| REPORT DOCUMENTATION PAGE | Form Approved OMB NO. 0704-0188 |
|---|---|

| | | |
|---|---|---|
| 1. REPORT DATE (DD-MM-YYYY)<br>11-01-2023 | 2. REPORT TYPE<br>Final Report | 3. DATES COVERED (From - To)<br>10-Dec-2018 - 9-Mar-2022 |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>Final Report: Harnessing the architecture of natural biological modules for templated-assembly | 5a. CONTRACT NUMBER<br>W911NF-19-1-0021 |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER<br>611102 |
| 6. AUTHORS | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |

| | |
|---|---|
| 7. PERFORMING ORGANIZATION NAMES AND ADDRESSES<br>University of Texas at Austin<br>101 East 27th Street<br>Suite 5.300<br>Austin, TX 78712 -1532 | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)<br>U.S. Army Research Office<br>P.O. Box 12211<br>Research Triangle Park, NC 27709-2211 | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>ARO<br>11. SPONSOR/MONITOR'S REPORT NUMBER(S)<br>73162-BB-YIP.16 |

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 15. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON<br>19b. TELEPHONE NUMBER |
|---|---|---|---|---|---|
| UU | UU | UU | UU | | David Taylor<br>512-471-9156 |

# RPPR Final Report
as of 12-Jan-2023

Agency Code: 21XD

Proposal Number: 73162BBYIP **Agreement Number: W911NF-19-1-0021**

**INVESTIGATOR(S):**

**Name:** David W. Taylor
**Email:** dtaylor@utexas.edu
**Phone Number:** 5124719156
**Principal:** Y

Organization: **University of Texas at Austin**
Address: 101 East 27th Street, Austin, TX 787121532
Country: USA
DUNS Number: 170230239 EIN: 746000203
**Report Date:** 09-Jun-2022 Date Received: 11-Jan-2023
**Final Report** for Period Beginning 10-Dec-2018 and Ending 09-Mar-2022
**Title:** Harnessing the architecture of natural biological modules for templated-assembly
**Begin Performance Period:** 10-Dec-2018 **End Performance Period:** 09-Mar-2022
**Report Term:** 0-Other
Submitted By: David Taylor Email: dtaylor@utexas.edu
Phone: (512) 471-9156
**Distribution Statement:** 1-Approved for public release; distribution is unlimited.

**STEM Degrees:** 6 **STEM Participants:** 10

**Major Goals:** The original goals of this project were to:
1. Solve structures of polyketide synthase assembly lines from natures, and
2. Use this information to create user defined modules that could be used for various applications (electron transfer, etc.).

In agreement with the sponsor, I have shifted into a more productive project that involves using computational biology to predict protein-protein interaction interfaces to achieve similar goals of creating user defined pipelines, driven by electrostatics, hydrophobic interactions, and evolutionary coupling. It includes the following

Aims:
Aim 1: 2D and 3D structural characterization of protein complexes using cryo-EM.
We will combine biochemical fractionation with cryo-EM to investigate structures of native protein complexes directly from cells. We hypothesize that multiple 3D structures of soluble megadalton scale protein complexes can be determined using a shotgun approach to cryo-EM.

Aim 2: Development and application of shotgun cryo-EM computational tools.
We propose that new algorithms for “computational purification” of cryo-EM data will allow for higher resolution 3D models to be produced from samples containing multiple distinct protein complexes.

Aim 3: Development of a partner-specific, protein-protein interaction interface prediction tool using image processing
We hypothesize that by reducing an intricate structure of a protein into a simplified geometric cube, we will be able to store intrinsic properties on the surface of this simplified shape for a rapid and efficient analysis.

**Accomplishments:** See PDF document.

**Training Opportunities:** Nothing to Report

**Results Dissemination:** Results from this project have been the basis for a citizen science/educational video game that is in development. This will undoubtedly have a major impact.

**Honors and Awards:** American Cancer Society Research Scholar

**Protocol Activity Status:**

**Technology Transfer:** Nothing to Report

**PARTICIPANTS:**

**Participant Type:** Postdoctoral (scholar, fellow or other postdoctoral position)
**Participant:** Zhongwu Zhou
**Person Months Worked:** 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

**Participant Type:** Graduate Student (research assistant)
**Participant:** Eric Verbeke
**Person Months Worked:** 9.00 **Funding Support:**
Project Contribution:
National Academy Member: N

**Participant Type:** PD/PI
**Participant:** David Taylor
**Person Months Worked:** 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

**Participant Type:** Graduate Student (research assistant)
**Participant:** Roisin O’Brien
**Person Months Worked:** 11.00 **Funding Support:**
Project Contribution:
National Academy Member: N

