Predicting the fitness landscape of β-lactamase deletion mutants

β-lactamase is an antibiotic resistance factor

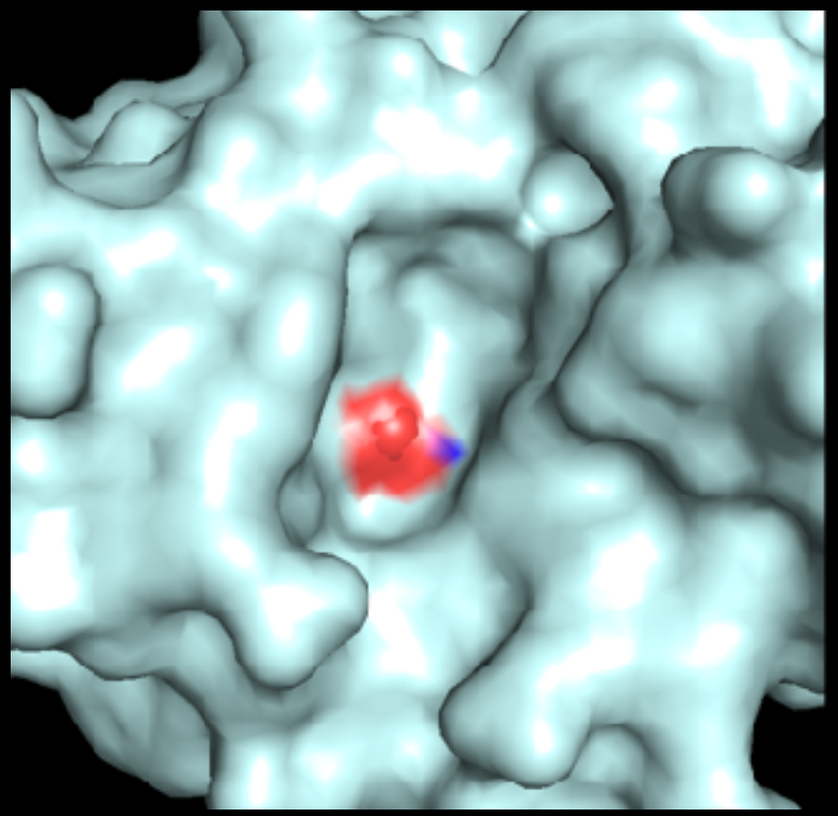
- confers resistance to β-lactam antibiotics
  - penicillin, cephalosporin, many more
  - most interfere with cell wall construction (growth)

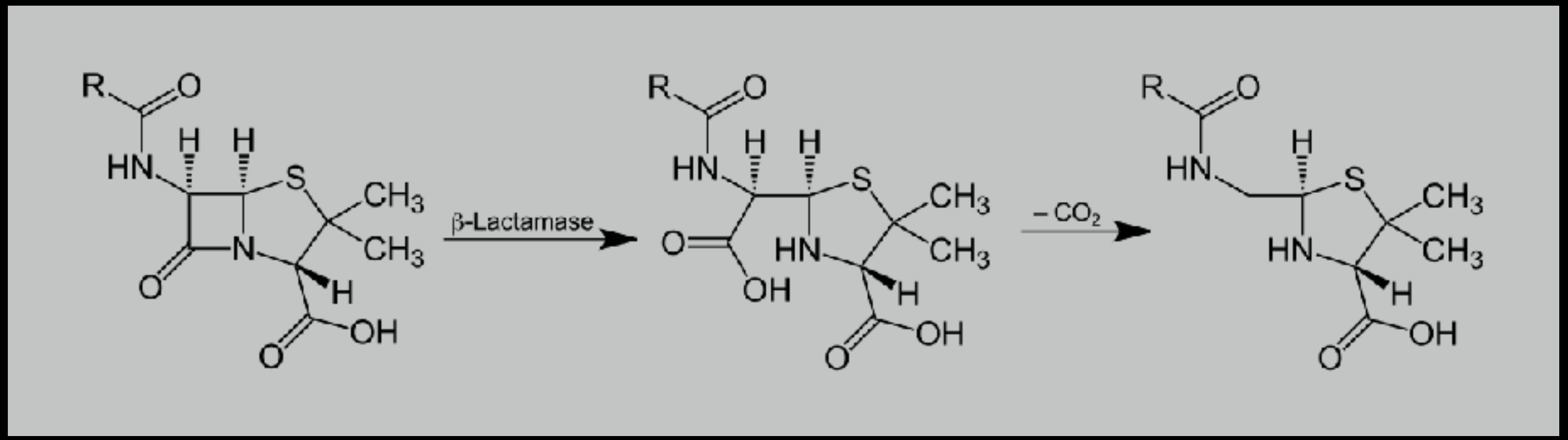
- here, TEM-1 β-lactamase was used (PDB:1BTL)
  - Isolated from E. coli

Illustration: Punnoose AR, Lynm C, Golub RM. Antibiotic Resistance. JAMA. 2012

The key residue is a serine at position 70

- Serine oxygen acts as a nucleophile for the ring-opening reaction
- Nearby Glu166 helps increase nucleophilicity





Protein Images: coordinates from crystal structure, Jelsch et.al., Reaction Structures: Wikimedia Commons

What is fitness?

- **Fitness** is the ability of an organism to survive and reproduce
- **Fitness landscapes** are a model for protein evolution where sequence changes and a local optimum is reached
- Understanding the **distribution of fitness effects** (the impact of mutating each residue) is the goal of this research -> understanding protein evolution

Mapping a protein's fitness landscape

Experiment: e Coli mutants are exposed to ampicillin -> deep sequencing to determine results

Rosetta: mutant proteins are created -> proxies for fitness are assessed (more on this later)

Previous fitness landscape work

A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape

Elad Firnberg,[1] Jason W. Labonte,[1] Jeffrey J. Gray,[1] and Marc Ostermeier*,[1]
[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University
***Corresponding author:** E-mail: oster@jhu.edu.
Associate editor: Howard Ochman

Abstract

Mutations are central to evolution, providing the genetic variation upon which selection acts. A mutation's effect on the suitability of a gene to perform a particular function (gene fitness) can be positive, negative, or neutral. Knowledge of the distribution of fitness effects (DFE) of mutations is fundamental for understanding evolutionary dynamics, molecular-level genetic variation, complex genetic disease, the accumulation of deleterious mutations, and the molecular clock. We present comprehensive DFEs for point and codon mutants of the *Escherichia coli TEM-1* β-lactamase gene and missense mutations in the TEM-1 protein. These DFEs provide insight into the inherent benefits of the genetic code's architecture, support for the hypothesis that mRNA stability dictates codon usage at the beginning of genes, an extensive framework for understanding protein mutational tolerance, and evidence that mutational effects on protein thermodynamic stability shape the DFE. Contrary to prevailing expectations, we find that deleterious effects of mutation primarily arise from a decrease in specific protein activity and not cellular protein levels.

