

Πρωτεϊνική Μηχανική

Υπολογιστικός πρωτεϊνικός σχεδιασμός

Ιωάννης Παυλίδης
Τμήμα Χημείας | Πανεπιστήμιο Κρήτης

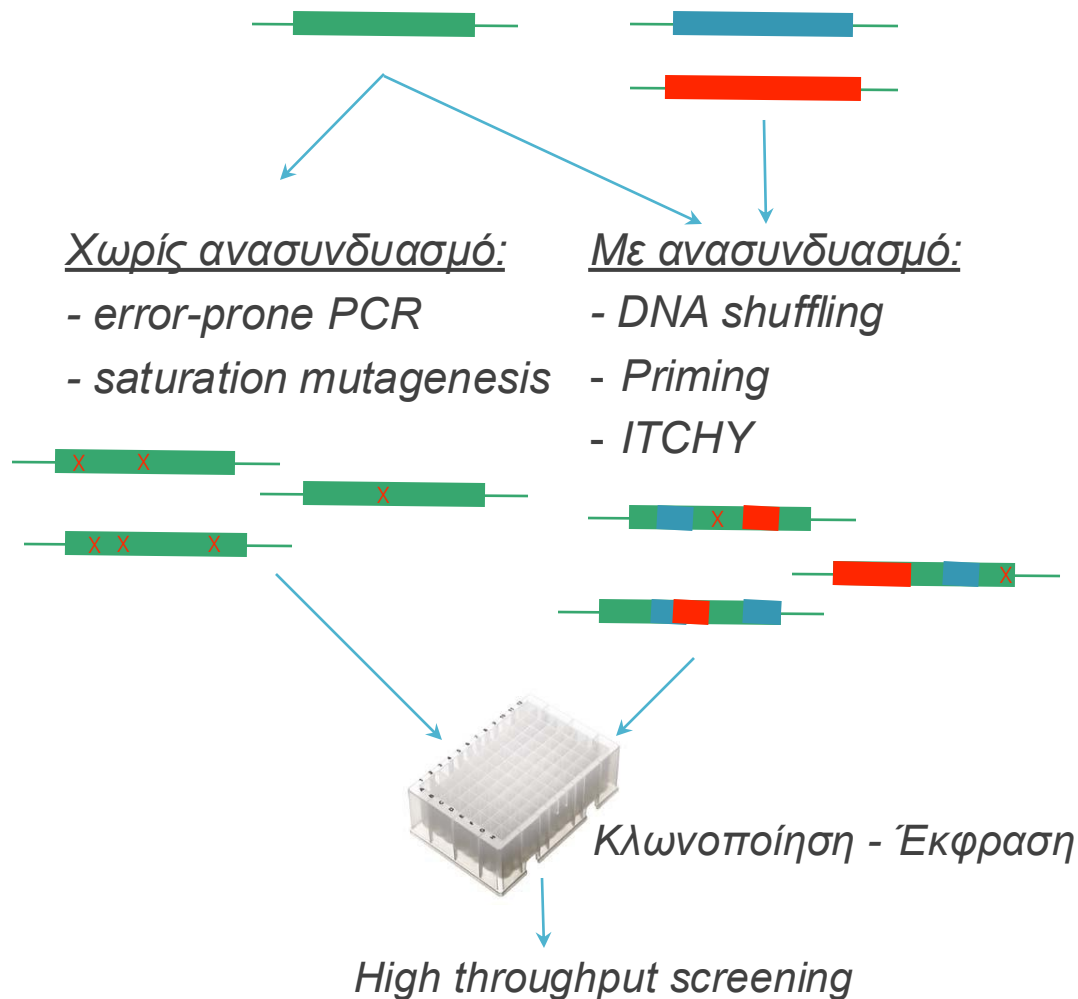
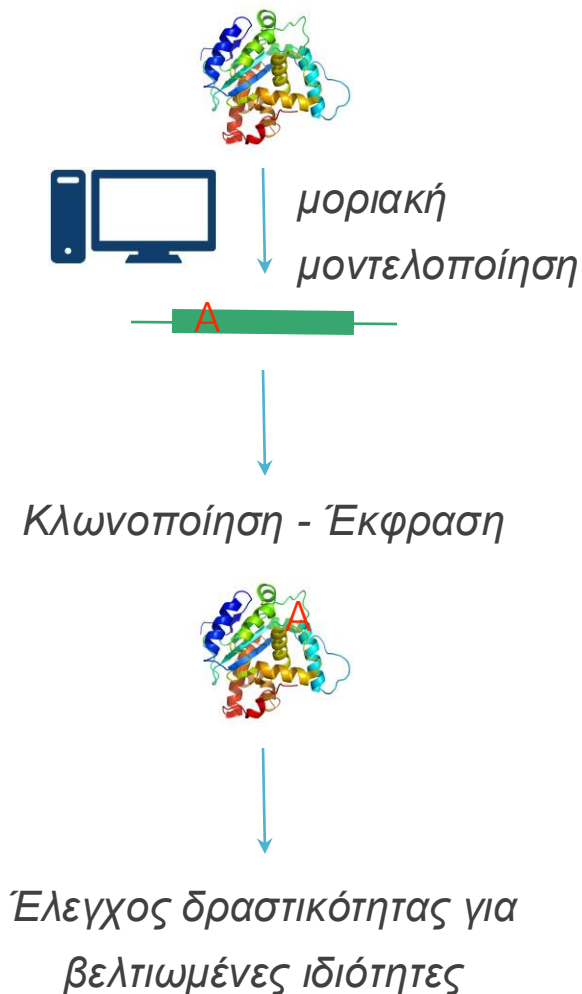
E-mail: ipavlidis@uoc.gr

Τηλ.: +30 2810 54-5130 | Γραφείο Γ211

ΕΠΙΣΗΜΗ ΣΕΛΙΔΑ: [HTTP://WWW.CHEMISTRY.UOC.GR/PAVLIDIS/](http://www.chemistry.uoc.gr/pavlidis/)

Ορθολογικός σχεδιασμός

Κατευθυνόμενη μεταλλαξιγένεση



Το πρόβλημα της πρόβλεψης

Η σύνδεση του DNA με την πρωτοταγή δομή (αλληλουχία αμινοξέων) έχει λυθεί

Η σύνδεση της πρωτοταγούς με την τεταρτοταγή δομή είναι ακόμα ένα σημαντικό πρόβλημα

Δύσκολη η ακριβής πρόβλεψη όλων των ασθενών αλληλεπιδράσεων (alpha-fold)

Computational design:

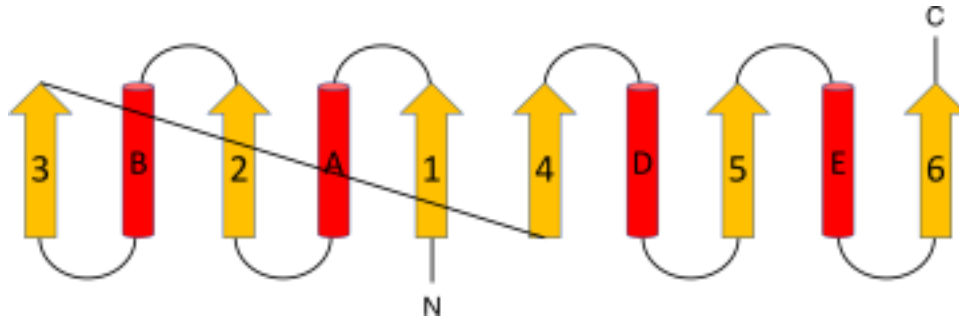
Σύνδεση της αλληλουχίας με ένα συγκεκριμένο τύπο αναδίπλωσης και σχετική λειτουργικότητα.

Αλλαγή ορισμένων (λίγων) αμινοξέων για την επίτευξη του επιθυμητού φαινοτύπου

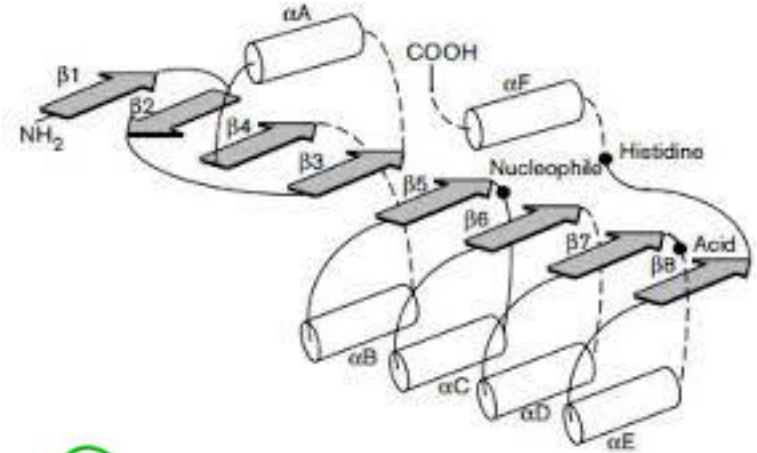
De novo design:

Σχεδιασμός νέας αλληλουχίας για πιθανές νέες αντιδράσεις

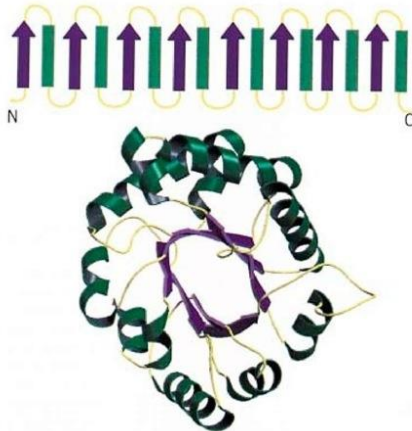
Δομικά μοτίβα



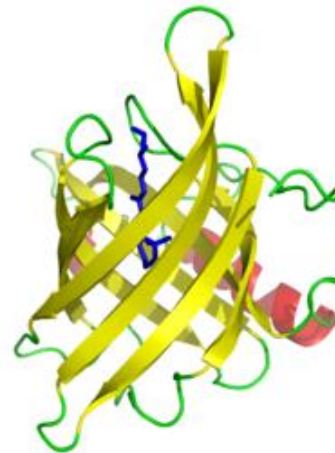
Rossmann fold



α/β -hydrolase fold



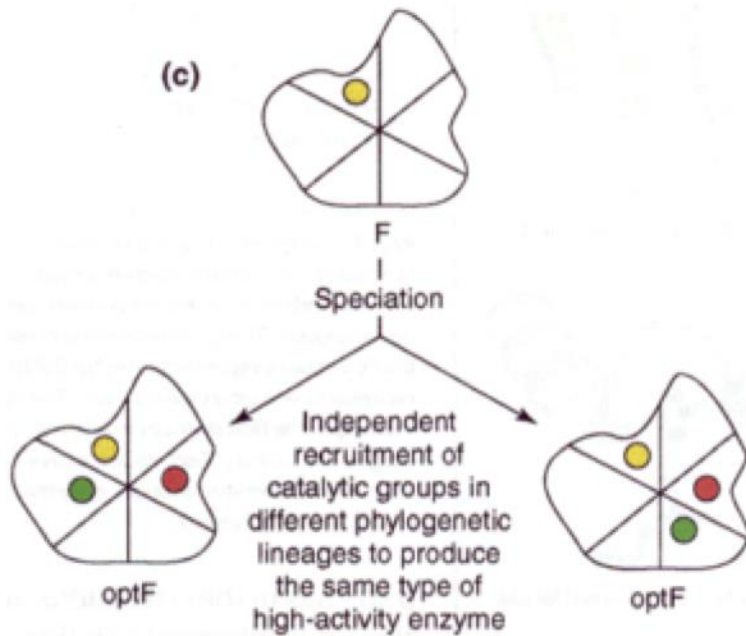
TIM-barrel



β -barrel

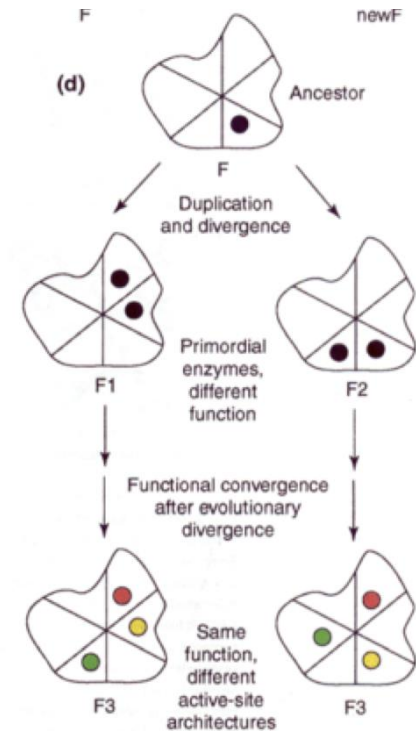
Εξέλιξη πρωτεϊνών

Διπλασιασμός και διαφοροποίηση



Ίδια αναδίπλωση
καταλύουν διαφορετικές αντιδράσεις
(π.χ. εστεράσες, αμιδάσες)