**ARTICLES:**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Nature Communications
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41467-020-17411-1
Volume: Issue: First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 7/17/20 10:00AM
Publication Location:
**Article Title:** Basis for discrimination by engineered CRISPR&amp;&#x23;x2f&#x3b;Cas9 enzymes
**Authors:** Mu-Sen Liu, Shanzhong Gong, Helen-Hong Yu, Kyungseok Jung, Kenneth A. Johnson, David W. Taylc
**Keywords:** CRISPR, Cas9, biochemistry, structure-based
**Abstract:** CRISPR/Cas9 is a programmable genome editing tool that has been widely used for biological applications. While engineered Cas9s have been reported to increase discrimination against off-target cleavage compared with wild type Streptococcus pyogenes (SpCas9) in vivo, the mechanism for enhanced specificity has not been extensively characterized. To understand the basis for improved discrimination against off-target DNA containing important mismatches at the distal end of the guide RNA, we performed kinetic analyses on the high-fidelity (Cas9-HF1) and hyper-accurate (HypaCas9) engineered Cas9 variants. While DNA unwinding is the rate-limiting step for on-target cleavage by SpCas9, we show that chemistry is seriously impaired by more than 100-fold for the high-fidelity variants. The high-fidelity variants improve discrimination by slowing the rate of chemistry without increasing the rate of DNA rewinding—the kinetic partitioning favors release rather than cleavage of a bound off-target sub
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Journal of Structural Biology
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.jsb.2019.107416
Volume: Issue: First Page #:
Date Submitted: 8/30/20 12:00AM Date Published: 1/1/20 6:00AM
Publication Location:
**Article Title:** Separating distinct structures of multiple macromolecular assemblies from cryo-EM projections
**Authors:** Eric J. Verbeke, Yi Zhou, Andrew P. Horton, Anna L. Mallam, David W. Taylor, Edward M. Marcotte
**Keywords:** cryo-EM, computational methods
**Abstract:** Cryo-electron microscopy is traditionally applied to samples purified to near homogeneity as current reconstruction algorithms are unable to handle heterogeneous mixtures of structures from many macromolecular complexes. We extend on long established methods and demonstrate that relating two-dimensional projection images by their common lines in a graphical framework is sufficient for partitioning distinct protein and multiprotein complexes within the same data set. Using this approach, we first group a large set of synthetic reprojections from 35 unique macromolecular structures ranging from ?30 – 3000 kDa into individual homogenous classes. We then apply our algorithm on cryo-EM data collected from a mixture of five protein complexes and use existing reconstruction methods to solve multiple three-dimensional structures ab initio. Incorporating methods to sort cryo-EM data from heterogeneous mixtures will alleviate the need for stringent purification and pave the way toward investigat
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Journal of Chemical Information and Modeling
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.jcim.9b01164
Volume: 60 Issue: 5 First Page #: 2424
Date Submitted: 8/30/20 12:00AM Date Published: 3/1/20 6:00AM
Publication Location:
**Article Title:** Structural Biology in the Multi-Omics Era
**Authors:** Caitlyn L. McCafferty, Eric J. Verbeke, Edward M. Marcotte, David W. Taylor
**Keywords:** Mixtures, Protein structure, Sample preparation, Protein identification, Mass spectrometry
**Abstract:** Rapid developments in cryogenic electron microscopy have opened new avenues to probe the structures of protein assemblies in their near native states. Recent studies have begun applying single -particle analysis to heterogeneous mixtures, revealing the potential of structural-omics approaches that combine the power of mass spectrometry and electron microscopy. Here we highlight advances and challenges in sample preparation, data processing, and molecular modeling for handling increasingly complex mixtures. Such advances will help structural-omics methods extend to cellular-level models of structural biology.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Langmuir
Publication Identifier Type: DOI Publication Identifier: 10.1021/acs.langmuir.0c00827
Volume: 36 Issue: 26 First Page #: 7345
Date Submitted: 8/30/20 12:00AM Date Published: 6/1/20 5:00AM
Publication Location:
**Article Title:** Functionalized Mesoporous Silicas Direct Structural Polymorphism of Amyloid-&#x3f; Fibrils
**Authors:** Michael J. Lucas, Henry S. Pan, Eric J. Verbeke, Lauren J. Webb, David W. Taylor, Benjamin K. Keitz
**Keywords:** fibrils, amyloid, functional
**Abstract:** The aggregation of amyloid-? (A?) is associated with the onset of Alzheimer’s disease (AD) and involves a complex kinetic pathway as monomers self-assemble into fibrils. A central feature of amyloid fibrils is the existence of multiple structural polymorphs, which complicates the development of disease-relevant structure? function relationships. Developing these relationships requires new methods to control fibril structure. In this work, we evaluated the effect that mesoporous silicas (SBA-15) functionalized with hydrophobic (SBA-PFDTS) and hydrophilic groups (SBA- PEG) have on the aggregation kinetics and resulting structure of A?1?40 fibrils. The hydrophilic SBA-PEG had little effect on amyloid kinetics, while as-synthesized and hydrophobic SBA-PFDTS accelerated aggregation kinetics. Subsequently, we quantified the relative population of fibril structures formed in the presence of each material using electron microscopy. Fibrils formed from A?1?40 exposed to SBA-PEG were structurally
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Proteins: Structure, Function, and Bioinformatics
Publication Identifier Type: DOI Publication Identifier: 10.1002/prot.26020
Volume: 89 Issue: 3 First Page #: 348
Date Submitted: 8/30/21 12:00AM Date Published: 11/1/20 11:00AM
Publication Location:
**Article Title:** Simplified geometric representations of protein structures identify complementary interaction interfaces
**Authors:** Caitlyn L. McCafferty, Edward M. Marcotte, David W. Taylor
**Keywords:** protein interactions, shape reduction, complexes
**Abstract:** Protein-protein interactions are critical to protein function, but three-dimensional (3D) arrangements of interacting proteins have proven hard to predict, even given the identities and 3D structures of the interacting partners. Specifically, identifying the relevant pairwise interaction surfaces remains difficult, often relying on shape complementarity with molecular docking while accounting for molecular motions to optimize rigid 3D translations and rotations. However, such approaches can be computationally expensive, and faster, less accurate approximations may prove useful for large-scale prediction and assembly of 3D structures of multi-protein complexes. We asked if a reduced representation of protein geometry retains enough information about molecular properties to predict pairwise protein interaction interfaces that are tolerant of limited structural rearrangements. Here, we describe a reduced representation of 3D protein accessible surfaces on which molecular properties such a
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Nature Communications