Key words: **protein evolution, fitness landscape, beta-lactamase.**

Introduction

The fitness landscape model for protein evolution, as first

significant limitations. First, most studies utilize nonnative reporter assays (e.g., phage display, cell surface display, and

So what about deletions?

- My work deals with deletions involving exactly one codon (three nucleotides, one amino acid) - “**in-frame, single codon deletion mutants**”
- Understanding these types of mutations, however rare and improbable they seem, has important clinical applications

Clinical implications of in-frame deletion mutants

nature genetics

An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy

Michael N Weedon[1,12], Sian Ellard[1,12], Marc J Prindle[2,12], Richard Caswell[1], Hana Lango Allen[1], Richard Oram[1], Koumudi Godbole[3,4], Chittaranjan S Yajnik[4], Paolo Sbraccia[5,6], Giuseppe Novelli[5,6], Peter Turnpenny[7], Emma McCann[8], Kim Jee Goh[9,10], Yukai Wang[9,10], Jonathan Fulford[1], Laura J McCulloch[1], David B Savage[9,10], Stephen O'Rahilly[9,10], Katarina Kos[1], Lawrence A Loeb[2,11], Robert K Semple[9,10] & Andrew T Hattersley[1]

DNA polymerase δ, whose catalytic subunit is encoded by *POLD1*, is responsible for lagging-strand DNA synthesis during DNA replication[1]. It carries out this synthesis with high fidelity owing to its intrinsic 3′- to 5′-exonuclease activity, which confers proofreading ability. Missense mutations affecting the exonuclease domain of POLD1 have recently been shown to predispose to colorectal and endometrial cancers[2]. Here we report a recurring heterozygous single-codon deletion in *POLD1* affecting the polymerase active site that abolishes DNA polymerase activity but only mildly impairs 3′- to 5′ exonuclease activity. This mutation causes a distinct

Because all reported individuals with MDP syndrome have unrelated parents and no other affected family members, we hypothesized that the syndrome was caused by a heterozygous *de novo* mutation in a single gene. We therefore performed exome sequencing on two probands with MDP syndrome (**Fig. 1** and **Supplementary Table 1**) and their unaffected parents to look for candidate *de novo* disease-causing mutations. Exonic sequences were enriched from genomic DNA using the Agilent SureSelect Human All Exon kit (version 4) and then sequenced on an Illumina HiSeq 2000 sequencer using 100-bp paired-end reads. We used Burrows-Wheeler aligner (BWA v0.6.2)[5] to

Clinical implications of in-frame deletion mutants

nature genetics

An in-frame deletion
POLD1 causes a mult

Michael N Weedon[1,12], Sian Ellard[1,12], Ma
Koumudi Godbole[3,4], Chittaranjan S Yajn
Emma McCann[5], Kim Jee Goh[9,10], Yukai
Stephen O'Rahilly[9,10], Katarina Kos[1], Law

DNA polymerase δ, whose catalytic subunit is e
POLD1, is responsible for lagging-strand DNA s
during DNA replication[1]. It carries out this synt
fidelity owing to its intrinsic 3′- to 5′-exonuclea
which confers proofreading ability. Missense mu
affecting the exonuclease domain of POLD1 ha
been shown to predispose to colorectal and end
cancers[2]. Here we report a recurring heterozyg
codon deletion in *POLD1* affecting the polymer
that abolishes DNA polymerase activity but only
3′- to 5′-exonuclease activity. This mutation cau

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Congenital Insensitivity to Pain: Novel *SCN9A* Missense and In-frame Deletion Mutations

James J. Cox [1*], Jony Sheynin [2,3*], Zamir Shorer [4*], Frank Reimann [5*], Adeline K. Nicholas [1], Lorena Zubovic[6], Marco Baralle [6], Elizabeth Wraige [7], Esther Manor [2,8], Jacov Levy [4], C. Geoffery Woods [1*§], and Ruti Parvari [2,3*#]

[1] *Department of Medical Genetics, University of Cambridge, UK;* [2] *Department of Virology and Developmental Genetics, Faculty of Health Sciences, and* [3]*National Institute of Biotechnology in the Negev, Ben Gurion University of the Negev, Israel;* [4] *Division of Pediatrics, Soroka Medical Center and Faculty of Health Sciences, Ben Gurion University of the Negev, Israel;*[5] *Department of Clinical Biochemistry, University of Cambridge, UK;* [6] *International Centre for Genetic Engineering and Biotechnology, Trieste, Italy;* [7] *Guy's and St Thomas' NHS Foundation Trust, London, UK;* [8] *Institute of Genetics, Soroka Medical Center and Faculty of Health Sciences, Beer Sheva, Israel.*

* *These authors equally contributed to this work*

Correspondence to Prof. Parvari and Dr. Woods. RP: Ben Gurion University of the Negev, Beer Sheva 84105, Israel; Tel.: 972 8 6479967; FAX.: 972 8 6276215; E-mail: ruthi@bgu.ac.il; CGW: Cambridge Institute for Medical Research, Cambridge, CB2 0XY, UK; Tel.: +44 1223 767811; FAX.: +44 1223 331206; E-mail: cw347@cam.ac.uk

Communicated by Claude Ferec

ABSTRACT: *SCN9A* encodes the voltage-gated sodium channel Na$_v$1.7, a protein highly expressed in pain-sensing neurons. Mutations in *SCN9A* cause three human pain disorders: bi-allelic loss of function mutations result in Channelopathy-associated Insensitivity to Pain (CIP), whereas activating mutations cause severe episodic pain in Paroxysmal Extreme Pain Disorder (PEPD) and Primary Erythermalgia (PE). To date, all mutations in *SCN9A* that cause a complete inability to experience pain are protein truncating and presumably lead to no protein being produced. Here, we