Συγκλίνουσα εξέλιξη



Διαφορετική αναδίπλωση
καταλύουν την ίδια αντίδραση
(π.χ. τρανσαμινάσες)

Βασικά στοιχεία υπολογιστικού σχεδιασμού πρωτεϊνών

1. Εύρεση δομής-στόχου που να κατέχει ευελιξία

(δομή υψηλής ανάλυσης - αν δεν είναι διαθέσιμη, μοντέλο ομολογίας)

2. Βαθμοί ελευθερίας

α. η αμινοξική αλληλουχία

β. η διαμόρφωση των αμινοξέων (γνωρίζουμε τη διευθέτηση;)

3. Συναρτήσεις ενέργειας (Amber, Gromos κτλ)

Ποσοτικοποίηση των φυσικοχημικών αλληλεπιδράσεων

Εύρεση της δομής με τη χαμηλότερη ενέργεια

4. Διαλυτοποίηση και σχηματοποίηση

Υδρόφοβες / υδρόφιλες αλληλεπιδράσεις

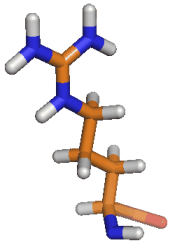
Συνήθως εξωτερικά υδρόφιλα αμινοξέα και εσωτερικά τα υδρόφοβα

5. Μέθοδοι αναζήτησης

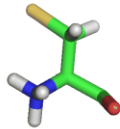
Αλγόριθμοι σχέσης αλληλουχίας / δομής - π.χ. Monte Carlo

Σύγκριση γνωστών δομών

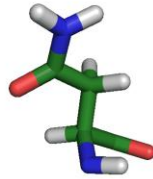
Rotamers



ARG



CYS



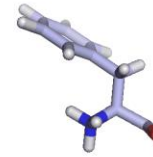
ASN



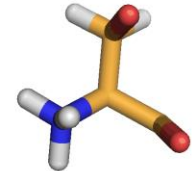
ILE



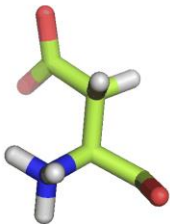
LEU



PHE



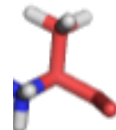
SER



ASP



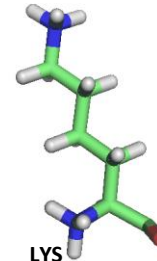
GLU



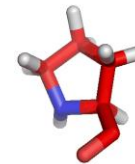
ALA



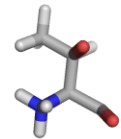
HIS



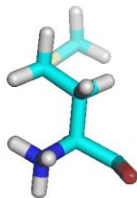
LYS



PRO



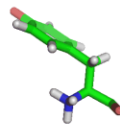
THR



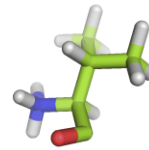
MET



TRP



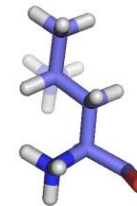
TYR



VAL



GLY



GLN

Επαναπρογραμματισμός πρωτεϊνών

Κύριοι στόχοι:

- ❖ Επίδραση στην δραστικότητα (ταχύτητα, εκλεκτικότητα)
- ❖ Επίδραση στην σταθερότητα
- ❖ Κατάλυση νέων αντιδράσεων / νέων υποστρωμάτων

Ανάγκη για:

- ❖ Δομή υψηλής ανάλυσης
- ❖ Κατανόηση του καταλυτικού μηχανισμού

Αρχές του σχεδιασμού πρωτεϊνών

Τα λογισμικά συνήθως έχουν δύο πακέτα:

- τις **συναρτήσεις ενέργειας**
- τις **μεθόδους εύρεσης**, για βελτιστοποίηση

Οι **συναρτήσεις ενέργειας (force fields)** εμπεριέχουν φυσικούς και εμπειρικούς όρους.

Πραγματικό πρόβλημα:

“determine the amino acid sequence that achieves the lowest value of the cost-function.”

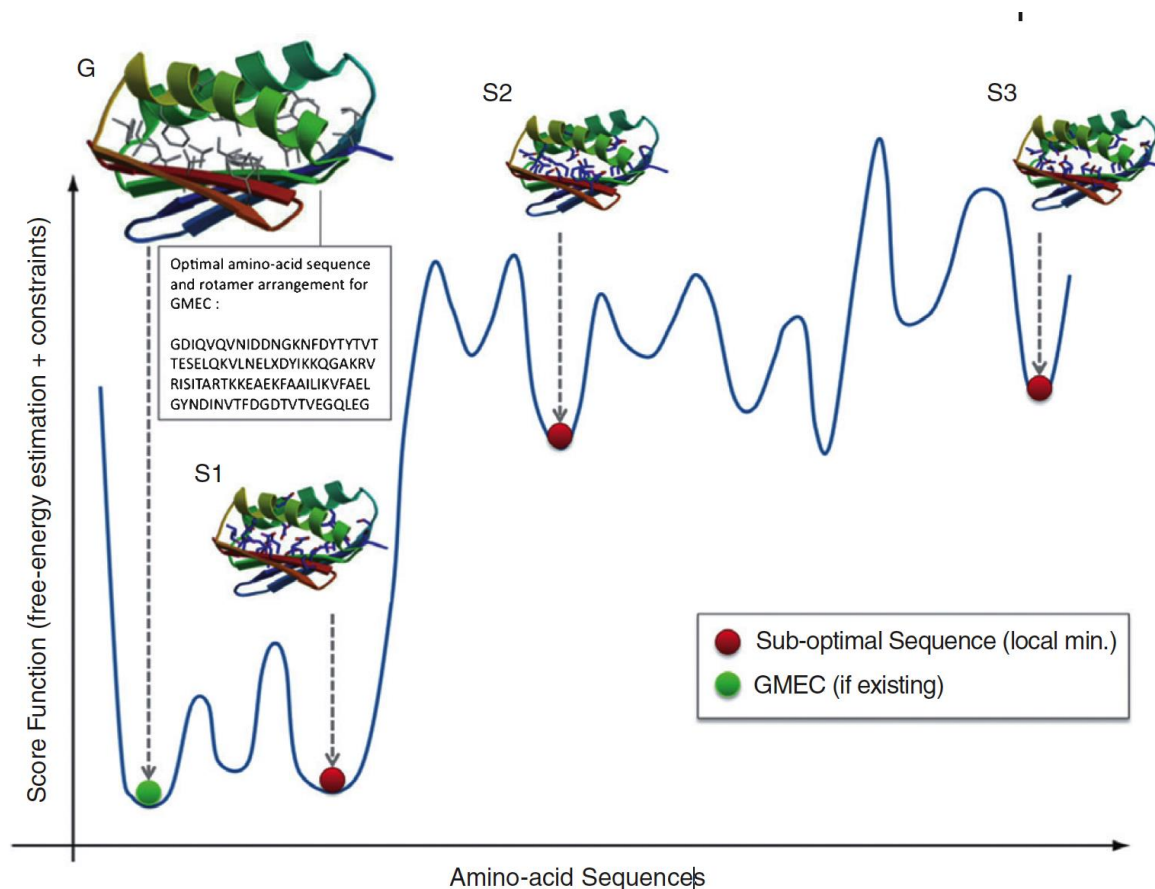
διαμορφώσεις σκελετού και πλευρικών αλυσίδων χρειάζεται να υπολογιστούν

Υπεραπλούστευση: “given the spatial coordinates of the backbone, determine the amino acid sequence that achieves the lowest value of the cost-function.”

μόνο οι διαμορφώσεις πλευρικών αλυσίδων χρειάζεται να υπολογιστούν

Μεγάλο πρόβλημα η συνδυαστική βελτίωση καθώς χρειάζεται αρκετά περίπλοκους υπολογισμούς και η πολυπλοκότητα αυξάνεται εκθετικά με τον αριθμό αμινοξέων.

Επίλυση με σταθερό σκελετό



Στόχος είναι να βρεθεί η *global minimum energy conformation* (GMEC, πράσινο σημείο).

Συνήθως υπάρχουν αρκετά τοπικά ελάχιστα που οι αλγόριθμοι μπορεί να σταματήσουν.