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41467-020-19785-8
Volume: 11 Issue: 1 First Page #:
Date Submitted: 8/30/21 12:00AM Date Published: 11/1/20 5:00AM
Publication Location:
**Article Title:** Structural basis for assembly of non-canonical small subunits into type I-C Cascade
**Authors:** Roisin E. O'Brien, Inês C. Santos, Daniel Wrapp, Jack P. K. Bravo, Evan A. Schwartz, Jennifer S. Brodb
**Keywords:** CRISPR-Cas, Cascade
**Abstract:** Bacteria and archaea employ CRISPR (clustered, regularly, interspaced, short palindromic repeats)-Cas (CRISPR-associated) systems as a type of adaptive immunity to target and degrade foreign nucleic acids. While a myriad of CRISPR-Cas systems have been identified to date, type I-C is one of the most commonly found subtypes in nature. Interestingly, the type I- C system employs a minimal Cascade effector complex, which encodes only three unique subunits in its operon. Here, we present a 3.1 Å resolution cryo-EM structure of the Desul- fovibrio vulgaris type I-C Cascade, revealing the molecular mechanisms that underlie RNA- directed complex assembly. We demonstrate how this minimal Cascade utilizes previously overlooked, non-canonical small subunits to stabilize R-loop formation. Furthermore, we describe putative PAM and Cas3 binding sites. These findings provide the structural basis for harnessing the type I-C Cascade as a genome-engineering tool.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Molecular Cell
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.molcel.2020.11.003
Volume: 80 Issue: 6 First Page #: 971
Date Submitted: 8/30/21 12:00AM Date Published: 12/1/20 6:00AM
Publication Location:
**Article Title:** Diverse CRISPR-Cas Complexes Require Independent Translation of Small and Large Subunits from a Single Gene
**Authors:** Tess M. McBride, Evan A. Schwartz, Abhishek Kumar, David W. Taylor, Peter C. Fineran, Robert D. Fa
**Keywords:** CRISPR-Cas, Cascade, genome editing
**Abstract:** CRISPR-Cas adaptive immune systems provide prokaryotes with defense against viruses by degradation of specific invading nucleic acids. Despite advances in the biotechnological exploitation of select systems, mul- tiple CRISPR-Cas types remain uncharacterized. Here, we investigated the previously uncharacterized type I-D interference complex and revealed that it is a genetic and structural hybrid with similarity to both type I and type III systems. Surprisingly, formation of the functional complex required internal in-frame translation of small subunits from within the large subunit gene. We further show that internal translation to generate small subunits is widespread across diverse type I-D, I-B, and I-C systems, which account for roughly one quarter of CRISPR-Cas systems. Our work reveals the unexpected expansion of protein coding potential from within single cas genes, which has important implications for understanding CRISPR-Cas function and evolution.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Molecular Cell
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.molcel.2021.01.035
Volume: 81 Issue: 7 First Page #: 1548
Date Submitted: 8/30/21 12:00AM Date Published: 4/1/21 5:00AM
Publication Location:
**Article Title:** Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication
**Authors:** Jack P.K. Bravo, Tyler L. Dangerfield, David W. Taylor, Kenneth A. Johnson
**Keywords:** COVID-19, remdesivir, drug action
**Abstract:** Remdesivir is a nucleoside analog approved by the US FDA for treatment of COVID-19. Here, we present a 3.9-A ? -resolution cryo-EM reconstruction of a remdesivir-stalled RNA-dependent RNA polymerase complex, revealing full incorporation of 3 copies of remdesivir monophosphate (RMP) and a partially incorporated fourth RMP in the active site. The structure reveals that RMP blocks RNA translocation after incorporation of 3 bases following RMP, resulting in delayed chain termination, which can guide the rational design of improved antiviral drugs.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** iScience
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.isci.2021.102201
Volume: 24 Issue: 3 First Page #: 102201
Date Submitted: 8/30/21 12:00AM Date Published: 3/1/21 6:00AM
Publication Location:
**Article Title:** Structure of a type IV CRISPR-Cas ribonucleoprotein complex
**Authors:** Yi Zhou, Jack P.K. Bravo, Hannah N. Taylor, Jurre A. Steens, Ryan N. Jackson, Raymond H.J. Staals,
**Keywords:** CRISPR-Cas, Cascade
**Abstract:** We reveal the cryo-electron microscopy structure of a type IV-B CRISPR ribonu- cleoprotein (RNP) complex (Csf) at 3.9-A ? resolution. The complex best resembles the type III-A CRISPR Csm effector complex, consisting of a Cas7-like (Csf2) fila- ment intertwined with a small subunit (Cas11) filament, but the complex lacks subunits for RNA processing and target DNA cleavage. Surprisingly, instead of assembling around a CRISPR-derived RNA (crRNA), the complex assembles upon heterogeneous RNA of a regular length arranged in a pseudo-A-form configuration. These findings provide a high-resolution glimpse into the assembly and function of enigmatic type IV CRISPR systems, expanding our understanding of class I CRISPR-Cas system architecture, and suggesting a function for type IV-B RNPs that may be distinct from other class 1 CRISPR-associated systems.
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Protein Science
Publication Identifier Type: DOI Publication Identifier: 10.1002/pro.4067
Volume: 30 Issue: 5 First Page #: 1006
Date Submitted: 8/30/21 12:00AM Date Published: 4/1/21 5:00AM
Publication Location:
**Article Title:** Improving integrative 3D modeling into low? to medium?resolution electron microscopy structures with evolutionary couplings
**Authors:** Caitlyn L. McCafferty, David W. Taylor, Edward M. Marcotte
**Keywords:** evolutionary couplings, protein interactions
**Abstract:** Electron microscopy (EM) continues to provide near-atomic resolution structures for well-behaved proteins and protein complexes. Unfortunately, structures of some complexes are limited to low- to medium-resolution due to biochemical or conformational heterogeneity. Thus, the application of unbiased systematic methods for fitting individual structures into EM maps is important. A method that employs co-evolutionary information obtained solely from sequence data could prove invaluable for quick, confident localization of subunits within these structures. Here, we incorporate the co-evolution of intermolecular amino acids as a new type of distance restraint in the integrative modeling platform in order to build three-dimensional models of atomic structures into EM maps ranging from 10-14 Å in resolution. We validate this method using four complexes of known structure, where we highlight the conservation of intermolecular couplings despite dynamic conformational changes using the BAM compl
**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

