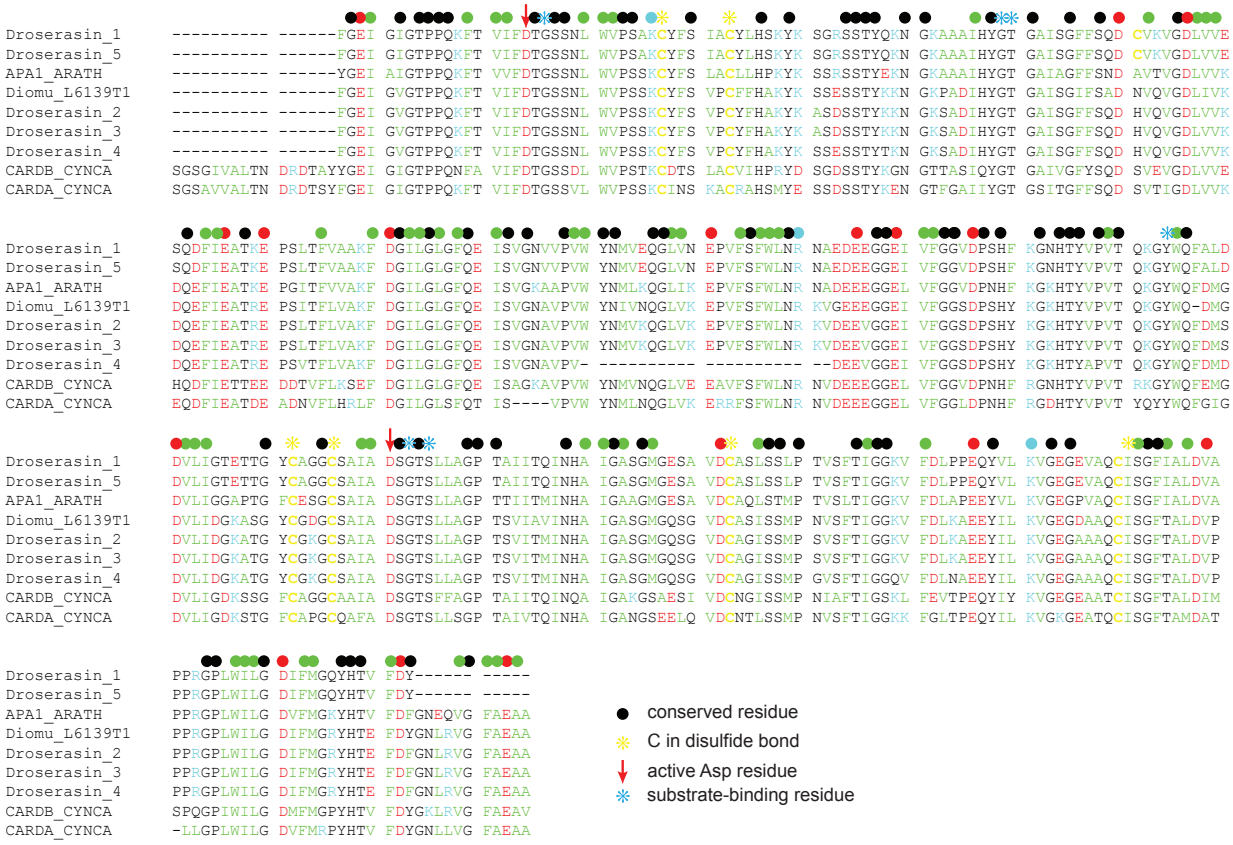


Figure 5.3: Sequence alignment of the constructs used for molecular modeling of droserasins and annotation reference sequences (mature sequences). The signal peptides, pro-sequences, and PSIs have been removed.

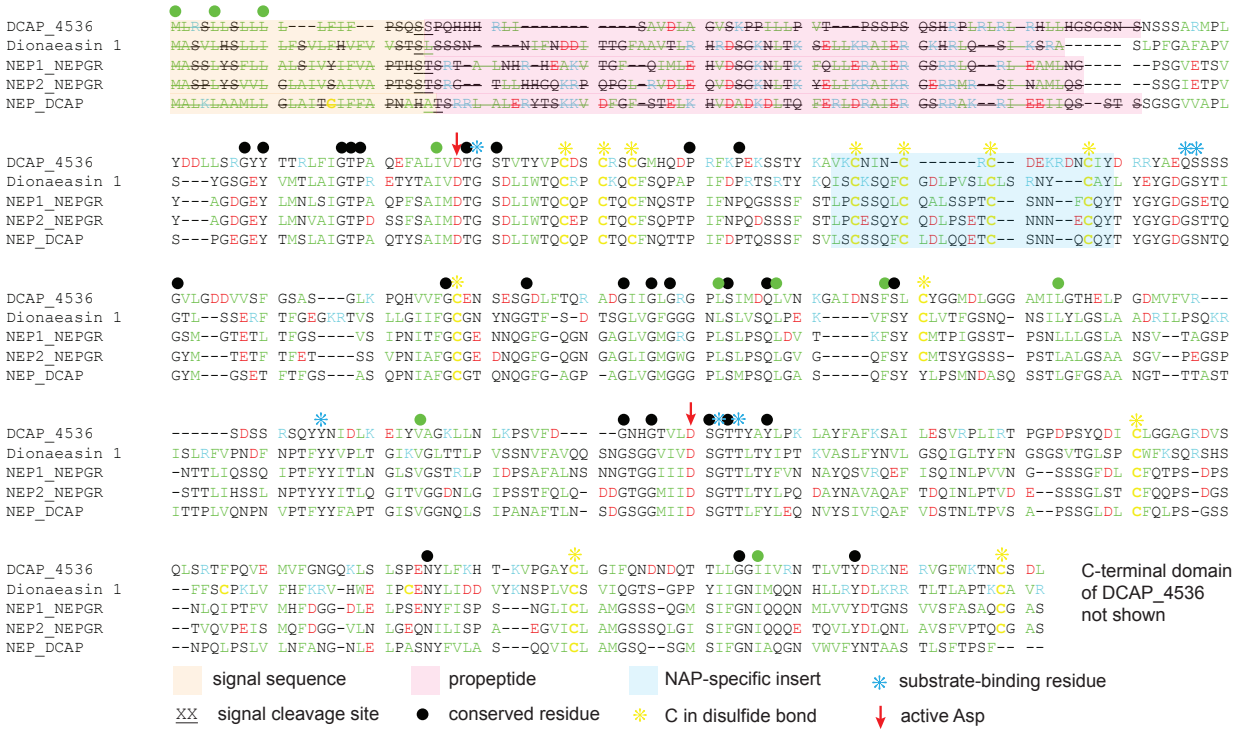


Figure 5.4: Sequence alignment for *D. capensis* proteases from MEROPS family A1B (nepenthesins), along with selected reference sequences. Annotations are the same as in Figure 5.2, except for the NAP-specific insert (blue). All signal peptides and pro-sequences were removed for molecular modeling of the mature sequences.

The sequences in the droserasin cluster have a high degree of sequence identity with each other, particularly in the catalytic domain and the plant-specific insert (PSI), a sequence region of approximately 100 residues that is excised during maturation in vivo, whereupon it acts as an antimicrobial peptide. The droserasins are also highly similar to APA1_ARATH, a vacuolar protein from *A. thaliana*, and to cardosins A and B, aspartic proteases previously characterized from the cardoon, *Cynara cardunculus*. Previous studies have shown that recombinantly expressed APA1_ARATH is maximally efficient at pH 5.3, and has a highly specific cleavage profile with respect to the insulin β-chain [173]. The full-length droserasins share important functional sequence features with APA1_ARATH and the cardosins, including the active site residues, the disulfide bonding pattern, and the PSI. Several of the proteins in this cluster contain the tetrapeptide FAEA near the C-terminal end, which is consistent with targeting to the vacuole [174]. However, the trafficking of aspartic proteases is compli-

cated, with the same protein undergoing secretion or transport to the vacuole depending on the type of cell and its developmental stage [140]. In the cardoon, cardosin A is transported to the vacuole, while cardosin B collects in the extracellular space, even though both proteins have the C-terminal vacuolar transport sequence. When heterologously expressed in model plants, both cardosin A [175] and cardosin B [176, 177] are mostly found in the vacuole as expected, indicating variations in the signaling and transport among different plant species. Thus far, neither the localization of aspartic proteases nor their transport mechanisms have been investigated in *Drosera*, although their expression in the digestive fluid is well-attested in other carnivorous plants from order Caryophylalles, specifically *Nepenthes* pitcher plants [154, 178] and the Venus flytrap, *Dionaea muscipula* [179].

The sequences in the nepenthesin cluster are more diverse, with only moderate sequence identiy among most members of the set, although they all have the key aspartic protease active site residues, disulfide bonding pattern, and substrate-binding residues. They also contain the characteristic nepenthesin aspartic protease (NAP)-specific insert, which is thought to increase the stability of nepenthesins via its three disulfide bonds. Also included for comparison are the endopeptidases Nepenthesin 1 (NEP1_NEPGR) and Nepenthesin 2 (NEP2_NEPGR), which digest prey proteins in the pitchers of *Nepenthes gracilis*, and Dioneaesin 1, which was identified in transcripts from *Dionaea muscipula* traps [179]. Based on sequence similarity, we hypothesize that NEP_DCAP plays a similar digestive role in *D. capensis*. Also included in this set is DCAP_4536, which shares the critical sequence and structural characteristics with the nepenthesins, but has an additional C-terminal domain.

5.3.2 Molecular Modeling Predicts Pepsin-like Enzymes

3D modeling was performed on these sequences to allow the development of hypotheses based on predicted 3D structure and activity. Few structures have been solved for plant aspartic proteases, making traditional homology modeling difficult in this case. Therefore, we use comparative modeling with all-atom refinement, implemented in Rosetta [160, 161].

This approach makes use of comparative modeling based on local sequence identity, combined with de novo structure prediction in regions of lower sequence coverage. We take the Rosetta structures as a starting point for all-atom MD simulation in explicit solvent, after *in silico* maturation. In this process, disulfide bonds are added and side chain protonation states are corrected to reflect the pH of the environment. For each putative functional protease, two structures were predicted using Rosetta: the structure of the full-length sequence, which includes the signal peptide, the pro-sequence, and the PSI if present, and the mature form, in which the signal peptide, pro-sequence, and PSI have been removed. In nature, the mature forms of the cardosins are two-chain enzymes, because the PSI is excised post-translationally. Here, we modeled them as single-chain enzymes, representing the form that would be produced by recombinant expression for biotechnology applications, as previously demonstrated for both cardosin A [150] and cardosin B [180]. The mature sequences were subject to an *in silico* maturation procedure where disulfide bonds were added based on experimental data for the reference sequences (APA1_ARATH, cardosins A and B, and nepenthesins 1 and 2), the pH adjusted to the expected cellular environment of the enzyme (vacuole, drosera mucilage, or *Nepenthes* pitcher fluid), and briefly equilibrated via MD simulation in explicit solvent. The *in silico* maturation and equilibration process refines the initial Rosetta structure predictions and applies biologically relevant modifications. Initial, cut, and mature structural models for a representative protease, Droserasin 2, are shown in Figure 5.5. The full-length structure (Fig. 5.5A) consists of the active region, a secretion signal peptide (light orange), and an N-terminal pro-sequence (pink). The core sequence making up the mature form of this enzyme (dark blue) is structurally similar to pepsin, with two domains of approximately equal size with the active site cleft between them. The cut and equilibrated structures of the mature form are shown in Figure 5.5B and C.

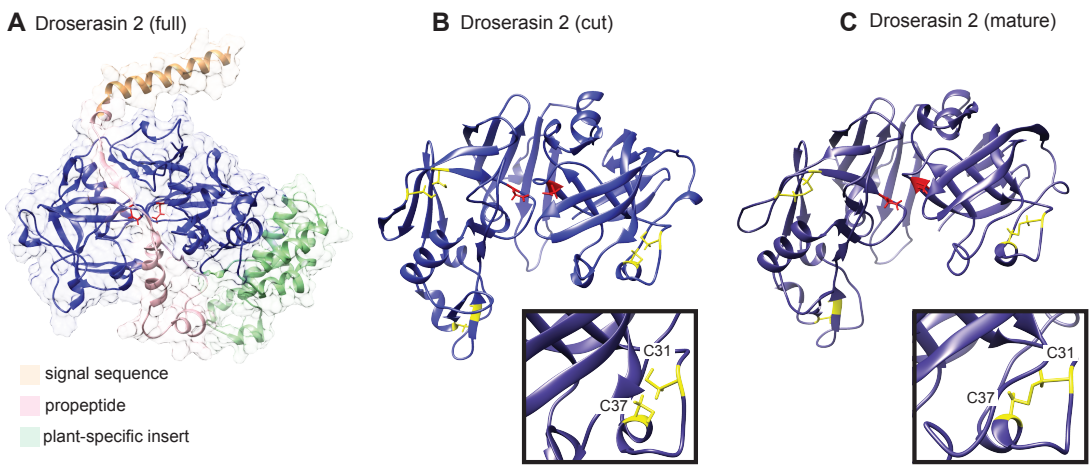


Figure 5.5: (A) Rosetta model of the full sequence of Droserasin 2. Color coding indicates sequence regions, with the catalytic domain shown in dark blue and peptide regions to be removed during maturation in lighter colors. (B) Rosetta model of the cut sequence thought to represent the active enzyme. The active site Asp residues are shown in red, while the Cys residues involved in disulfide bonds are shown in yellow. Inset: cysteine side-chains are not positioned optimally for forming disulfide bonds. (C) the active enzyme after the *in silico* maturation process. Inset: the energy minimization changes side-chain rotameric conformations; disulfide bonds are added.

A1 aspartic proteases are mostly organized in a two-domain, primarily β-sheet fold similar to pepsin. The active site is located in a deep cleft between two domains of roughly equal size, stabilized by an interdomain β sheet located behind the active site residues. Studies of other aspartic proteases, notably that of HIV-1, indicated that only one active Asp is protonated [181, 182], with the proton held between the active site residues in a planar configuration [183]. Depending on the specific protein, access to the active site is modulated to a greater or lesser degree by a β-hairpin "flap" protruding into the cleft. The positioning and dynamics of the flap play a major role in defining substrate specificity in mammalian pepsins as well as retroviral aspartic proteases [184]; because the enzymes are structurally similar and the relevant residues are conserved, we predicted that this structural feature would also be important in the specificity of plant aspartic proteases. The crystal structure of human cathepsin D complexed with the inhibitor pepstatin demonstrated that the substrate lies in the active site in an extended conformation reminiscent of a β-strand [185]. The active

site cleft is long enough to accommodate a hexapeptide, with the substrate held in place via specific hydrogen bonds with backbone and sidechain moieties of the enzyme. The particular resides involved are summarized in Table 5.1 and illustrated in Figure 5.6 for a representative droserasin and nepenthesin.

Table 5.1: Active site moieties involved in specific hydrogen bonds to the substrate

residue number (Dro. 1 numbering)	residue type (Dro.)	residue type (Nep.)	moiety / moieties
20	G	G	backbone carbonyl
63	G	G	backbone amide
64	S	T	backbone amide, side-chain hydroxyl
178	Y	Y	side-chain hydroxyl
207	G	G/E	backbone carbonyl
209	S	T	backbone amide, side-chain hydroxyl

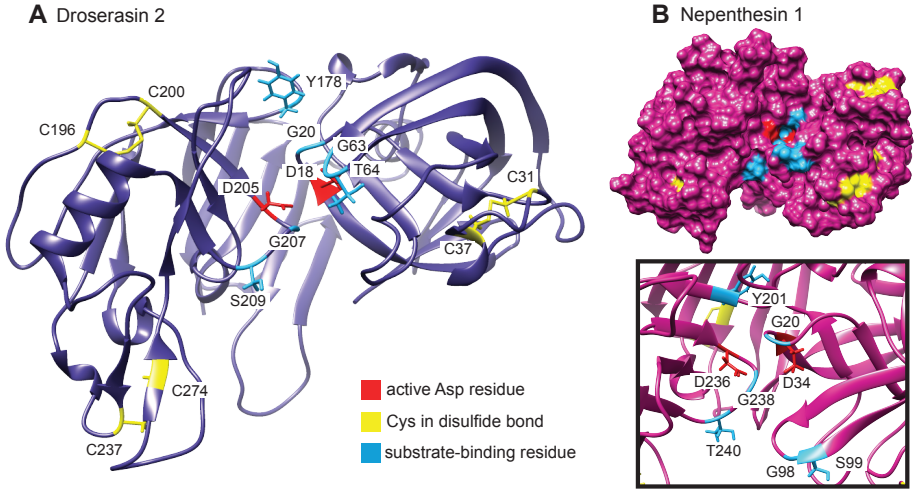


Figure 5.6: Key residues necessary for catalytic activity, stability, and substrate binding are shown for the mature forms of (**A**) Droserasin 2 and (**B**) Nepenthesin 1. Catalytic Asp residues are highlighted in red, Cys residues in disulfide bonds are shown in yellow, and residues involved in forming hydrogen bonds to the substrate are shown in turquoise.

5.3.3 Protein Structure Networks

5.3.4 Surface Residues

Recent literature suggests that surface residues are important for the stability of aspartic proteases [186, 173, 187]. Principal component analysis (PCA) of the surface residue properties yields a pattern where the proteins are mostly distributed by protein type, rather than by species. Overall, the droserasins, nepenthesins, and cardosins are well-separated in the PCA space based on surface residue properties, with the exception of Dionaeasin, which is a nepenthesin but occupies the same region of the space as the droserasins and APA1_ARATH. Nepenthesin 1, Nepenthesin 2, NEP_DCAP, and pig pepsin occupy a region of the space characterized by a lower mean surface charge, a lower fraction of charged residues, and relatively more negative charge on the surface, although they vary substantially on fraction of polar residues and hydrophobicity. This region includes the definitively known digestive enzymes except for Dionaeasin, which has a more positively charged surface. In contrast, the droserasins and their *D. muscipula* ortholog Diomu_L6139T1 cluster in a region characterized by more positively charged residues on the surface and a lesser fraction of polar surface residues. The *A. thaliana* vacuolar protein APA1_ARATH is also found in this region, consistent with the idea that the droserasins are closely related to vacuolar aspartic proteases. Somewhat surprisingly, Cardosins A and B, which share many sequence features with the droserasins and APA1_ARATH, occupy a different region of the space characterized by moderately low hydrophobicity and more polar and charged surface residues. DCAP_4536 is an outlier in surface properties among the proteins in this set, with the most hydrophilic surface residues, highest fraction of both charged and polar residues, and a relatively large number of positively charged surface residues. The construct analyzed here is the mature form, without the C-terminal domain; the surface residue differences observed here are localized to the catalytic domain rather than being added by the additional domain.

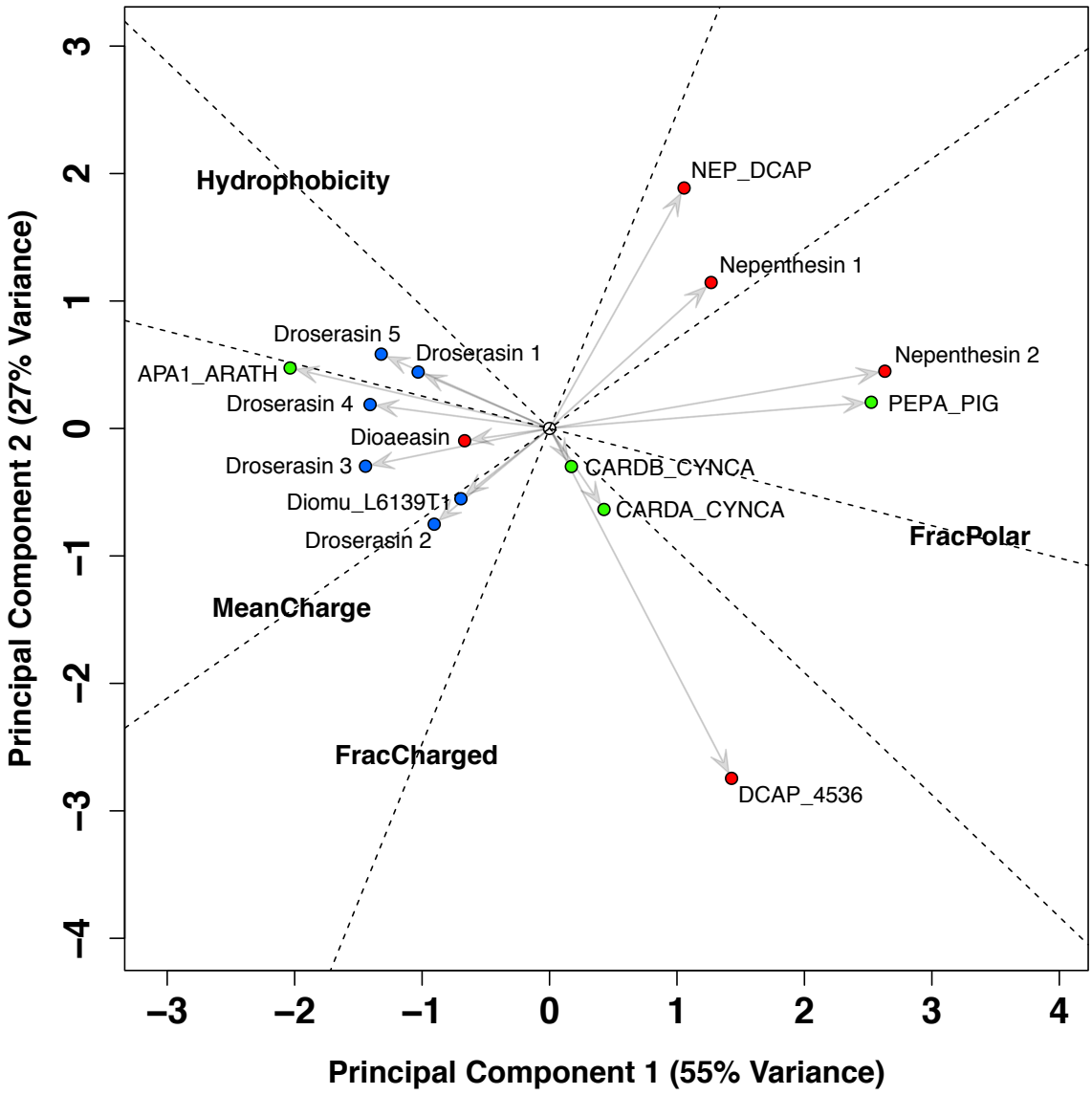


Figure 5.7: PCA of surface residue properties for mature aspartic proteases. Projections of the original variables into the PCA space are shown by dotted lines; arrows indicate positive direction. The original variables used are: overall mean residue charge (MeanCharge), fraction of residues that are charged (FracCharged), fraction of residues that are polar (FracPolar) and overall hydrophobicity (Hydrophobicity). Colors indicate the type of aspartic protease (blue: CP droserasin, red: CP nepenthesin, green: reference sequence.)

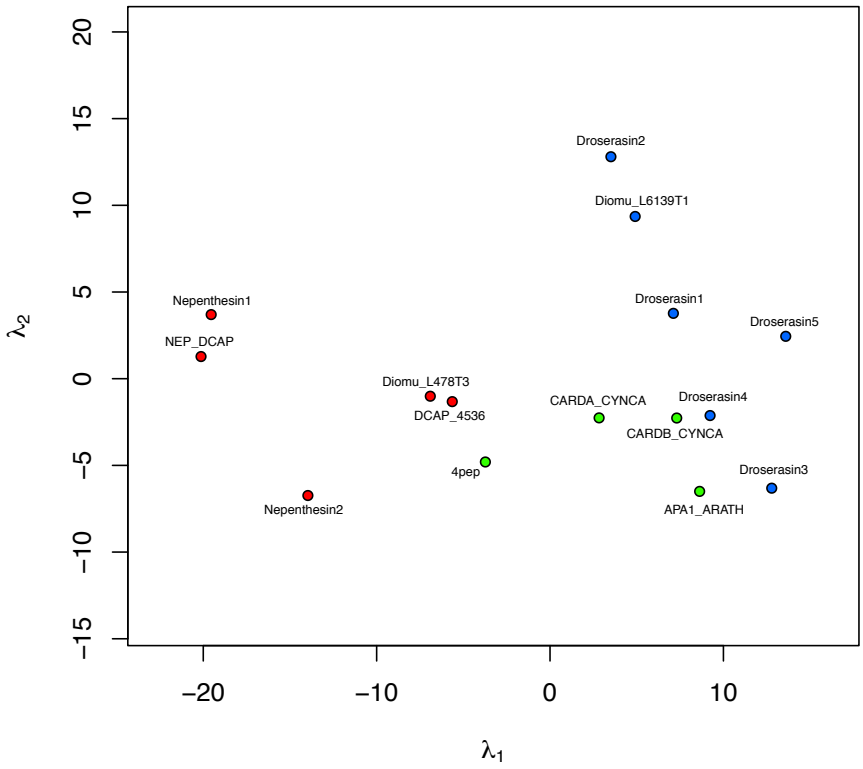


Figure 5.8: Substrate binding geometry. Colors indicate the type of aspartic protease (blue: CP droserasin, red: CP nepenthesin, green: reference sequence.)

In addition, it is also of interest how many of the charged surface residues are negatively charged (the acidic residues) as they have been previously obseved to play a role in protein stability [186]. A comparison of the surfaces of the nepethesins and pepsin can be found in Figure 5.9. It has be previously found that the negative residues are more scattered across the surface in nepenthesin 1 than in pig pepsin [186]. This seems to hold true for all of proteins in the nepenthesin cluster and, to a lesser extent, the droserasins as well. This reduces the chances of charge repulsion within the protein, which was previously hypothesized to have an effect on increasing the stability of the proteins [186].

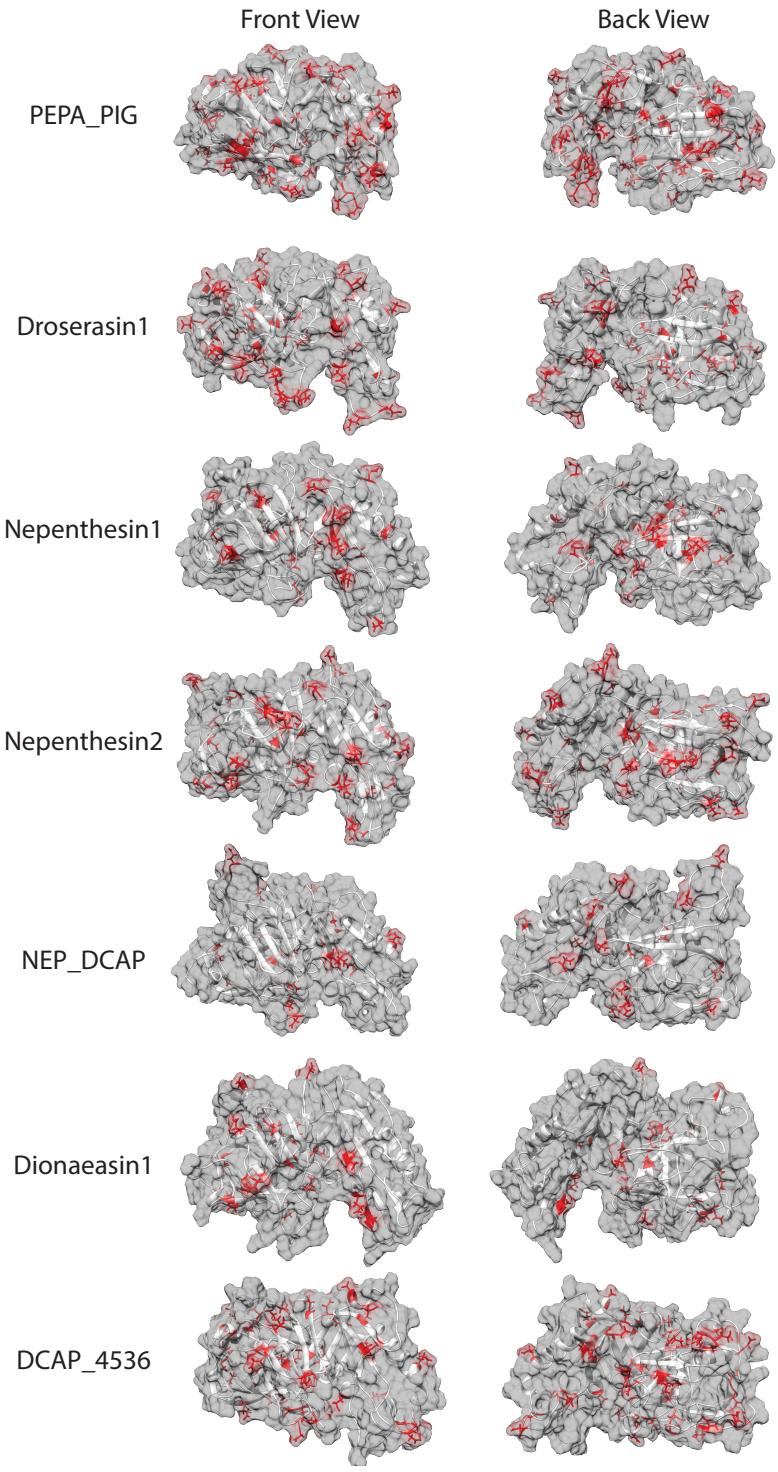


Figure 5.9: Comparison of the negative surface residues of pig pepsin versus the nepenthesins and a representative droserasin.