Clinical implications of in-frame deletion mutants

Haemophilia (2008), **14**, 1094–1098 DOI: 10.1111/j.1365-2516.2008.01816.x

ORIGINAL ARTICLE *Laboratory investigation*

Severe haemophilia A in a female resulting from an inherited gross deletion and a *de novo* codon deletion in the *F8* gene

A. VENCESLÁ,* P. FUENTES-PRIOR,† M. BAENA,* M. QUINTANA,‡ M. BAIGET* and E. F. TIZZANO*

**Department of Genetics, Hospital de la Santa Creu i Sant Pau, Barcelona; †Cardiovascular Research Center, CSIC-ICCC, Hospital de la Santa Creu i Sant Pau, Barcelona; and ‡Haemophilia Unit, Hospital La Paz, Madrid, Spain*

Summary. Haemophillia A (HA) is an X-linked bleeding disorder caused by mutations in the *F8* gene. While the disease affects 1 in 5000 males, phenotypic expression of haemophilia A is rare in females, similar to other X-linked recessive disorders. We describe a 5-year-old female with severe haemophilia A. We determined the underlying molecular defect in the *F8* genes of the proband and her closest family members by direct DNA sequencing, marker analysis and quantitative real-time polymerase chain reaction. The patient showed two different mutations while the maternally inherited gene showed a large deletion encompassing exons 1 to 22. The structural analysis of residues Phe652/Phe653 based on a three-dimensional model of activated factor VIII provides evidence of the impact of the mutant factor VIII protein in the clinical manifestations of the patient. This unusual finding highlights the need to perform a thorough molecular analysis including sequencing, marker and quantitative analyses to identify compound heterozygous females with HA.

Previous work with deletion mutants

Published in final edited form as:
Proteins. 2011 October ; 79(10): 2844–2860. doi:10.1002/prot.23109.

Computed structures of point deletion mutants and their enzymatic activities

Monica Berrondo and **Jeffrey J. Gray**
Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, 21218

Abstract

Point deletions in enzymes can vary in effect from negligible to complete loss of activity, however, these effects are not generally predictable. Deletions are widely observed in nature and often result in diseases such as cancer, cystic fibrosis, or osteogenesis imperfecta. Here, we have developed an algorithm to model the perturbed structures of deletion mutants with the ultimate goal of predicting their activities. The algorithm works by deleting the specified residue from the wild-type structure, creating a gap that is closed using a combination of local and global moves that change the backbone torsion angles of the protein structure. On a set of five proteins for which both wild-type and deletion mutant x-ray crystal structures are available, the algorithm produces deep, narrow energy funnels within 1.5 Å of the crystal structure for the deletion mutants. To

Identify secondary structure

Delete residue and shift gap to loop

Close gap in loop

Generate diversity

Relax structure to minimize energy

performed for each residue (263 in β-lactamase)

