

wykład 12

Przewidywanie struktur trzeciorzędowych

dr Jacek Śmietański

jacek.smietanski@ii.uj.edu.pl

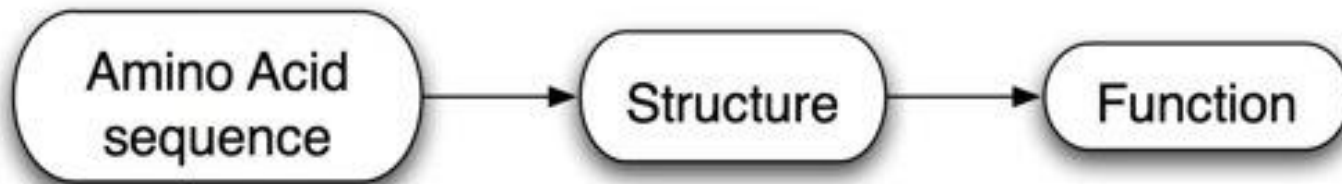
<http://jaceksmietanski.net>

1. Przewidywanie struktur trzeciorzędowych
 1. Modelowanie homologiczne
 2. Rozpoznawanie zwoju
 3. Przewidywanie *ab initio*
2. Nakładanie i porównywanie struktur
3. Ocena jakości algorytmów: konkursy CASP, CAFASP, CAPRI, RNA-Puzzles
4. Projekty rozproszone
5. Citizen Science



Przewidywanie struktur trzeciorzędowych

Do czego potrzebna nam struktura?



Znajomość funkcji białek pozwala zrozumieć organizm



<https://publications.nigms.nih.gov/structlife/chapter1.html>

Istnieje szereg technik eksperymentalnych. Do najważniejszych należą:

- **Krystalografia rentgenowska (X-Ray)**

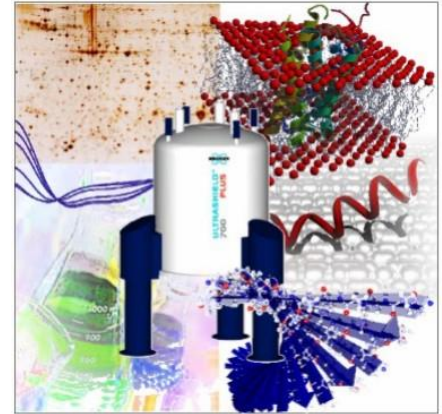
(ok. 90% wszystkich zidentyfikowanych struktur)

- **Jądrowy rezonans magnetyczny (NMR)**

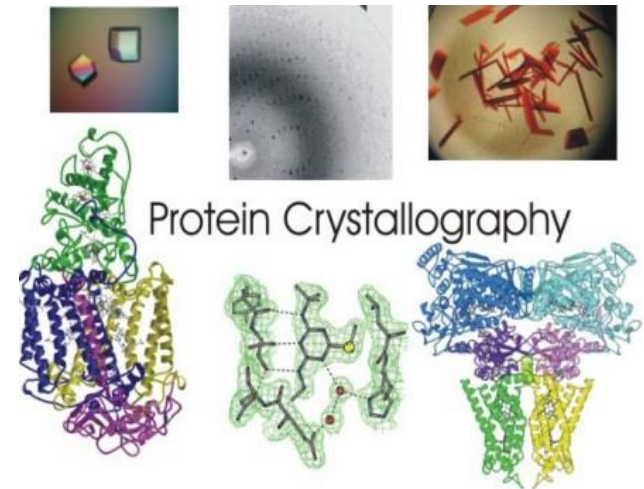
(ok. 10% skruktur)

- **Mikroskopia elektronowa (EM)**

(dopiero od niedawna pozwala na wystarczającą dokładność; popularność tej metody rośnie)



Metody eksperymentalne są bardzo drogie i czasochłonne. Dla wielu grup cząsteczek, mimo prób, do dziś nie udało się zakończyć eksperymentu sukcesem.



Podejście fizyczne (szkoła boltzmannowska):

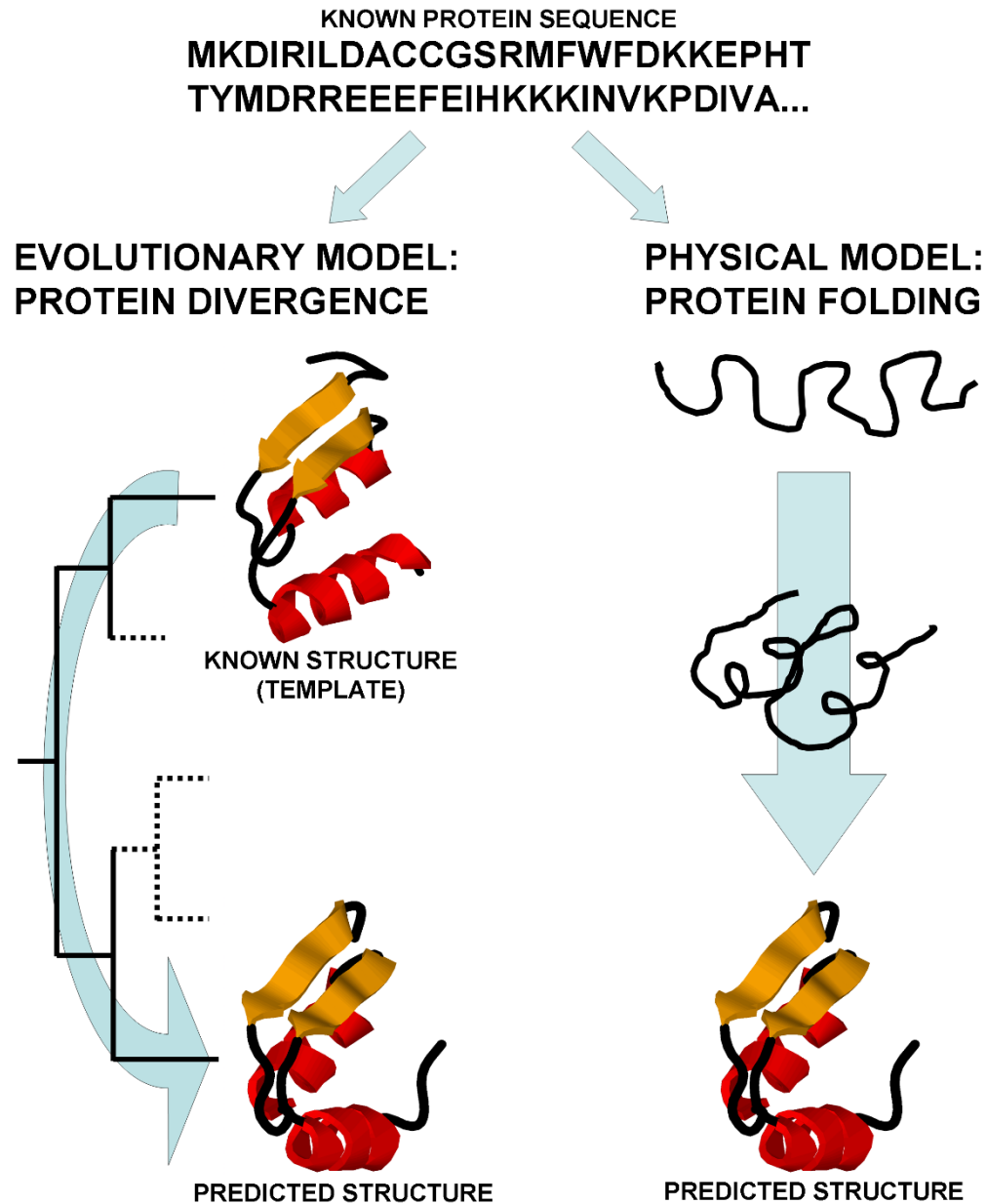
Modelowanie zwijania białka korzystając z praw fizyki statycznej. Jest to proces poszukiwania przez łańcuch konformacji o najniższej energii swobodnej. W komórkach trwa on zaledwie ułamki sekundy.

Podejście ewolucyjne (szkoła darwinowska):

Rekonstrukcja procesu powstania sekwencji i struktury białka na drodze ewolucji. Przyrodzie zabiera to miliony lat.

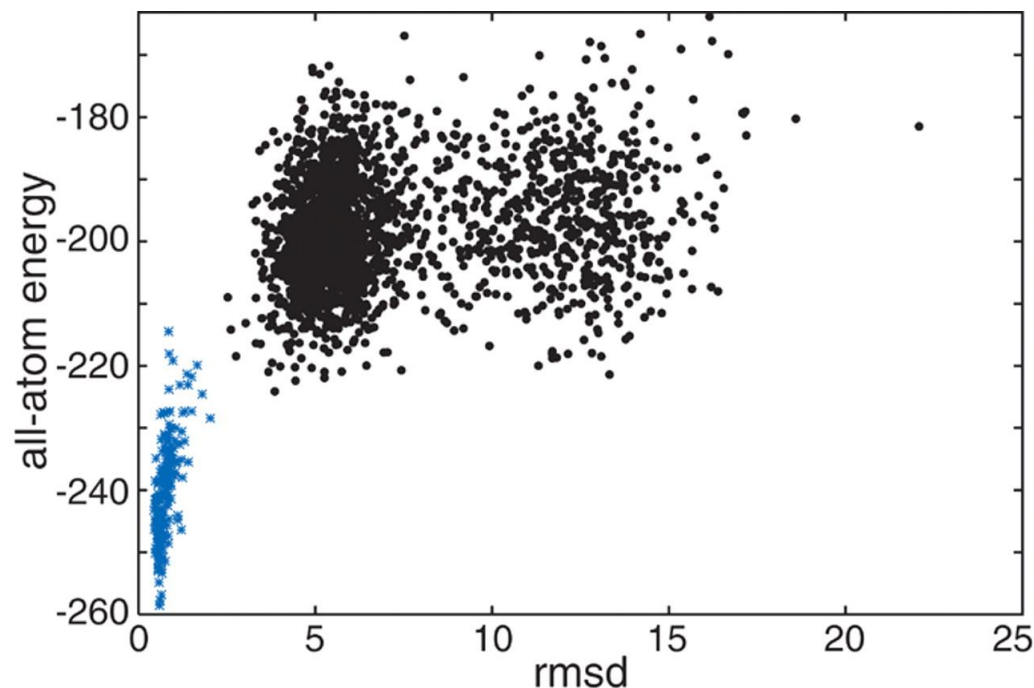
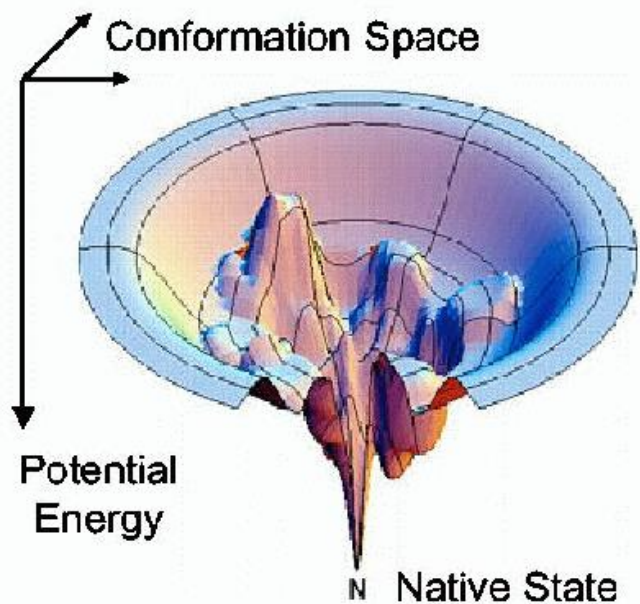


Podejścia do przewidywania *in silico* - ilustracja



Odnalezienie minimum energetycznego jest bardzo trudne

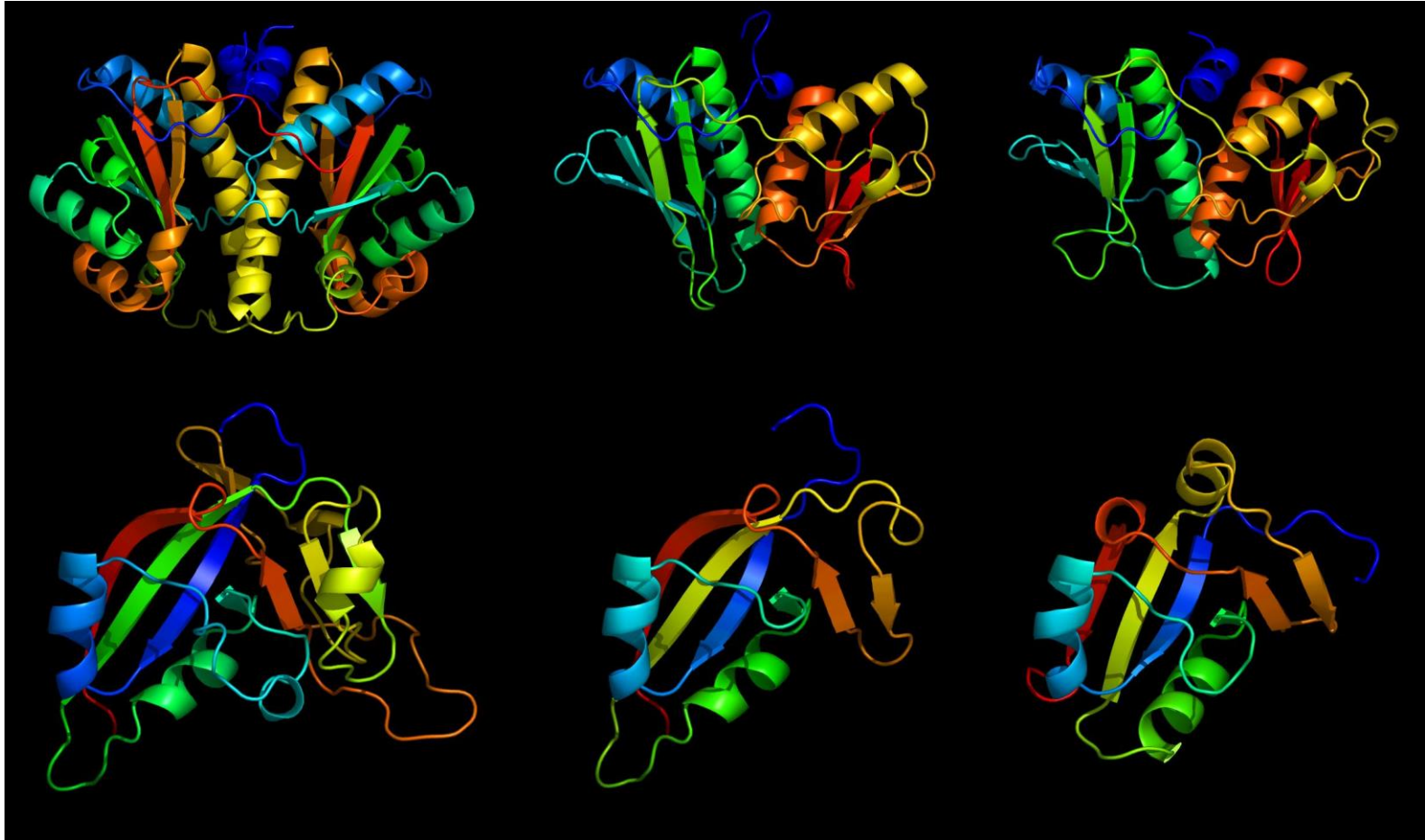
(wielowymiarowy problem optymalizacyjny o wykładniczej złożoności obliczeniowej względem liczby atomów; przeciętne białko składa się z kilkudziesięciu tysięcy atomów)



template
structure

modeled target
structure

native target
structure

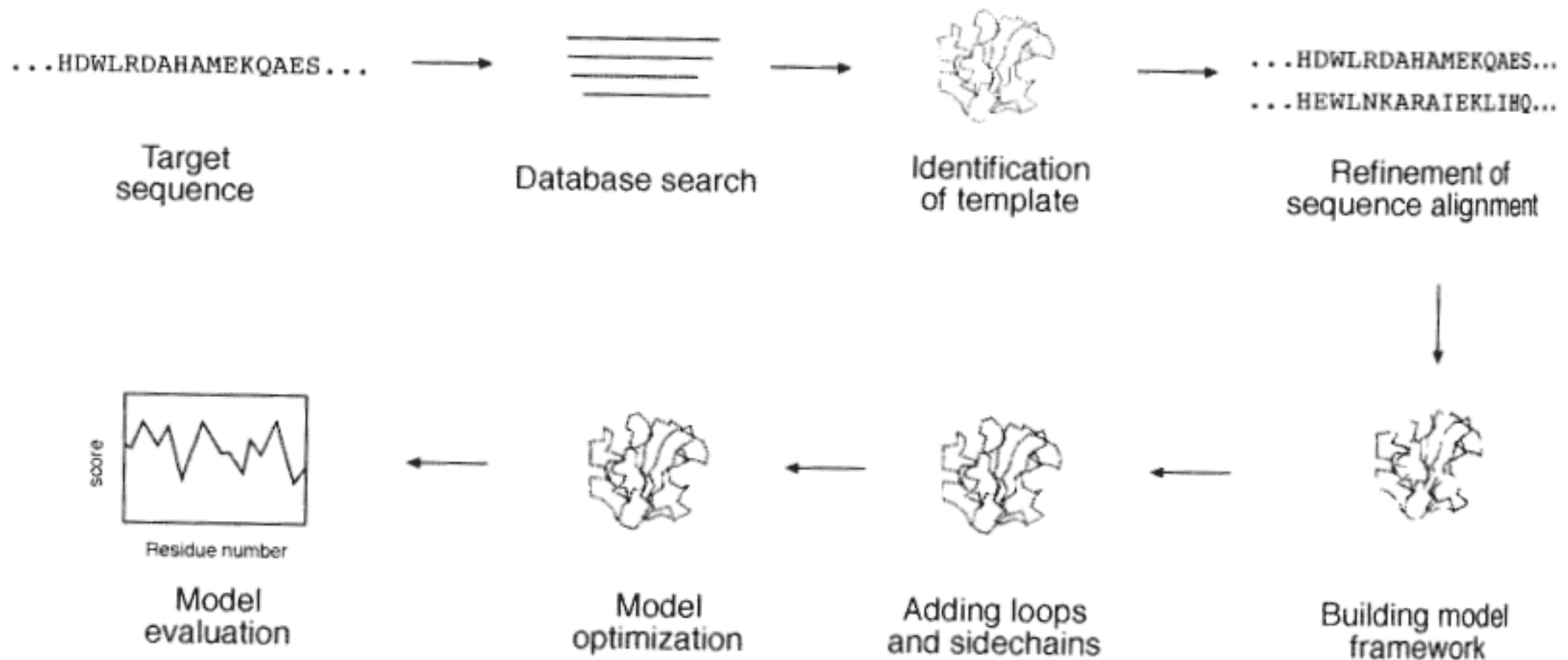


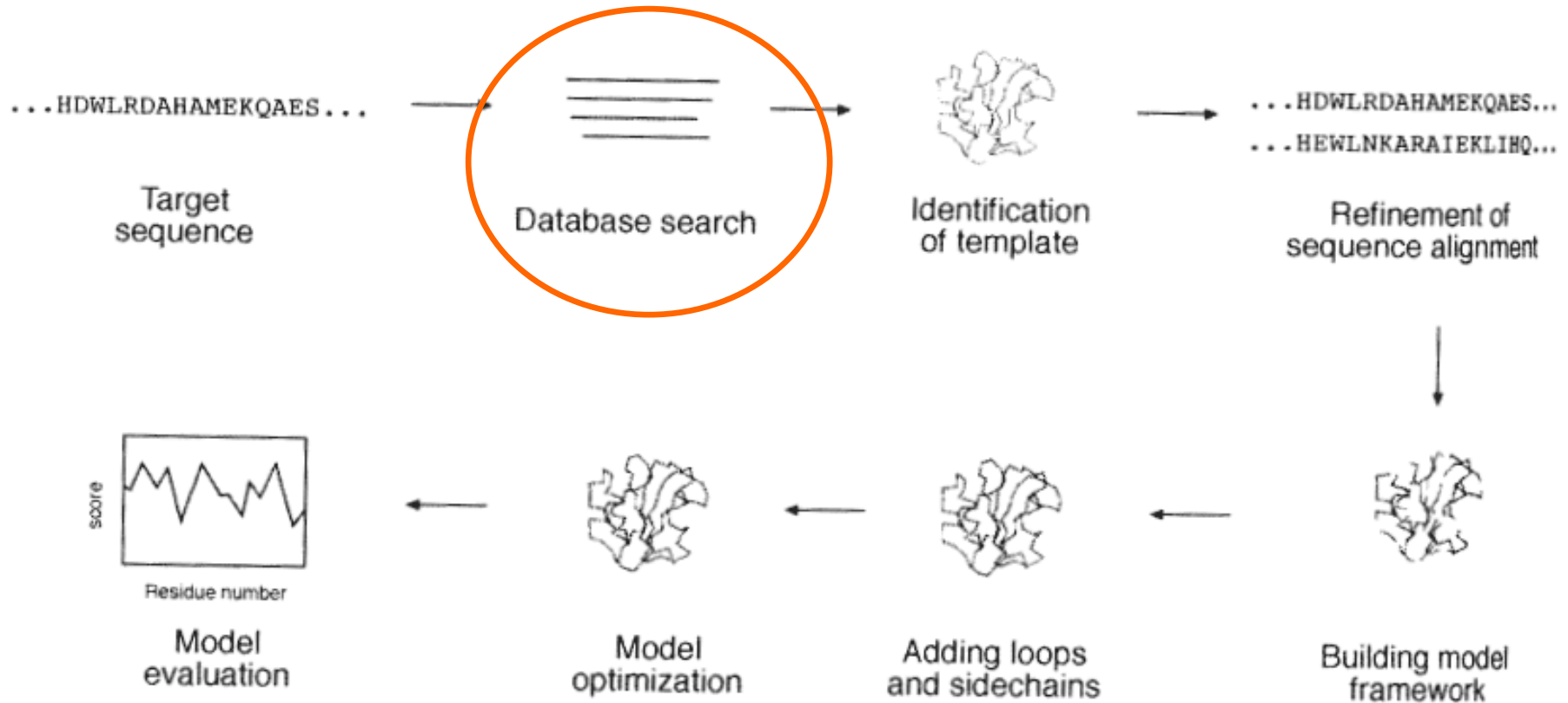
Modelowanie homologiczne

Jeżeli dwa białka wykazują podobieństwo **sekwencyjne**, to prawdopodobnie ich **struktury** przestrzenne też będą do siebie podobne.