5.4 Conclusion

In summary, seven aspartic proteases were identified directly from the genomic DNA of *Drosera capensis*, and sorted into clusters based on sequence identity to known plant aspartic proteases, including homologs from *Dionaea muscipula* and *Nepenthes gracilis*. Molecular modeling and network analysis indicate that these proteases have distinct structural properties suggesting potential diversity in functional characteristics (e.g., thermal stability, substrate affinity). These diverse properties make this class of proteins an attractive target for further characterization studies, with rich potential for biotechnology applications. We have already begun the characterization of the PSI from Droserasin 1, and I think going forward it would be interesting to characterize Droserasin 1 alongside its PSI to further investigate their joint purposes and if there is further interaction between them in the mucilage after the PSI has been cleaved off. As for the nepenthesins, I would like to characterize NEP_DCAP, not only is it from *D. capensis*, but it has one of the least dense distributions of negative residues. This is promising for its stability based on the literature. It is also closer to Nepenthesins 1 and 2 based on the PCA, and can presumed to have a similar functionality, but it will be interesting to see the differences since it has fewer negative residues than either Nepenthesin.

Chapter 6

Conclusion

This thesis work covers a very wide cross-section of science. The triple-resonance SAS probe was designed to enable new types of protein experiments to be run, particularly with the aim of solving the structures of membrane proteins embedded within a membrane mimetic or native membrane. The bioinformatics project for high-throughput screening of protein sequences allows for the faster and more informed selection of targets for structure determination and biochemical characterization. The tie between these projects, and the future direction of this line of research is to use this SAS probe to solve the structure of a membrane protein chosen from the bioinformatics pipeline. This next step is just beginning with the initial characterization of the Droserasin 1 PSI by other members of the Martin Lab.

Bibliography

[1] R. W. Martin, J. E. Kelly, K. A. Collier, Spatial Reorientation Experiments for NMR of Solids and Partially Oriented Liquids, Progress in Nuclear Magnetic Resonance Spectroscopy 90-91 (2015) 92–122.

[2] P. J. Grandinetti, J. H. Baltisberger, A. Llor, Y. K. Lee, U. Werner, M. A. Eastman, A. Pines, Pure-Absorption-Mode Lineshapes and Sensitivity in Two-Dimensional Dynamic-Angle Spinning NMR, Journal of Magnetic Resonance, Series A 103 (1) (1993) 72–81.

[3] S. L. Gann, J. H. Baltisberger, A. Pines, Dynamic-angle spinning without sidebands, Chemical Physics Letters 210 (4) (1993) 405–410.

[4] J. Klinowski, Spin-off for solid-state studies, Nature 346 (6284) (1990) 509–510.

[5] P. J. Grandinetti, Dynamic-Angle Spinning, in: The Encyclopedia of NMR, Wiley, 1995, pp. 1768–1776.

[6] D. B. Zax, Dynamic angle spinning and double rotation quadrupolar NMR, TrAC Trends in Analytical Chemistry 13 (8) (1994) 300–312.

[7] J. P. Amoureux, NMR Characterization of Half-integer Quadrupolar Nuclei in Solids, Zeitschrift für Naturforschung A 47 (5) (1992) 665–674.

[8] A. P. M. Kentgens, A practical guide to solid-state NMR of half-integer quadrupolar nuclei with some applications to disordered systems, Geoderma 80 (3) (1997) 271–306.

[9] A. Jerschow, From nuclear structure to the quadrupolar NMR interaction and high-resolution spectroscopy, Progress in Nuclear Magnetic Resonance Spectroscopy 46 (1) (2005) 63–78.

[10] S. E. Ashbrook, M. J. Duer, Structural information from quadrupolar nuclei in solid state NMR, Concepts in Magnetic Resonance Part A 28A (3) (2006) 183–248.

[11] J. P. Amoureux, J. Trébosc, L. Delevoye, O. Lafon, B. Hu, Q. Wang, Correlation NMR spectroscopy involving quadrupolar nuclei, Solid State Nuclear Magnetic Resonance 35 (1) (2009) 12–18.

[12] C. Fernandez, M. Pruski, Probing Quadrupolar Nuclei by Solid-State NMR Spectroscopy: Recent Advances, in: Solid State NMR, Topics in Current Chemistry, Springer, Berlin, Heidelberg, 2011, pp. 119–188.

[13] B. Deloche, E. T. Samulski, Short-range nematic-like orientational order in strained elastomers: A deuterium magnetic resonance study, Macromolecules 14 (3) (1981) 575–581.

[14] H.-J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes, Journal of Biomolecular NMR 18 (4) (2000) 303–309.

[15] R. Tycko, F. J. Blanco, Y. Ishii, Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR, Journal of the American Chemical Society 122 (38) (2000) 9340–9341.

[16] Y. Liu, J. H. Prestegard, A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins, Journal of Biomolecular NMR 47 (4) (2010) 249–258.

[17] A. Bax, A. Grishaev, Weak alignment NMR: A hawk-eyed view of biomolecular structure, Current Opinion in Structural Biology 15 (5) (2005) 563–570.

[18] A. Naito, Structure elucidation of membrane-associated peptides and proteins in oriented bilayers by solid-state NMR spectroscopy, Solid State Nuclear Magnetic Resonance 36 (2) (2009) 67–76.

[19] P. Thureau, A. C. Sauerwein, M. Concistrè, M. H. Levitt, Selective internuclear coupling estimation in the solid-state NMR of multiple-spin systems, Physical Chemistry Chemical Physics 13 (1) (2011) 93–96.

[20] A. Lange, S. Luca, M. Baldus, Structural Constraints from Proton-Mediated Rare-Spin Correlation Spectroscopy in Rotating Solids, Journal of the American Chemical Society 124 (33) (2002) 9704–9705.

[21] T. Manolikas, T. Herrmann, B. H. Meier, Protein Structure Determination from 13C Spin-Diffusion Solid-State NMR Spectroscopy, Journal of the American Chemical Society 130 (12) (2008) 3959–3966.

[22] H. Saitô, I. Ando, A. Ramamoorthy, Chemical shift tensor – The heart of NMR: Insights into biological aspects of proteins, Progress in Nuclear Magnetic Resonance Spectroscopy 57 (2) (2010) 181–228.

[23] J. Herzfeld, A. E. Berger, Sideband intensities in NMR spectra of samples spinning at the magic angle, The Journal of Chemical Physics 73 (12) (1980) 6021–6030.

[24] P. Hodgkinson, L. Emsley, The reliability of the determination of tensor parameters by solid-state nuclear magnetic resonance, The Journal of Chemical Physics 107 (13) (1997) 4808–4816.

[25] W. T. Dixon, Spinning-sideband-free NMR spectra, Journal of Magnetic Resonance (1969) 44 (1) (1981) 220–223.

[26] E. T. Olejniczak, S. Vega, R. G. Griffin, Multiple pulse NMR in rotating solids, The Journal of Chemical Physics 81 (11) (1984) 4804–4817.

[27] D. P. Raleigh, E. T. Olejniczak, S. Vega, R. G. Griffin, An analysis of sideband suppression techniques in magic-angle sample spinning NMR, Journal of Magnetic Resonance (1969) 72 (2) (1987) 238–250.

[28] N. C. Nielsen, H. Bildsoe, H. J. Jakobsen, TOSS in high-speed MAS NMR, Journal of Magnetic Resonance (1969) 80 (1) (1988) 149–154.

[29] W. T. Dixon, Spinning-sideband-free and spinning-sideband-only NMR spectra in spinning samples, The Journal of Chemical Physics 77 (4) (1982) 1800–1809.

[30] R. Tycko, G. Dabbagh, P. A. Mirau, Determination of chemical-shift-anisotropy Lineshapes in a two-dimensional magic-angle-spinning NMR experiment, Journal of Magnetic Resonance (1969) 85 (2) (1989) 265–274.

[31] S.-F. Liu, J.-D. Mao, K. Schmidt-Rohr, A Robust Technique for Two-Dimensional Separation of Undistorted Chemical-Shift Anisotropy Powder Patterns in Magic-Angle-Spinning NMR, Journal of Magnetic Resonance 155 (1) (2002) 15–28.

[32] R. M. Orr, M. J. Duer, Recoupling of chemical-shift anisotropy powder patterns in MAS NMR, Journal of Magnetic Resonance 181 (1) (2006) 1–8.

[33] L. Shao, C. Crockford, H. Geen, G. Grasso, J. J. Titman, Chemical shift anisotropy amplification, Journal of Magnetic Resonance 167 (1) (2004) 75–86.

[34] M. Carravetta, M. Edén, X. Zhao, A. Brinkmann, M. H. Levitt, Symmetry principles for the design of radiofrequency pulse sequences in the nuclear magnetic resonance of rotating solids, Chemical Physics Letters 321 (3) (2000) 205–215.

[35] G. Hou, I.-J. L. Byeon, J. Ahn, A. M. Gronenborn, T. Polenova, Recoupling of chemical shift anisotropy by R-symmetry sequences in magic angle spinning NMR spectroscopy, The Journal of Chemical Physics 137 (13) (2012) 134201.

[36] M. K. Pandey, M. Malon, A. Ramamoorthy, Y. Nishiyama, Composite-180° pulse-based symmetry sequences to recouple proton chemical shift anisotropy tensors under ultrafast MAS solid-state NMR spectroscopy, Journal of Magnetic Resonance 250 (2015) 45–54.

[37] G. Hou, S. Paramasivam, I.-J. L. Byeon, A. M. Gronenborn, T. Polenova, Determination of relative tensor orientations by γ -encoded chemical shift anisotropy/heteronuclear dipolar coupling 3D NMR spectroscopy in biological solids, Physical Chemistry Chemical Physics 12 (45) (2010) 14873–14883.

[38] J. C. C. Chan, R. Tycko, Recoupling of chemical shift anisotropies in solid-state NMR under high-speed magic-angle spinning and in uniformly 13C-labeled systems, The Journal of Chemical Physics 118 (18) (2003) 8378–8389.

[39] Y. Nishiyama, T. Yamazaki, T. Terao, Development of modulated rf sequences for decoupling and recoupling of nuclear-spin interactions in sample-spinning solid-state NMR: Application to chemical-shift anisotropy determination, The Journal of Chemical Physics 124 (6) (2006) 064304.

[40] B. J. Wylie, C. D. Schwieters, E. Oldfield, C. M. Rienstra, Protein Structure Refinement Using 13Cα Chemical Shift Tensors, Journal of the American Chemical Society 131 (3) (2009) 985–992.

[41] B. J. Wylie, L. J. Sperling, A. J. Nieuwkoop, W. T. Franks, E. Oldfield, C. M. Rienstra, Ultrahigh resolution protein structures using NMR chemical shift tensors, Proceedings of the National Academy of Sciences 108 (41) (2011) 16974–16979.

[42] Y. Tian, G. J. Lu, F. M. Marassi, S. J. Opella, Structure of the membrane protein MerF, a bacterial mercury transporter, improved by the inclusion of chemical shift anisotropy constraints, Journal of Biomolecular NMR 60 (1) (2014) 67–71.

[43] N. K. Sethi, D. M. Grant, R. J. Pugmire, 13C chemical shielding anisotropy studied by variable-angle sample spinning, Journal of Magnetic Resonance (1969) 71 (3) (1987) 476–479.

[44] E. O. Stejskal, J. Schaefer, R. A. McKay, High-resolution, slow-spinning magic-angle carbon-13 NMR, Journal of Magnetic Resonance (1969) 25 (3) (1977) 569–573.

[45] T. Wang, M. Hong, Investigation of the Curvature Induction and Membrane Localization of the Influenza Virus M2 Protein Using Static and Off-Magic-Angle Spinning Solid-State Nuclear Magnetic Resonance of Oriented Bicelles, Biochemistry 54 (13) (2015) 2214–2226.

[46] G. R. Marshall, D. D. Beusen, K. Kociolek, A. S. Redlinski, M. T. Leplawy, Y. Pan, J. Schaefer, Determination of a precise interatomic distance in a helical peptide by REDOR NMR, Journal of the American Chemical Society 112 (3) (1990) 963–966.

[47] T. G. Oas, R. G. Griffin, M. H. Levitt, Rotary resonance recoupling of dipolar interactions in solid-state nuclear magnetic resonance spectroscopy, The Journal of Chemical Physics 89 (2) (1988) 692–695.

[48] M. H. Levitt, Symmetry-Based Pulse Sequences in Magic-Angle Spinning Solid-State NMR, in: eMagRes, American Cancer Society, 2007.

[49] T. Gullion, R. A. McKay, A. Schmidt, Spin-echo, double-resonance NMR with flipped spinning (SEDORFS), Journal of Magnetic Resonance (1969) 94 (2) (1991) 362–369.

[50] P. D. Murphy, B. C. Gerstein, Principal components of the cadmium-113 shielding tensors in cadmium sulfate hydrates: A nuclear magnetic resonance study of cadmium coordinated with oxygen, Journal of the American Chemical Society 103 (12) (1981) 3282–3286.

[51] A. C. Kolbert, P. J. Grandinetti, M. Baldwin, S. B. Prusiner, A. Pines, Measurement of Internuclear Distances by Switched Angle Spinning NMR, The Journal of Physical Chemistry 98 (33) (1994) 7936–7938.

[52] R. Tycko, Normal Angle Spinning Dipolar Spectroscopy for Structural Studies by Solid-State Nuclear Magnetic Resonance, Journal of the American Chemical Society 116 (5) (1994) 2217–2218.

[53] A. Bax, N. M. Szeverenyi, G. E. Maciel, Chemical shift anisotropy in powdered solids studied by 2D FT NMR with flipping of the spinning axis, Journal of Magnetic Resonance (1969) 55 (3) (1983) 494–497.

[54] I. M. Litvak, C. A. Espinosa, R. A. Shapiro, A. N. Oldham, V. V. Duong, R. W. Martin, Pneumatic switched angle spinning NMR probe with capacitively coupled double saddle coil, Journal of Magnetic Resonance 206 (2) (2010) 183–189.

[55] D. I. Hoult, R. E. Richards, The signal-to-noise ratio of the nuclear magnetic resonance experiment, Journal of Magnetic Resonance (1969) 24 (1) (1976) 71–85.

[56] C. Qian, P. Thureau, R. W. Martin, Variable angle spinning (VAS) experiments for strongly oriented systems: Methods development and preliminary results, Magnetic Resonance in Chemistry 46 (4) (2008) 351–355.

[57] A. Medek, J. R. Sachleben, P. Beverwyk, L. Frydman, Multi-rank nuclear magnetic resonance studies of half-integer quadrupolar nuclei in solids by three-dimensional dynamic-angle correlation spectroscopy, The Journal of Chemical Physics 104 (14) (1996) 5374–5383.

[58] S. Idziak, U. Haeberlen, Design and construction of a high homogeneity rf coil for solid-state multiple-pulse NMR, Journal of Magnetic Resonance (1969) 50 (2) (1982) 281–288.

[59] Y. H. Sun, G. E. Maciel, The Tilted Coil for NMR Experiments, Journal of Magnetic Resonance, Series A 105 (2) (1993) 145–150.

[60] J. Alonso, A. Soleilhavoup, A. Wong, A. Guiga, D. Sakellariou, Double helix dipole design applied to magnetic resonance: A novel NMR coil, Journal of Magnetic Resonance 235 (2013) 32–41.

[61] J. A. Stringer, C. E. Bronnimann, C. G. Mullen, D. H. Zhou, S. A. Stellfox, Y. Li, E. H. Williams, C. M. Rienstra, Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes, Journal of Magnetic Resonance 173 (1) (2005) 40–48.

[62] B. Dillmann, K. Elbayed, H. Zeiger, M.-C. Weingertner, M. Piotto, F. Engelke, A novel low-E field coil to minimize heating of biological samples in solid-state multinuclear NMR experiments, Journal of Magnetic Resonance 187 (1) (2007) 10–18.

[63] T. G. Ajithkumar, E. R. H. van Eck, A. P. M. Kentgens, Homonuclear correlation experiments for quadrupolar nuclei, spinning away from the magic angle, Solid State Nuclear Magnetic Resonance 26 (3) (2004) 180–186.

[64] D. W. Alderman, D. M. Grant, An efficient decoupler coil design which reduces heating in conductive samples in superconducting spectrometers, Journal of Magnetic Resonance (1969) 36 (3) (1979) 447–451.

[65] C. V. Grant, Y. Yang, M. Glibowicka, C. H. Wu, S. H. Park, C. M. Deber, S. J. Opella, A Modified Alderman–Grant Coil makes possible an efficient cross-coil probe for high field solid-state NMR of lossy biological samples, Journal of Magnetic Resonance 201 (1) (2009) 87–92.

[66] S. A. McNeill, P. L. Gor'kov, K. Shetty, W. W. Brey, J. R. Long, A low-E magic angle spinning probe for biological solid state NMR at 750 MHz, Journal of Magnetic Resonance 197 (2) (2009) 135–144.

[67] F. D. Doty, J. Kulkarni, C. Turner, G. Entzminger, A. Bielecki, Using a cross-coil to reduce RF heating by an order of magnitude in triple-resonance multinuclear MAS at high fields, Journal of Magnetic Resonance 182 (2) (2006) 239–253.

[68] B. C. Gerstein, H.-J. Pan, M. Pruski, Stable system for multiple pulse, magic angle spinning and dynamic angle spinning NMR, Review of Scientific Instruments 61 (10) (1990) 2699–2701.

[69] K. T. Mueller, G. C. Chingas, A. Pines, NMR probe for dynamic-angle spinning, Review of Scientific Instruments 62 (6) (1991) 1445–1452.

[70] F. D. Doty, G. Entzminger, Y. A. Yang, Magnetism in high-resolution NMR probe design. II: HR MAS, Concepts in Magnetic Resonance 10 (4) (1998) 239–260.

[71] M. A. Eastman, P. J. Grandinetti, Y. K. Lee, A. Pines, Double-tuned hopping-coil probe for dynamic-angle-spinning NMR, Journal of Magnetic Resonance (1969) 98 (2) (1992) 333–341.

[72] M. C. Davis, D. C. Kaseman, S. M. Parvani, K. J. Sanders, P. J. Grandinetti, D. Massiot, P. Florian, Q(n) Species Distribution in K2O·2SiO2 Glass by 29Si Magic Angle Flipping NMR, The Journal of Physical Chemistry A 114 (17) (2010) 5503–5508.

[73] Novabraid, Spectra Fiber for Superior Cut-Resistant Rope from Novabraid, http://www.novabraid.com/rope-material/spectra-fiber-rope/ (2018).

[74] K. T. Mueller, B. Q. Sun, G. C. Chingas, J. W. Zwanziger, T. Terao, A. Pines, Dynamic-angle spinning of quadrupolar nuclei, Journal of Magnetic Resonance (1969) 86 (3) (1990) 470–487.

[75] F. D. Doty, High power flexible leads for DAS NMR (Mar. 2001).

[76] R. G. Larsen, Y. K. Lee, B. He, J. O. Yang, Z. Luz, H. Zimmermann, A. Pines, Carbon-13 chemical shift tensor correlation via spin diffusion in solid tropolone using switched-angle spinning spectroscopy, The Journal of Chemical Physics 103 (22) (1995) 9844–9854.

[77] F. D. Doty, P. D. Ellis, Design of high speed cylindrical NMR sample spinners, Review of Scientific Instruments 52 (12) (1981) 1868–1875.

[78] M. C. Davis, K. J. Sanders, P. J. Grandinetti, S. J. Gaudio, S. Sen, Structural investigations of magnesium silicate glasses by 29Si 2D Magic-Angle Flipping NMR, Journal of Non-Crystalline Solids 357 (15) (2011) 2787–2795.

[79] R. A. McKay, Transmission Line Probes, Vol. 6 of Encyclopedia of Nuclear Magnetic Resonance, Wiley, Chichester, 1996, pp. 3768 – 3771.

[80] R. W. Martin, E. K. Paulson, K. W. Zilm, Design of a triple resonance magic angle sample spinning probe for high field solid state nuclear magnetic resonance, Review of Scientific Instruments 74 (6) (2003) 3045–3061.

[81] E. Mihaliuk, A switched angle sample spinning probe and its application to solid state dipolar recoupling experiments, 2009.

[82] L. Chopin, R. Rosanske, T. Gullion, Simple Improvements in Spinning-Speed Control for MAS NMR Experiments, Journal of Magnetic Resonance, Series A 122 (2) (1996) 237–239.

[83] J. N. Lee, D. W. Alderman, J. Y. Jin, K. W. Zilm, C. L. Mayne, R. J. Pugmire, D. M. Grant, Cylindrical spinner and speed controller for magic angle spinning nuclear magnetic resonance, Review of Scientific Instruments 55 (4) (1984) 516–520.

[84] E. Hughes, T. Gullion, A simple, inexpensive, and precise magic angle spinning speed controller, Solid State Nuclear Magnetic Resonance 26 (1) (2004) 16–21.

[85] E. Mihaliuk, T. Gullion, Using heat to control the sample spinning speed in MAS NMR, Journal of Magnetic Resonance 212 (2) (2011) 249–253.

[86] R. A. Wind, F. E. Anthonio, M. J. Duijvestijn, J. Smidt, J. Trommel, G. M. C. de Vette, Experimental setup for enhanced 13C NMR spectroscopy in solids using dynamic nuclear polarization, Journal of Magnetic Resonance (1969) 52 (3) (1983) 424–434.

[87] S. F. Dec, R. A. Wind, G. E. Maciel, F. E. Anthonio, High-speed magic-angle spinning, Journal of Magnetic Resonance (1969) 70 (2) (1986) 355–359.

[88] J. Courtieu, D. W. Alderman, D. M. Grant, Spinning near the magic angle: A means of obtaining first-order dipolar NMR spectra of molecules dissolved in nematic liquid crystals, Journal of the American Chemical Society 103 (22) (1981) 6783–6784.

[89] J. Courtieu, D. W. Alderman, D. M. Grant, J. P. Bayles, Director dynamics and NMR applications of nematic liquid crystals spinning at various angles from the magnetic field, The Journal of Chemical Physics 77 (2) (1982) 723–730.

[90] T. Väänänen, J. Jokisaari, M. Seläntaus, A variable-angle spinning system for the determination of NMR parameters of liquid-crystalline samples, Journal of Magnetic Resonance (1969) 72 (3) (1987) 414–421.

[91] M. Tomaselli, B. H. Meier, M. Baldus, J. Eisenegger, R. R. Ernst, An rf-driven nuclear spin-diffusion experiment using zero-angle sample spinning, Chemical Physics Letters 225 (1) (1994) 131–139.

[92] C. Qian, A. Pines, R. W. Martin, Design and construction of a contactless mobile RF coil for double resonance variable angle spinning NMR, Journal of Magnetic Resonance 188 (1) (2007) 183–189.

[93] J. Courtieu, J. P. Bayle, B. M. Fung, Variable angle sample spinning NMR in liquid crystals, Progress in Nuclear Magnetic Resonance Spectroscopy 26 (1994) 141–169.

[94] L. Beguin, J. Courtieu, L. Ziani, D. Merlet, Simplification of the 1H NMR spectra of enantiomers dissolved in chiral liquid crystals, combining variable angle sample spinning and selective refocusing experiments, Magnetic Resonance in Chemistry 44 (12) (2006) 1096–1101.

[95] E. Mihaliuk, T. Gullion, Optical lever for monitoring of the magic angle, Journal of Magnetic Resonance 223 (2012) 46–50.

[96] G. Zandomeneghi, M. Tomaselli, J. D. van Beek, B. H. Meier, Manipulation of the Director in Bicellar Mesophases by Sample Spinning: A New Tool for NMR Spectroscopy, Journal of the American Chemical Society 123 (5) (2001) 910–913.

[97] F. D. Doty, Solid State NMR Probe Design, in: R. K. Harris, R. Wasylishen (Eds.), Encyclopedia of Magnetic Resonance, John Wiley & Sons, Ltd, Chichester, UK, 2007.

[98] P. L. Gor'kov, R. Witter, E. Y. Chekmenev, F. Nozirov, R. Fu, W. W. Brey, Low-E probe for 19F–1H NMR of dilute biological solids, Journal of Magnetic Resonance 189 (2) (2007) 182–189.

[99] P. L. Gor'kov, E. Y. Chekmenev, C. Li, M. Cotten, J. J. Buffy, N. J. Traaseth, G. Veglia, W. W. Brey, Using low-E resonators to reduce RF heating in biological samples for static solid-state NMR up to 900MHz, Journal of Magnetic Resonance 185 (1) (2007) 77–93.

[100] E. K. Paulson, R. W. Martin, K. W. Zilm, Cross polarization, radio frequency field homogeneity, and circuit balancing in high field solid state NMR probes, Journal of Magnetic Resonance 171 (2) (2004) 314–323.

[101] F. D. Doty, G. Entzminger, Y. A. Yang, Magnetism in high-resolution NMR probe design. I: General methods, Concepts in Magnetic Resonance 10 (3) (1998) 133–156.

[102] S. Mamone, A. Dorsch, O. G. Johannessen, M. V. Naik, P. K. Madhu, M. H. Levitt, A Hall effect angle detector for solid-state NMR, Journal of Magnetic Resonance 190 (1) (2008) 135–141.

[103] J. S. Frye, G. E. Maciel, Setting the magic angle using a quadrupolar nuclide, Journal of Magnetic Resonance (1969) 48 (1) (1982) 125–131.

[104] S. Idziak, U. Haeberlen, Design and construction of a high homogeneity rf coil for solid-state multiple-pulse NMR, Journal of Magnetic Resonance (1969) 50 (2) (1982) 281–288.

[105] Z. Zhang, P. C. Hammel, G. J. Moore, Application of a novel rf coil design to the magnetic resonance force microscope, Review of Scientific Instruments 67 (9) (1996) 3307–3309.

[106] D. Sakellariou, G. L. Goff, J.-F. Jacquinot, High-resolution, high-sensitivity NMR of nanolitre anisotropic samples by coil spinning, Nature 447 (7145) (2007) 694–697.

[107] A. Fiser, Comparative Protein Structure Modelling, in: From Protein Structure to Function with Bioinformatics, Springer, Dordrecht, 2017, pp. 91–134.

[108] N. C. Benson, V. Daggett, A chemical group graph representation for efficient high-throughput analysis of atomistic protein simulations, Journal of Bioinformatics and Computational Biology 10 (04) (2012) 1250008.

[109] C. T. Butts, X. Zhang, J. E. Kelly, K. W. Roskamp, M. H. Unhelkar, J. A. Freites, S. Tahir, R. W. Martin, Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, Drosera capensis, Computational and Structural Biotechnology Journal 14 (2016) 271–282.

[110] M. H. Unhelkar, V. T. Duong, K. N. Enendu, J. E. Kelly, S. Tahir, C. T. Butts, R. W. Martin, Structure prediction and network analysis of chitinases from the Cape sundew, Drosera capensis, Biochimica et Biophysica Acta (BBA) - General Subjects 1861 (3) (2017) 636–643.