Identify secondary structure

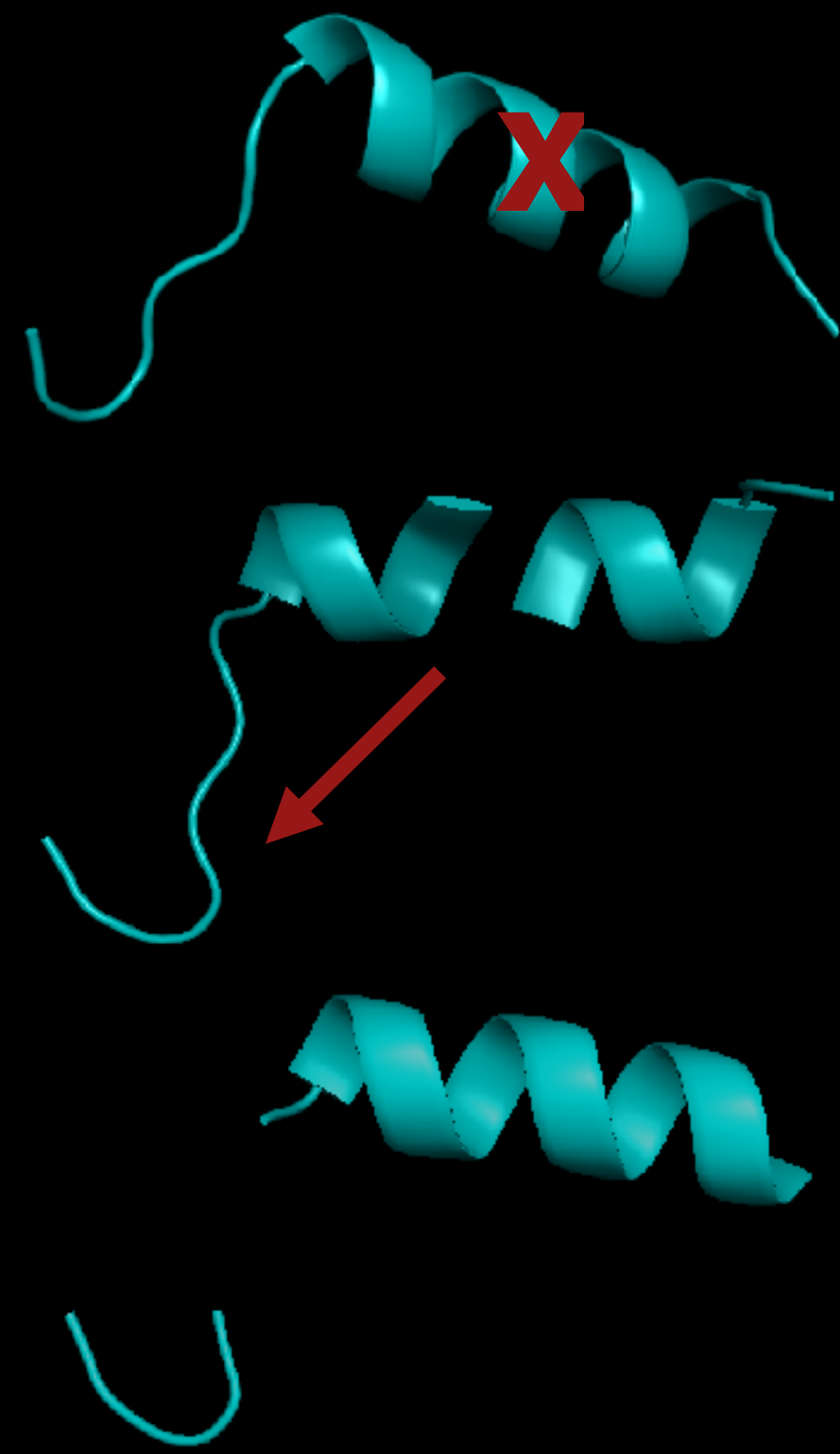
Delete residue and shift gap to loop

Close gap in loop

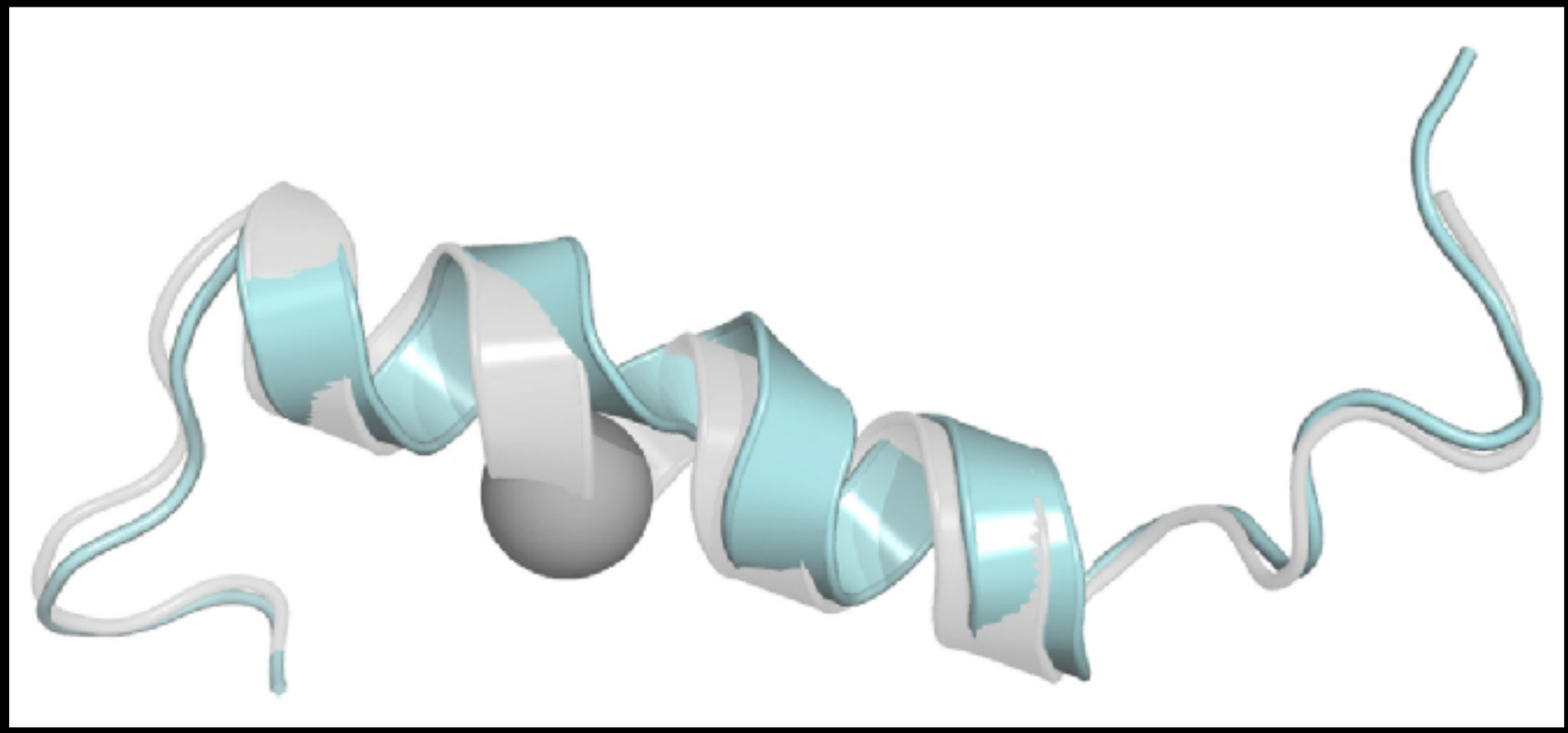
Generate diversity

Relax structure to minimize energy

"Thread-and-Close"



Why move to the loop?



- “loop” closing causes helices with unrealistic torsion angles and the ‘wrong’ H-bonding pattern (not i -> i+4)

