# CHEM 4700

# Advanced Biochemistry Laboratory

**CHEM 4700 Advanced Biochemistry Laboratory (3)L** (Formerly 002.470) (Lab required) A laboratory and workshop consisting of lectures, problem solving, and advanced instrumental techniques such as magnetic resonance spectroscopy, mass spectrometry, circular dichroism, x-ray crystallography, fluorescence spectroscopy and computer analysis of protein sequences. This course is required for all final year Honours students in Biochemistry. Prerequisite or concurrent requirement: CHEM 4620 or CHEM 4621; and CHEM 4630 or CHEM 4631.

## Announcements

**Please** read this important notice from the Faculty of Science: Registration Advisory.

Note also that Wednesday September 21 is the last day for adding courses and VW day is Friday March 16.

## Course Outline 2011-12

### Times:

- Tuesdays 2:30-5:30
- Please note that there is no laboratory during the first week of classes.

### Locations:

See the schedule below:

- CHEM 4700 Laboratory Schedule

### Instructors:

Dr. Mazdak Khajehpour
Room 468 Parker
Phone: 272-1546
E-mail:khajehpo@cc.umanitoba.ca

Dr. Brian Mark
Room 420 Buller Building
Phone: 480-1430
E-mail: brian_mark@umanitoba.ca

Dr. Sean McKenna
Room 380 Parker Building
Phone: 272-1562

Dr. Joe O'Neil
Room 390 Parker Building
Phone: 474-6697

E-mail: mckenna@cc.umanitoba.ca          E-mail: joneil@cc.umanitoba.ca

Dr. Hélène Perreault                      Dr. Jörg Stetefeld
Room 550 Parker                           Room 426 Parker
Phone: 474-7418                           Phone: 474-9731
E-mail:perreau@cc.umanitoba.ca            E-mail:stetefel@cc.umanitoba.ca

## Laboratory Manual:

- CHEM 4700 Laboratory Manual

## Recommended References:

- *Principles of Physical Biochemistry 2nd Ed.* by K. van Holde, W. C. Johnson, P S. Ho 2006, Pearson / Prentice Hall QP517.P49V36.

- *Biophysical Chemistry Part 2* R. Cantor & P. R. Schimmel Freeman & Co. QH345.C36

- *Instrumental Methods of Analysis* H. Willard *et al*. 1988 QD79.I5I52

- In my office is a collection of articles which provide some of the theory behind the experiments we will be carrying out. You are welcome to borrow this from me.

- Copies of previous years' examinations are available on-line at: Online Examinations, and below. Working through them will help prepare you for the final examination.

| | | |
|---|---|---|
| 1991 Final | 1992 Final | 1993 Final |
| 1994 Final | 1995 Final | 1996 Final |
| 1997 Final | 1998 Final | 1999 Final |
| 2000 Final | 2001 Final | 2002 Final |
| 2003 Final | 2004 Final | 2005 Final |
| 2006 Final | 2007 Final | 2008 Final |
| 2009 Final | 2010 Final | |

## Evaluation:

| | |
|---|---|
| Lab reports and problem sets 5% each: | 45% |
| Research Proposal: | 20% |
| Final Exam-scheduled by Student Records (3 hours): | 35% |

## Lab Report Format:

Lab reports should be written in the style of a research paper published in a scientific journal, *e.g.* Biochemistry. Typically, your lab report will be subdivided into Introduction, Materials and Methods, Results, Discussion, and References. **Don't recopy the lab handouts in your lab report.**

## Final Exam:

Examinable material includes that presented in 4700 workshops, and experiments, including handout material and problem sets, <u>and</u> closely related material presented in the 4630 lectures; for example, the lectures on circular dichroism, NMR, mass spectrometry *etc*. Some problem set material <u>will</u> appear on the final exam.

## Policy:

- Unless a medical certificate is presented attendance at all labs is <u>mandatory</u>. Lab reports will not be accepted from anyone not attending a lab.

- Lab reports are due 3 weeks after the day the experiment was completed.

- Late penalty is 20%.

- Research proposal (see below) due: Tuesday March 13.

- Penalty for late papers is 10% <u>per</u> <u>day</u>.

---

## Final Exam Review

Students in the Faculty of Science are permitted to review their final exams before the deadline for appealing final grades ([Final Grade Appeal](#)). If you wish to view your final exam please go to the Department of Chemistry general office (360 Parker Building), fill in an application form, and pay the $5.00 fee.

---

**Academic Dishonesty**: Please visit the Faculty of Science web site [Cheating, Plagiarism etc.](#)

---

## Research Proposal:

Due Tuesday March 13.

Read the paper by Oloo *et al*., (**2011** *J. Biol. Chem. 286*, 12133 - 40) entitled "*Structure-guided antigen engineering yields pneumolysin mutants suitable for vaccination against pneumococcal disease.*" Then, suppose you are preparing an application to a granting agency. Write a research proposal describing how you would extend the research described in this paper. You will notice that the authors apply a number of methods that are not covered in our course. You might propose to apply one or more of their methods to a different protein, or you might prefer to apply one of the methods used in our course to Pneumolysin or another protein. As long as there is some connection to the paper by Oloo *et al*., any proposal will be acceptable. Your approach may be pure, research driven aimed at learning more about a protein, or it might be more applied, with a human health goal similar to the approach of Oloo and colleagues. It is a good idea to state a formal hypothesis and then explain how you will test your hypothesis rather than simply proposing to collect data. Don't restrict yourself to using only the methods used by Oloo *et al*. Remember, your experiments should be scientifically sound. That does not mean you must be sure they would work; research projects often do not work! You will be judged mainly on the scientific merit of your proposal but organization, spelling, and grammar will also count.

Download the Oloo paper here: [Oloo *et al*. Paper](#)

## Length:

2,000-3,000 words, typed or very neatly handwritten. Do not write a chapter for a book; most grant application forms provide only limited space for a proposal. Diagrams are usually helpful.

**References:**

Use the format of one of the above papers; put references at the end and refer properly in the text.

---

**Useful WWW Sites for CHEM 4700:**

On-line Medical Dictionary

On-line Webster Dictionary

Material Safety Data Sheets

---

Return to the Chemistry Department Course Descriptions

Return to Joe O'Neil's Home Page

---

Last update on 09/06/2011

http://home.cc.umanitoba.ca/~joneil/CHEM4700.Course.outline.htm

Maintained by J. O'Neil

---

**9/6/2011**
**4:44 P.M.**
Free JavaScripts provided
by The JavaScript Source

---

**University of Manitoba**
**Chemistry**

---

# Fall Schedule for CHEM 4700 Advanced Biochemistry Laboratory

| Date | Course Intro & NMR Workshop J. O'Neil Room 539 Parker | Protein Purification Workshop S. McKenna Room 539 Parker | Calmodulin Purification J. O'Neil E. Dzananovic Room 562 Parker | Circular Dichroism: Nucleic Acids J. O'Neil E. Dzananovic Room 386 Parker | Circular Dichroism: Proteins J. O'Neil E. Dzananovic Room 386 Parker | H-Exchange Workshop J. O'Neil Room 539 Parker |
|---|---|---|---|---|---|---|
| Sept 13 | All | - | - | - | - | - |
| Sept 20 | - | All | - | - | - | - |
| Sept 27 | - | - | All | - | - | - |
| Oct 4 | - | - | All | - | - | - |
| Oct 11 | - | - | - | 3,6 | - | - |
| Oct 18 | - | - | - | 4,7 | - | - |
| Oct 25 | - | - | - | 1,2,5 | - | - |
| Nov 1 | - | - | - | - | 1,3 | - |
| Nov 8 | - | - | - | - | 2,7 | - |
| Nov 15 | - | - | - | - | 4,5,6 | - |
| Nov 22 | - | - | - | - | - | All |
| Nov 29 | - | - | - | - | - | - |

*Please bring a hardcover laboratory notebook to each workshop and laboratory.*

1 = Hartmann, Riley J.       4 = McDougall, Matthew D.       7 = Wang, Zhinan       10 =

2 = Liao, Catherine R.       5 = Shpiruk, Tanner A.       8 =       11 =

3 = Martens, Michael R.       6 = Urena Ramirez, Viridiana       9 =

# Winter Schedule for CHEM 4700 Advanced Biochemistry Laboratory

| Date | H-Exchange *via* NMR J. O'Neil T.Wolowiec Rooms 392/338 Parker | Protein Fluorescence M. Khajehpour E. Dzananovic Room 122 Parker | X-ray Diffraction of Proteins B. Mark E. Dzananovic Room 118 Buller & Room 539 Parker | Applied Structural Biology Workshop J. Stetefeld Room 539 Parker |
|---|---|---|---|---|
| Jan 10 | 4,5 | - | - | - |
| Jan 17 | 2,6 | - | - | - |
| Jan 24 | 1,3,7 | - | - | - |
| Jan 31 | - | 1,6 | - | - |
| Feb 7 | - | 5,7 | - | - |
| Feb 14 | - | 2,3,4 | - | - |
| Feb 28 | - | - | All | - |
| Mar 6 | - | - | All | - |
| Mar 13 | - | - | - | All |
| Mar 20 | - | - | - | - |
| Mar 27 | - | - | - | - |

Last update on 09/06/2011

http://home.cc.umanitoba.ca/~joneil/CHEM4700/Labs.htm

**9/6/2011
3:49 P.M.**

# SAFETY IN THE LABORATORY

**The Chemistry 4700 laboratories operate within Workplace Hazardous Materials Information System (WHMIS) guidelines.  This requires that you are responsible for learning safe-handling techniques for controlled products.  Supervisory personnel may be consulted for information, and material safety data sheets (MSDS) are available.  If you have any concerns regarding safety, or procedures used in laboratories, you are urged to contact your laboratory supervisor, or Dr. Joe O'Neil  in Room 390 Parker Building, who will contact the Department Safety Committee and/or the University Occupational Health and Safety Office.**

WHMIS is Canada's right-to-know legislation developed after consultation among Government, Labour and Industry.  In Manitoba this legislation is enforced for students just as it is for anyone in a workplace setting.   As part of your *training* in Chemistry you must:
   -become familiar with and abide by all SAFETY regulations in our laboratories
   -learn about *supplier* and *workplace* LABELS and how to interpret the information on these labels
   -know where to obtain MSDS for substances used in your experiments and how to read these
   sheets.

1.  SAFETY - There is a list of precautions and instructions provided in your laboratory manual.  Your Laboratory Supervisor and/or T.A. will provide further information.

2.  SUPPLIER LABELS - Controlled products have labels of a prescribed design when sold.  The label will have **Product Identifier, Supplier Identifier, MSDS Reference, Hazard Symbols, Risk Phrases, Precautionary Measures, First Aid Measures.**  (See sample)

WHMIS border — — — — — — — Paint Stripper

Product Identifier

Hazard Symbols — — — —

Flammable liquid
Risk Phrases — — — — — — — May cause birth defects

Precautionary                    Keep away from source of ignition
Measures — — — — — — Avoid prolonged breathing of vapour
                                 Avoid skin contact

First Aid — — — — — — — If in contact with skin wash repeatedly
                               with soap and water

Reference to MSDS — — — — See Material Safety Data Sheet

Supplier Identifier — — — — — Sunny Chemical Company

Except for excluded products, all materials are classified and use the symbols illustrated below:

Classes and Divisions                                    Hazard Symbols

Class A - Compressed Gas
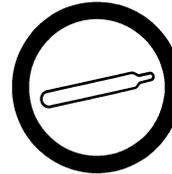


Class B - Flammable and Combustible Material



Class C - Oxidizing Material

Class D - Poisonous and Infectious Material                Hazard Symbols

   1. Materials Causing Immediate and Serious Toxic Effects



   2. Materials Causing Other Toxic Effects



   3. Biohazardous Infectious Material



Class E - Corrosive Material



Class F - Dangerously Reactive Material



WORKPLACE LABELS - For all chemicals not dispensed from the original supplier's container, a new label must be attached which includes **Identity of Product, Precautionary Measures** (usually simplified)**, and Reference to Availability of MSDS.**  Where the sample container is too small to attach a workplace label it must be coded and/or the rack (or other multiple sample container) must have an identifier label.

**WORKPLACE LABEL**

PRODUCT IDENTIFIER

SAFE HANDLING/PERSONAL PROTECTION

See MATERIAL SAFETY DATA SHEET

---

UNIVERSITY of MANITOBA

**PRODUCT IDENTIFIER:**

**PRECAUTIONARY MEASURES:**
- ☐ Avoid prolonged or repeated exposure
- ☐ Avoid any exposure – readily absorbs through skin
  - ☐ corrosive   ☐ skin, eye or lung irritant
  - ☐ toxic   ☐ carcinogen
  - ☐ Wear chemical safety goggles or face mask
  - ☐ Wear chemically resistant gloves
    - ☐ latex   ☐ vinyl   ☐_____
- ☐ Inhalant Hazard –
  - ☐ wear respirator
  - ☐ confine handling to a fume hood
- ☐ Flammable – avoid potential ignition sources
- ☐ Potential Explosive
- ☐ Contents under pressure   ☐ Asphyxiant

**FIRST AID**
- ☐ Contact with skin and eyes – Flush with water for 15 minutes then seek medical attention
- ☐ Inhalation – Remove to fresh air and seek medical attention
- ☐ Ingestion – Rinse mouth with water and seek medical attention

**REFER TO MATERIAL SAFETY DATA SHEET FOR MORE INFORMATION**

---

3. MSDS for **Chemistry 4700** are all available in **ROOM 392** Parker Chemistry Building. Instructions are posted on the wall.

# Structure-guided Antigen Engineering Yields Pneumolysin Mutants Suitable for Vaccination against Pneumococcal Disease[S]

**Eliud O. Oloo**[1,2]**, Jeremy A. Yethon**[1]**, Martina M. Ochs, Bruce Carpick, and Raymond Oomen**
*From sanofi pasteur, Toronto, Ontario M2R 3T4, Canada*

**Pneumolysin (PLY) is a cholesterol-binding, pore-forming protein toxin. It is an important virulence factor of *Streptococcus pneumoniae* and a key vaccine target against pneumococcal disease. We report a systematic structure-driven approach that solves a long-standing problem for vaccine development in this field: detoxification of PLY with retention of its antigenic integrity. Using three conformational restraint techniques, we rationally designed variants of PLY that lack hemolytic activity and yet induce neutralizing antibodies against the wild-type toxin. These results represent a key milestone toward a broad-spectrum protein-based pneumococcal vaccine and illustrate the value of structural knowledge in formulating effective strategies for antigen optimization.**

Infections by *Streptococcus pneumoniae* cause an estimated 1 million deaths worldwide, and the development of an accessible and universal vaccine against pneumococcal disease is a global health priority (1). Several efforts are currently under way to design and develop vaccines based on conserved *S. pneumoniae* protein antigens (2). In contrast to successes with small-molecule drugs, the structure-based design of vaccines is still in its infancy (3). This is understandable given that vaccine antigens are generally complex biological macromolecules. The role of structure-driven methods in vaccinology is poised to expand, however, owing to recent advances in experimental and computational structural biology (3). The vast wealth of information currently available in bioinformatics databases can be mined and applied to the prediction of unknown three-dimensional protein structures as well as the design of new proteins with novel functionality (4–6). Furthermore, emerging molecular dynamics simulation techniques have made it possible to venture beyond static conformational snapshots and into computational modeling of the functionally relevant motions of proteins in their natural environments (7, 8). As proof of concept that these tools are now sophisticated enough to enable vaccine antigen design, we have engineered the major cholesterol-dependent cytolysin of *S. pneumoniae* into a detoxified form that retains its protective antigenicity.

To date, vaccination strategies against *S. pneumoniae* have focused on its capsular polysaccharide as the source of protective antigen in vaccine formulations (9). However, this approach covers only a limited number of serotypes, and production issues make polysaccharide conjugates inaccessible in sufficient quantity to the population most in need in developing countries (10). As a result, research efforts are now being directed toward recombinant protein vaccines, which are cheaper, more robust, and easier to manufacture in comparison with conjugates and also have the potential to provide coverage against all serotypes (2, 10).

A leading candidate for such a protein-based vaccine is pneumolysin (PLY)[3]: it is an important virulence factor of *S. pneumoniae* (11, 12), is highly conserved across all serotypes (13), is immunogenic in humans (14, 15), and has shown protection in numerous animal models of infection (16–18). However, PLY is a toxin that oligomerizes to form large pores in the membranes of target host cells (19). The native protein is therefore undesirable for use as a human vaccine antigen. Less toxic mutant forms have been described but have generally retained some degree of hemolytic activity (20). Historically, the best characterized of these is PdB (PLY(W433F)), which possesses ~1% of the hemolytic activity of native PLY (21). More recently, completely inactive PLY mutants have been discovered by scanning mutagenesis (22); however, the mechanism of their attenuation is not well understood.

The objective of this study was to identify and introduce amino acid modifications that completely detoxify PLY by arresting its transitioning to the pore-forming state without compromising its ability to elicit neutralizing antibodies against the native toxin. To achieve this goal, we applied a targeted genetic detoxification approach based on insights into the toxin's putative three-dimensional structure and functional mechanics. The first step in the computational design process was the prediction of atomic resolution structures of PLY in its water-soluble, prepore, and pore-forming states. Contact maps and dynamic domain analyses were then used to formulate strategies for stabilization and detoxification. The resulting protein designs were cloned, purified, and validated experimentally by their ability to elicit antibodies that neutralized the wild-type toxin.

---

[3] The abbreviations used are: PLY, pneumolysin; PFO, perfringolysin O; DSC, differential scanning calorimetry; MACPF, membrane attack complex/perforin.

## EXPERIMENTAL PROCEDURES

*Construction of Structural Models*—A model of soluble wild-type PLY was constructed by comparative modeling using a published x-ray crystallographic three-dimensional structure of perfringolysin O (PFO) as the template (Protein Data Bank code 1PFO) (23). Template coordinates were obtained from the RSCB Protein Data Bank (24). Models of the prepore and pore-forming conformations were based on Cα traces of PFO fit into cryo-electron microscopy maps of prepore and pore-forming states of PLY (Protein Data Bank codes 2BK1 and 2BK2) (25). Computational modeling was performed using Discovery Studio® 2.1 (Accelrys). Sequence alignments were generated using the align multiple sequences protocol, and homology models were constructed with MODELLER (26). Twenty-five models of each conformation (soluble, prepore, and pore-forming) were predicted, and the best models of each category were selected based on the probability density function energy scores.

*Dynamic Domain Analyses*—Dynamic domain analyses were carried out using the Hingefind algorithm (27).

*Contact Map Calculation*—Contact matrices for the soluble and pore-forming conformations were independently generated using Discovery Studio® 2.1 with the Cα-to-Cα cutoff distance set at 7 Å. A superposition of the two contact maps was then obtained by plotting on common axes.

*Cloning*—The *ply* gene was PCR-cloned from the genome of *S. pneumoniae* strain R36A into the NdeI/XhoI sites of pET-28a (Novagen) for expression of His-tagged protein. Expression constructs for PLY mutants were created from this plasmid template by site-directed mutagenesis using the QuikChange Multi-Site-directed mutagenesis kit (Stratagene). All mutations were verified by sequencing.

*Protein Purification*—Recombinant His-tagged proteins were expressed in *Escherichia coli* BL21(DE3) grown in Overnight Express Autoinduction System 1 medium (Novagen) at 25 °C for 24 h. Protein from the soluble fraction was purified using $Ni^{2+}$-nitrilotriacetic acid-agarose (Qiagen) according to the manufacturer's instructions. Protein purity was assessed by SDS-PAGE (≥85% in all cases).

*Differential Scanning Calorimetry (DSC)*—DSC measurements were performed on a VP-Capillary DSC microcalorimeter with an autosampler (MicroCal, Northampton, MA). Protein samples were loaded onto a 96-well plate and then loaded automatically and sequentially into a 0.325-ml analysis cell. Samples were heated at a scan rate of 1 °C/min (temperature range of 20–80 °C) at a constant pressure of 42–44 p.s.i. The heat capacity profiles were recorded using VP Viewer Version 1.2 software. The transition temperature and enthalpy of transitions were determined using MicroCal Origin™ Version 7.0 software.

*CD Spectroscopy*—Samples were analyzed using a Jasco Model 810 CD spectropolarimeter. The CD spectra were recorded in the range of 170–280 nm using a 0.1-cm path length quartz cell. Spectra were obtained by scanning at 50 nm/min. Data were converted to molar ellipticity [θ] expressed in degrees cm²/dmol and plotted using the instrument-specific software.

*Hemolysis Assay*—Membranolytic activity was measured using an *in vitro* hemolysis assay. Purified PLY mutants were serially diluted (2-fold) in a round-bottom microtiter plate in Hanks' buffered saline (Invitrogen) supplemented with 0.5% bovine serum albumin (Sigma). Washed sheep red blood cells (Rockland Immunochemicals) were added to each well (1.5% (v/v) final concentration), and the plate was incubated for 30 min at 37 °C. Intact cells were removed by centrifugation, and supernatants were transferred to a flat-bottom plate to measure hemoglobin release by absorbance at 540 nm. Percent lysis values were graphed *versus* protein concentration (milligrams/milliliter), and specific activity was taken as the inverse of the concentration at which 50% erythrocyte lysis was achieved.

*Antiserum Generation*—Antisera were generated in rats using protocols approved by an internal Animal Care Committee. Animal handling adhered to all guidelines outlined by the Canadian Council on Animal Care. Antibody titers were assessed by ELISA using wild-type PLY as the coating antigen.

*Neutralization Assay*—Toxin neutralization was tested using an *in vitro* hemolysis inhibition assay. Because free cholesterol inhibits PLY activity (28), antisera were first depleted of cholesterol by precipitation with dextran sulfate (Dextralip® 50, Sigma). Antisera were then serially diluted (2-fold) across a round-bottom microtiter plate in Hanks' buffered saline supplemented with 0.5% bovine serum albumin. Wild-type PLY was added to each well (10 ng), and the plate was incubated for 30 min at 37 °C to allow antibody binding. Washed sheep red blood cells were then added to each well (0.75% final concentration), and the plate was incubated for another 30 min at 37 °C. Intact cells were removed by centrifugation, and supernatants were transferred to a flat-bottom plate to measure hemoglobin release by absorbance at 540 nm. Percent lysis values were graphed *versus* antiserum dilution, and the neutralizing titer was taken as the inverse of the dilution at which 50% inhibition of lysis was achieved.

## RESULTS

*Prediction of PLY Three-dimensional Structures*—Distinct conformations of PLY were modeled to analyze the series of inter- and intradomain rearrangements that drive its transformation from the water-soluble state to the membrane-associated prepore state and ultimately the membrane-inserted pore-forming state (29). Both low-resolution structures of PLY and high-resolution coordinates of its homolog, PFO, were used as templates to guide model construction (25, 30, 31). These models represent snapshots of key conformational states sampled by a molecule of PLY as it converts from a soluble monomer to an oligomeric transmembrane pore-forming structure (Fig. 1). The transition from the water-soluble monomer to the membrane-bound prepore oligomer is characterized by significant structural reorganization (25). One of the initial steps may involve modest deformation of the D2 domain and the twisting and bending of the D4 domain relative to D1–D3 (Fig. 1, *STEP 1*). The most remarkable internal structural rearrangements are those associated with the transformation of the prepore state to the membrane-inserted pore-forming conformation (Fig. 1, *STEP 2*). These rearrangements include crumpling of the D2 β-sheet, separation of D3 from D2, and uncoiling of two helical
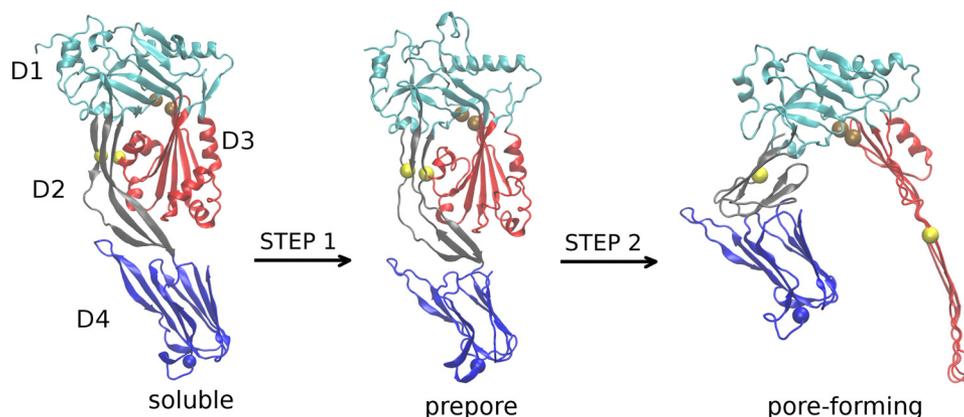
FIGURE 1. **Homology models of three distinct conformations of PLY illustrate a stepwise mechanistic model for the transformation of PLY from the water-soluble to the pore-forming conformation.** The structural domains of PLY are labeled *D1*, *D2*, *D3*, and *D4*, and colored *cyan*, *gray*, *red*, and *blue*, respectively. For clarity, oligomeric partners and the host cell membrane are excluded from the rendering of the prepore and pore-forming conformations. The locations of some amino acids pairs that were linked via disulfide bridging are shown as *spheres*. The *yellow spheres* represent amino acid pair Glu-159 and Thr-348, and the *brown spheres* represent Thr-65 and Gly-293. The *blue sphere* represents Cys-428. Rendering was done using VMD (47).

regions in D3 into an extended membrane-traversing $\beta$-sheet that constitutes the interior lining of the pore. Forty-four oligomeric partners, contributing a total of 176 $\beta$-strands, assemble into a pore measuring $\sim$260 Å in diameter (25).

