Enzyme Design Primer
Overview of the current standard protocol

Florian Richter
(floric@u.washinton.edu)
Baker Group

In silico Enzyme Design - *de novo* vs. redesign of existing enzymes

0. Description of minimal Active Site/Theozyme — *de novo*

1. Grafting minimal Active Site residues into protein scaffold — *de novo*

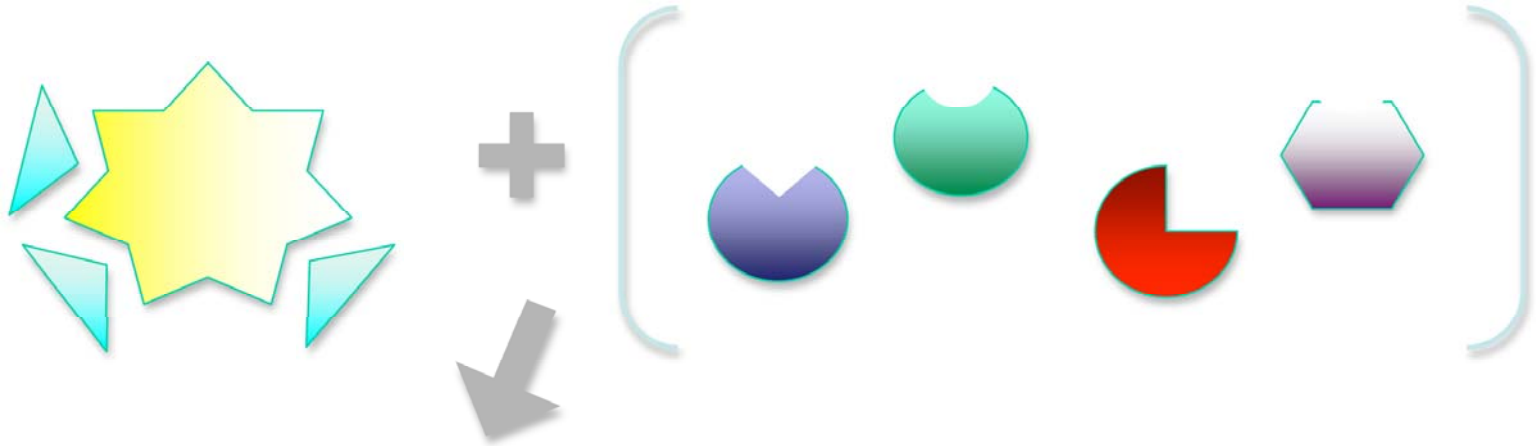
2. Designing surrounding scaffold residues for high affinity binding — *de novo / redesign*

Focus of this talk

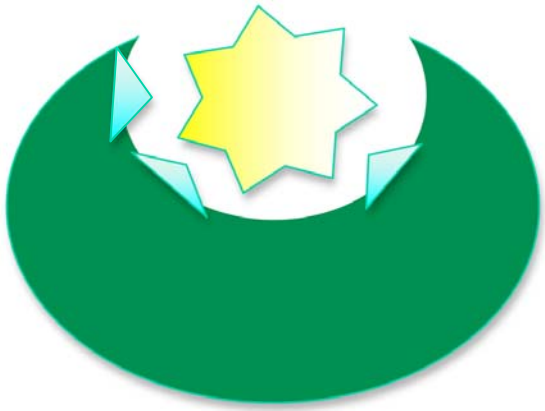
Stage 1: Grafting of theozyme (“matching”)