[111] V. T. Duong, M. H. Unhelkar, J. E. Kelly, S. H. Kim, C. T. Butts, R. W. Martin, Protein Structure networks provide insight into the active site flexibility in esterase/lipases from the carnivorous plant \textit{Drosera capensis}, submitted.

[112] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Molecular Systems Biology 7 (1) (2011) 539.

[113] T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions, Nature Methods 8 (10) (2011) 785–786.

[114] C. T. Butts, J. C. Bierma, R. W. Martin, Novel proteases from the genome of the carnivorous plant Drosera capensis: Structural prediction and comparative analysis, Proteins: Structure, Function, and Bioinformatics 84 (10) (2016) 1517–1533.

[115] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, The Journal of Physical Chemistry B 102 (18) (1998) 3586–3616.

[116] J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton, Journal of Molecular Biology 285 (4) (1999) 1735–1747.

[117] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.

[118] J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein, Journal of Molecular Biology 157 (1) (1982) 105–132.

[119] R. Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2017).

[120] G. Smith, A. Macias-Muñoz, A. D. Briscoe, Gene Duplication and Gene Expression Changes Play a Role in the Evolution of Candidate Pollen Feeding Genes in Heliconius Butterflies, Genome Biology and Evolution 8 (8) (2016) 2581–2596.

[121] D. Harpel, D. A. Cullen, S. R. Ott, C. D. Jiggins, J. R. Walters, Pollen feeding proteomics: Salivary proteins of the passion flower butterfly, Heliconius melpomene, Insect Biochemistry and Molecular Biology 63 (2015) 7–13.

[122] L. E. Gilbert, Pollen Feeding and Reproductive Biology of Heliconius Butterflies, Proceedings of the National Academy of Sciences 69 (6) (1972) 1403–1407.

[123] M. Z. Cardoso, Patterns of pollen collection and flower visitation by Heliconius butterflies in southeastern Mexico, Journal of Tropical Ecology 17 (5) (2001) 763–768.

[124] P. Rodbumrer, D. Arthan, U. Uyen, J. Yuvaniyama, J. Svasti, P. Y. Wongsaengchantra, Functional expression of a Bombyx mori cocoonase: Potential application for silk degumming, Acta Biochimica et Biophysica Sinica 44 (12) (2012) 974–983.

[125] F. C. Kafatos, J. H. Law, A. M. Tartakoff, Cocoonase II. SUBSTRATE SPECIFICITY, INHIBITORS, AND CLASSIFICATION OF THE ENZYME, Journal of Biological Chemistry 242 (7) (1967) 1488–1494.

[126] P. Geng, L. Lin, Y. Li, Q. Fan, N. Wang, L. Song, W. Li, A novel fibrin(ogen)olytic trypsin-like protease from Chinese oak silkworm (Antheraea pernyi): Purification and characterization, Biochemical and Biophysical Research Communications 445 (1) (2014) 64–70.

[127] G. Smith, J. E. Kelly, A. Macias-Muñoz, C. T. Butts, R. W. Martin, A. D. Briscoe, Evolutionary and structural analyses uncover a role for solvent interactions in the diversification of cocoonases in butterflies, Proc. R. Soc. B 285 (1870) (2018) 20172037.

[128] L. Hedstrom, L. Szilagyi, W. J. Rutter, Converting trypsin to chymotrypsin: The role of surface loops, Science 255 (5049) (1992) 1249–1253.

[129] T. Kurth, D. Ullmann, H.-D. Jakubke, L. Hedstrom, Converting Trypsin to Chymotrypsin: Structural Determinants of S1‘ Specificity, Biochemistry 36 (33) (1997) 10098–10104.

[130] W. Ma, C. Tang, L. Lai, Specificity of Trypsin and Chymotrypsin: Loop-Motion-Controlled Dynamic Correlation as a Determinant, Biophysical Journal 89 (2) (2005) 1183–1193.

[131] J. J. Perona, C. S. Craik, Structural basis of substrate specificity in the serine proteases, Protein Science 4 (3) (1995) 337–360.

[132] X. Zhang, T. Perica, S. A. Teichmann, Evolution of protein structures and interactions from the perspective of residue contact networks, Current Opinion in Structural Biology 23 (6) (2013) 954–963.

[133] M. M. Krem, E. D. Cera, Molecular markers of serine protease evolution, The EMBO Journal 20 (12) (2001) 3036–3045.

[134] V. G. H. Eijsink, S. Gåseidnes, T. V. Borchert, B. van den Burg, Directed evolution of enzyme stability, Biomolecular Engineering 22 (1) (2005) 21–30.

[135] H. Ogino, T. Uchiho, N. Doukyu, M. Yasuda, K. Ishimi, H. Ishikawa, Effect of exchange of amino acid residues of the surface region of the PST-01 protease on its organic solvent-stability, Biochemical and Biophysical Research Communications 358 (4) (2007) 1028–1033.

[136] A. Silvério, J. E. de Araujo Mariath, Comparative structure of the pollen in species of Passiflora: Insights from the pollen wall and cytoplasm contents, Plant Systematics and Evolution 300 (2) (2014) 347–358.

[137] G. J. Székely, M. L. Rizzo, Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method, Journal of Classification 22 (2) (2005) 151–183.

[138] P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics 16 (6) (2000) 276–277.

[139] W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso, R. Lopez, The EMBL-EBI bioinformatics web and programmatic tools framework, Nucleic Acids Research 43 (W1) (2015) W580–W584.

[140] I. Simões, C. Faro, Structure and function of plant aspartic proteinases, European Journal of Biochemistry 271 (11) (2004) 2067–2075.

[141] A. Schaller, A cut above the rest: The regulatory function of plant proteases, Planta 220 (2) (2004) 183–197.

[142] X. Yao, W. Xiong, T. Ye, Y. Wu, Overexpression of the aspartic protease ASPG1 gene confers drought avoidance in Arabidopsis, Journal of Experimental Botany 63 (7) (2012) 2579–2593.

[143] M. García-Lorenzo, A. Sjödin, S. Jansson, C. Funk, Protease gene families in Populus and Arabidopsis, BMC Plant Biology 6 (1) (2006) 30.

[144] M. W. Risør, L. R. Thomsen, K. W. Sanggaard, T. A. Nielsen, I. B. Thøgersen, M. V. Lukassen, L. Rossen, I. Garcia-Ferrer, T. Guevara, C. Scavenius, E. Meinjohanns, F. X. Gomis-Rüth, J. J. Enghild, Enzymatic and Structural Characterization of the Major Endopeptidase in the Venus Flytrap Digestion Fluid, Journal of Biological Chemistry 291 (5) (2016) 2271–2287.

[145] Y. Takeuchi, M. M. Salcher, M. Ushio, R. Shimizu-Inatsugi, M. J. Kobayashi, B. Diway, C. von Mering, J. Pernthaler, K. K. Shimizu, In Situ Enzyme Activity in the Dissolved and Particulate Fraction of the Fluid from Four Pitcher Plant Species of the Genus Nepenthes, PLOS ONE 6 (9) (2011) e25144.

[146] W. Adlassnig, M. Peroutka, T. Lendl, Traps of carnivorous pitcher plants as a habitat: Composition of the fluid, biodiversity and mutualistic activities, Annals of Botany 107 (2) (2011) 181–194.

[147] K. Rost, R. Schauer, Physical and chemical properties of the mucin secreted by Drosera capensis, Phytochemistry 16 (9) (1977) 1365–1368.

[148] E. Gout, R. Bligny, R. Douce, Regulation of intracellular pH values in higher plant cells. Carbon-13 and phosphorus-31 nuclear magnetic resonance studies., Journal of Biological Chemistry 267 (20) (1992) 13903–13909.

[149] N. D. Rawlings, M. Waller, A. J. Barrett, A. Bateman, MEROPS: The database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Research 42 (D1) (2014) D503–D509.

[150] P. Castanheira, B. Samyn, K. Sergeant, J. C. Clemente, B. M. Dunn, E. Pires, J. V. Beeumen, C. Faro, Activation, Proteolytic Processing, and Peptide Specificity of Recombinant Cardosin A, Journal of Biological Chemistry 280 (13) (2005) 13047–13054.

[151] P. J. M. Reis, F. X. Malcata, Current state of Portuguese dairy products from ovine and caprine milks, Small Ruminant Research 101 (1) (2011) 122–133.

[152] S. V. Silva, F. X. Malcata, Proteolysis of ovine caseins by cardosin A, an aspartic acid proteinase from Cynara cardunculus L., Le Lait 78 (5) (1998) 513–519.

[153] S. V. Silva, F. X. Malcata, On the activity and specificity of cardosin B, a plant proteinase, on ovine caseins, Food Chemistry 67 (4) (1999) 373–378.

[154] C.-I. An, E.-i. Fukusaki, A. Kobayashi, Aspartic proteinases are expressed in pitchers of the carnivorous plant Nepenthes alata Blanco, Planta 214 (5) (2002) 661–667.

[155] S. B. P. Athauda, K. Matsumoto, S. Rajapakshe, M. Kuribayashi, M. Kojima, N. Kubomura-Yoshida, A. Iwamatsu, C. Shibata, H. Inoue, K. Takahashi, Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases (2004) 12.

[156] K. Takahashi, S. B. P. Athauda, K. Matsumoto, S. Rajapakshe, M. Kuribayashi, M. Kojima, N. Kubomura-Yoshida, A. Iwamatsu, C. Shibata, H. Inoue, Nepenthesin, a Unique Member of a Novel Subfamily of Aspartic Proteinases: Enzymatic and Structural Characteristics, Current Protein and Peptide Science 6 (6) (2005) 513–525.

[157] A. Kadek, V. Tretyachenko, H. Mrazek, L. Ivanova, P. Halada, M. Rey, D. C. Schriemer, P. Man, Expression and characterization of plant aspartic protease nepenthesin-1 from Nepenthes gracilis, Protein Expression and Purification 95 (2014) 121–128.

[158] M. Rey, M. Yang, K. M. Burns, Y. Yu, S. P. Lees-Miller, D. C. Schriemer, Nepenthesin from Monkey Cups for Hydrogen/Deuterium Exchange Mass Spectrometry, Molecular & Cellular Proteomics 12 (2) (2013) 464–472.

[159] F. Buch, W. E. Kaman, F. J. Bikker, A. Yilamujiang, A. Mithöfer, Nepenthesin Protease Activity Indicates Digestive Fluid Dynamics in Carnivorous Nepenthes Plants, PLOS ONE 10 (3) (2015) e0118853.

[160] D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Research 32 (suppl_2) (2004) W526–W531.

[161] S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. DiMaio, O. Lange, L. Kinch, W. Sheffler, B.-H. Kim, R. Das, N. V. Grishin, D. Baker, Structure prediction for CASP8 with all-atom refinement using Rosetta, Proteins: Structure, Function, and Bioinformatics 77 (S9) (2009) 89–99.

[162] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, Journal of Computational Chemistry 26 (16) (2005) 1781–1802.

[163] A. D. Mackerell, M. Feig, C. L. Brooks, Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, Journal of Computational Chemistry 25 (11) (2004) 1400–1415.

[164] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, The Journal of Chemical Physics 79 (2) (1983) 926–935.

[165] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, Journal of Computational Chemistry 25 (13) (2004) 1605–1612.

[166] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, Journal of Molecular Graphics 14 (1) (1996) 33–38.

[167] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, M. Morris, Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data, Journal of statistical software 24 (1) (2008) 1548–7660.

[168] C. T. Butts, **Network** : A Package for Managing Relational Data in *R*, Journal of Statistical Software 24 (2).

[169] C. T. Butts, K. M. Carley, Some Simple Algorithms for Structural Comparison, Computational & Mathematical Organization Theory 11 (4) (2005) 291–305.

[170] C. T. Butts, Social Network Analysis with **sna**, Journal of Statistical Software 24 (6).

[171] M. Campbell, M. Law, C. Holt, J. Stein, G. Moghe, D. Hufnagel, J. Lei, R. Achawanantakun, D. Jiao, C. Lawrence, D. Ware, S.-H. Shiu, K. Childs, Y. Sun, N. Jiang, M. Yandell, MAKER-P: A tool-kit for the rapid creation, management, and quality control of plant genome annotations, Plant Physiology (2013) pp.113.230144.

[172] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: Protein domains identifier, Nucleic Acids Research 33 (suppl_2) (2005) W116–W120.

[173] M. A. Mazorra-Manzano, T. Tanaka, D. R. Dee, R. Y. Yada, Structure–function characterization of the recombinant aspartic proteinase A1 from Arabidopsis thaliana, Phytochemistry 71 (5) (2010) 515–523.

[174] C. Pereira, S. Pereira, B. Satiat-Jeunemaitre, J. Pissarra, Cardosin A contains two vacuolar sorting signals using different vacuolar routes in tobacco epidermal cells, The Plant Journal 76 (1) (2013) 87–100.

[175] P. Duarte, J. Pissarra, I. Moore, Processing and trafficking of a single isoform of the aspartic proteinase cardosin A on the vacuolar pathway, Planta 227 (6) (2008) 1255–1268.

[176] D. S. da Costa, S. Pereira, I. Moore, J. Pissarra, Dissecting cardosin B trafficking pathways in heterologous systems, Planta 232 (6) (2010) 1517–1530.

[177] D. S. da Costa, S. Pereira, J. Pissarra, The heterologous systems in the study of cardosin B trafficking pathways, Plant Signaling & Behavior 6 (6) (2011) 895–897.

[178] N. Hatano, T. Hamada, Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant Nepenthes alata, Journal of Proteomics 75 (15) (2012) 4844–4852.

[179] W. X. Schulze, K. W. Sanggaard, I. Kreuzer, A. D. Knudsen, F. Bemm, I. B. Thogersen, A. Brautigam, L. R. Thomsen, S. Schliesky, T. F. Dyrlund, M. Escalante-Perez, D. Becker, J. Schultz, H. Karring, A. Weber, P. Hojrup, R. Hedrich, J. J. Enghild, The protein composition of the digestive fluid from the Venus flytrap sheds light on prey digestion mechanisms, Molecular & Cellular Proteomics (2012) mcp.M112.021006.

[180] C. M. Almeida, D. Gomes, C. Faro, I. Simões, Engineering a cardosin B-derived rennet for sheep and goat cheese manufacture, Applied Microbiology and Biotechnology 99 (1) (2015) 269–281.

[181] W. E. Harte, D. L. Beveridge, Prediction of the protonation state of the active site aspartyl residues in HIV-1 protease-inhibitor complexes via molecular dynamics simulation, Journal of the American Chemical Society 115 (10) (1993) 3883–3886.

[182] T. D. McGee, J. Edwards, A. E. Roitberg, pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State, The Journal of Physical Chemistry B 118 (44) (2014) 12577–12585.

[183] D. B. Northrop, Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases, Accounts of Chemical Research 34 (10) (2001) 790–797.

[184] M. Cascella, C. Micheletti, U. Rothlisberger, P. Carloni, Evolutionarily Conserved Functional Mechanics across Pepsin-like and Retroviral Aspartic Proteases, Journal of the American Chemical Society 127 (11) (2005) 3734–3742.

[185] E. T. Baldwin, T. N. Bhat, S. Gulnik, M. V. Hosur, R. C. Sowder, R. E. Cachau, J. Collins, A. M. Silva, J. W. Erickson, Crystal structures of native and inhibited forms of human cathepsin D: Implications for lysosomal targeting and drug design, Proceedings of the National Academy of Sciences 90 (14) (1993) 6796–6800.

[186] K. Kubota, Y. Metoki, S. B. P. Athauda, C. Shibata, K. Takahashi, Stability Profiles of Nepenthesin in Urea and Guanidine Hydrochloride: Comparison with Porcine Pepsin A, Bioscience, Biotechnology, and Biochemistry 74 (11) (2010) 2323–2326.

[187] A. Kadek, H. Mrazek, P. Halada, M. Rey, D. C. Schriemer, P. Man, Aspartic Protease Nepenthesin-1 as a Tool for Digestion in Hydrogen/Deuterium Exchange Mass Spectrometry, Analytical Chemistry 86 (9) (2014) 4287–4294.

LIST OF FIGURES IN APPENDIX

Appendix A

CAD Drawings

All dimensions are in inches.

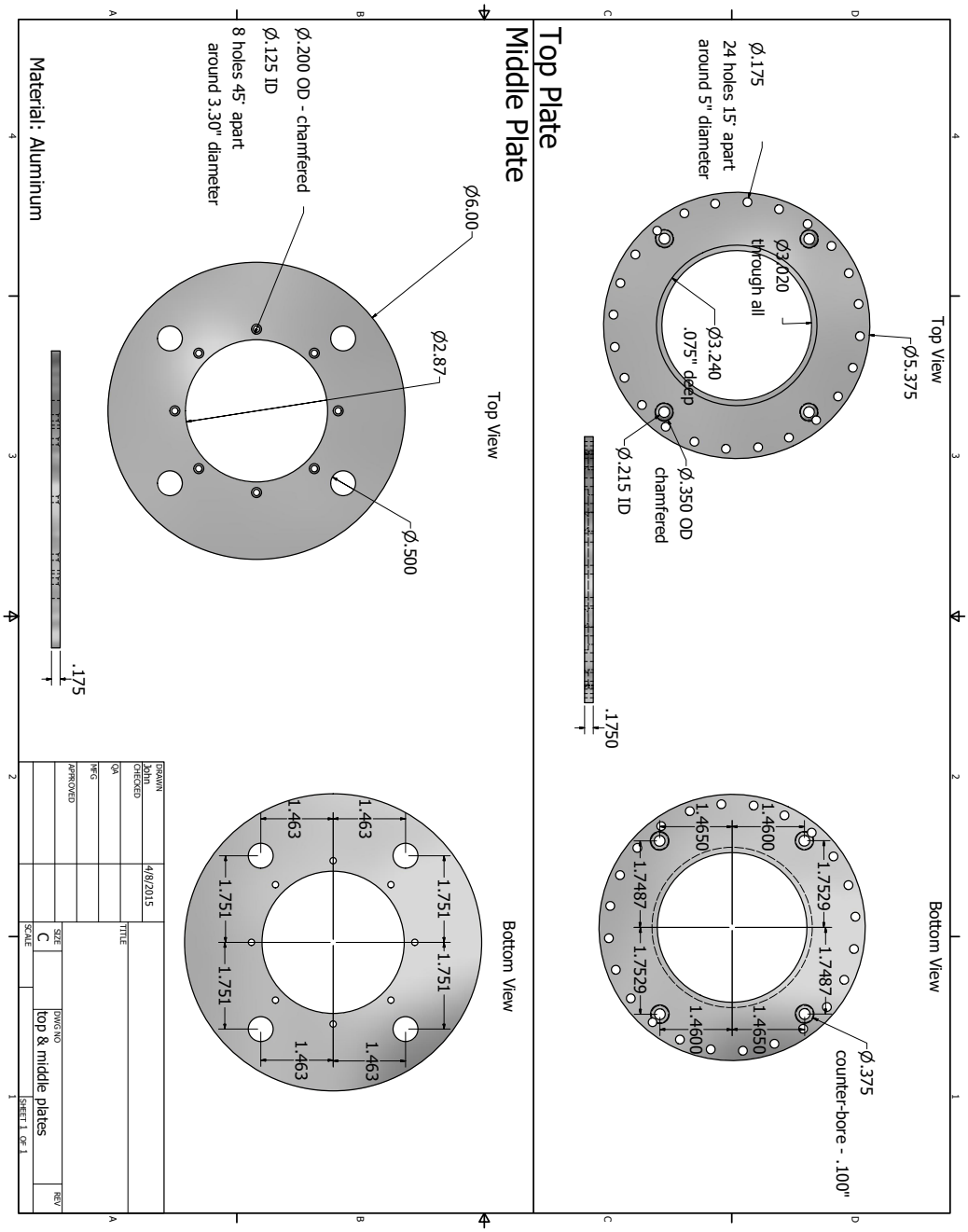


Figure A.1: Top and middle structural plates.

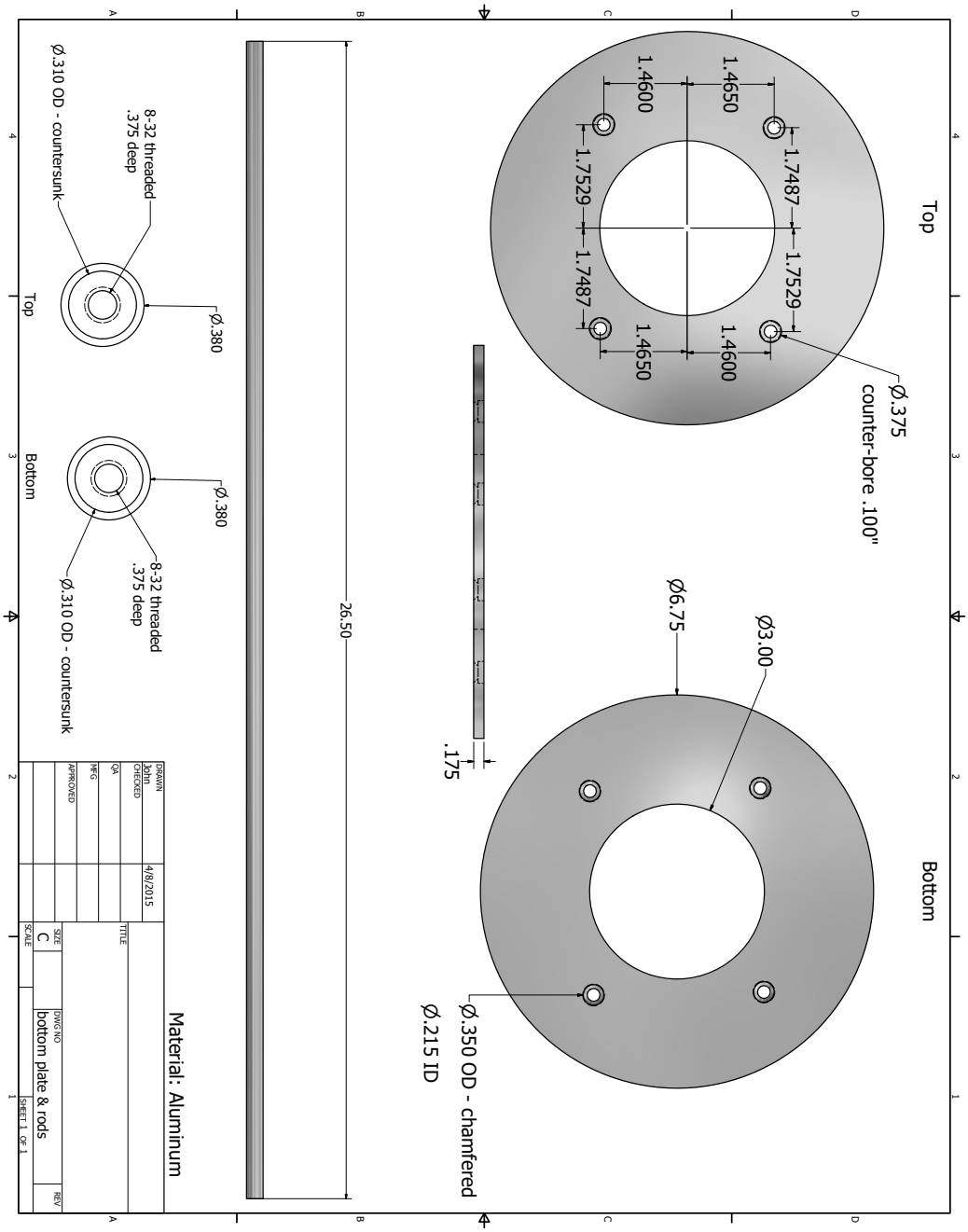


Figure A.2: Bottom structural plate and rods.

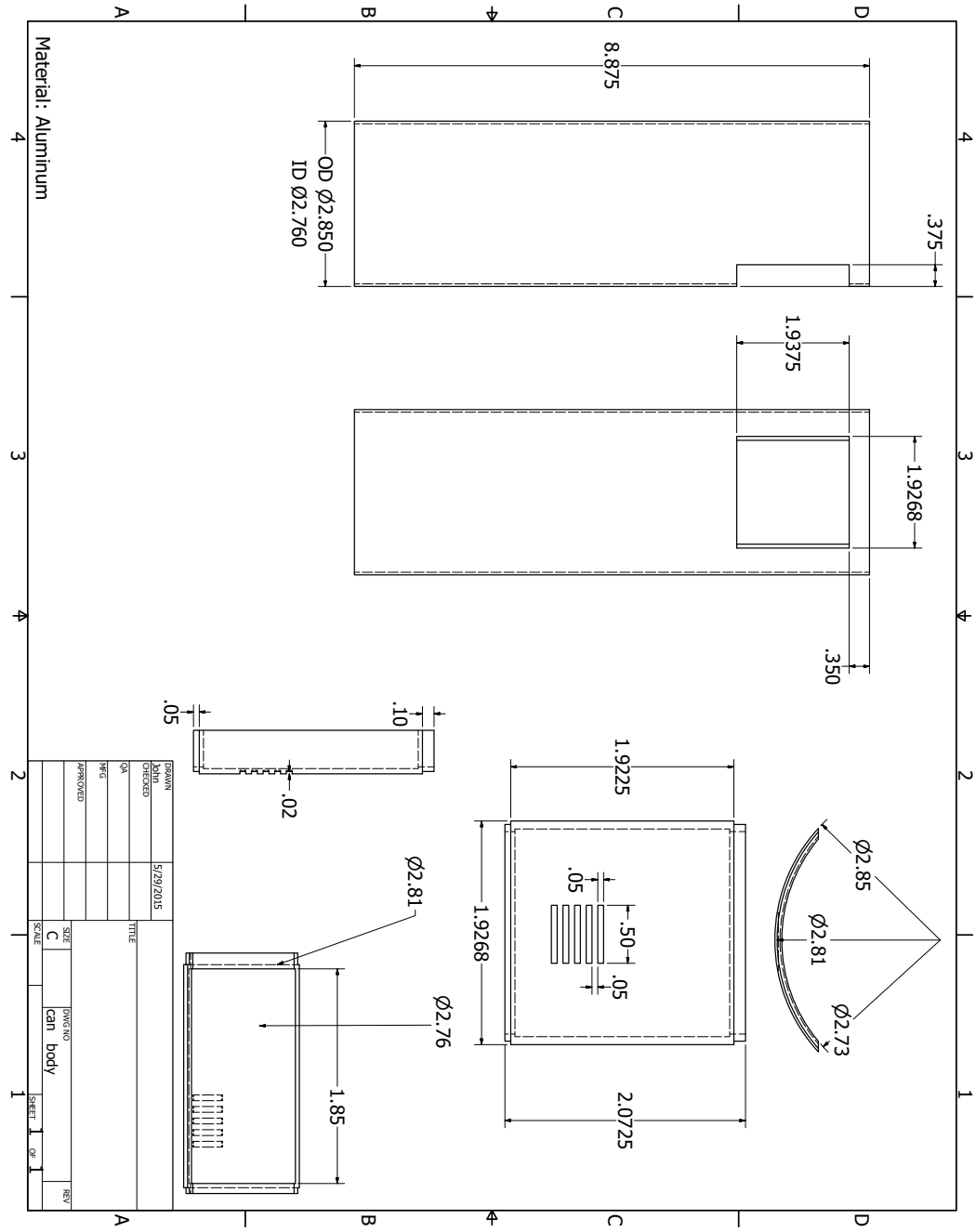


Figure A.3: Probe can and can door.