*Analyses of Structural Models and Design of Detoxified Mutants*—The computationally modeled structures of various conformations of PLY were used to develop strategies for its detoxification. The hypothesis was that PLY could be rendered inactive by trapping it in the soluble or prepore conformation. The three main approaches adopted to prevent structural rearrangement of the toxin to the pore-forming state were the following: (*a*) introduction of conformation-trapping intramolecular disulfide bridges, (*b*) engineering of rigidity-conferring mutations at mechanical hinge regions, and (*c*) introduction of steric hindrance at conserved Gly residue positions.

*Conformational Trapping Using Disulfide Linkages*—This approach has been previously applied to structural studies on the related toxin PFO (32). It was shown that a mutant form of PFO featuring a disulfide linkage between domains 2 and 3 was able to assemble into a prepore complex but did not convert to the pore-forming state. Similarly, it was demonstrated that disulfide cross-linking of two $\beta$-stands in domain 3 of PFO prevented prepore oligomerization and membrane insertion even though membrane binding was not significantly affected (33).

The typical C$\alpha$-C$\alpha$ distance for disulfide-bridged cysteine residues in proteins is reported to be $\sim$6.5 Å (34, 35). To accommodate inherent protein flexibility, we used a slightly higher C$\alpha$-C$\alpha$ value of 7 Å to calculate contact matrices for both the soluble and pore-forming conformations of PLY (Fig. 2). These matrices enabled the identification of residue pairs that could potentially be linked by disulfide bridges. The suitability of residue pairs as candidates for disulfide-trapping mutations was further assessed by visual examination of the structural models. Particular attention was paid to residue pairs at interdomain boundaries based on the expectation that the introduction of interfacial disulfide bridges would prevent the relative displacement of domains. An important consideration, besides the magnitude of C$\alpha$-C$\alpha$ separation, was the directionality of C$\alpha$-C$\beta$ bond vectors. Residue pairs fulfilling one of the following criteria were targeted for mutation to Cys: (i) within the 7-Å
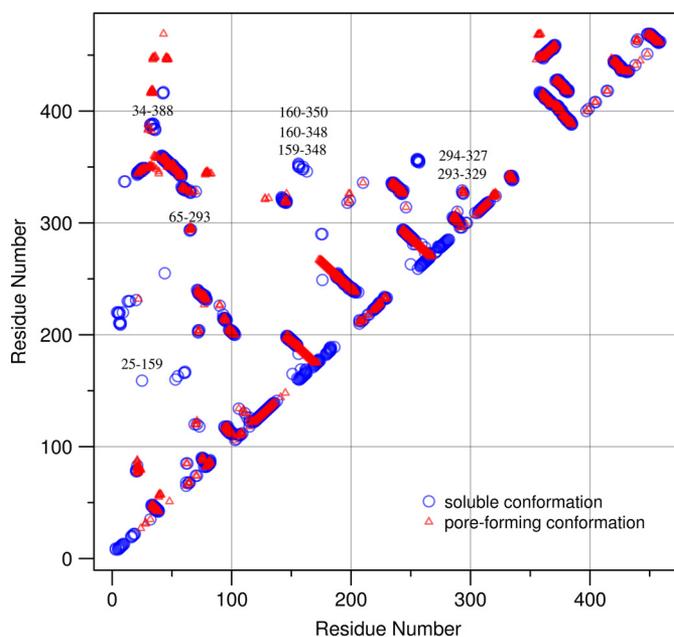


FIGURE 2. **Overlay of contact maps of the soluble and pore-forming conformations of PLY.** Points shown as *blue circles* represent residue pairs whose C$\alpha$ atoms are within 7 Å of each other in the three-dimensional structural model of the soluble form. *Red triangles* represent residues in similarly close proximity in the pore-forming conformation. Off-diagonal points are particularly noteworthy because they represent close three-dimensional proximity between residues that are distant in the amino acid linear sequence. Isolated *blue* or *red* points represent contacts present in only one of the two conformations. The positions of cysteine substitutions in the disulfide-trapped mutants designed in this study are labeled.

C$\alpha$-C$\alpha$ cutoff distance and with side chains suitably oriented for disulfide bonding in the initial soluble state but spatially too distant for disulfide linking in the final pore-forming conformation or (ii) within the disulfide cutoff distance in both the soluble and pore-forming conformations but with side chain orientations suitable for disulfide bridging only in the soluble state. To prevent formation of unintended intermolecular disulfide linkages, the sole free Cys at position 428 of native PLY was mutated to Ala (known to have no effect on toxin activity (36)). Designs 1, 2, and 6–9 in Table 1 fulfilled the first criterion defined above, whereas Designs 3–5 fulfilled the second crite-

**TABLE 1**

**List of mutants designed in this study**

For each design, the amino acid modifications made and the rationale behind the changes are specified.

| Design no. | Mutation(s) | Modified domains | $C\alpha$-$C\alpha$ (Å) | | Rationale |
|---|---|---|---|---|---|
| | | | Soluble | Pore-forming | |
| 1 | G25C, E159C, C428A | D2–D3 | 5.7 | 50.4 | Constrained via disulfide bond |
| 2 | E159C, T348C, C428A | D3–D2 | 6.8 | 52.0 | Constrained via disulfide bond |
| 3 | G294C, L327C, C428A | D3–D1 | 4.9 | 4.2 | Constrained via disulfide bond |
| 4 | G293C, I329C, C428A | D3–D1 | 6.4 | 5.7 | Constrained via disulfide bond |
| 5 | T65C, G293C, C428A | D1–D3 | 5.7 | 4.6 | Constrained via disulfide bond |
| 6 | Q160C, Y350C, C428A | D3–D2 | 6.6 | 56.7 | Constrained via disulfide bond |
| 7 | Q160C, T348C, C428A | D3–D2 | 6.7 | 55.6 | Constrained via disulfide bond |
| 8 | G25C, Q160C, C428A | D2–D3 | 7.7 | 54.0 | Constrained via disulfide bond |
| 9 | K34C G388C | D2–D4 | 4.1 | 18.2 | Constrained via disulfide bond |
| 10 | E29P | D2 | | | Hinge stiffening |
| 11 | E352P | D2 | | | Hinge stiffening |
| 12 | E29P, E352P | D2 | | | Hinge stiffening |
| 13 | G243V | D1 | | | Introduce steric hindrance at conserved Gly positions |
| 14 | G293A | D3 | | | Introduce steric hindrance at conserved Gly positions |
| 15 | G293T | D3 | | | Introduce steric hindrance at conserved Gly positions |
| 16 | G293V | D3 | | | Introduce steric hindrance at conserved Gly positions |
| 17 | G294P | D3 | | | Introduce steric hindrance at conserved Gly positions |
| 18 | G293C | D3 | | | Introduce steric hindrance at conserved Gly positions |

rion. It is evident from Table 1 that, in some cases, $C\alpha$-$C\alpha$ separations in the soluble form differ from corresponding distances in the pore-forming state by as much as 50 Å. Such dramatic disparities suggest that disulfide trapping could effectively prevent the structural rearrangements required for pore formation, and hence toxicity, as hypothesized.

*Rigidity-conferring Mutations at Mechanical Hinge Regions*—Protein domains that move relative to each other during pore formation and the associated hinge regions were identified using the Hingefind algorithm (27). Coordinates of the soluble and pore-forming conformations were provided as input. Through least-squares superposition of the two configurations, PLY was partitioned into geometrically preserved domains that move as rigid bodies relative to each other (Fig. 3*A*). Hinge axes between the dynamic domains were then determined. Considering movements between physically connected regions only, the domains colored *yellow* and *orange* in Fig. 3 undergo the largest relative displacement. The 117° effective rotation between them reflects the extensive deformation of D2 during pore formation. The interface between these two regions was thus chosen as the focal point for the introduction of hinge-stiffening mutations. A pictorial view of the Hingefind result showing the axis of rotation of the *yellow* domain relative to the *orange* domain is presented in Fig. 3*B*. When designing hinge-stiffening mutants of PLY, candidate amino acids for mutation were selected by taking into account the following: (i) proximity to the hinge axis, (ii) the magnitude of change the backbone dihedral angles undergo in transitioning from the soluble to the pore-forming conformation, and (iii) the compatibility of backbone dihedrals in the soluble state with $\varphi$ and $\psi$ angles typical of Pro. The range of $\varphi$ and $\psi$ angles accessible to Pro is limited by its pentagonal ring that connects the $C\alpha$ atom to the backbone nitrogen atom. As a result, the $\varphi$ angle is restricted to approximately −60°, whereas the $\psi$ angle clusters around two regions in the Ramachandran plot, −35° ($\alpha$-region) and +150° ($\beta$-region) (37). Listed in Table 2 are backbone dihedral angles of nine residues in closest proximity to the rotation axis described in Fig. 3*B*. Table 2 shows that Glu-29, Glu-50, and Glu-352 best match the previously stated selection criteria for Pro substitu-
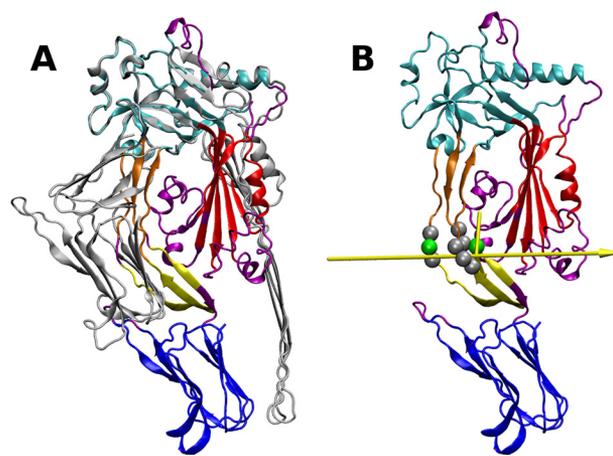


FIGURE 3. **Structural transitions of PLY.** *A*, structural alignment of the soluble and pore-forming conformations of PLY. The largest dynamic domain (colored *cyan* in the soluble conformation) was used as the reference for superposition. Domain partitioning was done using the Hingefind algorithm with a tolerance of 5.0 Å (27). Other dynamic domains identified are shown in *blue*, *red*, *orange*, and *yellow*, whereas regions that could not be portioned into a dynamic domain are *purple*. *B*, the soluble conformation is used to illustrate the movement of the *orange* domain relative to the *yellow* domain. The axis of rotation is shown as a *yellow arrow*, with its orientation representing a left-handed 117° rotation about the axis. The pivot point in the middle of the arrow is connected to domain centroids to illustrate the rotation angle. Residues in close proximity to the axis are rendered as *spheres*, with the position of the two hinge residues modified to prevent interdomain rotation (Glu-29 and Glu-352) highlighted in *green*. The images were produced with VMD (47).

tion. Three variants of PLY were thus designed based on the result of the Hingefind analysis: PLY(E29P), PLY(E352P), and PLY(E29P,E352P). These are designated as Designs 10–12 in Table 1. In all three, the modified positions are located in D2, where the backbone flexibility restrictions imposed by Pro are expected to prevent the twisting and bending of the $\beta$-sheet linking D1 to D4, which appears to be necessary for the conformational transition from the soluble to the pore-forming state (see Figs. 1 and 3).

*Introduction of Steric Hindrance at Conserved Glycines*—Two recent studies have revealed the existence of structural similarity between the bacterial cholesterol-dependent cytolysin family of proteins (to which PLY belongs) and proteins con-

**TABLE 2**
**Mechanical hinges of PLY identified by the Hingefind algorithm**
Comparisons of backbone $\varphi$ and $\psi$ dihedral angles (in degrees) for each hinge residue in the soluble and pore-forming conformation are shown.

| | Hinge residue | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Ile-28 | Glu-29 | Asn-30 | Arg-48 | Lys-49 | Lys-50 | Tyr-350 | Val-351 | Glu-352 |
| $\varphi$ | | | | | | | | | |
| Soluble | −104 | −68 | −112 | −115 | −127 | −79 | −110 | −89 | −82 |
| Pore | −137 | −77 | −56 | −179 | −78 | −144 | −149 | −156 | −41 |
| $\psi$ | | | | | | | | | |
| Soluble | 122 | −47 | 136 | 134 | 132 | 116 | 152 | 105 | 110 |
| Pore | 171 | −111 | 120 | −166 | −51 | 106 | −117 | −169 | −43 |
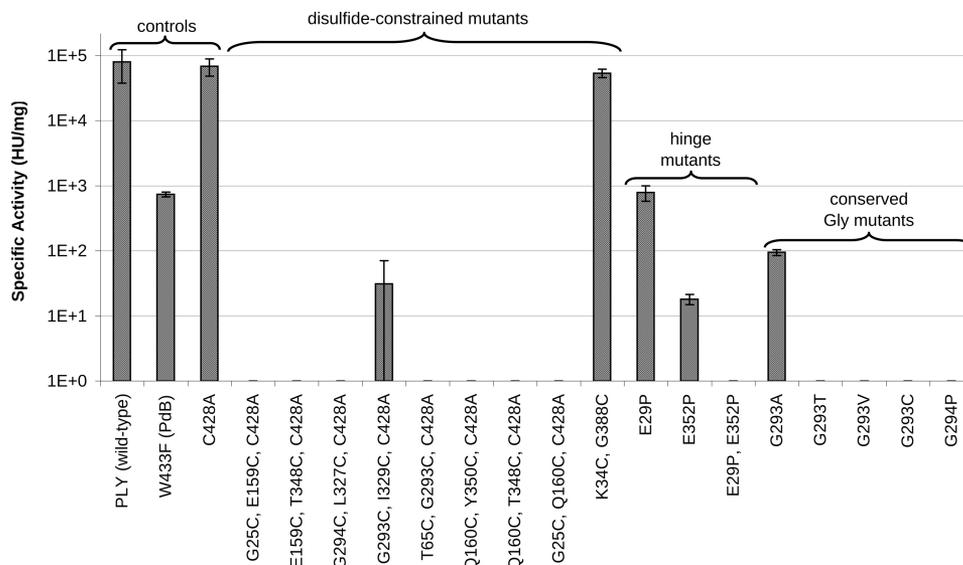


FIGURE 4. **Specific activity of engineered PLY mutants under nonreducing conditions.** An absence of a *gray bar* indicates undetectable levels of activity. For activity under reducing conditions, see supplemental Fig. S2. *HU*, hemolytic units.

taining membrane attack complex/perforin (MACPF) domains (38, 39). Despite an overall lack of sequence similarity between cholesterol-dependent cytolysin toxins and the MACPF domain, some core structural elements and three Gly residues are conserved (supplemental Fig. S1). In PLY, the conserved glycines correspond to Gly-243, Gly-293, and Gly-294. The structural and functional similarities between these two distinct classes of pore-forming proteins suggest that they share a common mechanism and that the three conserved glycines may play an important role in pore formation (39). These Gly residues were therefore targeted for substitution with amino acids of varying side chain sizes, with the goal of sterically hindering the conformational transition from the soluble to the pore-forming state. Substitution of a small residue like Gly with a larger one was predicted to increase the energy barrier to conformational transition from the soluble to the pore state by reducing backbone dihedral space and restricting the torsional rotation of nearby side chains of surrounding residues. The mutations evaluated for this strategy are listed in Table 1 (Designs 13–18), with the rationale summarized briefly as follows. Because Gly-243 is located at a bend in a long $\beta$-strand that connects D1 and D3 (supplemental Fig. S1), it was substituted with Val, a larger amino acid that is also typically a $\beta$-strand former and hence likely to preserve local secondary structure (40). On the other hand, Gly-293 and Gly-294 are located in a loop connecting two short $\beta$-strands in D3. Several amino acids (Ala, Thr, Val, and Cys) were substituted for Gly-

293 to explore the effect of side chain size. The G294P mutation was inspired by previous studies with unrelated proteins demonstrating that Pro residues in loop regions increase conformational stability by enhancing the rigidity of the loop (41, 42).

*Generation of Recombinant Proteins*—Mutant proteins were expressed and purified using standard techniques (see "Experimental Procedures"). Initially, the quality of the mutant proteins was assessed by comparing their DSC profiles with those obtained using the wild-type active protein (supplemental Table S1). In the case of disulfide-constrained mutants, it was also possible to show the maintenance of correct overall conformation by reducing the disulfide constraints and demonstrating the reacquisition of hemolytic activity (supplemental Fig. S2). Ultimately, the ability to elicit antibodies that neutralized native PLY was the critical test of mutant protein quality (discussed below).

*Toxin Activity of Mutant PLY Proteins*—The toxin activity of the PLY mutants was measured using a hemolysis assay designed to monitor the pore-induced release of hemoglobin from red blood cells by absorbance spectroscopy (see "Experimental Procedures"). Specific hemolytic activities are summarized in Fig. 4, with results grouped according to class of mutation. In agreement with published data (21), PdB was 100-fold less active than wild-type PLY, and the mutation C428A (36) did not affect toxin activity. Consequently, PdB hemolytic activity was used as the benchmark for improvement, and the PLY(C428A) substitution was incorporated into the disulfide-

constrained designs to prevent the unwanted formation of disulfides with the endogenous Cys. Most of the novel PLY mutants were less hemolytic than PdB, and many were completely inactive ($\geq$100,000-fold less active than wild-type PLY, below the assay limit of detection) (Fig. 4).

*Antigenic Integrity of Mutant PLY Proteins*—To test whether the attenuated mutants were still able to elicit neutralizing antibodies, antisera were generated against these proteins. Immunogenicity of the mutant proteins was confirmed by determination of PLY-specific IgG titers and by testing their ability to neutralize the wild-type toxin (see "Experimental Procedures"). All mutants elicited IgG titers similar to wild-type PLY (supplemental Table S2). Neutralization titers of antisera were similar to that of the control serum raised against wild-type PLY with the exception of two mutants; the neutralization titers of anti-PLY(E352P) and anti-PLY(E29P,E352P) antisera were 4- and 64-fold less effective than that of the control serum, respectively (supplemental Table S2). This observation suggests that PLY(E352P) and PLY(E29P,E352P) may have adopted non-native conformations compared with wild-type PLY and would thus be less effective vaccine antigens. This interpretation is supported by thermal stability data obtained by DSC: the thermal transition midpoint temperature ($T_m$) of PLY(E29P,E352P) was nearly 7 °C lower than that of wild-type PLY ($T_m = 43.0$ °C *versus* 49.8 °C for wild-type PLY) (Fig. 5A). In addition, the profile of the PLY(E29P,E352P) thermal transition (Fig. 5A) suggests that the protein would be at least partially unfolded at the body temperature of the immunized rats. This is further supported by CD spectroscopy experiments that showed PLY(E29P,E352P) had a different secondary structural profile compared with wild-type PLY (Fig. 5B). Based on these results, PLY(E352P) and PLY(E29P,E352P) were deemed to be structurally compromised and unsuitable as vaccine candidates.

## DISCUSSION

Using a systematic structure-driven approach, we have solved a long-standing problem in pneumococcal vaccine development: detoxification of the key virulence factor PLY with retention of its antigenic integrity. Our objective was to identify and introduce modifications into PLY that could prevent its transition from the soluble to the pore-forming state while retaining its ability to elicit neutralizing antibodies. In all, 11 of our designs met these criteria.

The premise that Gly-293 would be critical to PLY activity was inspired by the recently demonstrated commonality of structural folds between cholesterol-dependent cytolysins and MACPF domain proteins and the observation that specific Gly residues also appeared to be conserved between these protein families (38, 39). This hypothesis was validated by experiments in which we substituted Gly-293 and Gly-294 with side chains larger than Ala to generate mutants that lost hemolytic activity (Fig. 4). Both Gly residues are well conserved in cholesterol-dependent cytolysins and MACPF domain proteins (38, 39). Therefore, for this class of mutants, we propose a mechanism of inactivation in which large side chains introduce steric interference. This would impede the conformational progression of PLY to the competent pore-forming state. The structural model of PLY indicates that large amino acid residues at posi-
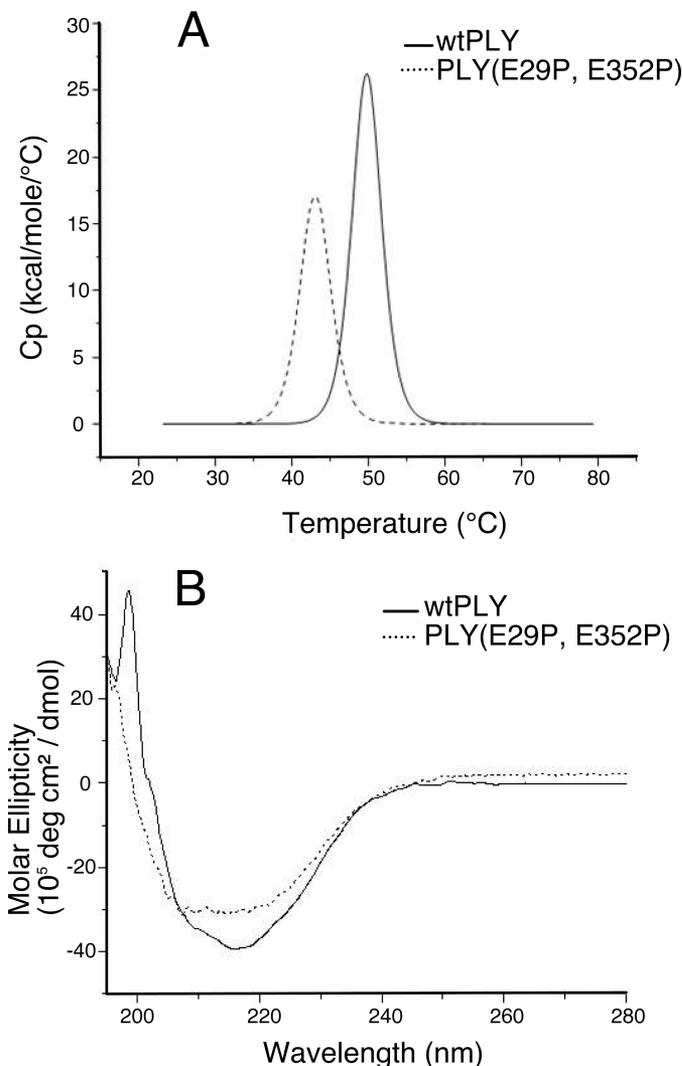


FIGURE 5. *A*, overlay of DSC thermograms for wild-type PLY (*solid line*) and mutant PLY(E29P,E352P) (*dashed line*). $T_m$ values were 49.8 °C and 43.0 °C, respectively. *Cp*, specific heat capacity at constant pressure. *B*, overlay of CD spectra for wild-type PLY (*solid line*) and mutant PLY(E29P,E352P) (*dashed line*). *deg*, degrees.

tion 293 are subject to steric hindrance from three neighboring side chains: Thr-65, Asn-66, and Ile-329. Extrapolating our findings, we predict that the pore-forming activity of MACPF proteins would be similarly impeded by the introduction of large residues at the positions equivalent to Gly-293 and Gly-294 of PLY (*e.g.* Gly-270 and Gly-271 in *Photorhabdus luminescens* MACPF).