Założenie to ma silne podłoże ewolucyjne: białka mają podobne sekwencje, czyli mają wspólnego przodka. Skoro są spokrewnione, zapewne pełnią podobne funkcje, a więc i ich struktura musi być podobna.

Etapy modelowania homologicznego

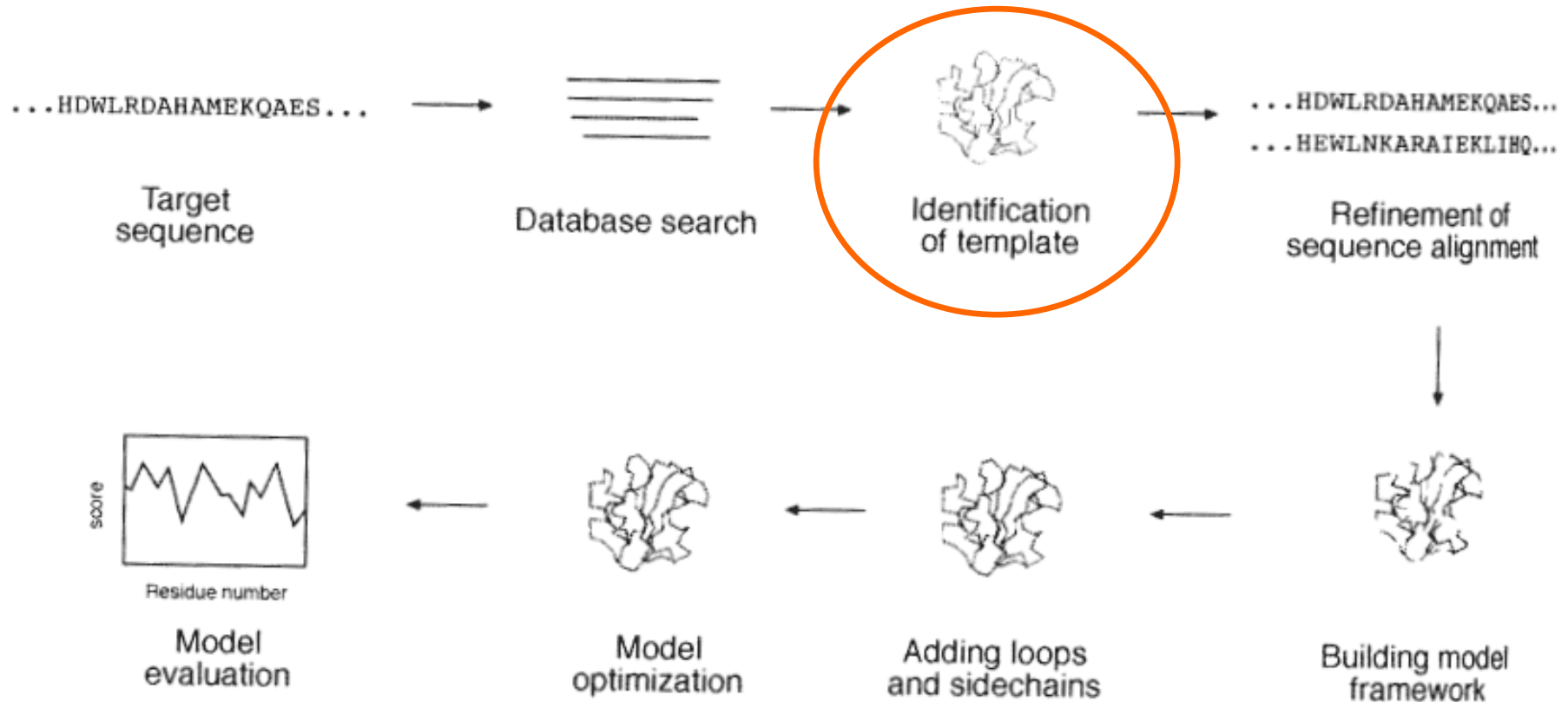




Przeszukujemy bazę struktur w celu znalezienia białek homologicznych o znanej strukturze (BLAST, FASTA, SSEARCH).

SSEARCH – dopasowanie optymalne (algorytm Smitha-Watermana). Ze względu na niewielką liczbę znanych struktur, takie dokładne przeszukiwanie może być przeprowadzone w rozsądnym czasie.

<http://www.ebi.ac.uk/Tools/sss/>



Szablonem zwykle może być białko, którego podobieństwo sekwencyjne do białka-celu jest wyższe niż 30%.

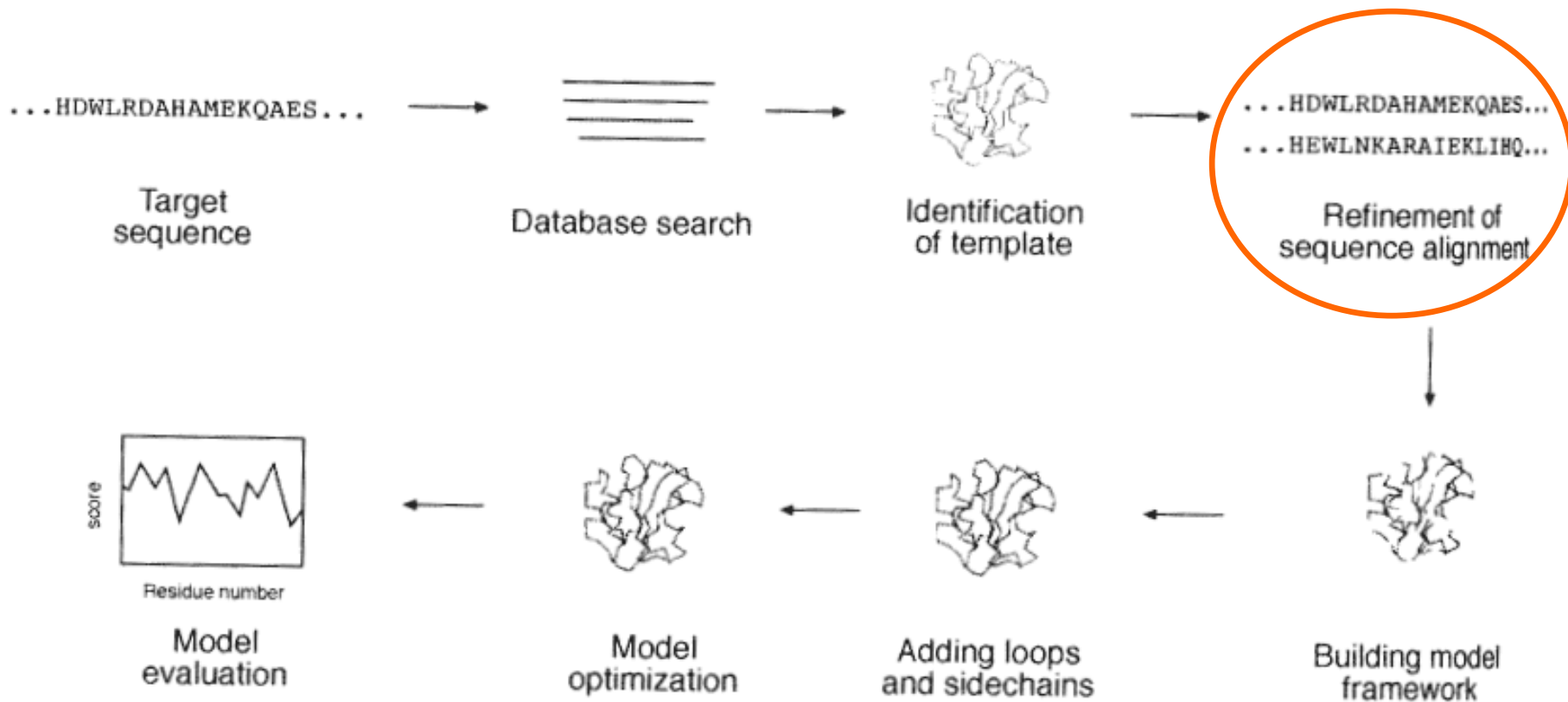
Gdy podobieństwo sekwencyjne jest niższe niż 30%, muszą istnieć inne dowody na homologię.

Gdy znamy wiele podobnych struktur, preferujemy te z lepszą rozdzielczością czy podobnymi kofaktorami.

- przeszukiwanie profili (PSI-BLAST);
- poszukiwanie odległych homologów (rozpoznawanie zwoju)

Przypuszczalnie uda się znaleźć jedynie odległe podobieństwa.

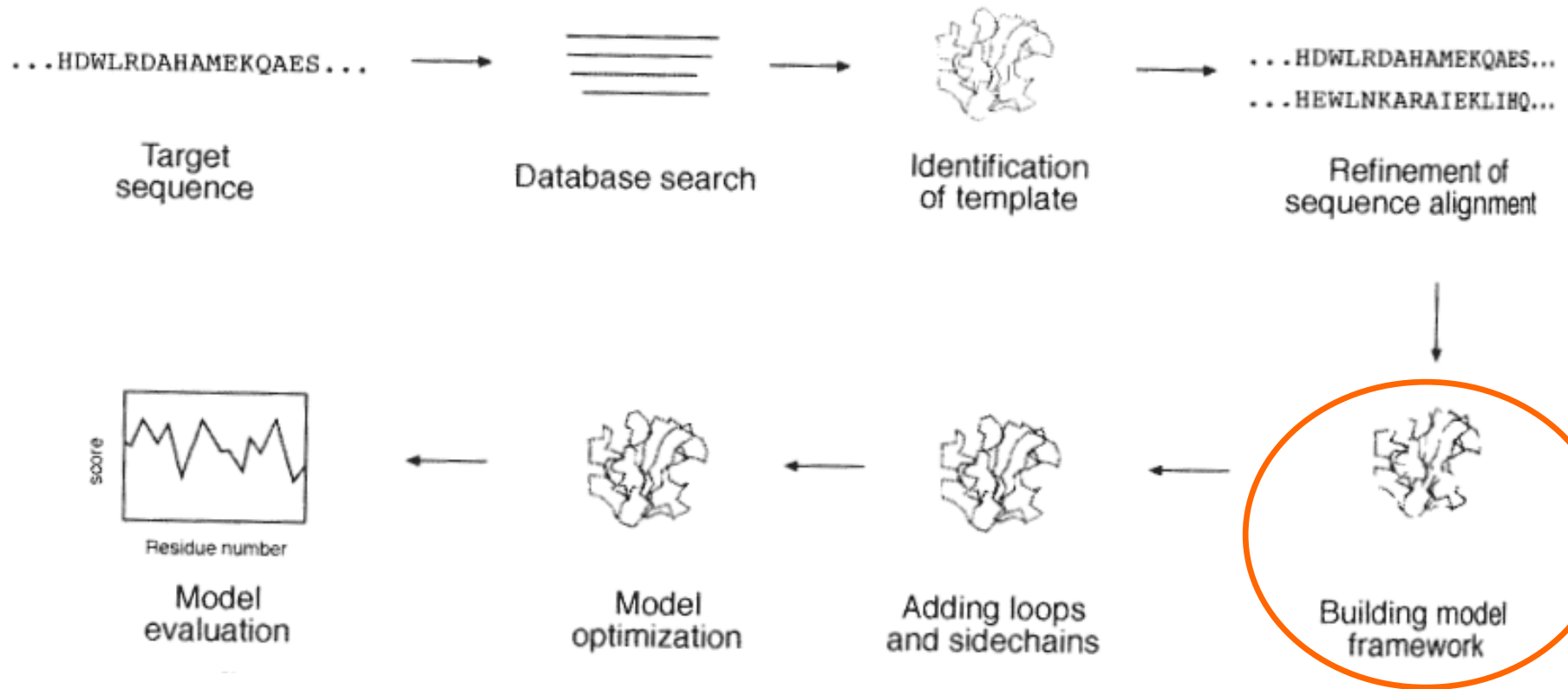
Przyrównanie sekwencyjne szablonu i celu



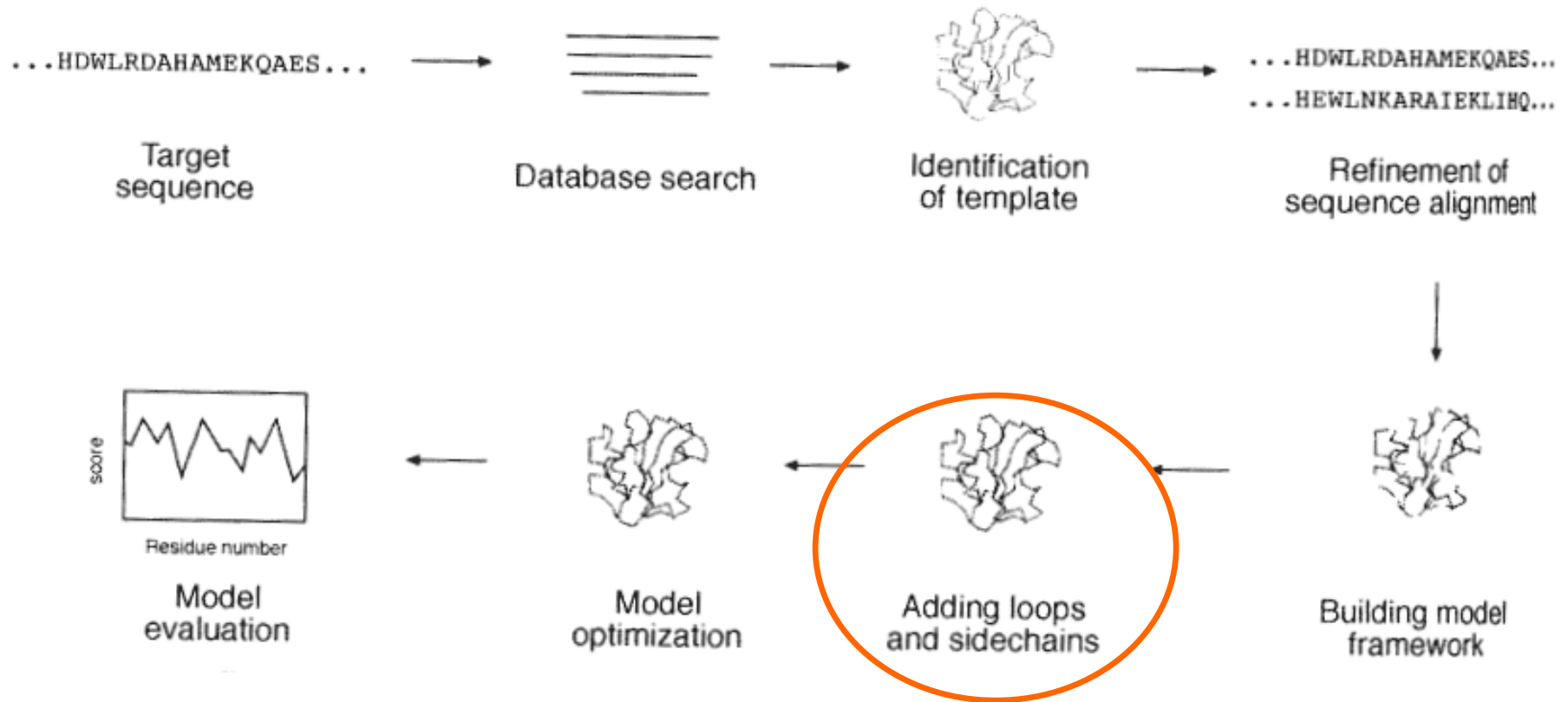
Jest to etap kluczowy dla jakości modelu. Skutków błędnego przyrównania nie da się później naprawić.

Nie ma jednak jednoznacznej wskazówki, jakiego algorytmu przyrównania użyć.

Modelowanie łańcucha głównego



Kopiuujemy współrzędne atomów łańcucha głównego szablonu na współrzędne atomów celu. Jeśli na przyrównanej pozycji aminokwasy są identyczne, kopiujemy także współrzędne atomów reszty.



Uzupełnianie luk w przyrównaniu (gapy odpowiadające za insercję lub delecję).

Przeszukiwanie baz danych lub *ab initio*.

Szukamy pętli dającej najmniej zawał sterycznych z modelowanym białkiem.

FREAD – podejście z wykorzystaniem baz danych

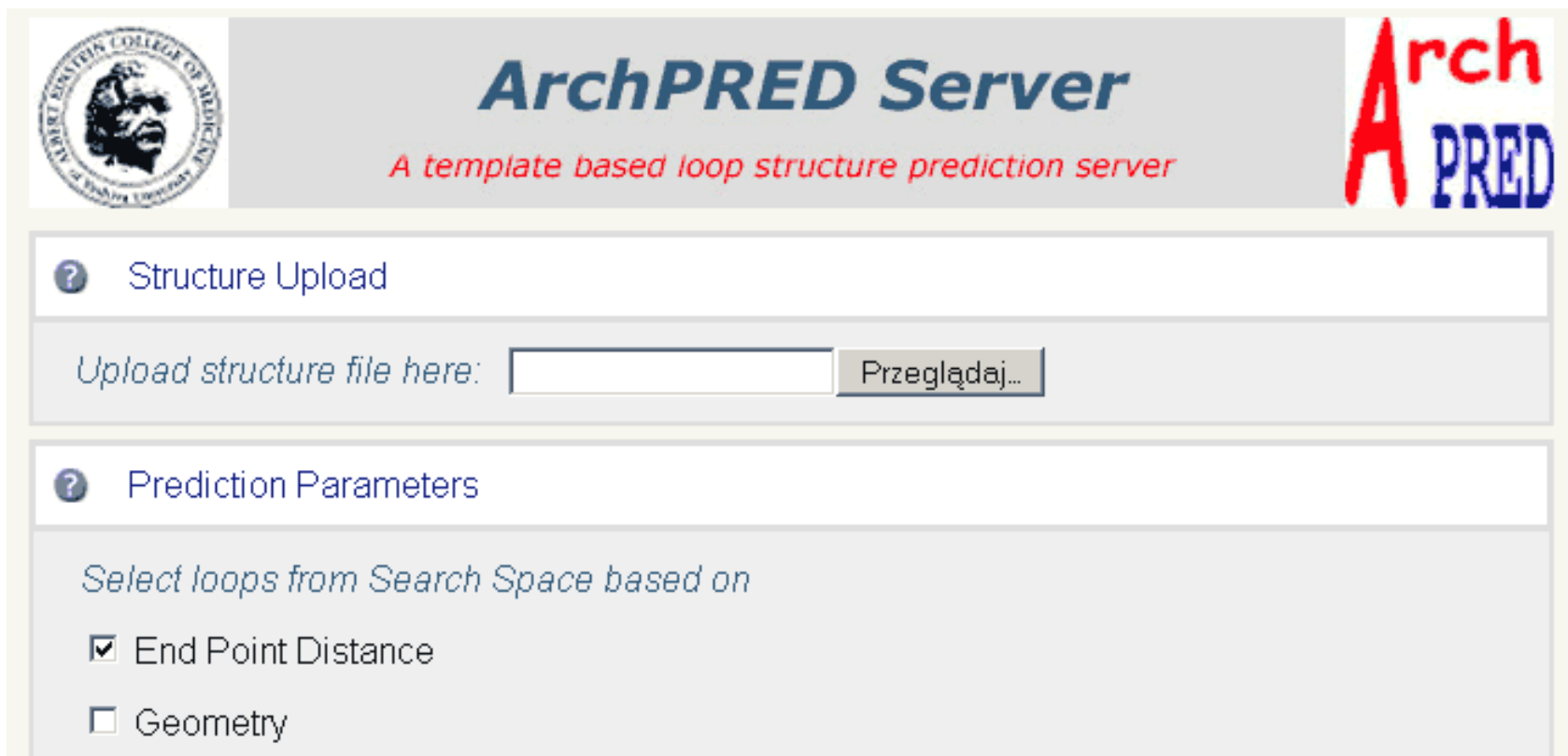
<http://opig.stats.ox.ac.uk/webapps/fread/php/>





FREAD is a database search loop modelling algorithm. Its primary use is to fill in the gaps in incomplete 3D models of protein structures. The input is a 3D model of a protein structure as well as the location and amino acid sequence of the loop region to be modelled. For more information, please refer to the following paper: **Yoonjoo Choi and Charlotte M. Deane (2009). FREAD Revisited: Accurate loop structure prediction using a database search algorithm. Proteins, 78(6):1431-1440.**

ArchPRED – podejście z wykorzystaniem baz danych

<http://manaslu.fiserlab.org/loopred/>



The image shows the ArchPRED Server web interface. At the top, there is a header with the Albert Einstein College of Medicine logo on the left, the text "ArchPRED Server" in a large, bold, blue font in the center, and a red and blue "ArchPRED" logo on the right. Below the header, the text "A template based loop structure prediction server" is displayed in red. The main content area is divided into two sections. The first section, titled "Structure Upload" with a question mark icon, contains the text "Upload structure file here:" followed by a text input field and a "Przeglądaj..." button. The second section, titled "Prediction Parameters" with a question mark icon, contains the text "Select loops from Search Space based on" followed by two checkboxes: "End Point Distance" (checked) and "Geometry" (unchecked).