# RPPR Final Report
as of 12-Jan-2023

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** PLOS ONE
Publication Identifier Type: DOI Publication Identifier: 10.1371/journal.pone.0245710
Volume: 16 Issue: 5 First Page #:
Date Submitted: 8/30/21 12:00AM Date Published: 5/1/21 5:00AM
Publication Location:
**Article Title:** Isolation of the Buchnera aphidicola flagellum basal body complexes from the Buchnera membrane
**Authors:** Matthew J. Schepers, James N. Yelland, Nancy A. Moran, David W. Taylor, Chih-Horng Kuo
**Keywords:** macromolecular machine

**Abstract:** Buchnera aphidicola is an intracellular bacterial symbiont of aphids and maintains a small genome of only 600 kbps. Buchnera is thought to maintain only genes relevant to the symbiosis with its aphid host. Curiously, the Buchnera genome contains gene clusters coding for flagellum basal body structural proteins and for flagellum type III export machinery. These structures have been shown to be highly expressed and present in large numbers on Buchnera cells. No recognizable pathogenicity factors or secreted proteins have been identified in the Buchnera genome, and the relevance of this protein complex to the symbiosis is unknown. Here, we show isolation of Buchnera flagellum basal body proteins from the cellular membrane of Buchnera, confirming the enrichment of flagellum basal body proteins relative to other proteins in the Buchnera proteome. This will facilitate studies of the structure and function of the Buchnera flagellum structure, and its role in this model symbiosis.

**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Nature Communications
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41467-021-25337-5
Volume: 12 Issue: 1 First Page #:
Date Submitted: 12/26/22 12:00AM Date Published: 8/1/21 8:00PM
Publication Location:
**Article Title:** SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation
**Authors:** Jurre A. Steens, Yifan Zhu, David W. Taylor, Jack P. K. Bravo, Stijn H. P. Prinsen, Cor D. Schoen, Bart
**Keywords:** CRISPR, diagnostic

**Abstract:** Red blood cells (RBCs) (erythrocytes) are the simplest primary human cells, lacking nuclei and major organelles and instead employing about a thousand proteins to dynamically control cellular function and morphology in response to physiological cues. In this study, we define a canonical RBC proteome and interactome using quantitative mass spectrometry and machine learning. Our data reveal an RBC interactome dominated by protein homeostasis, redox biology, cytoskeletal dynamics, and carbon metabolism. We validate protein complexes through electron microscopy and chemical crosslinking and, with these data, build 3D structural models of the ankyrin/Band 3/Band 4.2 complex that bridges the spectrin cytoskeleton to the RBC membrane. The model suggests spring-like compression of ankyrin may contribute to the characteristic RBC cell shape and flexibility. Taken together, our study provides an in-depth view of the global protein organization of human RBCs and serves as a comprehensive resource