Figure from M. Berrondo and J.J.Gray 2011

Identify secondary structure

Delete residue and shift gap to loop

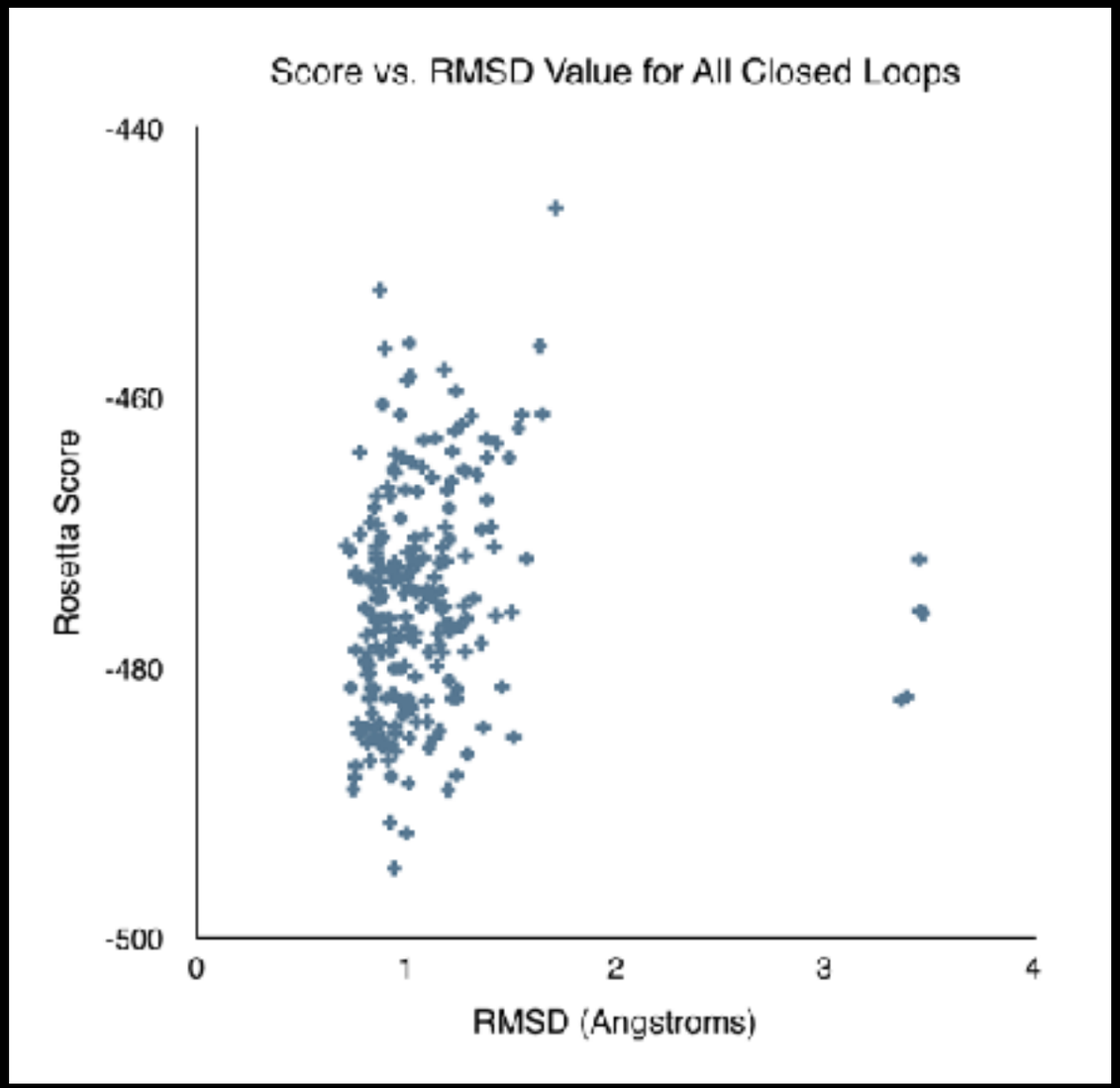
Close gap in loop

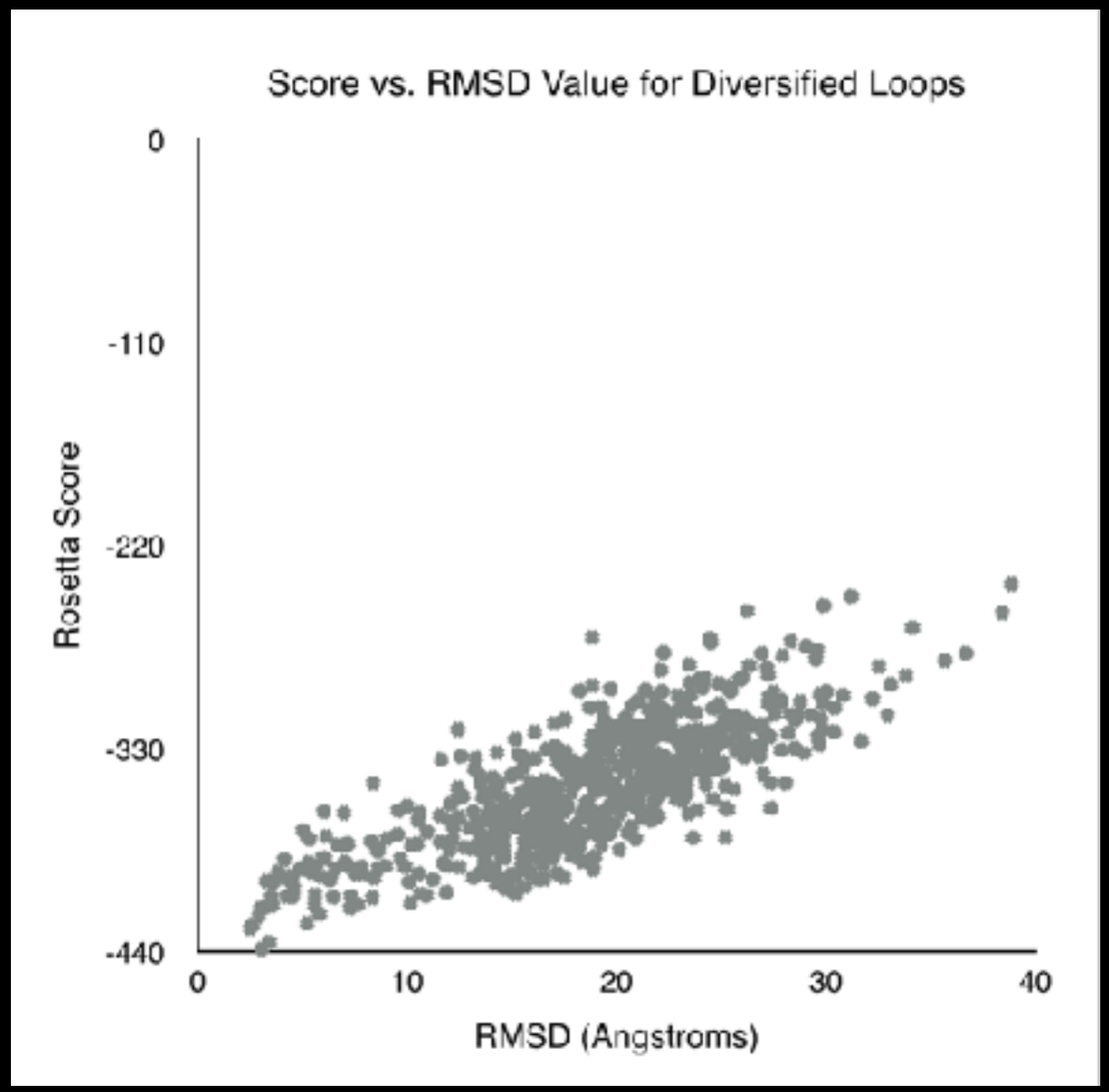
Generate diversity

Relax structure to minimize energy

performed for each residue (263 in β-lactamase)