Γιατί παγιδεύονται σε τοπικά ελάχιστα;

Η αρχική δομή παίζει σημαντικό ρόλο

Παράδειγμα γλουταμίνης:

Μπορεί να έχει ανάστροφη διευθέτηση

Είναι πακεταρισμένη ανάμεσα σε άλλα αμινοξέα

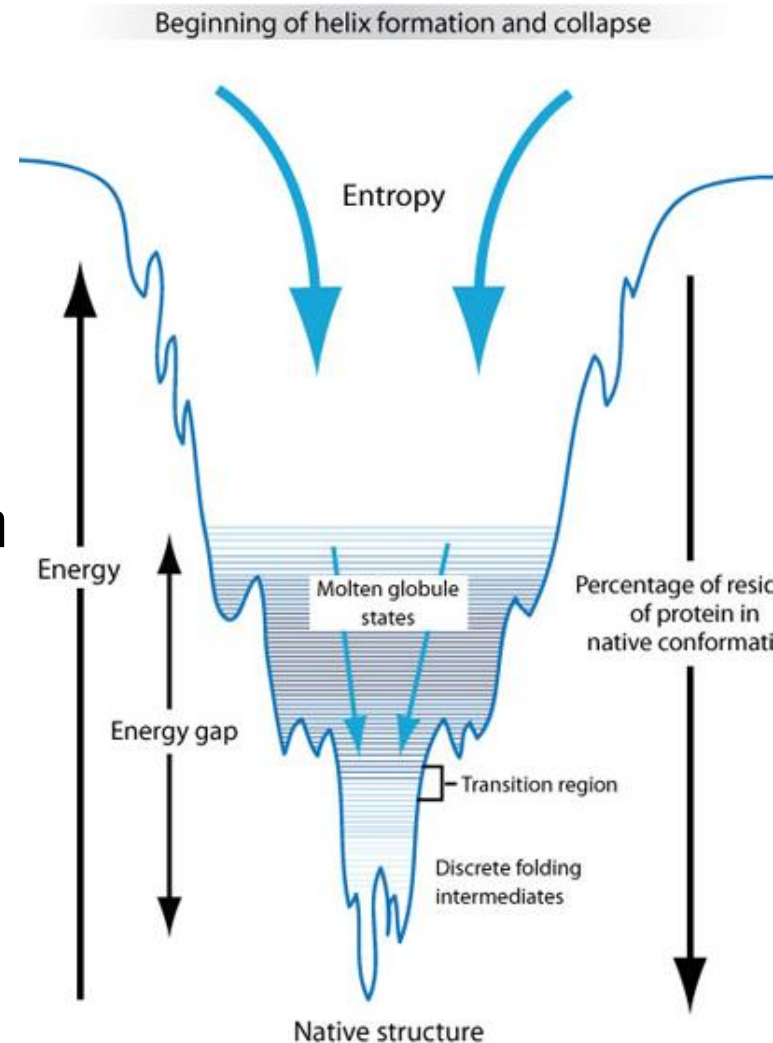
Για να στραφεί 180° έχει μεγάλη ενεργειακή ποινή

Δύσκολο να ξεπεράσει αυτή την ποινή και έτσι

εγκλωβίζεται σε αυτές τις δομές

Πρέπει να δοθούν περισσότεροι βαθμοί

ελευθερίας για να το αποφύγει



Loop grafting

Ευέλικτες περιοχές, μη συντηρημένες, που μπορεί να παίζουν σημαντικό ρόλο

- Στερικές παρεμποδίσεις στο ενεργό κέντρο
- Διάχυση υποστρωμάτων
- Σταθερότητα δομής

Έτοιμοι βρόγχοι από την PDB μπορούν να δοκιμαστούν στην δομή που έχουμε:

- Οι βέλτιστες επιλέγονται
- Γίνεται ελαχιστοποίηση ενέργειας με τα γειτονικά αμινοξέα

Προβλήματα:

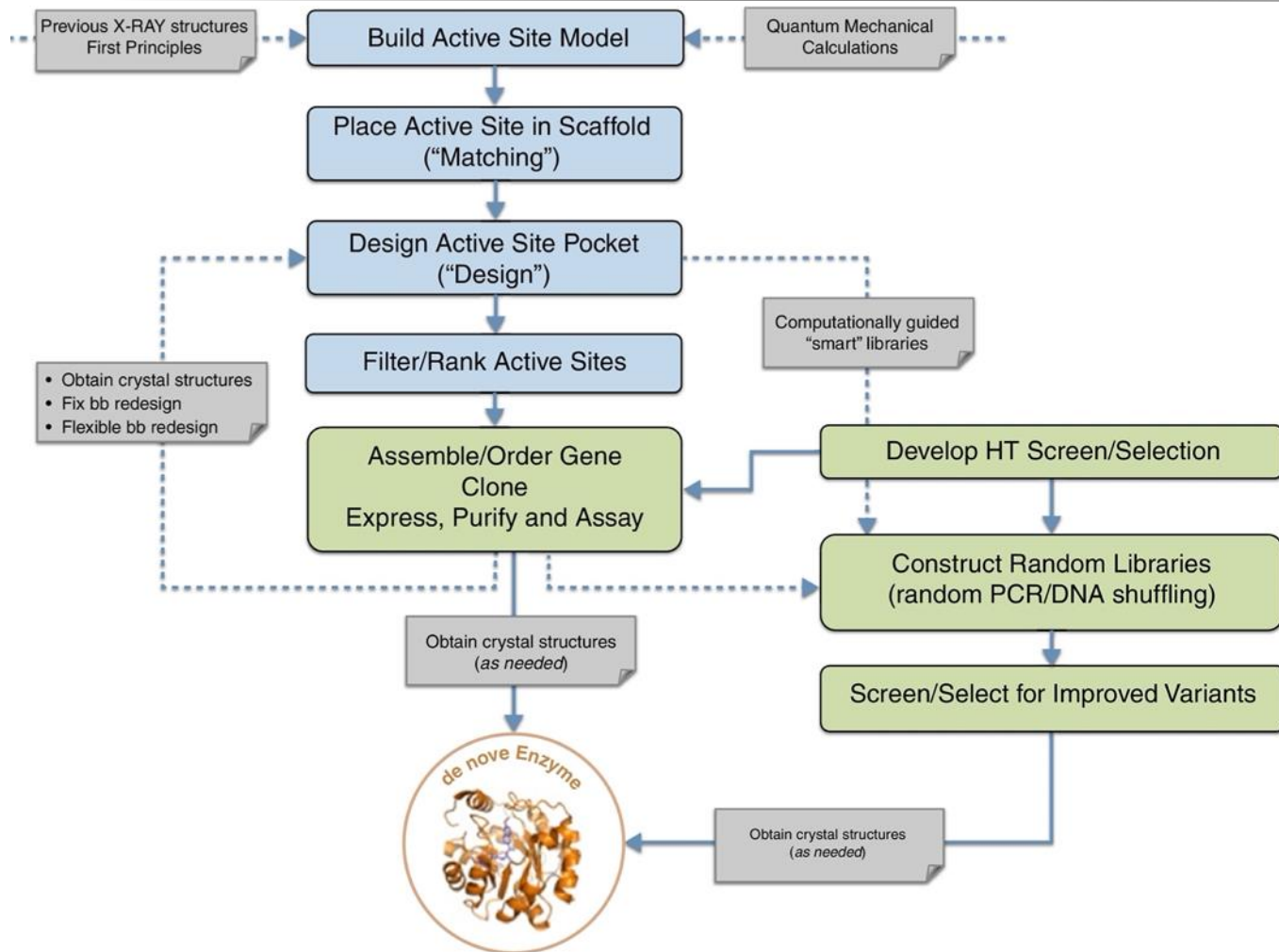
1. Δεν υπολογίζεται αν ο βρόγχος μπορεί να ταιριάξει στη συγκεκριμένη δομή
2. Λαμβάνεται ως δεδομένη η αναδίπλωση της πρωτεΐνης, το οποίο δεν είναι απαραίτητο ότι ισχύει με την εισαγωγή ενός ξένου βρόγχου

Δημιουργία ενζύμων από το μηδέν μέσω υπολογιστικών μεθόδων

Βασικές αρχές:

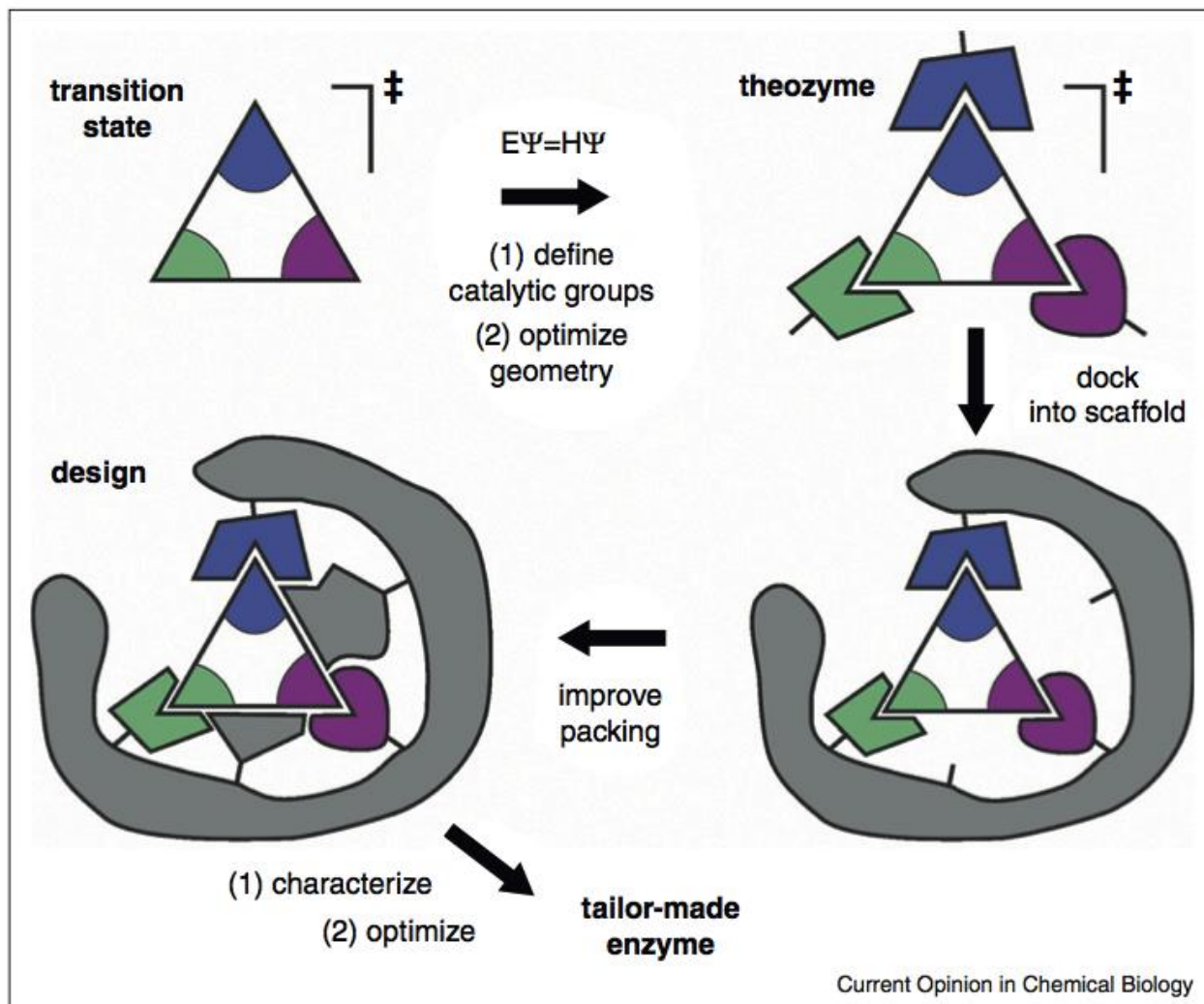
1. Καθορισμός αντίδρασης ενδιαφέροντος
2. Καθορισμός καταλυτικού μηχανισμού για την αντίδραση ενδιαφέροντος
3. Σχεδιασμός ιδεατού πιθανού ενεργού κέντρου, με βελτιστοποίηση της θέσης των καταλυτικών αμινοξέων για σταθεροποίηση της μεταβατικής κατάστασης
4. Αυτό δημιουργεί μία δομή “θεωζύμου (**theozyme** = theoretical enzyme)”
5. Ταίριασμα (docking) του θεωζύμου σε διάφορες πρωτεϊνικές δομές
6. Ανάλυση / φιλτράρισμα των καλύτερων αποτελεσμάτων και βελτίωση με σημειακές μεταλλάξεις
7. Παραγωγή πρωτεΐνης (συνθετικό γονίδιο), καθαρισμός, βιοχημικός χαρακτηρισμός
8. Βελτίωση του σχεδιασμού μέσω επιπλέον ανάλυσης και τυχαίας μεταλλαξιγένεσης