**Distribution Statement:** 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Cell Reports
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.celrep.2022.111103
Volume: 40 Issue: 3 First Page #: 111103
Date Submitted: 12/26/22 12:00AM Date Published: 7/1/22 5:00AM
Publication Location:
**Article Title:** The protein organization of a red blood cell
**Authors:** Wisath Sae-Lee, Caitlyn L. McCafferty, Eric J. Verbeke, Pierre C. Havugimana, Ophelia Papoulas, Clair
**Keywords:** protein complexes
**Abstract:** Red blood cells (RBCs) (erythrocytes) are the simplest primary human cells, lacking nuclei and major organelles and instead employing about a thousand proteins to dynamically control cellular function and morphology in response to physiological cues. In this study, we define a canonical RBC proteome and interactome using quantitative mass spectrometry and machine learning. Our data reveal an RBC interactome dominated by protein homeostasis, redox biology, cytoskeletal dynamics, and carbon metabolism. We validate protein complexes through electron microscopy and chemical crosslinking and, with these data, build 3D structural models of the ankyrin/Band 3/Band 4.2 complex that bridges the spectrin cytoskeleton to the RBC membrane. The model suggests spring-like compression of ankyrin may contribute to the characteristic RBC cell shape and flexibility. Taken together, our study provides an in-depth view of the global protein organization of human RBCs and serves as a comprehensive resource
**Distribution Statement:** 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

**Publication Type:** Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
**Journal:** Nature
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41586-022-04470-1
Volume: 603 Issue: 7900 First Page #: 343
Date Submitted: 12/26/22 12:00AM Date Published: 3/1/22 6:00AM
Publication Location:
**Article Title:** Structural basis for mismatch surveillance by CRISPR–Cas9
**Authors:** Jack P. K. Bravo, Mu-Sen Liu, Grace N. Hibshman, Tyler L. Dangerfield, Kyungseok Jung, Ryan S. McC
**Keywords:** CRISPR
**Abstract:** CRISPR–Cas9 as a programmable genome editing tool is hindered by off-target DNA cleavage1,2,3,4, and the underlying mechanisms by which Cas9 recognizes mismatches are poorly understood5,6,7. Although Cas9 variants with greater discrimination against mismatches have been designed8,9,10, these suffer from substantially reduced rates of on-target DNA cleavage5,11. Here we used kinetics-guided cryo-electron microscopy to determine the structure of Cas9 at different stages of mismatch cleavage. We observed a distinct, linear conformation of the guide RNA–DNA duplex formed in the presence of mismatches, which prevents Cas9 activation. Although the canonical kinked guide RNA–DNA duplex conformation facilitates DNA cleavage, we observe that substrates that contain mismatches distal to the protospacer adjacent motif are stabilized by reorganization of a loop in the RuvC domain. Mutagenesis of mismatch-stabilizing residues reduces off-target DNA cleavage but maintains rapid on-target DNA cleavag
**Distribution Statement:** 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **N**

**Partners**

,

I certify that the information in the report is complete and accurate:
Signature: David W. Taylor
Signature Date: 1/11/23 10:11AM

The initial proposal was to use natural building blocks from bacteria to create user-defined structures that could perform various processes. It stemmed from a fruitful collaboration with Andrew Ellington, who was funded by the Army Research Office. We demonstrated a simple, robust strategy to assemble monomeric proteins into entirely synthetic, well-defined oligomers (Simon, Zhou, Ramasubraman, Glaser, Pothukuchy, Gerberich, Leggere, Morrow, Golihar, Jung, Glotzer, **Taylor**, Ellington, *Nature Chemistry*, 2019). This strategy centers on driving protein–protein interactions by combining engineered, oppositely supercharged variants of an initially monomeric protein. Generally, oppositely charged proteins associate through simple electrostatic interactions. Previously, this propensity for association has been exploited in artificial biosystems to engineer binary protein crystals, capsules, and Matryoshka-like cages from naturally charged or self-assembling proteins. We suspected that, on combining pairs of highly oppositely charged, monomeric-engineered protein variants, shape and physicochemical features would also favor assembly along particular geometrically and physiochemically favored interfaces, producing well-defined architectures rather than amorphous aggregates. The ability to simply engineer proteins to assemble into such synthetic scalable molecular assemblies via supercharging may prove useful for technologies ranging from pharmaceutical targeting to artificial energy harvesting and ‘smart’ sensing and building materials.

We initially set out to use polyketide synthases as our molecular modules. Unfortunately, all attempts at reconstituting our own user-defined assemblies failed. We even failed at reconstituting naturally occurring polyketide synthase assembly lines. Instead, we build on work where we have created a pipeline for high-throughput structural biology using cryo-EM. Traditional structural biology involves first isolating and purifying a protein or protein complex of interest before characterization by methods such as x-ray crystallography or cryo-EM. Although these methods are effective on homogeneous, isolated samples, many complexes have yet to be structurally annotated. In the current era of big data, structural biology lags behind other biological fields (e.g. genomics and proteomics) by not making use of its high-throughput potential. We designed a method that takes advantage of single particle analysis by EM and protein identification by mass spectrometry to characterize multiple protein complexes from cellular extracts. We showed three-dimensional structures of native proteasomes emerge directly from ab initio classification of a heterogeneous mixture of protein complexes. This study suggests high-throughput cryo-EM could be a valuable research tool to uncover structures and functions of macromolecular machines in a highly parallel manner.