 **ArchPRED Server** 

A template based loop structure prediction server

? Structure Upload

Upload structure file here: Przeglądaj...

? Prediction Parameters

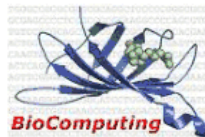
Select loops from Search Space based on

☒ End Point Distance

☐ Geometry

LOBO – podejście *ab initio*

<http://protein.bio.unipd.it/lobo/>



Loop Modelling with LOBO

Version 1.2

[Quick Help](#)

[Examples](#)

[References](#)

[Server Statistics](#)

General Informations

E-Mail address

Name of sequence (optional)

E-mail notification of results ☐

PDB file

Select a valid PDB ID:

- OR - Upload a file:

- OR - Fetch a structure from a PID: and Model #:

Chain ID (if present and not first in file):

Options

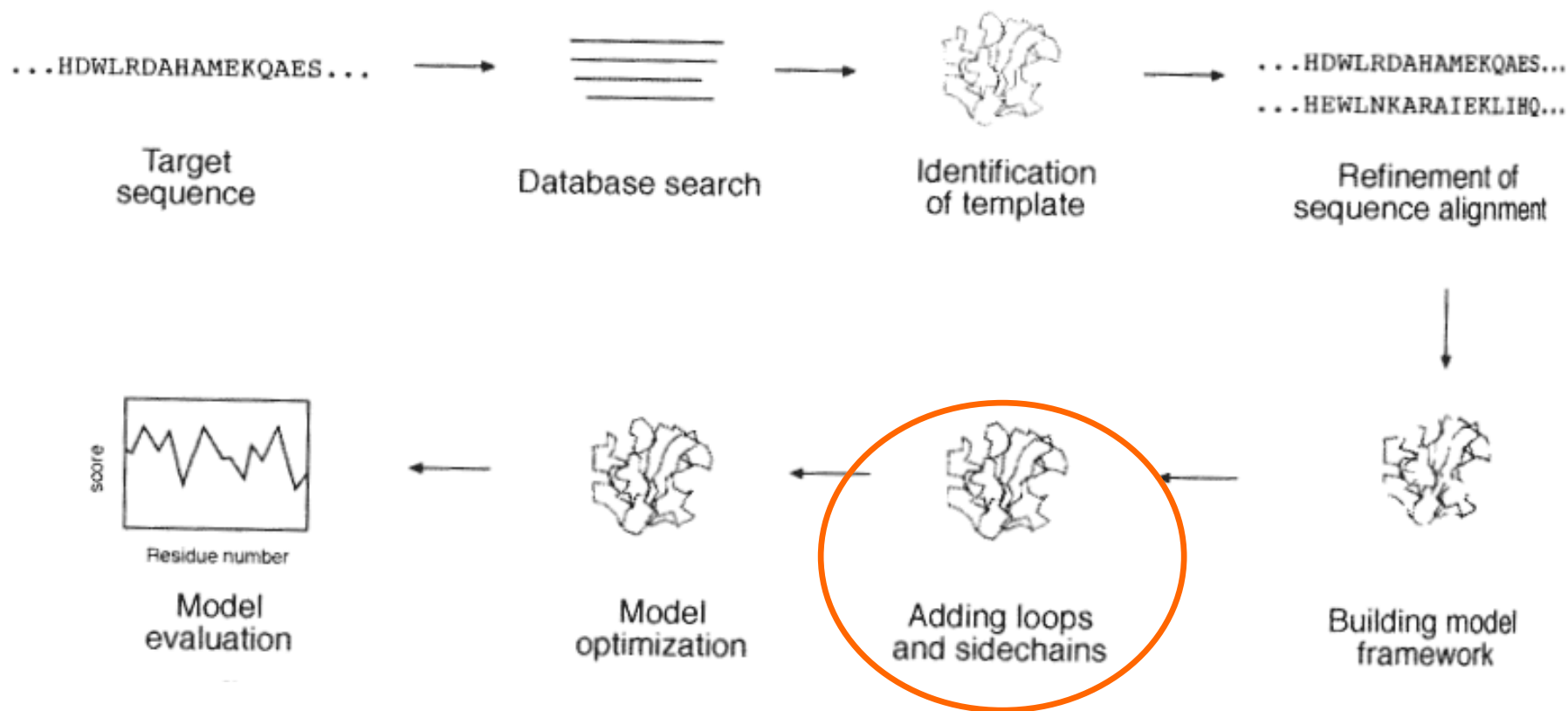
Solutions: (1 - 30)

Loop anchors

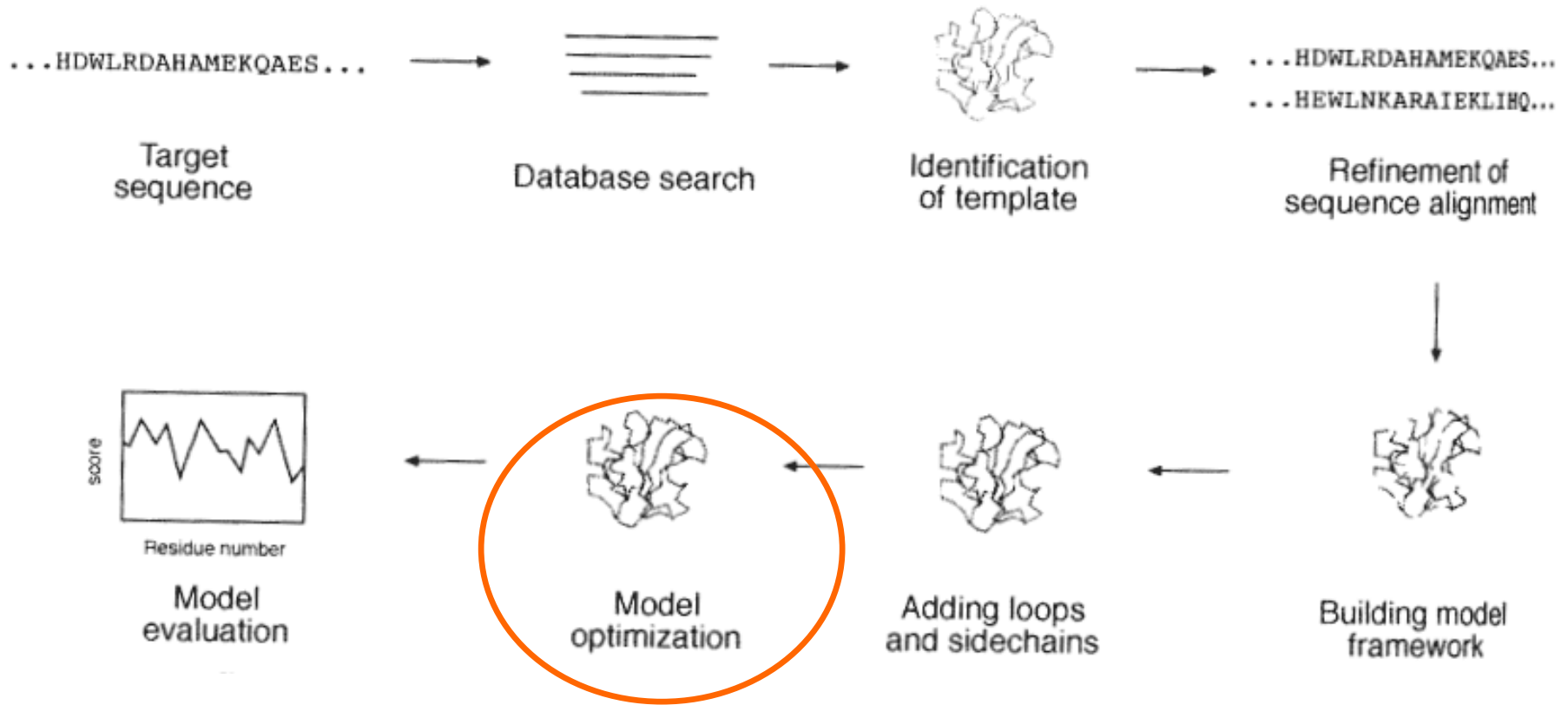
N-terminal loop residue:

C-terminal loop residue:

Uzupełnianie łańcuchów bocznych

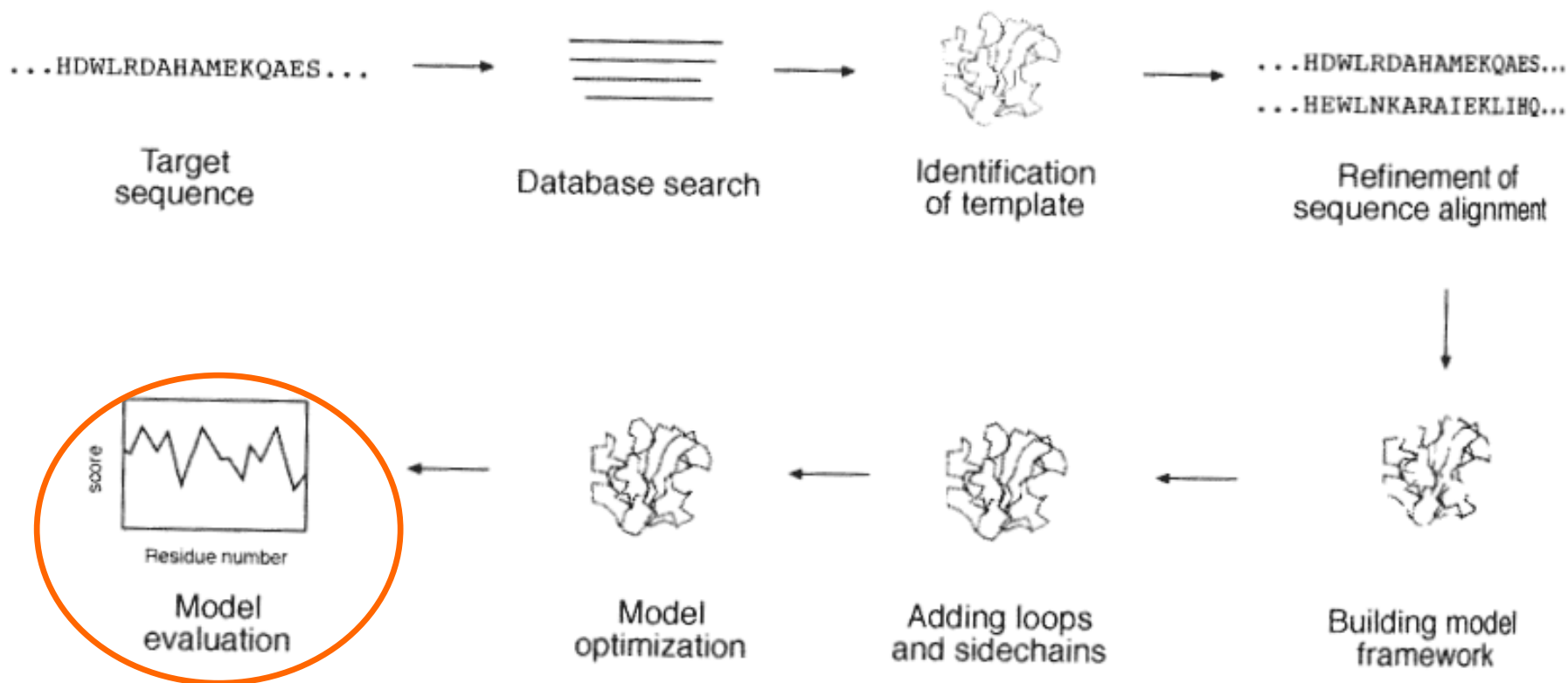


Rotamery – konformacje łańcuchów bocznych uzyskane ze znanych struktur krystalicznych → Szukamy preferowanych kątów torsyjnych, dających najniższą energię oddziaływania z sąsiadującymi atomami.



- Usuwanie konfliktów sterycznych.
- Minimalizacja energii całego układu.
- Symulacja dynamiki molekularnej.

Ewaluacja modelu



Sprawdzenie poprawności kątów, długości wiązań, oddziaływań, itp.

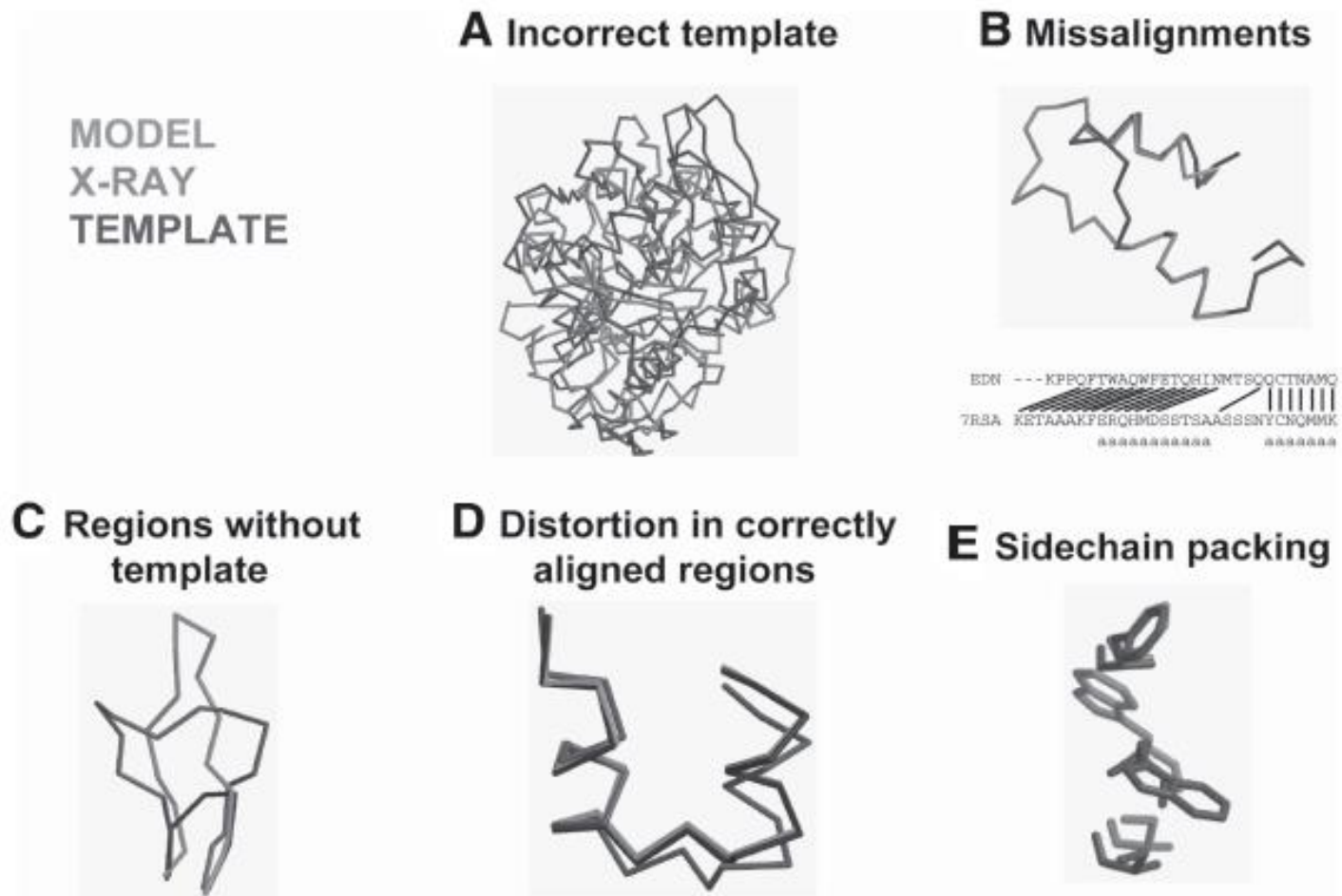


Fig. 4. Typical errors in comparative modeling.

PROCHECK

<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>

- **Home**
- Download
- Manual
- NMR Manual
- References
- Contact

EBI > Groups > Thornton > Software > PROCHECK

PROCHECK and PROCHECK-NMR

PROCHECK checks the stereochemical quality of a protein structure, producing a number of PostScript plots analysing its overall and residue-by-residue geometry. It includes **PROCHECK-NMR** for checking the quality of structures solved by NMR.

PROCHECK

Download

PROCHECK is available free. Download details are given [here](#). We request that you complete and sign the **Confidentiality Agreement** (see below) and return by post, fax or e-mail (see [Contact details](#)). Non-academic users can strike out Clause 7 of the agreement.

Confidentiality Agreement



PROCHECK Confidentiality Agreement

Notes

- ➔ You can upload your structure to [PDBsum](#) to have a full set of **PDBsum** analyses, including PROCHECK plots, generated for it. (Use the **Generate** option in the left-hand menu).

Page last modified: 14 January 2010

WHAT IF

<http://swift.cmbi.ru.nl/servers/html/index.html>

Classes

- [Help](#)
- [Administration](#)
- [Build/check/repair model](#)
- [Structure validation](#)
- [Analyse a residue](#)
- [Protein analysis](#)
- [2-D graphics](#)
- [3-D graphics](#)
- [Hydrogen \(bonds\)](#)
- [Accessibility](#)
- [Atomic contacts](#)
- [Coordinate manipulations](#)
- [Rotamer related](#)
- [Cysteine related](#)
- [Water](#)
- [Ions](#)
- [Docking](#)
- [Crystal symmetry](#)
- [mutation prediction](#)
- [Other options](#)



WHAT IF Web Interface

Click on one of the classes in the left side column to activate those servers.

You can also read the [help page](#), or you can read about the [output formats](#) used by all servers. We also made one [page](#) with notes for people who want to make their own servers.

W A R N I N G. Results are only kept on this server till Saturday, midnight in The Netherlands. This is mainly a safety feature, but also saves us disk space. Feel free to look at the [WHAT IF](#) writeup too.

In case a server fails, first check your input PDB file "List sequence of a PDB file" server that can be found in the "Administration" section of the servers. If this server gives you a long list of administrative information, and no obvious error messages, then look at the [PDBREPORT](#) database to see if there are other obvious problems with the PDB file. If that also does not provide an answer, mail the PDB file (and other files when needed) to Gert Vriend. Please mention the server class (mentioned in the left-hand column of the server page) and the title of the actual server (you click on that in the right-hand part of the server page to get at that server).

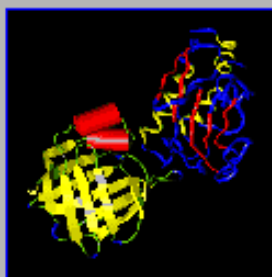
But PLEASE, do not submit the same file over and over again! If the thing fails once, it fails twice, and if it is very slow, submitting more jobs will make it even slower. Remember that the servers run just on my private machine and not some supercomputer at a big centre....

If you have detected any errors, or have any question or suggestion, please send an Email to Gert Vriend.
Roland Krause, Jens Erik Nielsen, [Gert Vriend](#).

Last modified Sat Sep 17 21:44:57 2011

ANOLEA

<http://protein.bio.puc.cl/cardex/servers/anolea/>



ANOLEA

(**A**tom**i**c **N**on-**L**ocal **E**nvironment **A**ssessment)

Francisco Melo and Ernest Feytmans.