Minimal active site

Scaffold library



Active site placement

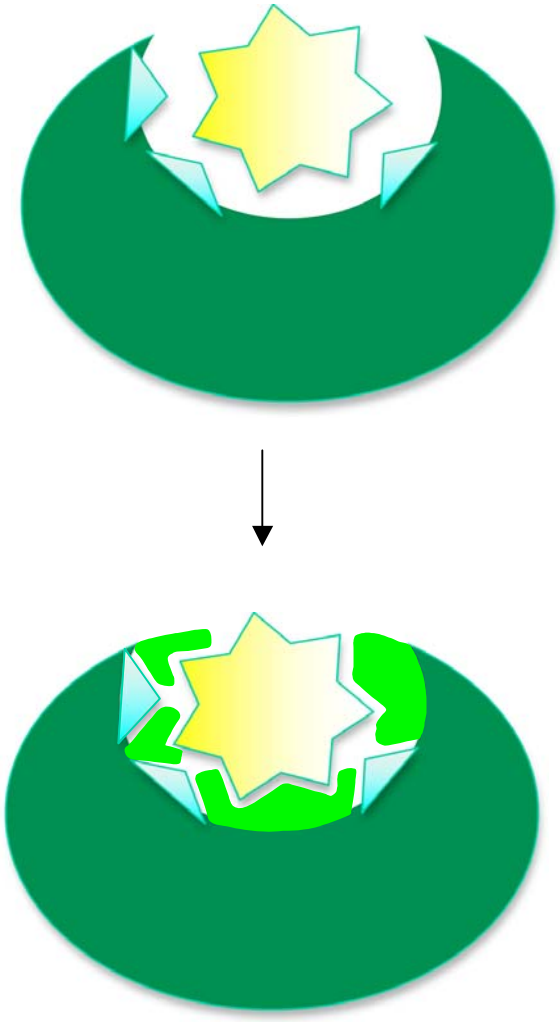


Alexandre Zanghellini

Stage 2: (Re)-Designing the active site

Starting model from matching (*de novo)* or xtal (*redesign)*

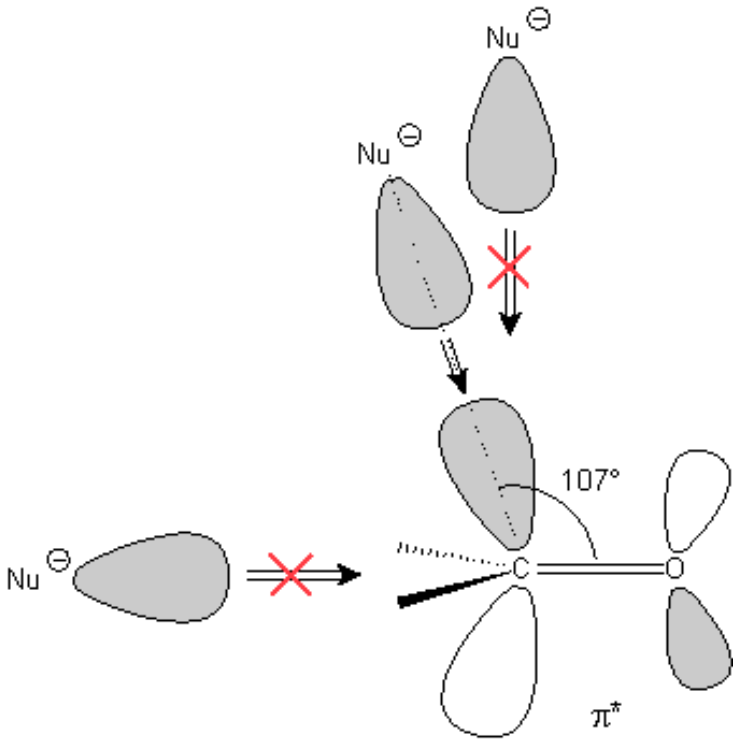
1. Optimize ligand position w/ respect to catalytic residues
2. Iterative rounds of rotamer packing/design and minimization of active site
3. Verification / Ranking of designs



Treatment of catalytic interactions I

A residue is termed catalytic if it plays a chemical role in the proposed reaction mechanism

- Rosetta employs classical empirical energy function
- No quantum terms, no bond rearrangements
- Often can't differentiate catalytic from non-catalytic conformations