We had successfully applied this technique to crude cell extracts (Verbeke, Mallam, Drew, Marcotte, **Taylor**, *Cell Reports*, 2018). We developed a simple, adaptable method combining microfluidic single-cell extraction with single-particle analysis by EM to characterize protein complexes from individual Caenorhabditis elegans embryos (Yi, Verbeke, Chang, Dickinson, **Taylor**, *J. Biol. Chem.*, 2019). Using this approach, we uncovered 3D structures of ribosomes directly from single cell extracts. Moreover, we investigated structural dynamics during development by counting the number of ribosomes per polysome in early and late embryos (Fig. 1). This approach has significant potential applications for counting protein complexes and studying protein architectures from single cells in developmental, evolutionary, and disease contexts. We also realized that understanding the structures and interfaces could allow us to build the structures that we initially proposed in the initial proposal to the ARO.

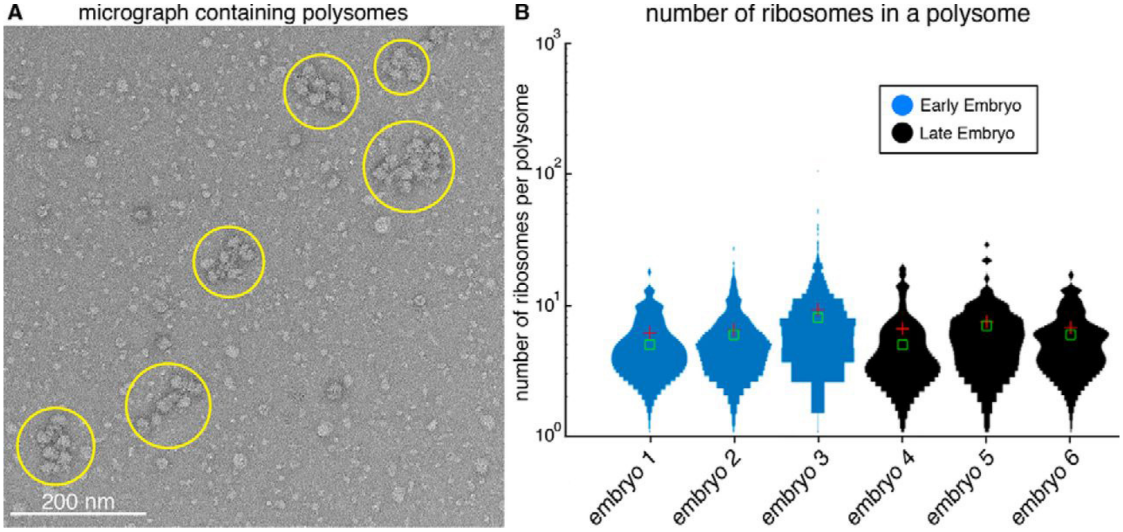

**Fig. 1. Counting ribosomes in polysomes from early- and late-stage C. elegans embryos**. (A) representative raw electron micrograph of negatively stained single-cell lysate showing several distinct polysome clusters of varying size (yellow circles). (B) distribution of the number of ribosomes in a polysome across three early- and three late-stage embryos. The average numbers of ribosomes for early- and late-stage embryos are eight and seven, respectively. The red cross-hair is the mean value, and the green box is the median (n = 81, 513, 319, 31, 71, and 52 for embryos 1–6, respectively).

<u>Current singe particle reconstruction algorithms are not designed to handle heterogeneous mixtures of structures from many distinct macromolecular complexes</u>. We demonstrated that relating two-dimensional projection images by their common lines in a graphical framework (SLICEM) is sufficient for partitioning distinct protein and multiprotein complexes within the same data set (Verbeke, Zhou, Horton, Mallam, **Taylor**, Marcotte, *J. Struct. Biol.*, 2020). The feasibility of this approach was demonstrated on a large set of synthetic reprojections from 35 unique macromolecular structures (Fig. 2.). We also used our algorithm on cryo-EM data collected from a mixture of five protein complexes and use existing methods to solve multiple three-dimensional structures ab initio. Incorporating methods to sort single particle cryo-EM data from extremely heterogeneous mixtures will alleviate the need for stringent purification.

<u>Protein-protein interactions are critical to protein function, but 3D arrangements of interacting proteins have proven hard to predict, even given the identities and 3D structures of the interacting partners.</u> We developed a reduced representation of 3D protein accessible surfaces on which molecular properties such as charge, hydrophobicity, and evolutionary rate that can be easily mapped (McCafferty, Marcotte, **Taylor**, *Proteins*, 2021). On two available benchmarks of 185 overall known protein complexes, we observe predictions comparable to other structure-based tools at correctly identifying protein interaction surfaces. This technique of shape reduction of protein surfaces retains considerable information about surface complementarity is completely novel, offers enhanced speed compared to the large number of other prediction software, and exhibits tolerance to conformational changes. We recently used an integrated modeling pipeline that incorporates evolutionary couples between proteins (Fig. 3) (McCafferty, **Taylor**, Marcotte, *Protein Sci.*, 2021). This work is quickly becoming capable of modeling 3D structures of large macromolecular complexes ab initio.