This server was originally developed at:

[Unité de Recherche en Biologie Moléculaire \(URBM\)](#)
[The University of Namur, Belgium.](#)

Now it is running at:

[Laboratorio de Bioinformática Molecular](#)
[Pontificia Universidad Católica, Chile.](#)

Verify3D

http://nihserver.mbi.ucla.edu/Verify_3D/



Verify3D Structure Evaluation
Server

[\[Servers Home\]](#)

People ♦ Seminars
Lectures ♦ Webmail
Links ♦ Facilities
Software ♦ Home



The UCLA-DOE Structure Evaluation server is a tool designed to help in the refinement of crystallographic structures. It will provide you with a visual analysis of the quality of a putative crystal structure for a protein. Verify3D expects this crystal structure to be submitted in PDB format. Please note that Verify3D works best on proteins with at least 100 residues. To submit a crystal structure for analysis, simply select it with the file dialog which is activated by clicking on the Browse button below, then click the Send File button.

Form Based PDB File Upload:

Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc). A collection of good structures is used as a reference to obtain a score for each of the 20 amino acids in this structural class. The scores of a sliding 21-residue window (from -10 to +10) are added and plotted for individual residues.

[Obtain your own standalone copy of Profile Search/Environments program/Verify 3D](#)

Metaserwer: Saves

<http://services.mbi.ucla.edu/SAVES/>

PROCHECK	Checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. [Reference]
WHAT_CHECK	Derived from a subset of protein verification tools from the WHATIF program (Vriend, 1990), this does extensive checking of many stereochemical parameters of the residues in the model. [Reference]
ERRAT	Analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. [Reference]
VERIFY_3D	Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures. References: [Bowie et al., 1991] ; [Luethy et al., 1992] .
PROVE	Calculates the volumes of atoms in macromolecules using an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures. [PUBMED Reference] .
CRYST1 record matches	We take the CRYST1 record and search the entire PDB for matches and report these as possibly similar structures.
Ramachandran Plot	We produce an interactive Ramachandran plot. Also a standalone server linked above.
WedMol Viewer	We provide a structure viewer in the web page, although not all browsers support it. We are working to provide a more robust viewer

A user can run all 6 programs to get a collective view of the input structure, or individual programs can be selected. The evaluations the server makes of the outputs of each program are labeled by a simple 3 color scheme:

Modeller

<http://salilab.org/modeller/>

Program uruchamialny z poziomu pythona

Swiss-Model

<http://swissmodel.expasy.org/>

Modelowanie interaktywne (on-line)

3D-JIGSAW

<http://bmm.cancerresearchuk.org/~3djigsaw/>

(aktualnie niedostępny)

Bazy w pełni automatycznie wymodelowanych struktur dla dużej liczby białek (ok. 1/3 wszystkich znanych białek).

Modele te mają zwykle niższą jakość niż modele tworzone indywidualnie z udziałem człowieka.

SwissModel Repository:

<http://swissmodel.expasy.org/repository/>

ModBase:

<http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>

Rozpoznawanie zwoju

Fakt:

Struktury są znacznie bardziej konserwowane niż sekwencje.

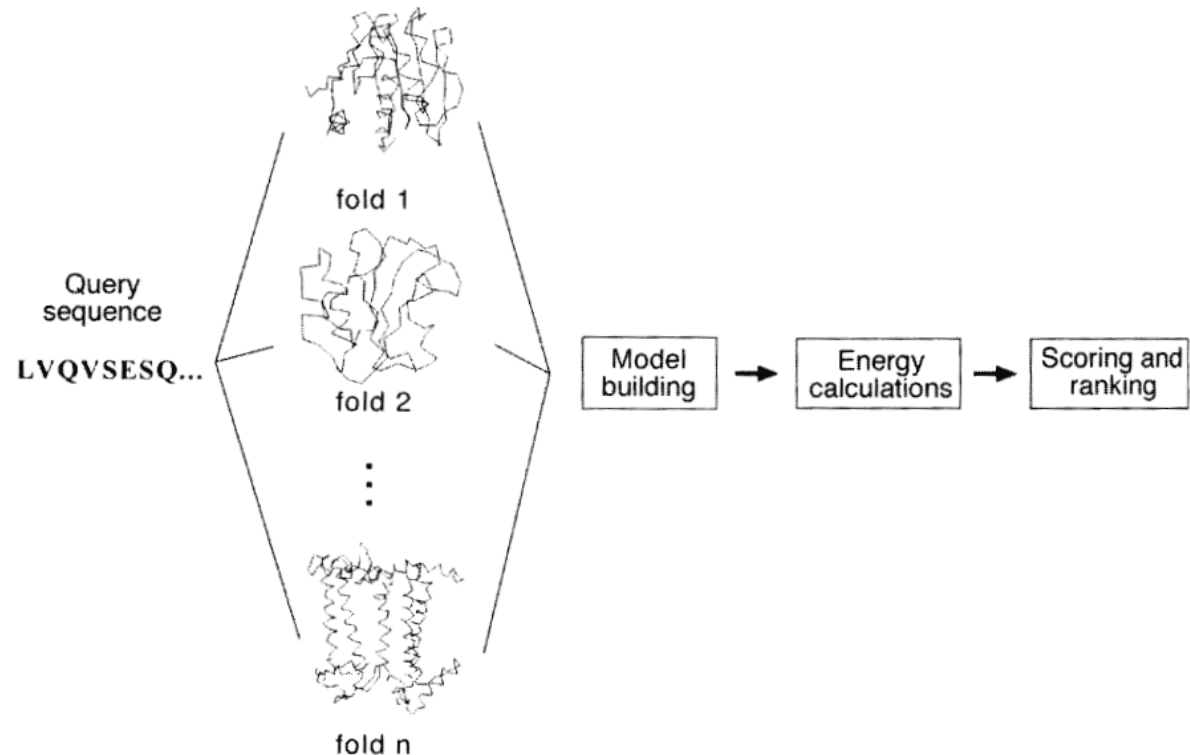
Znamy miliony sekwencji, ale tylko ok. 1200 różnych zwojów.

→ wiele białek ma ten sam zwój, pomimo braku podobieństwa sekwencyjnego.

1. Porównywanie energii parami (przewlekanie)
2. Analiza profili

Przegląd biblioteki zwojów:

- budowa modelu
- obliczenia energetyczne
- ocena, tworzenie rankingu



1. Wyższa czułość (pozwała odkryć nawet odległe homologii)
2. Niższa specyficzność
3. Nie działa dla nieznanych zwojów
4. Generuje niedokładne, niepełnoatomowe modele
5. Umożliwia jednak zgrubne oszacowanie topologii poszukiwanej struktury.

PHYRE2 (dawniej 3D-PSSM)

<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

PSIPRED

<http://bioinf.cs.ucl.ac.uk/psipred/>

Przewidywanie *ab-initio*

Białko w procesie fałdowania dąży do uzyskania struktury, dla której energia cząsteczki byłaby najmniejsza.

Problem:

znalezienie globalnego minimum energetycznego cząsteczki wymaga obliczenia energii dla wszystkich możliwych konformacji
→ niewykonalne

Próba rozwiązania:

symulacja procesu zwijania się białka.

Kompletny opis układu składa się z opisu białka i roztworu, oraz oddziaływań między nimi.

Teoretycznie poprawnym, zgodnym z fundamentalnymi prawami fizyki jest opis układu przy użyciu mechaniki kwantowej.

W przybliżeniu nierelatywistycznym układ można opisać równaniem Schroedingera:

$$i\hbar \frac{d\Psi(r,t)}{dt} = -\frac{\hbar^2}{2m} \Delta \Psi(r,t) + V\Psi(r,t)$$

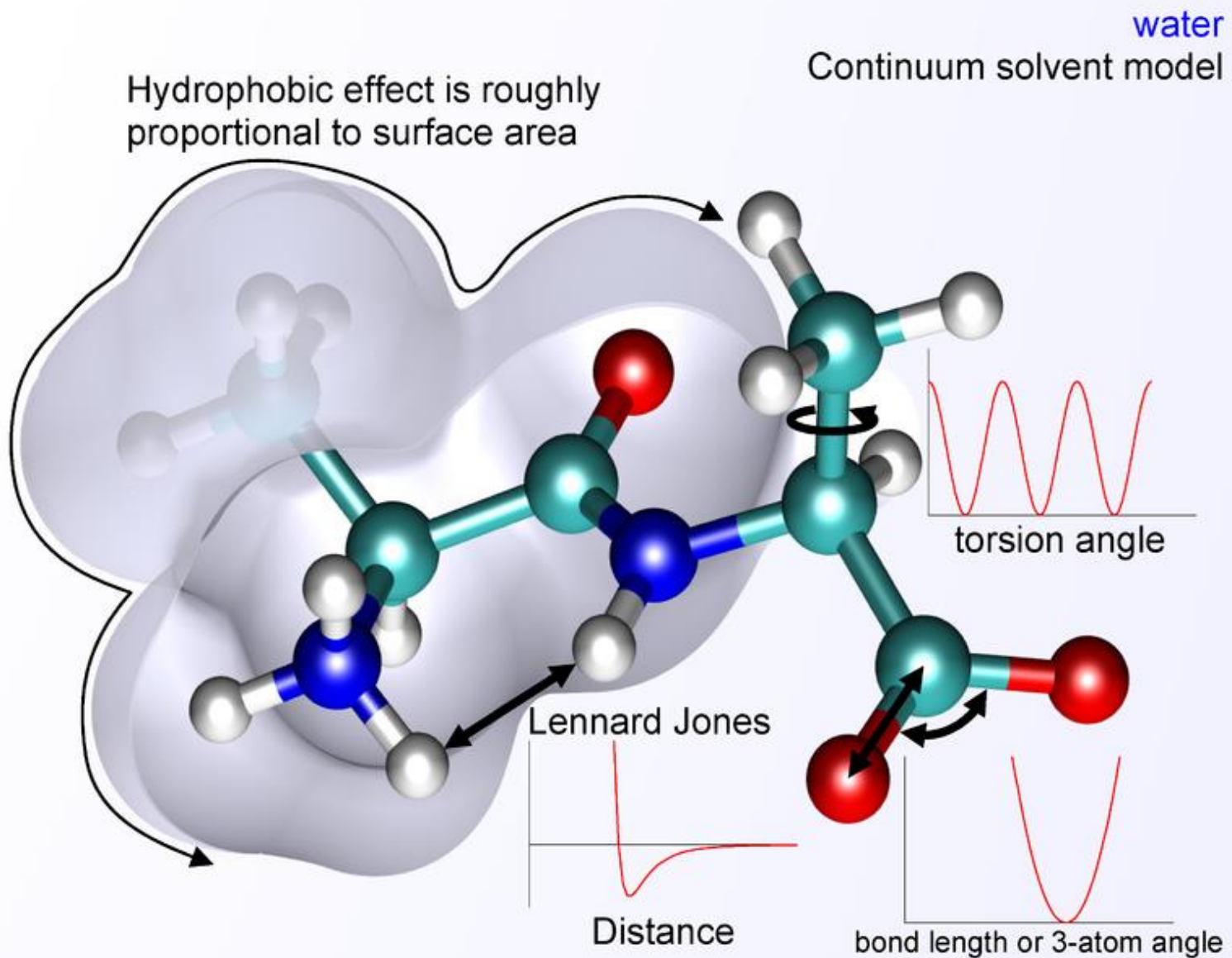
$$E = E_{\text{kowalencyjna}} + E_{\text{niekowalencyjna}}$$

$$E_{\text{kowalencyjna}} = E_{\text{wiazanie}} + E_{\text{kat}} + E_{\text{kat torsyjny}}$$

$$E_{\text{niekowalencyjna}} = E_{\text{elektrostatyczna}} + E_{\text{van der Waalsa}}$$

$$E_{\text{tot}} = \underbrace{E_l + E_\theta + E_\omega}_{\text{oddziaływania „wiązań”}} + \underbrace{E_{vdw} + E_{coulomb}}_{\text{oddziaływanie „nie-związane”}}$$

energia całkowita (entalpia)



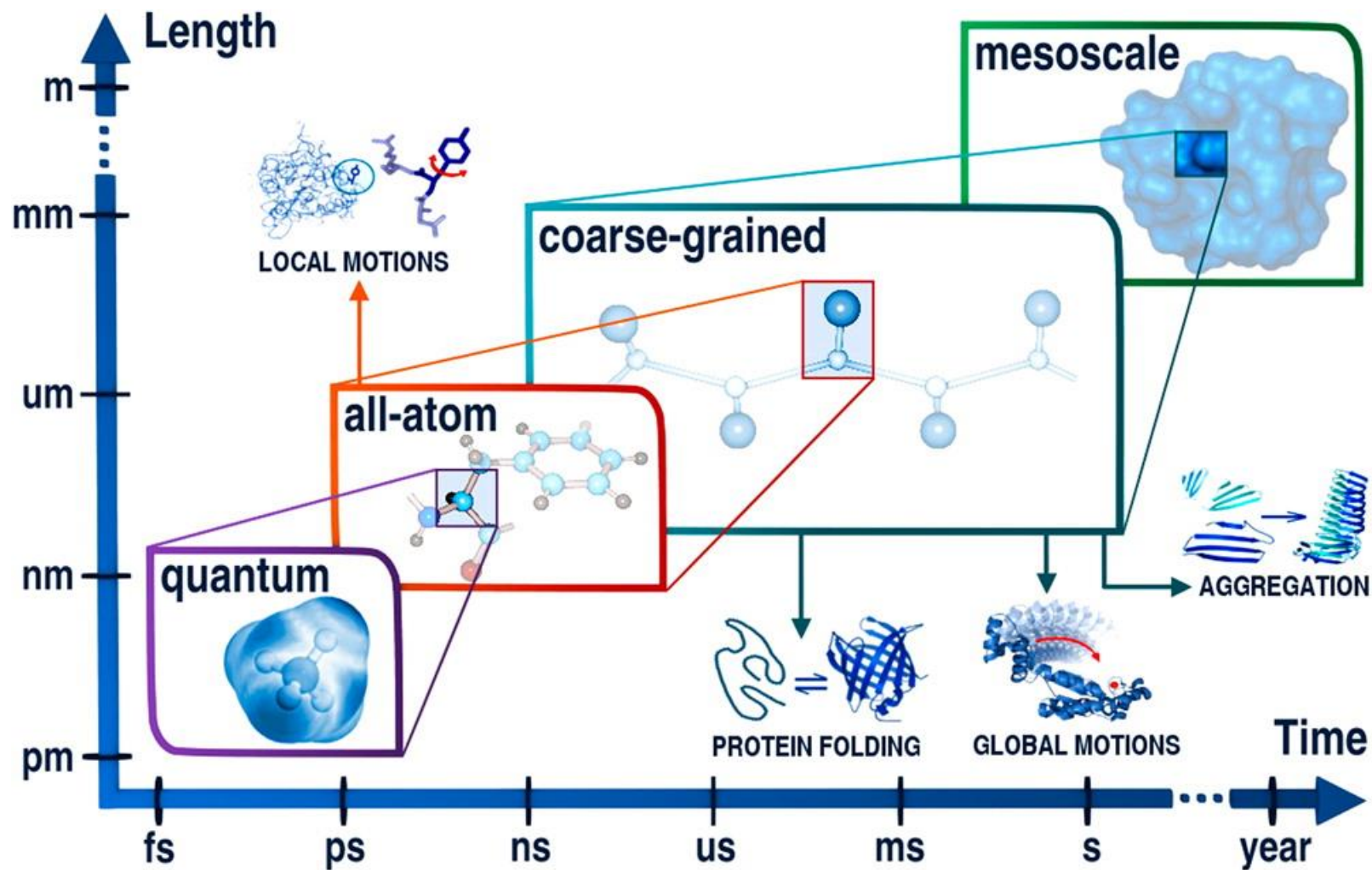
Funkcja energii potencjalnej może być parametryzowana na wiele sposobów. Różne parametryzacje są stosowane do różnych celów, występują również konkurencyjne parametryzacje dla tych samych zastosowań.

W żargonie modelarzy molekularnych parametryzacja funkcji energii potencjalnej nazywa się polem siłowym (*force field*).

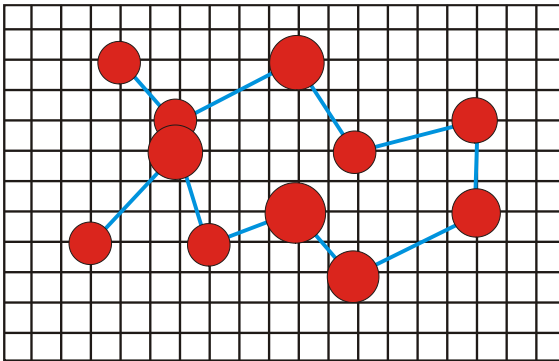
W badaniach obiektów biologicznych stosowane są m.in. pola siłowe:

- AMBER (Kollman i współpracownicy, UCSF)
- CHARMM (Karplus i współpracownicy, Harvard)
- Gromos (Berendsen i współpracownicy, Nijmegen)
- CVFF i CFF9X (Accelrys)
- TRIPOS (Tripos)