de novo design of enzymes

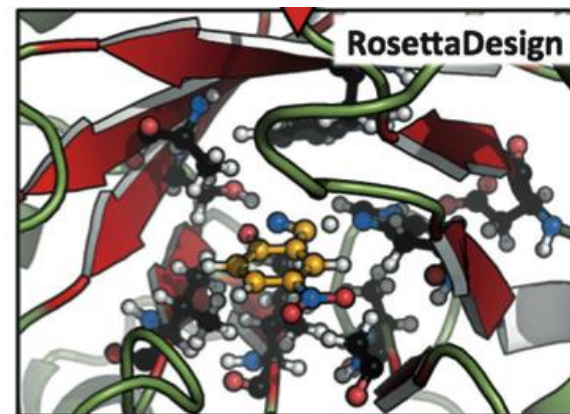
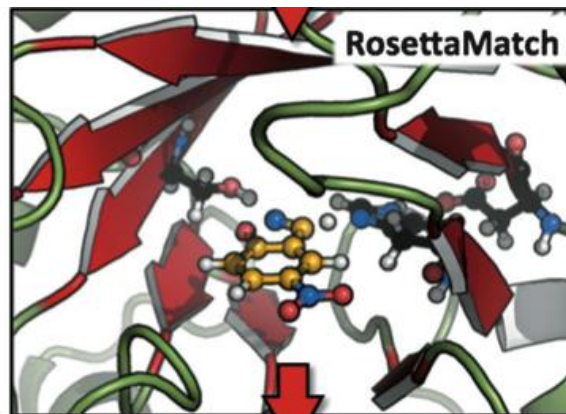
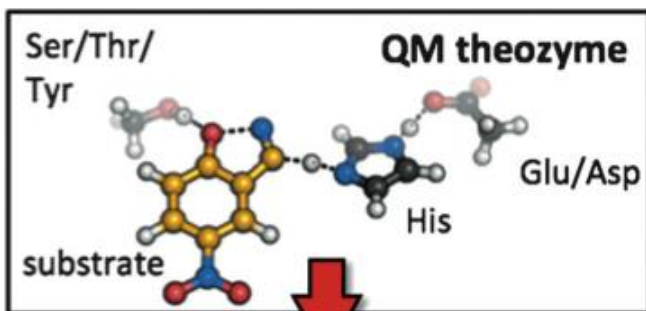


Zanghellini (2014) *Curr. Opin. Biotechnol.* **29**: 132-138

Σχηματική απεικόνιση



Προσέγγιση “Inside-out”



Baker και Houk: Kemp eliminase

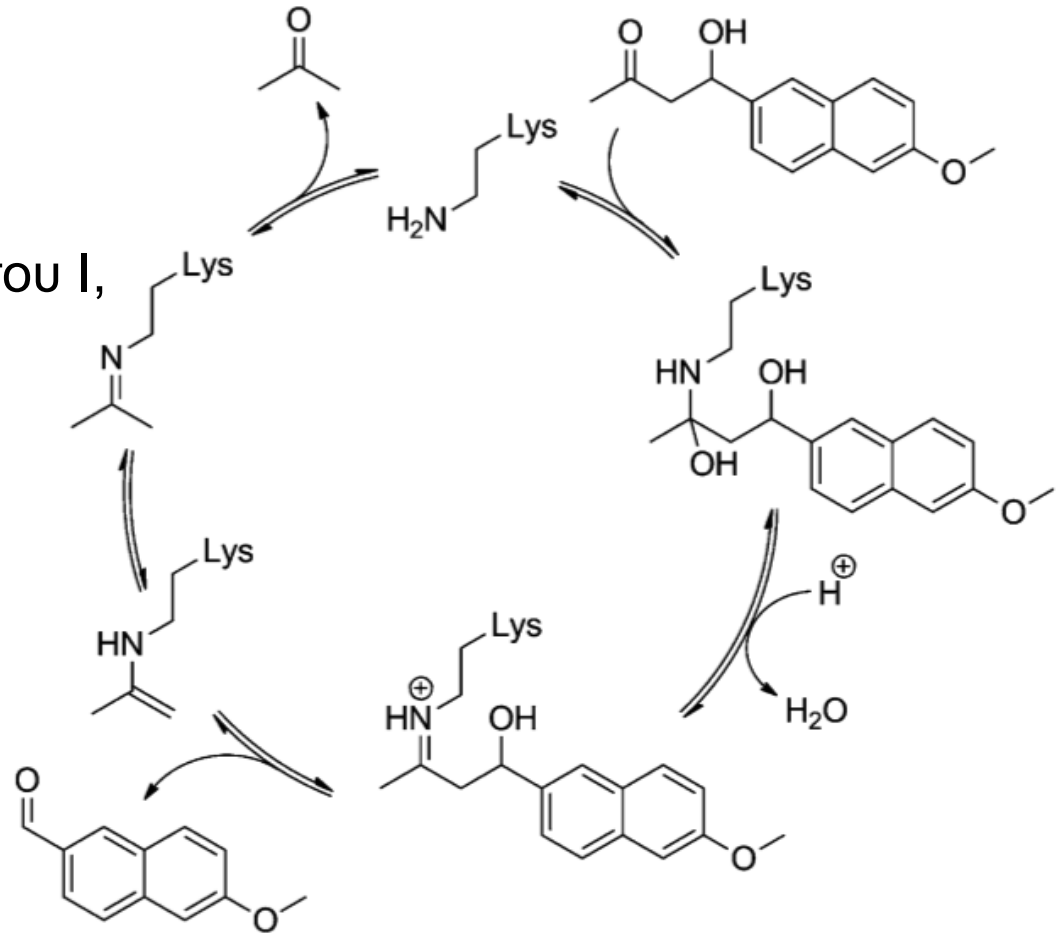
Από το QM θεώζυμο, στο πρωτεϊνικό φορέα, και μετά στη βελτιστοποίηση

Rosetta Match και **SABER** (Selection of Active/Binding sites for Enzyme Redesign) για την επιλογή πρωτεϊνικής μήτρας

RosettaDesign χρησιμοποιείται για τη βελτιστοποίηση των αμινοξέων γύρω από τα καταλυτικά

Παράδειγμα της ρетро-αλδολάσης

Η ρέτροαλδολάση
καταλύει τη λύση δεσμών C-C,
παρόμοια με τις αλδολάσες τύπου I,
Ο μηχανισμός βασίζεται
σε ένα πυρηνόφιλο κατάλοιπο
λυσίνης

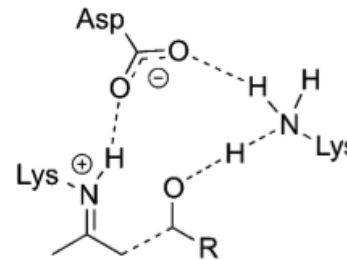


Παράδειγμα της ρетро-αλδολάσης

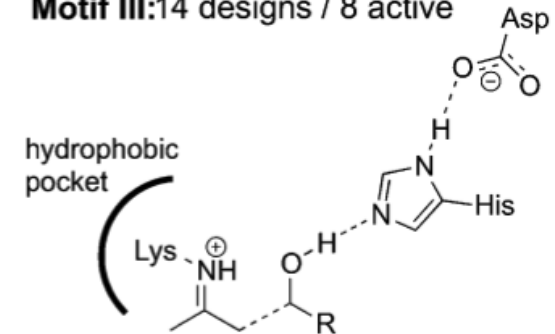
4 διαφορετικά θεώζυμα
η λυσίνη ως βάση κατά Schiff
και ως γενικό οξύ/βάση

- I: Lys/Asp дуάδα
- II: Tyr
- III: His/Asp дуάδα
- IV: H₂O

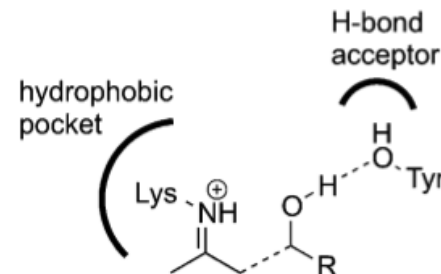
Motif I: 12 designs / 0 active



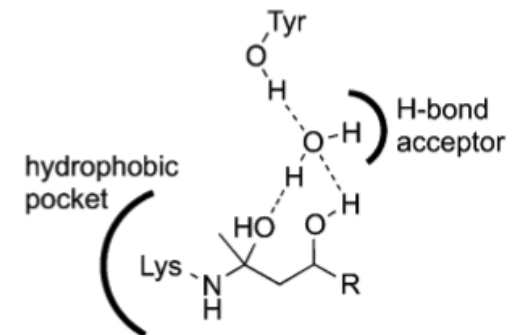
Motif III: 14 designs / 8 active



Motif II: 10 designs / 0 active



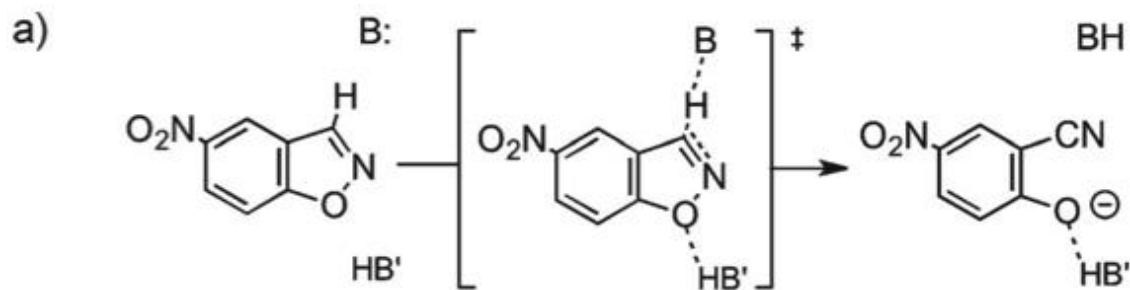
Motif IV: 38 designs / 22 active



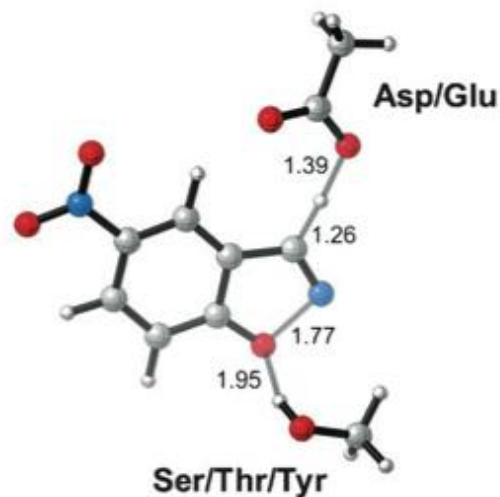
Παράδειγμα της ρетро-αλδολάσης

- Βελτιστοποίηση (refinement) της μεταβατικής κατάστασης
- Προσανατολισμός πλευρικών αλυσίδων σε σχέση με την μεταβατική δομή
- Διαμόρφωση καταλυτικών αμινοξέων και των πλευρικών τους αλυσίδων
- ~ 1.000 γεωμετρίες του ενεργού κέντρου υπολογίστηκαν για κάθε μοτίβο
- 71 πρωτεϊνικές μήτρες, υπολογιστικά βρέθηκαν 180.000 ταιριάσματα
- 72 σχέδια σε 10 διαφορετικές μήτρες μελετήθηκαν πειραματικά
- 70 μπόρεσαν να εκφραστούν σε διαλυτή μορφή
- 32 είχαν ενεργότητα σε μετρήσεις φθορισμού
- Οι 3D δομές των καλύτερων ενζύμων αναλύθηκαν
- Χαμηλή καταλυτική δραστητικότητα (k_{cat}/K_M 0.02-0.74 M⁻¹s⁻¹)