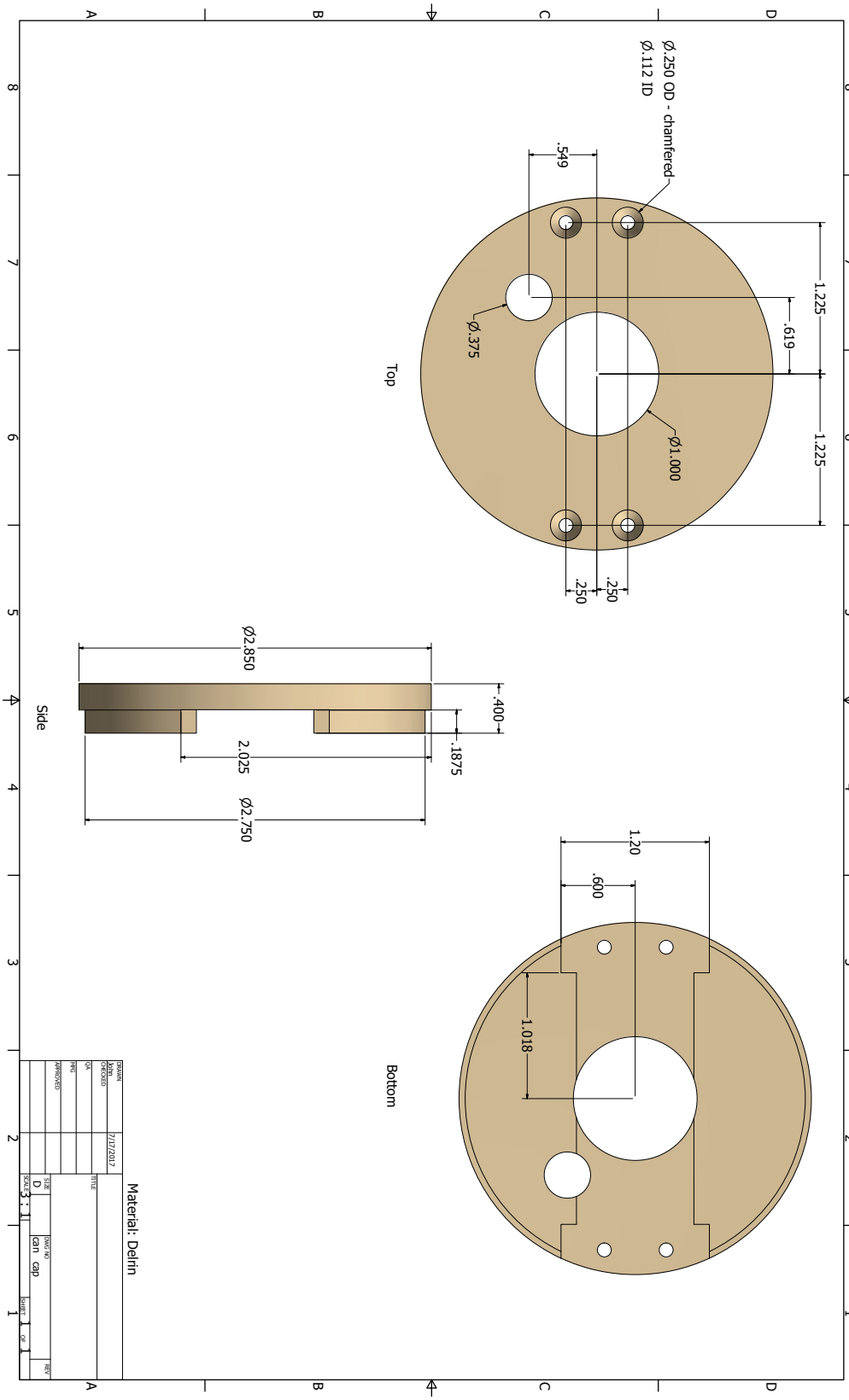


Figure A.4: Probe can cap.

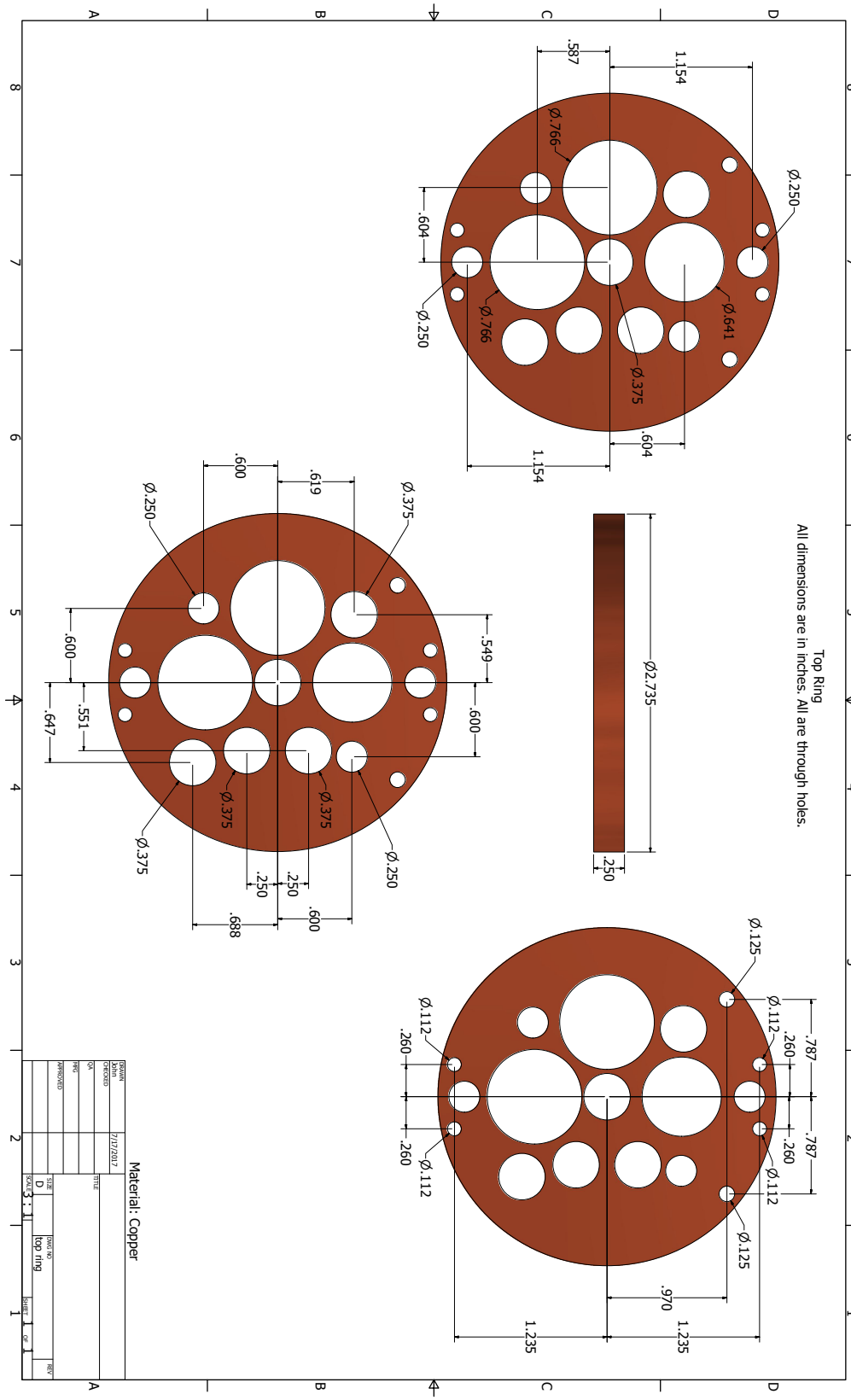


Figure A.5: Top ring.

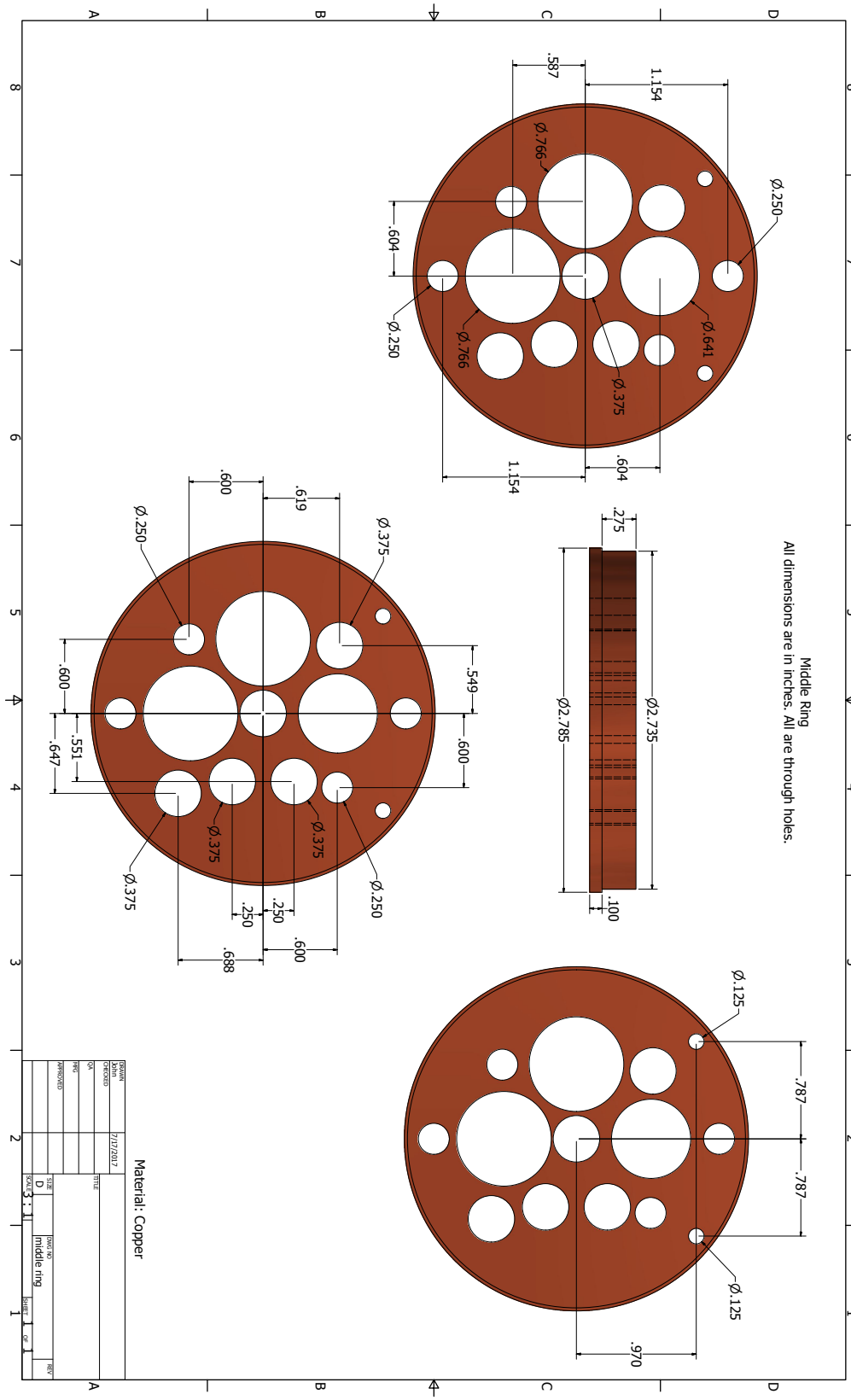


Figure A.6: Middle ring.

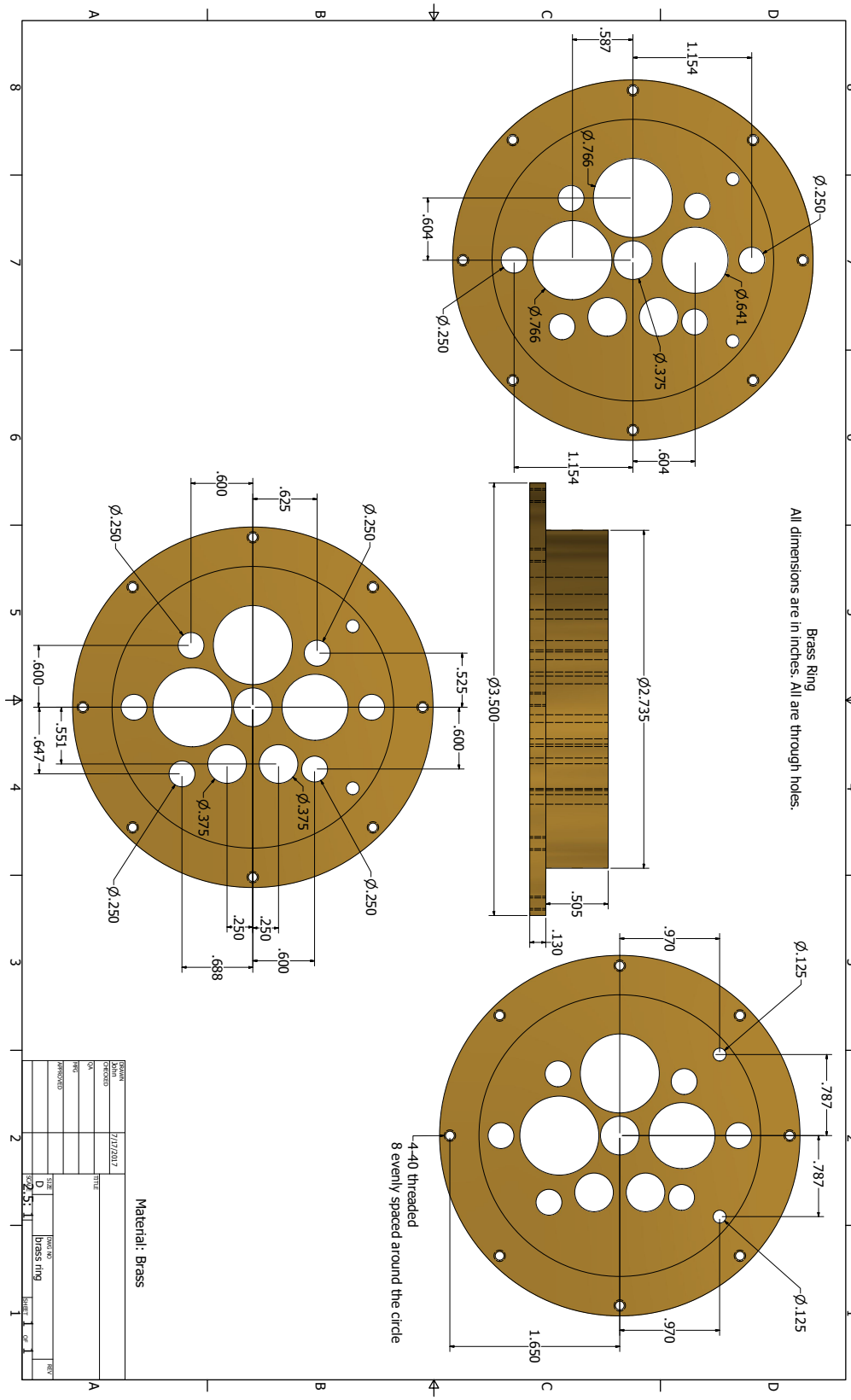


Figure A.7: Brass ring (bottom ring).

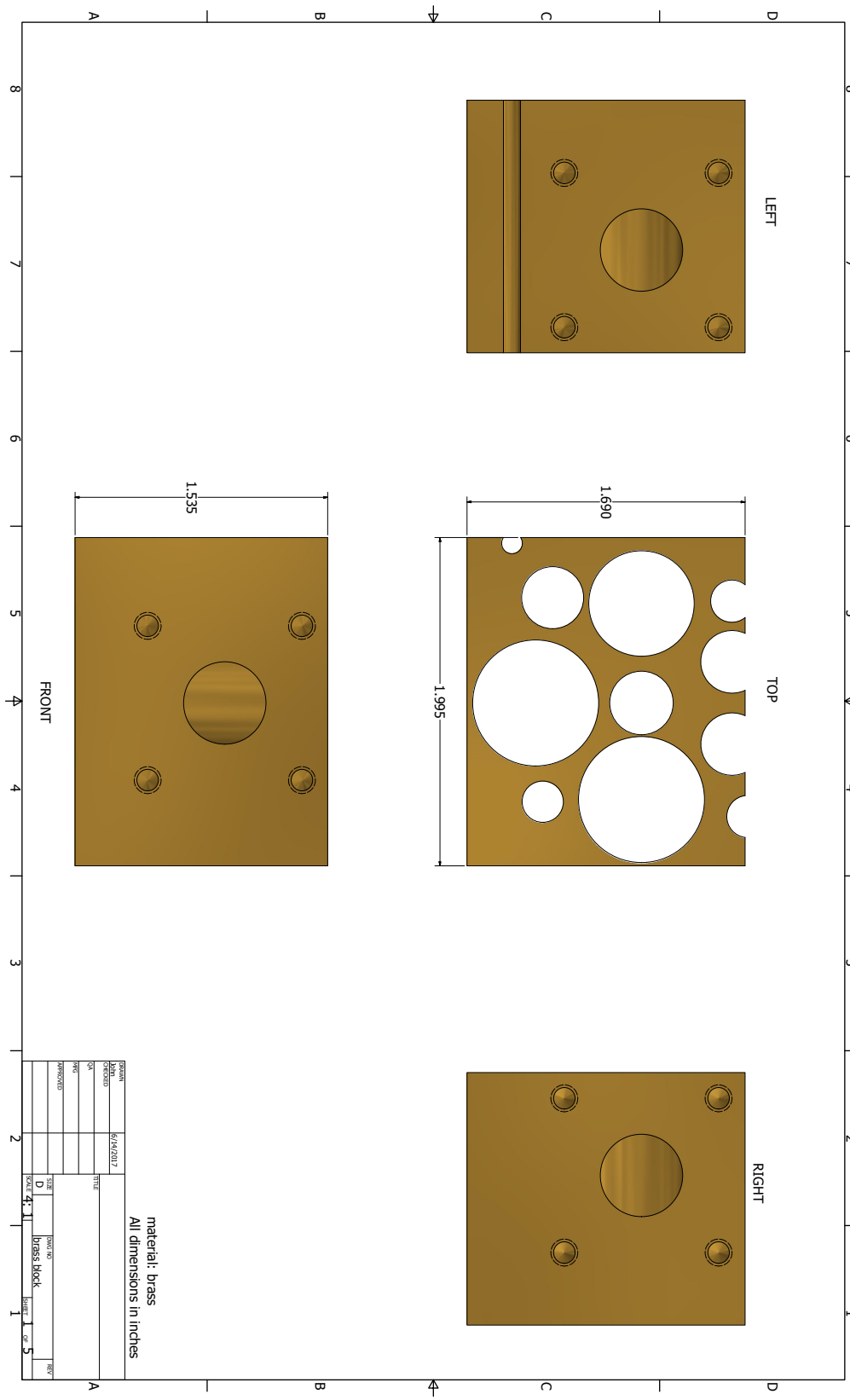


Figure A.8: Brass block full view.

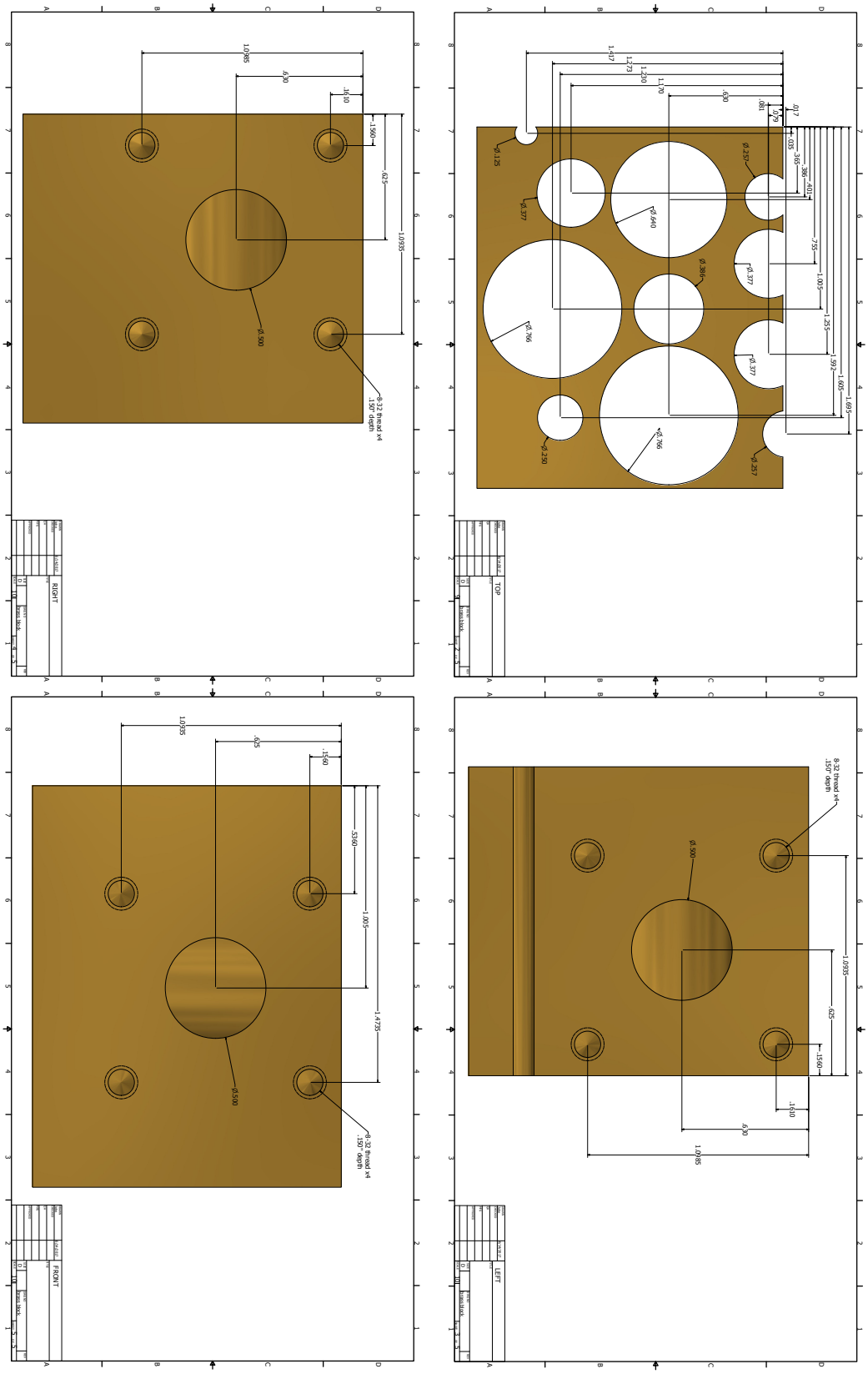


Figure A.9: Brass block side dimensions.

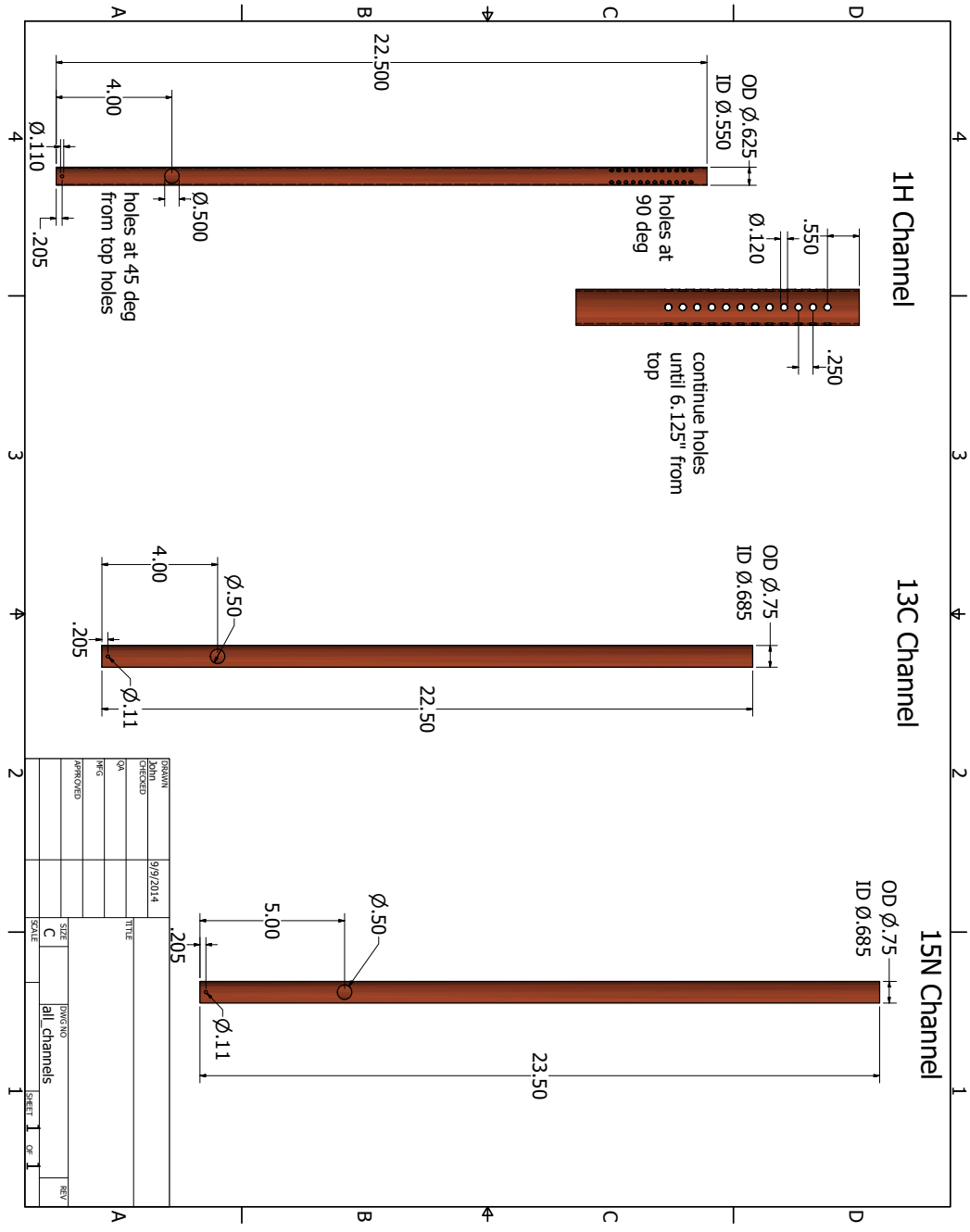


Figure A.10: Probe channels.

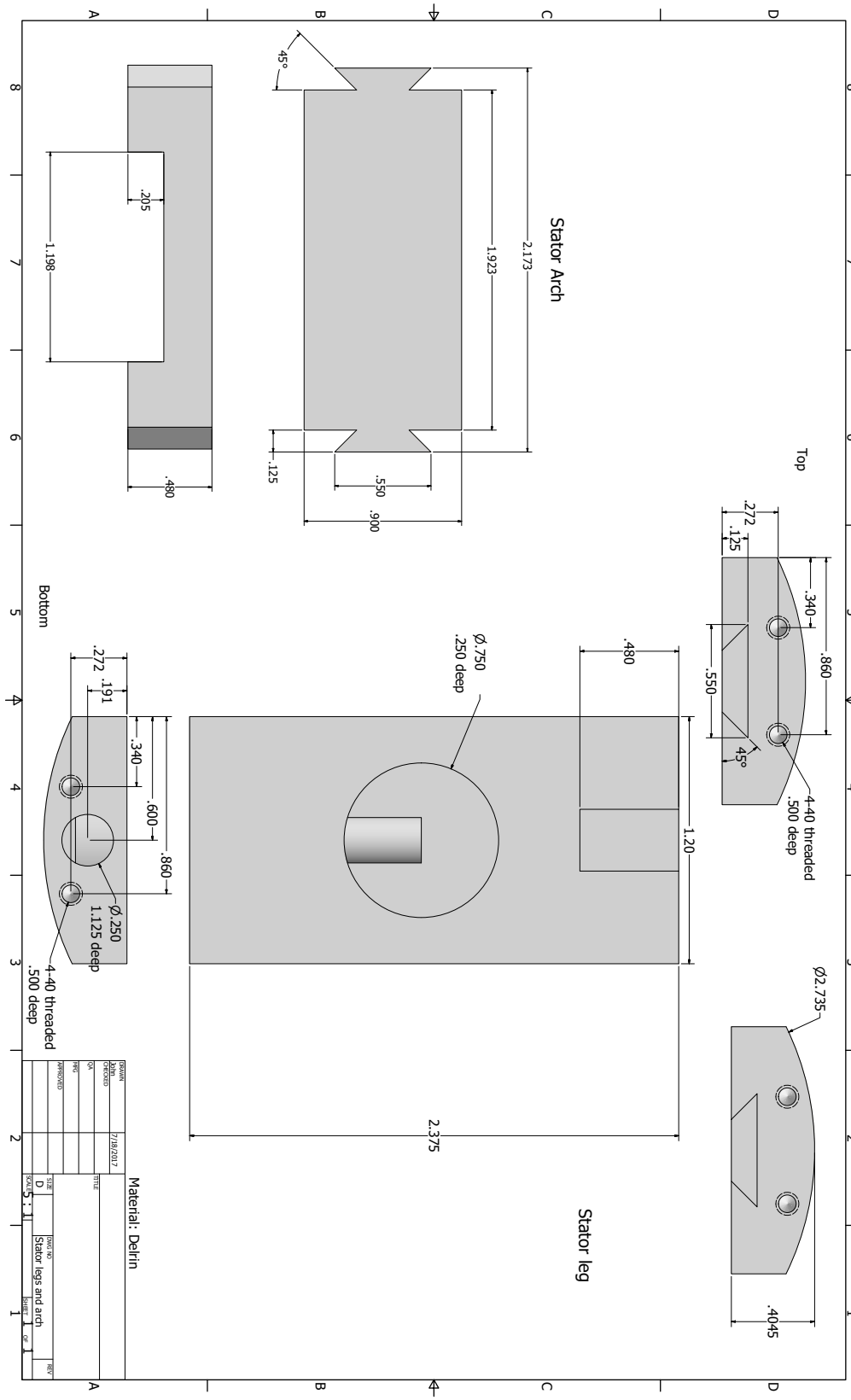


Figure A.11: Stator legs and arch.