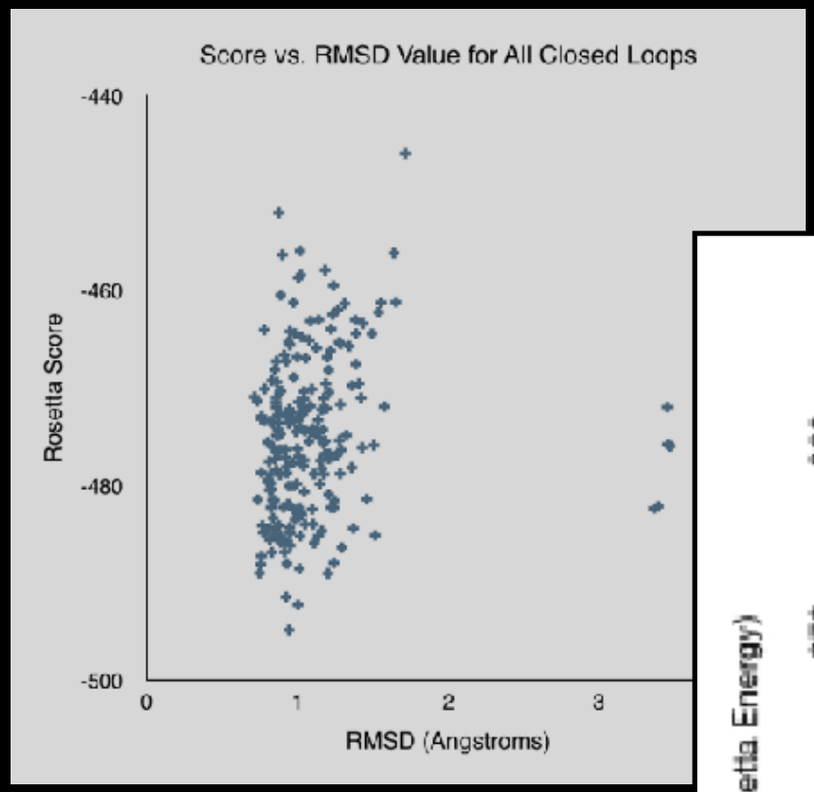
Preliminary Data - Does my diversity make sense?