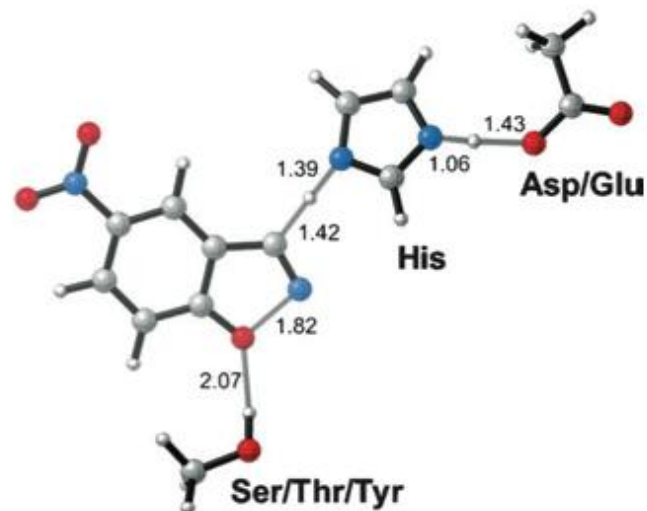
Παράδειγμα της Kemp-eliminase



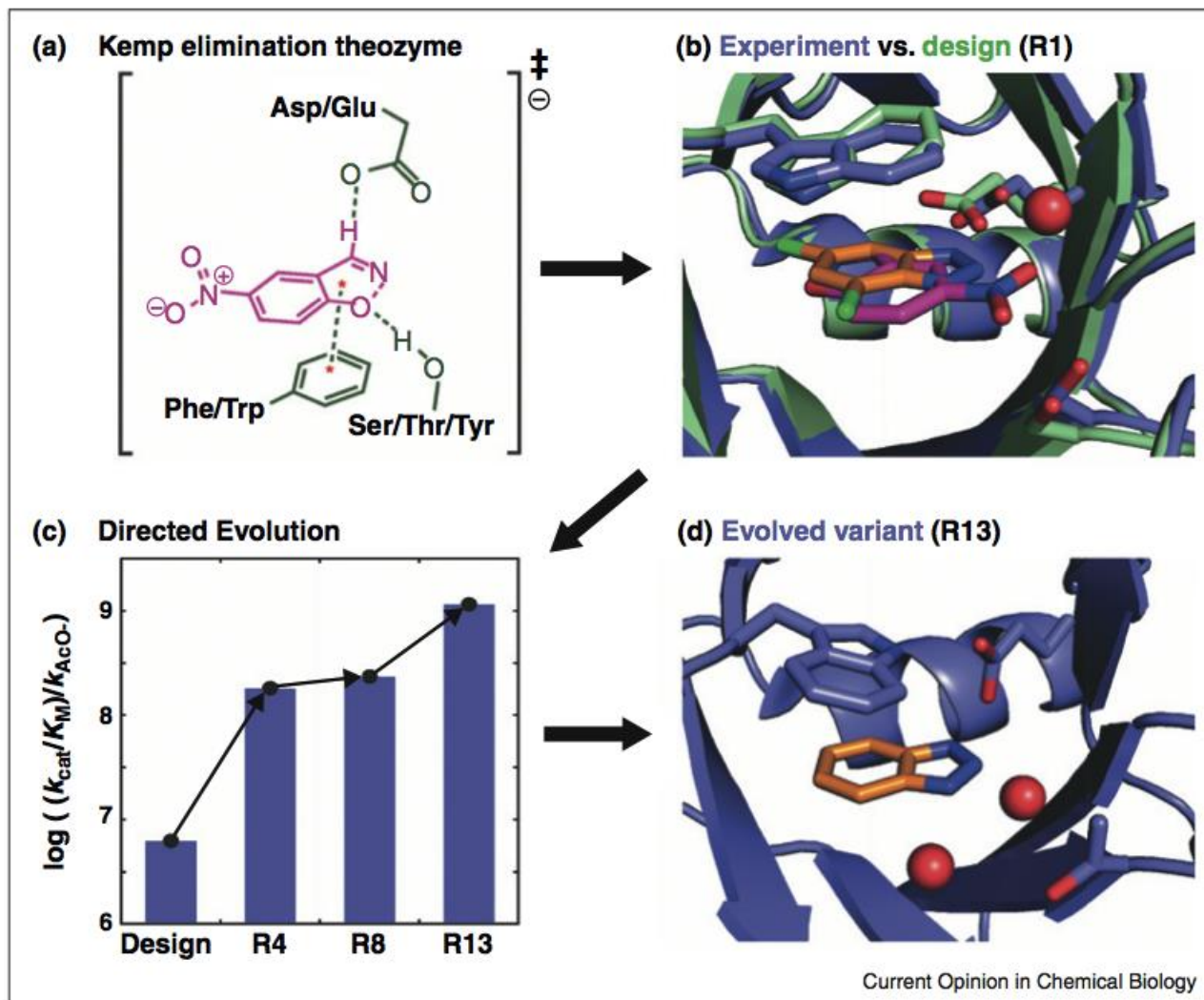
b) **theozyme I**
39 expressed / 6 active



theozyme II
20 expressed / 2 active



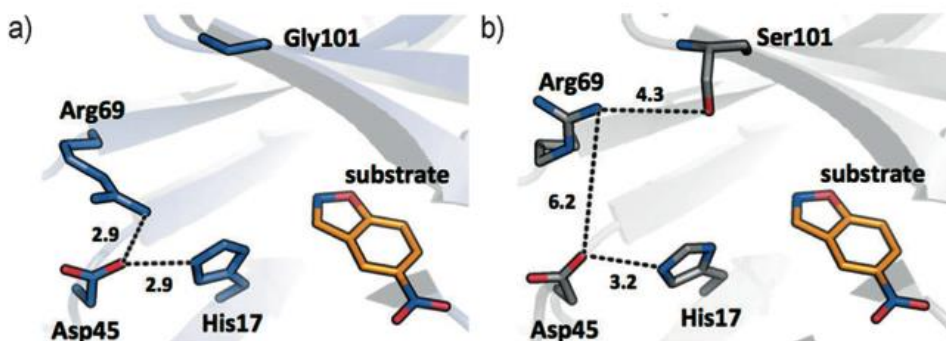
Παράδειγμα της Kemp-eliminase



Παράδειγμα της Kemp-eliminase

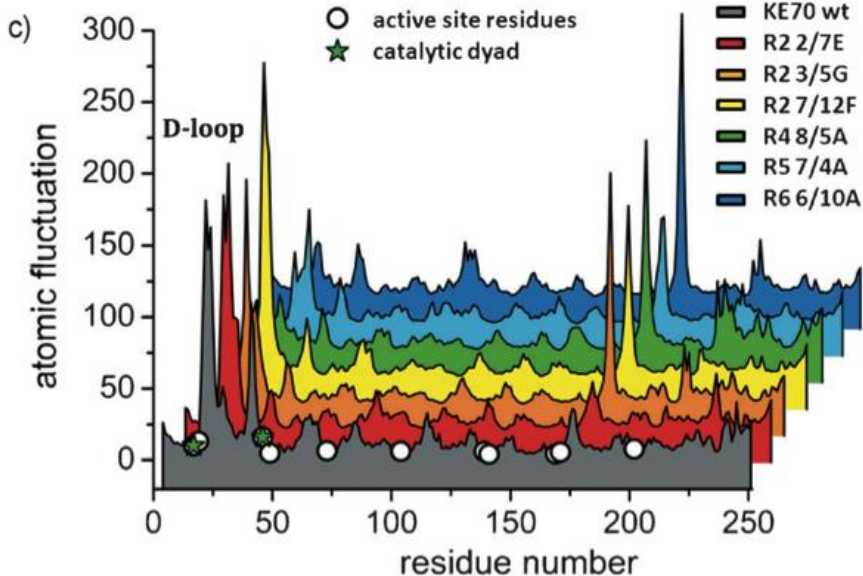
- Τα πρώτα σχέδια μέσω του **Rosetta Match** έδωσαν 8 ενεργά ένζυμα από 59 σχέδια που εκφράστηκαν, καλύτερο k_{cat}/K_M $12.2 \text{ M}^{-1}\text{s}^{-1}$
- **Κατευθυνόμενη εξέλιξη** τριών σχεδίων βελτίωσε περαιτέρω την δραστικότητα: KE07: 200 φορές, KE70: 400 φορές, **KE59: 2000 φορές**
- Περαιτέρω ανάλυση με υπολογισμούς MD-QM/MM βοήθησαν στην περαιτέρω κατανόηση της δραστικότητας

Παράδειγμα της Kemp-eliminase



a) Κρυσταλλική δομή του KE70.

b) Κρυσταλλική δομή μίας μεταλλαγμένης πρωτεΐνης (R66/10A). Η G101S σταθεροποιεί την R69 σε διαφορετική θέση και δεν επηρεάζει την καταλυτική δυάδα D45–H17.



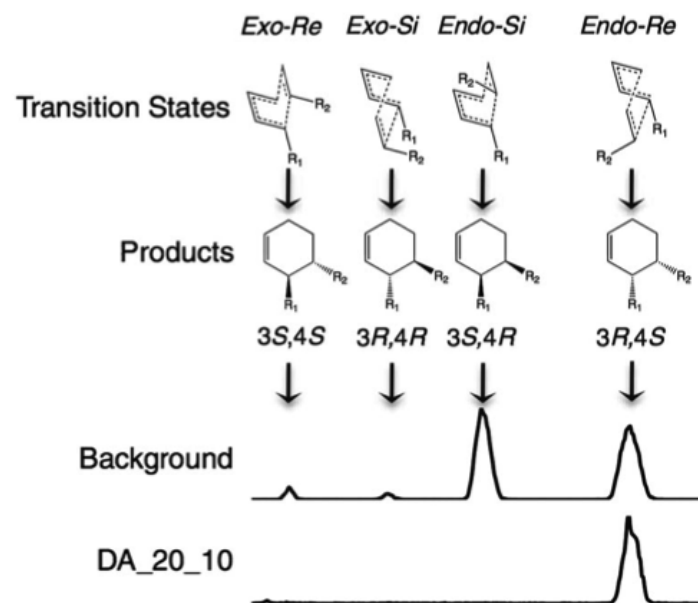
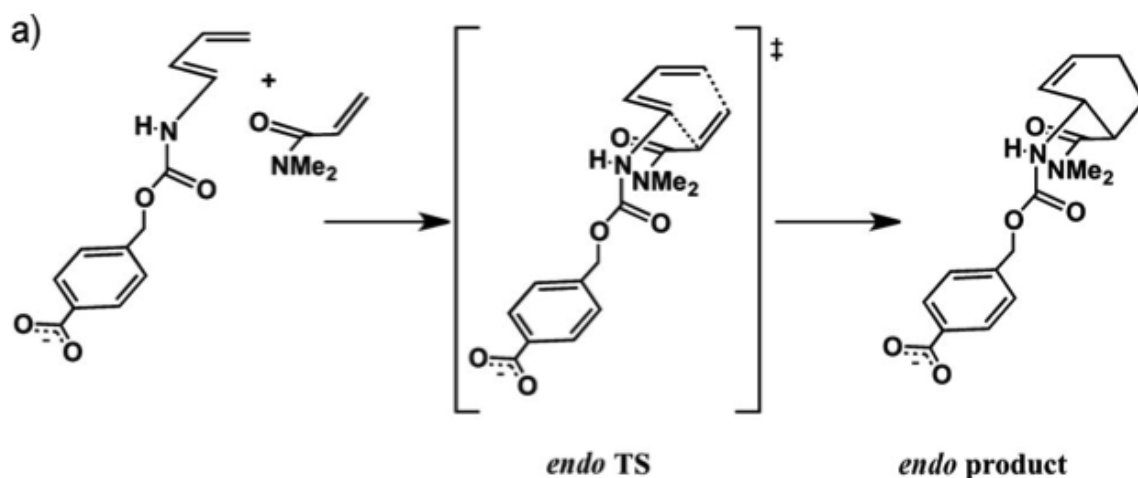
c) Ατομική διακύμανση κατά τις προσομοιώσεις MD. Τα αμινοξέα ενεργού κέντρου (κύκλοι) και η καταλυτική δυάδα (αστέρια) φαίνονται. Οι κορυφές αντιστοιχούν σε βρόγχους με αυξημένη ευκαμψία.