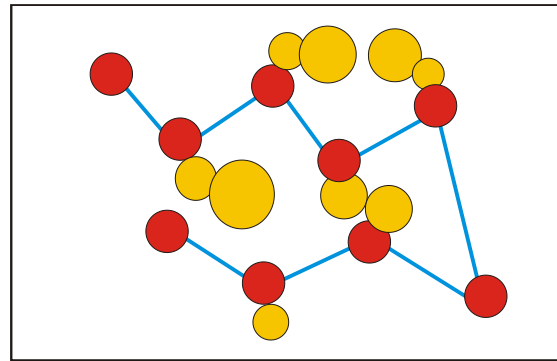
Skala modelowania



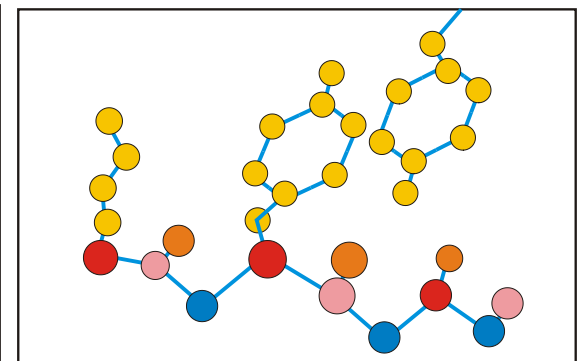
**Uprozczone
modele siatkowe**



**Uprozczone
modele ciągłe**



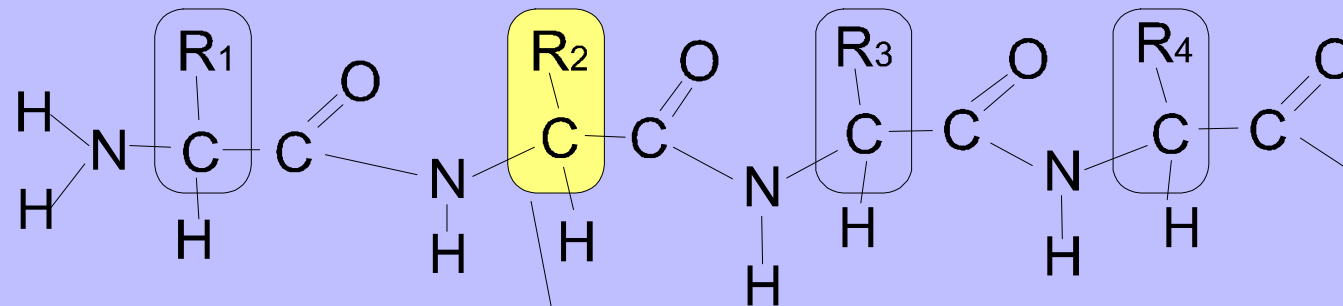
**Pełnoatomowe
modele ciągłe**



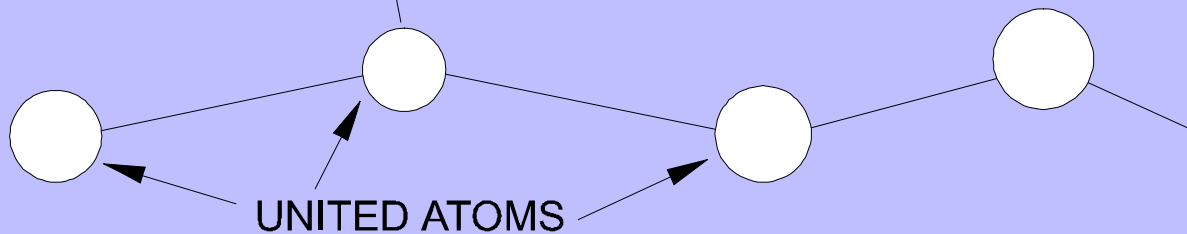
SZCZEGÓŁOWOŚĆ REPREZENTACJI



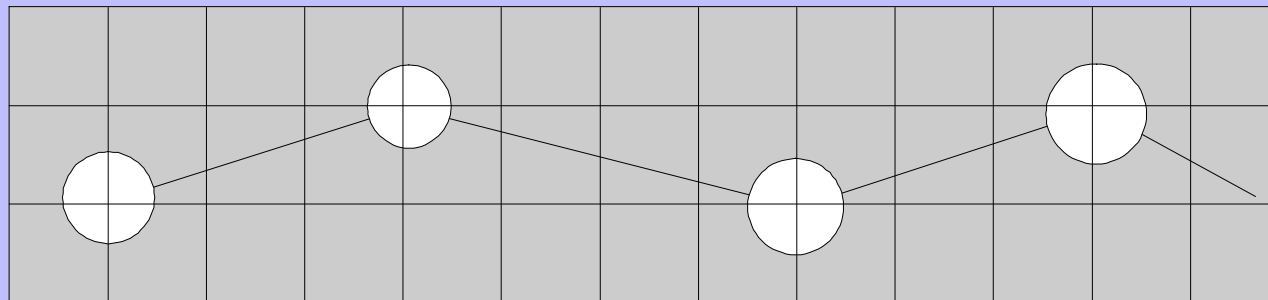
WYMAGANIA OBLICZENIOWE



FULL ATOMIC
MODEL

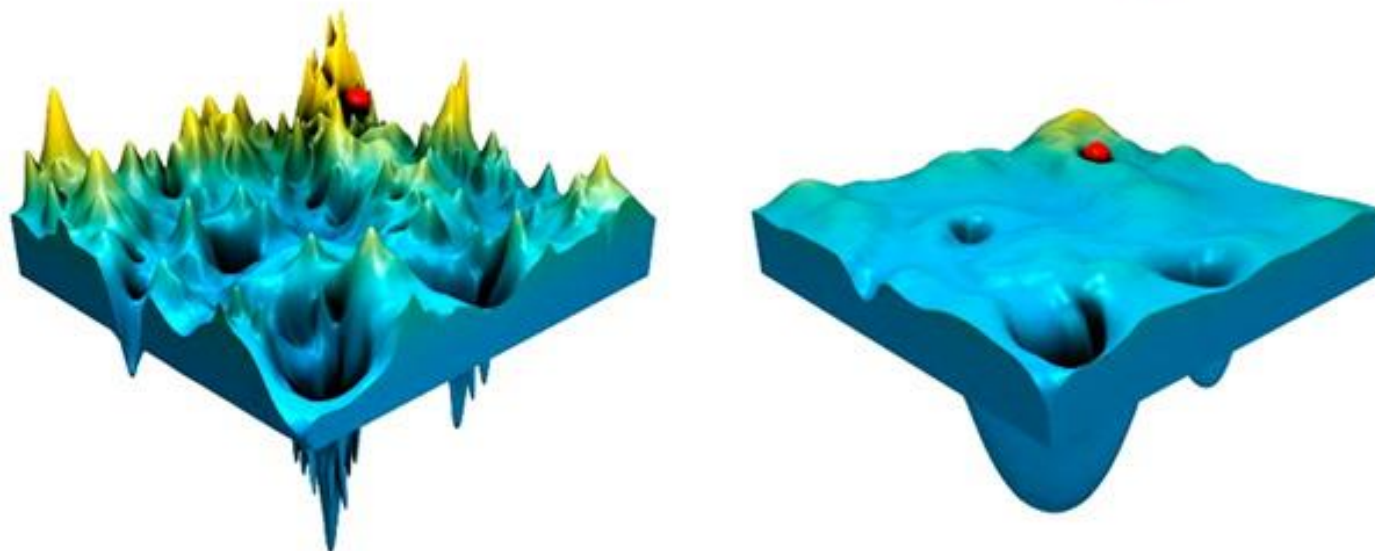
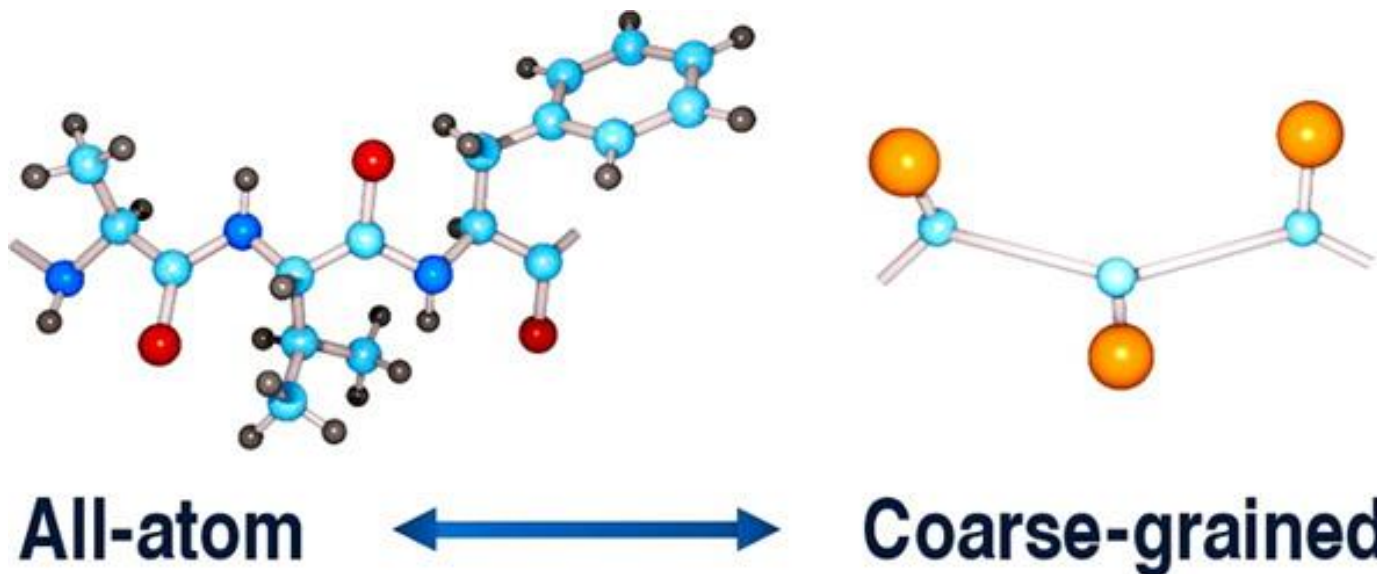


REDUCED
MODEL

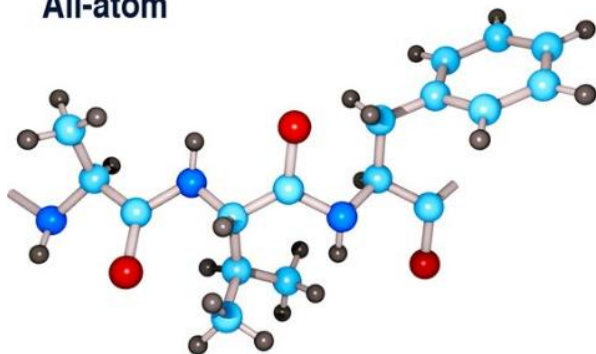


REDUCED
LATTICE
MODEL

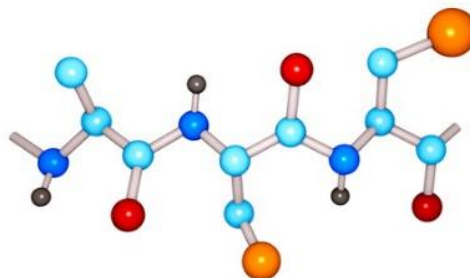
Różnica pomiędzy modelem pełnoatomowym a gruboziarnistym



All-atom

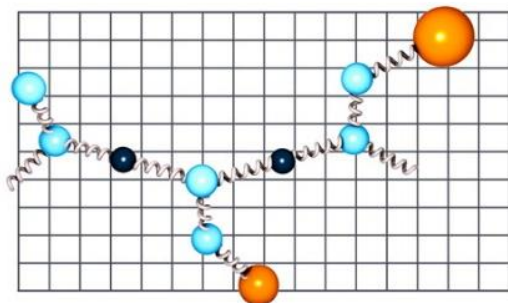


Rosetta CEN

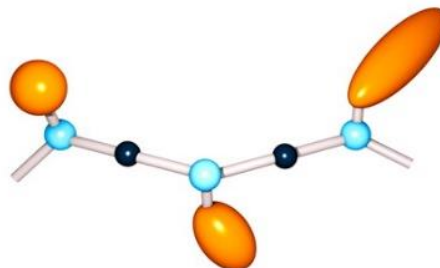


All-atom representation of a tripeptide and the corresponding coarse-grained models. Various coarse-grained models are presented: Rosetta centroid mode (CEN) representation, CABS, UNRES, SICHO, and Levitt and Warshel model. United side chain atoms are colored in orange. Pseudobonds of fluctuating length are shown as springs and lattice models are shown on the underlying lattice slide.

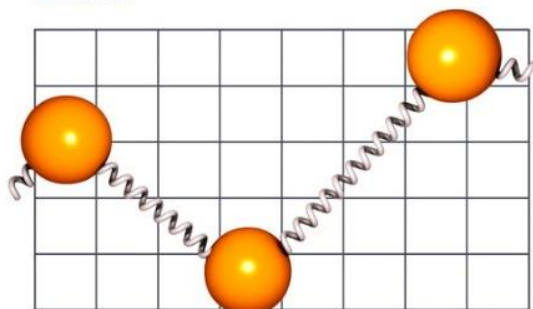
CABS



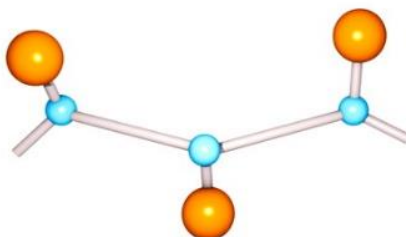
UNRES

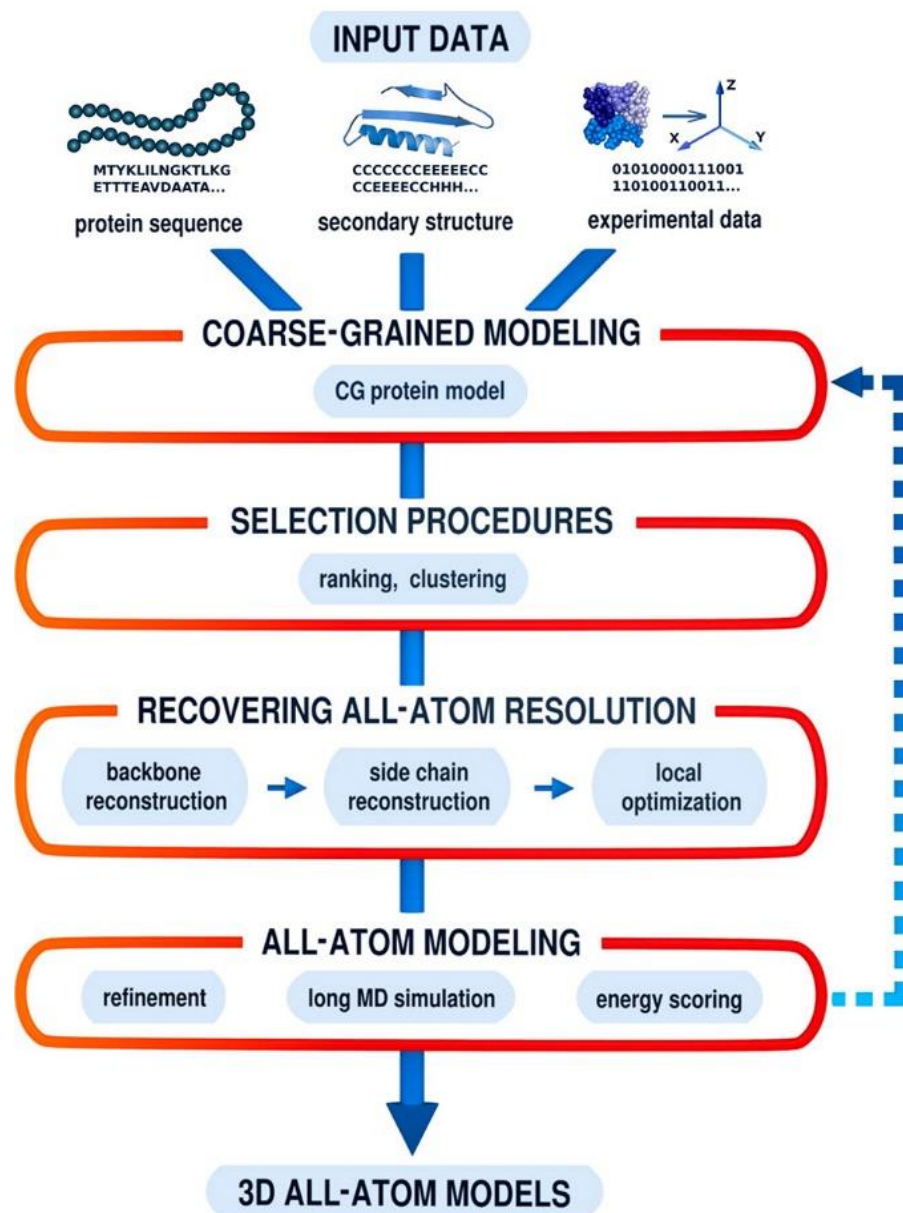


SICHO



Levitt-Warshel

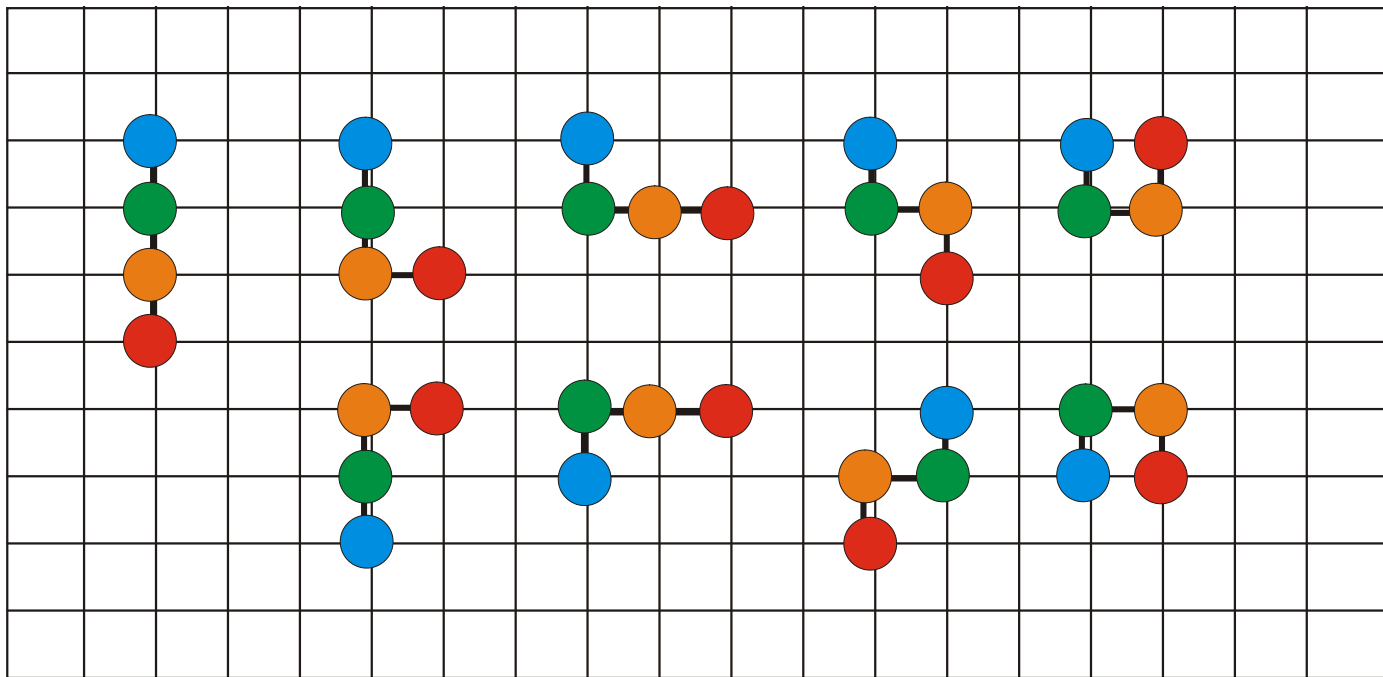




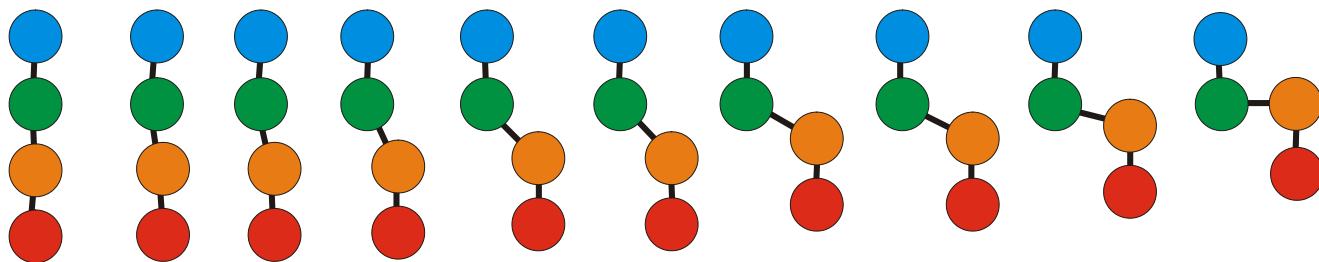
Typical multiscale modeling scheme that merges coarse-grained and all-atom modeling.

In specific tasks, the resulting all-atom structures could be used as an input for the next stage of coarse-grained simulations.

Skończona przestrzeń konformacyjna



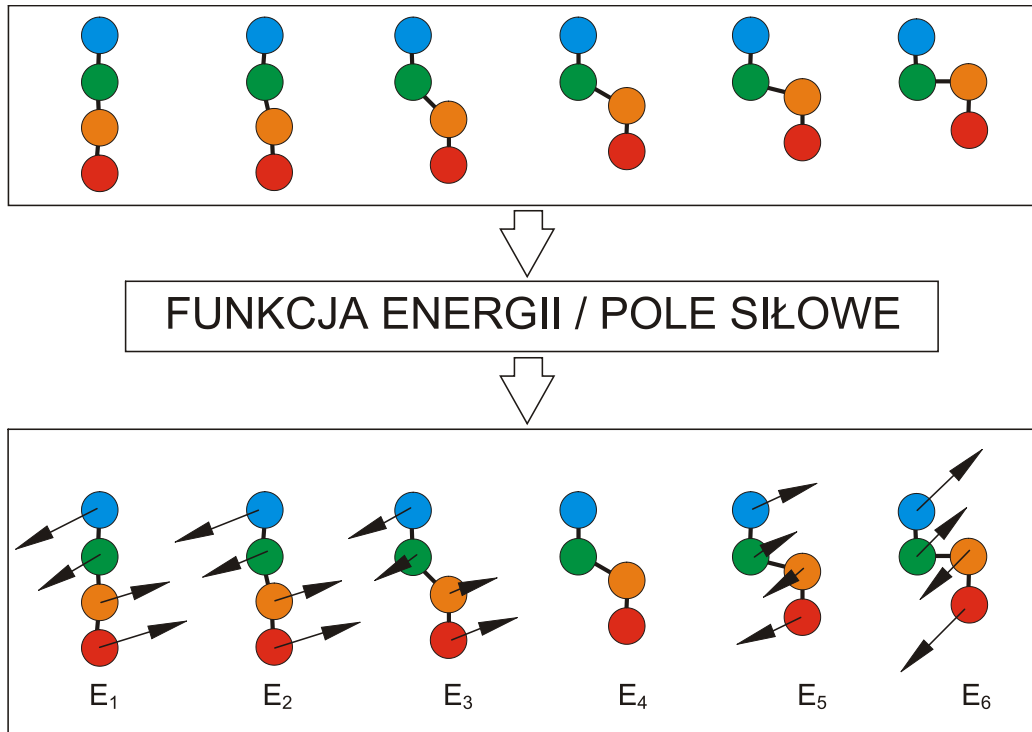
Nieskończona przestrzeń konformacyjna



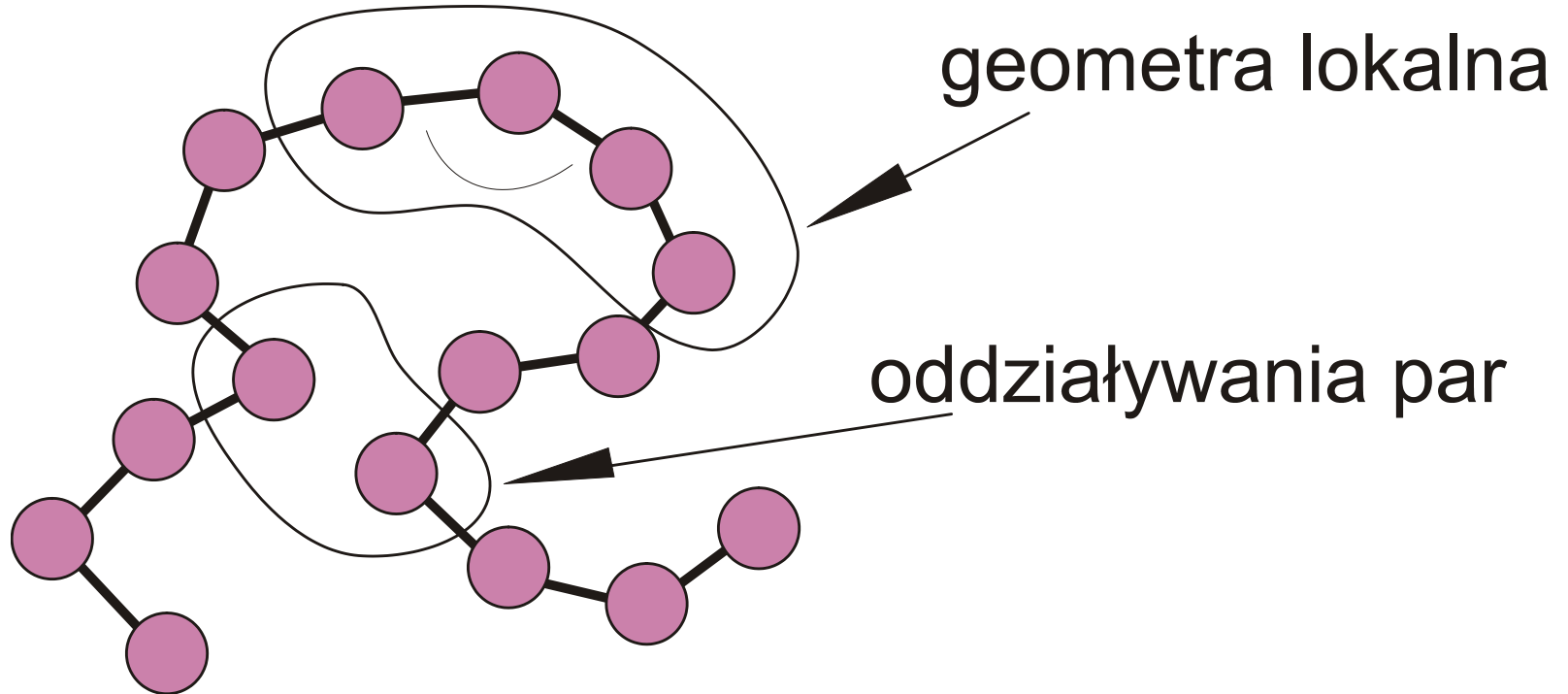
Funkcja energii

Idealna funkcja energii powinna posiadać globalne minimum dla konformacji natywnej