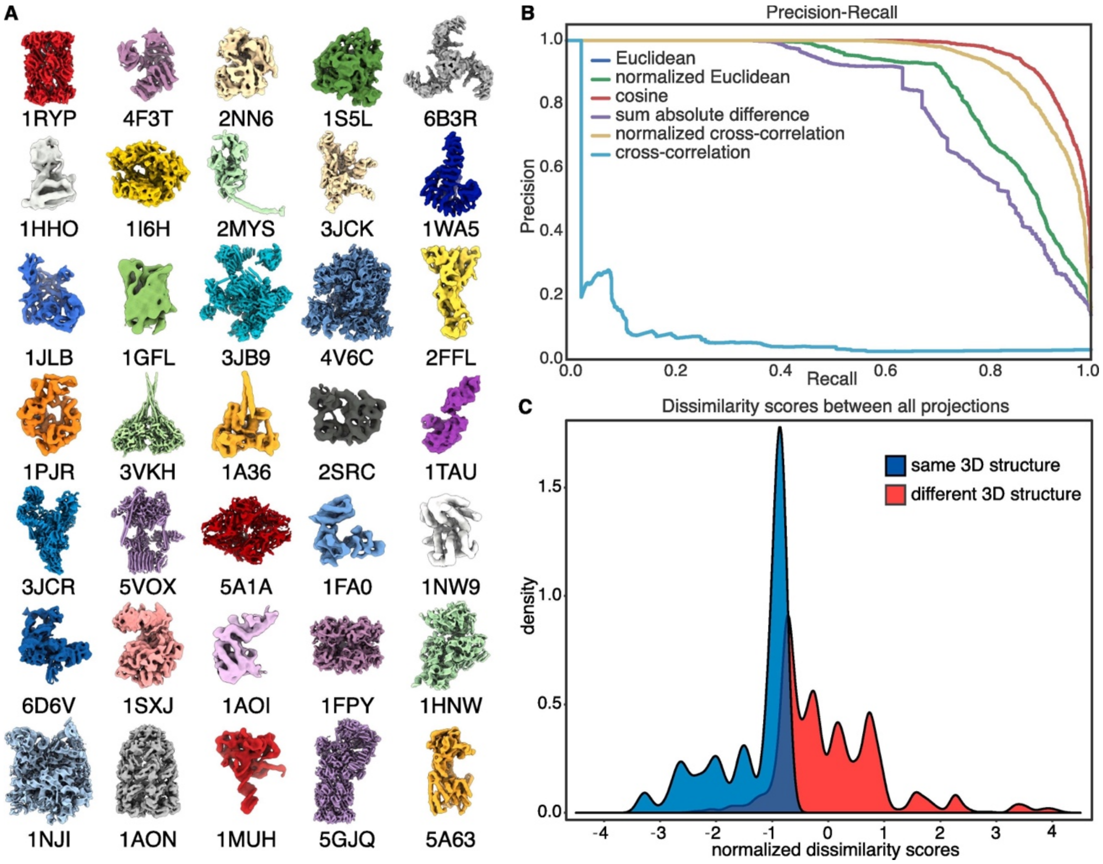

**Fig. 2. Separating mixtures of synthetic 2D reprojections.** Synthetic reprojections were generated from 35 distinct PDB structures low-pass filtered to 9 Å from protein and protein assemblies ranging in molecular weight from ~30 to 3000 kDa, prior to separation using SLICEM. (A) Low-pass filtered models of each PDB structure. (B) Precision-recall plot ranking 6 different metrics at scoring the similarity between 1D line projections from each 2D reprojection. (C) Distribution of scores calculated using Euclidean distance for reprojections belonging to the same structure and reprojections belonging to different structures.

In the final stage of the project, we used protein complexes from red blood cells (RBCs) to test our modeling ability. Most of the complexes in RBCs are involved in energy metabolism, structural integrity, redox biology, and proteostasis. Furthermore, we performed chemical crosslinking on these native complexes and used integrative structural modeling to shed light on the molecular organization of integral membrane proteins, channels, cytoskeletal proteins, and metabolic enzymes at the RBC membrane. Our findings provide a comprehensive blueprint of the cell surface and subcellular architecture of a key blood cell type and suggest biophysical mechanisms underlying its adaptability (Sae-Le, McCafferty, Verbeke, Havugimana, Papoulas, McWhite, Houser, Vanuytsel, Murphy, Drew, Emili, Taylor, Marcotte, *Cell Rep*orts, 2022).