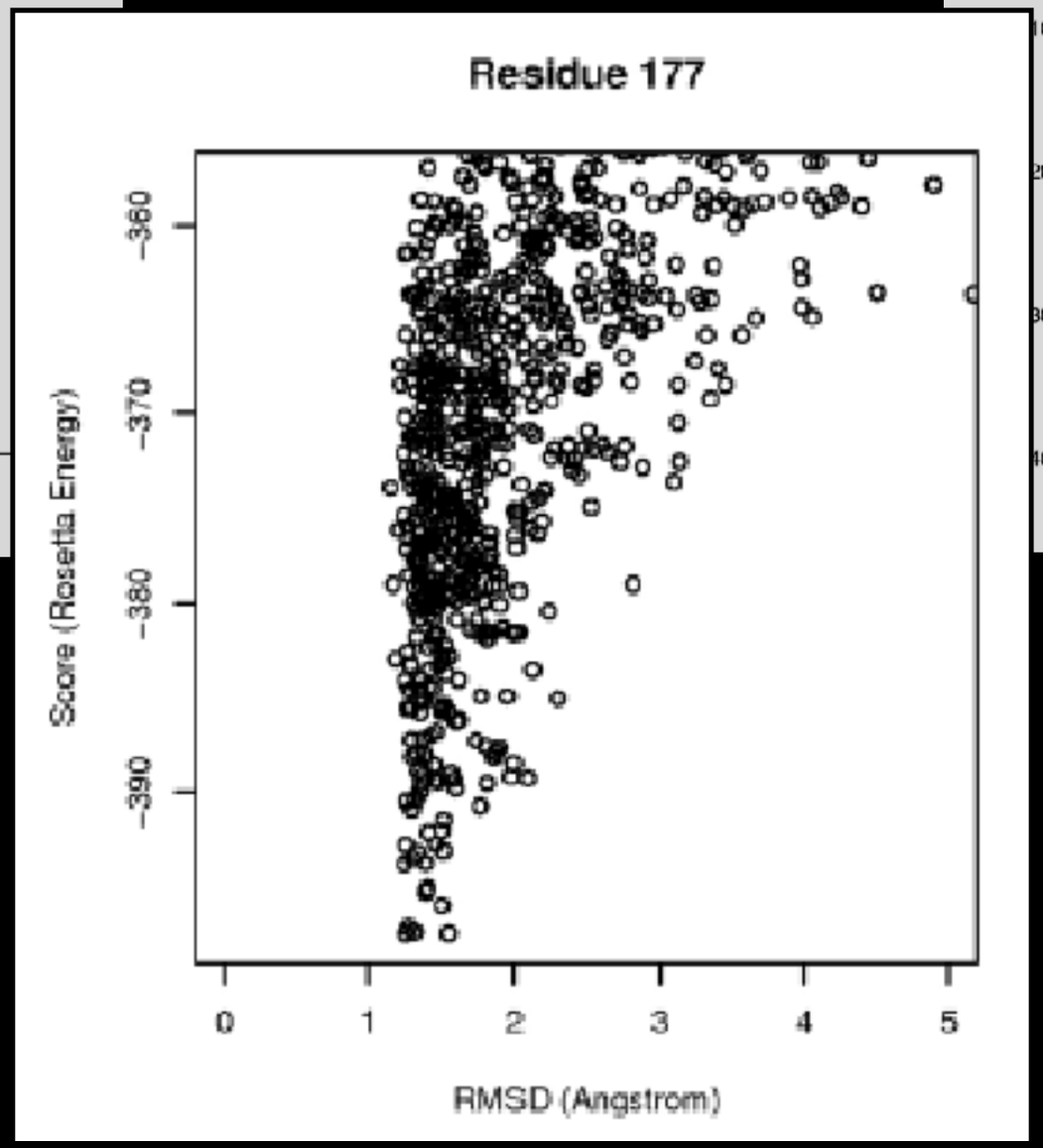


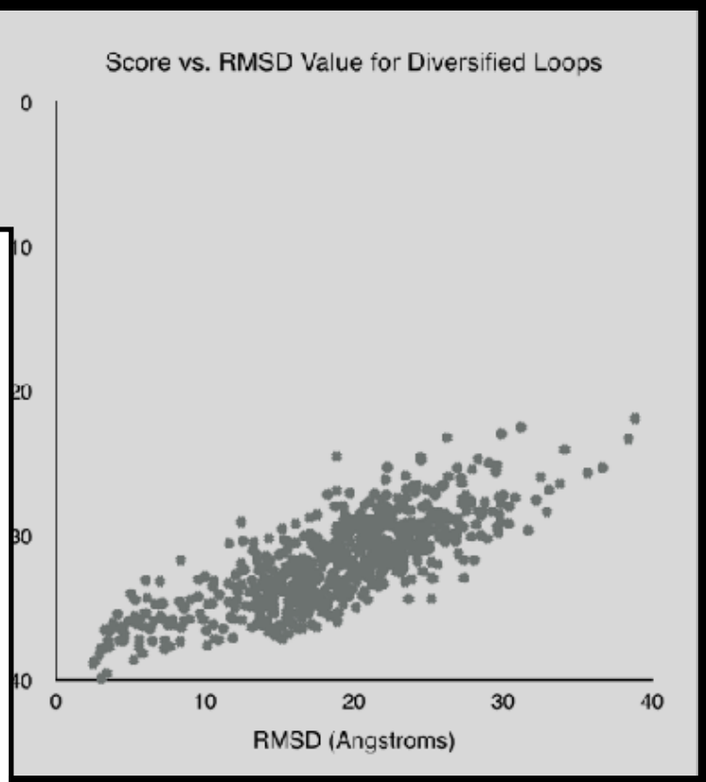


- Diversification procedure is too extreme?

Preliminary Data

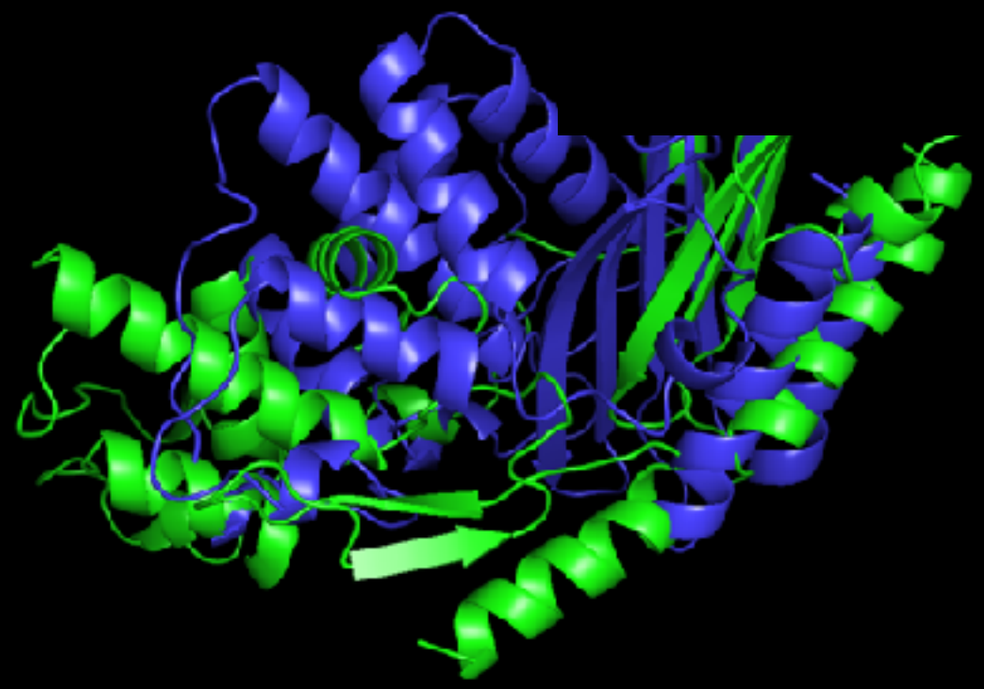






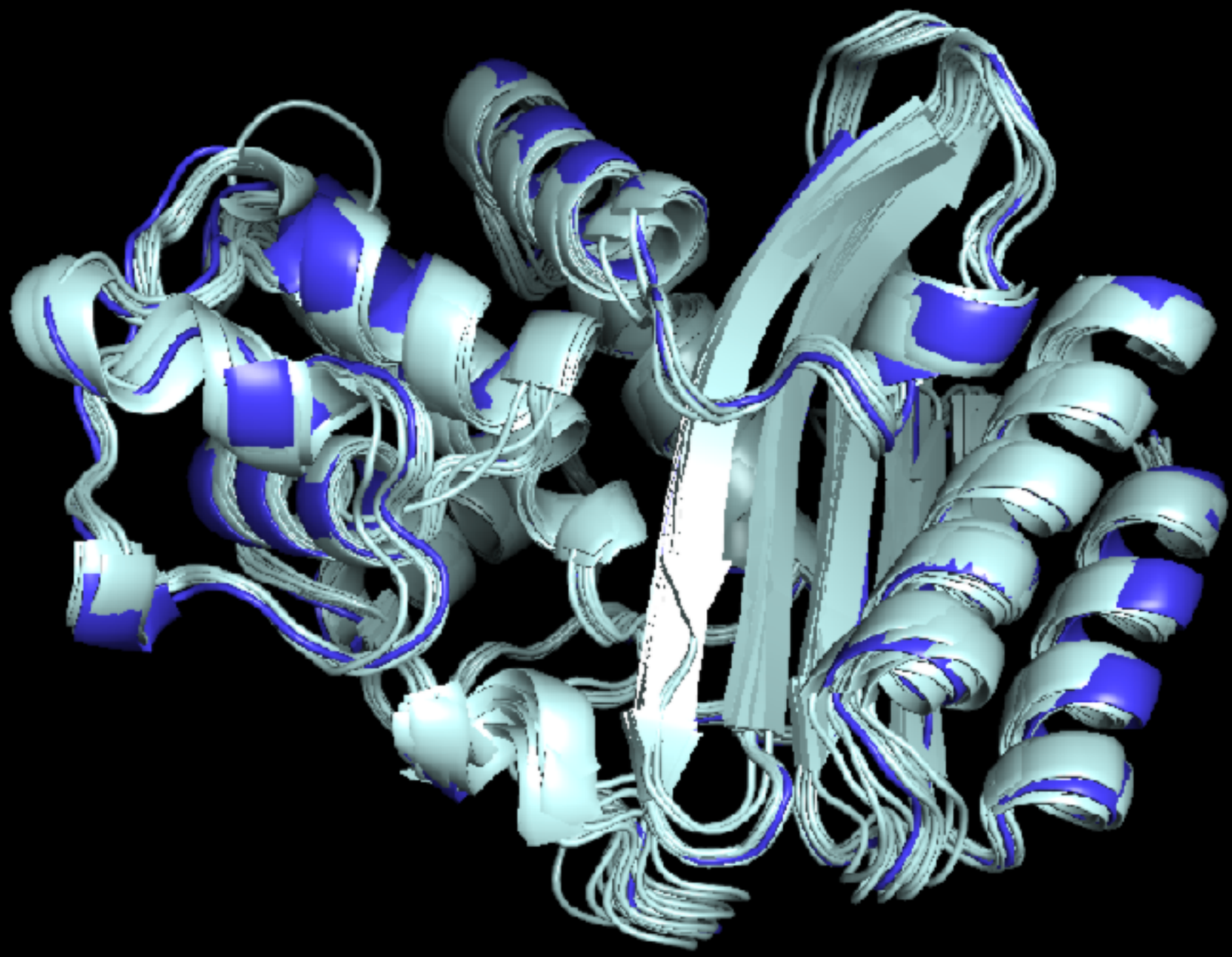
Center figure from M. Berrondo and J.J.Gray 2011