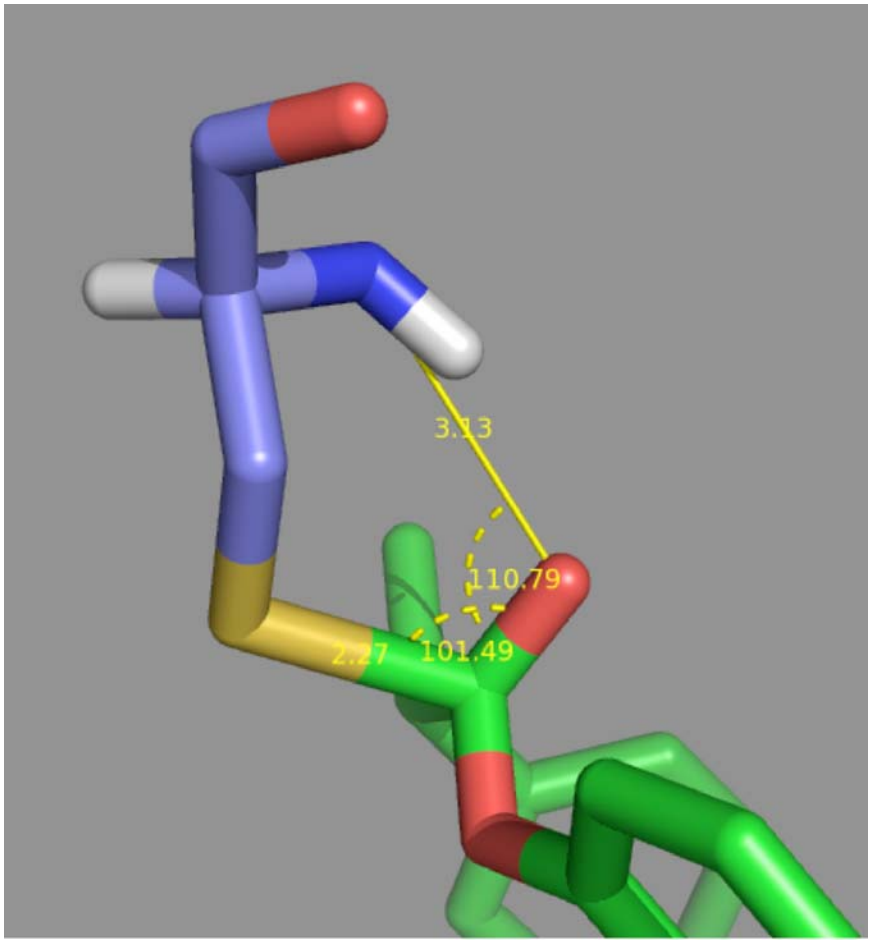


Treatment of catalytic interactions II

Workaround: use penalty functions to disfavor non-catalytic conformations

<u>For every catalytic residue:</u>

- Determine ideal catalytic geometry (i.e. distances/angles between key atoms)
- Harmonic restraining potential on these parameters (catalytic constraints/CCs)
- Possibly exclude atoms from LJ clash calculations