Computer models of the multiple PLY conformational states were instrumental to our antigen-engineering efforts involving disulfide bonds. Of nine *in silico* designs featuring disulfide constraints, seven were rendered completely inactive, another had its hemolytic activity reduced by 3 orders of magnitude relative to wild-type PLY, and one mutant (PLY(K34C,G388C)) was not attenuated as intended (Fig. 4). Exposure of these mutants to reducing conditions led to the recovery of hemolytic activity where expected (supplemental Fig. S2). We conclude from these data that the mechanism of detoxification involves the disulfide trapping of PLY in conformations that are incapable of proceeding through one or more steps of

association and assembly into oligomeric membrane-spanning pores. Of particular note is the disulfide-constrained mutant PLY(T65C,G293C,C428A), which seems to be attenuated via a dual mechanism, one aspect of which depends on disulfide trapping and the other of which depends on alteration of the conserved Gly-293 residue. Such redundancy in detoxification mechanisms is highly attractive from a vaccine safety point of view. The successful formation of a disulfide bond in this mutant was verified by DSC (supplemental Table S1) and free thiol-scavenging experiments using Ellman's reagent (data not shown). We attribute its lack of activity under reducing conditions to the effect of the G293C mutation alone (Fig. 4) because T65C by itself does not affect PLY activity (data not shown). Also noteworthy is PLY(G293C,I329C,C428A), whose hemolytic activity was 3 orders of magnitude lower than that of wild-type PLY under nonreducing conditions. The partial inactivation of this mutant indicates that the Cys-293–Cys-329 disulfide bond impedes but does not completely prevent conformational changes required for membrane insertion. In addition, a disulfide bond between residues 293 and 329 is predicted to pull the side chain of residue 293 away from two of its nearest and largest neighbors, Thr-65 and Asn-66, thus reducing steric hindrance to pore-forming conformational change. Upon reduction of the disulfide bond, restoration of steric interference between the side chains of Cys-293, Thr-65, and Asn-66 results in complete inactivation. We therefore conclude that steric effects resulting from the G293C modification are a more effective inactivation mechanism than the linking of D3 domain residue 293 to D1 domain residue 329 via a disulfide bond.

The final class of mutants in this study had rigidity-conferring Pro substitutions introduced at hinge residue sites as a strategy for detoxification. Although computational tools for dynamic domain and hinge prediction have been in existence for over a decade, examples of the practical application of these tools in protein design are rare (27). All PLY mutants designed based on this approach showed reduced hemolytic activity, with PLY(E29P,E352P) being completely attenuated. Despite the success of this strategy in attenuating toxin activity, the concomitant decrease in neutralizing antibodies elicited by PLY(E352P) and PLY(E29P,E352P) makes these mutations less interesting from a vaccine development perspective.

Achievement of safety and efficacy is the ultimate challenge in vaccinology. Antigen modifications intended to eliminate toxicity may have unexpected effects on the correct presentation of protective epitopes. Therefore, our engineered PLY mutants were prioritized not only by loss of hemolytic activity but also by their ability to elicit neutralizing antibodies. PLY(T65C,G293C,C428A), which is attenuated by a dual mechanism and elicits the desired antibody response, is currently in Phase I clinical trials.

Although genetic detoxification of bacterial toxins has been reported (43, 44), the application of rigorous structure-driven methods to identify suitable mutation sites is yet to be fully exploited. As demonstrated herein, structure-based antigen design is a viable approach for eliminating toxin activity while retaining the features necessary for inducing neutralizing antibodies. Historically, chemical methods like formalin inactivation have been used to detoxify virulence factors (45). However, these methods are less desirable from a regulatory and production standpoint due to their nonspecific nature. A systematic structure-guided approach has the potential to save time and resources in vaccine development and holds promise for making progress on difficult but critical antigen targets that have thus far defied traditional approaches (46).

## REFERENCES

1. World Health Organization (2003) *Wkly. Epidemiol. Rec.* **78,** 110–119
2. Paton, J. C., and Boslego, J. W. (2008) in *Pneumococcal Vaccines: The Impact of Conjugate Vaccine* (Siber, G. R., Klugman, K. P., and Makela, P. H., eds) pp. 421–436, ASM Press, Washington, D. C.
3. Dormitzer, P. R., Ulmer, J. B., and Rappuoli, R. (2008) *Trends Biotechnol.* **26,** 659–667
4. Berman, H. M., Westbrook, J. D., Gabanyi, M. J., Tao, W., Shah, R., Kouranov, A., Schwede, T., Arnold, K., Kiefer, F., Bordoli, L., Kopp, J., Podvinec, M., Adams, P. D., Carter, L. G., Minor, W., Nair, R., and La Baer, J. (2009) *Nucleic Acids Res.* **37,** D365–D368
5. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) *Annu. Rev. Biophys.* **37,** 289–316
6. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D. (2005) *Science* **310,** 638–642
7. Ash, W. L., Zlomislic, M. R., Oloo, E. O., and Tieleman, D. P. (2004) *Biochim. Biophys. Acta* **1666,** 158–189
8. Lindahl, E., and Sansom, M. S. (2008) *Curr. Opin. Struct. Biol.* **18,** 425–431
9. Pletz, M. W., Maus, U., Krug, N., Welte, T., and Lode, H. (2008) *Int. J. Antimicrob. Agents* **32,** 199–206
10. Black, S., Eskola, J., Whitney, C., and Shinefiled, H. (2008) in *Vaccines* (Plotkin, S. A., Orenstein, W., and Offit, P., eds) pp. 531–567, W. B. Saunders Co., Philadelphia
11. Paton, J. C. (1996) *Trends Microbiol.* **4,** 103–106
12. Feldman, C., Munro, N. C., Jeffery, P. K., Mitchell, T. J., Andrew, P. W., Boulnois, G. J., Guerreiro, D., Rohde, J. A., Todd, H. C., Cole, P. J., and (1991) *Am. J. Respir. Cell Mol. Biol.* **5,** 416–423
13. Mitchell, T. J., Mendez, F., Paton, J. C., Andrew, P. W., and Boulnois, G. J. (1990) *Nucleic Acids Res.* **18,** 4010
14. Ferrante, A., Rowan-Kelly, B., and Paton, J. C. (1984) *Infect. Immun.* **46,** 585–589
15. Paton, J. C., and Ferrante, A. (1983) *Infect. Immun.* **41,** 1212–1216
16. Alexander, J. E., Lock, R. A., Peeters, C. C., Poolman, J. T., Andrew, P. W., Mitchell, T. J., Hansman, D., and Paton, J. C. (1994) *Infect. Immun.* **62,** 5683–5688
17. Lock, R. A., Paton, J. C., and Hansman, D. (1988) *Microb. Pathog.* **5,** 461–467
18. Paton, J. C., Lock, R. A., and Hansman, D. J. (1983) *Infect. Immun.* **40,** 548–552
19. Gilbert, R. J. (2002) *Cell. Mol. Life Sci.* **59,** 832–844
20. Berry, A. M., Alexander, J. E., Mitchell, T. J., Andrew, P. W., Hansman, D., and Paton, J. C. (1995) *Infect. Immun.* **63,** 1969–1974
21. Korchev, Y. E., Bashford, C. L., Pederzolli, C., Pasternak, C. A., Morgan, P. J., Andrew, P. W., and Mitchell, T. J. (1998) *Biochem. J.* **329,** 571–577
22. Kirkham, L. A., Kerr, A. R., Douce, G. R., Paterson, G. K., Dilts, D. A., Liu, D. F., and Mitchell, T. J. (2006) *Infect. Immun.* **74,** 586–593
23. Rossjohn, J., Feil, S. C., McKinstry, W. J., Tweten, R. K., and Parker, M. W. (1997) *Cell* **89,** 685–692
24. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242
25. Tilley, S. J., Orlova, E. V., Gilbert, R. J., Andrew, P. W., and Saibil, H. R.

(2005) *Cell* **121,** 247–256

26. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) *Proteins* **23,** 318–326

27. Wriggers, W., and Schulten, K. (1997) *Proteins* **29,** 1–14

28. Nöllmann, M., Gilbert, R., Mitchell, T., Sferrazza, M., and Byron, O. (2004) *Biophys. J.* **86,** 3141–3151

29. Walz, T. (2005) *Mol. Cell* **18,** 393–394

30. Rossjohn, J., Polekhina, G., Feil, S. C., Morton, C. J., Tweten, R. K., and Parker, M. W. (2007) *J. Mol. Biol.* **367,** 1227–1236

31. Polekhina, G., Giddings, K. S., Tweten, R. K., and Parker, M. W. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 600–605

32. Hotze, E. M., Wilson-Kubalek, E. M., Rossjohn, J., Parker, M. W., Johnson, A. E., and Tweten, R. K. (2001) *J. Biol. Chem.* **276,** 8261–8268

33. Ramachandran, R., Tweten, R. K., and Johnson, A. E. (2004) *Nat. Struct. Mol. Biol.* **11,** 697–705

34. Reiter, Y., Brinkmann, U., Jung, S. H., Pastan, I., and Lee, B. (1995) *Protein Eng.* **8,** 1323–1331

35. Sowdhamini, R., Srinivasan, N., Shoichet, B., Santi, D. V., Ramakrishnan, C., and Balaram, P. (1989) *Protein Eng.* **3,** 95–103

36. Saunders, F. K., Mitchell, T. J., Walker, J. A., Andrew, P. W., and Boulnois, G. J. (1989) *Infect. Immun.* **57,** 2547–2552

37. MacArthur, M. W., and Thornton, J. M. (1991) *J. Mol. Biol.* **218,** 397–412

38. Hadders, M. A., Beringer, D. X., and Gros, P. (2007) *Science* **317,** 1552–1554

39. Rosado, C. J., Buckle, A. M., Law, R. H., Butcher, R. E., Kan, W. T., Bird, C. H., Ung, K., Browne, K. A., Baran, K., Bashtannyk-Puhalovich, T. A., Faux, N. G., Wong, W., Porter, C. J., Pike, R. N., Ellisdon, A. M., Pearce, M. C., Bottomley, S. P., Emsley, J., Smith, A. I., Rossjohn, J., Hartland, E. L., Voskoboinik, I., Trapani, J. A., Bird, P. I., Dunstone, M. A., and Whisstock, J. C. (2007) *Science* **317,** 1548–1551

40. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13,** 222–245

41. Choi, E. J., and Mayo, S. L. (2006) *Protein Eng. Des. Sel.* **19,** 285–289

42. Vieille, C., Burdette, D. S., and Zeikus, J. G. (1996) *Biotechnol. Annu. Rev.* **2,** 1–83

43. Giannini, G., Rappuoli, R., and Ratti, G. (1984) *Nucleic Acids Res.* **12,** 4063–4069

44. Del Giudice, G., and Rappuoli, R. (1999) *Schweiz. Med. Wochenschr.* **129,** 1744–1748

45. Rappuoli, R. (1994) *Vaccine* **12,** 579–581

46. Douek, D. C., Kwong, P. D., and Nabel, G. J. (2006) *Cell* **124,** 677–681

47. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J. Mol. Graph.* **14,** 33–38

**Introduction to Nuclear Magnetic Resonance Spectroscopy of Proteins**

**Structure**                    **Dynamics**

**Biological Function**

Most protein structures in the Protein Data Bank were determined by X-ray diffraction.

NMR can determine the structure of a small protein, on the order of 200 – 300 residues.

An advantage of NMR is that proteins can be studied in solution, and dynamics over a wide range of timescales can be measured from ns → s.

---

Structure determination by NMR amounts to measuring inter-atomic **distances** and dihedral **angles**.



*http://cmgm.stanford.edu/biochem201/Slides/Protein%20Structure/Phi-Psi%20Angles.JPG*
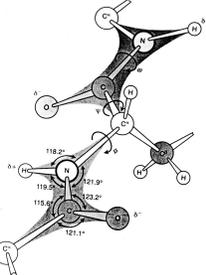
*Figure 1*

---

Information is contained in the NMR parameters:

**Chemical Shift**:  sensitive to the **chemical environment** – provides information about secondary structure.

**J-Coupling**:  contains **dihedral angle** information about atoms connected by covalent bonds.

**Nulcear Overhauser Effect**: contains information about atoms connected through **space** and about molecular motion.

The NMR measurements are entered into a computer program that calculates protein structures that satisfy the experimental restraints.

## An Overview of NMR Theory

**Magnetism**: A property of moving electric charge.
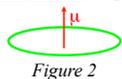


*Figure 2*
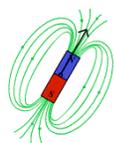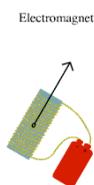
Electromagnet

Bar Magnet

(c) Copyright 2003 Genevieve Tauxe
*http://magician.ucsd.edu/Lab_tour/Vectors/moments.gif*

*Figure 3*

---

A rotating charged body (*Figure 4*) will have associated with it an angular momentum **J** and a magnetic moment **μ** where:

$$\vec{\mu} = \gamma \vec{J}$$

**μ** units are $J \cdot T^{-1}$ **J** units are $J \cdot s$ γ units are $s^{-1} \cdot T^{-1}$

γ = magnetogyric ratio, a property of the body.
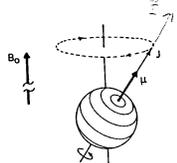


*A. Derome: Modern NMR Techniques for Chemistry Research*

*Figure 4*

---

Such particles behave just the same as a bar magnet.

## Nuclear Magnetism:

Atomic particles possess a property analogous to spin angular momentum called the *intrinsic spin*. It is described by the quantum mechanical spin angular momentum operator (vector) **I**.

$$\vec{J} = \hbar\sqrt{\vec{I}(\vec{I}+1)}$$

$\hbar$ = Planck's constant divided by $2\pi \sim 10^{-34}$ J·s

Electrons, neutrons, and protons all have spin angular momentum and an associated magnetic moment.

Table 1

| | Z | I | $\gamma \times 10^7$ | $\omega_0$ @11.7T | Abund | Sensitiv |
|---|---|---|---|---|---|---|
| $^{-1}e$ | 0 | ½ | −17,766 | | 100 | |
| $^1_0 n$ | 0 | ½ | 18 | | 100 | |
| $^1H$ | 1 | ½ | 27 | 500 | 99.98 | 1 |
| $^2H$ | 1 | 1 | 4.1 | 76.8 | 0.016 | 0.009 |
| $^{12}C$ | 6 | 0 | 0 | 0 | 98.9 | 0 |
| $^{13}C$ | 6 | ½ | 6.7 | 126 | 1.11 | 0.016 |
| $^{14}N$ | 7 | 1 | 1.9 | 36 | 99.6 | 0.001 |
| $^{15}N$ | 7 | ½ | −2.7 | 50.7 | 0.37 | 0.001 |
| $^{16}O$ | 8 | 0 | 0 | 0 | 99.76 | 0 |
| $^{31}P$ | 15 | ½ | 10.8 | 202 | 100 | 0.07 |
| $^{32}S$ | 16 | 0 | 0 | 95.1 | 0 | 0 |

Not all nuclei are magnetic.

---

What happens when a spinning top with angular momentum in a uniform gravitational field is displaced from the vertical?



http://hyperphysics.phy-astr.gsu.edu/HBASE/mechanics/imgmech/topp.gif

*Figure 5*

---
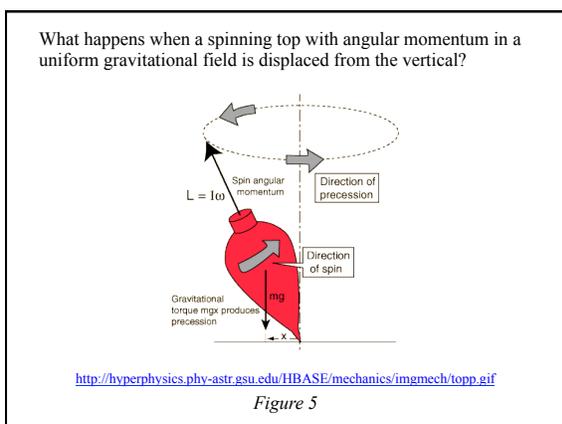
The gravitational force **F=mg** produces a horizontal torque **τ** causing the top to precess about the vertical axis at an angular velocity **ω**. See Figures 5 and 6.

This is a good analogy for what happens to the nucleus of an atom when it is placed in an external magnetic field **B$_o$**.

$$\frac{d\vec{\mu}}{dt} = \vec{\tau} = -\gamma \vec{B}_0 \times \vec{\mu}$$

This is a cross product.

Thus, the change in direction of **μ** is perpendicular to the plane formed by **μ** and **B$_o$**.



*Figure 6*

The nucleus of an atom possessing spin, immersed in a magnetic field, $B_o$, will **precess** about the axis of the magnetic field with angular velocity $\vec{\omega}_0 = -\gamma\vec{B}_0$ ⟵ magnetic field strength

precession frequency    property of nucleus

Figures 4 and 6 show the behaviour of 1 nucleus.

Figure 7 shows the magnetic moments of a large population of nuclei precessing in a magnetic field.

$B_0$

$M_0$

*Figure 7*

http://www.nmr.uni-duesseldorf.de/sets/theory.html

---
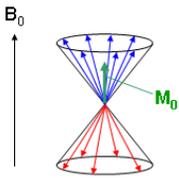
The NMR experiment measures the precession frequency, $\omega_0$, of an atom in a magnetic field.

For atoms with spin $I = 1/2$, the energy of a magnetic moment $\mu$ in $B_o$ is:

$$E_1 = \frac{-\gamma h B_0}{2} \qquad E_2 = \frac{+\gamma h B_0}{2}$$

Here is the E-level diagram:

β ___↓↓↓___  $E_2$

α ___↑↑↑___  $E_1$
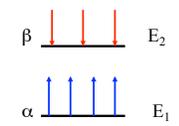
*Figure 8*

---

A classical picture of this is two populations of spins aligned **with** and **against** the applied field.

NMR measures the net magnetic moment $M_0$

$B_0$

$M_0$
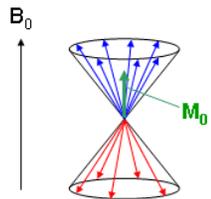
*Figure 7*

Since $\omega_0 = -\gamma B_0$ then $\nu_0 = \dfrac{-\gamma B_0}{2\pi}$

If $B_o = 11.7$ Tesla and $\gamma$ for $^1H = 268.5 \times 10^6$ rad·Tesla$^{-1}$·sec$^{-1}$

Then $\nu_o = 500 \times 10^6 = 500$ MHz

So why don't all protons resonate at 500 MHz?

The equation above is not quite complete. In addition to $B_o$, each nucleus experiences a small magnetic field, $B_i$, due to the orbiting electrons in its local environment.

We should write: $\vec{\omega}_0 = -\gamma(\vec{B}_0 + \vec{B}_i)$

---

Generally, $B_o$ induces electrons to circulate and creates a $B_i$ which **opposes** $B_o$.

For *e.g.*, because O is more electronegative than H, it will pull bonding electrons away from an attached H causing it to experience nearly the full $B_o$ – we say the H is **de-shielded**.

A H attached to a C will be **shielded** from $B_o$ by the nearby bonding electron's $B_i$. $\omega_0(H_B) > \omega_0(H_A)$

So H attached to O resonates at a lower field, and higher frequency than H attached to C:

$$O—H_B \qquad C—H_A$$

O-H        C-H

10        5        1    0    ppm

*Figure 9*

---
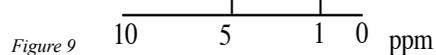
The resonance frequency is determined by the chemical "*environment*" of an atom and is called the **Chemical Shift**. It can be used to extract information about protein conformation.

But there is a problem. Atoms precess at different frequencies in different fields. This is solved by reporting the difference between the resonance frequency of an atom and that of a standard atom divided by the frequency of the standard. For $^1H$ and $^{13}C$ the standard is DSS; disodium 2,2-dimethyl-2-silapentane-5-sulfonate.

$$CH_3—Si(CH_3)(CH_3)—CH_2—CH_2—CH_2—S(=O)(=O)—O^-$$
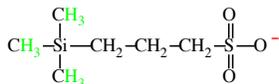
*Figure 10*

The DSS protons resonate at 300,000,000 Hz in a 7 Tesla magnet and at 500,000,000 Hz in an 11.7 Tesla field. A proton resonating 600 Hz faster than the DSS protons in a 7 Tesla magnet will resonate 1000 Hz faster that the reference in an 11.7 Tesla magnet:

$$\frac{300,000,600 - 300,000,000}{300,000,000} = 0.000002 = 2\,ppm$$

$$\frac{500,001,000 - 500,000,000}{500,000,000} = 0.000002 = 2\,ppm$$

So in any field in which DSS protons resonate at 0 ppm, the atom will resonate at 2 ppm. The symbol **δ** is often used to represent chemical shift.

---

**Chemical Shift Index**: The chemical shifts of a large number of protein $^1$H, $^{13}$C, and $^{15}$N resonances have been measured. There is a strong correlation between the measured chemical shift and the backbone conformation.

In Figure 11, the $^{13}$Cα and $^{13}$Cβ resonances of 442 residues from 4 proteins fall into distinct ranges for the α-helix and β-strand. Note that the chemical shifts are calculated as **differences from their random coil values**.
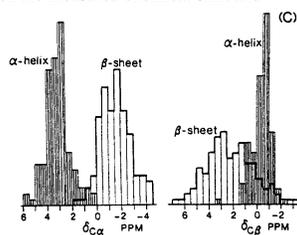


Cavanagh et al., Protein NMR Spectroscopy
*Figure 11*

---

The random coil values were determined from measurements of unstructured peptides and the unstructured regions of proteins.

*Figure 12* shows how CSI is sometimes used. Hα chemical shifts are determined to be either α-helical or β-strand.



TERRRQQLDKDGDGTIDEREIKIWFQNKRAKIK
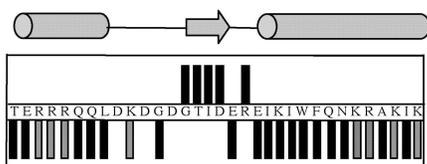
*Figure 12*

Welch J T *et al. PNAS* 2003;100:3725-3730

During structure determination (below) the difference between the measured and random coil $^{13}C\alpha$ chemical shifts can be used to restrain the $\phi$-angle (-50º +/-50º and 90º +/-50º) in a computer program.

**J- or Scalar or Spin-Spin Coupling:**
An atom with I = 1/2 can exist in two energy states, one aligned parallel to $\mathbf{B_o}$ ($\alpha$) and one aligned antiparallel to $\mathbf{B_o}$ ($\beta$):

β         $E_2$

α         $E_1$
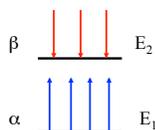
*Figure 8*

---

The energy difference is very small so $N_\alpha \sim N_\beta$.

In the molecule H—F we have:  50% H—F ↑ and 50% H—F ↓

The orientation of the F nucleus subtly affects the bonding electrons at H in the H—F bond which changes $\mathbf{B_l}$ at H.

H—F ↑      H—F ↓

So the Hs have a different resonance frequency depending on the orientation of F.

$$\omega_1 = \gamma(B_0 + B_{up}) \neq \omega_2 = \gamma(B_0 + B_{down})$$

---

The $^1H$ NMR spectrum is a doublet.

$^1J$

*Figure 13*

ppm
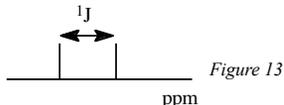
For the same reasons the $^{19}F$ NMR spectrum is also a doublet.

The separation of the peaks in the doublet is called the **J-coupling constant**.