Energia konformacyjna powinna rosnąć wraz ze zmniejszającym się podobieństwem danej konformacji do konformacji natywnej



- człony bliskiego zasięgu - geometria lokalna
- człony dalekiego zasięgu (oddziaływania par)



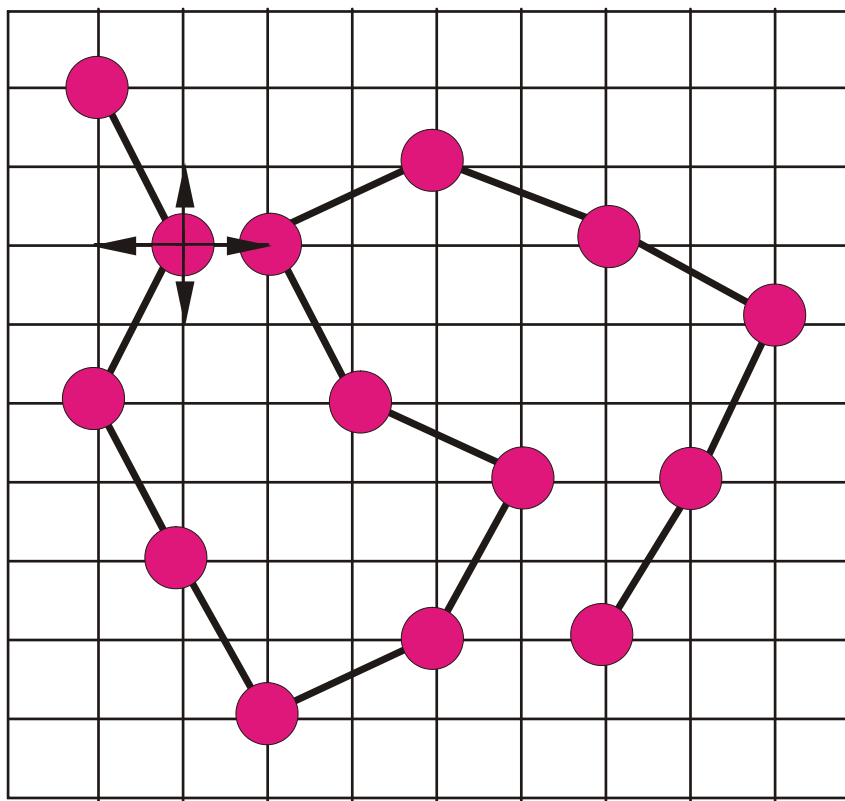
Ze względu na sposób wyprowadzania potencjały dzieli się na:

- statystyczne - wyprowadzone na podstawie bazy danych struktur natywnych;
- fizyczne - wyprowadzone na podstawie fundamentalnych praw fizyki.

Potencjały mogą być też wyznaczone na podstawie danych eksperymentalnych.

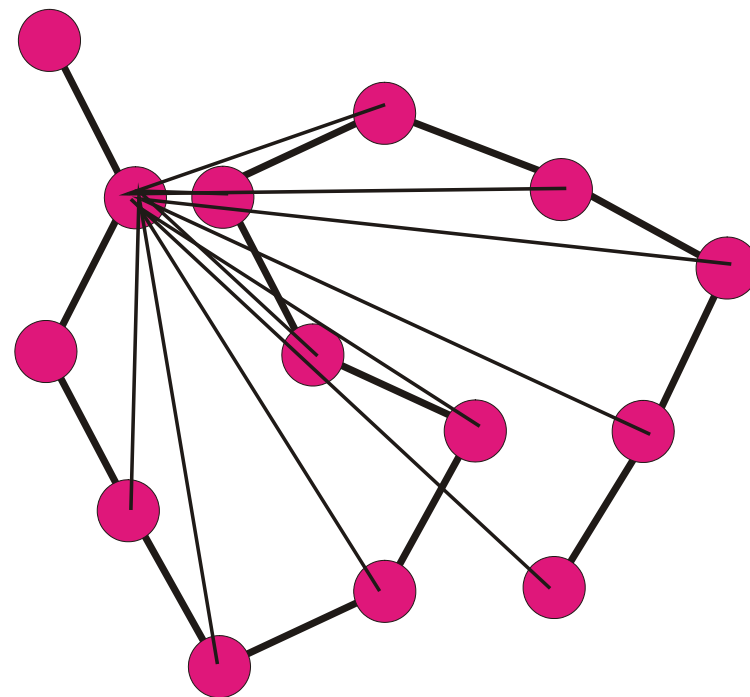
W ogólności istotne jest uwzględnienie reprezentacji łańcucha.

Koszt obliczenia oddziaływań – znaczenie siatki



Atom w położeniu $[i,j]$
sprawdzenie położenia sąsiednich:
 $[i+1,j]$, $[i-1,j]$, $[i,j-1]$, $[i,j+1]$

Koszt obliczenia energii dla
pojedynczego atomu jest
niezależny od długości łańcucha



Dla danego atomu trzeba policzyć
prawie n odległości.

Koszt obliczenia energii dla
pojedynczego atomu jest liniowo
zależny od długości łańcucha. Możli-
we jest obniżenie kosztu do $\log(n)$

1. Atomy jako punkty materialne
2. Oddziaływania typu atom-atom opisywane odpowiednio sparametryzowaną funkcją energii potencjalnej.
3. Badany za pomocą trzech klas metod:
 - I. Mechanika molekularna (MM)
 - II. Dynamika molekularna (MD)
 - III. Monte Carlo (MC)

Porównywanie struktur

1. Oparta na odległościach atomów między cząsteczkami (RMSD).
2. Oparta na odległościach atomów wewnątrz cząsteczek.
3. Metoda mieszana.



Root-mean-square deviation (RMSD)

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2}$$

N – liczba odpowiadających sobie reszt;

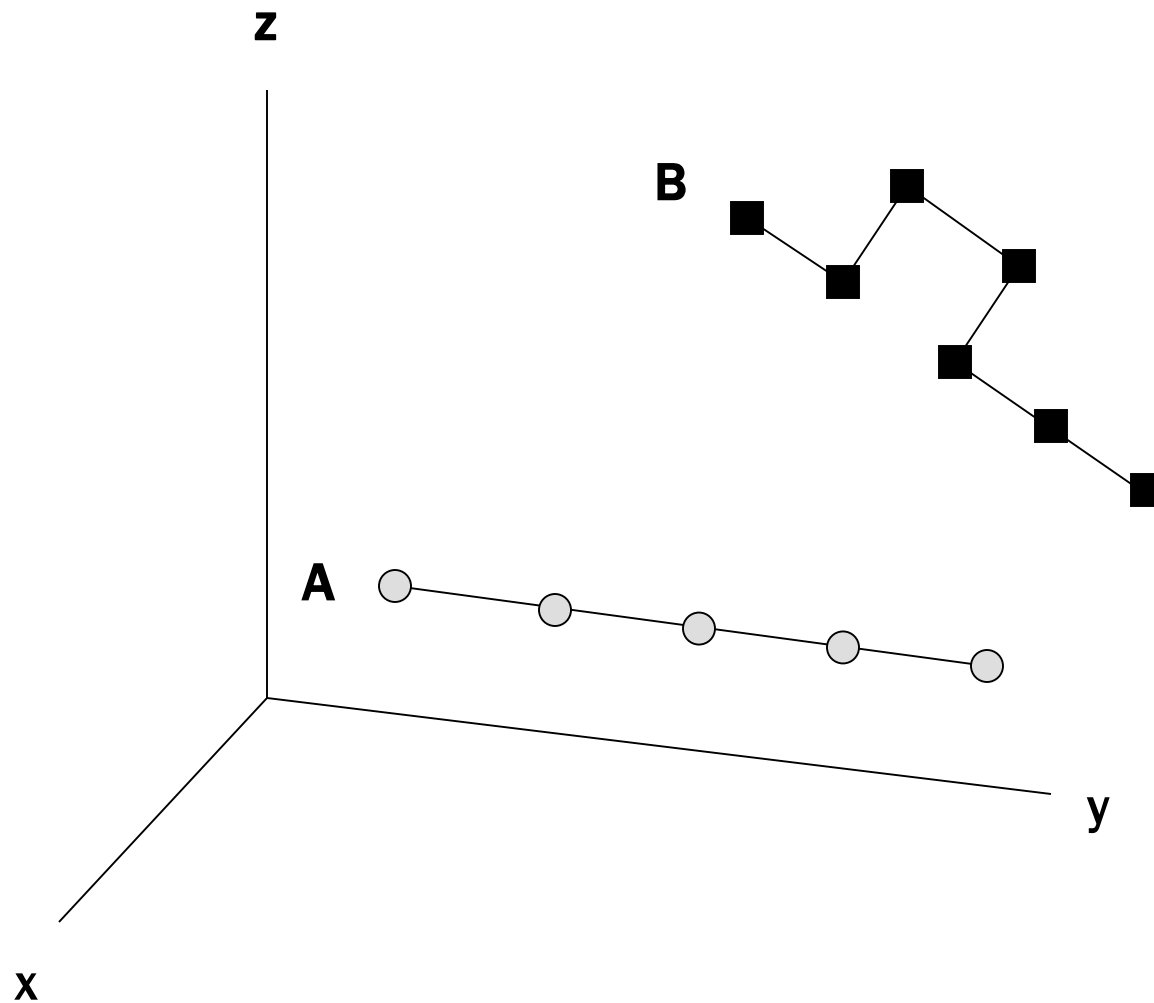
δ – odległość pomiędzy wybranymi atomami (zwykle C α) odpowiadających sobie reszt.

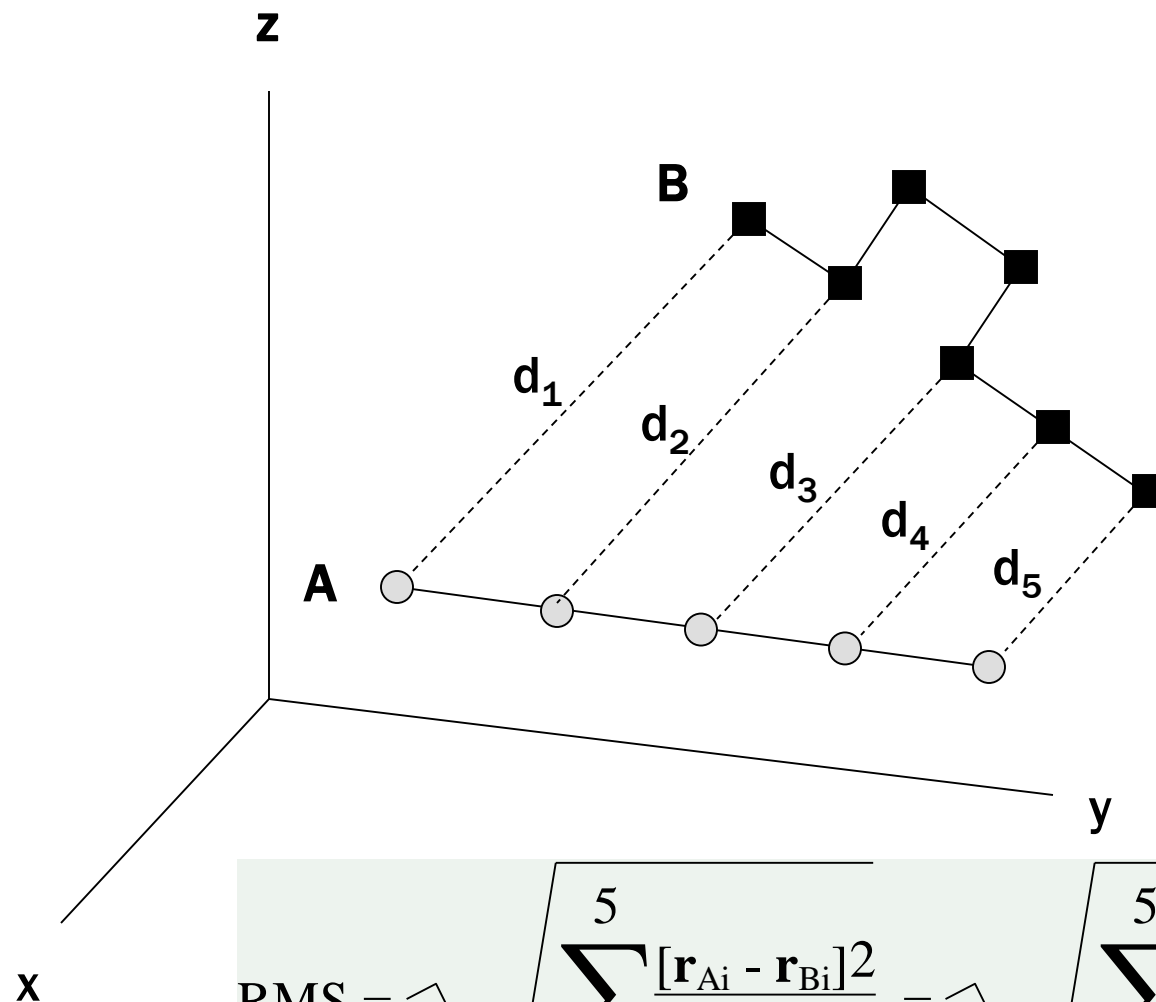
RMSD_{100} – korekcja ograniczająca wpływ wielkości białka na wartość miary.

$$\text{RMSD}_{100} = \text{RMSD} / (-1,3 + 0,5 \ln(N))$$

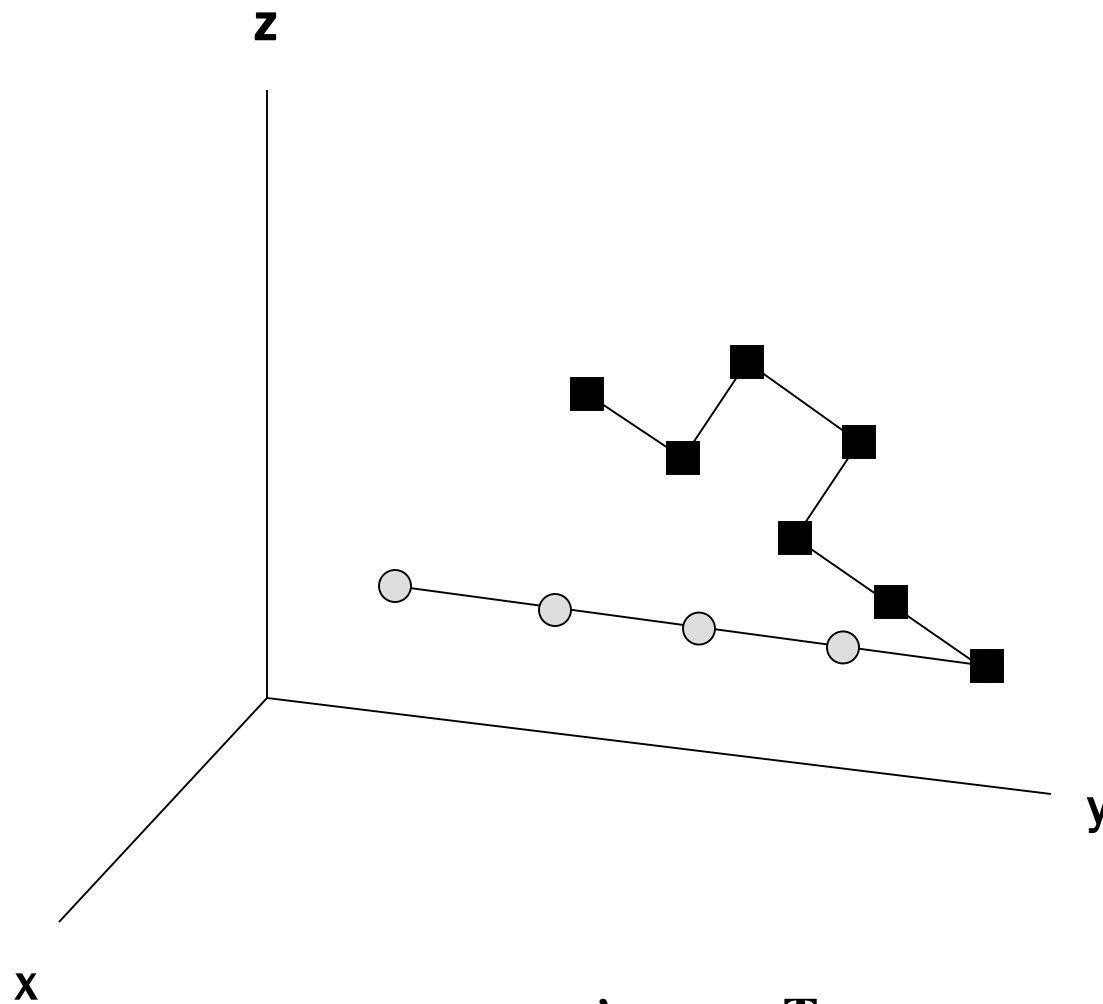
N – liczba odpowiadających sobie atomów

Nałożenie bazujące na RMSD

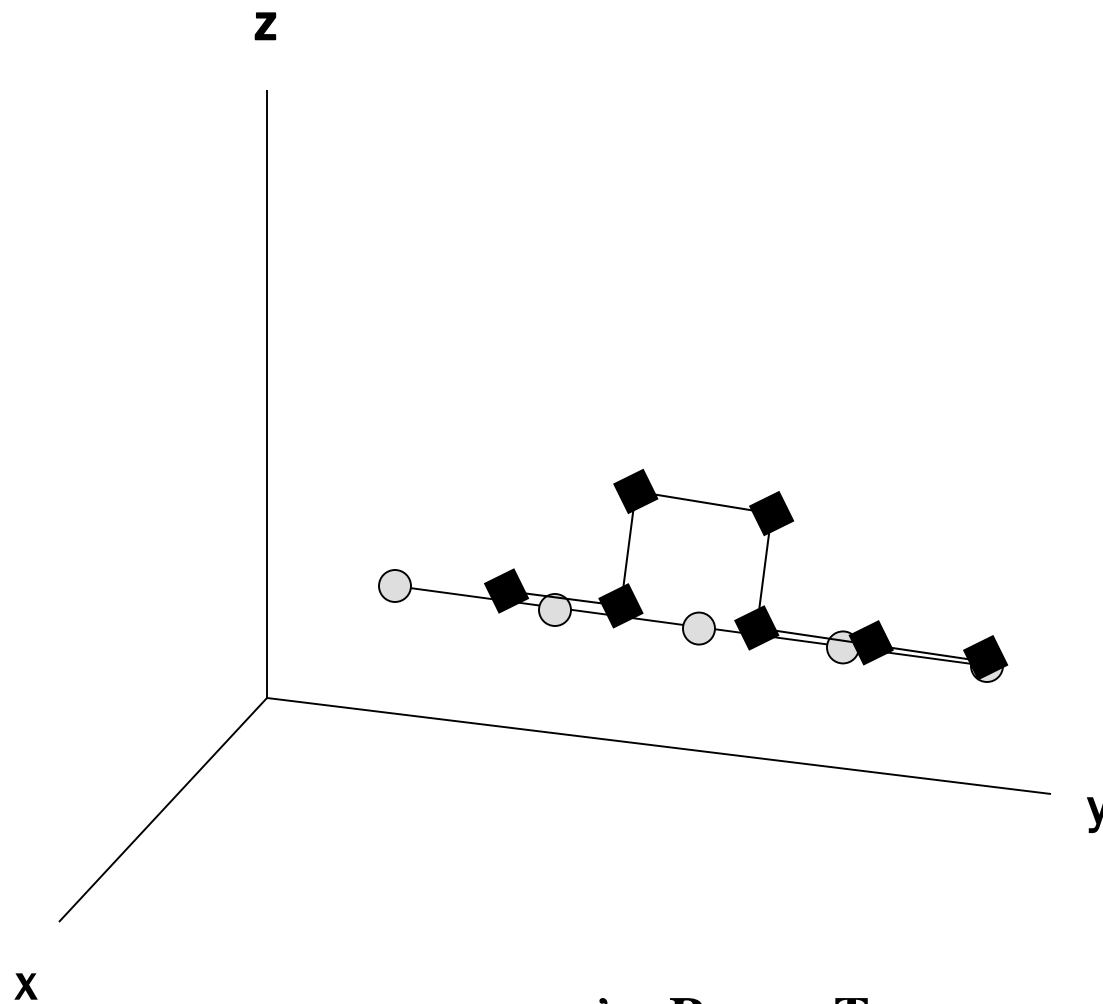




$$\text{RMS} = \sqrt{\sum_{i=1}^5 \frac{[\mathbf{r}_{Ai} - \mathbf{r}_{Bi}]^2}{5}} = \sqrt{\sum_{i=1}^5 \frac{d_i^2}{5}}$$