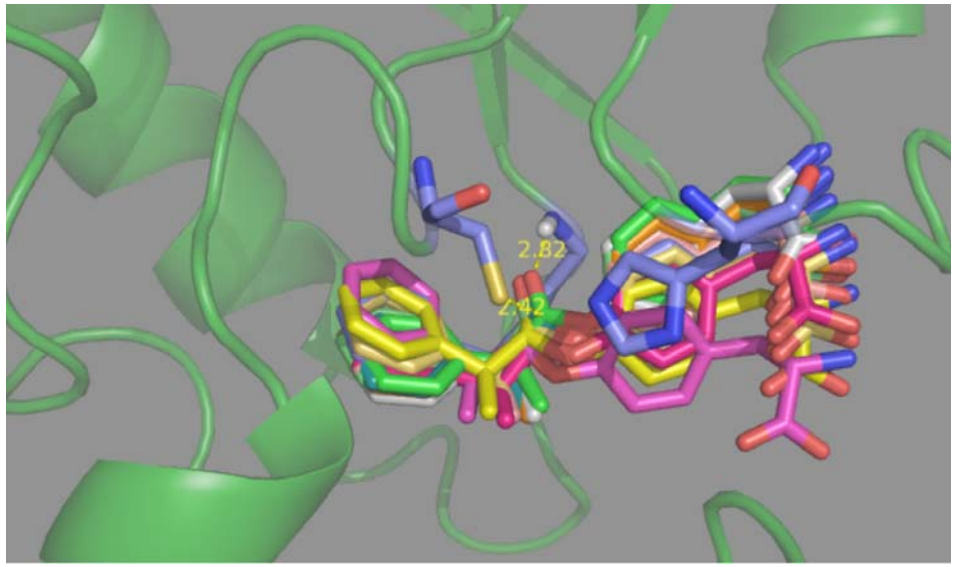


Step 1: Optimization of ligand position

Optimal starting models should have all catalytic interactions ideal, yet be diverse

<u>2 means of idealising geometries:</u>

1. Gradient based minimization of ligand with CCs
2. Random perturbations of ligand (“docking”) with CCs



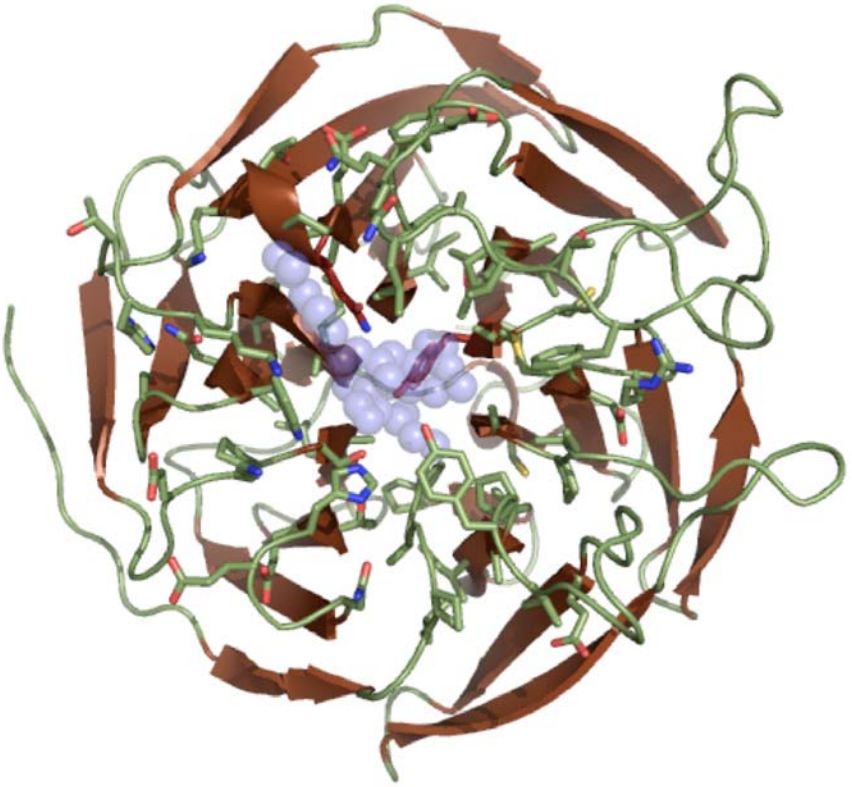
All design positions mutated to Ala at this stage

Step 2: Sequence Design

Usually 2-4 iterative rounds of:

- Sequence Design by standard Rosetta Monte Carlo algorithm
- Gradient-based minimization of ligand position and protein sidechain and backbone DOFs

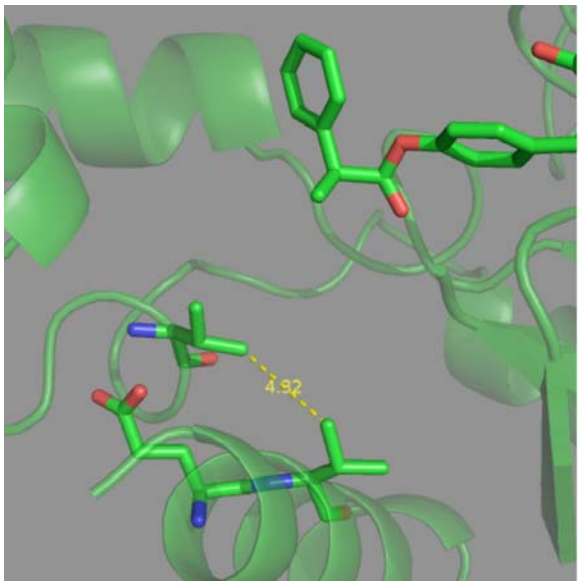
2 tricks used to facilitate good protein-ligand contacts



