DNASTAR®
Software for Life Scientists

Structure-guided molecular cloning for improving site-directed mutagenesis and stability in protein design

Steven J. Darnell*, Matthew R. Larson, Constantin Tivis, Martin Riese, Matthew Keyser, Katie Maxfield, Frederick R. Blattner

DNASTAR, Inc. 3801 Regent St, Madison, WI USA
www.dnastar.com

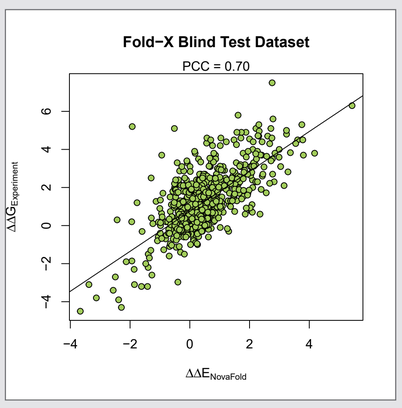
Abstract

Protein stability is essential when optimizing the expression, purification, and formulation conditions for protein functional studies. Unstable proteins are often co-expressed with chaperones or fused to affinity tags to improve solubility and expression. These approaches are ineffective when constructs are prone to degradation or thermodynamic instability. In these cases, established protocols suggest removing protease sites or introducing mutations that improve thermodynamic stability. However, predicting the contextual effects of these changes requires a 3D structure and an approach to evaluate many sequence perturbations within the protein fold.

We present a deterministic protein design algorithm that optimizes stability in a protein by identifying the sequence in a given search space associated with the Global Minimum Energy Conformation (GMEC) and pinpointing potential liabilities in the resulting molecule. When searching for the GMEC, variant and surrounding side chains are repacked and backbone atoms are perturbed in a chemically-plausible manner to accommodate the changes. We also report the accuracy of the algorithm using a benchmark of over 650 mutations with thermodynamically characterized changes in fold stability. In addition, we present a case study where we redesign the core of a protein to improve its fold stability and design cloning primers to synthesize the new molecule using an array of recombinant DNA technologies. The entire workflow can be completed within an hour using a Windows or macOS workstation, highlighting the approach's practicality for use by any life scientist.

NovaDesign Benchmarks

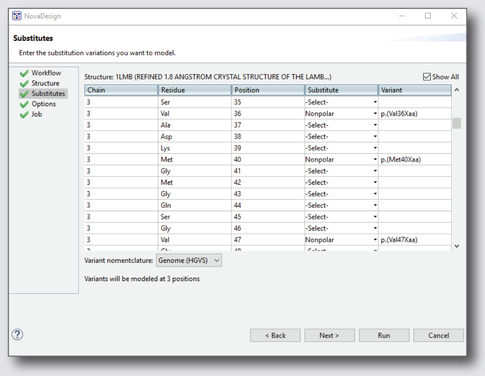
NovaDesign uses the DEE/A* method to deterministically find the global minimum energy conformation (GMEC) for a protein design problem. Unlike stochastic approaches, NovaDesign guarantees it will find the GMEC for a given energy function and flexibility model. Our approach currently uses the Penultimate rotamer library to model side chain flexibility, the Backrub model to model backbone flexibility, and the DFIRE statistical potential to evaluate variants and conformations. NovaDesign can instantly model specific variants, while larger searchers for optimizing fold stability can take minutes to hours. We are also planning to introduce backbone-dependent rotamers and molecular mechanics force fields in the future.