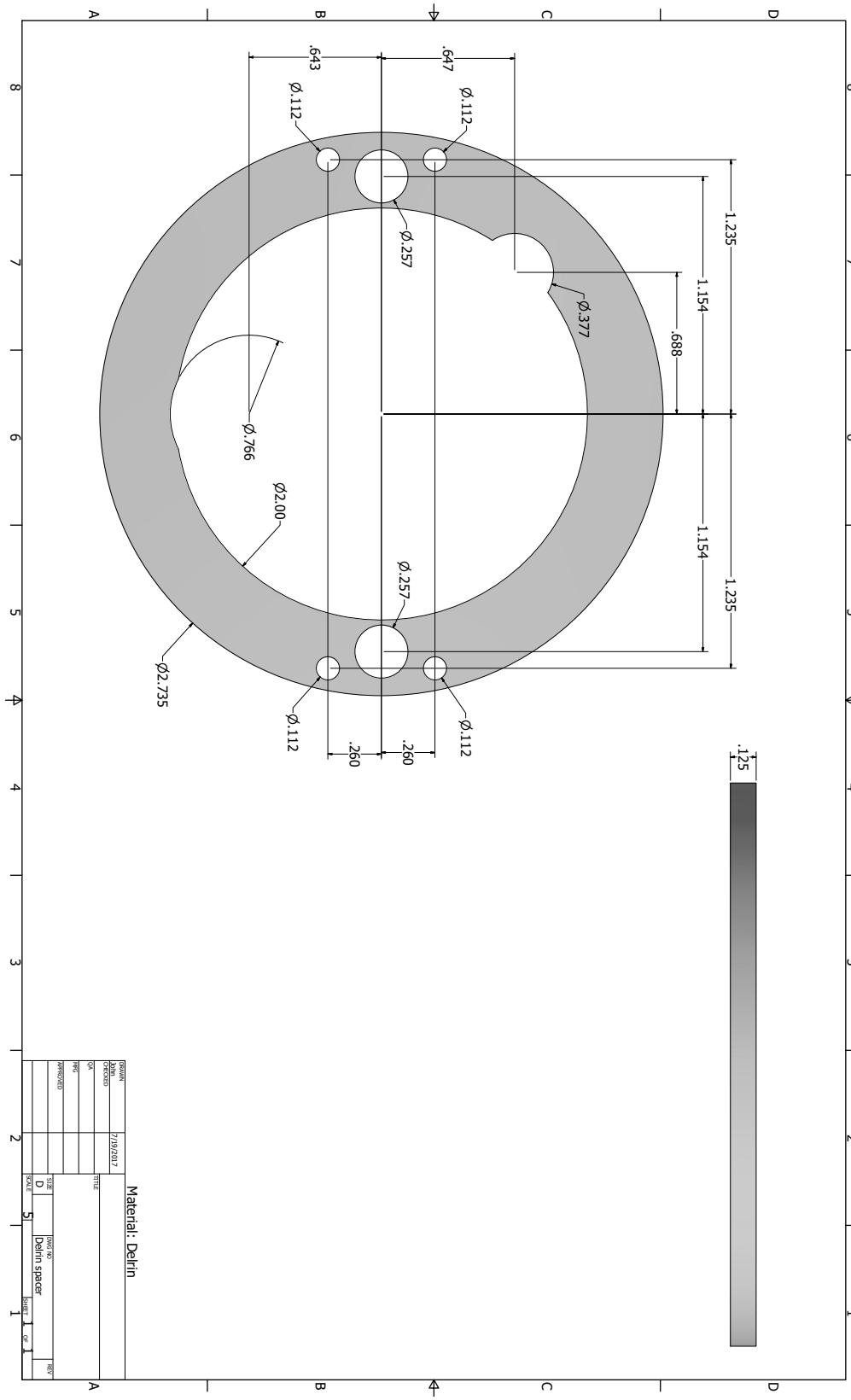


Figure A.12: Delrin spacer for probe top.

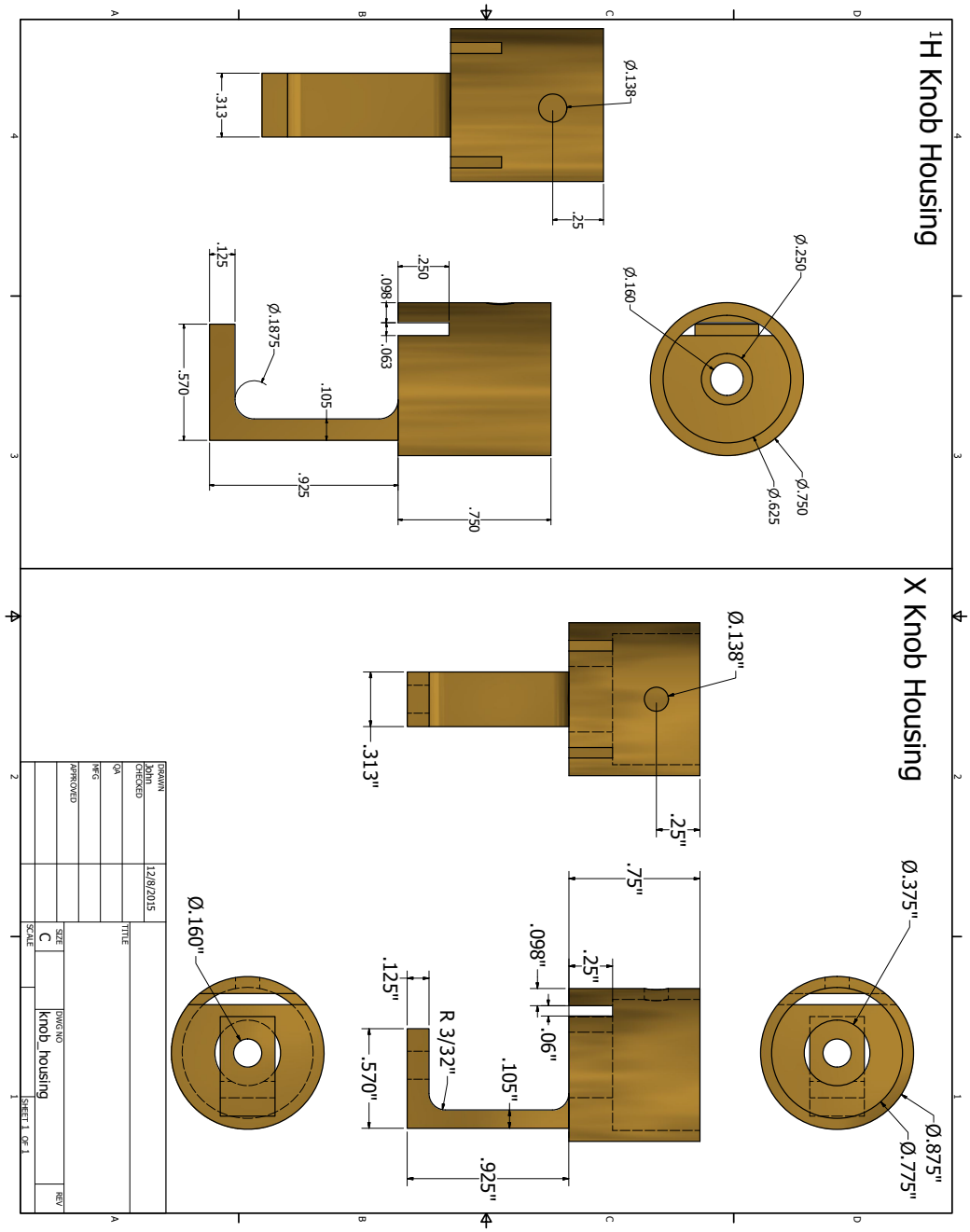


Figure A.13: Knob housing for the bottom of the channels.

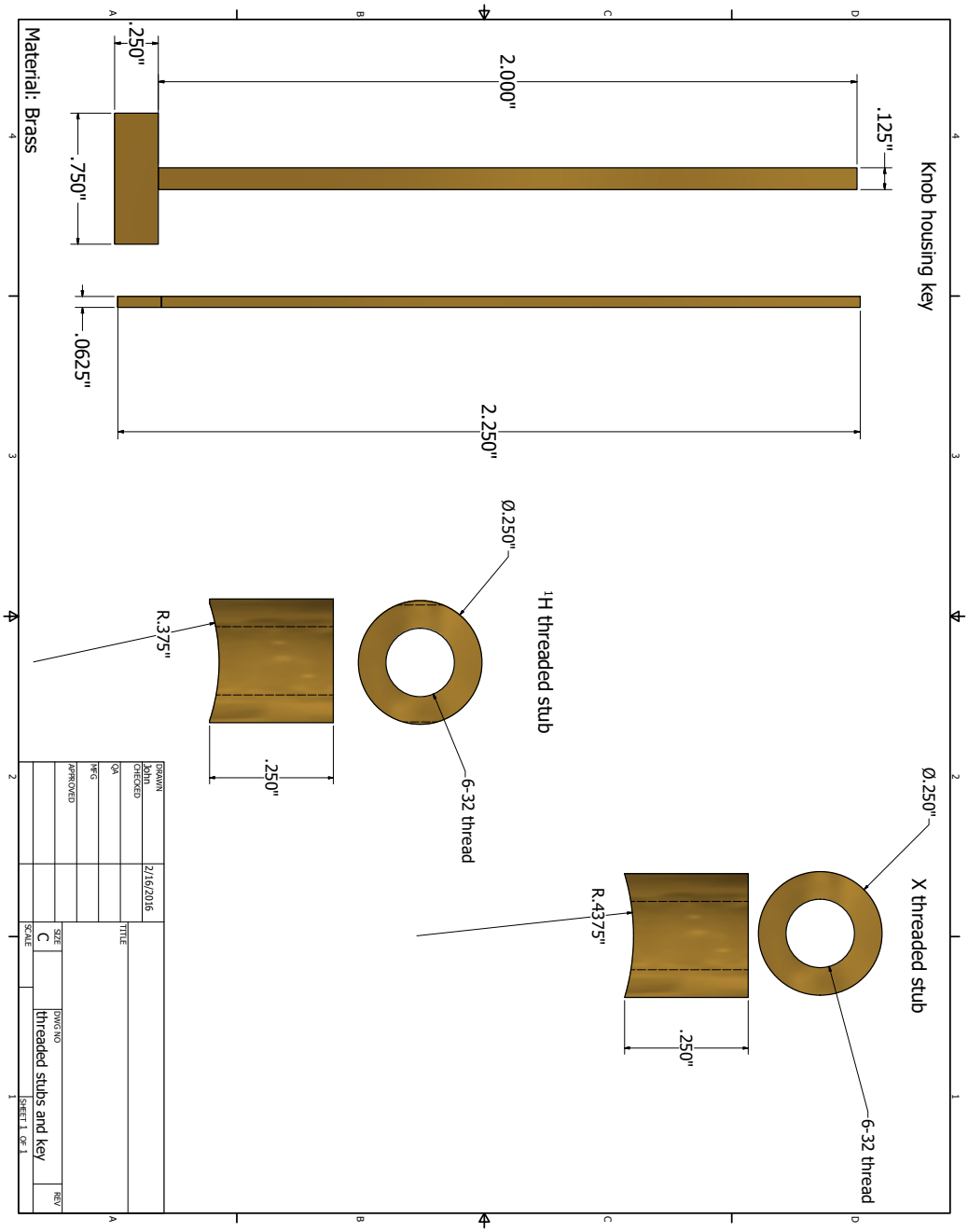


Figure A.14: Brass key and threaded stubs for the knob housings.

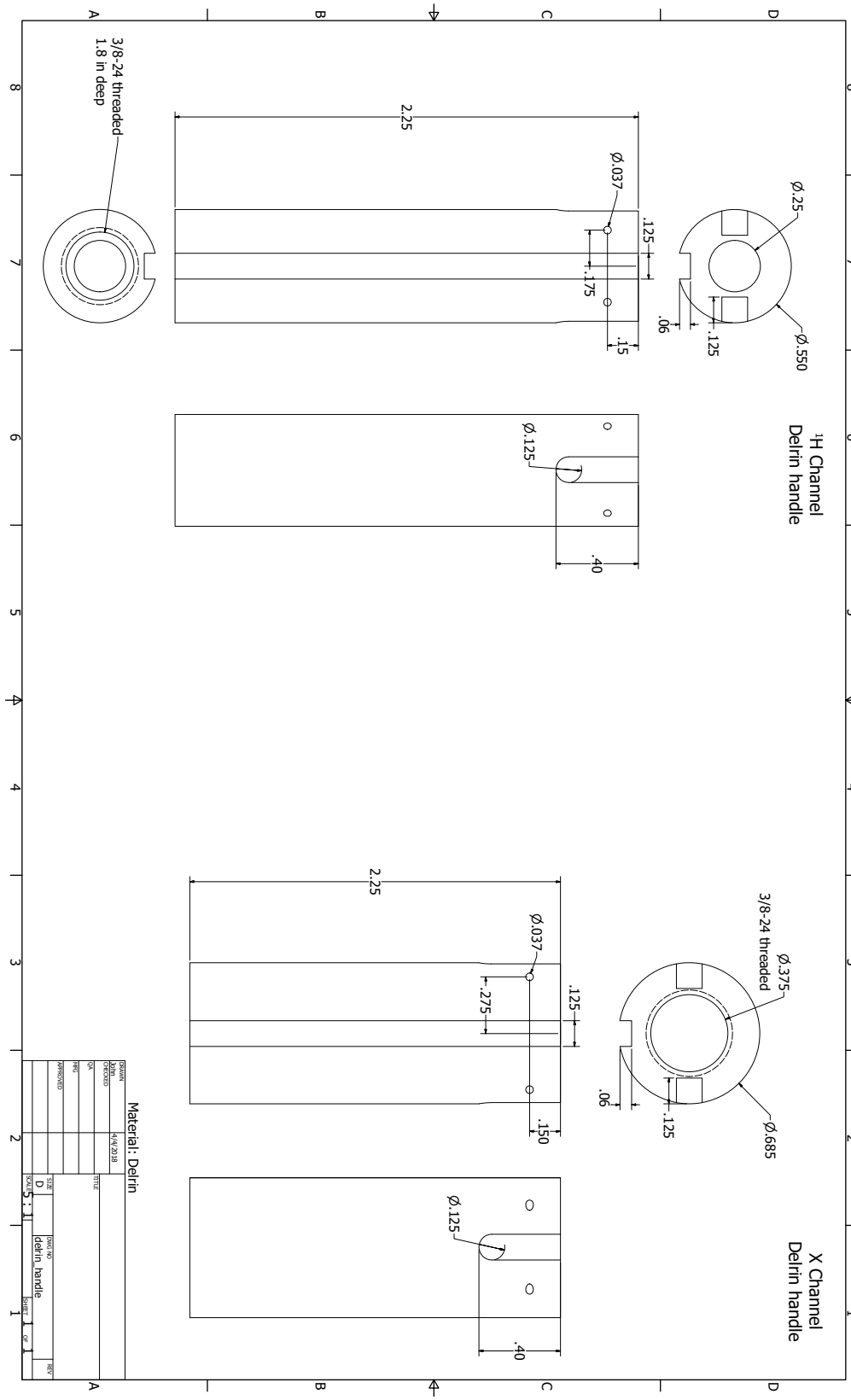


Figure A.15: Delrin handle connecting to the tune capacitor.

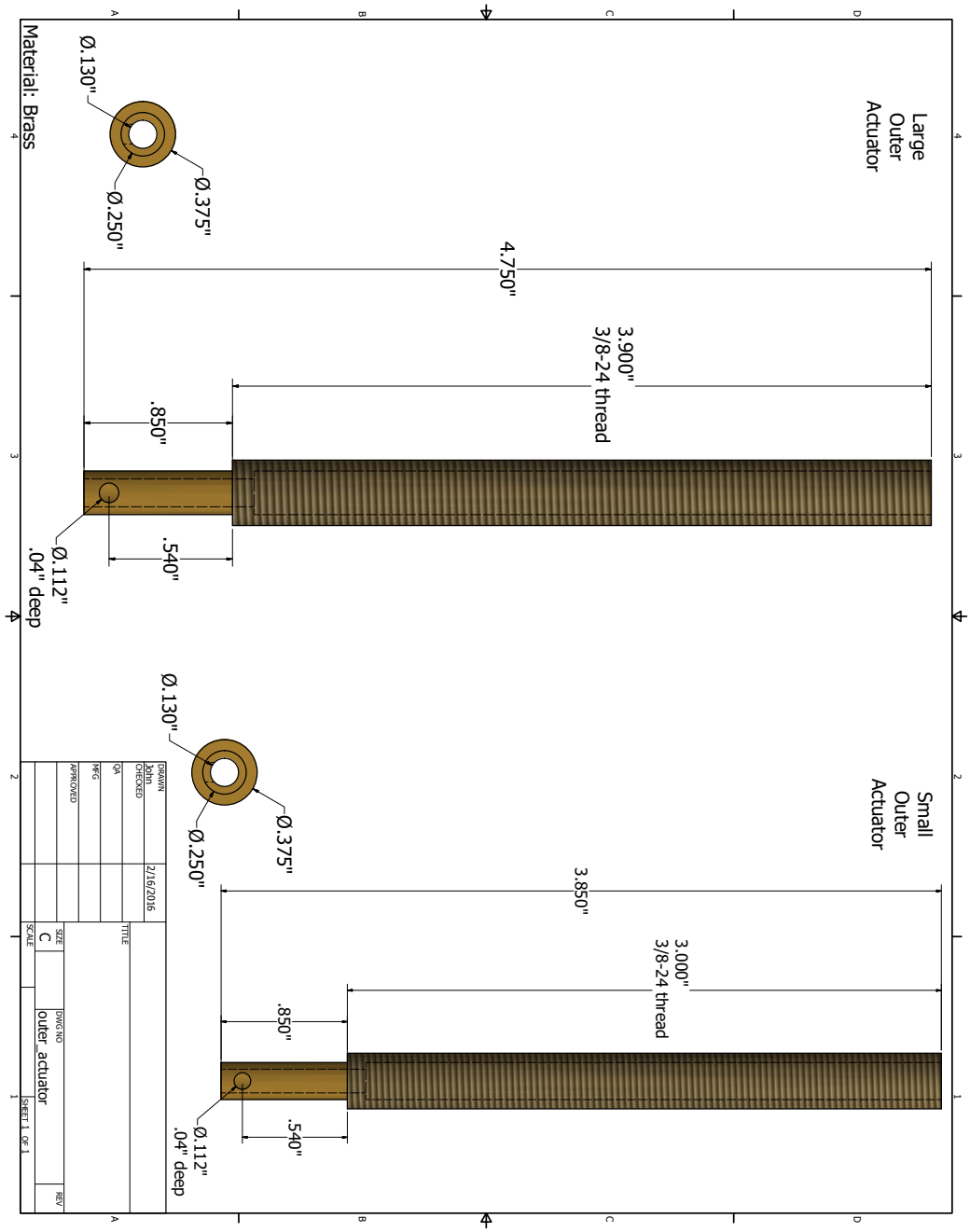


Figure A.16: Outer actuator (tune).

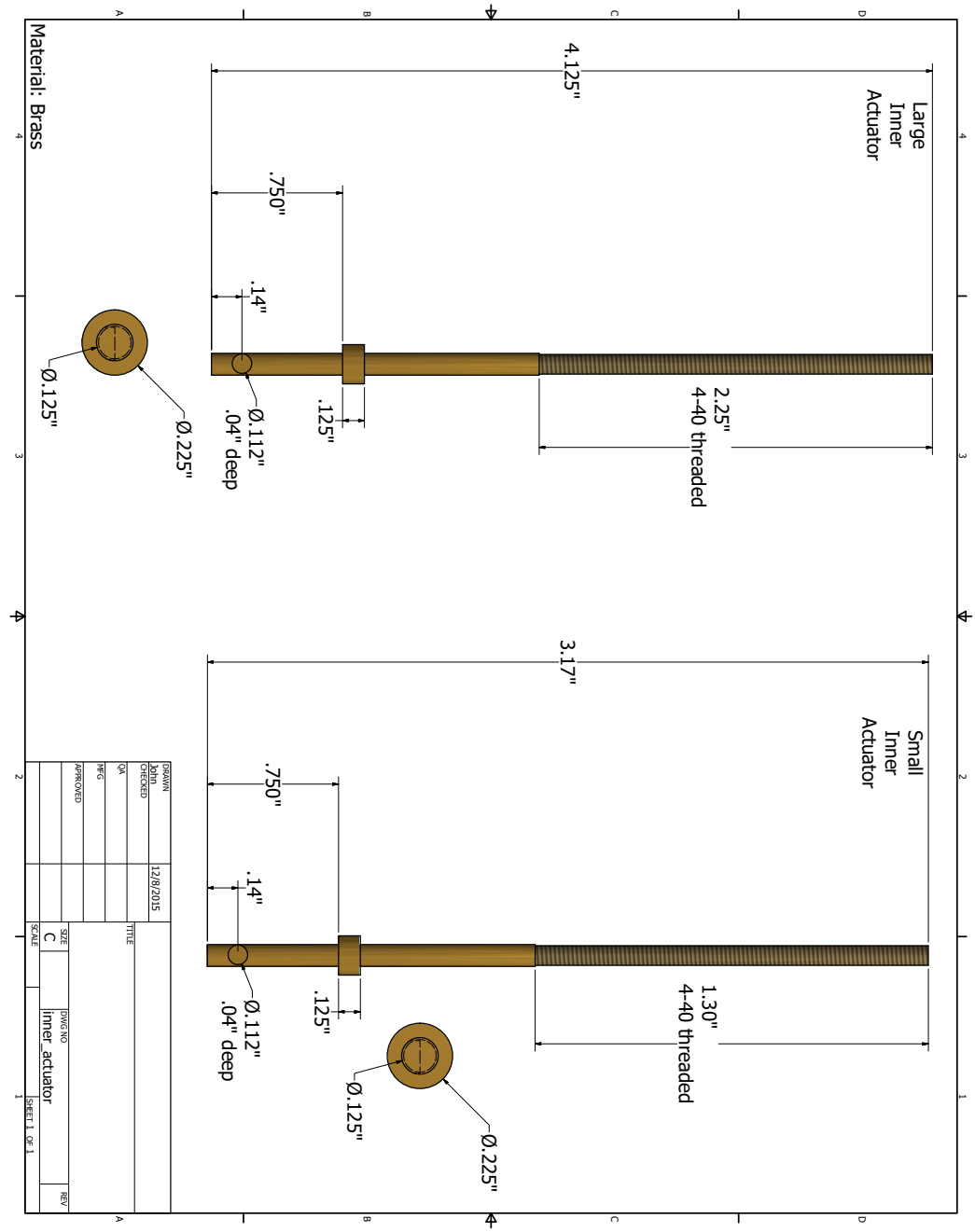


Figure A.17: Inner actuator (match).

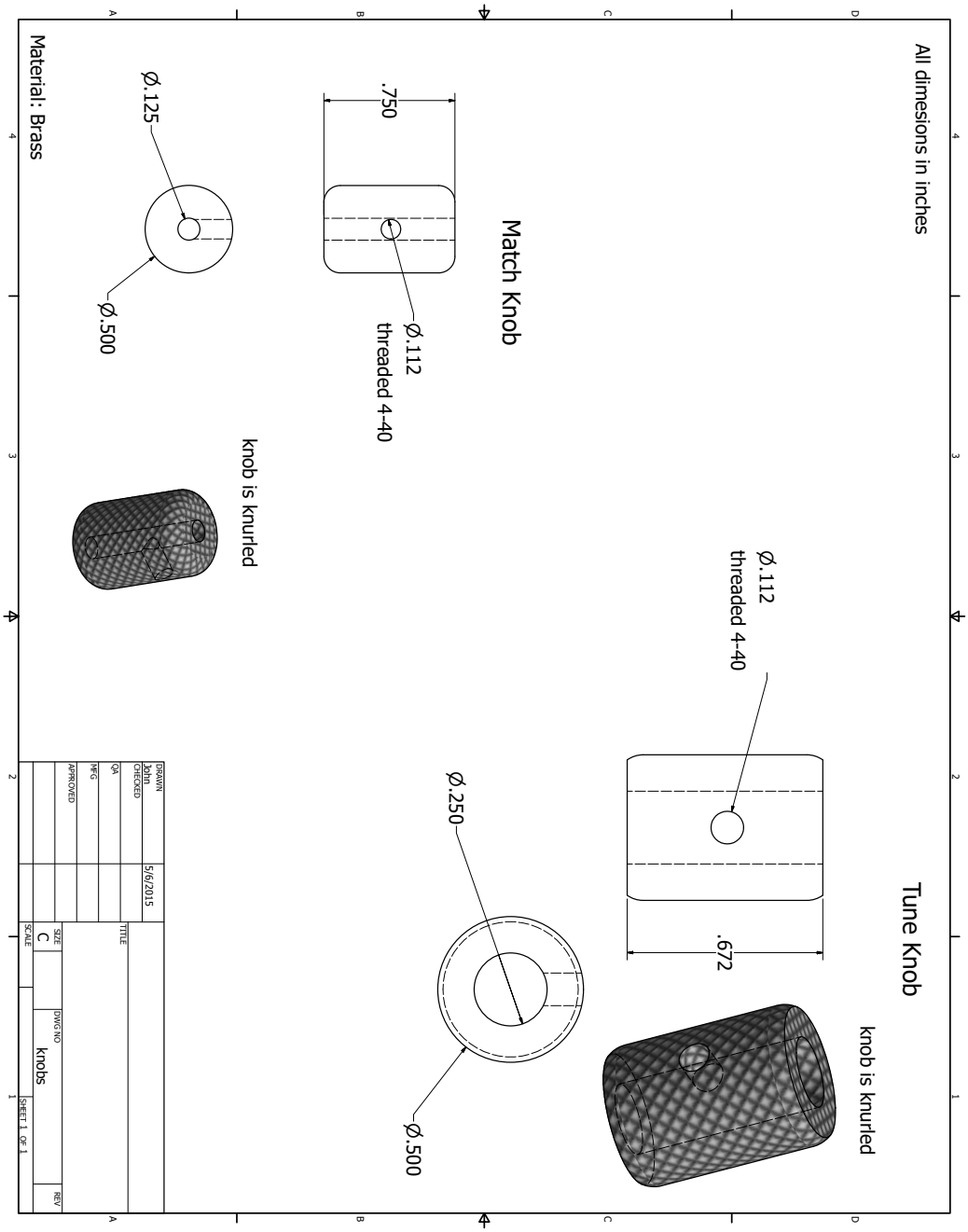


Figure A.18: Tune and Match adjust knobs.