Movie courtesy of Justin Siegel

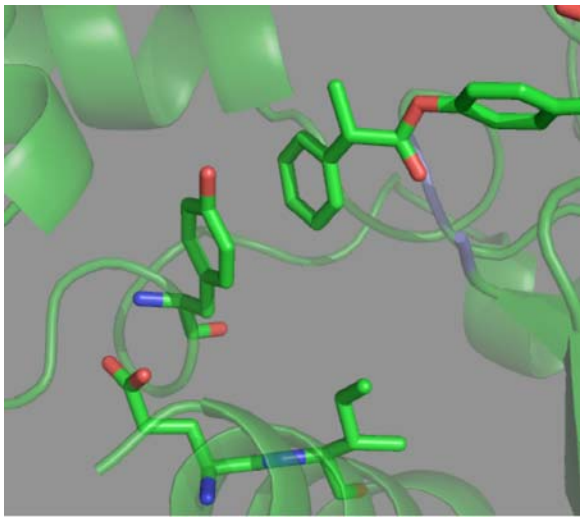
Step 2: Ensuring good ligand contacts I

•During rotamer design/ packing, ligand-protein interactions count more than protein-protein interactions



vs

•By upweighting protein-ligand interactions, the design algorithm is more likely to converge on a sequence complementary to the ligand



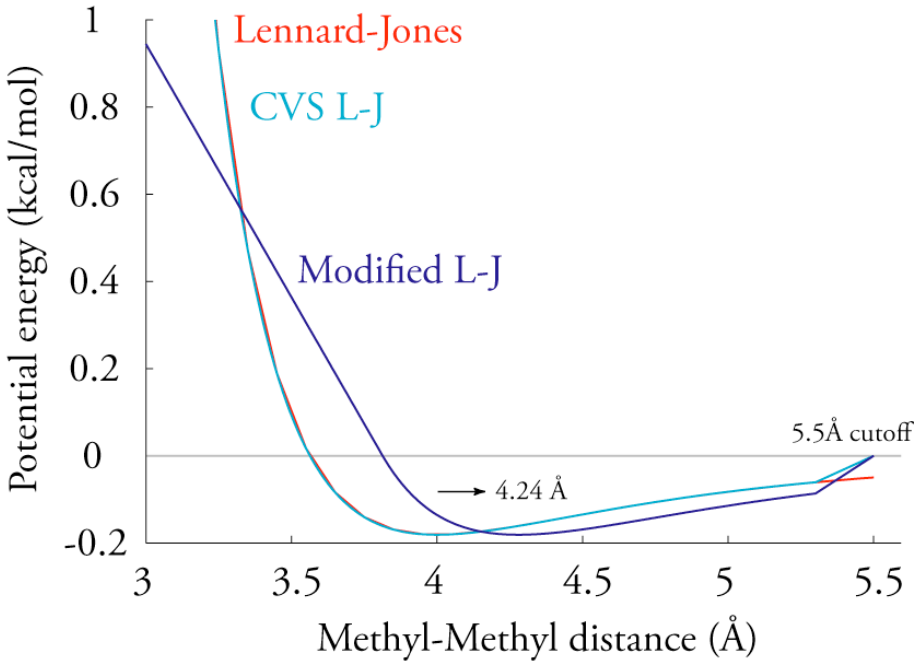
Step 2: Ensuring good ligand contacts II

Rotamer approximation leads to discretization of conf. space

Problem: small changes in rotamer-chis can make the difference between clash and tight packing

Solution:

- More rotamers (*slow*)
- "Soft-repulsive" potential, allows small overlaps between atoms