Παράδειγμα Diels-Alderase

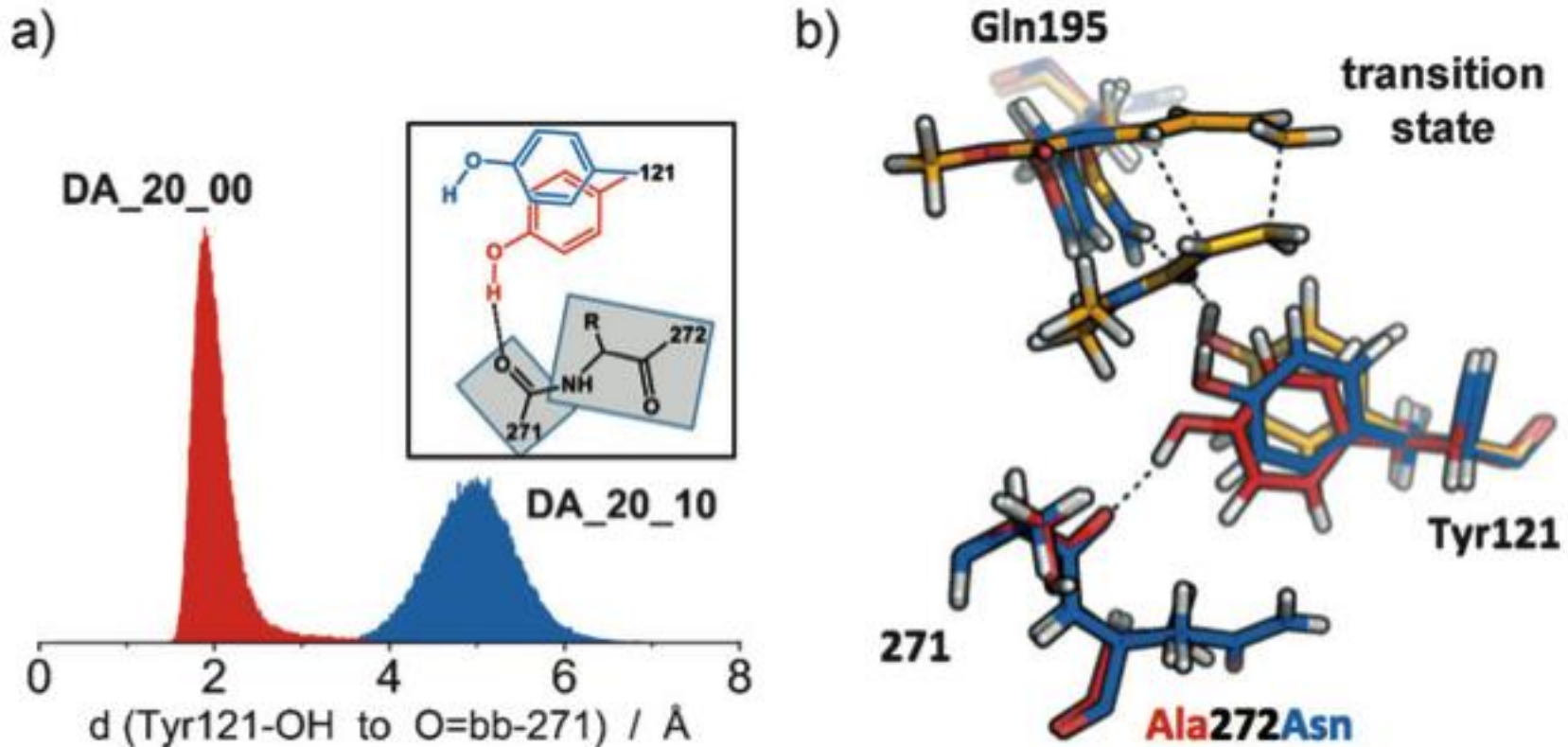
Μόνο ένα ένζυμο φυσικού τύπου είναι γνωστό για αυτή την αντίδραση

(*macrophomate synthase*, T. Ose et al., *Nature*, 2002, **422**: 185-189)

- Η αντίδραση είναι στερεοεκλεκτική και με υψηλή εξειδίκευση στο υπόστρωμα
- Ο σχεδιασμός βασίστηκε σε καταλυτικά αντισώματα (ενεργό κέντρο D, N, Y)
- Τα θεώζυμα που παρουσίασαν ενεργότητα είχαν στο ενεργό κέντρο Q, Y, Y



Παράδειγμα - Diels-Alderase

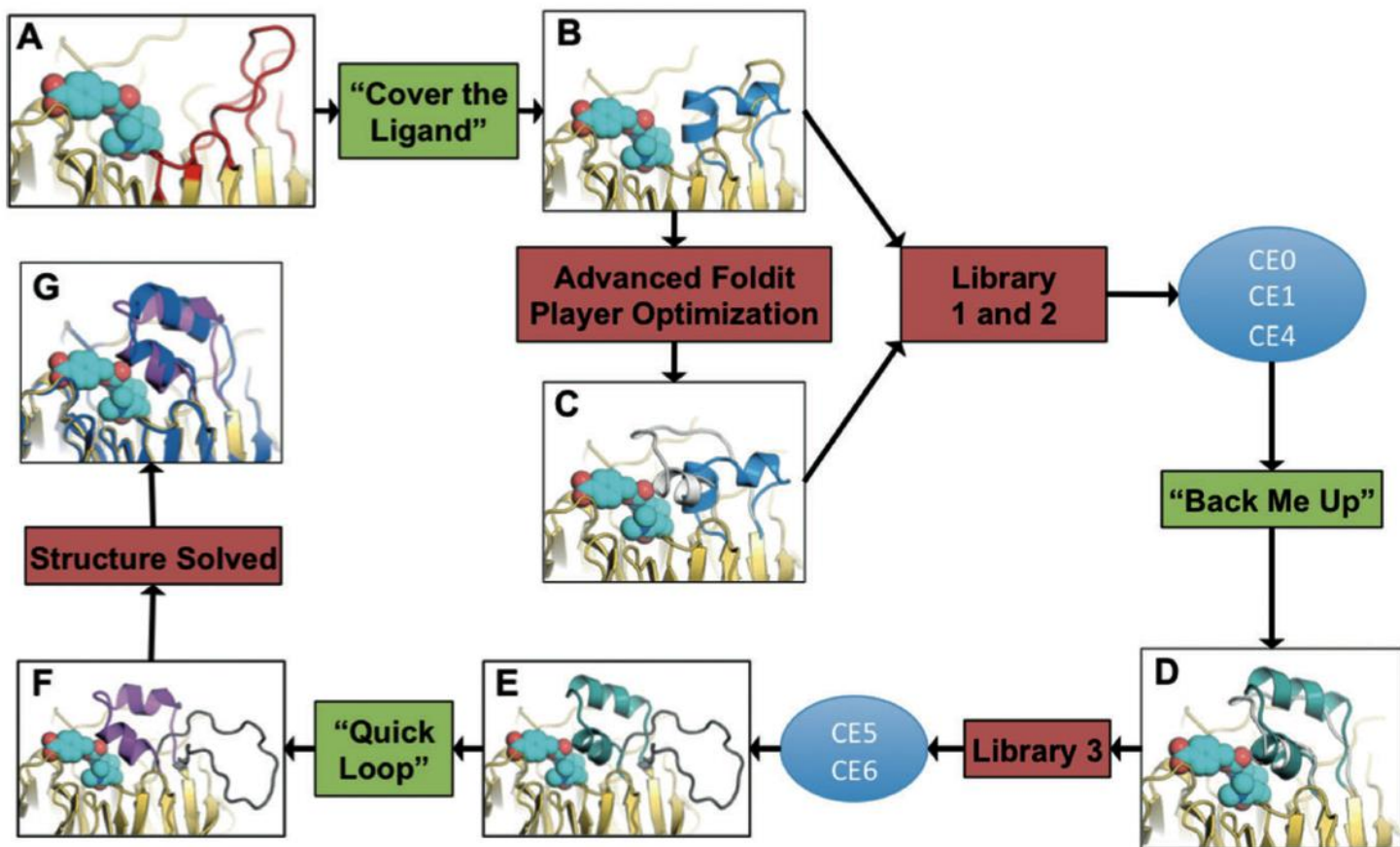


- DA20: k_{cat} : 0.1 h^{-1}
- DA20-10: k_{cat} : 2.1 h^{-1}

Καταλυτική δραστηριότητα

	$k_{\text{cat}}^{[a]}$	$K_M^{[b]}$	$k_{\text{cat}}/K_M^{[c]}$
nat. enzymes	av 10^5	av 10^{-4}	10^6 – 10^9
cat. antibodies	10^{-2} – 1	av 10^{-4}	10^2 – 10^4
Kemp elim.			
cat. antibodies	10^{-1} – 1	10^{-3} – 10^{-4}	10^2 – 10^3
comp. designs	10^{-2} – 1	av 10^{-3}	10 – 10^2
evolved designs	1 – 20	10^{-3} – 10^{-5}	10^3 – 10^6
Retro-Aldol			
cat. antibodies	10^{-3} – 10^{-1}	10^{-4} – 10^{-5}	10 – 10^3
comp. designs	10^{-2} – 10^{-1}	av 10^{-4}	10^{-2} – 10^{-1}
Diels–Alder^[d]			
cat. antibodies	av 10^{-5}	av 10^{-3}	av 10
comp. designs	10^{-5} – 10^{-4}	10^{-1} – 10^{-4}	1 – 10^2

Σχεδιασμός μέσω fold it



Διάγραμμα εργασίας σχεδιασμού βρόγχων με το Foldit για το DA20_10. Οι δυσκολίες που εμφανίστηκαν στην online κοινότητα εμφανίζονται με πράσινο. Οι καλύτεροι σχεδιασμοί από την κοινότητα χρησιμοποιήθηκαν για πειραματικές δοκιμές και για τα επόμενα βήματα στις δοκιμασίες του Fold it.