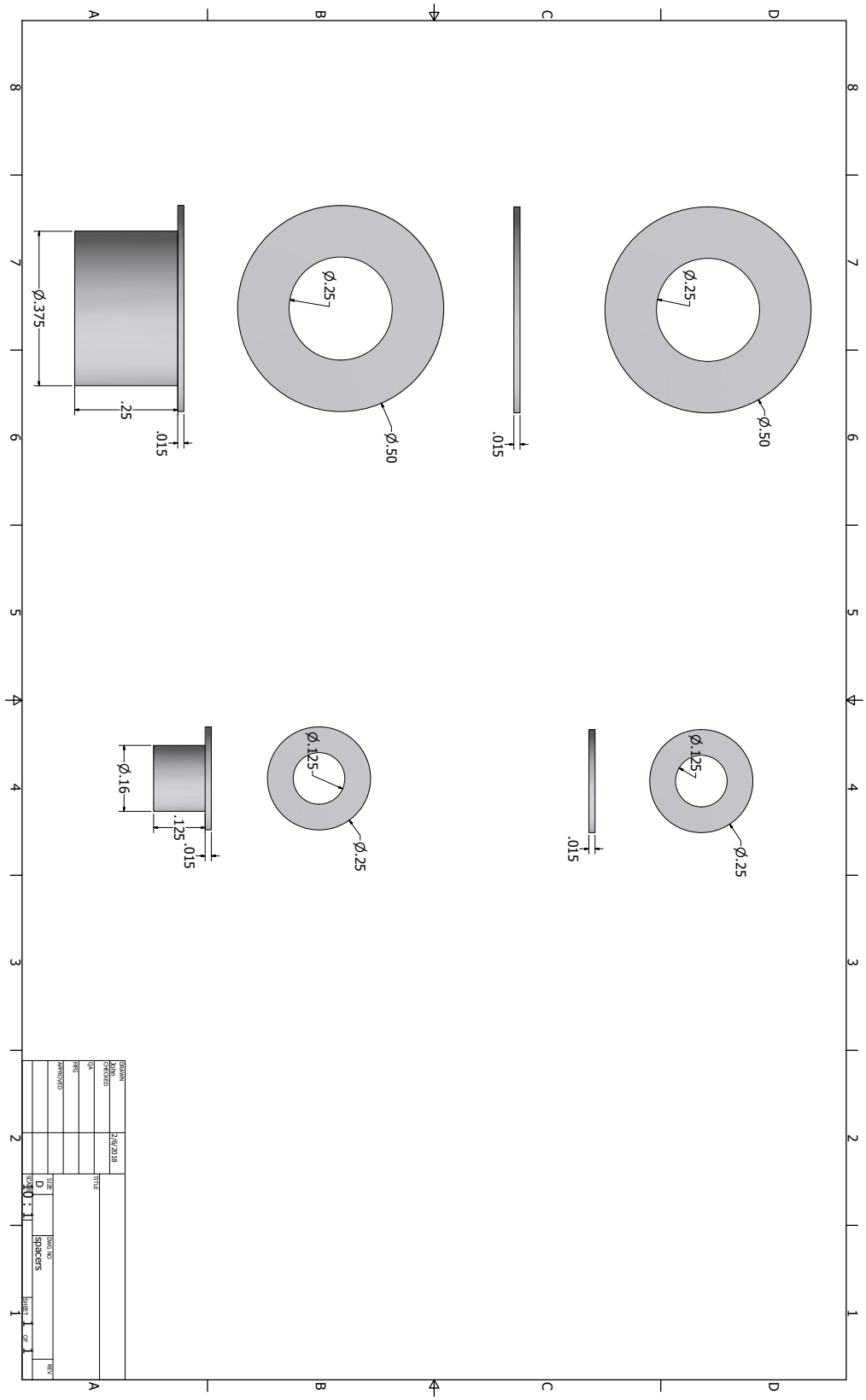


Figure A.19: Teflon spacers for the knob housing.

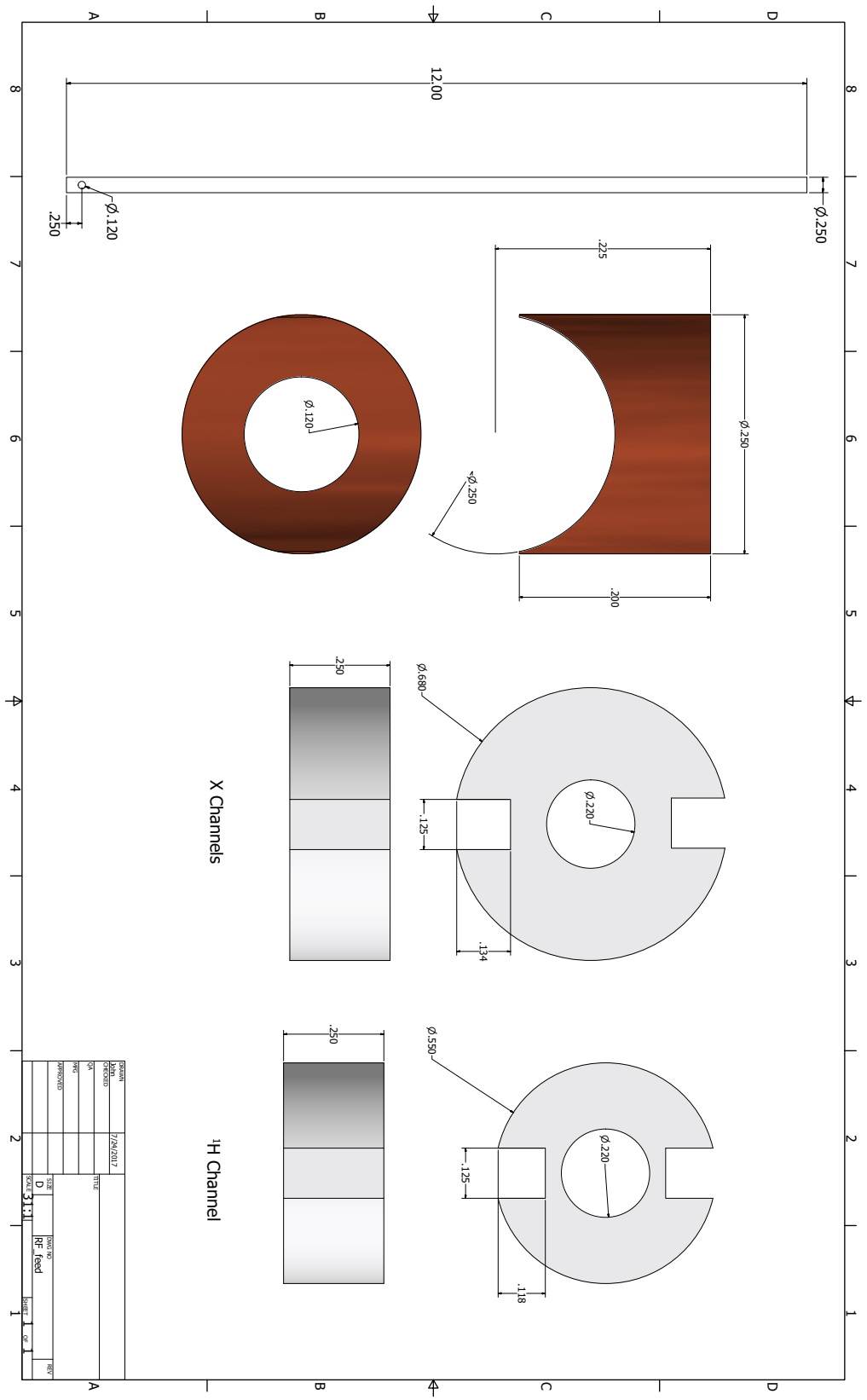


Figure A.20: RF feed and stub with spacers for each channel size.

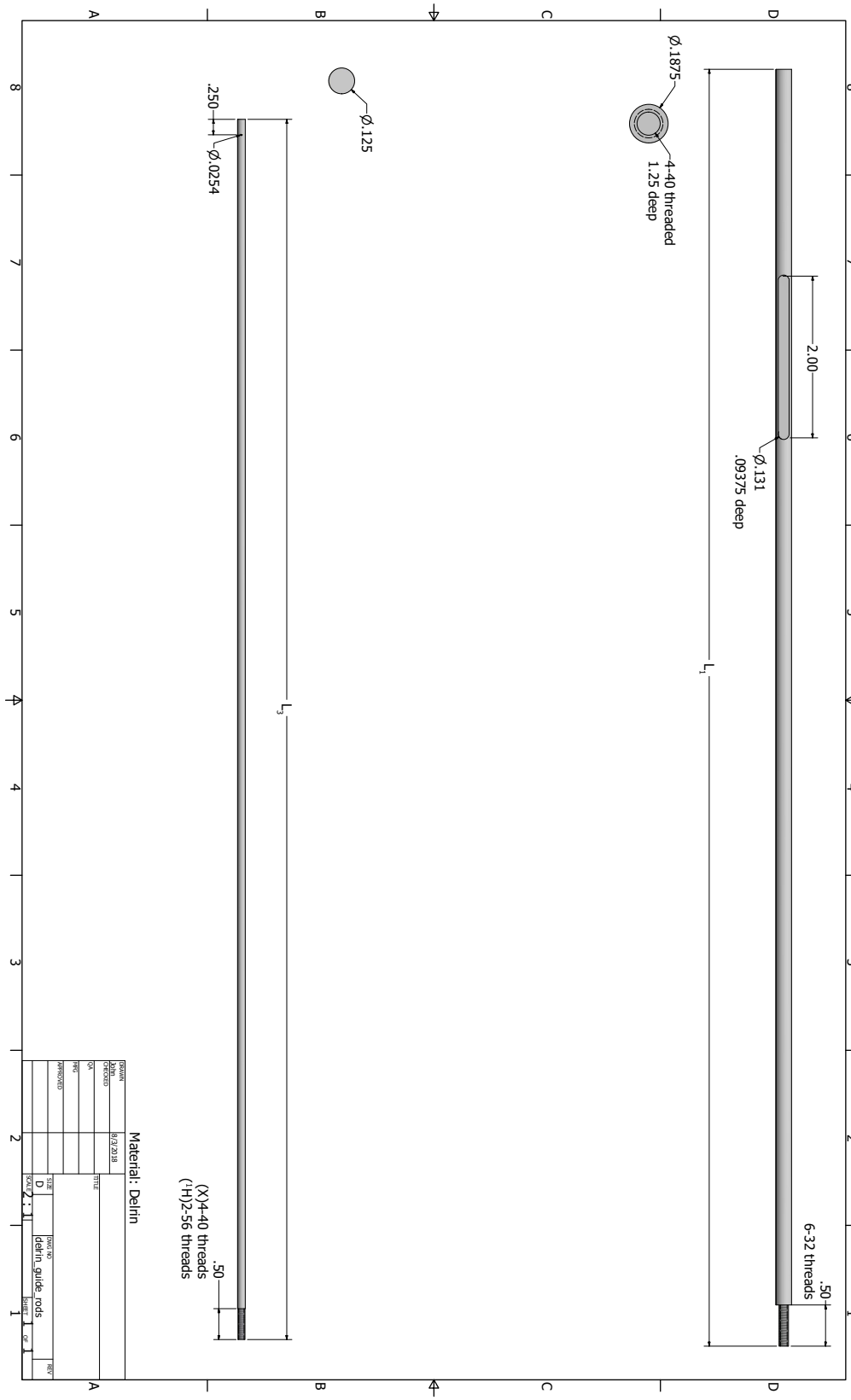


Figure A.21: Delrin guide rods for tune (bottom) and match (top).

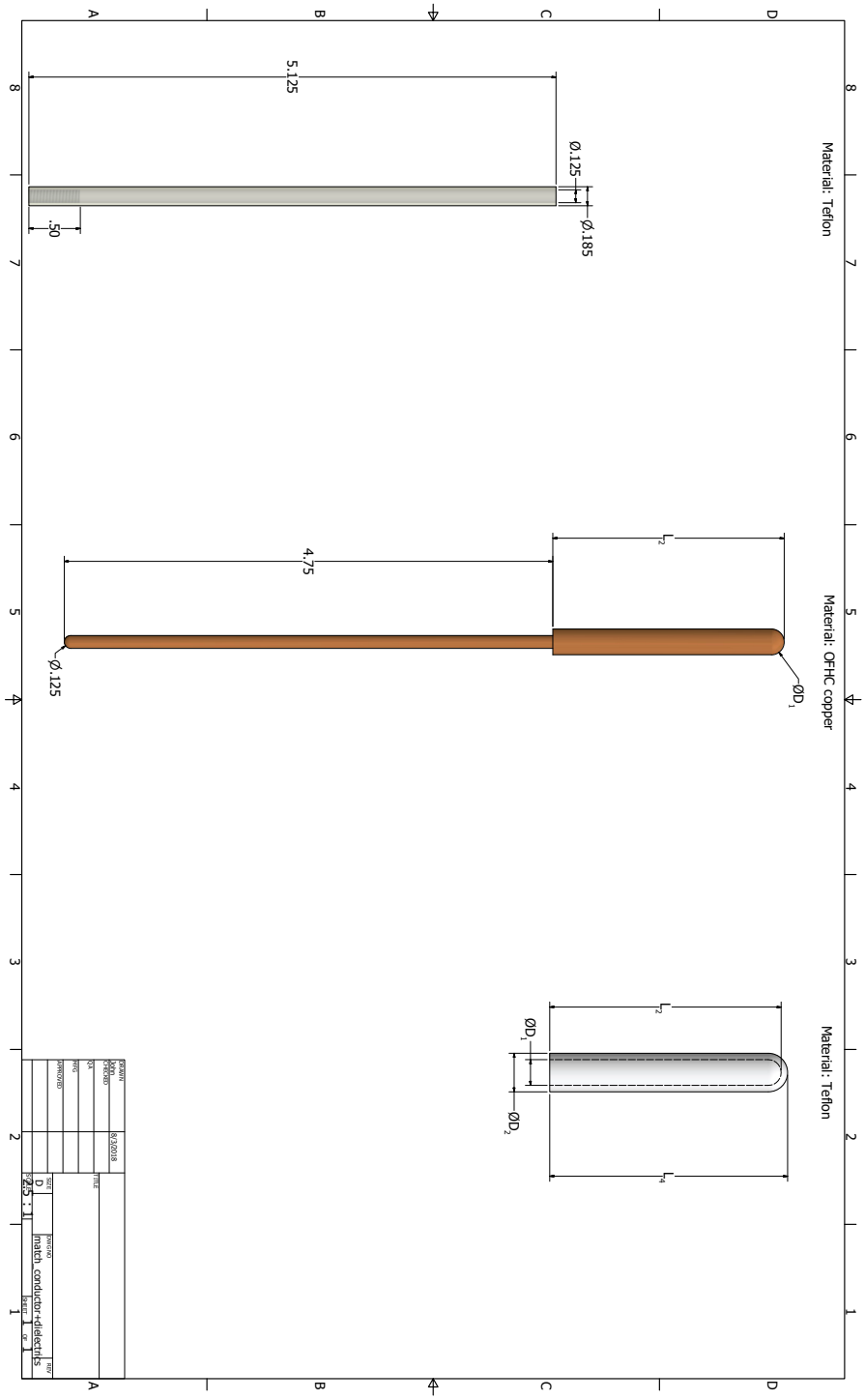


Figure A.22: Match conductor, match dielectric, and match sleeve for the XY channels; L_2 is the length of the match conductor as well as the depth of the hole in the match dielectric, L_4 is the length of the match dielectric (equal to the depth of the hole in the tune conductor), D_1 is the diameter of the match conductor as well as the diameter of the hole in the match dielectric, D_2 is the diameter of the outer diameter of the match dielectric (equal to the hole in the tune conductor). All variable values will be particular to the chosen nucleus/frequency.

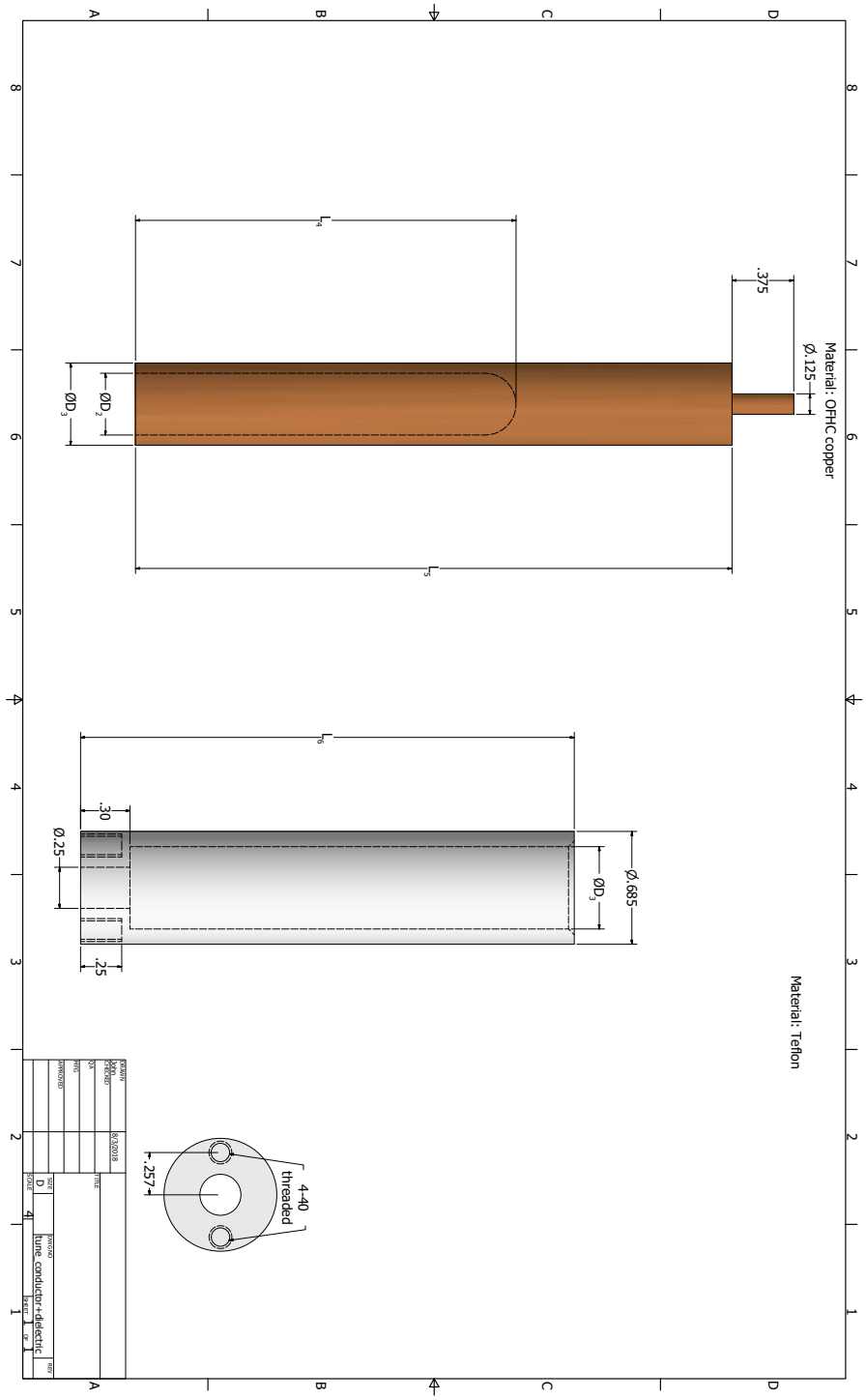


Figure A.23: Tune conductor and dielectric for the XY channels; L_4 is the depth of the hole in the tune conductor (equal to the length of the match dielectric), L_5 is the length of the tune conductor, L_6 is the length of the tune dielectric, D_2 is the diameter of the hole in the tune conductor (equal to the outer diameter of the match dielectric), D_3 is the outer diameter of the tune conductor as well as the inner diameter of the tune dielectric. All variable values will be particular to the chosen nucleus/frequency.

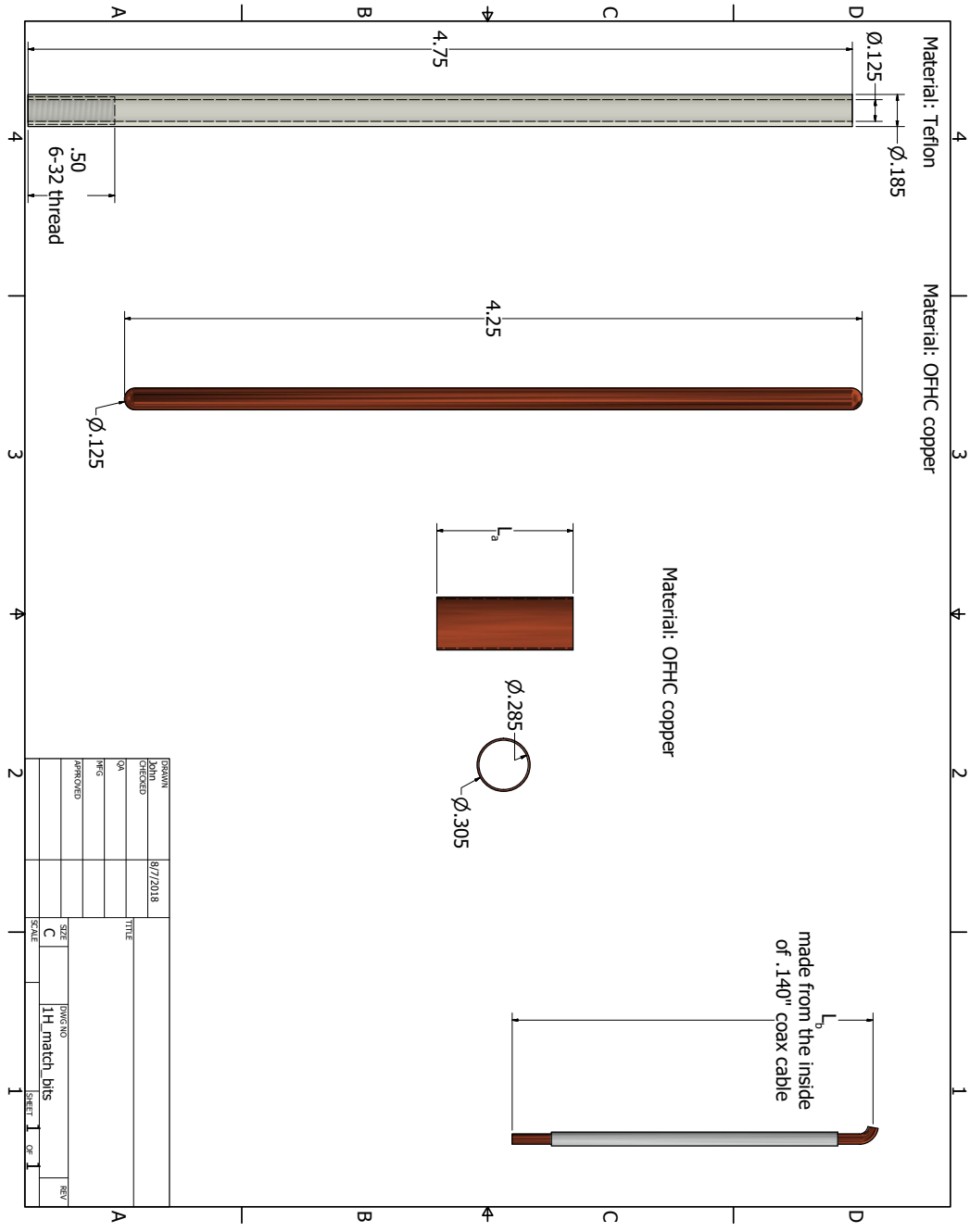


Figure A.24: Match rod, dielectric sleeve, match sleeve, and match connective wire for the ^{1}H channel. The connective wire is made from the inside of a coaxial cable and runs through the holes in the match plate and inductive stub before attaching to the match sleeve.

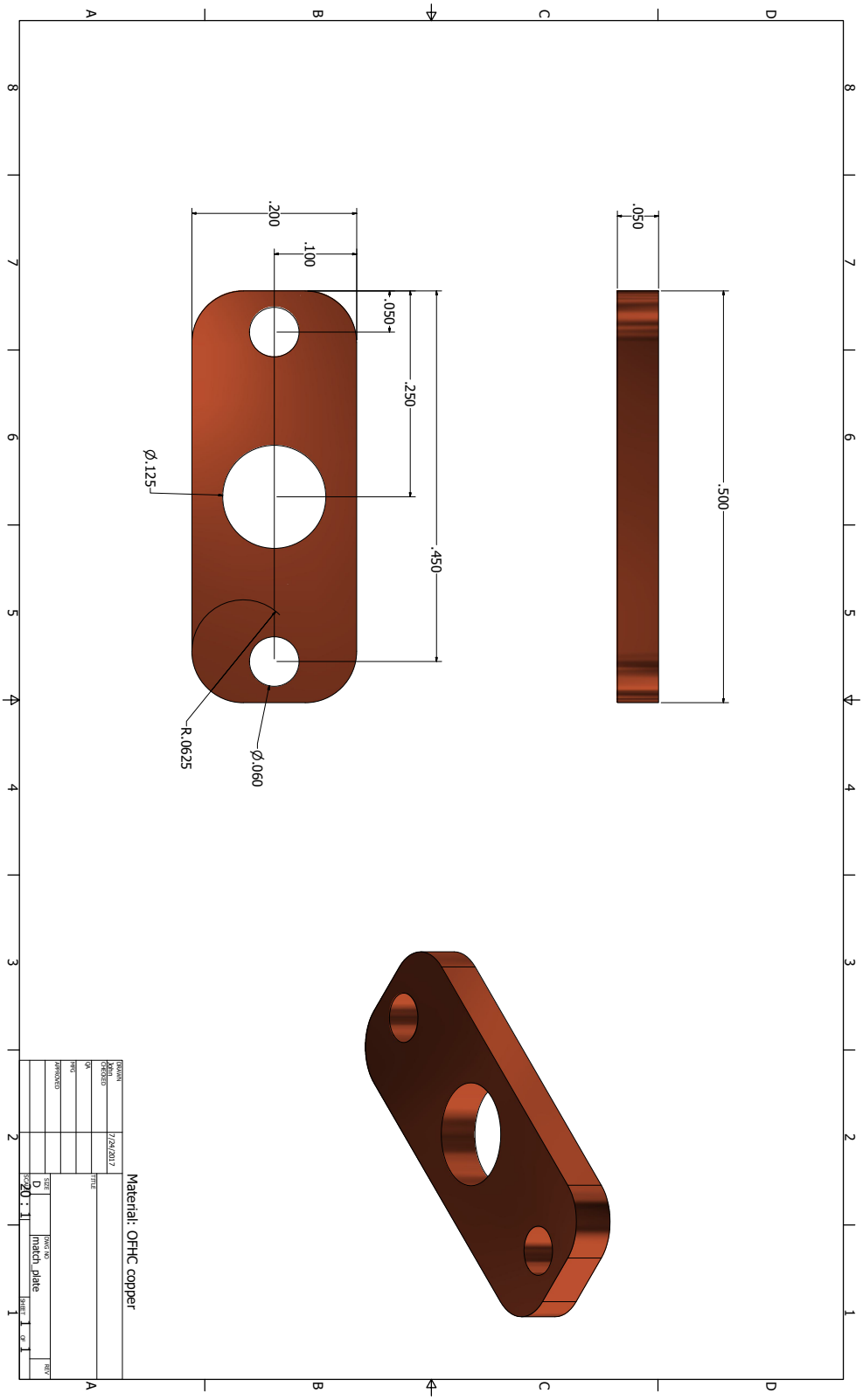


Figure A.25: Match plate for the ^{1}H channel, it is soldered to the match rod and the match connective wires are soldered into the small side holes.