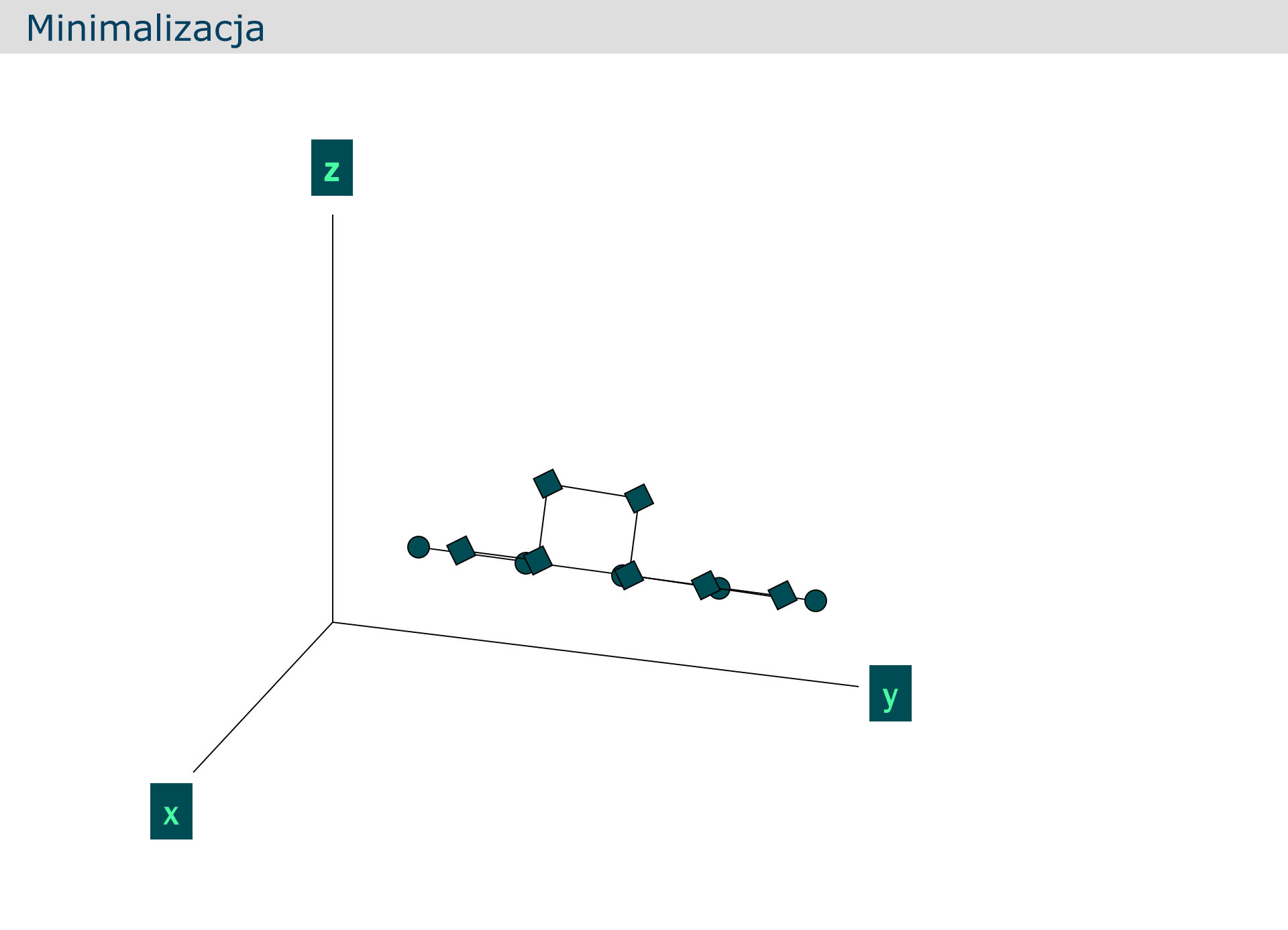


$$\mathbf{r}_{Bi}' = \mathbf{r}_{Bi} + \mathbf{T}$$



$$\mathbf{r}_{Bi}' = \mathbf{R} \mathbf{r}_{Bi} + \mathbf{T}$$

Minimalizacja



nałożenie (*superimpose*) – znamy odpowiadające sobie pary aminokwasów (stosunkowo proste)

dopasowanie (*alignment*) – musimy odnaleźć odpowiadające sobie fragmenty (trudne)



Obserwacja:

Podobnie pofałdowane białka nie muszą zachowywać widocznego podobieństwa sekwencji (strefa zmierzchu).

Dlatego identyfikacja odpowiadających sobie reszt na podstawie przyrównania sekwencji (np. algorytmem N-W) nie jest najlepszym pomysłem.

Rozwiązania praktyczne:

- ograniczenie przestrzeni poszukiwań przez usunięcie fragmentów nie tworzących struktury drugorzędowej
- podział struktury na mniejsze fragmenty (6-9 reszt) i dopasowywanie ich po kolei
- optymalizacja interaktywna: szkielet na podstawie przyrównania sekwencji; uzupełnianie w drodze analizy przestrzennej

Distance based methods :

DALI (Holm and Sander): Aligning scalar distance plots

STRUCTAL (Gerstein and Levitt): Dynamic programming using pair wise inter-molecular distances

SSAP (Orengo and Taylor): Dynamic programming using intramolecular vector distances

MINAREA (Falicov and Cohen): Minimizing soap-bubble surface area

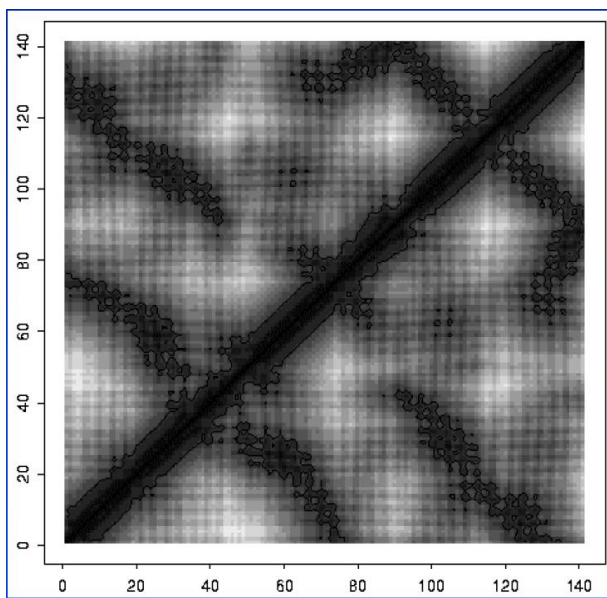
Vector based methods:

VAST (Bryant): Graph theory based secondary structure alignment

3dSearch (Singh and Brutlag): Fast secondary structure index lookup

Both vector and distance based:

LOCK (Singh and Brutlag): Hierarchically uses both secondary structure vectors and atomic distances



- Dopasowuje macierze odległości wewnątrzcząsteczkowych
- Oblicza najlepszy podzbiór odpowiadających sobie reszt na podstawie maksymalizacji podobieństwa między macierzami
- Przy przeszukiwaniu przestrzeni możliwych dopasowań wykorzystywana jest metoda Monte-Carlo.

http://ekhidna.biocenter.helsinki.fi/dali_server/start

Dali server

Institute of Biotechnology

SERVICES & TOOLSGROUP MEMBERSNEWS & VACANCIESRESEARCHPUBLICATIONS

Protein Structure Database Searching by DaliLite v. 3

The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank (PDB). You receive an email notification when the search has finished. In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

Requests can also be submitted by e-mail to *dali-server at helsinki dot fi*. The body of the e-mail message must contain atomic coordinates in PDB format.

If you want to know the structural neighbours of a protein already in the Protein Data Bank (PDB), you can find them in the [Dali Database](#).

If you want to superimpose two particular structures, you can do it in the [pairwise DaliLite](#) server.

Upload a structure:

Przeglądaj...

Or enter PDB identifier: chain: (optional)
([Keyword search](#) for PDB identifiers)

Job name:
 (optional)

Enter email address for notification:
 (recommended)

☐ lower priority queue

<http://ekhidna.biocenter.helsinki.fi/dali/start>

Dali Database

Institute of Biotechnology

SERVICES & TOOLSGROUP MEMBERSNEWS & VACANCIESRESEARCHPUBLICATIONS

Dali structural neighbours

The Dali Database is based on all-against-all 3D structure comparison of protein structures in the Protein Data Bank (PDB). The structural neighbourhoods and alignments are automatically maintained and regularly updated using the [Dali](#) search engine.

- Please note that PDB structures released after the last update will not be in the database! If you wish to find structural neighbours of these proteins, you are advised to submit the structure to the [Dali Server](#) instead.
- If you want to superimpose two particular structures, you can do it in the [pairwise DaliLite](#) server.

* Last Update: 7 March 2011
Update frequency: twice a year

Enter PDB identifier: chain: (optional)

([Keyword search](#) for PDB identifiers)

Dali Database entries are retrieved on demand, and formatting the results page may take up to one minute. Return visits to an existing results page are much faster.

Example

Structural neighbours of [1tu9](#), a globin-like protein in bacteria. [Tutorial](#)

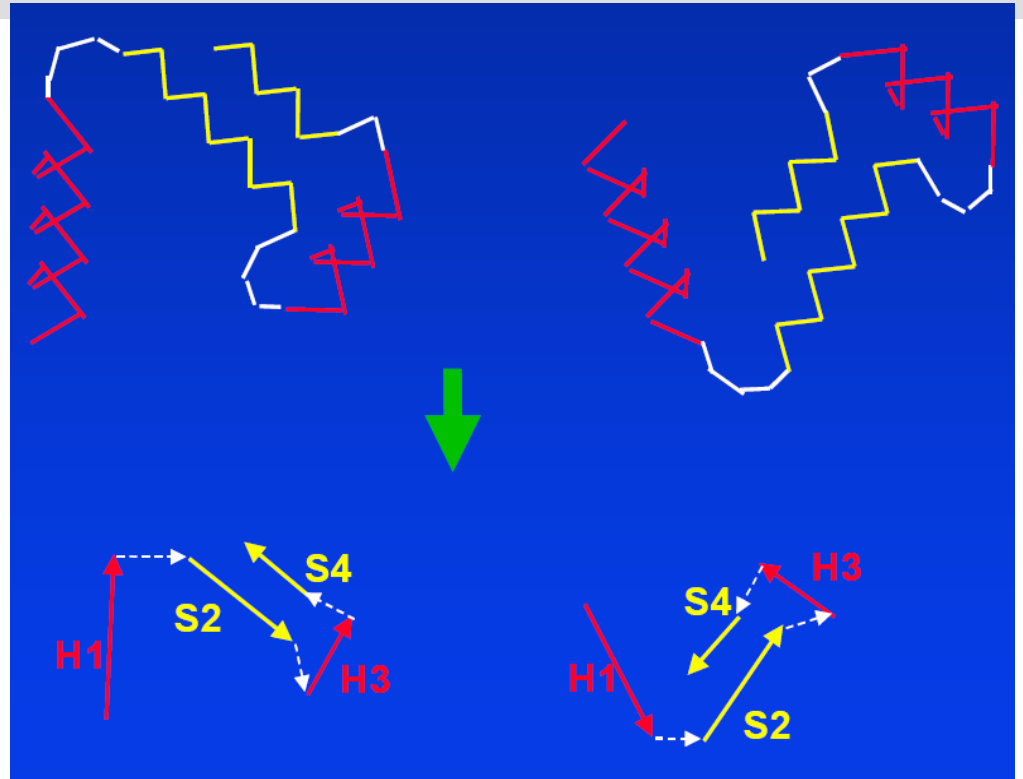
Downloading

Usage statistics

Reference

Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. [Nucl. Acids Res.](#) 38, W545-549.

Last Updated: Saturday 2nd April 2011, 23:55:23 GMT+2
Pages designed by Christopher Wilton, maintained by [webmaster](#)
© University of Helsinki, 2007



1. Dopasowuje elementy struktury drugorzędowej (*SSE - secondary structure elements*)
2. Każda struktura reprezentowana jest jako wektor
3. Znajduje wszystkie pary podobnych wektorów
4. Za pomocą algorytmu grafowego identyfikuje największy podzbiór podobnych wektorów.
5. Oblicza wartość dopasowania bazującą na liczbie podobnych par wektorów pomiędzy dwiema strukturami.

NCBI Structure

Structure Home 3D Macromolecular Structures Conserved Domains PubChem BioSystems

Search Entrez Structure for Go

VAST Help

Comprehensive help and frequently asked questions

VAST Search

Submit structure database search

VAST Search Help

Help on submitting VAST Searches

VAST Search FAQ

More help on VAST Search

Linking to VAST

direct WWW access to the VAST server

nr-PDB

non-redundant protein structure subsets

MMDB

NCBI's structure database

Vector Alignment Search Tool try:

Protein structure neighbors in Entrez are determined by direct comparison of 3-dimensional protein structures with the VAST algorithm. Each of the more than 87,804 domains in MMDB is compared to every other one. From the MMDB Structure

Structure Summary via PDB/MMDB Code: Get

<http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>

know a PDB/MMDB-Id you can try this at once, using the input form in the right column.

On the Structure summary page, use "3d Domains" or "Protein" to retrieve a list of similar structures. For example, click on a bar with a chain identifier such as "B", or the bar below the Chain B with a domain identifier such as "1", to get a list of neighbors. The results of the precompiled VAST search will then present structural neighbors graphically. Using the check boxes in the leftmost column of this graph, select those structures you would like to see superimposed and click on "View 3D Structure" to view these with the mime-typed helper application you have installed (e.g., Cn3D).

Install and test structure alignment viewers:

[Get Cn3D v4.1](#) and [look at this example](#) to test!

[Read a bit more about VAST...](#)

VAST Search is a service that allows searching for structural neighbors starting with a set of 3D-coordinates specified by the user. This service is meant to be used with newly determined protein structures that are not yet part of MMDB. Structure neighbors for proteins already in MMDB have been pre-computed and can simply be looked up from [MMDB's](#) Structure summary pages!

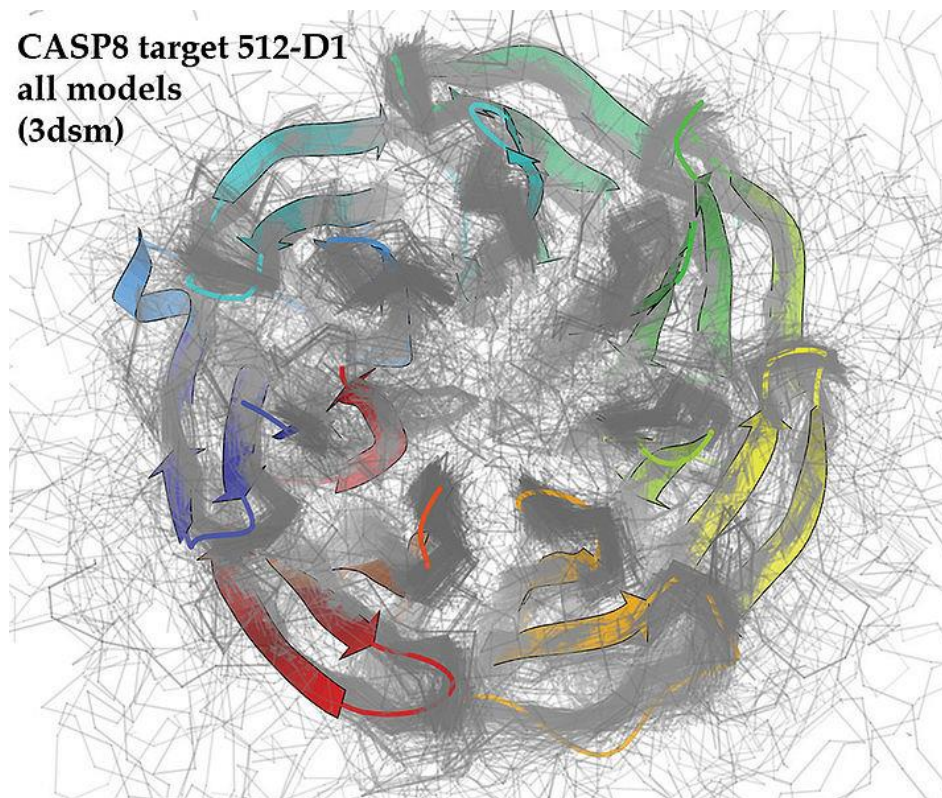
CASP

CASP (Critical Assessment of techniques for protein Structure Prediction)

C
A
S
P
10



CASP8 target 512-D1
all models
(3dsm)



<http://predictioncenter.org/>

Scope

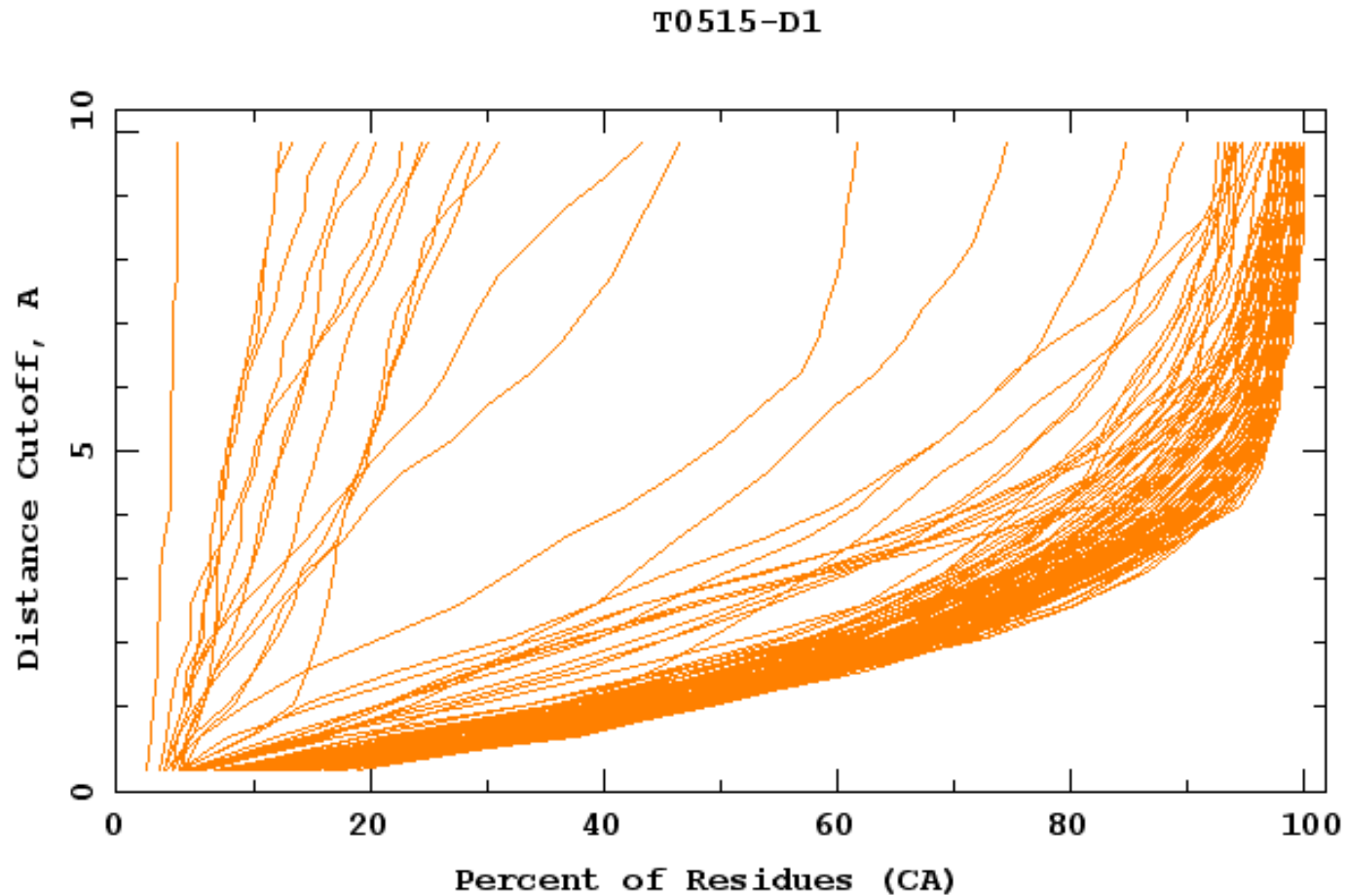
Tertiary structure predictions (TS):

- The *Template Based Modeling (TBM)* category will include domains where a suitable template can be identified that covers all or nearly all of the target.
- The *Template free modeling (FM)* category will include models of proteins for which no suitable template can be identified.
- The *Refinement* category will include selected targets from among those released in the main modeling experiment to analyze success in refining models beyond the quality obtained by simply copying from a single template. For suitable CASP10 targets, we will select one of the best models received during the prediction season, and reissue it as a starting structure for refinement.
- The *contact-assisted structure modeling* category will show how the knowledge of a few (usually 3 to 5) long-range contacts influences the ability of predictors to model the complete structure. This experiment will be carried out only for the more challenging CASP10 targets where we can get coordinates in advance and have at least two weeks for re-prediction.
- The *chemical shifts guided modeling* of NMR structures will be performed on the selected CASP10 targets, for which we can get a chemical shifts table from the NMR-spectroscopists at least two weeks ahead of public release of the target.
- The *structure modeling based on molecular replacement with ab initio models and crystallographic diffraction data* will be carried out for selected targets provided we get the structure factors from the crystallographers.