Diversified structures



Diversification too extreme?

Closed loop (no diversity)



Refining this algorithm

- Fine tuning diversity generation
- Creating large, statistically significant data sets (1000+/position)

Determining fitness in-silico

What measures of fitness are appropriate for computational contexts?

- **Stability** (Rosetta score) - previous validation
- **Active site geometry** (position of nucleophile, etc)?
- **Pocket** size, shape, accessibility?
- ? - ideas welcome!

(implementing and validating these is a work in progress)

Rosetta score has been shown to correlate well with experimental measures of fitness

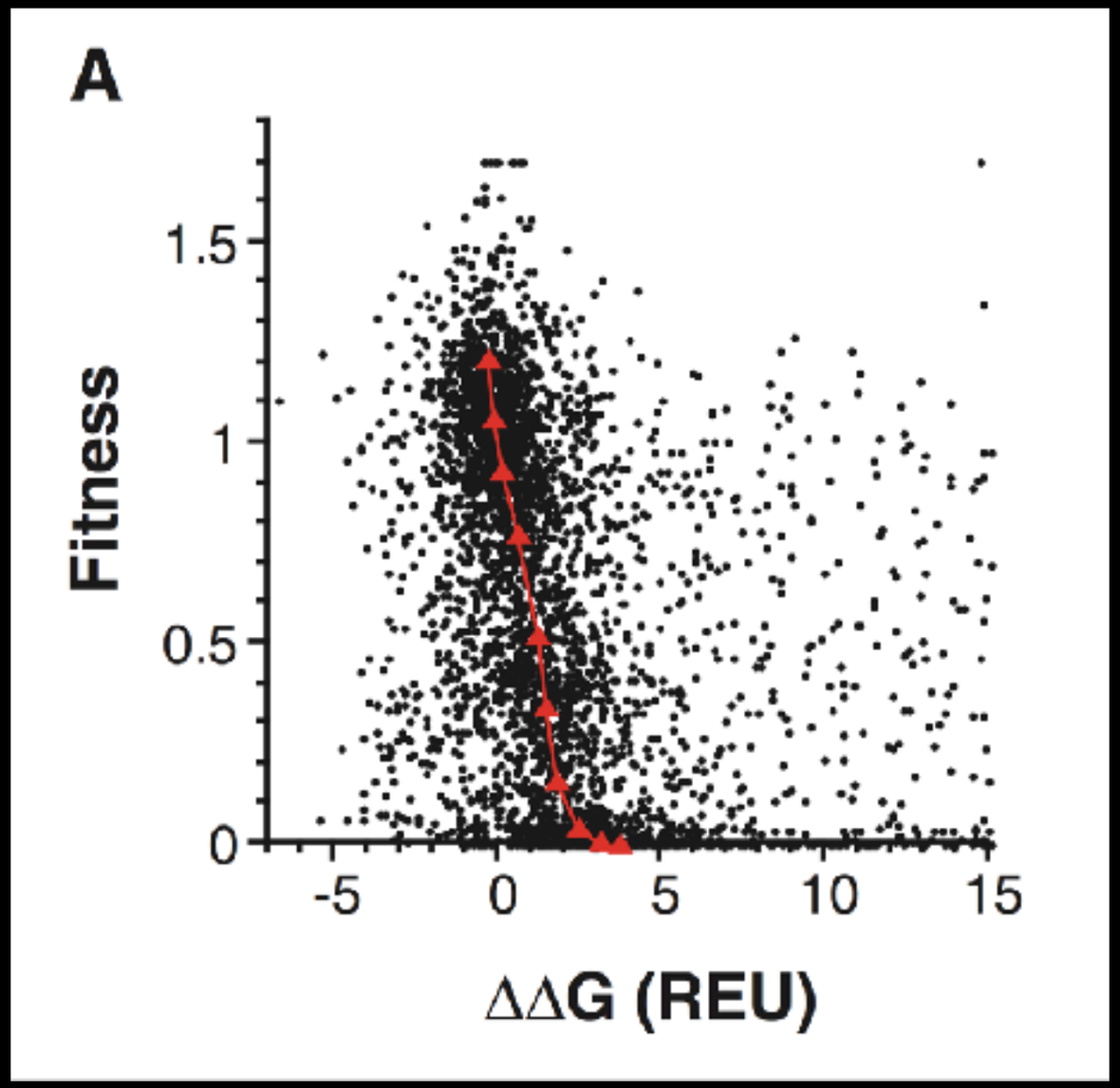


Figure from Firnberg, et al. 2014

Toward an improved overall algorithm for handling indels in secondary structural elements

- Proteins in-vivo don’t fold first, and have their mutations removed later.
- Peptides coming off the ribosome tend to fold in a somewhat domain-wise fashion.

Find and anchor ends of helix

Delete residue

Break secondary structure

Close gap in loop

Constrain backbone to retrieve ss.

Docking?

Helix

Identify secondary structure

Loop

Close gap in loop

Generate diversity

Final Relax

thanks!