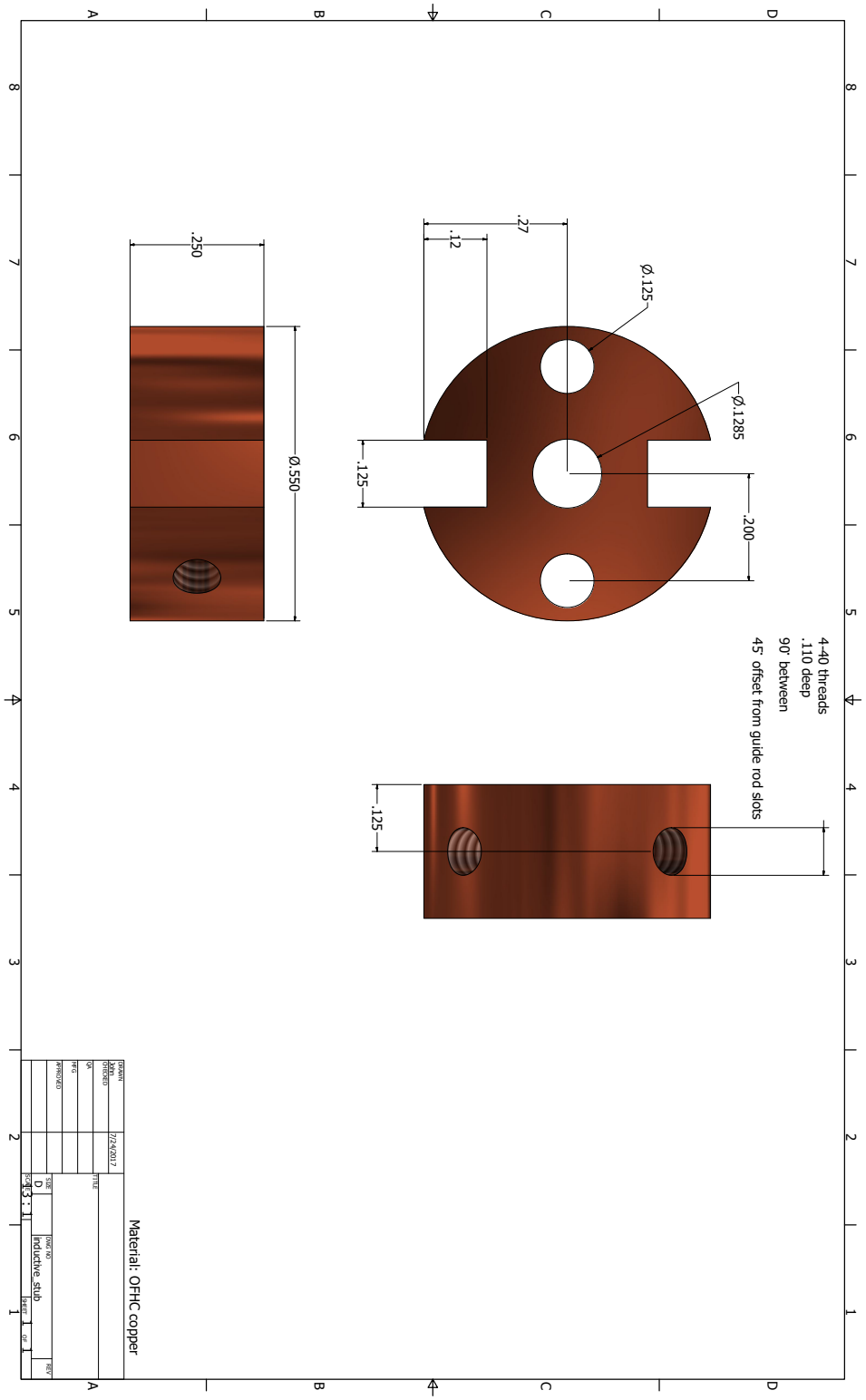


Figure A.26: Inductive stub for the ^{1}H channel. The match connective wires run through the side holes, the guide rods run though the side slots and the tune conductor solders into the center hole. The threaded side holes are used to screw the piece to the outer ground plane, causing this inductance to be in parallel with signal coil.

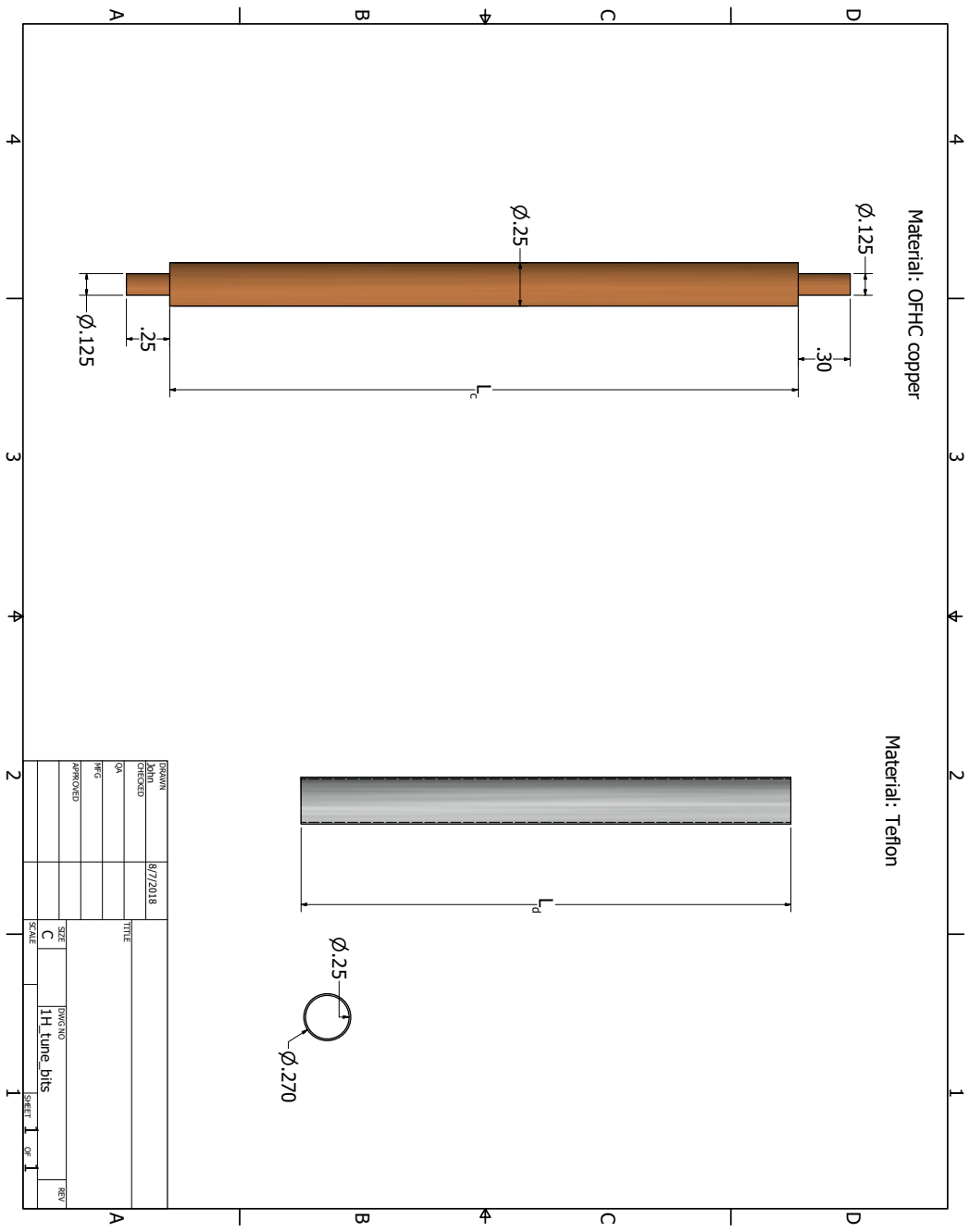


Figure A.27: Tune conductor and dielectric sleeve for the ^{1}H channel. The tune conductor is soldered into the inductive stub and the dielectric sleeve goes over the tune conductor but under the match sleeve.

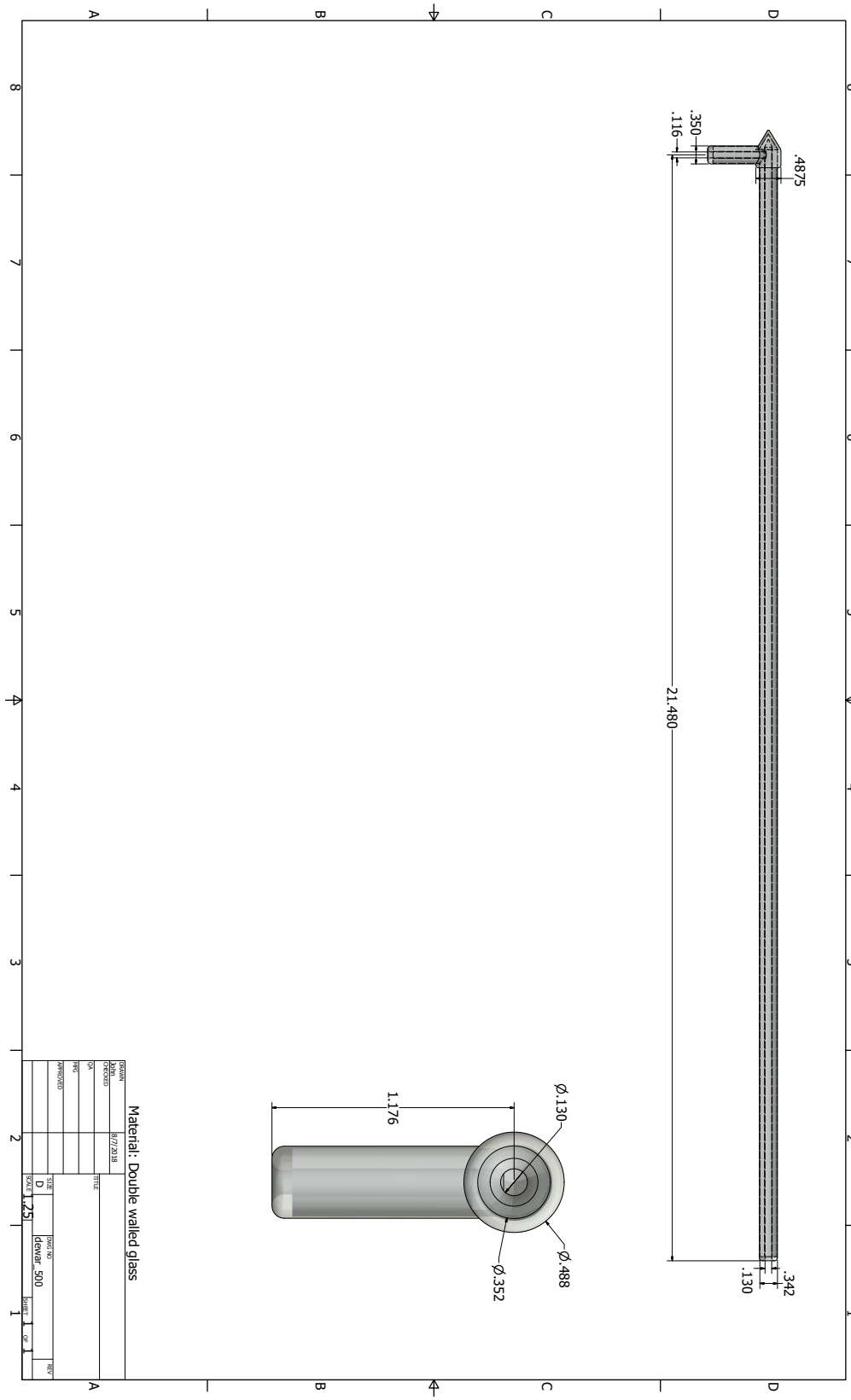


Figure A.28: Double-walled glass dewar used for the VT air.

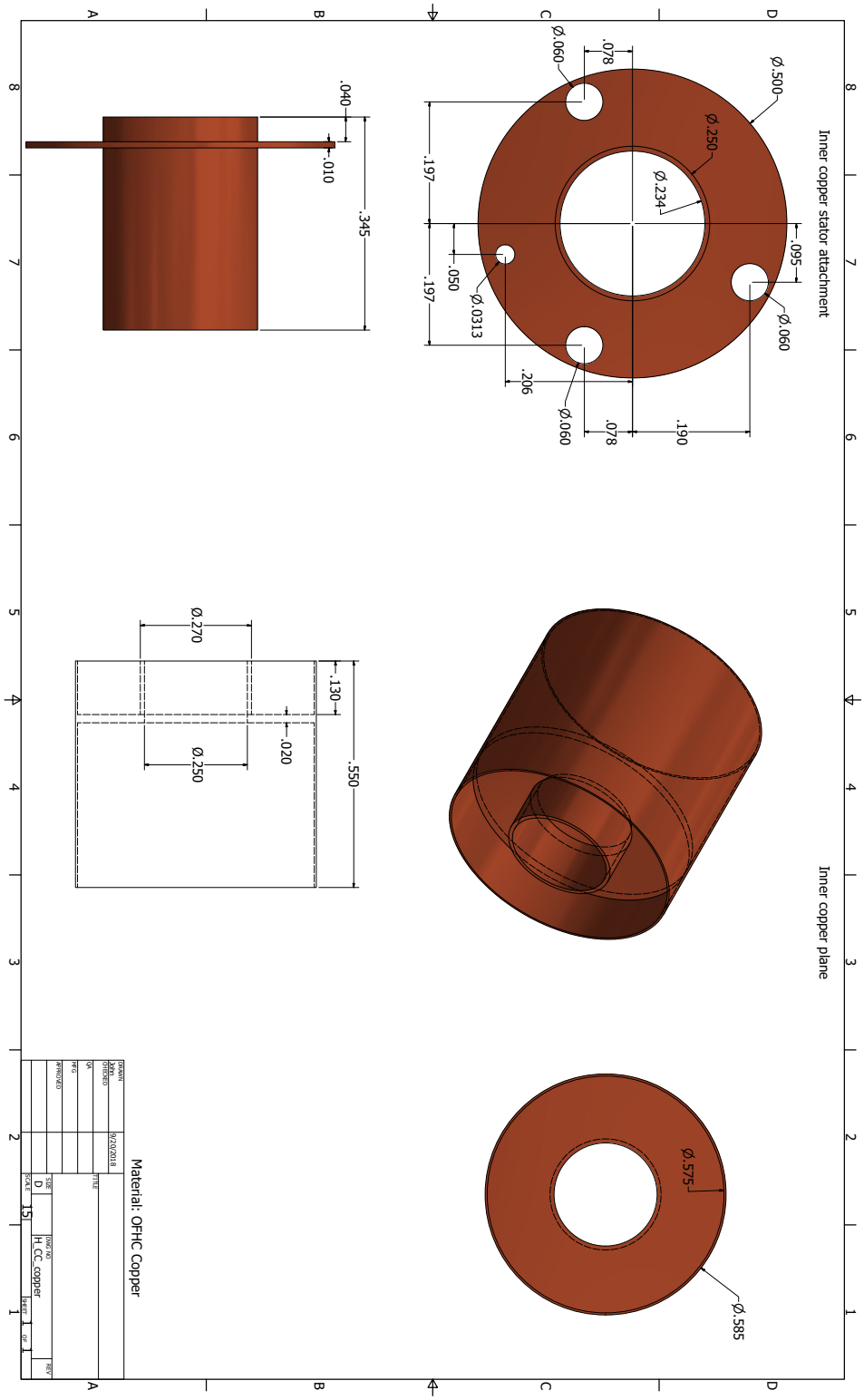


Figure A.29: The copper attachment to the stator and the inner copper plane of the capacitor for the capacitive coupling on the ^{1}H side.

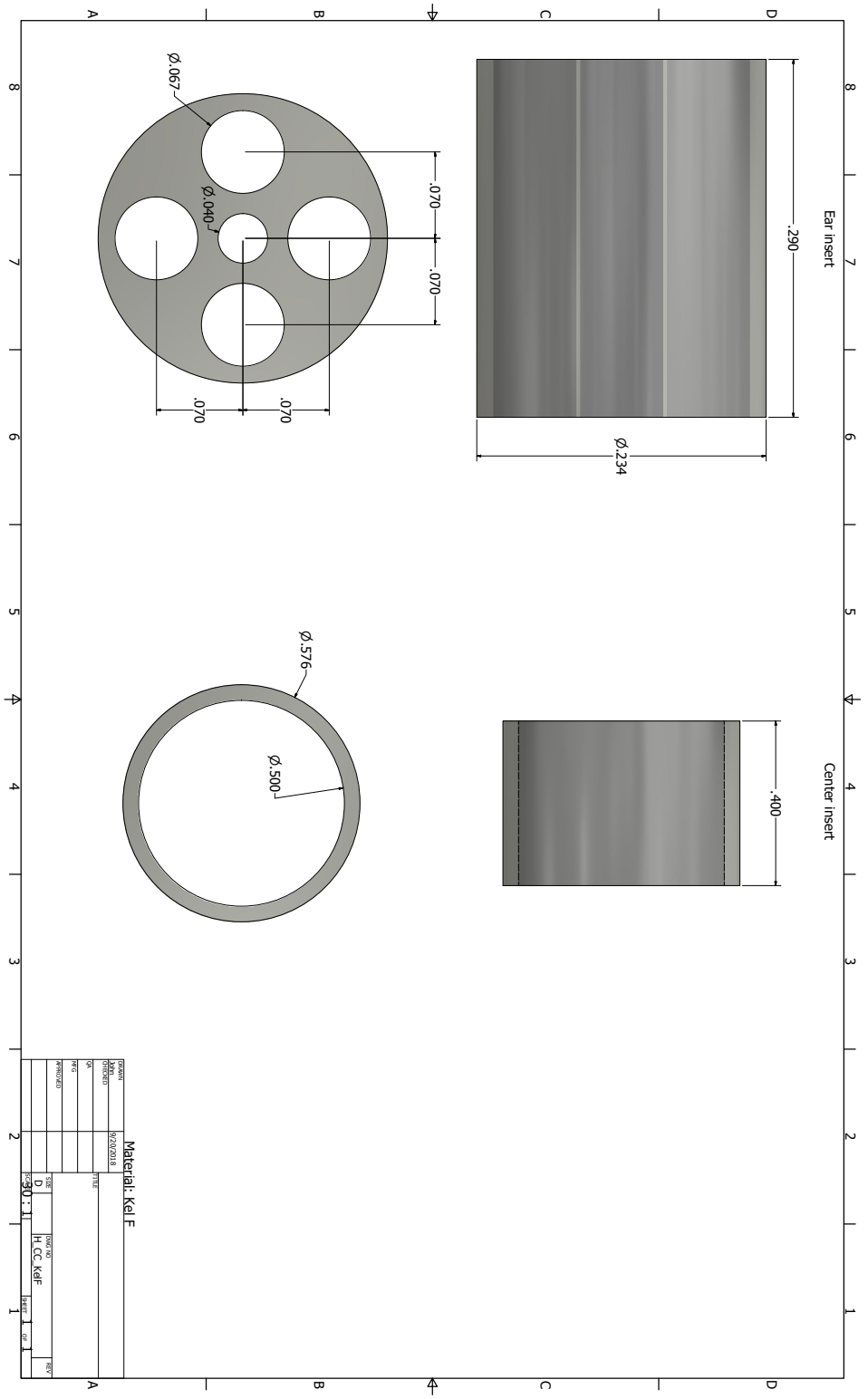


Figure A.30: The Kel F pieces that support the inner ear and inner plane for the capacitive coupling on the ^{1}H side.

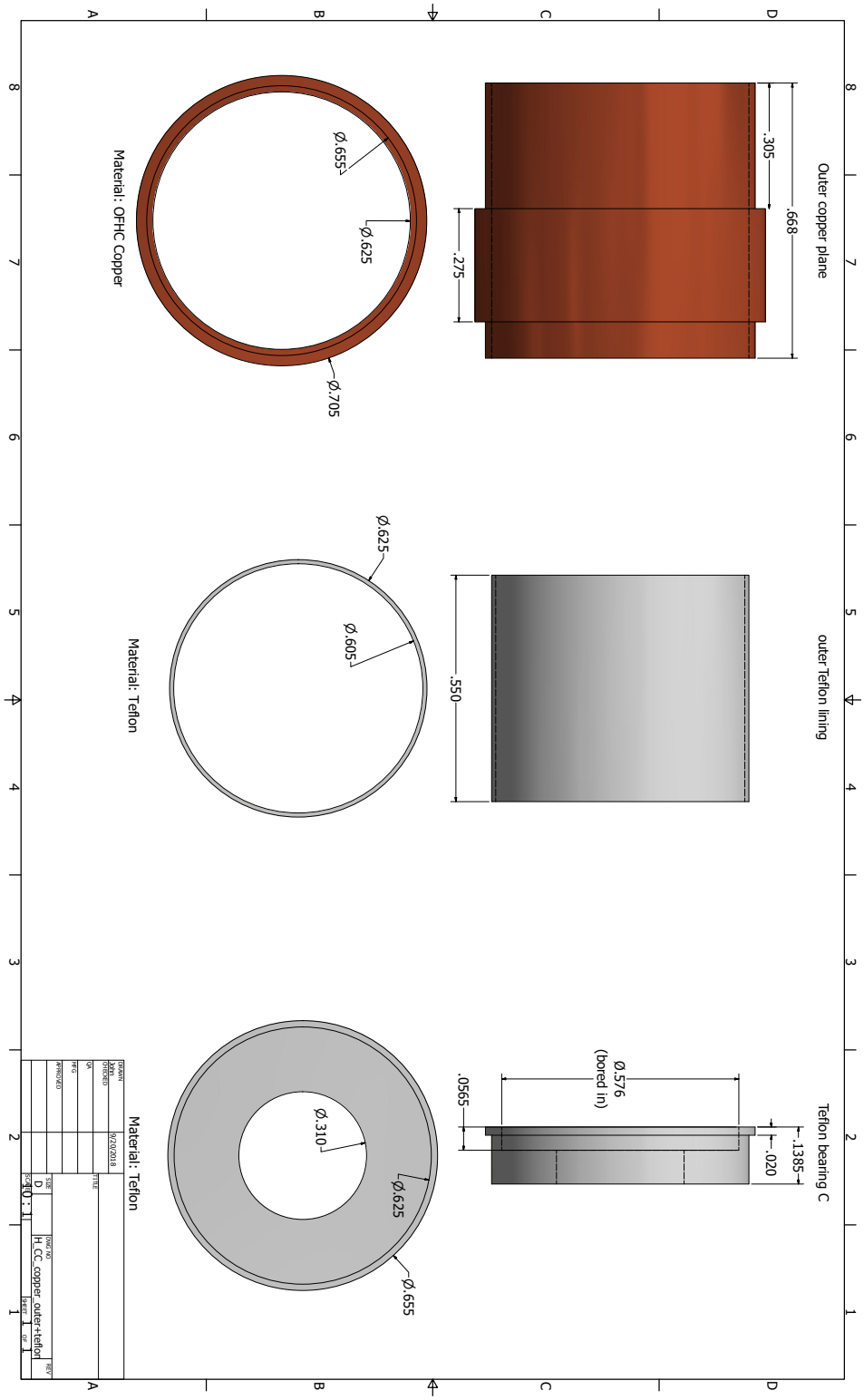


Figure A.31: The outer copper plane of the capacitor, the outer Teflon lining, and Teflon bearing C for the capacitive coupling on the ^{1}H side.

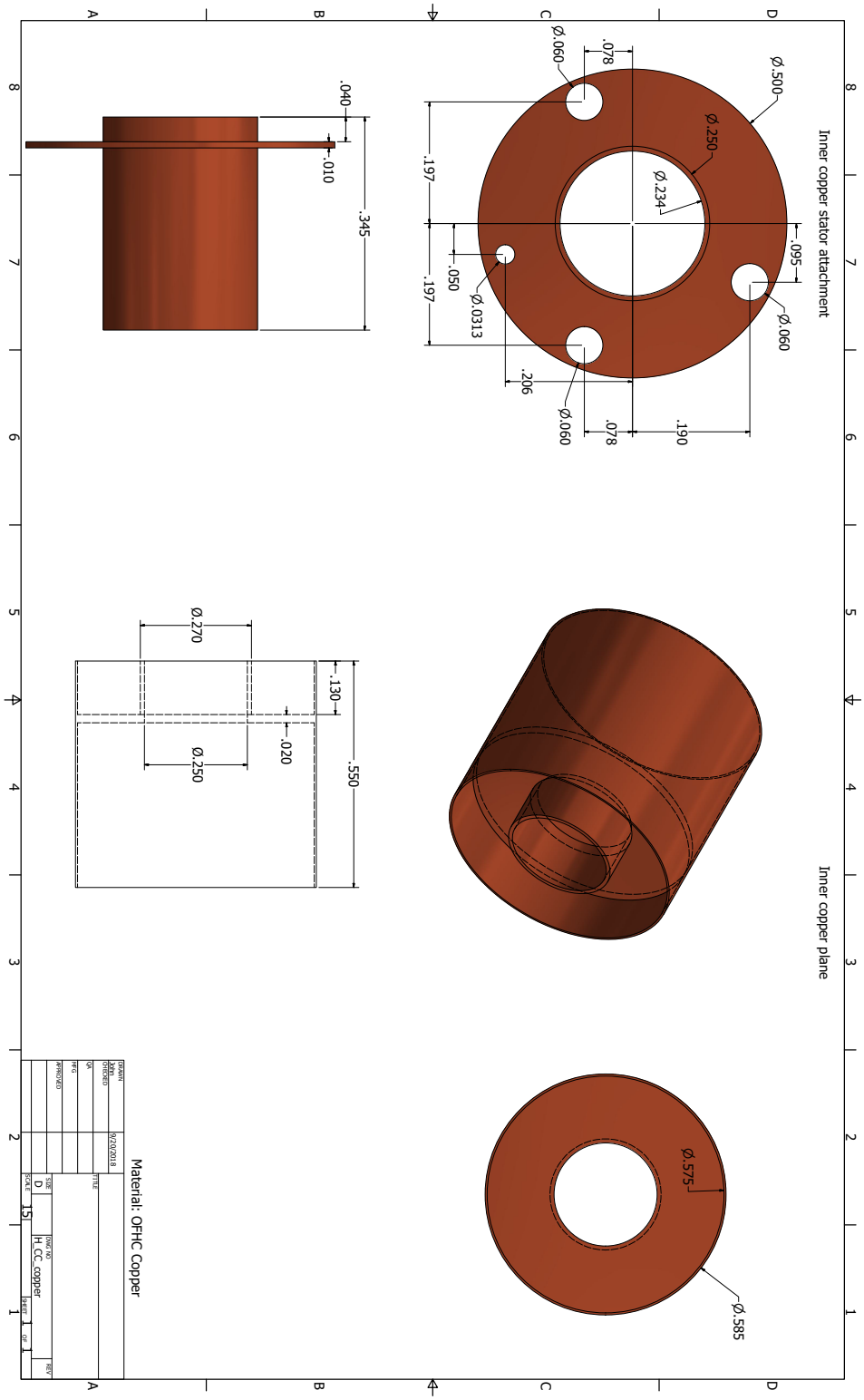


Figure A.32: The Teflon bearings A and B, and the inner Teflon lining for the capacitive coupling on the ^{1}H side.