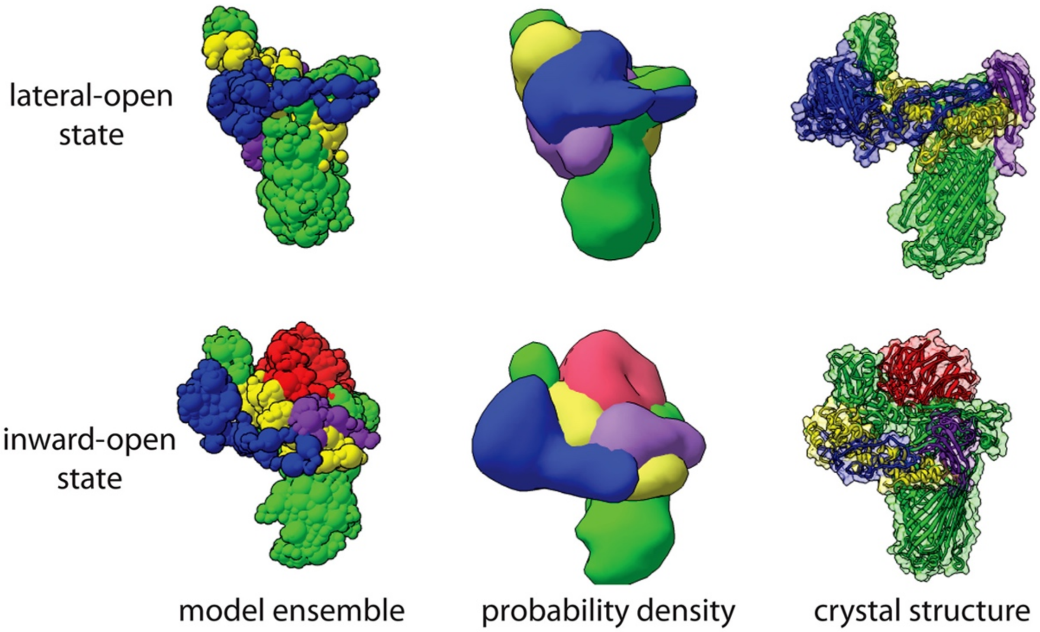

**Fig. 3. Evaluation of evolutionary coupling restraints on a dynamic structure.** The BAM complex illustrates the robustness of evolutionary couplings against conformational changes in dynamic complexes. The BAM complex exists in a lateral-open state (PDB: 5D0Q) and an inward open state (PDB: 5D0O), which includes an additional subunit. (c) The combination of the EM map and evolutionary couplings was enough data to produce each of the states as compared to the crystal structures.

We integrated known and predicted interaction interfaces to build models of several of the major RBC membrane and cytoskeletal complexes, whose structures have remained elusive. Our modeling approach provided us with a detailed molecular view of both the Band 3 (Fig. 4, top) and Band 4.1 complexes (Fig. 4, bottom). We used crosslinking mass spectrometry to validate these predications. Indeed, the Band3-ankyrin1 complex has been proposed to locate in the middle of the spectrin tetramer chain to anchor spectrins to the plasma membrane, while the Band 4.1 complex connects the end of the tetrameric spectrin chain to one of six binding sites on actin in the actin junctional complex. Our model of the Band3-ankyrin1 complex shows the association of Band 3 with Band 4.2, Rh proteins, GlutI, ankyrin, spectrin, and glycophorin A. Intriguingly, this suggests that we could create user defined structures anchored in membranes for energy transport, like those in plants. Thus, while our proposed research diverted in a different direction, it ended at the same destination. As a result of this ARO funded research, we are closer to our objective of creating user defined modules for technological applications.

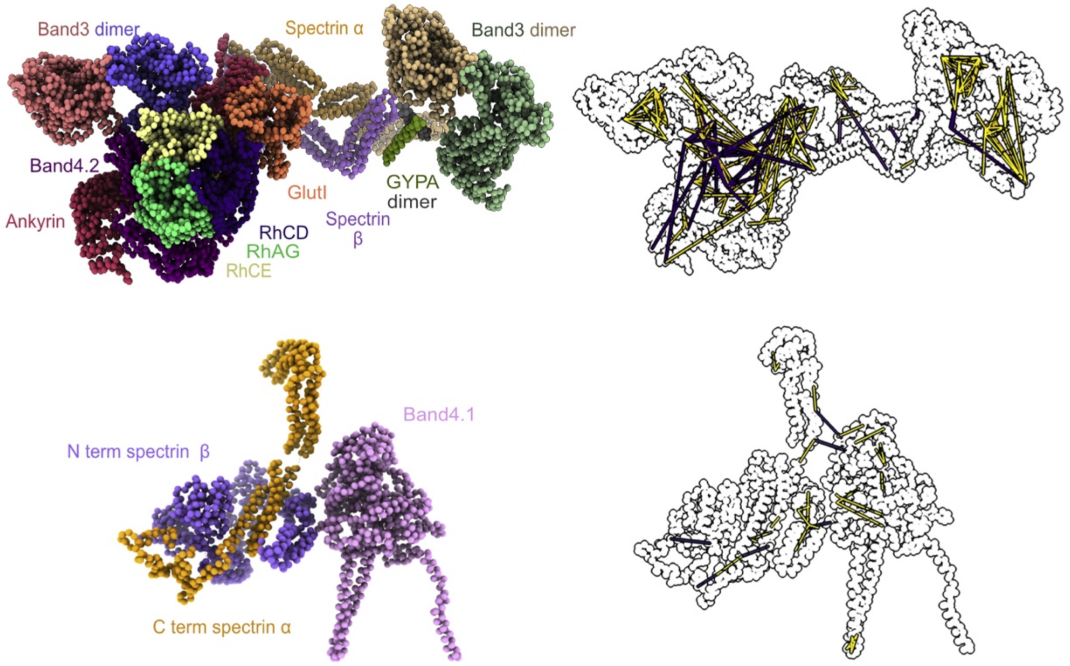

**Fig. 4. 3D modeling of membrane and cytoskeletal complexes.** Top, Side view of integrative structure of Band 3-ankyrin1 complex and band 3-GYPA complex. Our model suggests that GlutI competes with the Rh proteins for binding with Band 3 and Band 4.2. The outline figure on the right shows intramolecular and intermolecular crosslinks that are overlaid onto the structure. Bottom, Integrative structure of Band 4.1-spectrin complex. Six of these complexes are proposed to link spectrin heterodimers with actin.