If an atom is bonded to more than 1 other atom the spin states of all the nuclei can cause splitting of the main peak:

The relative intensities of the multiplets can be determined from Pascal's triangle where n = # of coupled atoms:

| | | | | | | Pascal's Triangle |
|---|---|---|---|---|---|---|

$n = 0$        1

$n = 1$      1    1

$n = 2$    1    2    1

$n = 3$   1    3    3    1

*Figure 14*

*e.g.*   Alanine in a peptide:



*Figure 15*

---

The $^1$H-NMR spectrum looks like:



*Figure 16*

The **areas** under the peaks are proportional to the number of atoms of each type.

In proteins, $^4$J are usually too small to affect the experiment but $^1$J, $^2$J, and $^3$J are important.

---

The magnitude of J is dependent upon the intervening **dihedral angles**.

*E.g.* $^3$J$_{HNC\alpha H}$ is determined by $\phi$ and $^3$J$_{H\alpha H\beta}$ is determined by $\chi_1$.

*Figure 17* shows a peptide fragment in the extended conformation where $\phi = \psi = \omega = 180°$



*Figure 17*

*Figure 18* shows Newman projections for $\phi = \psi = \omega = 0°$.
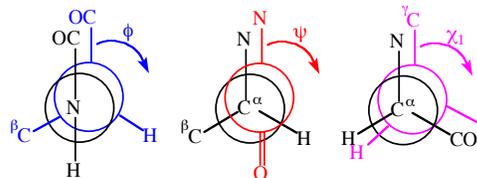
Figure 18

*Figure 19* shows spatial and Newman projections for the dihedral angle $\phi$ between the H-N-CA and N-CA-H planes.
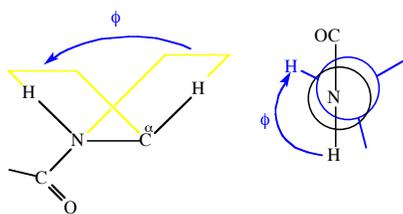
Figure 19

*Figure 20* shows the **Karplus relationship** for $^3J_{HNC\alpha H}$. This is based on $\phi$ angles measured from X-ray diffraction structures of hundreds of proteins and the $^3J$ values were determined from the proteins in solution by NMR.
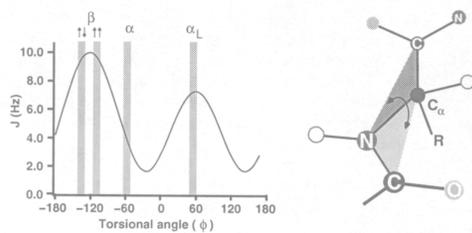
Rule and Hitchens (2006 Fundamentals of Protein NMR Spectroscopy

Figure 20

Based on analysis of hundreds of proteins the best Karplus equation is:

$$^3J = 6.4\cos^2\theta - 1.4\cos\theta + 1.9 \qquad \theta = \phi - 60^o$$

1. There are 4 different values for some $^3$J. Structure calculations must include all possibilities.

2. For $\phi$ = -57$^o$ ($\alpha$-helix) $^3$J$_{HNNA}$ = 3.9 Hz
   For $\phi$ = -119$^o$ (parallel $\beta$-sheet) $^3$J$_{HNHA}$ = 9.7 Hz.

3. Measurements of consecutive angles can give an indication of **secondary structure**.

4. Dynamic averaging of the structure results in $^3$J which are averages over the structures sampled. **Flexible peptides** yield $^3$J which are 7-8 Hz.

5. In structure calculations (below) $^3$J$_{HNHA}$ measurements are used to restrict $\phi$ by ±60$^o$.

**Spectroscopy**: Application of electromagnetic radiation with energy equal to the difference in energy between the states $\alpha$ and $\beta$ will cause an absorption.

$$\Delta E = h\nu_0 = \hbar\omega_0 = \hbar\gamma B_0$$

Thus, $\omega_0$ is the precession frequency of a nucleus in B$_o$ (classical picture) and the frequency of radiation absorbed in a transition (QM picture).

The number of nuclei in each energy state are described by the Boltzmann law:

$$\frac{N_\beta}{N_\alpha} = e^{(-\Delta E/kT)} = e^{(-\gamma\hbar B_0/kT)}$$

At room temperature $\Delta E$ is very small (N$_\alpha$ ~ N$_\beta$) so NMR is **not** a very sensitive technique and the electromagnetic radiation is in the radiofrequency (rf) portion of the spectrum.

**Resonance**: *"A sound produced by a body vibrating in sympathy with a nearby sound; tuning fork."*

10

"*A condition of a system in which there is a sharp maximum probability for the absorption of electromagnetic radiation.*"

Remember that electromagnetic radiation is an oscillating, propagating, electric and magnetic field.



http://micro.magnet.fsu.edu/primer/java/electromagnetic/electromagneticjavafigure1.jpg
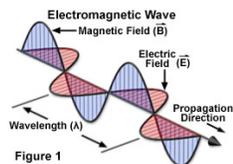
*Figure 21*

---

In this case, the applied EM radiofrequency resonates with the precessing nuclei, if their energies match.

To detect absorption of energy by precessing atoms we could vary the rf at constant $B_o$, or vary $B_o$ at constant $\omega$, to achieve resonance.

In **Pulse-Fourier Transform NMR** a broad range of rf is applied at constant $B_o$ bringing a wide range of nuclei into resonance in a few microseconds.

Then, they are all detected simultaneously.

The precessing atoms induce a time-dependent signal that lasts about 1s and is detected by the NMR spectrometer probe.

---

The signal is converted to a frequency spectrum by the **Fourier Transformation**.



http://chemistry.umeche.maine.edu/CHY431/NMR/NMR-9.jpg
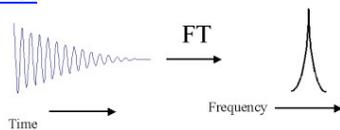
*Figure 22*

Adding the signals from many acquisitions can increase the sensitivity of the experiment because the signal increases in proportion to the number of scans, whereas the noise increases in proportion to the $\sqrt{N}$.

So, after 16 scans the S/N has increased by a factor of 4.



**How Many Scans Should I Collect?**

**Intensity vs. Number of Scans**

1  8  16  32        64                    128
**Number of Scans**

*Figure 23*  http://u-of-o-nmr-facility.blogspot.com/2007/11/how-many-scans-should-i-collect.html

---

**The Nuclear Overhauser Effect:**

Consider two atoms close in space ($H_a$ and $H_b$) but not necessarily coupled by covalent bonds. If the atoms are close enough together the magnetic moments of the two nuclei will interact. See *Figures 24 and 25*.



*A. Derome: Modern NMR Techniques for Chemistry Research*

*Figure 24*                    *Figure 25*

---

This is called **dipolar coupling**. It is not ordinarily observable in liquid samples because the motion of the molecule averages the effect to zero. (*Figure 26*) It is observable in liquid samples when the populations of the spin states are "*perturbed*".

*The NOE is a change in the intensity of a resonance when another resonance is perturbed.*

The effect depends on the distance (r) between the two atoms:

$$NOE \propto \frac{1}{r^6}$$



*Figure 26*

12

The NOE can be used to find out which atoms are close to a perturbed atom. Because of the $1/r^6$ dependence this effectively limits the distances that can be measured to about **5Å**.
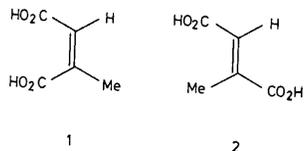


*Figure 27*

1 = Citraconic acid          2 = Mesaconic acid

In *Figure 28*, the methyl Hs were "*saturated*". The olefin Hs were enhanced.
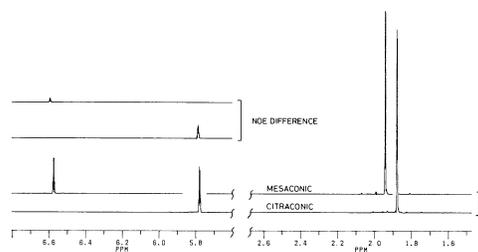


*Figure 28*

The citraconic H enhancement is greater than the mesaconic H because the internuclear distance is smaller in the former.

**Multi-dimensional NMR Spectroscopy**

A serious problem for NMR spectroscopy of large biomolecules is that the NMR spectra will contain thousands of resonance lines that **overlap** one another and are impossible to separate in a 1-dimensional spectrum. This problem is partly solved by 2D, 3D, and 4D NMR spectroscopy.

Two types of spectra are measured:

**COSY   Co**rrelation **S**pectroscop**y**

The **diagonal** contains all the information of a 1-D spectrum. The **cross-peaks** occur between atoms connected by **3** or fewer bonds (**J-coupled**).

13

Different amino acids exhibit different patterns of cross-peaks.
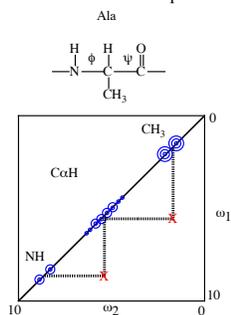
*Figure 29*

Here is a spectrum of lysozyme:

http://www.jeolusa.com/news/News28a/html/products/48-2.gif
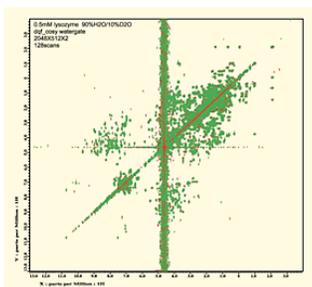
*Figure 30*

**NOESY**   **N**uclear **O**verhauser **E**ffect **S**pectroscop**y**

In this experiment the cross-peaks connect dipolar coupled atoms (**through space**).

Here is a NOESY spectrum of lysozyme.

http://www.bruker.com.cn/index/PRODUCS/NMR/magnet-image/mag3-4.jpg

*Figure 31*

The intensity (I) of the cross-peaks is proportional to $1/r^6$ between the atoms so r can be determined from I if a calibration $r_{ref}$ and $I_{ref}$ are known *e.g.* methylenes, aromatics.

$$r = r_{ref}\left(\frac{I_{ref}}{I}\right)^{\frac{1}{6}}$$

If the distance (r) between the atoms fluctuates (conformation) the NOE may be reduced or eliminated. For this, and other reasons, the NOE cannot be used to measure precise distances.

Thus, NOEs are usually classified as strong (X-2.5Å), medium (X-3.5Å), or weak (X-5.0Å).

X is the covalent bond distance or the van der Waals distance between the atoms.

---

*Figure 32* is a summary of sequential and medium range NOEs and their nomenclature.



K. Wuthrich: NMR of Proteins and Nucleic Acids

*Figure 32*

---

An α-helix is the most compact form of a polypeptide whereas the β-strand is the most stretched out form of the polypeptide. As a result, the pattern of interatomic distances in the two types of secondary structure are different.



*Figure 33*

Rule and Hitchens (2006 Fundamentals of Protein NMR Spectroscopy

Table 2 shows the interatomic distances that characterize secondary structure.

*Table 2*

| $Dis\tan ce$ | $\alpha-helix$ | $3_{10}-helix$ | $\beta_A$ | $\beta_P$ | $turnI$ | $turnII$ |
|---|---|---|---|---|---|---|
| $d_{\alpha N}$ | 3.5 | 3.4 | 2.2 | 2.2 | 3.4 | 2.2 |
| $d_{\alpha N}(i,i+2)$ | 4.4 | 3.8 | | | 3.6 | 3.3 |
| $d_{\alpha N}(i,i+3)$ | 3.4 | 3.3 | | | | |
| $d_{\alpha N}(i,i+4)$ | 4.2 | | | | $3-4$ | $4-5$ |
| $d_{NN}$ | 2.8 | 2.6 | 4.3 | 4.2 | 2.6 | 4.5 |
| $d_{NN}(i,i+2)$ | 4.2 | 4.1 | | | 3.8 | 4.3 |
| $d_{\beta N}$ | $2.5-4$ | $3-4.4$ | $3-4.5$ | $4-5$ | $3-4.4$ | $3.6-4.6$ |
| $d_{\alpha\beta}(i,i+3)$ | $2.5-3.4$ | $3-5$ | | | | |

A single NOE measurement is not sufficient to determine secondary structure. If several different NOE contacts are measured in consecutive residues, secondary structure is reliably identified. See *Figure 34*. Note that α- and $3_{10}$ helices are difficult to distinguish; turns look like short helices.
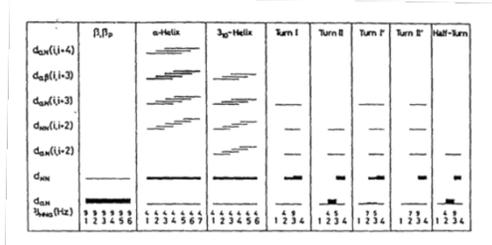


*Figure 34*

In larger proteins, 3D and 4D NMR measurements are used to separate the many peaks. Below is a 2D $^1$H-$^1$H NOESY spectrum separated into a $3^{rd}$ $^{15}$N dimension.
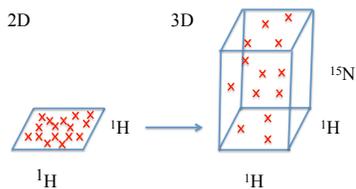


*Figure 35*

**Strategies for Structure Determination**

A number of different approaches are possible. All involve assigning the $^1$H chemical shifts, measuring NOEs and J-coupling constants between $^1$Hs, and using the NMR structural information in a computer algorithm to generate structures that agree with the data measured.

**Strategy 1**:    NOE-based sequential assignment of $^1$H resonances in unlabelled proteins.  This works for proteins up to about 20 kDa.

**Strategy 2**:  Uniform labelling with $^{15}$N and / or $^{13}$C and / or $^2$H can extend the $M_r$ limit to 25-50 kDa by increasing the dimensionality of the experiments.

---

$^1$H, $^{15}$N, and $^{13}$C atoms are assigned *via* triple resonance 3D and 4D experiments.  This can be done without using NOESY.

**Structure Determination**:

**1st Stage**:  Determine secondary structure from:

1.  Chemical Shift Index:

2.  Scalar Coupling Constants:

3.  NOE Patterns:

---

**2nd Stage:  Global Fold Determination**:

After thorough secondary structure determination, and with the use of a few long-range NOEs, a low-resolution picture of the **global fold** of a protein can be generated.    *Figure 35* shows the fold of ubiquitin.
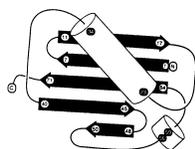


*Figure 36*
*Cavanagh et al., Protein NMR Spectroscopy*

### 3rd Stage:  Tertiary Structure Determination:

The principal structural information for structure calculation is the NOE.  Dihedral angles can be restrained by scalar coupling measurements and the CSI.  The restraints are used by computer algorithms that calculate families of structures that satisfy the experimental restraints. *Figure 37* outlines a strategy for tertiary structure determination.



*Figure 37*

Rule and Hitchens (2006 Fundamentals of Protein NMR Spectroscopy
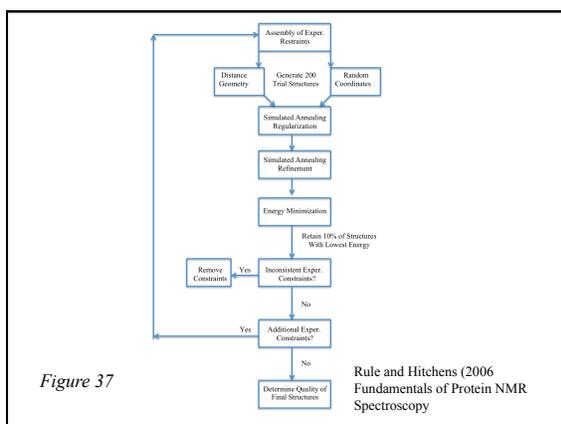
### Distance Geometry Calculations:

This is computationally expensive for large proteins.

If a molecule has N atoms then there are $\dfrac{N(N-1)}{2}$ interatomic distances that define the structure.

If the distances are not well determined, then N(N-1) upper and lower distance bounds are needed to define the structure.  These are stored in an N x N "bounds matrix".

| Atom | 1 | 2 | 3 | 4 | . | . | . | . | N |
|---|---|---|---|---|---|---|---|---|---|
| 1 | – | 4.4 | 9.6 | u | u | u | u | u | u |
| 2 | 1.7 | – | u | u | u | u | u | u | u |
| 3 | 3.5 | l | – | u | u | u | u | u | u |
| 4 | l | l | l | – | u | u | u | u | u |
| . | l | l | l | l | – | u | u | u | u |
| . | l | l | l | l | l | – | u | u | u |
| . | l | l | l | l | l | l | – | u | u |
| . | l | l | l | l | l | | l | – | u |
| N | l | l | l | l | l | l | l | l | – |

The lower distances (*l*) are determined by the covalent structure and van der Waals interactions.

The upper limits (*u*) are set by the covalent structure and experimentally-determined distances from NOEs, $^3$Js, and δs.

In a 10 amino acid peptide there are about 12,720 interatomic distances, or 25,440 upper and lower bounds.

Only about 586 distances (2.3%) will be determined by the covalent geometry. An excellent NMR structure would yield 20-30 structural constraints per residue or an additional 1.5% of the distances.

A direct calculation of Cartesian coordinates from the bounds matrix is not feasible.

Distance geometry calculates structures that are highly distorted (See *Figure 38*) and require refinement. Several procedures are possible:
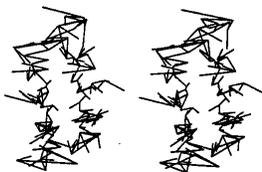


*Figure 38*

1. **Energy Minimization**

Distance-based (pseudo) energy violations are calculated:

$$P_{tot} = P_{cov} + P_{exp} + P_{nonbonded}$$ Where

$$P_{cov} = \Sigma K_{bonds}(d - d_0)^2 + \Sigma K_{angles}(d - d_0)^2 + \Sigma K_{rigids}(d - d_0)^2$$
$$+ \Sigma K_{chirals}(D)^2 + \Sigma K_{linears}(D)^2$$

$$P_{exp} = \Sigma K_{NOEs}(d - d_0)^2 + \Sigma K_{dihedrals}(d - d_0)^2 + \Sigma K_{ChemShifts}(d - d_0)^2$$

$$P_{nonbonded} = \Sigma K_{vdW}(d - d_0)^2$$

d = distance in structure

$d_o$ = upper or lower distance in bounds matrix.

A least-squares minimization of the penalty function is performed based on the atomic coordinates.

Poor structures will change rapidly during minimization and structures close to a minimum will change little.

---

As the minimizer runs only downhill, if a local minimum is reached there is no way to improve the structure of that molecule.

So a large number of minimizations are done and a family of structures with the lowest energies are retained.

**Simulated Annealing and Molecular Dynamics**:

These are used to push a molecule out of a local minimum so that it has the opportunity to reach the global energy minimum.

The dynamics of the molecule are simulated by heating up the molecule slowly and solving Newton's equations of motion for all atoms.

---

The molecule samples a wide conformational space and then is slowly cooled to its new energy minimum.

The heating and cooling are repeated many times to explore all possible conformations which satisfy the experimental and theoretical constraints and leading ideally to a global minimum.

**Restrained Molecular Dynamics**:

NMR constraints are added to MD force fields consisting of bond lengths, bond angles, repulsive and attractive van der Waals, and electrostatic terms.

The NMR constraints are added as square-well pseudoenergy terms. The structure is heated and cooled and the potentials drive the structure toward a conformation that reduces the violations of the restraints.

Once structures have been calculated they may be refined. The average structure may be used to back-calculate a theoretical NOESY spectrum using a complete dipolar relaxation rate matrix analysis. The agreement between experiment and theory, the **R-factor**, is calculated and can be used as a pseudoenergy term in the MD force field. This is exactly analogous to the refinement of X-ray diffraction structures.

Usually about 20 structures are calculated. The conformation may be represented by the **family** of conformations superimposed or by an **average structure**:

---

Let's examine an example from the literature.

*Solution structure and backbone dynamics of the pleckstrin homology domain of the human protein kinase B (PKB/Akt). Interaction with inositol phosphates.* By Daniel Auguin *et al*. ***Journal of Biomolecular NMR 28***: 137–155, 2004.

PKB is involved in cell signal transduction pathways, cell proliferation, and cancer.

The PH-domain is a 111 residue protein.

In the NMR spectra, peaks for the residues between $Pro_{42}$-$Pro_{51}$ are duplicated suggesting two conformations in slow equilibrium. This is likely due to *cis-trans* Pro isomerization.

---

Missing from the spectra are peaks for residues $Glu_{59}$-$Gln_{61}$ and $Arg_{76}$-$Thr_{87}$.

Chemical analysis with DTNB showed that $Cys_{60}$ and $Cys_{71}$ are in a disulphide bridge suggesting that isomerization ($\pm90°$) of the disulphide bond causes conformational exchange and **broadening** of the resonances.

*Figure 39A* below shows the best 20 structures in the family of NMR sructures. Only the N, Cα, and C' atoms are shown.

The protein is a 7 strand β-sandwhich flanked by a single α-helix.

The green segments are unconstrained because they are unassigned.

A

*Figure 39A*

Panel B shows a ribbon diagram of the lowest energy structure.

B

*Figure 39B*

The precision of the structures calculated depends on the number of experimental restraints added. *Figure 40* is a graph of the number of NMR constraints per resudue.

A

Number of constraints

Residue number

*Figure 40*
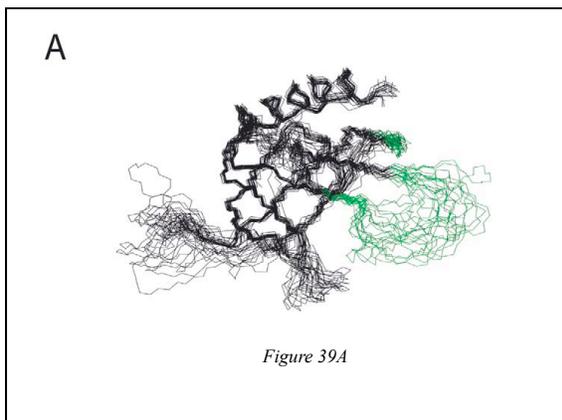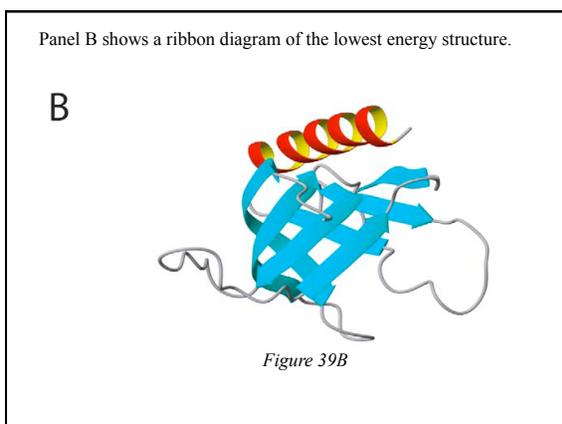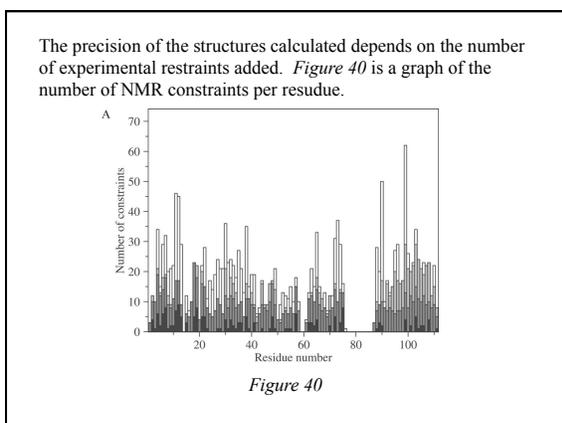
The agreement between the structures can be quantified by the **root mean squared deviations** of the positions of the atoms.

$$RMSD = \sqrt{\frac{\Sigma(r_i - r_j)^2}{n}}$$

Where $r_i$ - $r_j$ is the distance between 2 equivalent atoms and n is the number of pairs of atoms compared.

| Restraints/Residue | Backbone RMSD |
| --- | --- |
| 10 | 1.0-1.5Å |
| 15 | 0.5-1.0Å |
| 25 | 0.3-0.5Å |

*Figure 41*, is a plot of the RMSD per residue in the final NMR structures.
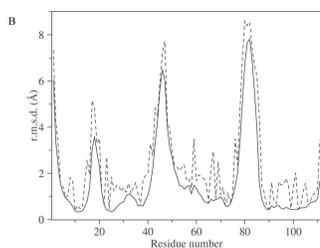


*Figure 41*

In this structure, some regions of the protein are very well-determined but some regions, such as the loops, do not converge to a single structure. This was shown by dynamic NMR experiments to be due to conformational fluctuations in solution.