Πρόβλεψη πρωτεϊνικής αναδίπλωσης

Βασική ιδέα:

- Η σωστή πρόβλεψη της δομής από την αμινοξική αλληλουχία είναι το άγιο δισκοπότηρο στην πρωτεϊνική μηχανική
- Η πρόβλεψη είναι αρκετά δύσκολη καθώς υπάρχουν αναρίθμητες “λύσεις” για μη ομοιοπολικές αλληλεπιδράσεις στην αναδιπλωμένη πρωτεΐνη (van der Waals, δεσμοί υδρογόνου, υδρόφοβες αλληλεπιδράσεις κ.τ.λ.)
- Πως μπορείς η “φυσική” δομή να επιλεγεί και να αποκλειστούν οι “μη φυσικές” αλληλεπιδράσεις που δεν οδηγούν σε ενεργό ένζυμο;
- Πως μπορούν οι προβλέψεις της δευτεροταγούς δομής να μας οδηγήσουν σε τριτοταγή και τεταρτοταγή δομή;

Πρόβλεψη πρωτεϊνικής αναδίπλωσης

Νέες προσεγγίσεις:

- Εστίαση σε μικρές μοτίβα τριτοταγούς δομής που προέρχονται από 2-3 συνεχόμενα στοιχεία δευτεροταγούς δομής
- Χαρτογράφηση αυτών των δομικών στοιχείων με το πρόγραμμα Rosetta και σύγκριση με φυσικές δομές στην PDB
- Πολλαπλές προσομοιώσεις αναδίπλωσης ώστε να “διδασχθεί” το σύστημα και να μπορέσει να κάνει καλύτερες προβλέψεις
- Αυτό οδήγησε σε τρεις γενικούς κανόνες

Πρόβλεψη πρωτεϊνικής αναδίπλωσης

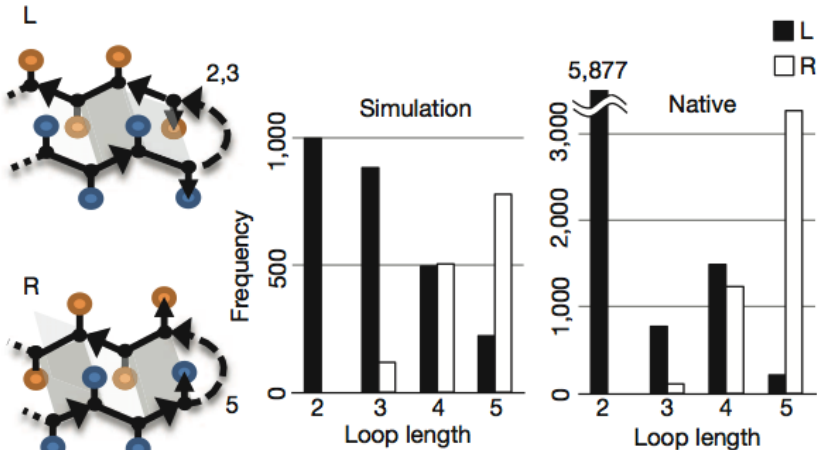
Κανόνες για τις συνδέσεις στις αβ-αναδιπλώσεις (ασχέτως αμινοξικής αλληλουχίας!):

- **ββ-κανόνας:** βασισμένο στην χειρομορφία, προκύπτουν L και R ββ-μονάδες
Βρόγχοι με 2-3-αμινοξέα δίνουν L-φουρκέτες, 5-αμινοξέα R-φουρκέτες
- **βα-κανόνας:** βασισμένο στον προσανατολισμό P (παράλληλες) και A (αντιπαράλληλες) βα-μονάδες
Βρόγχοι 2-αμινοξέων δίνουν P-προσανατολισμό, 3-αμινοξέα A-προσανατολισμό
- **αβ-κανόνας:** βασισμένο στον προσανατολισμό, P αβ-μονάδες
προτιμώμενος προσανατολισμός P, ειδικά όταν δεσμός υδρογόνου σταθεροποιεί την έλικα

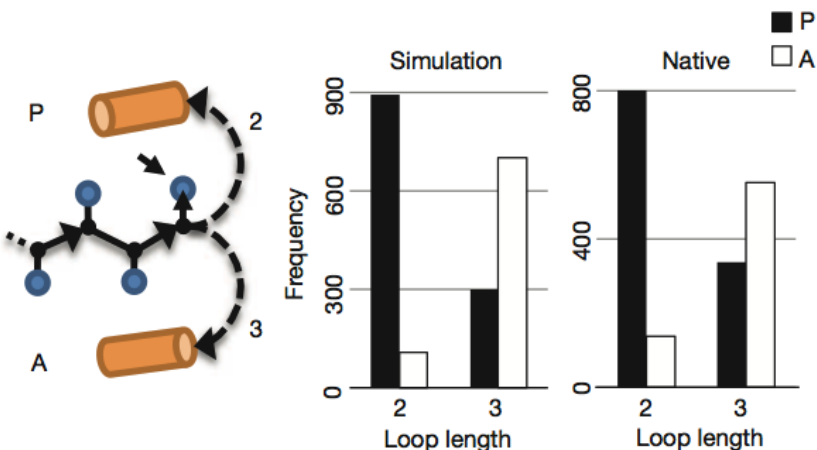
Οι κανόνες επεκτάθηκαν για τρία στοιχεία (ββα, αββ, βαβ)

Πρόβλεψη πρωτεϊνικής αναδίπλωσης

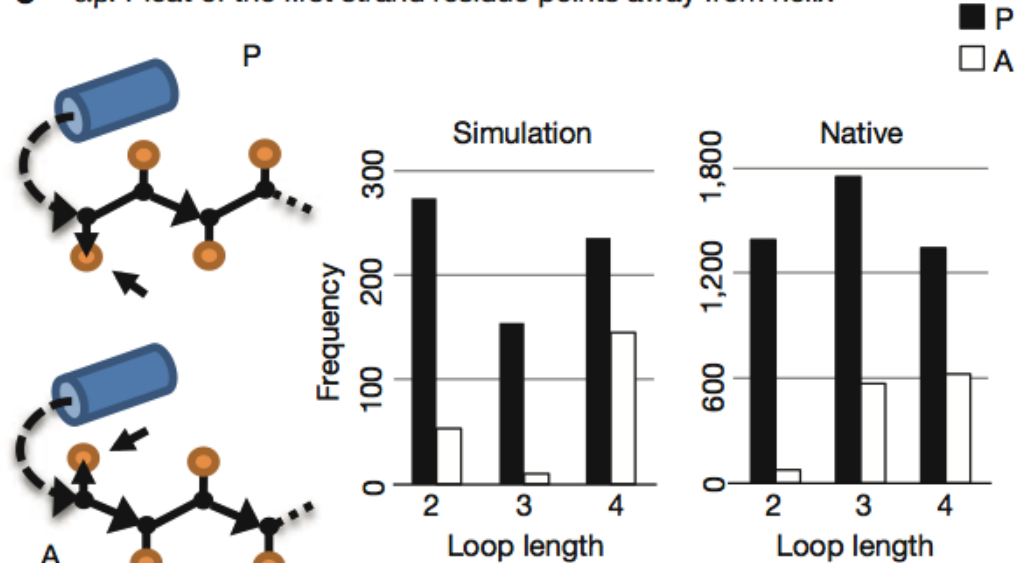
a $\beta\beta$: Chirality of β -hairpins is determined by loop length



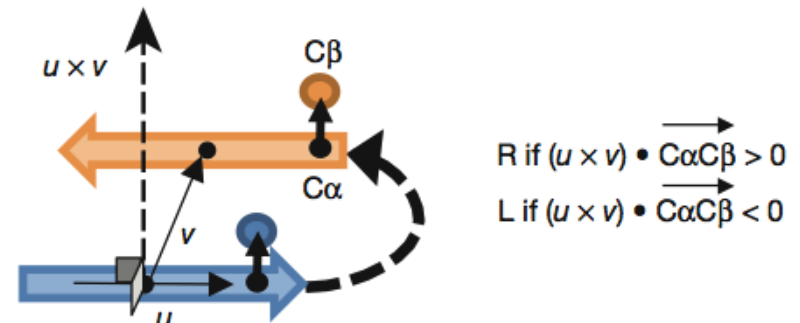
b $\beta\alpha$: Helix direction is determined by pleat direction of the last strand residue and loop length



c $\alpha\beta$: Pleat of the first strand residue points away from helix

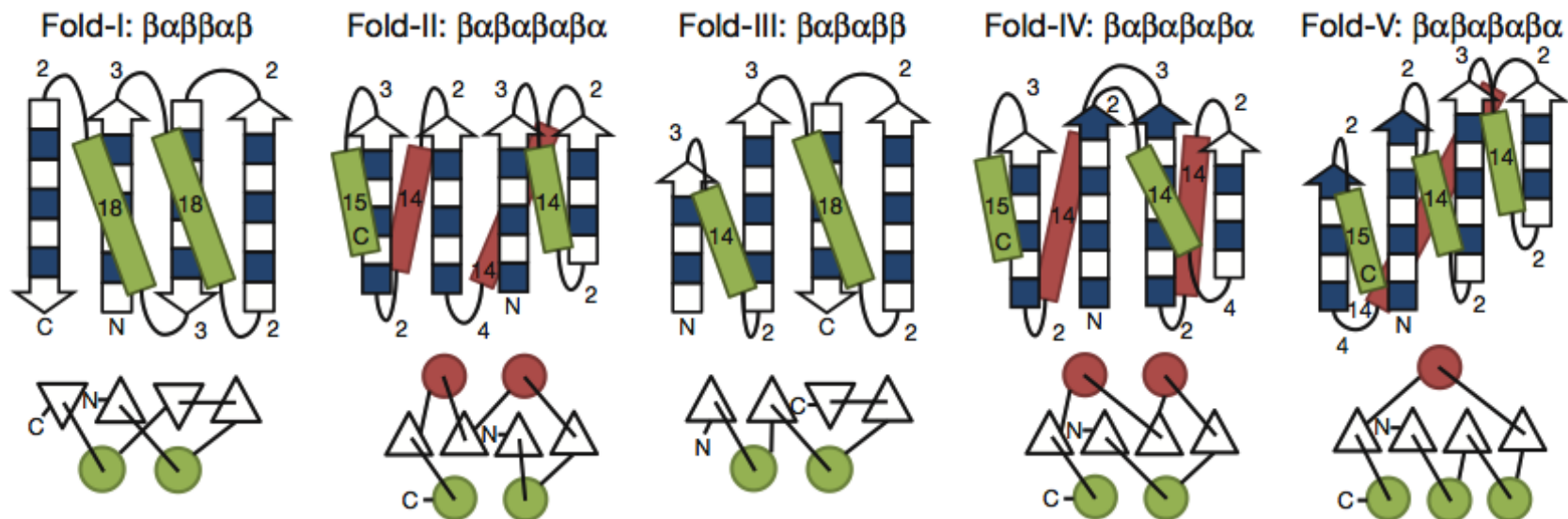


d



Πρόβλεψη πρωτεϊνικής αναδίπλωσης

- Επιλογή μήκους δευτεροταγούς δομής που ευνοούν την επιθυμητή αναδίπλωση (ή δισχεραίνουν τις εναλλακτικές)
- Ανάθεση μήκους βρόγχων και πτυχών σύμφωνα με τους κανόνες
- Εκτέλεση προσομοιώσεων αναδίπλωσης με Rosetta, προσομοιώσεις Monte-Carlo, και RosettaDesign, με αρνητικούς σχεδιασμούς, φιλτράροντας τις προβλέψεις, πειραματικό χαρακτηρισμό



Πρόβλεψη πρωτεϊνικής αναδίπλωσης

- >50 συνθετικά γονίδια από τις 5 κλάσεις, καμία γνωστή πρωτεΐνη!
- Έκφραση, καθαρισμός, CD- και NMR-φασματομετρία
- Οι περισσότερες εκφράστηκαν καλά σε διαλυτή μορφή
- NMR δομές ταίριαζαν με τις θεωρητικές
- Ο υπολογιστικός σχεδιασμός δούλεψε!