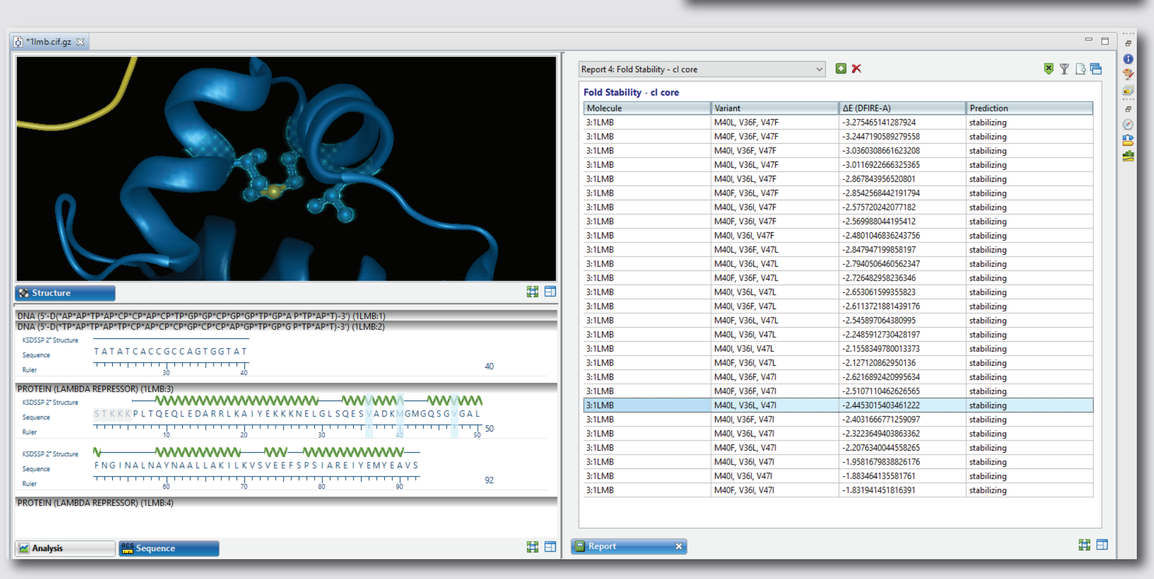


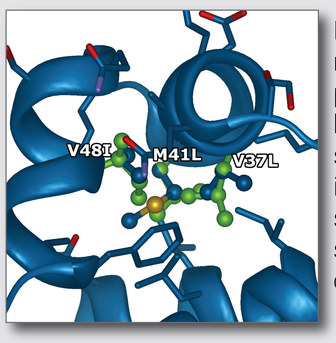
The Fold-X Blind Test benchmark contains 664 substitution variants, each with an experimentally-determined change in free energy (ΔΔG) relative to wild type. NovaDesign calculated the change in fold stability (ΔΔE) for the entire data set in less than 10 minutes on a single workstation with a Pearson's linear correlation coefficient (PCC) of 0.70, indicating a strong correlation between experimental and calculated values. Backrub backbone flexibility was used and side chains of neighboring residues were repacked.

Protein Design Tools to Improve Fold Stability

Nature commonly evolves proteins to provide sufficient fold stability to optimize its activity—not its energetic stability. Buried core residues can be altered to improve a protein's developability and stability, as is the case with Repressor protein cI. The variant p.(Val37Leu; Met41Leu; Val48Ile) (LLI) confers 0.5 kcal/mol more stability than wild type (Lim, et al., 1994, Proc. Natl. Acad. Sci. USA, 91:423-7). Using the wild type structure (PDB ID: 1LMB), NovaDesign reproduces this observation and suggest several other variant combinations that could provide even greater thermostability.







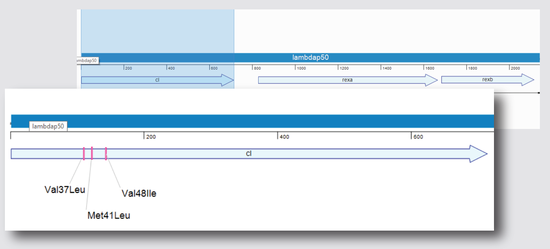
First, positions Val 37, Met 41, and Val 48 are annotated to allow the residues to be replaced with any hydrophobic residue (AGILMPV). Next, pairwise energies are precalculated for all potential combinations, high-energy conformations are excluded, and the GMEC and alternate sequence permutations are identified in energetic order. NovaDesign identified the LLI sequence as the 21st lowest-energy sequence in its search, suggesting 20 other permutations that have even greater fold stability. All computations were completed in under a minute on a laptop computer, highlighting the accessibility of the software.

Molecular Cloning Workflow

The SeqBuilder Pro primer workbench allows users to create PCR primer pairs at the exact desired location with respect to features and translational frames. The point-and-click "Introduce Mutation" tool allows a user to easily modify triplet codons within primers to introduce the desired mutations. SeqBuilder Pro provides multi-fragment PCR based cloning options so that the mutated gene of interest can be cloned into an appropriate expression vector. The automated virtual cloning process creates all the required PCR primers, a cloning history, and the clone sequences. After virtual cloning, users can visualize all the of the vector and insert feature annotations and apply ORF detection and translation to confirm that the gene of interest remains in-frame with the fusion protein.

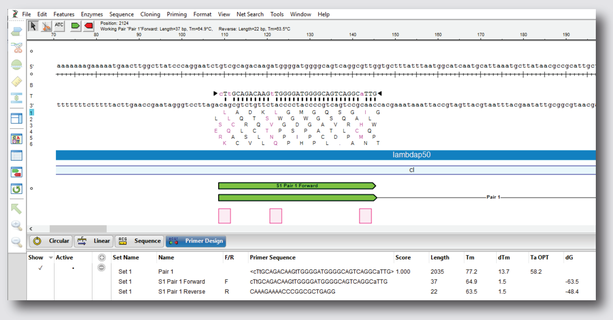
IDENTIFY VARIANTS

The cI gene is the first of three in the lambdap50 operon. There are three predicted mutations in the cI gene: Val37Leu; Met41Leu; and Val48Ile.



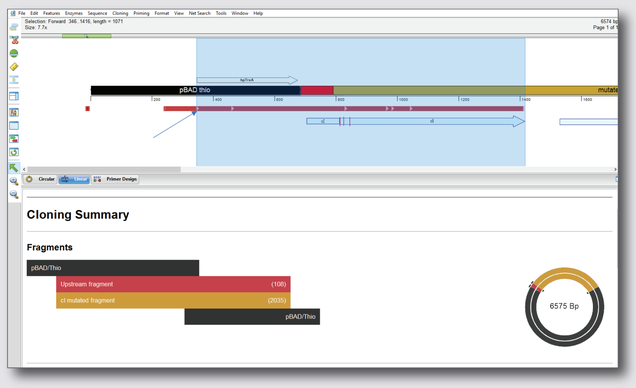
CREATE & MUTATE PRIMERS

Mutated bases and amino acids are colored pink and bases are displayed in lower case in the primer list view. Both primers can be further adjusted to improve Tm differences and minimize primer-dimer formation. Primer statistics are updated instantly in header.



CREATE VIRTUAL CLONES

The mutated cI gene within the lambdap50 operon and the upstream (non-mutated) fragment of the cI gene have been cloned into an *E.coli* expression vector system (pBAM/Thio) using Gibson Assembly. The uninterrupted ORF (arrow) points to the atg start site of hpTrxA, a mutated version of the *E. coli* protein thioredoxin (trxA) used to create a metal-binding domain that allows purification of thioredoxin fusion on metal- chelating resins.



www.dnastar.com

Research reported in this publication was supported by the National Human Genome Research Institute of the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R44GM119985. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.