*Table 3*

Experimental and structural statistics for the family of 20 structures of PKBβ-PHa

No. of distance constraints

| | |
| --- | --- |
| Intraresidue ($i - j = 0$) | 183 |
| Sequential ($i - j = 1$) | 365 |
| Medium-range ($i - j \leq 4$) | 157 |
| Long-range ($i - j > 4$) | 329 |
| Total | 1034 |

*Table 3*

No. of dihedral angle constraints

| | |
|---|---|
| $\Phi$ | 56 |
| $\psi$ | 56 |
| $\chi^1$ | 15 |
| Total | 127 |

| | |
|---|---|
| No. of hydrogen bonds | 34 |
| No. of disulfide bridges | 1 |

Mean r.m.s. deviations from experimental restraints

| | |
|---|---|
| NOE (Å) | $0.028\pm 0.003$ |
| Dihedrals (deg) | $1.162\pm 0.171$ |

*Table 3*

Mean r.m.s. deviations from idealized covalent geometry

| | |
|---|---|
| Bonds (Å) | $0.003\pm 0.001$ |
| Angles (deg) | $0.565\pm 0.004$ |
| Impropers (deg) | $0.360\pm 0.008$ |

Mean energies (kcal·mol$^{-1}$)

| | |
|---|---|
| $E_{NOE}$ | $12.92\pm 2.33$ |
| $E_{cdih}$ | $2.68\pm 0.78$ |
| $E_{vdW}$ | $-340.85\pm 15.42$ |
| $E_{bond}$ | $21.21\pm 0.79$ |
| $E_{improper}$ | $19.79\pm 0.89$ |

*Table 3*

| | |
|---|---|
| $E_{angle}$ | $165.02\pm 2.41$ |
| $E_{total}$ | $-602.27\pm 19.94$ |

Pairwise atomic rms differences (Å)

Residues 1–111 (all) BA[b]/HA[c]    $2.39\pm 0.57$ / $3.24\pm 0.49$

Residues 1–58, 62–75, 88–111 BA[b]/HA[c] $1.82\pm 0.49$ / $2.55\pm 0.44$

Residues from secondary structures BA[a]/HA[b] $0.76\pm 0.19$ / $1.53\pm 0.16$

*Table 3*

[a]For these calculations, the XPLOR all-hydrogen force fields topoallhdg and parallhdg were used. The final minimization of the 20 structures was carried out with force constants of 15 kcal mol$^{-1}$ Å$^{-2}$ and 50 kcal mol$^{-1}$ rad$^{-2}$ for the NOE and dihedral angle potentials, respectively.

[b]Backbone atoms.

[c] All heavy atoms.

At the bottom of the Table it is clear that the secondary structures are well defined whereas the loops are poorly defined.

# CHEM 4700 NMR Problems

## Due Tuesday October 4, 2011

1. Draw the 1 dimensional $^1$H-NMR spectrum of Asp in a polypeptide *i.e.* assume that the rest of the resonances are invisible and show only the peaks from Asp. You can assume that the exchangeable hydrogens on the carboxylic acid groups are not visible in the NMR spectrum. Draw and label the structure of Asp in a polypeptide and label the peaks in your spectrum. Draw a 2D-COSY and a 2D NOESY spectrum of the same compound.

2. The Department of Chemistry owns NMR spectrometers in which hydrogen atoms resonate at 300, 500, and 600 MHz. What are the field strengths of the magnets in Tesla units?

3. Canada's National High Field Biomolecular NMR Centre in Edmonton contains an 18.78 Tesla magnet. Calculate the resonance frequencies of the $^1$H, $^2$H, $^{13}$C, $^{15}$N, $^{31}$P, and $^{33}$S atoms.

4. Florida State University in Tallahassee operates the National High Field Magnet Laboratory that researches and designs very high magnetic fields for a variety of purposes including NMR spectroscopy. They have in operation a 900 MHz spectrometer with the largest bore in the world at 100 mm. The probe contains 95 miles of wound superconducting wire! A wide bore permits them to study larger samples than possible in standard magnet bores. They are currently working on developing a 25 Tesla magnet. At what frequency would protons resonate in such a field?

   FSU also designs pulsed magnets that can reach fields of 89 Tesla for periods of milliseconds. At what frequency would protons resonate in a 89 Tesla field?

5. In June of 2009, Bruker introduced the highest field commercially-available NMR spectrometer to date. In it, protons resonate at 1GHz. What is the field strength in Tesla? The spectrometer costs 11.7 megaEuros. What is its cost in Canadian dollars?

6. The earth's magnetic field is about 0.00005 Tesla. What is the resonance frequency of a $^1$H atom on earth?

7. In an 18.8 Tesla magnet the resonances of an amide H doublet of an amino acid in an α-helix are located at 7.980 and 7.975 ppm. What is the $^3$J-coupling constant between the amide H and the α-H? In the same magnetic field the resonances of a different amide H doublet of an amino acid residing in a β-strand are located at 9.67 and 9.66 ppm. What is the J-coupling constant between the amide H and the α-H?

8. Unpaired electrons in atoms make NMR measurements difficult. Which parameter in Table 1 is pertinent to this? Why?

9. A peak on an NMR spectrum has a signal / nose intensity of 1600. The spectrum was acquired by adding together 64 scans. What would the signal / noise intensity have been if 256 scans had been added together? What would the ratio have been for just 1 scan?

10. Explain how the chemical shift index works and what information about protein structure it provides.

11. The Boltzmann distribution of spin 1/2 nuclei in the lower ($\alpha$) and upper ($\beta$) energy states can be written:

$$N\alpha/N\beta = \exp\text{-}[(E\alpha\text{-}E\beta)/kT] = \exp\text{-}[(\text{-}\gamma hB_0)/2\pi kT]$$

From this it can be shown that, in the high temperature limit where $kT \gg \gamma hB_0/2\pi$, the population difference between the two states is:

$$N\alpha - N\beta = N\gamma hB_0/4\pi kT$$

Calculate the population difference between the two states for a population of 5000 thousand $^1$H atoms in an 11.7 Tesla field and in an 18 Tesla field at 325 K. Calculate the population difference for 5000 thousand $^{13}$C atoms in the two fields.

10 bonus marks (out of 100) will be awarded to anyone who can show how to <u>derive</u> the second equation from the first.

# Purification of GFP-Calmodulin

Here is a summary of what you should learn in this laboratory:

DNA shearing, centrifugation, heat denaturation-precipitation, hydrophobic interaction chromatography, UV absorption spectrophotometry, dialysis, and freeze-drying.

Calmodulin (CaM) is a small ($M_r$ = 16,690), acidic, Ca(II)-dependent regulatory protein found in all eucaryotic cells.  It contains four high-affinity calcium-binding sites that each are constructed of two α-helices connected by a 13-residue binding loop.  When cellular calcium levels rise, CaM binds calcium and undergoes a conformational change exposing a previously hidden hydrophobic surface.  CaM uses this hydrophobic surface to bind a large variety of enzymes that contain CaM-binding sites.  Upon binding CaM, many of these enzymes are activated and some are deactivated.  We will take advantage of the calcium-induced conformational change to purify a Green Fluorescent Protein-Calmodulin fusion protein using hydrophobic interaction chromatography.  Here is a www site with more information about CaM: http://structbio.vanderbilt.edu/cabp_database/general/prot_pages/calmod.html
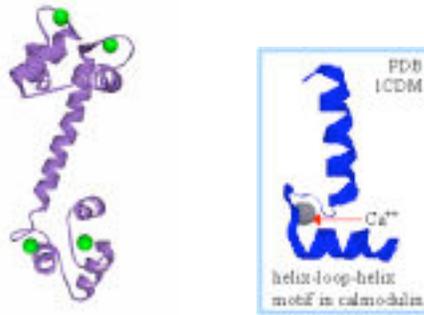


Green Fluorescent Protein (GFP) is a product of the Pacific Jellyfish, *Aequoria victoria*.  It transduces the chemiluminescent blue light produced by aequorin into green fluorescence.  It does this using an intrinsic chromophore constructed by cyclization and dehydrogenation of its polypeptide backbone. The protein is a powerful protein expression and localization reporter when fused to other proteins of interest.  It is formed by 11 β-strands arranged in a barrel structure surrounding a single α-helix that contains the chromophore.  We thank Ray Zielinski (University of Illinois) who constructed and sent us the GFP-Calmodulin expression vector.
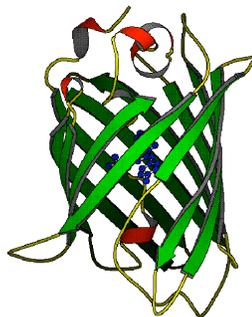
# Materials and Methods

*E. coli* cells containing a recombinant GFP-CaM fusion have been stored in a centrifuge bottle at -20 °C and thawed 2 h prior to the experiment.  They have been through 2 freeze-thaw cycles to lyse the cells and the enzymes Dnase , Rnase, and lysozyme have been added to reduce the viscosity of the cell extract.

## First Laboratory Period

1.  Incubate the cells in approximately 20 volumes of Buffer A per volume of cells at room temperature with gentle rocking for 15 min.

2.  Sonicate the suspended cells to shear the DNA.  Give short bursts of 10-15s with an equal time on ice between bursts.  Test after each burst whether the suspension will form drops at the end of a Pasteur pipette.  Once drops can form proceed to step 3.

3.  Transfer the lysate to an appropriate number of 50-mL centrifuge tubes.  Don't overfill them or they will spill in the centrifuge. The filled bottles must be of equal mass so that they balance properly in the centrifuge.  Use a Harvard Trip Balance for this purpose.

4.  Centrifuge the 50 mL tubes at 15,000 x g (13,500 rpm for the small rotor filled with 50 ml tubes) at 4 °C using the refrigerated centrifuge in Dr. Duckworth's laboratory (526 Parker) for 30 min.

5. GFP-CaM is water-soluble and will be in the aqueous supernatant. After centrifugation, collect the supernatant into an Erlenmeyer flask. The pellet can be discarded.  Ask the demonstrator about how to do this properly.

6.  Stir the contents of the Erlenmeyer flask with a magnetic stirrer.  Heat the protein extract at $65^{o}C$ for 3 minutes, then immediately chill the solution by pouring the extract into a clean centrifuge tube chilled in an ice-water bath.

7.  Add $CaCl_2$ to a total concentration of 3 mM and mix thoroughly.  That is a 2 mM excess over the 1 mM EDTA that is present in the buffer. The $M_r$ of $CaCl_2 \cdot H_2O$ is 147.02 g.

8.  Transfer the slurry to centrifuge tubes and again centrifuge at 15,000 x g for 30 min.  Decant the supernatant into a beaker. The pellet can be discarded.

11. Place your solution in the freezer until the next laboratory period.  It must be removed from the freezer about 2 h before the laboratory begins.

# Second Laboratory Period

1.  Prepare the column.  Make a small plug of glass wool and push it to the bottom of the column to retain the resin.  Pour 10 mL of Buffer I into the column and mark the level on the column with a marker.  This will be the height of the resin.  This volume is the bed volume of the column.

2.  Make a slurry of phenyl sepharose in Buffer I and pour some into the column.  Let it stand.  When it starts to settle, open the stopcock and allow some buffer to flow out.  Add more phenyl sepharose to the column.  Do not let the column completely settle while you are adding the phenyl sepharose.  You will have to guess when to stop adding the phenyl sepharose to get the final volume required.  It does not have to be perfect.

3.  Once the resin has settled, equilibrate the column by rinsing it with 50 mL of Buffer I.  Stop the flow as soon as the buffer level is at the top of the column.  Do not allow the column to dry out.

4.  Add the GFP-CaM solution to the column and open the stopcock.  Allow the protein to go slowly through the column.  Keep adding more solution to the top of the resin.  (The higher the sample level, the faster the flow rate.)  The flow rate should be about 100 mL/h, so it should take about 30 min for the sample to go onto the column.  The eluant may be cloudy; collect it into a beaker.

5.  Once the entire sample has entered the top of the column, wash the column with 5 bed volumes (50 mL) of Buffer I.  Adjust the flow rate to about 150 mL/h so that it takes about 20 min for the buffer to flow through.  Collect the eluant into a beaker.

6.  Repeat this procedure with Buffer I with Salt.  This buffer has added NaCl to help dislodge any basic proteins sticking to the GFP-CaM by ionic attractions.

7.  Set up and number about 15 test tubes.  Elute the GFP-Calmodulin using Buffer I with EGTA.  Collect 2.5-mL fractions.

8.  Look for the GFP-CaM-containing fractions by monitoring the UV absorbance at 280 nm. You should see low absorbance and then a sudden rise in absorbance, which will peak and taper off slowly. Record the absorbance for each fraction number.

9.  The GFP-CaM-containing fractions should be pooled, added to dialysis tubing, and twice dialyzed exhaustively against 0.1 M ammonium bicarbonate at 4 °C.

10.  The demonstrator will shell-freeze and then freeze-dry the dialysed CaM.  Your CaM will be available for further analysis in future experiments.

11.  Wash your column with 5 column volumes of 20% ethanol.

## **Buffers**

Buffer A   (1.5 litres)

- 250 mM Tris
- 1 mM 2-mercaptoethanol
- 1 mM EDTA
- pH 7.5

Buffer I   (1 litre)

- 250 mM Tris
- 1 mM 2-mercaptoethanol
- 0.1 mM CaCl·2H$_2$O
- pH 7.5

Buffer I with salt (1 litre)

- 250 mM Tris
- 1 mM 2-mercaptoethanol
- 0.1 mM CaCl·2H$_2$O
- 500 mM NaCl
- pH 7.5

Buffer I with EGTA  (0.5 litre)

- 250 mM Tris
- 1 mM 2-mercaptoethanol
- 1 mM EGTA
- pH 7.5

# Introduction to Hydrophobic Interaction Chromatography

From *Hydrophobic Interaction Chromatography: Principles and Methods* published by Amersham Pharmacia Biotech, 1990.

In a classical paper published in 1948 and entitled: "*Adsorption Separation by Salting Out*" Tiselius [1] laid down the foundation for a separation method which is now popularly known as **hydrophobic interaction chromatography** (HlC). He noted that, "*...proteins and other substances which are precipitated at high concentration of neutral salts (salting out), often are adsorbed quite strongly already in salt solutions of lower concentration than is required for their precipitation, and that some adsorbents which in salt-free solutions show no or only slight affinity for proteins, at moderately high salt concentrations become excellent adsorbents*". Since then, great strides have been made in developing almost ideal stationary phases for chromatography (such as cellulose, cross-linked dextran (Sephadex$^{TM}$), cross-linked agarose (Sepharose CL, Sepharose High Performance and Sepharose Fast Flow), and in developing coupling methods for immobilizing ligands of choice [2, 3] to such matrices. It was a combination of these two events which, in the beginning of 1970's, led to synthesis of a variety of hydrophobic adsorbents for biopolymer separations based on this previously rarely exploited principle.

The first attempt at synthesizing such adsorbents was made by Yon [4] followed by Err-el *et al*. [5], Hofstee [6] and Shaltiel & Er-el [7]. Characteristically, these early adsorbents showed a mixed ionic-hydrophobic character [8]. Despite this, Halpern *et al*. [9] claimed that protein binding to such adsorbents was predominantly of a hydrophobic character. Porath *et al*. [10] and Hjertén *et al*. [11], later synthesized charge-free hydrophobic adsorbents and demonstrated that the binding of proteinA was enhanced by high concentrations of neutral salts, as previously observed by Tiselius [1], and that elution of the bound proteins was achieved simply by washing the column with salt-free buffer or by decreasing the polarity of the eluent [6, 10, 11].

## Theory

The many theories that have been proposed for HIC are essentially based upon those derived for interactions between hydrophobic solute and water [22,23], but none of them has enjoyed universal acceptance. What is common to all is the central role played by the structure-forming salts and the effects they exert on the individual components (i.e., solute, solvent and adsorbent) of the chromatographic system to bring about the binding of solute to adsorbent. In view of this, Porath (24) proposed *"salt-promoted adsorption"* as a general concept for HIC and other types of solute-adsorbent interactions occurring in the presence of moderately high concentrations of neutral salts.

Hofstee [6] and later Shaltiel [7] proposed *"hydrophobic chromatography"* with the implicit assumption that the mode of interaction between proteins and the immobilized hydrophobic ligands is similar to the self-association of small aliphatic organic molecules in water. Porath *et al*. [10] suggested a salting-out effect in hydrophobic adsorption, thus extending the earlier observations of Tiselius [1]. They also suggested that " . . .*the*

*driving force is the entropy gain arising from structure changes in the water surrounding the interacting hydrophobic groups"*. This concept was later extended and formalized by Hjertén [25] who based his theory on the well known thermodynamic relationship $\Delta G = \Delta H - T\Delta S$. He proposed that the displacement of the ordered water molecules surrounding the hydrophobic ligands and the proteins leads to an increase in entropy ($\Delta S$) resulting in a negative value for the change in free energy ($\Delta G$) of the system. This implies that the hydrophobic ligand-protein interaction is thermodynamically favourable, as is illustrated in Fig. 1. An alternative theory is based on the parallelism between the effect of neutral salts in salting out (precipitation) and H1C [26,27]. According to Melander and Horvath [27], hydrophobic interaction is accounted for by increase in the surface tension of water arising from the structure - forming salts dissolved in it. In fact, a combination of these two mechanisms seems to be an obvious extension and has been exploited long Before HIC adsorbents were synthesized [28]. Finally, Srinivasan and Ruckenstein (29) have proposed that HIC is due to van der Waals attraction forces between proteins and immobilized ligands. The basis for this theory is that the van der Waals attraction force between protein and ligand increase as the ordered structure of water increases in the presence of salting out salts.

*Figure 1*.
Close to the surface of the hydrophobic ligand and solute (L and H), the water molecules are more highly ordered than in the bulk water and appear to "shield off" the hydrophobic ligand and solute molecules. Added salt interacts strongly with the water molecules leaving less water available for the "shielding off" effect, which is the driving force for L and H to interact with each other.
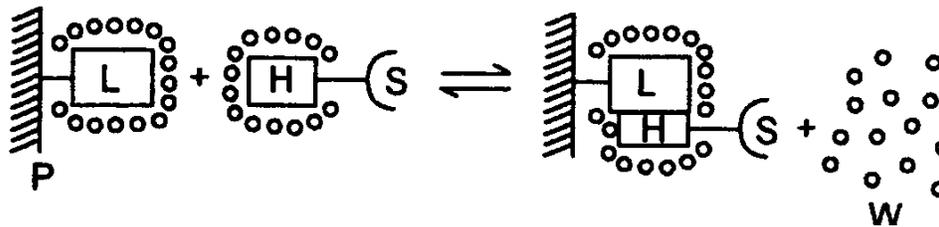


P = polymer matrix
S = solute molecule
L = Ligand attached to polymer matrix
H = Hydrophobic patch on surface of solute molecule
W = Water molecules in the bulk solution

## HIC vs RPC

In theory, HIC and ***reverse-phase chromatography*** (RPC) are closely related LC techniques. Both are based upon interactions between solvent-accessible non-polar adsorbents (hydrophobic patches) on the surface of biomolecules and the hydrophobic ligands (alkyl or aryl groups) covalently attached to the gel matrix. In practice, however, they are different. Adsorbents for RPC are more highly substituted with hydrophobic ligands than HIC adsorbents. The degree of substitution of HIC adsorbents is usually in

the range of 10-50 μmoles/mL gel of $C_2$-$C_8$ alkyl or simple aryl ligands, compared with several hundred μmoles/mL gel of $C_4$-$C_{18}$ alkyl ligands usually used for RPC adsorbents. Consequently, protein binding to RPC adsorbents is usually very strong, which requires the use of non-polar solvents for their elution. RPC has found extensive applications in analytical and preparative separations of mainly peptides and low molecular weight proteins that are stable in aqueous-organic solvents.

In summary, HIC is an alternative way of exploiting the hydrophobic properties of proteins, working in a more polar and less denaturing environment.

Compared with RPC, the polarity of the complete system of HIC is increased by decreased ligand density on the stationary phase and by adding salt to the mobile phase.

## Factors Affecting HIC

The main parameters to consider when selecting HIC media and optimizing separation processes on HIC media are:

- Ligand type and degree of substitution
- Type of base matrix
- pH
- Temperature
- Additives.

## Effects of Salts

The addition of various structure-forming ("salting-out") salts to the equilibration buffer and sample solution promotes ligand-protein interactions in HIC (10, 12, 36, 65, 66). As the concentration of such salts is increased, the amount of proteins bound also increases almost linearly up to a specific salt concentration and continues to increase in an exponential manner at still higher concentrations.

The effects of salts in HIC can be accounted for by reference to the Hofmeister series for the precipitation of proteins or for their positive influence in increasing the molal surface tension of water (for extensive review, see refs. 27, 29). These effects are summarized in Tables 1 and 2.

Table 1: The Hofmeister series on the effect of some anions and cations in precipitating proteins.

← Increasing precipitation ("salting-out") effect
Anions: $PO_4^{3-}$, $SO_4^{2-}$, $CH_3$, $COO^-$, $Cl^-$, $Br^-$, $NO_3^-$, $ClO_4^-$, $I^-$, $SCN^-$
Cations: $NH_4^+$, $Rb^+$, $K^+$, $Na^+$, $Cs^+$, $Li^+$, $Mg^{2+}$, $Ca^{2+}$, $Ba^{2+}$
Increasing chaotropic ("salting-in") effect →

Table 2: Relative effects of some salts on the molal surface tension of water.

$Na_2SO_4 > K_2SO_4 > (NH_4)_2SO_4 > Na_2HPO_4 > NaCl > LiCl. . . > KSCN$

In both instances, sodium, potassium or ammonium sulphates produce relatively higher "salting-out" (precipitation) or molal surface tension increment effects. It is also these salts that effectively promote ligand-protein interactions in HIC. Most of the bound proteins are effectively desorbed by simply washing the HIC adsorbent with water or dilute buffer solutions at near neutral pH.

### *Effect of pH*

The effect of pH in HIC is also not straightforward. In general, an increase in pH weakens hydrophobic interactions [10, 41], probably as a result of increased titration of charged groups, thereby leading to a decrease in the hydrophobicity of the proteins. On the other hand, a decrease in pH results in an apparent increase in hydrophobic interactions. Thus, proteins which do not bind to a HIC adsorbent at neutral pH bind at acidic pH [9]. Hjertén *et al*. [42] found that the retention of proteins changed more drastically at pH values above 8.5 and/or below 5 than in the range pH 5-8.5.

These findings suggest that pH is an important separation parameter in the optimization of hydrophobic interaction chromatography and it is advisable to check the applicability of these observations to the particular separation problem at hand.

## *Effect of Temperature*

Based on theories developed for the interaction of hydrophobic solutes in water [22,37], Hjerten [38] proposed that the binding of proteins to HIC adsorbents is entropy driven, which implies that the interaction increases with an increase in temperature. [$\Delta G = \Delta H - T\Delta S \sim T\Delta S$]. However, also to be considered are the differential effects exerted by temperature on the conformational states of different proteins and their solubilities in aqueous solutions.

# Circular Dichroism of Nucleic Acids
## CHEM 4700 Advanced Biochemistry Lab

Molecular Chirality

Most biological molecules are chiral, that is, they have neither a mirror plane nor a centre of symmetry. A chiral molecule cannot be superimposed on its mirror image, and the two forms have the same relationship as do our left and right hands. The simplest examples of chirality result from the presence of tetrahedral carbons to which 4 distinct groups are bonded ('chiral' or 'asymmetric' carbons); such carbons are commonly the sole source of the chirality of small molecules, such as nucleotides, amino acids, and monosaccharides. Chirality, however, has no absolute requirement for the presence of chiral C atoms. Equally interesting sources of chirality are left-handed (LH) and right-handed (RH) secondary folding of the helical segments of nucleic acids, proteins, and polysaccharides, as well as the LH and RH tertiary folding of these polymers.