Soft-repulsive potential, courtesy of Jim Havranek

Step 3: Verification / Ranking of designs

A good design must satisfy three criteria

1. Catalysis: all CCs must have low penalties, otherwise active site residues not in competent conformation

2. Binding: ligand must have a low score, otherwise ligand unlikely to be in active site

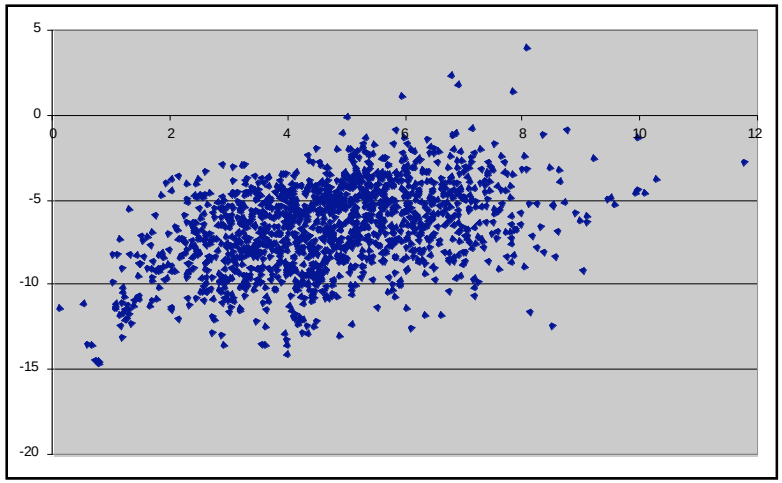
3. Scaffold integrity: the protein scaffold must not be perturbed too much, otherwise protein will not fold/express