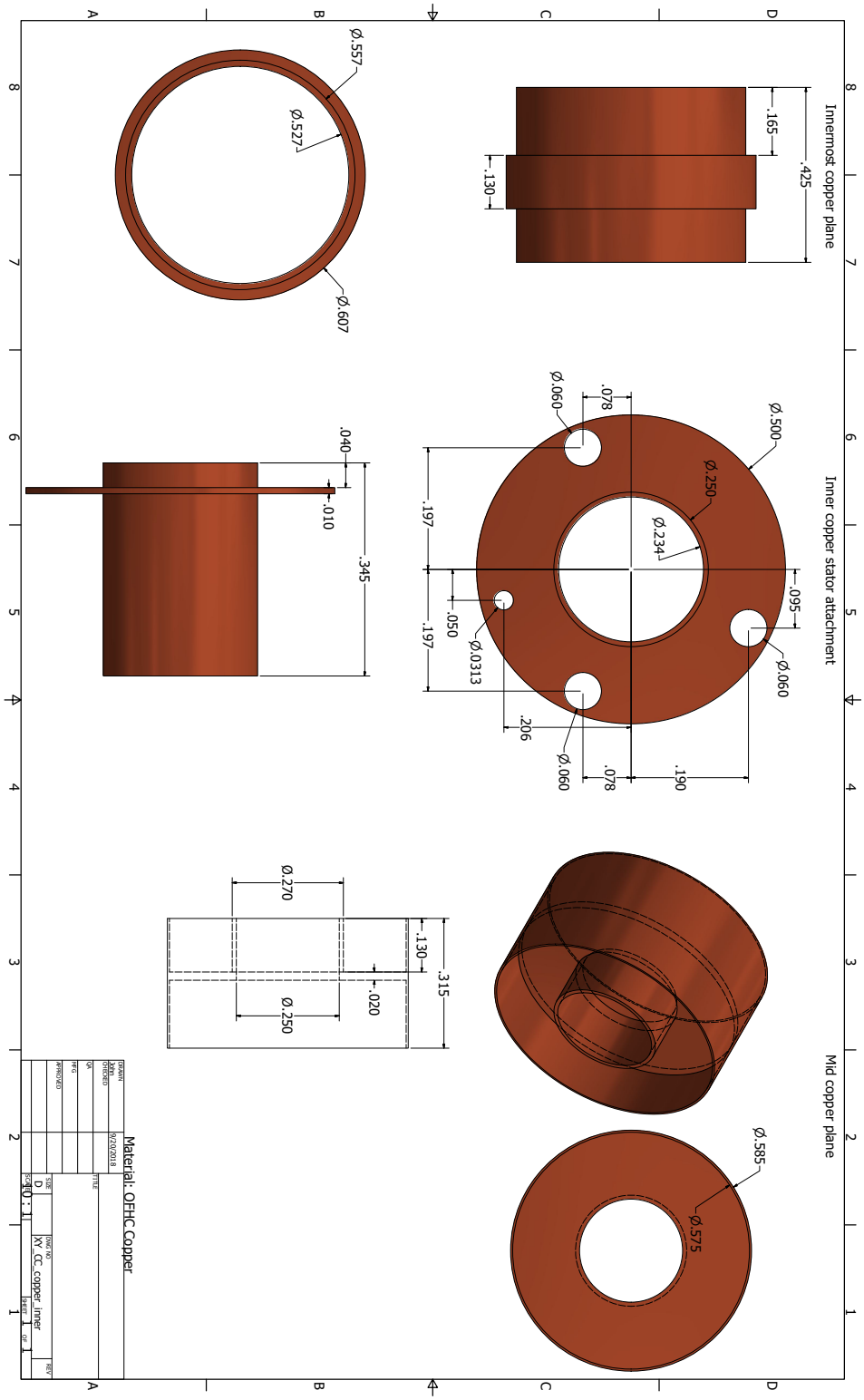


Figure A.33: The innermost copper plan, the copper attachment to the stator and the middle copper plane of the capacitors for the capacitive coupling on the ^{13}C / ^{15}N side.

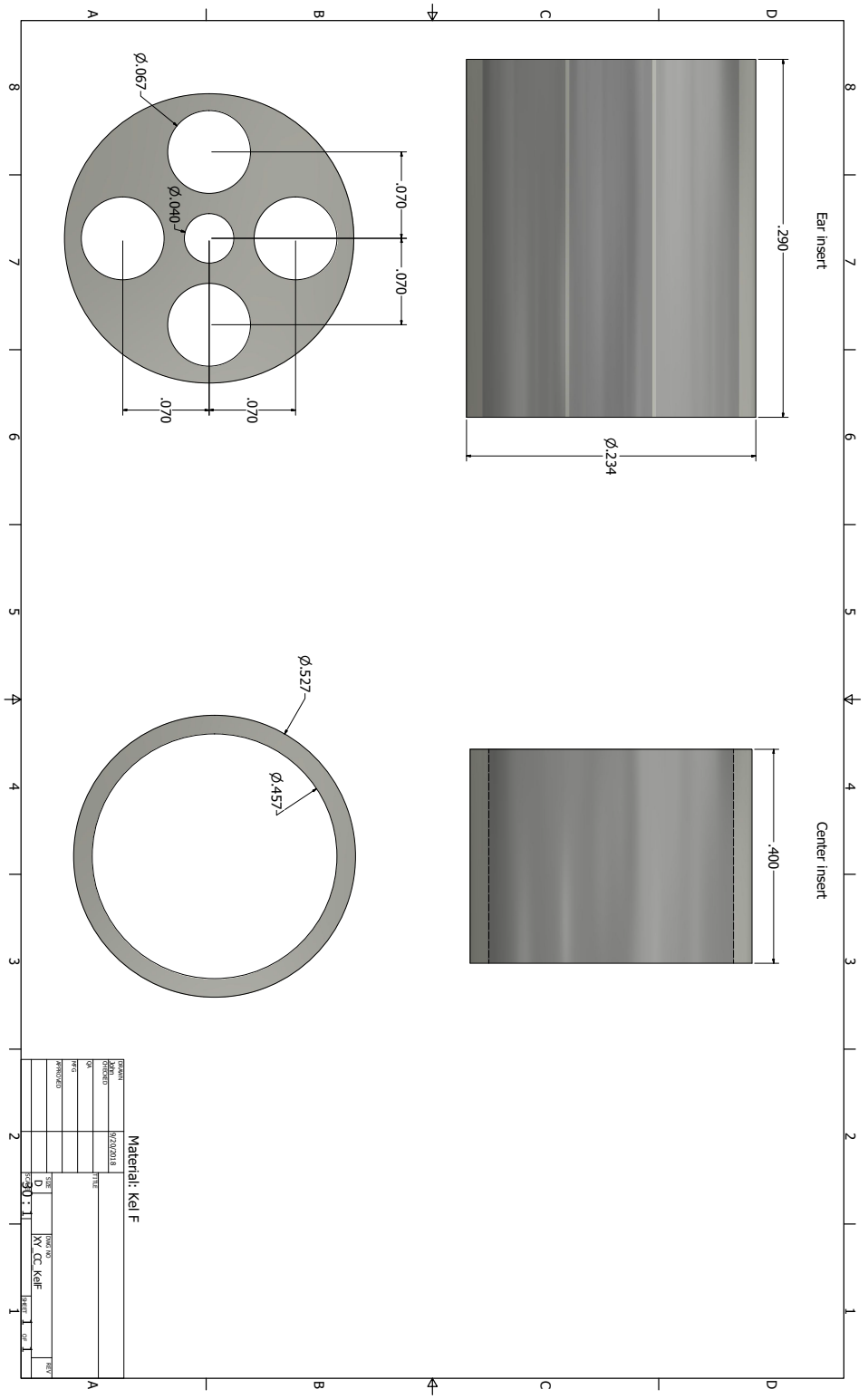


Figure A.34: The Kel F pieces that support the inner ear and innermost plane for the capacitive coupling on the ^{13}C / ^{15}N side.

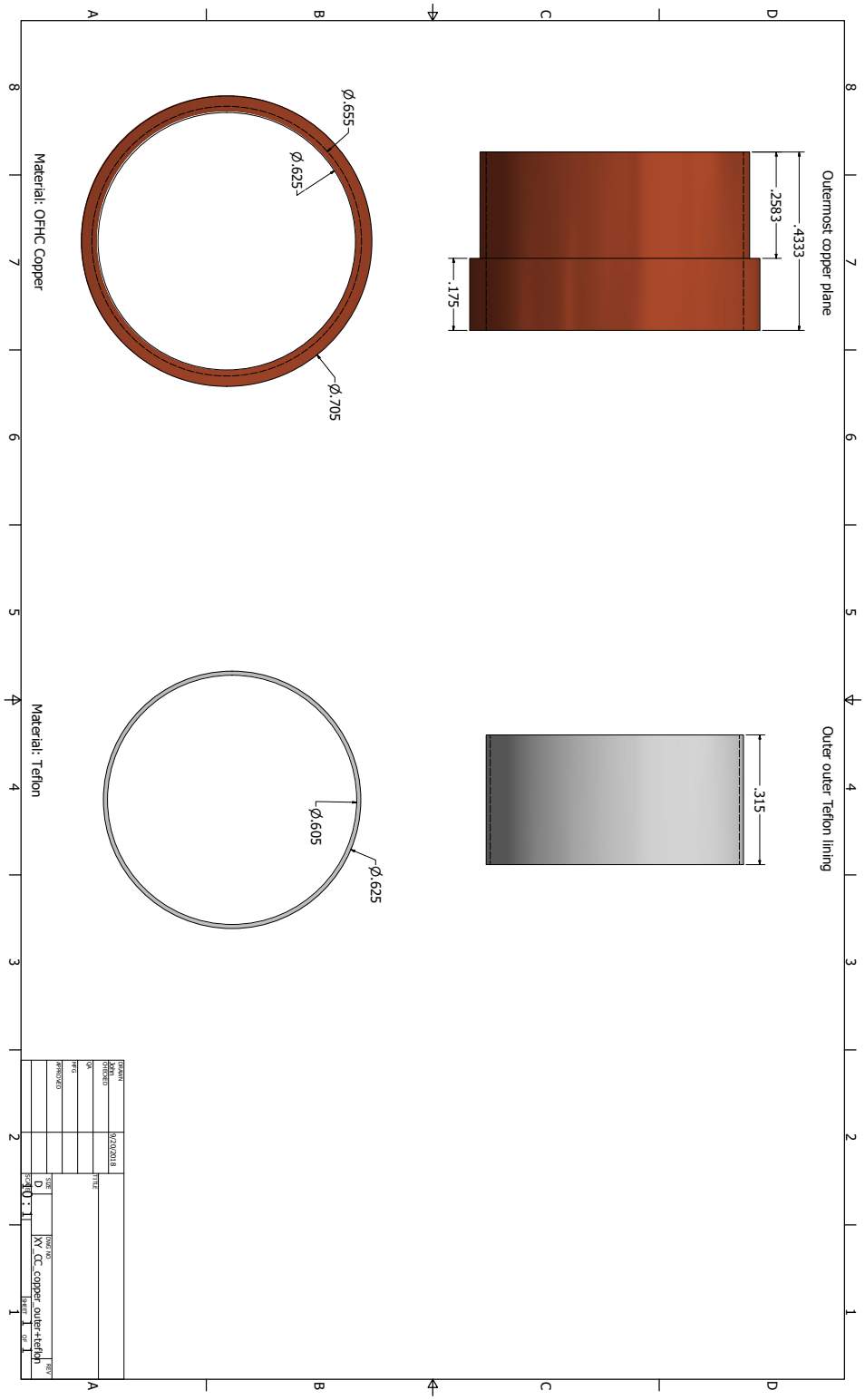


Figure A.35: The outermost copper plane of the capacitors and the outer Teflon lining of the outer capacitor for the capacitive coupling on the ^{13}C / ^{15}N side.

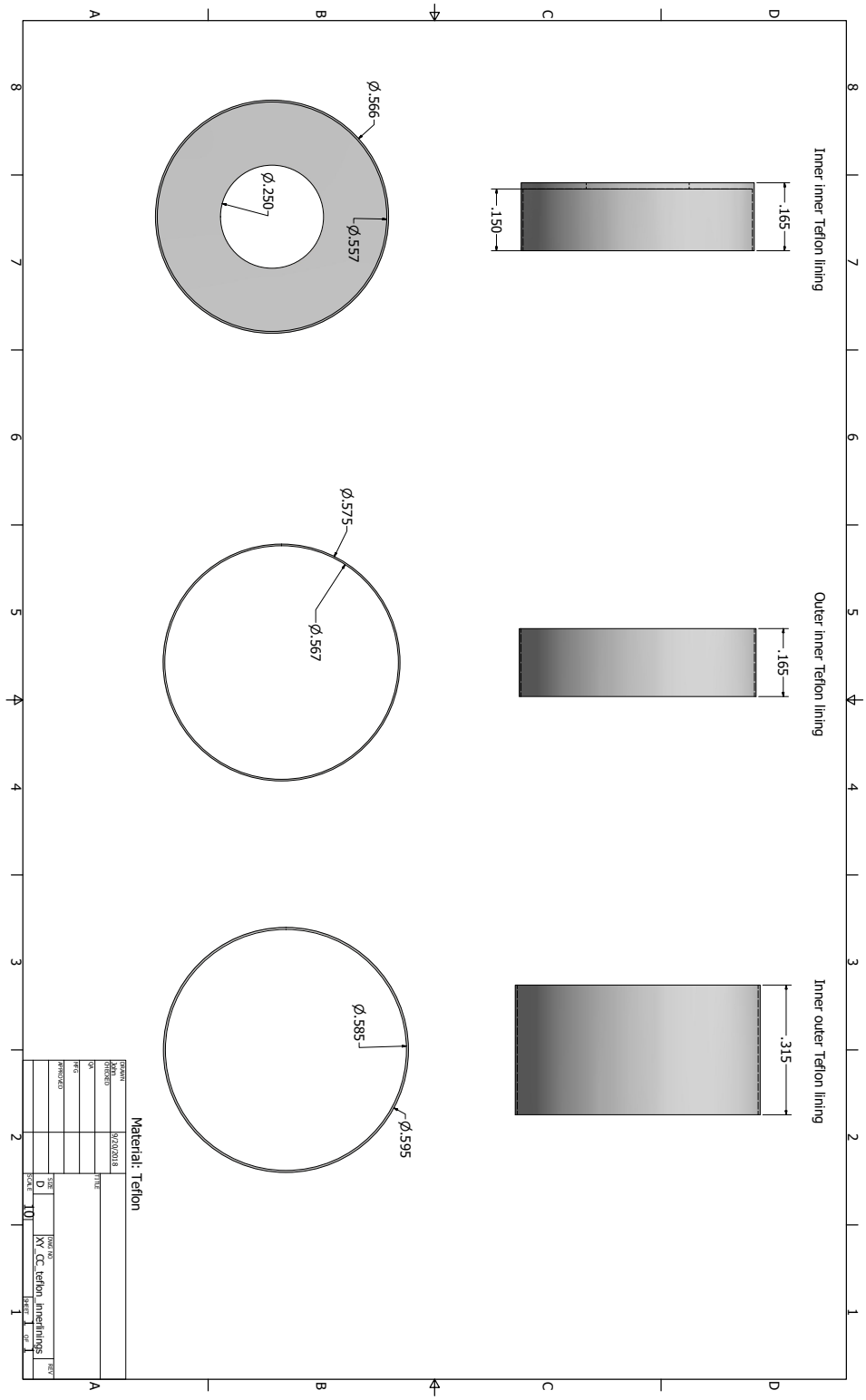


Figure A.36: The inner Teflon lining of the inner capacitor, the outer Teflon lining of the inner capacitor, and the inner Teflon lining of the outer capacitor for the capacitive coupling on the ^{13}C / ^{15}N side.

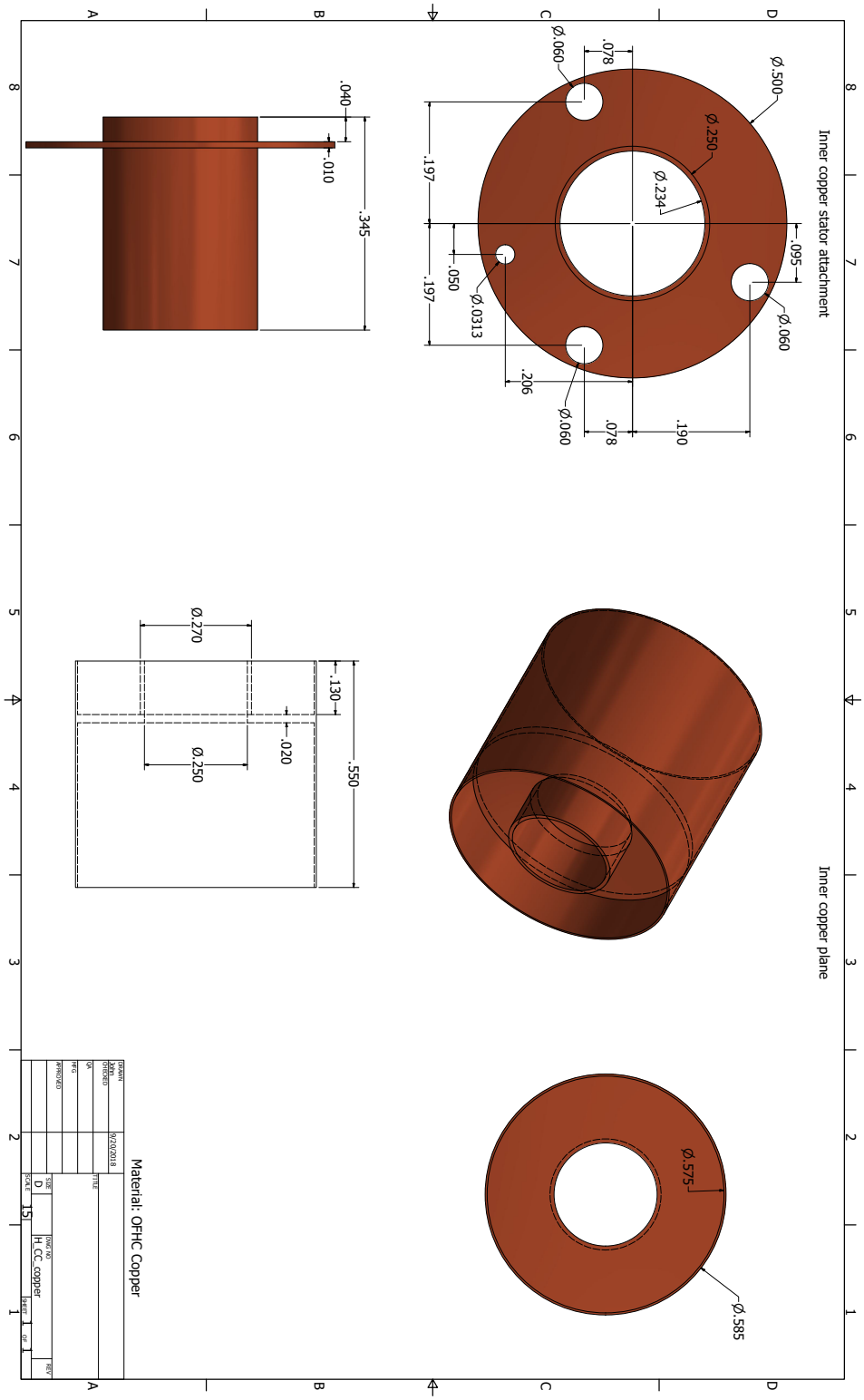


Figure A.37: The Teflon bearings A, B, and C for the capacitive coupling on the ^{13}C / ^{15}N side.

Appendix B

Alignment Script

B.1 Plain Alignment (PC Version)

Listing B.1: Alignment script that takes the default .clustal output and rearranges it into lines of 100 characters with a space every 10 for easier reading. It also renames the proteins to match their UniProt or PDB identifiers and identifies the proteins from the *Drosera capensis* as DCAP_####.

```
#Code to change the length of sequence displayed on each line using for a sequence alignment file

from string import ascii_letters


####### DATA IN FILE GOES HERE ############
alignedFile = "C:\\Users\\Username\Desktop\\test.clustal"
##########################################


####### DATA OUT FILE GOES HERE ############
outFile = 'C:\\Users\\Username\Desktop\\test_pretty.rtf'
##########################################


with open(alignedFile, 'r') as infile:
    DataIn = infile.readlines()

ID_dict = list()   #A list for all of the sequences
Full_ID_List=list()   #A list for all of the headers that identify the sequences
counter = 0        #Using a counter to index the location of the header and sequence in both lists
```

```
Key_Ref = {} #Have this to quickly check to see if the entry is new or has already
    been created

for line in DataIn:
   if len(line) != 1:
      line=line.strip()
      item=line.split()
      proteinName = "ERROR"

      if len(item) == 2:

         if all(c in ascii_letters+'-' for c in item[1]):


            #Going to change the name if it matches the "AKJ", "sp|", "gi|", "tr|"
                identifier
            identifier_list = ["sp|", "gi|", "tr|"]

            if item[0][0:3] == "AKJ":
               proteinName = "DCAP_" + item[0][9:13]
            elif item[0][0:3] in identifier_list:
               tempHolder = item[0].split("|")
               proteinName = tempHolder[2]
            elif "|" in item[0]:
               if item[0].split("|")[1] == "PDBID":
                  proteinName = item[0][0:4]
            else:
               proteinName = item[0]


            if proteinName not in Full_ID_List:

               ID_dict.insert(counter, item[1])
               Full_ID_List.insert(counter, proteinName)
               Key_Ref[proteinName] = counter
               counter = counter + 1

            else:
               temp = Key_Ref[proteinName]
               ID_dict[temp] = ID_dict[temp] + item[1]


#Make sure all the strings are the same length
max = 0
for x in range(0, counter):
   if len(ID_dict[x]) > max:
      max = len(ID_dict[x])

for x in range(0, counter):
   if len(ID_dict[x]) < max:
      addSpaceLen = max - len(ID_dict[x])
      addSpace = ' ' * addSpaceLen
      ID_dict[x] = ID_dict[x] + addSpace


#Make sure all the strings of the IDs are the same length
max = 0
for x in range(0, counter):
   if len(Full_ID_List[x]) > max:
```

```
      max = len(Full_ID_List[x])

for x in range(0, counter):
   if len(Full_ID_List[x]) < max:
      addSpaceLen = max - len(Full_ID_List[x])
      addSpace = ' ' * addSpaceLen
      Full_ID_List[x] = Full_ID_List[x] + addSpace


f = open(outFile, 'w')

Running = True  #Here to make sure we run through the entire string
while Running:


   if len(ID_dict[0]) <= 100:
      Running = False

   for x in range(0, counter):

      printSeq = ID_dict[x]
      keepSeq = ID_dict[x]

      printSeq = printSeq[0:100]
      keepSeq = keepSeq[100:len(ID_dict[x])]

      #Now add spaces between every 10 characters of the print sequences
      printSeqWithSpaces = ""
      while len(printSeq) > 10:
         printSeqWithSpaces = printSeqWithSpaces + printSeq[0:10] + " "
         printSeq = printSeq[10:len(printSeq)]
      printSeqWithSpaces = printSeqWithSpaces + printSeq

      ID_dict[x] = keepSeq

      f.write(Full_ID_List[x] + "\t" + printSeqWithSpaces + "\n")

   f.write("\n")
```

B.2 Alignment with Colors (PC Version)

Listing B.2: Alignment script that in addition to the formatting from the previous script, also colors the the amino acids by type as follows: hydrophobic residues, A, F, M, L, I, V, W, are Green; negatively charged residues, D, E, are Red; positively charged residues, R, K, are Cyan; cysteines, C, are Yellow and Bold; all other residues are Black.

```
#Code to change the length of sequence displayed on each line using for a sequence
    alignment file and color code amino acids

from string import ascii_letters
```

```
####### DATA IN FILE GOES HERE ############
alignedFile = "C:\\Users\\Username\Desktop\\test.clustal"
##########################################


####### DATA OUT FILE GOES HERE ############
outFile = 'C:\\Users\\Username\Desktop\\test_pretty.rtf'
##########################################


with open(alignedFile, 'r') as infile:
    DataIn = infile.readlines()

ID_dict = list()   #A list for all of the sequences
Full_ID_List=list()   #A list for all of the headers that identify the sequences
counter = 0        #Using a counter to index the location of the header and
    sequence in both lists

Key_Ref = {} #Have this to quickly check to see if the entry is new or has already
     been created

for line in DataIn:
   if len(line) != 1:
      line=line.strip()
      item=line.split()
      proteinName = "ERROR"

      if len(item) == 2:

         if all(c in ascii_letters+'-' for c in item[1]):


            #Going to change the name if it matches the "AKJ", "sp|", "gi|", "tr|"
                identifier
            identifier_list = ["sp|", "gi|", "tr|"]

            if item[0][0:3] == "AKJ":
               proteinName = "DCAP_" + item[0][9:13]
            elif item[0][0:3] in identifier_list:
               tempHolder = item[0].split("|")
               proteinName = tempHolder[2]
            elif "|" in item[0]:
               if item[0].split("|")[1] == "PDBID":
                  proteinName = item[0][0:6]
            else:
               proteinName = item[0]


            if proteinName not in Full_ID_List:

               ID_dict.insert(counter, item[1])
               Full_ID_List.insert(counter, proteinName)
               Key_Ref[proteinName] = counter
               counter = counter + 1

            else:
               temp = Key_Ref[proteinName]
```

```
            ID_dict[temp] = ID_dict[temp] + item[1]

#Make sure all the strings are the same length
max = 0
for x in range(0, counter):
   if len(ID_dict[x]) > max:
      max = len(ID_dict[x])

for x in range(0, counter):
   if len(ID_dict[x]) < max:
      addSpaceLen = max - len(ID_dict[x])
      addSpace = ' ' * addSpaceLen
      ID_dict[x] = ID_dict[x] + addSpace


#Make sure all the strings of the IDs are the same length
max = 0
for x in range(0, counter):
   if len(Full_ID_List[x]) > max:
      max = len(Full_ID_List[x])

for x in range(0, counter):
   if len(Full_ID_List[x]) < max:
      addSpaceLen = max - len(Full_ID_List[x])
      addSpace = ' ' * addSpaceLen
      Full_ID_List[x] = Full_ID_List[x] + addSpace


f = open(outFile, 'w')

#print the lines that need to go at the head of every file
f.write("{\\rtf1\\ansi\deff0 {\\fonttbl {\\f0 Courier New;}}" + "\n")
f.write("{\colortbl;\\red0\green0\\blue0;\\red0\green255\\blue0;\\red255\green0\\
    blue0;\\red0\green255\\blue255;\\red255\green255\\blue0;}" + "\n")
f.write("\landscape" + "\n")
f.write("\paperw15840\paperh12240\margl1440\margr1440\margt1440\margb1440" + "\n")
f.write("\\fs14" + "\n")

#define the relation of colors to letters
color_dict = {" ":" ", "-":"\cf1 -", "*":"\cf1 *", "G":"\cf1 G", "H":"\cf1 H", "N":
    "\cf1 N", "P":"\cf1 P", "Q":"\cf1 Q", "S":"\cf1 S", "T":"\cf1 T", "Y":"\cf1 Y",
        "A":"\cf2 A", "F":"\cf2 F", "M":"\cf2 M", "L":"\cf2 L", "I":"\cf2 I", "V":
           "\cf2 V", "W":"\cf2 W", "D":"\cf3 D", "E":"\cf3 E", "R":"\cf4 R", "K":"
           \cf4 K", "C":"\cf5 \\b C\\b0 "}

Running = True  #Here to make sure we run through the entire string
while Running:


   if len(ID_dict[0]) <= 100:
      Running = False

   for x in range(0, counter):

      printSeq = ID_dict[x]
      keepSeq = ID_dict[x]

      printSeq = printSeq[0:100]
```

```
      keepSeq = keepSeq[100:len(ID_dict[x])]

      #Now add spaces between every 10 characters of the print sequences
      printSeqWithSpaces = ""
      while len(printSeq) > 10:
         printSeqWithSpaces = printSeqWithSpaces + printSeq[0:10] + " "
         printSeq = printSeq[10:len(printSeq)]
      printSeqWithSpaces = printSeqWithSpaces + printSeq

      ID_dict[x] = keepSeq

      color_printSeqWithSpaces = ""
      for z in range(0, len(printSeqWithSpaces)):
         val = printSeqWithSpaces[z]
         color_printSeqWithSpaces = color_printSeqWithSpaces + color_dict[val]
         print(val + "\t" + color_printSeqWithSpaces)


      f.write("\cf1" + Full_ID_List[x] + "\t" + color_printSeqWithSpaces + "\line"
          + "\n")

   f.write("\line" + "\n")
f.write("}" + "\n")
```

B.3 Alignments (Mac Version)

Listing B.3: The only change from the PC to the Mac versions of these scripts is the file path to read from and save to.

```
####### DATA IN FILE GOES HERE #############
alignedFile = "/Users/Username/Desktop/test.clustal"
##########################################


####### DATA OUT FILE GOES HERE #############
outFile = '/Users/Username/Desktop/test_pretty.rtf'
##########################################
```