Polarized Light

Chiral structures can be distinguished and characterized by polarized light. Electromagnetic radiation can be described by perpindicular electric (E) and magnetic (M) field vectors that oscillate with a characteristic frequency in time and in space. For electronic absorption spectroscopy the magnetic component can be ignored.

http://www.cps-amu.org/sf/notes/mw4-2-3.htm
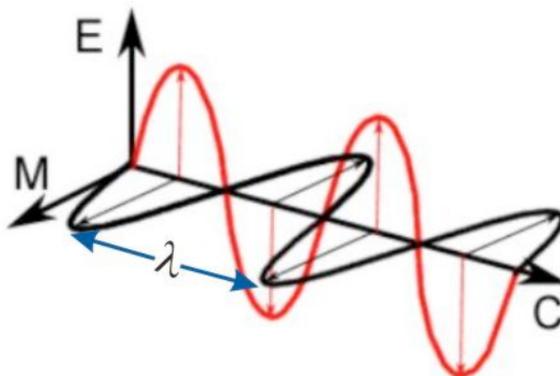
In unpolarized light, E oscillates in all directions perpendicular to the direction of propagation of the light beam. In plane polarized light (PPL), E oscillates in a single plane that includes the direction of propagation: at a particular point in space, E appears to be oscillating along a line and for this reason PP is also called linearly polarized light (LPL). For an excellent animation of this please visit the www site: http://www.enzim.hu/~szia/cddemo/edemo0.htm. The author will also permit you to download the animations so you can run them on your computer without having to be connected to the internet.

When two plane polarized waves that are polarized in perpindicular planes but are exactly 90$^o$ out of phase are superimposed the result is  circularly polarized light (CPL). Here, the E vectors at any given time define a helix in space: a RH helix in right circularly polarized light (RCPL) and a LH helix in left circularly light (LCPL). In fact, PPL may be viewed as having 2 components (with the same frequency and intensity), one component being LCPL, the other, RCPL.
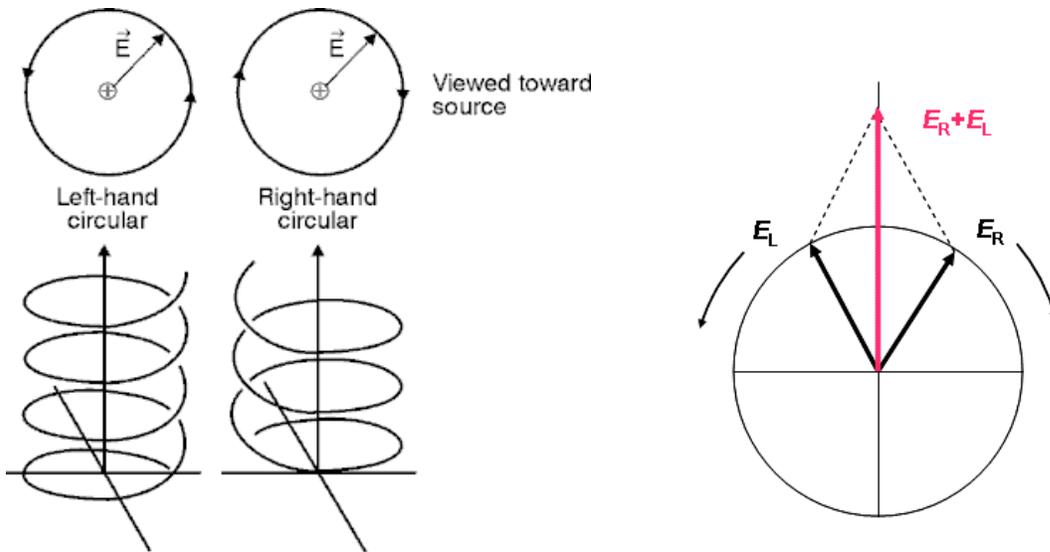
Optical Activity

Chiral molecules are optically active, that is, they interact differently with LCPL & RCPL. First, in a medium with chiral molecules, LCPL & RCPL (see EL and ER in the figure below) will be absorbed to different extents ($\varepsilon_L \neq \varepsilon_R$, is the extinction coefficient). Second, they will travel with different velocities through the medium, that is, the refractive index (n) of the medium will be different for the LCPL & the RCPL ($n_l \neq n_R$). The *circular dichroism* (CD) of a solution is defined as $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$, whereas *circular birefringence* of the solution is defined as: $\Delta n = n_L - n_R$. In passing through a chiral medium, a PPL beam with be affected in two ways. First, because of the CD of the solution, it will exit the solution as a beam of *elliptically polarized light* (EPL), either right EPL (REPL) or as left EPL (LEPL). Second, because of the circular birefringence of the solution, the beam will exit with the major axis of the ellipse rotated—relative to the plane of the incident beam of PPL—in the clockwise sense (positive angle α, or dextrorotatory solution) or in the counterclockwise sense (negative angle α, or levorotatory solution).
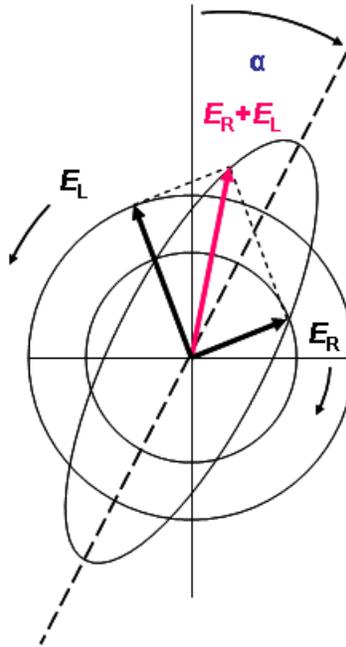
Circular Dichroism (CD)

CD results from the differential absorption of the LCPL and RCPL components of the PPL incident on the medium. With A defined as the absorbance of a solution, we can write:

1) $\Delta A = A_L - A_R$, where
2) $A_L = \varepsilon_L cl$ and $A_R = \varepsilon_R cl$ ($l$ = cell length in cm; $c$ = concentration in moles/liter).
3) $\Delta A = \Delta\varepsilon cl$.

When PPL passes through a chiral medium, the emerging beam is found to be elliptically polarized. The *ellipticity* of the emerging beam ($\Theta$) is directly proportional to $\Delta A$ and is related to it by the expression

4) $\Theta$ (degrees) $= 2.303(A_L - A_R)180°/4\pi$

Note that, because $A_L - A_R$ can be positive or negative, $\Theta$ is accordingly positive or negative.

Though a spectropolarimeter measures $A_L - A_R$, CD data are normally converted (*via* Eqn 4) into the corresponding ellipticities, and the output is calibrated in degrees (actually, millidegrees since $\Theta$ is generally very small; $<1°$).

Presentation of Data
You will see from the CD spectra (plots of $\Theta$ versus $\lambda$) for 5'AMP and for ApC that $\Theta$ varies with $\lambda$; it also varies with the number of chiral molecules that the beam encounters in its passage through the cell (l) and the concentration of the material, (c). So, if we wish to compare $\Theta$ values for 2 molecules at some $\lambda$, we must define quantities for unit length and unit concentration. The following definitions are commonly used:

5a)  Specific ellipticity: $[\Theta] = \dfrac{\Theta}{dc} = \dfrac{100\Theta}{lc}$

where d is the cell length in decimeters (1dM = 10 cm), l is the cell length in cm, and c is the concentration in g/mL

5b) Molar ellipticity: $[\Theta]_M = \dfrac{M[\Theta]}{100} = \dfrac{M\Theta}{10lc}$     Units: See 5c below.

where M is the molecular weight of the material in g/mole, and the 'subscripted M' is a label which defines 'Molar Ellipticity'. The divisor is added for convenience, that is, to make smaller (scale down) the numbers. It can be shown (with some difficulty) that:

5c) $[\Theta]_M = \dfrac{100\Theta}{Cl}$ Units: deg•C$^{-1}$•cm$^{-1}$.

Where C is the concentration in moles/liter, and l is the cell length in cm. For polymers, it is useful to know the *average* contribution to $[\Theta]_M$ from the residues in the chain. This is obtained by:

5d) The Mean Ellipticity: $[\Theta]_n = \dfrac{[\Theta]_M}{n}$

where n = the number of residues in the chain, and the 'subscripted' n is a label.
Once you have calculated $[\Theta]_M$ from eqn. 5c, you can calculate $\Delta\varepsilon$ from:

6) $[\Theta]_M = 3{,}300\Delta\varepsilon$

which gives $\Delta\varepsilon$ in the familiar units C$^{-1}$•cm$^{-1}$, or liter•mol$^{-1}$•cm$^{-1}$. (See CHEM4620 notes on UV absorption.)

Procedure in the Laboratory
The object of this experiment to introduce you to a CD spectropolarimeter, and to show that chiral biological molecules have characteristic CD spectra, which can change dramatically when the temperature or the nature of the solvent is changed.

1. You will have prepared for you the following solutions: a) 5'AMP in water; b) 5'AMP in a water-ethylene glycol mixture (v/v = 40/60), c) ApC in water, d) ApC in the water-ethylene glycol mixture.

2. A CD spectrum will be obtained for a), b) and d) at room temperature, and for c) at low temperature (10ºC), room temperature, and high temperature (60ºC).

3. You will be provided with spectra in both millidegrees and $[\Theta]_M$ or $[\Theta]_n$. In your report, provide a sample calculation of the latter units from the former.

4. Submit all of your original spectra (or copies) with your report.


Lab Report

Your report should contain the following sections:

**I** Introduction: This should contain a brief discussion which includes a statement of the purpose of the experiment.

**II** Experimental: A brief discussion of how the experiment was carried out and sample calculations. Please do not re-copy the laboratory manual. Include the name and model number of the spectropolarimeter and the manufacturer.

**III** Results: This section includes tables & spectra, and about 1 page describing the general shapes of the plots. So, compare the relative intensities for the ApC, the 5'AMP spectra, and the DNA spectra, and note any changes, if any, in position (λ), and intensity of maxima and minima when solvent or temperature is changed.

**IV** Discussion: Normally this would be the longest section. Here you discuss your observations in **III** in terms of molecular structure, that is, comment on the effect of temperature and solvent on molecular conformation.

**V** List of references. Your report should be appropriately referenced with articles or books that you have read in preparing your report. See a recent issue of the *Canadian Journal of Chemistry* or of the *Journal of the American Chemical Society* for acceptable formats for listing references.

Marking

Spelling and composition will be considered.


References

Bloomfield, V, Crothers, D & Tinoco, I, Jr. *Physical Chemistry of Nucleic Acids*.

Cantoni GL & Davies, DR. *Procedures in Nucleic Acid Research Vol 2*.

Cantor, C & Schimmel, P, *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure & Function*.

Fasman, GD & Timasheff, SN. *Fine Structure of Protein & Nucleic Acids, Vol 4*.

Freifelder, D. *Physical Biochemistry*.

Hennessey & Johnson, *Analytical Biochemistry*, 125, 177-88 (1982).

Johnson, WC. *Methods in Biochemical Analysis* 31, 61 (1987). Nicolai L & Baron, M. *Device to describe wave polarization*. J. Chem. Educ. 67. 647 (1990).

Ts'o. POP. *Basic Principles of Nucleic Acid Chemistry. Vols. I & II*.

Van Holde, KE. *Physical Biochemistry, 2nd edition*.

Circular Dichroism and Fluorescence of Proteins

Introduction:

The purpose of this laboratory is to introduce you to circular dichroism (CD) and fluorescence as tools for obtaining structural information about proteins. You will also conduct a protein denaturation experiment. Much of the background material is covered in the CHEM 4620 and CHEM 4630 lectures.

In brief, the side chains of Trp, Tyr, and Phe absorb ultra-violet light in the near UV region (250-300 nm). All 3 aromatic residues also emit light of lower energy. Trp has by far the highest quantum yield and it dominates the fluorescence spectra of most proteins. Trp emits light with a maximum intensity at 350 nm when dissolved in water and at lower wavelengths when dissolved in organic solvents. Trp side chains located in the hydrophobic core of a protein have fluorescence spectra that are shifted to the blue, or lower wavelengths, compared to Trp that reside on the surface of a protein and are in contact with water. Changes in Trp fluorescence can be used to follow protein folding and unfolding. Green fluorescent protein has additional fluorescence in the visible region of the light spectrum. It absorbs UV blue light and emits visible green light. In a folded protein Trp, Tyr, and Phe residues usually exhibit a weak circular dichroism signal owing to asymmetry in their environments. Observations of near UV CD signals are used as evidence that a protein has a well packed tertiary structure.

Protein peptide bonds absorb UV light below 230 nm. If the polypeptide is folded into an asymmetric structure such as a right handed α-helix or a twisted β-sheet, the polypeptide will exhibit a circular dichroism spectrum characteristic of its secondary structure.

All asymmetric molecules are said to be optically active. This includes many of the molecules made by living organisms. Optical activity originates with asymmetric atoms like the α-carbons of α-amino acids and the helical structures that are found in proteins. It can be detected by optical rotation (OR), optical rotatory dispersion, ORD, (the dependence of OR on wavelength), or circular dichroism (CD) spectroscopy.

CD and ORD depend on the existence of groups which are both optically active and chromophoric. Proteins have such groups. The side-chains of Trp, Tyr, and Phe all absorb ultraviolet light around 260-290 nm ("the aromatic " and many side-chains and the peptide bond absorb at 230 nm and below ("the peptide"). Their optical activity is the result of asymmetry in their environments.

Circular dichroism can be thought of as follows: Light polarized in a plane can be broken down into 2 components which are said to be circularly polarized *i.e.* the electric vectors of the light follow the path of right- and left-hand screws equal in magnitude. When they encounter an optically active chromophore each component will be absorbed to a different extent. As a result, the electric vectors are no longer equal in magnitude and the light leaving the sample is elliptically polarized. The ellipticity, $\theta$, is proportional to the difference in absorbance of the 2 components, $A_L - A_R$.

<u>Definitions:</u>

<u>Molar Ellipticity:</u>

The molar ellipticity $[\theta_\lambda]_M$ of a protein is defined as:

$$[\theta_\lambda]_M = \frac{\theta_\lambda M}{10lc} \qquad \text{......1}$$

where $\theta_\lambda$ is the measured ellipticity in millidegrees at the wavelength $\lambda$ (nm); M is the relative atomic mass of the protein; l is the cell path length (cm); and c is the protein concentration (mg/ml). The units for $[\theta_\lambda]_M$ are thus $\text{deg} \cdot \text{cm}^2 \cdot \text{dmole}^{-1}$.

<u>Mean Residue Ellipticity</u>

The relative masses of proteins range from about 1000 to several millions. However, the mean residue mass of most proteins is about $110 \pm 5$. The mean residue mass is just the mass per amino acid residue, M/n, where n is the number of amino acids in the protein.

We can now rewrite equation 1 as:

$$[\theta_\lambda] = \frac{\theta_\lambda M_o}{10lc} \qquad \text{........2}$$

where $M_o$ is M/n and $[\theta_\lambda]$ is the mean residue ellipticity of a protein. This value expresses the molar ellipticity of a protein in the far UV region in terms of the molar concentration of peptide units. Sometimes this is unclear in the literature and mean residue ellipticity is called molar ellipticity where it is understood that in the far UV region the molarity referred to involves peptide bonds.

<div align="center"><u>Protein Secondary Structures by CD</u></div>

The broadest classification of protein secondary structure involves 3 distinct types, namely helical (including $\alpha$ and $3_{10}$ helices), $\beta$-structure (parallel and antiparallel) and irregular or unordered structures which is a class broadly defined here as those parts of the polypeptide backbone which are not helix or $\beta$-sheet. The CD spectrum of a protein in the far UV region (250-180 nm) can be considered to be comprised of ellipticity arising from each of the three types of structure. This idea is expressed in equation 3 below:

$$[\theta_\lambda] = f_h[\theta_\lambda]_h + f_\beta[\theta_\lambda]_\beta + f_i[\theta_\lambda]_i \qquad \text{.......3}$$

where $f_h$ is the fraction of protein that is helical, $f_\beta$ is the fraction of protein that is in $\beta$–sheet, $f_i$ is the irregular fraction, and the $[\theta_\lambda]_{h,\beta,i}$ are the corresponding ellipticities arising from each of the 3 types of secondary structure. These can be obtained from CD spectra of model polypeptides

entirely in one conformation or they can be deduced from the decomposition of the CD spectrum of a protein of known, or presumed, secondary structure. Of course each $f_x \geq 0$ and $\Sigma f_x = 1$.

Equation 3 contains three unknowns and is impossible to solve by itself. Measurements at three different wavelengths would yield three equations still with only three unknowns which can be solved. In practice, computer programs can use the entire measured CD spectrum to decompose the protein into $f_h$, $f_\beta$, and $f_i$. In effect, CD spectra are calculated for different contents of secondary structure until the minimum variance between the calculated and experimental spectrum is obtained.

During the laboratory, the demonstrator will assist you in the measurement of some protein CD and fluorescence spectra using the Jasco 810 spectropolarimeter-fluorometer. Make a note of the parameters used so you can show an example calculation of the unit conversions in your report.

Experimental:

1. With the help of the demonstrator, measure a fluorescence spectrum of GFP-Calmodulin.

2. With the help of the demonstrator, measure a far UV CD spectrum of GFP-Calmodulin.

3. Measure a far UV CD spectrum of native, folded ribonuclease (RNase).
   $M_o = 110$; $M = 13,700$;

4. Follow the denaturation of RNase by measuring far UV CD spectra of the protein dissolved in 1 M, 2 M, 3 M, 4 M, 5 M, and 6 M guanidine HCl.

5. Follow the thermal denaturation of RNase by measuring far UV CD spectra between 20ºC and 80ºC.

Data Analysis for the Lab Reports.

1. Your CD spectra will be deconvoluted by the demonstrator using software available on-line at www.cryst.bbk.ac.uk/cdweb/html/home.html. Unfortunately, you can only do this yourself if you apply for permission, as the site is password protected.

2. Plot a denaturation curve for RNase in guanidine hydrogen chloride *i.e.* plot $[\theta_\lambda]$ or $\theta_\lambda$ vs. guanidine concentration.

3. Make an appointment to see J. O'Neil who will assist you to curve fit your thermal denaturation results to obtain the melting point using the software *Mathematica*.
   In your lab report comment on whether or not the curve is reversible.

4. Compare the fluorescence spectra you obtained for GFP-CaM and the CD spectra you obtained for ribonuclease, with those in the literature. For CD spectra, see the Handbook of Biochemistry and Molecular Biology, third edition.

5. Show an example calculation of the conversion of millidegrees to molar or mean residue ellipticity using data from the spectra you were given.

**HYDROGEN EXCHANGE LABORATORY**

1. OVERVIEW OF PROTEIN DYNAMICS.

2. AMIDE HYDROGEN EXCHANGE CHEMISTRY.

3. DATA ANALYSIS FOR THE EXPERIMENT.

1. **PROTEIN DYNAMICS**

The central dogma of structural biology states that molecular structure determines biological function.
There are several problems with this with respect to proteins.

---

i. X-ray diffraction structures are deduced under *non*-native conditions, *e.g.* high salt concentrations. Proteins are not crystalline in the cell.

NMR spectroscopy determines protein structures in solution but is applicable only to small proteins that are soluble at high concentrations.

ii. Most protein function is expressed *via* protein *dynamics*. The active sites of enzymes are generally more flexible than other parts of the proteins.

The active parts of some proteins are disordered; *e.g.* calmodulin binds to a disordered region in calcineurin, a serine/threonine phosphatase.

---

Perhaps as much as 25% of encoded polypeptide is disordered.

To understand protein function we need high-resolution dynamic information on timescales ranging from femtoseconds to minutes.

All atoms in a protein undergo motion above absolute zero according to the Maxwell-Boltzmann distribution. What is the evidence for this?

**A. Hydrogen Exchange**  In the early 1950's Linderstrom-Lang dissolved proteins in $D_2O$.

Protein L
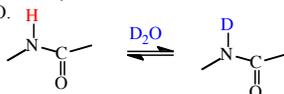


http://www.npaci.edu/online/v5.5/protein-folding.html
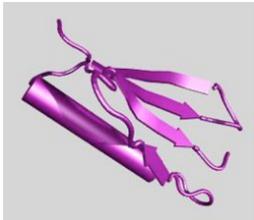
Amide hydrogen exchange times lasted from seconds to months.

---

Amides in unfolded proteins exchange in less than 1 second.



http://www.npaci.edu/online/v5.5/protein-folding.html
http://parasol-www.cs.tamu.edu/groups/amatogroup/research/folding/proteinL.php
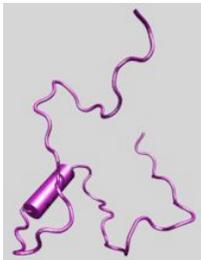
∴ the 3D conformation slows down H-exchange.

---

Because exchange requires contact between solvent and amides, conformational fluctuations are required for exchange to occur.

**HOW**?

1. Proteins are close-packed in their interiors.

*e.g.* On average, there is enough space for 1 methyl group.
So the conformation of the protein excludes solvent and slows exchange. This is called "*Solvent Exclusion*"

2. Even amides on the surface of a protein can be slow to exchange, especially if they are H-bonded.

**N-H |||||||||| O=C**

Usually, the H-bond must be broken in order for exchange to occur.

So hydrogen exchange rates measure "structural fluctuations".

**OTHER METHODS:**

**B  Steady-State Fluorescence**

In the 1970's Gregorio Weber demonstrated Tryptophan fluorescence quenching by $O_2$.

---

$O_2$ diffuses through a protein at 20-50% of its rate in water.

$$k_t = 1.2 \times 10^{10} \text{ M}^{-1} \cdot \text{sec}^{-1}$$

He concluded that a protein structure is dynamic.

This is because proteins are held together by a large # of weak bonds (van der Waals, H-bonds, electrostatic) and these are easily and reversibly disrupted by thermal energy.

∴ in solution proteins exist in a large # of "conformations".

Recently, **single-molecule fluorescence** has permitted the observation of dynamics in **individual** proteins.

---

*e.g.* the individual steps in the rotation of the g-subunit of the $F_oF_1$-ATP synthase were observed by single molecule **Fluroescence Resonance Energy Transfer.**
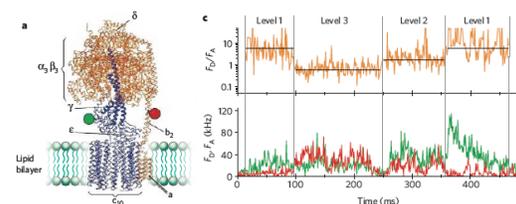
C. **Fluorescence Decay**

Measurements of the **excited state lifetimes** of Trp residues gives information about the dynamics of the side-chain.

D. **NMR Spectroscopy**:

$T_1$ and $T_2$ relaxation rate measurements are sensitive to dynamics on the nanosecond timescale: *e.g.* Tyr, Trp, Phe ring "flips".

$^{15}$N and $^{13}$C relaxation rates can be used to measure backbone and side-chain dynamics from the picosecond to the millisecond timescale. Relaxation dispersion experiments yield information on the microsecond to millisecond timescales

E. **X-ray Diffraction**:

Temperature **B-factors** can be used to obtain dynamics of atoms in crystals. The B-factor is the "*smearing out*" of the electron density. For small molecules, the value is about 3-4Å$^2$ but for proteins it is 10-20 Å$^2$ suggesting average vibrations of 0.3-0.5 Å and frequent deviations of greater than 1 Å. Some of the B-factor is due to lattice disorder. However, many protein structures are measured at liquid $N_2$ temperatures so some motions are damped out.

Neutron diffraction has measured H-exchange in protein crystals.

F. **Computational Methods**

Classical Molecular Dynamics Simulations.

Atomic Coordinates from an X-ray structure are used.

A hypothetical energy is calculated which takes into account:

*van der Waals forces, coulombic forces, dihedral angles, bond angles, bond lengths*

Each atom is given an initial velocity so that:

$$E_k = \frac{1}{2}\sum_i^N m_i v_i^2$$

Newton's equations of motion are solved over very small time steps $\Delta t$.