Table 1 | Summary of experimental results for designed proteins

	Designs tested	Expressed*	Soluble*	$\alpha\beta$ -protein circular dichroism spectrum	Stable† ($T_m \geq 95^\circ\text{C}$)	Monomeric‡	Well-resolved NMR§	Success
Fold-I	11	9	8	6	3	2	3	1 (9%)
Fold-II	12	12	12	10	10	4	4	4 (33%)
Fold-III	14	13	11	9	7	6	5	3 (21%)
Fold-IV	5	4	4	4	2	4	3	2 (40%)
Fold-V	12	11	10	3	3	1	1	1 (8%)

The second column shows the number of designs experimentally tested for the fold in the leftmost column. The subsequent columns give the number of designs that satisfy each criterion. The successful designs are defined as those that satisfy all criteria. The details of the results are shown in Supplementary Tables 1–5.

* Expression and solubility were assessed by SDS–polyacrylamide gel electrophoresis and mass spectrometry.

† Stability was measured by thermal denaturation; T_m is the melting temperature.

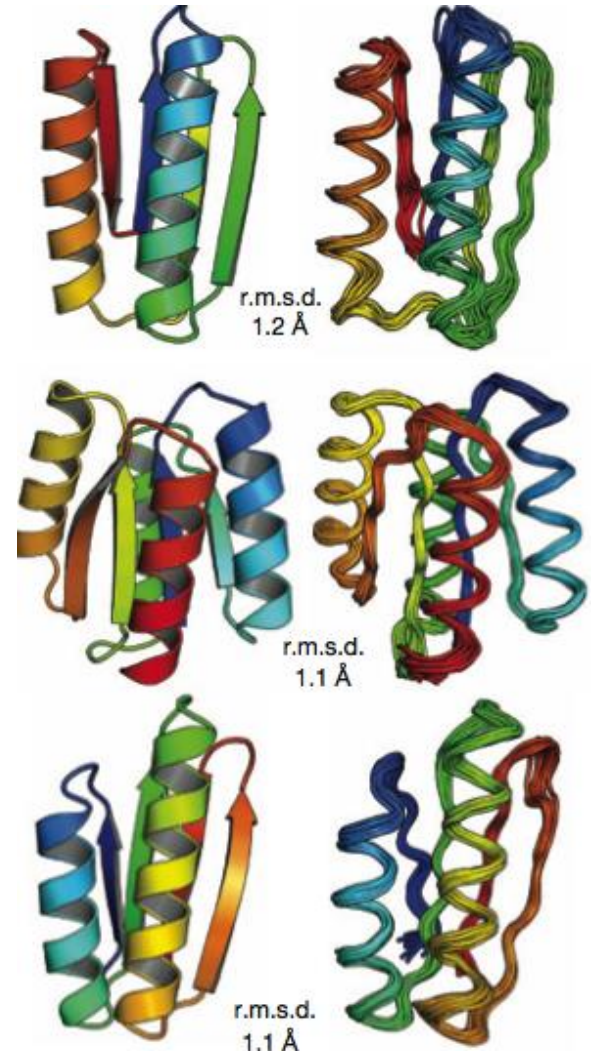
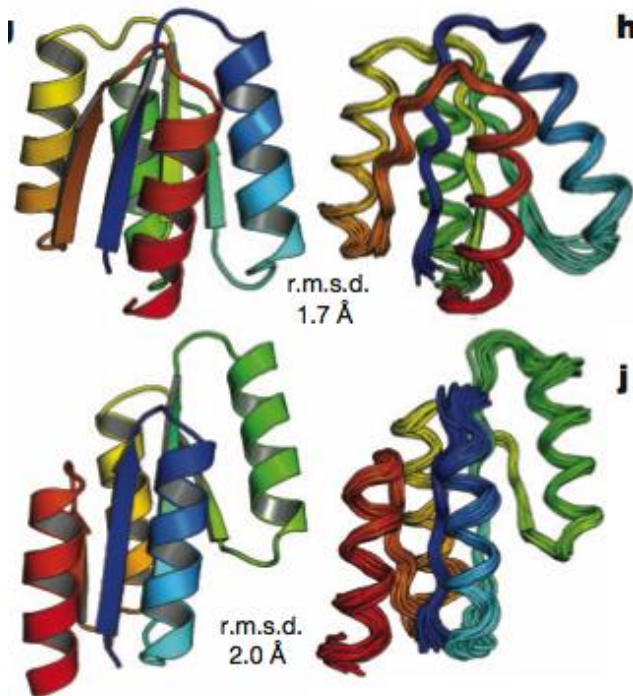
‡ SEC-MALS was used to determine oligomerization state. We counted the number of designs in which the main peak of the absorbance at 280 nm corresponds to the monomeric state.

§ For Folds-I and -II, one-dimensional NMR spectra were collected, and for Folds-III, -IV and -V, two-dimensional ^1H – ^{15}N HSQC spectra were collected.

Πρόβλεψη πρωτεϊνικής αναδίπλωσης

Σύγκριση συνολικής τοπολογίας μοντέλων
(αριστερά) με πειραματικές δομές (δεξιά).

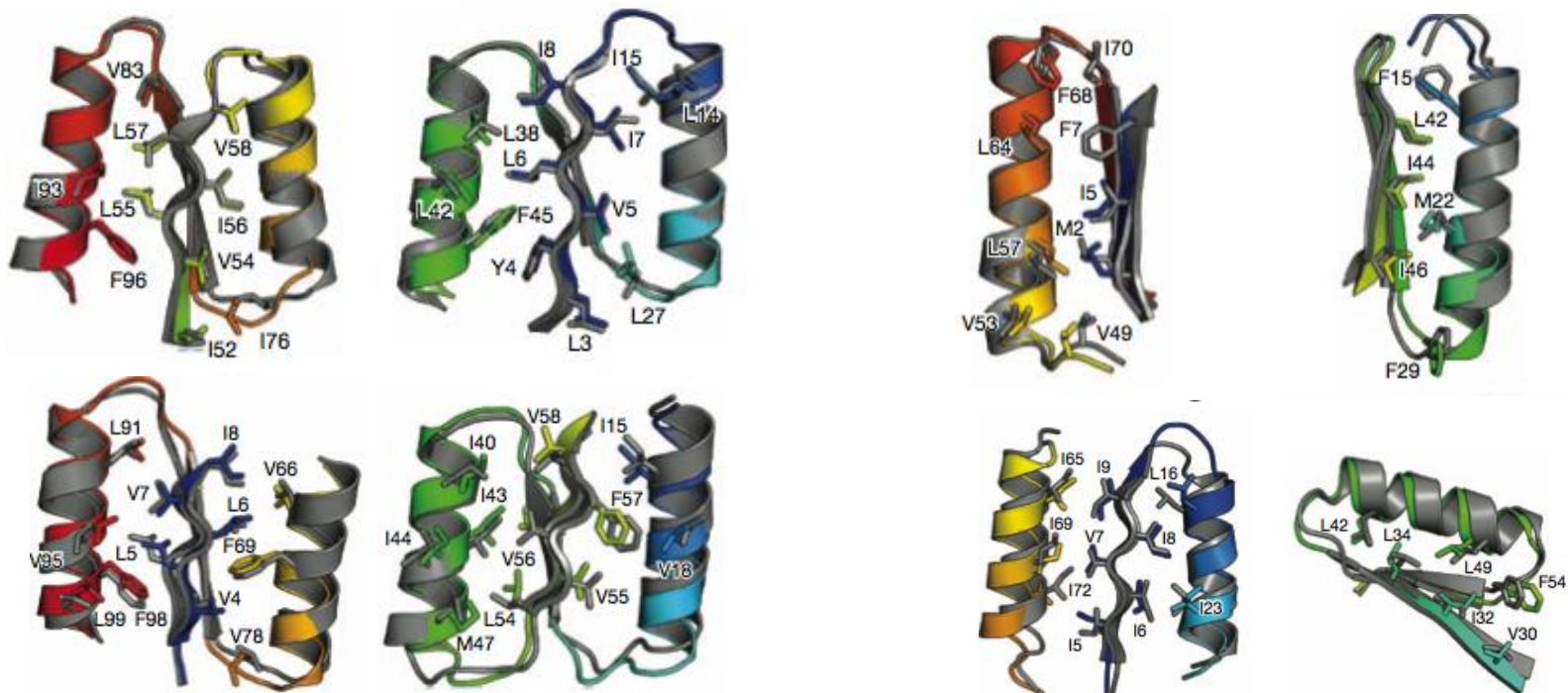
Το r.m.s.d. του Ca δίνεται.



Πρόβλεψη πρωτεϊνικής αναδίπλωσης

Σύγκριση του πακεταρίσματος των πλευρικών ομάδων των αμινοξέων πυρήνα (μοντέλο χρώμα, πειραματικό γκρι).

Αριστερά και δεξιά δίνονται οπτικές του πακεταρίσματος του πυρήνα



Επιπλέον βιβλιογραφία

Bornscheuer & Höhne - Protein Engineering - 2018:

- 3. Computational Analysis of Protein Tunnels and Channels
- 5. A Computational Library Design Protocol for Rapid Improvement of Protein Stability: FRESCO.

Park & Cohran - Protein Engineering and Design - 2009:

- 10 - Computer graphics, homology modeling, bioinformatics
- 11 - Knowledge-based protein design

Bornscheuer & Lutz - Protein Engineering Handbook - (different dates)

volume 1 12 - Computational Protein Design

volume 3 15 - Progress and Challenges in Computational Protein Design

volume 3 17 - Engineering of Protein Tunnels: The Keyhole–Lock–Key Model for Catalysis by Enzymes with Buried Active Sites

- Jumper et al. Highly accurate protein structure prediction with AlphaFold, *Nature* **2021**, 596: 583-589.
- Kiss et al., Computational enzyme design, *Angew. Chem. Int. Ed.* **2013**, 52: 5700-5725.
- Kries et al., De novo enzymes by computational design, *Curr. Opin. Chem. Biol.* **2013**, 17: 221-228.
- Bjelic et al., Computational design of enone-binding proteins with catalytic activity for the Morita–Baylis–Hillman reaction, *ACS Chem. Biol.* **2013**, 8: 7549-7567.
- Koga et al., Principles for designing ideal protein structures, *Nature* **2012**, 491: 222-227.
- Cooper et al, Predicting protein structures with a multiplayer online game, *Nature*, **2010**, 466: 756-760.
- <http://fold.it/portal/>