Step 3: Verification of ligand binding I

Docking substrate into design

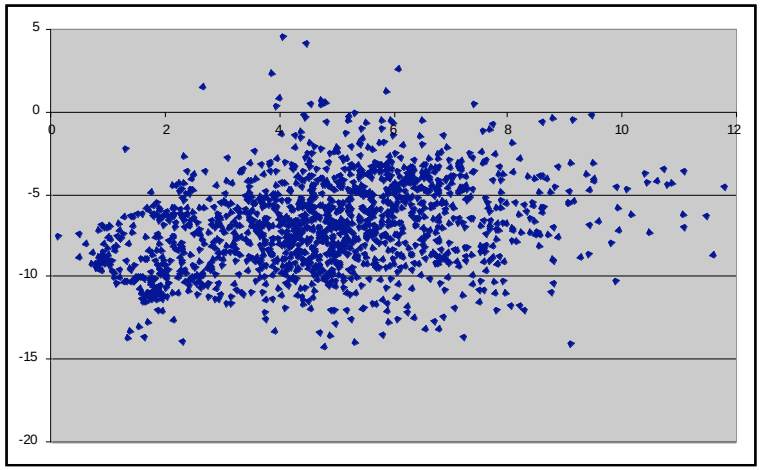
•Designs should be self-consistent

use docking to test if designed pose is lowest in score



•Designs should be high affinity

docking results should show deep funnel

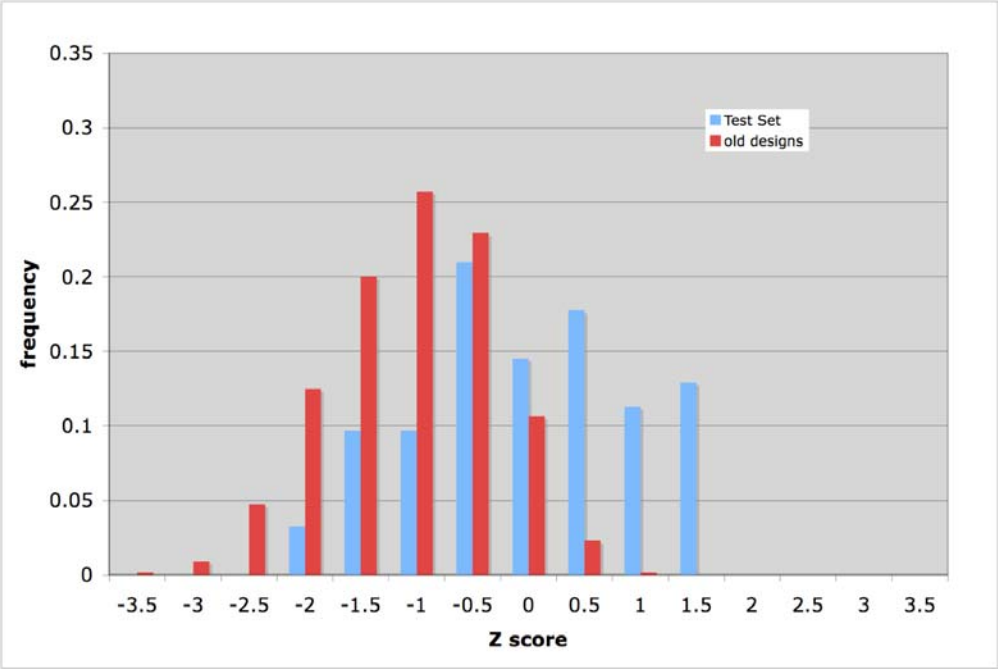


Drawback: docking into every designed sequence very expensive

Step 3: Verification of ligand binding II

Cheap alternative to docking every design: ligand Z-score

- Dock ligand against ~10-100 random proteins
- Calculate mean+SD of ligand score
- Only accept designs that have a Z-score of -2 SD or lower



Indirect assessment of binding site quality

Step 3: Verification of scaffold integrity I

Philosophy: ensure similarity to native starting scaffold

We've heard: rosetta score not always correlates with stability, solubility and expressability

Several other metrics evaluated (somewhat *ad-hoc*):

- Rosetta packstat (design tight enough?)
- # buried unsatisfied polar atoms (everything matched?)
- Solubility score (greasy patches introduced?)
- # Tertiary contacts (structurally critical interactions kept?)

Step 3: Verification of scaffold integrity II

How are good values for the additional metrics estimated?

→ no absolute cutoffs exist

But: design usually based on well-behaved starting structure

- 1st approximation: compared to the wt scaffold, every design must have:
  - No more than 5 additional buried unsatisfied polars
  - No more than 5 tertiary contacts lost
  - Packstat score difference no worse than 0.1
  - Roughly the same solubility score as the scaffold
  - Better Rosetta score

Documentation

After next release:

http://www.rosettacommons.org/manuals/rosetta3_user_guide/app_enzyme_design.html

Code - some relevant classes

General protocol:

protocols::enzdes::EnzdesBaseProtocol

protocols::enzdes::EnzdesFixBBProtocol

protocols::ligand_docking::LigandBaseProtocol

Catalytic constraints:

protocols::enzdes::EnzConstraintParameters

protocols::enzdes::EnzCstTemplateRes

protocols::enzdes::EnzConstraintIO

core::scoring::constraints::MultiConstraint

core::scoring::constraints::AmbiguousConstraint

Upweighting:

core::pack::task::IGEdgeReweightContainer

protocols::toolbox::IGEdgeReweighters

Evaluation:

protocols::ligand_docking::LigandDockProtocol

protocols::enzdes::DesignVsNativeComparison

In protocols::toolbox::PoseMetricCalculators:

NumberHBondsCalculator

BuriedUnsatisfiedPolarsCalculator

NonLocalConstactsCalculator

InterfaceDeltaEnergeticsCalculator

Thank You

Developers/testers: Baker Lab Enzdes team past and present:

- Eric Althoff
- Lin Jiang
- Daniela Grabs
- Alexandre Zanghellini
- Ling Wang
- Jennifer Bui
- Kui Chan
- Sagar Khare
- Andrew Wollacott
- Justin Siegel
- Sinisa Bjelic
- Lucas Nivon
- Paul Murphy
- Matthew Smith/Austin Day

Other contributors to Rosetta Enzdes:

Ian Davis, Andrew Leaver-Fay, Possu Huang, Andrew Ban

David Baker