*i.e.* atomic motions are simulated $\Delta t \sim 2 \, fsec$

Using massively parallel supercomputers protein dynamics as long as $10^{-6}$ s have been simulated for a small protein. Stanford scientists are also using a distributed computing network and the "Folding-at-home" program to fold small proteins.

Quantum Mechanical Dynamics can be calculated only on small parts of proteins such as active sites.

**Dynamics Today**

The picture we have today is of proteins existing in a large number of very similar conformations that can interconvert on the picosecond to nanosecond timescales. These involve small changes such as methyl rotations, side-chain motions and small loop movements.

Rarely, proteins can undergo collective motions involving larger domains on the microsecond to second timescale. This is usually the time domain that expresses the "activity".
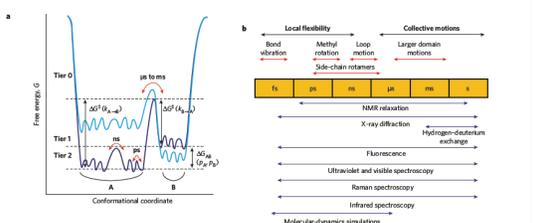


Two examples: In cyclophilin A (and calmodulin) all of the dynamic motions and all of the conformations of the active (bound) forms of the enzyme (protein) are present in the inactive (unbound) forms.

The idea is that the substrate (target) selects from among the pre-existing conformations in the enzyme (calmodulin) using the binding energy to shift the equilibrium. Recall that calmodulin binds hundreds of different proteins.

2. **HYDROGEN EXCHANGE**

The timescale is milliseconds to months. So in general it is sensitive to slower / rarer motions.

Those dynamics are relevant to the biological functions of proteins.

*e.g.* In Hemoglobin many amide exchange rates change when $O_2$ is released.

---

For small proteins (up to 300 amino acids) it is possible to measure amide exchange rates for *every* backbone amide using NMR spectroscopy.

This gives dynamic information in solution at high resolution.

To determine the contribution to exchange from protein dynamics we need to know all the other factors that can affect hydrogen exchange rates.

*Temperature, pressure, chemical induction, catalyst concentration, solvent viscosity, pH, and polarizability,* all determine the rates.

---

∴ in studying protein hydrogen exchange rates exchange of:

1. The protein in the denatured state,

2. poly-*d,l*-alanine,

3. small unstructured peptides,

are measured as controls.

We will study the rates of exchange of a primary amide, a secondary amide, and the indole NH in the side-chain of **N-Acetyl-Trp-amide** Make sure you know the structure of this molecule.

## Hydrogen Exchange: Kinetic Mechanism

$$XH + Y \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} XH\cdots Y \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X\cdots HY \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X + HY \rightleftharpoons$$
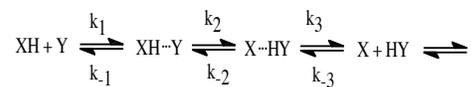
1. Diffusion-limited encounter of catalyst (Y) and the protein (XH) forms a complex that may involve a H-bond.

2. Proton transfer occurs within the complex.

Equilibrium ($K_2 = k_2/k_{-2}$) within the complex depends upon the relative pKa's of donor and acceptor (HX and HY).

3. Dissociation of the complex with the proton carried away by the acceptor.

4. Re-protonation of X by solvent follows a mechanism similar to that shown above.

For OH⁻ catalysis of exchange of an amide proton, the rate-limiting step is the transfer of the proton from XH to Y.

The chemical exchange rate ($k_{ce}$) can be shown to be the collision rate ($k_1[XH][Y]$) $x$ the fraction of successful collisions (f):

$$k_{ce} = k_1[XH][Y] \bullet f$$

Generally, the protein is in large excess over catalyst and its concentration does not change during the reaction. We can disregard the protein concentration by dividing the reaction rate by [XH]:

Now:
$$k_{HX} = \frac{k_{ce}}{[XH]} = k_1[Y] \bullet f$$

where $k_1$, formerly a second-order rate constant, is now a pseudo-first-order rate constant and

$$f = \frac{[X\cdots HY]}{[X\cdots HY]+[XH\cdots Y]} = \frac{\dfrac{[X\cdots HY]}{[XH\cdots Y]}}{\dfrac{[X\cdots HY]}{[XH\cdots Y]}+\dfrac{[XH\cdots Y]}{[XH\cdots Y]}}$$

But $K_2 = \dfrac{[X\cdots HY]}{[XH\cdots Y]} = \dfrac{k_2}{k_{-2}}$   Therefore $f = \dfrac{K_2}{K_2+1}$

Measuring *[XH··Y]* and *[X··HY]* and $K_2$ is difficult. Measuring the overall $K_{eq}$ is easy.

First, we write the $K_{eq}$ for the reaction in terms of the relative acidities ($K_a$) of *X* and *Y*.

$$K_{eq} = K_1 \cdot K_2 \cdot K_3 = \frac{k_1}{k_{-1}} \cdot \frac{k_2}{k_{-2}} \cdot \frac{k_3}{k_{-3}} = \frac{[X][HY]}{[XH][Y]}$$

For the reactions:

$$XH \rightleftharpoons X + H \qquad\qquad HY \rightleftharpoons Y + H$$

$$K_a^X = \frac{[H][X]}{[HX]} \qquad\qquad K_a^Y = \frac{[H][Y]}{[HY]}$$

---

And 
$$\frac{K_a^X}{K_a^Y} = \frac{[X][HY]}{[XH][Y]} = K_{eq}$$

Thus, the overall equilibrium constant depends on the relative affinities of X and Y for protons.

Next, consider the reaction mechanism:

$$XH + Y \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} XH{\cdots}Y \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X{\cdots}HY \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X + HY \rightleftharpoons$$
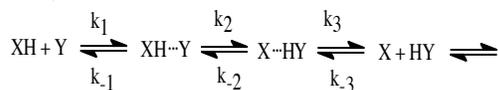
$k_1$ and $k_{-3}$ will both have a magnitude of about $10^{10} M^{-1} \cdot s^{-1}$ as both are diffusion-limited rate constants. Furthermore, the diffusion-limited separation rate constants $k_{-1}$ and $k_3$ are both expected to be about $10^{10}\ s^{-1}$.

---

Thus, when $k_1 = k_{-3}$ and $k_{-1} = k_3$, then $K_{eq} = K_2$.

$$k_{HX} = k_1[Y] \cdot f = \frac{k_1[Y]K_{eq}}{1 + K_{eq}} = \frac{k_1[Y]\dfrac{K_a^X}{K_a^Y}}{1 + \dfrac{K_a^X}{K_a^Y}}$$

From the definition of $pK_a = -log[K_a]$

$$\log\left[\frac{K_a^X}{K_a^Y}\right] = \log[K_a^X] - \log[K_a^Y] = -pK_a^X + pK_a^Y = \Delta pK_a^{Y-X}$$

Note that $\Delta pK_{Y-X}$ is defined as proton Acceptor (*Y*) - proton Donor (*X*).

---

Thus, $\log K_{eq} = \Delta pK_a^{Y-X}$ and $K_{eq} = 10^{\Delta pK_a^{Y-X}}$

Finally,

$$k_{HX} = \frac{k_1[Y] \bullet 10^{\Delta pK_a^{Y-X}}}{1 + 10^{\Delta pK_a^{Y-X}}}$$

**Base-Catalysed Exchange:**

Consider catalysis by $OD^-$.

e.g. $HY = D_2O$    $pK_a = 15$;    $HX = TyrOH$   $pK_a = 9.8$

$\Delta pK_{Y-X} = 15 - 9.8 = 5.2$

at $pD = 7.5$    $[OD^-] = 10^{-7.5}$ M

---

$k_{HX} = 10^{10} \cdot 10^{-7.5} = 10^{2.5}$ $sec^{-1}$   so exchange is very fast

For CONH   $pK = 18.5$

$\Delta pK_{Y-X} = 15 - 18.5 = -3.5$

so   at $pD = 7.5$    $[OD^-] = 10^{-7.5}$ M

$k_{HX} = 10^{10} \cdot 10^{-7.5} \cdot 10^{-3.5} \sim 0.1$ $sec^{-1}$

and exchange is much slower.

In proteins:  Asp, Glu, C-terminus, N-terminus, His, Tyr, Lys, Arg, Ser, Thr, Trp, Gln, Asn, backbone CONH all contain exchangeable hydrogens.

---

The slowest, and easiest to measure, are the backbone CONH, Gln, Asn, and Trp
since $\Delta pK^{Y-X}$ is < 0.

**Acid Catalysis**:   Hydrogen Exchange can also be acid catalysed by the N- or O-protonation mechanisms.
Imidic Acid Mechanism:  O-protonation

It is not straightforward to write the kinetic equation for this mechanism.

N-protonation: $\quad D_3O^+ \rightleftharpoons D^+ + D_2O$

$K_a^X = \dfrac{[D^+][D_2O]}{[D_3O^+]}$ and $\quad pK_a^X \cong -1.7$

$$ND_2^+ \rightleftharpoons ND + D^+$$

$K_a^Y = \dfrac{[D^+][ND]}{[ND_2^+]}$ and $\quad pK_a^Y \cong -12.7$

$\therefore \Delta pK_a^{Y-X} \cong -12.7 + 1.7 = -11$

---

$\therefore \quad [D_3O^+]$ must be very high for acid-catalysed exchange to occur.

Exchange is the sum of reactions catalysed by acid, base, and water. Water catalysis is negligible.

$$k_{ex} = \sum k_{HX}[cat]_i = k_{OH}[OH^-] + k_H[H_3O^+]$$

---
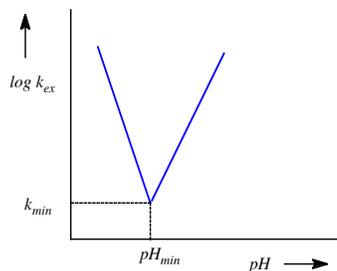


For proteins $pH_{min}$ is about 3. Why?

$$k_{HX} = \frac{k_1[Y] \cdot 10^{\Delta pK_a^{Y-X}}}{1 + 10^{\Delta pK_a^{Y-X}}}$$

At pH 3     $k_{OH} = 10^{10} \cdot 10^{-11} \cdot 10^{-3} = 10^{-4} s^{-1}$

$k_H = 10^{10} \cdot 10^{-3} \cdot 10^{-11} = 10^{-4} s^{-1}$

$pH_{min}$ of proteins depends on the amino acid sequence.

---

Electron-withdrawing side-chains stabilize negatively charged transition states (base catalysis) and lower the $pH_{min}$.



$log\ k_{ex}$

$pH_{min}$          $pH$

---
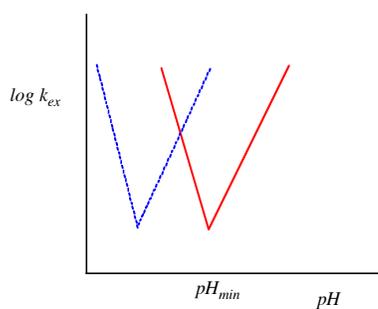
Electron-donating side-chains stabilize positively charged transition states (acid catalysis) and elevate the $pH_{min}$.



$log\ k_{ex}$

$pH$
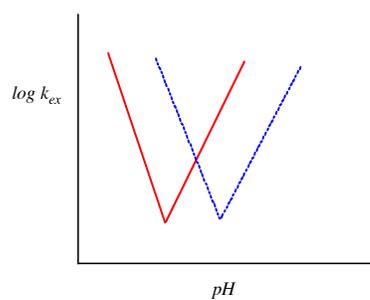
3. **EXPERIMENTAL**:

How to measure hydrogen exchange.

1. a. Peptide bond ultraviolet light absorption changes (200-220 nm).

b. Appearance of deuterium in protein crystals by neutron diffraction.

c. NH signal disappears in a $^1$H NMR spectrum.

d. Mass spectrometry detects the difference in mass between H and D.

These are slow methods *i.e.* $t_{1/2}$ = 5 *min*.

e. **Dynamic NMR spectroscopy** can measure much faster exchange rates.

**Dynamic NMR Spectroscopy**

NMR is the measurement of the magnetization of a sample due to the alignment of protons with a magnetic field.

Outside the magnet $M_{NH} = 0$.

In the magnet, equilibrium magnetization $= \bar{M}_{NH}$.

$T_{INH}$ = time constant for approach to $\bar{M}_{NH}$
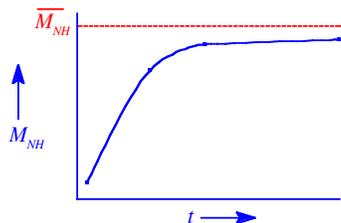
Rate constant = $R_{NH} = 1/T_{INH}$

12

$$\frac{dM_{NH}}{dt} = R_{NH}(\bar{M}_{NH} - M_{NH}) = \frac{\bar{M}_{NH} - M_{NH}}{T_{1NH}}$$

The integral of a first-order differential equation is an exponential:

$$\frac{M_{NH}}{\bar{M}_{NH}} = 1 - e^{-t/T_{1NH}}$$

At steady-state: $\frac{dM_{NH}}{dt} = 0$ and $\bar{M}_{NH} = M_{NH}$

When H-exchange is occurring: $NH \underset{k_r}{\overset{k_f}{\rightleftharpoons}} HOH$

---

$$\frac{dM_{NH}}{dt} = \frac{\bar{M}_{NH} - M_{NH}}{T_{1NH}} - k_f M_{NH} + k_r M_W$$

$M_w$ represents the water signal. Using a technique of NMR spectroscopy called "**saturation**" we can eliminate $M_w$.

Then at Steady-state: $\frac{\bar{M}_{NH} - M_{NH}}{T_{1NH}} = k_f M_{NH}$ and

$$M_{NH} = \frac{\bar{M}_{NH}}{1 + k_f T_{1NH}}$$

---

That is, the amide intensity will depend on the hydrogen exchange rate:

1. $k_f << 1/T_{1NH}$     $M_{NH} = \bar{M}_{NH}$

2. $k_f >> 1/T_{1NH}$     $M_{NH} = 0$

3. $k_f = 1/T_{1NH}$     $M_{NH} = \bar{M}_{NH}/2$

$$\frac{M_{NH}}{\bar{M}_{NH}} = \frac{1}{1 + k_f T_{1NH}} = F = \text{the fractional amide intensity.}$$

$k_f$ is what we called $k_{HX}$ earlier. We can manipulate $k_{HX}$ by changing the *pH*.

Remember:

$pH = -log[H_3O^+]$ and $[H_3O^+] = 10^{-pH}$

$pK_w = pH + pOH$

$[OH^-] = 10^{-pOH} = 10^{pH-pKw}$

$k_{ex} = k_{OH}[OH^-] + k_H[H_3O^+] = k_{OH}10^{pH-pK_w} + k_H 10^{-pH}$

And

$$F = \frac{1}{1 + (k_{OH}10^{pH-pK_w} + k_H 10^{-pH})T_{1NH}}$$

___

You will fit the amide intensities that you measure to this equation and obtain the acid and base-catalysed rate constants.

For you to solve:        What is the $pHmin$?

at the $pH_{min}$  →      $k_{OH} \bullet 10^{pH-pK_w} = k_H \bullet 10^{-pH}$

What is the rate of exchange at the $pH_{min}$?

$k_{min} = ?$

14

# Hydrogen Exchange Rates Measured by Saturation Transfer NMR Spectroscopy

Amide hydrogen exchange rates are catalysed by $H^+$ and $OH^-$ and depend on other factors such as temperature, proximity to charged groups, and through-bond induction. In proteins exchange rates also reflect dynamic fluctuations in secondary and tertiary structure. In this experiment you will measure the acid- and/or base-catalysed exchange rate constants and calculate the pH of minimum exchange for the amides in an amino-acid derivative. Since the molecule will be 'unstructured' the exchange rates will reflect the chemical environment of the amide group and not fluctuations in structure. Exchange measurements on unstructured peptides are important for the correct interpretation of exchange rates measured in proteins.

Exchange rates will be measured by pulsed, Fourier transform $^1H$ NMR spectroscopy. The molecule will be dissolved in a solution consisting of 80-90% $H_2O$/ 10-20% $D_2O$. The strong proton signal from the solvent will be eliminated by selectively irradiating or 'saturating' the water signal. Because the concentration of water is much larger than that of the amino acid any amide protons which are in fast exchange with water, on the NMR timescale, will end up spending most of their time as water protons. The water signal is eliminated by saturation so amide protons which are in fast exchange exhibit zero intensity in the NMR spectrum. Amides which are in slow exchange on the NMR timescale will spend all their time as amide protons, will be unaffected by saturation of the water, and therefore will be at full intensity in the spectrum. Amide protons in intermediate exchange will exhibit fractional intensities in the NMR spectra and these can be used to calculate the rate of exchange. Since exchange is catalysed by $H^+$ and $OH^-$ a pH titration of the amide intensity will yield acid- and base-catalysed exchange rate constants as well as the pH of minimum exchange for the amides.

------------------

Wet chemistry: 392 Parker.
NMR experiments: 338 Parker.
Data analysis on Macintosh computer using MATHEMATICA software: 390 Parker.

Buffer composition: 10-20% $D_2O$/80-90% $H_2O$; 5 mM Phosphate; 5 mM Acetate-$_{d4}$; 5 mM DSS; pH approx. 2.

Titrants: 1.1 N NaOH in 10-20% $D_2O$/80-90% $H_2O$;  0.22 N NaOH in 20% $D_2O$/80% $H_2O$;  1.1 N HCl in 20% $D_2O$/80% $H_2O$.

1. Calibrate the pH meter using standard solutions.

2. Weigh out approx. 1 mg of amino acid and place it in a 1.5 ml eppendorf tube.  Add 1.0 ml of buffer, mix thoroughly.

3. If any particulate matter is observed in the solution centrifuge it for 2 min.  Measure and record the pH.  Then add 0.55 ml of amino acid solution to an NMR tube, cap the tube, go to room 338 and, with the help of the technician, record an NMR spectrum.

4. Return to Room 392 and carefully remove the amino acid solution from the NMR tube using an extended Pasteur pipette, returning the solution to the eppendorf tube.  Measure and record the pH.

5. You will now raise the pH of your sample in steps of about 0.5 pH units and acquire NMR spectra after of each change.  Four to six additions of 3-5 microliters of 0.16 N NaOH (20% $D_2O$/80% $H_2O$) will bring the pH to about 4.5.  Take NMR spectra after each increment of the pH. Continue to make similar additions until the amide intensity is eliminated. Measure the pH before and after each spectrum is acquired.

6. With the help of the demonstrator plot all your spectra including one of the entire chemical shift range from -1-10 ppm,  and measure the peak heights of the amide resonance(s).  Tabulate the normalized peak heights vs. pH.  Then use the non-linear least-squares fitting routine in the program MATHEMATICA to determine the acid- and base-catalysed exchange rate constants as well as the pH of minimum exchange for your peptide, and to plot your results.  With the use of the $k_H$ and $k_{OH}$ values convert your amide intensities to exchange rates ($k_{ex}$) at each pH value measured and then plot pH vs. log[$k_{ex}$].

7. In your report discuss the significance of the $pH_{min}$ with respect to the chemical structure of your molecule.  Show how $pH_{min}$ can be calculated from the $k_H$ and $k_{OH}$  values.  Do the same for $k_{min}$.  List the $pH_{min}$ and $k_{min}$ values for each exchanging amide.  What is the significance of the chi-squared values?  Also explain the composition of the buffer solution.  Obtain a complete ^{1}H NMR spectrum from the demonstrator and assign all the resonance peaks.

## Advanced Biochemistry Laboratory CHEM 4700

## X-ray Diffraction

During the first week of this laboratory you will prepare crystallization trials of hen egg white lysozyme.  Instructions will be given in the lab and an outline can be found on the next page.  Following you will hear the first part of a workshp describing the determination of the structure of a biological macromolecule.

In the second laboratory we will examine the crystals produced over the week and then finish the workshop.

You will hand in a lab report in the usual style.  At the end of this laboratory you should be familiar with the following concepts, that you may include descriptions of in your lab report.

Diffraction
Crystal lattice
Reciprocal lattice
Unit cell
Isomorphous replacement
Temperature factor (B)
R-factor

# Lab Experiment: Crystallization of Hen Egg White Lysozyme by Two Different Methods

*By Terese Bergfors in Protein Crystallization:  Techniques, Strategies, and Tips.*

Lysozyme is cheap and easy to crystallize. Many of the experiments in this book will be using lysozyme as the model protein. Below you will find two recipes for your first crystallization efforts.

**Experiment 2.1.** Vapor diffusion hanging drop technique:

In the following experiment hen egg white lysozyme is crystallized using the hanging drop technique. The effect of different salt concentrations will be tested here. Other parameters such as pH, protein homogeneity, etc., are equally important.

**Materials:**

1. Q plate I or II
2. siliconrized cover slips, 18 mm
3. sealing tape
4. lysozyme: chicken egg white, SERVA Fein Biochemica GmbH & Co KG, cat. no. 28262

**Prepare the following solutions:**

1. Lysozyme:
Make 100 μl each of 15 mg/ml and 20 mg/ml lysozyme in 20 mM Na-acetate, pH 4.2.

2. Precipitant solutions:
A: NaCl, 5% (w/v) in 0.1 M Na-acetate, pH 4.2
B: NaCl, 6% (w/v) in 0.1 M Na-acetate, pH 4.2
C: NaC1, 7% (w/v) in 0.1 M Na-acetate, pH 4.2

**Experiment:**

1.  With a pipette place 500 μl of precipitant solution A in the first reservoir. Carefully place one drop of 5 μl lysozyme ( 15 mg/ml) on a cover slip. Try to make the droplet as spherical as possible. Add 5 μl of the reservoir solution to the lysozyme droplet. It is important that the precipitant is added to the protein solution and not in the reverse order. Exposure to high precipitant concentration can result in precipitation of the protein, which might destroy the experiment. Invert the cover slip and put it in place on the ledges in the Q plate. Prepare another two experiments with the same protein concentration and the precipitant solutions B and C.

2. Similarly prepare 3 new reservoirs with the solutions A, B, and C, and place 5 μl droplets of the 20 mg/ml lysozyme solution on the cover slips. Now seal the Q Plate with clear sealing tape.

   Crystals will appear after 16 hours in the experiment with the highest salt and protein concentration. The quality of the crystals is highly dependent on the purity of the lysozyme. Purification of the lysozyme on a size exclusion column and concentration of the material with ammonium sulfate precipitation results in well-formed crystals with no packing disturbances. Even if you do not further purify the lysozyme, it should be filtered through a 0.22 μm filter immediately before setting up the crystallization trials.

**Observation Sheet**        S - Small        M - Medium        L - Large

**Concentration of Lysozyme = 15 mg/ml**

| Name | Salt Concetration % | pH = 4.0 | pH = 4.2 | pH = 4.4 |
|---|---|---|---|---|
| Andriy | 5 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |
| William | 6 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |
| Andrea | 7 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |

**Concentration of Lysozyme = 20 mg/ml**

| Name | Salt Concetration % | pH = 4.0 | pH = 4.2 | pH = 4.4 |
|---|---|---|---|---|
| Stephanie | 5 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |
| Melissa | 6 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |
| Taras | 7 | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: | # of crystal ___<br>Size ___<br>Shape: |

# A Study in Hen Egg White Lysozyme Crystal Growth

Christopher J. Hodgson '99, Sean N. Maduck '00, Dean S. Rahman '01

**Abstract**

This study examined the effects of pH and buffer type on the crystal growth of hen egg white lysozyme (HEWL). Effects were measured by monitoring both the number of crystals grown and the axial ratios of the crystals as functions of pH. Previous studies on the same topic do not explore buffer effects, thus this variation is included here. Acetic acid yielded both decreased axial ratios and fewer crystals with increasing pH. Formic acid under the same conditions showed no change in axial ratio, but did show an overall decrease in the number of crystals. Additionally, X-ray diffraction analysis was performed in an attempt to obtain a rudimentary idea of the atomic structure of the crystals.

## Introduction

Proteins are an essential part of every biological process and are the basis of life itself. The study of protein crystallization—which has been carried out for over 150 years (Durbin & Feher, 1996)—has many important applications, including understanding biological processes and creating new drugs for the treatment of illnesses and injuries. The types of studies done on protein crystals range from the atomic level, where the interest is how changes in the protein structure affect its behavior, to a larger scale of whole crystals and the effects of different environmental conditions on their nature. The most popular protein used for investigation of general crystalline properties is the tetragonal form of HEWL. For this reason, it is the protein used here.

HEWL has a tetragonal Bravais lattice. This is a rectangular-type structure with two of its three dimensions equal to each other and both different from the third. Figure 1 shows a structural schematic:



Figure 1

$a = b \neq c$
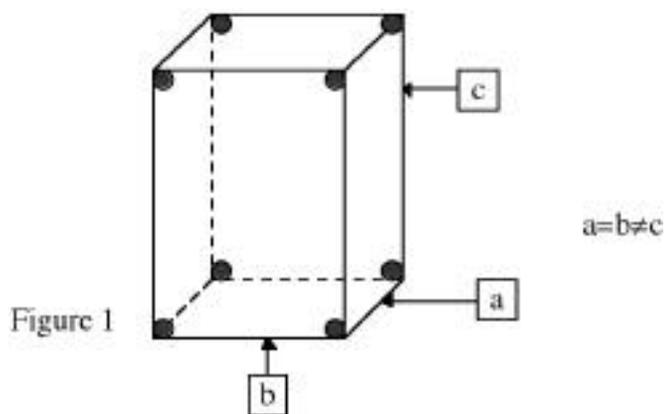
In the case of HEWL, the 'c' dimension is one lattice constant and is equal to 78 Å. The 'a' and 'b' dimensions are the other lattice constants and are equal to 39 Å.

This analysis focuses on the effects of varying the conditions under which the crystals are grown. Changes in pH and buffer solutions used in crystal growth will be examined for their effects on both numbers of crystals grown and the shape of the crystals. We took this approach because "few studies . . . have focused on measuring the effect of solution conditions on nucleation rates by specifically counting and sizing crystals" (Judge, *et al.*, in press). Protein concentration is also a variable which could greatly affect crystal growth, but for this project, this effect will not be examined in detail.

pH is of particular interest due to its importance to the crystallization process itself. Crystal structure is determined by the interaction of amino acids to form protein molecules (HEWL is comprised of 129 amino acids) and then the interaction of these molecules with each other. All of this is dependent upon van der Waals forces and hydrogen bonds. pH is important because it determines the net surface charge carried by the polar groups in the molecules (Rosenberger, 1996), and thus the bonding strength.

A study entitled "The Effect of Temperature and Solution pH on Tetragonal Lysozyme Nucleation Kinetics" by Russell A. Judge and colleagues is similar to this project. Judge and colleagues found that changing the levels of pH and supersaturation had large effects on the number of crystals formed, all other factors being held constant. The measured effects of temperature, however, were much less significant. One of the more prominent effects measured came as a result of changing the pH from 4.6 to 4.8. This range showed a drop in the number of crystals by a factor of approximatly 20. Additionally, a change in pH from 4.0 to 4.6 reduced the number of crystals by a factor of 1.5, and from 4.8 to 5.2, resulted in a reduction by a factor of 2. As for the crystal shape, there was a decreasing trend in the axial ratios—defined as the ratio of crystal length in the 'c' direction to the length in the 'a' direction (c:a)—with the lowest axial ratios being found at a pH of 5.2.

This study concluded that, "Of the crystallization parameters supersaturation, temperature and pH, it is pH that has the strongest effect on crystal number and size, while for the axial ratio both pH and supersaturation exhibit strong influences" (Judge, *et al.*, in press).

X-ray diffraction is another analytical method which is sometimes used as a means of examining the atomic structure of the protein crystals. The relevant parameter is the angle of diffraction, q, which can be determined from the Bragg equation:

$$n l = 2d \sin (q)$$

where l is the wavelength of the source's x-rays and d is the lattice parameter (distance between adjacent planes) which is determined according to the lattice constants, the structure of the specific crystal, and n = 1. A prediction of the peak angles in x-ray diffraction can be made given the lattice parameters of the crystal.

## Experimental Equipment and Procedure

Dialysis

The first step in the protein preparation for the initial experiment, measuring the effect of the buffer's pH on crystal growth, was dialysis. 1.258 ± 0.005 grams of commercially made, dehydrated lysozyme (which was 3x crystallized, dialyzed and lyophilized—containing approximately 95% protein) was dissolved in 50 mL of distilled water. To purify the protein, the solution was placed in a semi permeable membrane with a molecular weight cut-off of 6000 to 8000 g/mol. This membrane was sealed and left in a tank of distilled water overnight. By doing so, impurities were removed from the sack via osmosis. The amount of protein inside remained constant since the membrane was not permeable to the protein molecules. Part of the solution was run through filter paper after dialysis (this portion of the solution is referred to as prefiltered and the rest as non-prefiltered).

Concentration

The solution was then filtered through an Amicon® Stirred Ultrafiltration Cell which was subject to high pressure as supplied by compressed helium. The protein, which was above the membrane's molecular weight cut-off, was retained in the cell while excess water and dissolved impurities below the cut-off were forced through the filter and out of the cell.

buffer in the first trial. In the second trial, formic acid (0.1 M), also with 5% NaCl, was used as the buffer. The NaCl served as a precipitating agent in crystal growth. The desired pH range for each buffer is near the acid's pKwo ranges were made by titrating the respective acids with concentrated NaOH, increasing the pH in small increments. The buffer solutions obtained for acetic acid were of pH: 4.08, 4.20, 4.41, 4.60, 4.82, 5.00, 5.23, 5.42, 5.67, 5.92, and 9.81. For formic acid, the pHs were: 2.98, 3.21, 3.39, 3.62, 3.81, 4.01, 4.19, 4.48, 4.62, 4.89, 5.29, and 6.25. Accuracy for pH values (based upon experience with the pH meter) is ± 0.10.

Plate Preparation

For growing the protein crystals, special plates such as the one shown in Figure 2 were used. Prior to preparation, the plates were cleaned with distilled water and dried with Dust Away®, an aerosol drying product. 1 mL of buffer solution was placed in each outer well. The four wells in each column (columns are labeled 1-6) contained the same pH level. In each of the inner wells, 20 μLof protein solution and 20 μLof buffer solution were deposited. Once this plate preparation was complete, the plates were covered with clear tape to prevent dehydration and were placed in a refrigerator at 6° C for growth. The time between the
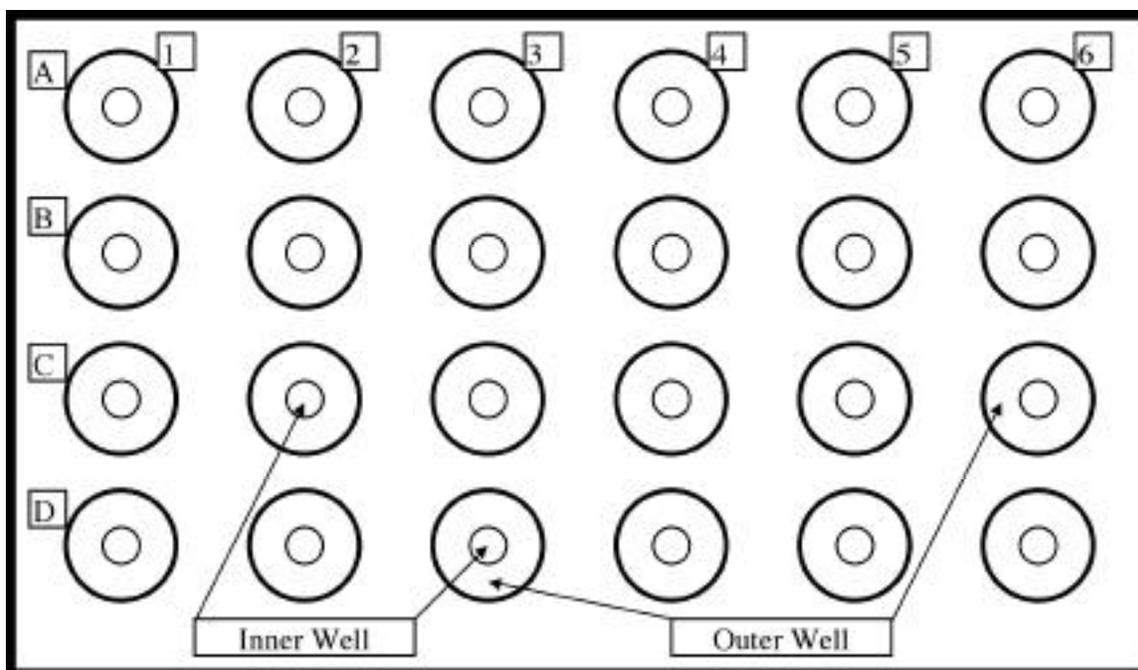


Figure 2: Crystal Tray Schematic

Measurement of Concentration

Measurement of the concentration was performed using an ultraviolet spectrometer: $l = 285$ nm and $c = 1$ cm. The initial concentration was measured at 35 mg/mL and then diluted to 25 mg/mL for use in the experiment.

The Buffer

Acetic acid (0.1 M) with 5% NaCl by weight was used as the

beginning of growth and examination of the crystals was no less than 3 days. The length of the interval did not have any noticable effect.

Investigation of Crystals

Once the crystals formed, a high magnification Edmund Scientific microscope was used to photograph the crystals at magnifications between 1000x and 3000x. The microscope was connected to a computer on which the program *NIH Image* was used

to take the photographs. From the photographs, the following were determined:
  • length to width ratios, or axial ratios
  • an estimate of the number of crystals grown per well

X-Ray Diffraction

X-ray diffraction was carried out on the protein crystals with a CuK source (l = 1.54 Å, 200 planes) using a Simens Kristalloflex Diffraktometer d 5000. Crystal samples were placed on slides and covered with Saran wrap to prevent dehydration (as an initial test, salt crystals were tested under Saran wrap to make sure that the wrap did not affect the testing process).

Sources of Error

The numbers given are intended to show only relations between the respective pH levels, and the amount of crystal grown at each level. For this goal number of crystals counted is only accurate to within ± 10%. In determining the axial ratios, the length and width of the crystals are on the order of 100 μm to 400 μm, with an accuracy of ± 10 μm in each measurement.

Results and Discussion

pH Variation Effects on Number of Crystals

In the first attempt to grow protein crystals (using acetic acid as the buffer solution), pH was varied from well to well in order to monitor possible effects. Other papers written on this subject indicated that the ideal growing condition for this protein is a pH between 4.0 and 6.0. For this experiment, pH was increased along this indicated range by increments of 0.2. Four wells were used for each pH, and the most clearly defined crystals from each level were selected for analysis.

To analyze the crystal growth in each well, a microscope with a magnification of 1500x was connected to a computer, and the program *NIH Image* was used to capture still images of the crystals. Because of the large number of crystals in each cell, only a 4 in$^2$ section of each photograph was used. The total area of the well surface relative to the size of the 4 in$^2$ viewing area was then determined and used to find the total number of crystals. This is only an approximation of the total number of crystals, but because the comparison between the different pH levels is more important than the actual number of crystals formed at each pH, the estimation is sufficient. Figure 3 shows the results of crystal growth in acetic acid.

Analysis of the data shows a wide range of growing results. A pH of 4.08 produced the greatest number of crystals, however, they were smaller in size and nonuniform in structure. Some had the expected hexagonal shape, but most looked fractured. The "cleanest" crystals with uniform structure appeared in the wells where there was the least amount of growth. In a pH range from 4.8 to 5.2, the best tetragonal crystals appeared. At a pH of 9.81 (not shown in Figure 3) a small number of crystals grew, but they had a random structure and were not uniform. This high pH was used as a test to show that using a pH from 4 to 6 is ideal.

For the next experiment, the same procedure was followed using formic acid rather than acetic acid. Similar growth results were expected and were observed. This time, however, the experiment began at pH 3 and was increased by increments of approximatly 0.2 through the ideal pH 6.25 for the growth range of the protein. The same analysis was performed. Figure 4 shows the
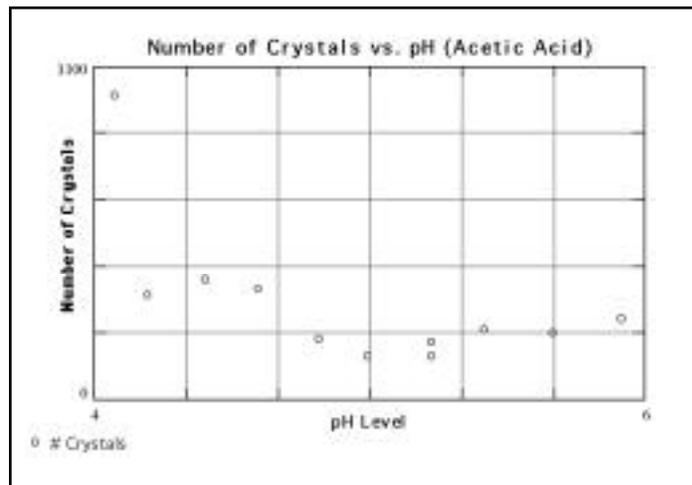


Figure 3: Crystal Growth in Acetic Acid
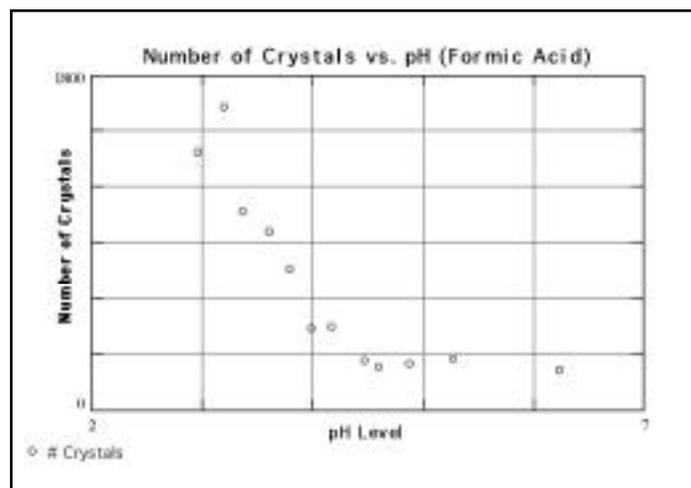


Figure 4: Crystal Growth in Formic Acid

crystal growth results for Trial 2:

Crystal growth was also very successful in the formic acid. The crystal growth followed the established pattern of Trial 1. Between a pH of 3 and 4, there were a large number of crystals; however, they were small in size and not well developed. As the pH climbed over 4, there were fewer and fewer crystals, and they were much more defined in the hexagonal shape expected. The optimal crystal growth was observed between a pH of 4.48 and 4.62. There were minimal numbers of crystals, and they were clean, well separated, and defined. This was very close to the results from Trial 1. At a pH of 4.89 and higher, the crystal development started to deteriorate. The crystals appeared blurred and fuzzy, and they definitely did not have the sharp hexagonal shape observed in the optimal growth range of pH 4.4 to 4.6.

Both crystal growth trials using acetic acid and formic acid were very successful. They supported each others' findings, and gave proof that the optimal pH for crystal growth was somewhere within a pH of 4 to 6. They also showed a dramatic drop in the number of crystals at a pH of 4.6 to 4.8, which was expected (in accordance with Judge, *et al.*). Combining the results of the two trials, it can be hypothesized that in order to get the best crystal growth, a pH in the range of 4.4 and 5.2 should be used (again, this agrees with the findings of Judge, *et al.*). This will produce the fewest number of crystals, and they will all have the desired hexagonal shape.

pH Variation Effects on Shape

A sketch of the general shape of one face of the crystals (side view):
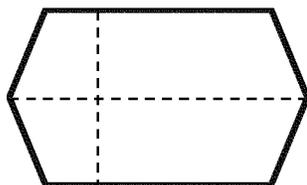


The measurement of interest is the ratio of the length of the horizontal axis/dotted line (defined as 'c') to the length of the vertical axis/dotted line (defined as 'a'). The change in this axial ratio was measured as a function of changing pH level. Figure 5 shows a graph of the results from the first trial in which acetic acid was used for the buffer solution. Here the axial ratio is graphed as a function of pH. Included in the graph are the standard deviations of the axial ratios. For each pH level, multiple crystals were examined (anywhere between 4 and 16 crystals per pH level), with varying axial ratios. The standard deviations serve to give an idea of the variation in axial ratio for each individual pH that was examined.

The pH range of the acetic acid buffer examined was 4.08 to 5.92 (9.81 did not give crystals suitable for examination). Above 4.60, the axial ratio (c:a) showed an inverse linear relationship with pH. At pH 5.42, that ratio was approximately 1.0. Above that pH, the ratio was below 1.0, meaning that the crystals were longer along the vertical axis than they were along the horizontal axis. In the pH range of 4.08 to 4.60 the results were roughly linear, but here the axial ratio did not change with pH—it remained approximately constant.

These data follow the results found by Judge *et al.*, in that they show a decreasing trend in the axial ratio as a function of pH. That study did not indicate that they found a flat region at the low pH, but the overall trend is in good agreement with their findings. Additional runs of this same experiment might reveal that the flat region at the low pH was an anomaly, and that in this region, axial ratio does decrease with increasing pH.

The data collected for crystallization using a formic acid buffer were analyzed in the same manner. Figure 6 shows the results of this trial, with axial ratio graphed as a function of the pH level of the buffer solution. In this case, the pH level ranges from 2.98 up to 4.62. The three higher pH values were not used because they did not form crystals that were suitable for determining an axial ratio. The results of this trial give an interesting contrast to the previous trial. This time there was no significant change in axial ratio with changing pH. As can be seen from the plot of the data, the axial ratio remained near 1.5 for all pH levels. This shows that crystal growth can be noticeably affected by the buffer solution used in the growth process. Standard deviations and a best-fit line are shown as before (this time, anywhere from 15 to 33 crystals per pH level were used).

These data do not follow the results found by Judge and colleagues like those from the acetic acid trial. In these data, there is no decreasing trend of axial ratio with pH. In fact, the axial ratio remains almost constant over the range examined (pH approxi-
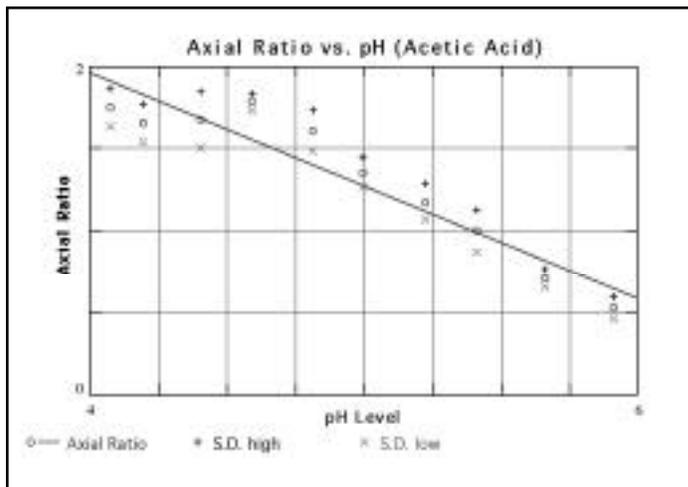


Figure 5: Crystal Shape in Acetic acid

mately 3 to 4.6). This seems to be an indication that the acid used for the buffer solution in growth can have a profound effect on the outcome. The crystals are similar to those grown with acetic acid; however, it seems that there is not nearly the same influence of pH on axial ratio.

Another interesting result on which the formic acid trial sheds more light, is the point at which the crystals become unsuitable (described earlier as random) for axial ratio comparisons. In the first trial, this happened somewhere between pH 5.92 and 9.81. But, in the second trial, this occurred between pH 4.62 and 4.89. From the acetic acid trial, there is no way of knowing if this change occurs rapidly or gradually. However, the formic acid trial seemed to indicate that the change occurs fairly rapidly. At pH 4.62, there were no signs of a change in growth, and at pH 4.89, it was clear that the change had taken place.

For visualization purposes, the crystals with a high axial ratio can be thought of as having a "sausage-like" shape, and the crystals with a low axial ratio, a "pancake-like" shape, roughly as follows:

SAUSAGE:                    PANCAKE:



X-Ray Diffraction

X-ray diffraction was performed in an attempt to analyze the atomic structure of the HEWL crystals. However, the tests were inconclusive and provided no solid results for presentation.

Conclusions

In using acetic acid as a buffer, changes in pH (between 4.0 and 6.0) were found to have a considerable effect on crystal growth, in terms of both number of crystals and axial ratio. The axial ratios show a definite decreasing trend with increasing pH. For the pH range of this specific analysis, this trend was linear. While they are not as convincing as the axial ratio data, the num-

ber of crystals formed at each pH showed an overall decrease over the pH range. For a formic acid buffer in a pH range of approximately 3.0 to 5.0, there were no significant changes in axial ratio as a function of pH; however, the number of crystals showed similar results to those for the acetic acid buffer, decreasing with increasing pH.

These findings, with the exception of axial ratios for the formic acid buffer, are in agreement with the findings of Judge, *et al*. (both axial ratio and number of crystals decrease with increasing pH). However, the study done by Judge, *et al*. did not examine the effects of changing the buffer solution. This study did, and it found a pronounced difference in the effect on axial ratio between acetic and formic acid buffers, leading to the conclusion that changes in the buffer solution do indeed have an effect on crystal growth.
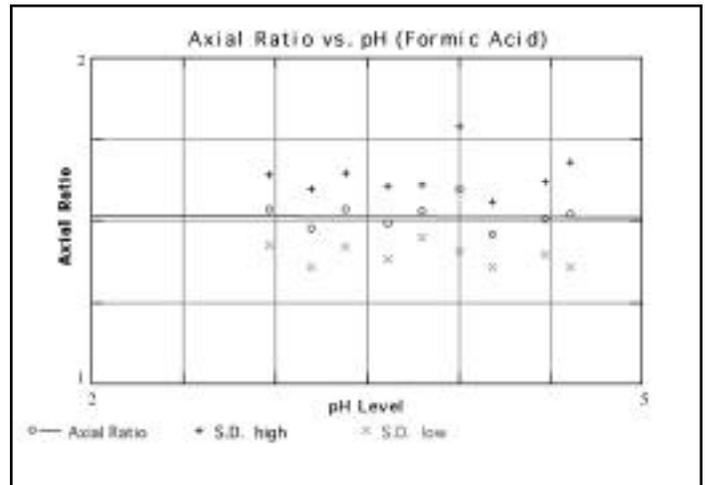


Figure 6: Crystal Shape in Formic Acid

### References

Cullity, B. D. (1978). *Elements of X-Ray Diffraction*. Reading: Addison-Wesley Publishing Co., Inc.

Durbin, S.D. & Feher, G. (1996). Protein Crystallization. *Annual Review of Physical Chemistry,* 47:171-204.

Judge, R. A., Jacobs, R. S., Frazier, T., Snell, E. H., & Pusey, M. L. (in press). The Effect of Temperature and Solution pH on Tetragonal Lysozyme Nucleation Kinetics.

Nadarajah, A. & Pusey, M. L. (in press). Growth Mechanism and Morphology of Tetragonal Lysozyme Crystals. *Acta Crystallographica D.*

Rosenberger, F. (1996). Protein Crystallization. *Journal of Crystal Growth,* 166:40-54.

Schaffer, J. P., Saxena, A., Antolovich, S, Sanders, T. H. Jr., & Warner, S. B. (1999). *The Science and Design of Engineering Materials*. New York: McGraw-Hill.

Schall, C. A., Arnold, E., Wiencek, J. M. (1996). Enthalpy of Crystallization of Hen Egg White Lysozyme. *Journal of Crystal Growth,*

### Acknowledgments

### ABOUT THE AUTHORS

Christopher J. Hodgson '99 is an engineering sciences major from Grand Rapids, MN. His career interests include bridge design and transportation engineering.

Sean N. Maduck '00 is majoring in engineering sciences (biomedical engineering). He is from Calgary, Alberta, Canada.

Dean S. Rahman '01 is an engineering major from Bangledesh.