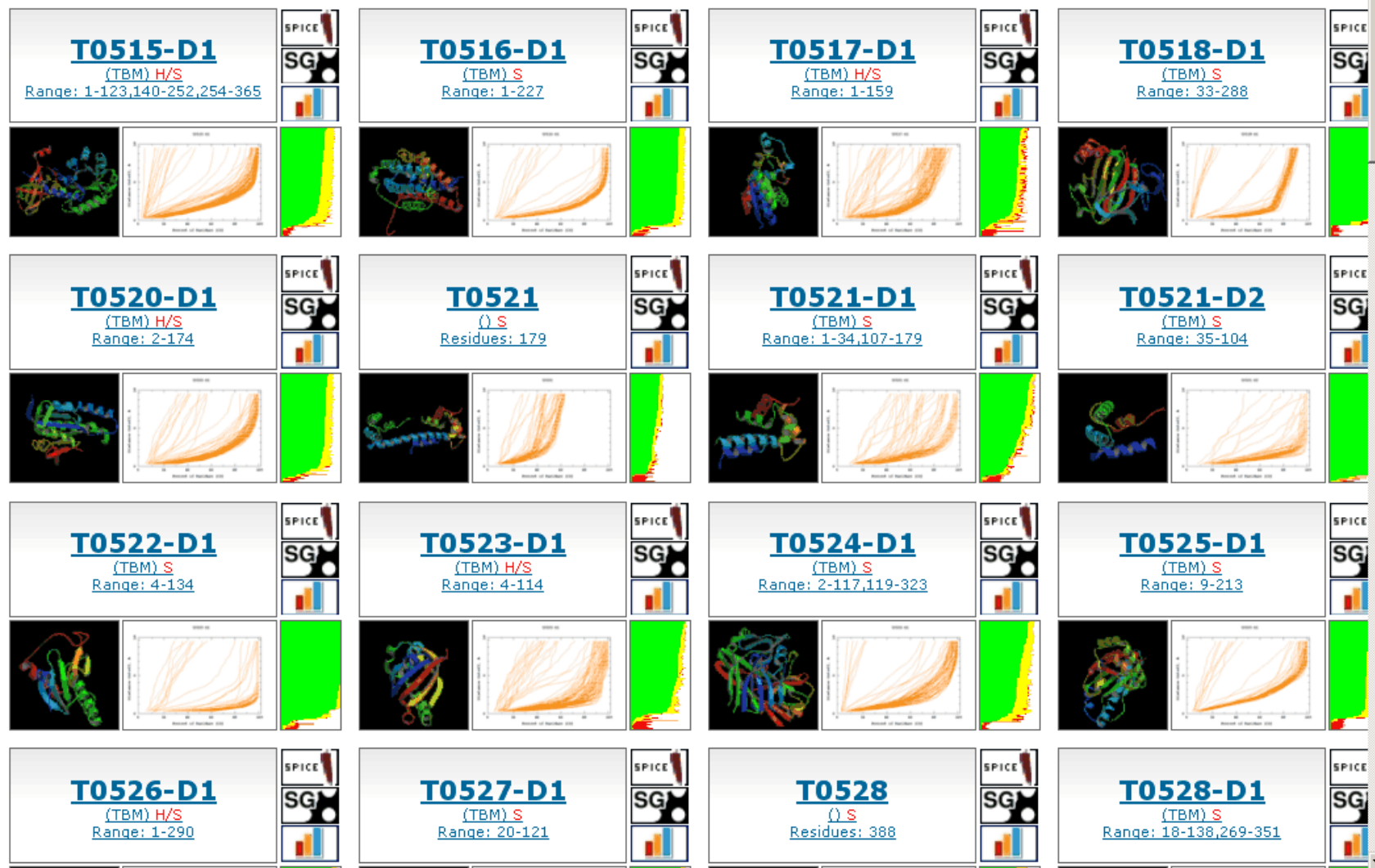
Other prediction categories:

- Detecting residue-residue contacts in proteins (RR).
- Identifying disordered regions in target proteins (DR).
- Function prediction (prediction of binding sites) (FN).
- Quality assessment of models in general (without knowing native structures) and the reliability of predicting certain residues in particular (QA).

<http://predictioncenter.org/casp9/results.cgi>

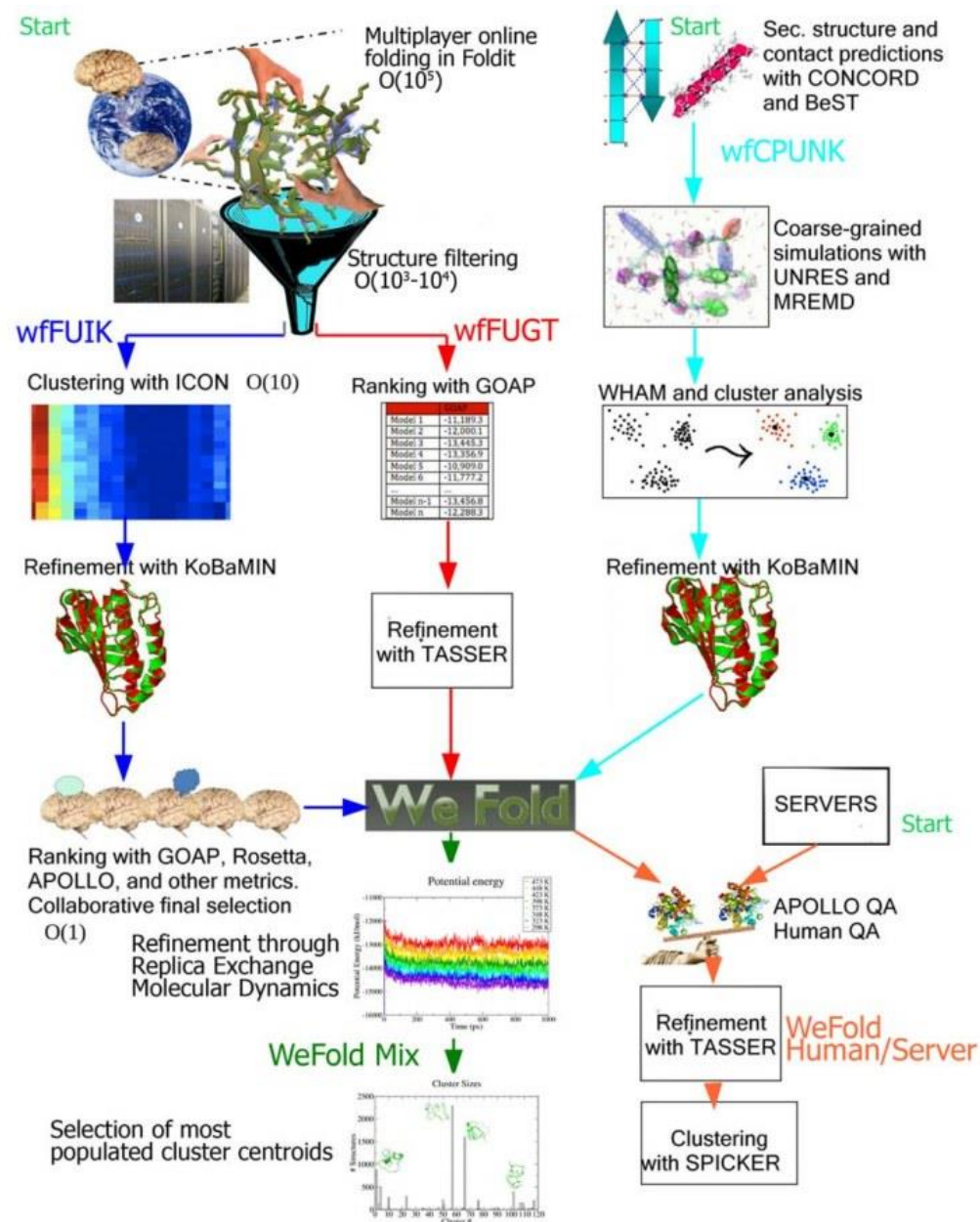


<http://predictioncenter.org/casp9/results.cgi>



Portal współpracy i
współzawodnictwa.

Skojarzony z konkursem CASP.



Projekty rozproszone

Protein folding simulation



Protein structure prediction.



Czym zajmuje się Rosetta@home?



Rosetta@home potrzebuje twojej pomocy, aby ustalić trójwymiarowy model białka potrzebny do badań, które mogą w przyszłości pomóc w wynalezieniu leków na choroby ludzkości. Uruchamiając aplikację Rosetty na swoim komputerze pomożesz nam w przyspieszeniu i poszerzeniu naszych badań, które bez twojego udziału nie byłyby możliwe. Możesz także pomóc nam w staraniach nad stworzeniem nowych białek, które posłużą do walki z chorobami nękającymi ludzkość takimi jak HIV, Malaria, Rak, oraz Alzheimer (Zobacz także artykuł [Badania związane z chorobami](#), gdzie znajdziesz więcej informacji). Prosimy [dołączyć do nas](#) w naszych staraniach! **Rosetta@home jest projektem non-profit.**

Site search

Łącz do projektu Rosetta@home

- [Zasady i reguły](#)
- [Wymagania systemowe](#)
- [Ściągnij, zainstaluj i uruchom BOINC](#)
- When prompted, select Rosetta@home from the list of projects.
- [Pamiętaj na początek od David'a Baker'a](#)
- [Donate](#)

Projekty

- 10 powodów, dlaczego uczestnicy liczą w Rosetta@home
- Krótki przewodnik po projekcie i wydawczu
- Play the interactive Rosetta game, FoldIt!
- Rosetta@home FAQ - odnośnie projektu
- Rosetta@home FAQ - odnośnie zagadnień naukowych
- Badania związane z chorobami
- Przegląd naszych badań
- Publications
- Nowinki i artykuły o projekcie Rosetta
- Dziennik Rosetta@home David'a Baker'a
- Rosetta@home promo video
- Nowinki techniczne

Ważne dla uczestników

- [Twoje konto](#) - zobacz statystyki, zmień ustawienia
- [Zespoły](#) - stwórz albo dołącz do zespołu
- [Aplikacje](#)
- [Statystyki Serwera](#)
- [Dodatki](#)
- [How to view your structure predictions](#)
- [Certificate](#)

Archiwum

- [Forum dyskusyjne](#)
- [Pytania i odpowiedzi](#)
- [Profil uczestników](#)
- [Grafiki](#)
- [Języki](#)

Statystyki

- [Najlepsi uczestnicy](#)

Uczestnik dnia

[AF>HFR] David
member of L'Alliance Francophone since 2010,
member of BOINC since 2009.

Statystyki Serwera as of 8 Dec 2014 8:05:55 UTC

[Scheduler pracuje]

Total queued jobs: **5,932,578**

In progress: 388,893

Successes last 24h: 146,278

Users [↓](#) (last day [↓](#)) : 685,999 (+75)

Hosts [↓](#) (last day [↓](#)) : 1,553,297 (+161)

Credits last 24h [↓](#) : 15,390,055

Total credits [↓](#) : 31,387,251,356

TeraFLOPS estimate: 153.901

Dec 07, 2014

Predictor of the day: Congratulations to [Lance Stringham](#) for predicting the lowest energy structure for workunit

FFF_846b36cdf4fe5cf199fcd797fea3ff01_aar28Ktest1_02_Sat_Nov_22_04_26_07_PST_2014_0000367_relax_SAVE_ALL_OUT_226888_0

[...więcej nowinek](#)

[XML](#) Available as an [RSS feed](#).

Tweets

[Follow](#)

R@h Rosetta@HOME @rosettaathome 18 Mar
just launched abinitio calculations for 95 protein sequences so we can have a set of decoys for training the energy function
[Expand](#)

R@h Rosetta@HOME @rosettaathome 1 Mar
Just queued up a new batch of @Foldit protein designs for ab initio folding!
[Expand](#)

Nowinki


Aug 15, 2014

Here's an update from David Baker on the Institute for Protein Design's progress: [Letter from the Director – IPD Update](#)

Aug 05, 2014

Citizen Science

<http://fold.it/portal/>



foldit
BETA
 Solve Puzzles
 for Science

09:15:08 GMT

[PUZZLES](#)
[BLOG](#)
[CATEGORIES](#)
[FEEDBACK](#)
[GROUPS](#)
[FORUM](#)
[PLAYERS](#)
[WIKI](#)
[FAQ](#)
[RECIPES](#)
[ABOUT](#)
[CONTESTS](#)
[CREDITS](#)

The Science Behind Foldit

Foldit is a revolutionary new computer game enabling you to contribute to important scientific research. This page describes the science behind Foldit and how your playing can help.

Page Contents:

- [What is protein folding?](#)
- [Why is this game important?](#)
- [Foldit Scientific Publications](#)
- [News Articles about Foldit](#)
- [News Articles about Rosetta](#)
- [Rosetta@Home Screensaver](#)
- [Community Rules](#)
- [Let's Foldit Podcast](#)
- [Instructions for Educators](#)
- [Terms of Service and Consent](#)
- [Credits](#)


What is protein folding?

What is a protein? Proteins are the workhorses in every cell of every living thing. Your body is made up of trillions of cells, of all different kinds: muscle cells, brain cells, blood cells, and more. Inside those cells, proteins are allowing your body to do what it does: break down food to power your muscles, send signals through your brain that control the body, and transport nutrients through your blood. Proteins come in thousands of different varieties, but they all have a lot in common. For instance, they're made of the same




Folded up Streptococcal Protein Puzzle
(+) Enlarge This Image

GET STARTED: DOWLOAD




Win Beta

Windows
(XP/Vista/7/8)



Mac Beta

OSX
(10.7 or later)



Linux Beta

Linux
(64-bit)

[Are you new to Foldit? Click here.](#)

[Are you an educator? Click here.](#)

SEARCH

☒ Only search fold.it

RECOMMEND FOLDIT

USER LOGIN

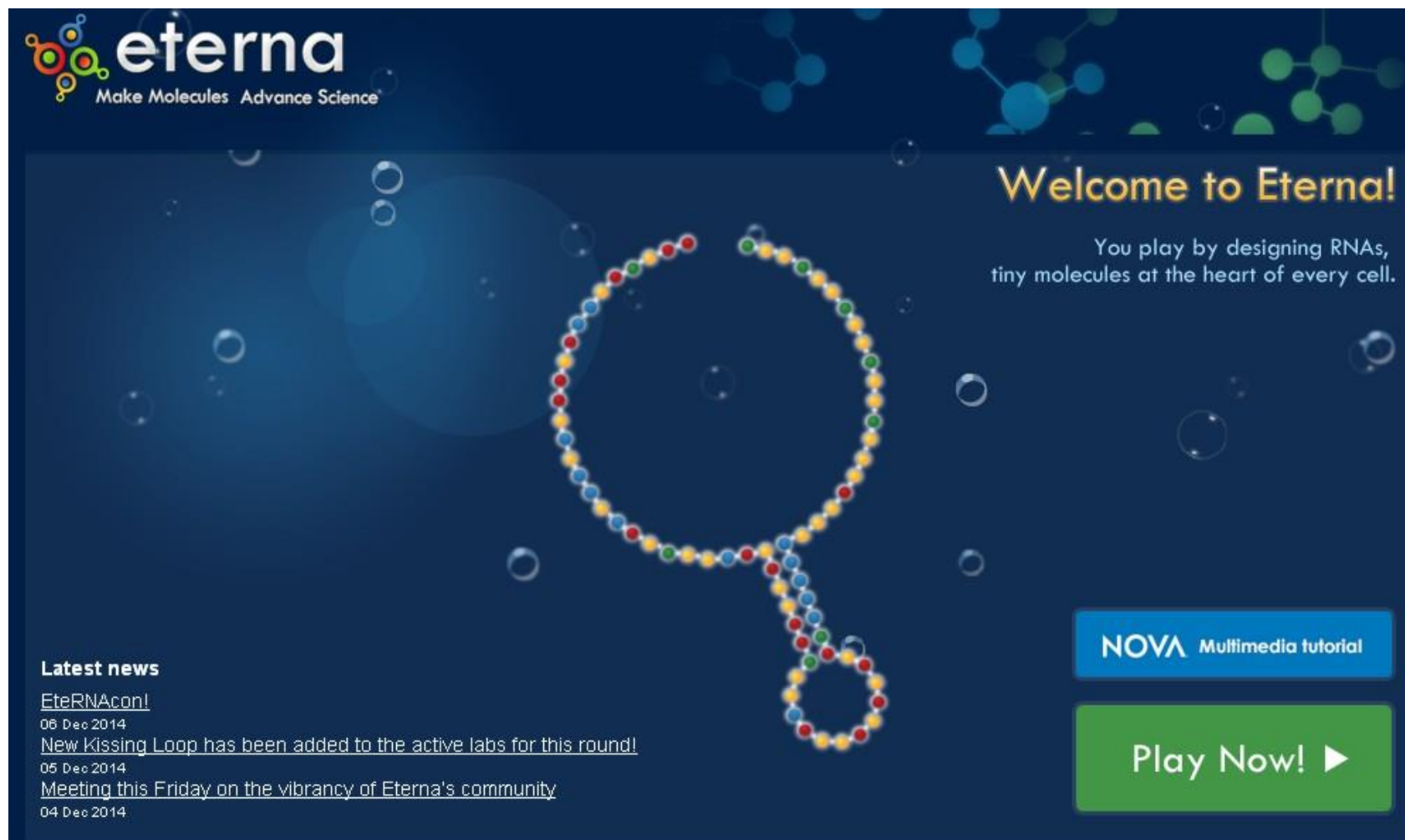
Username: *

Password: *

- [Create new account](#)
- [Request new password](#)

<http://eterna.cmu.edu/web/>

<http://www.pbs.org/wgbh/nova/labs/lab/rna/>

The image is a screenshot of the Eterna game interface. At the top left is the Eterna logo with the tagline "Make Molecules Advance Science". In the center is a large, colorful RNA molecule structure. To the right, the text "Welcome to Eterna!" is displayed in a yellow font, followed by the description "You play by designing RNAs, tiny molecules at the heart of every cell." Below this, there are two buttons: a blue "NOVA Multimedia tutorial" button and a green "Play Now!" button with a right-pointing triangle. In the bottom left corner, there is a "Latest news" section with three entries, each with a date and a link.

eterna
Make Molecules Advance Science

Welcome to Eterna!

You play by designing RNAs,
tiny molecules at the heart of every cell.

NOVA Multimedia tutorial

Play Now! ►

Latest news

- [EteRNAcon!](#)
06 Dec 2014
- [New Kissing Loop has been added to the active labs for this round!](#)
05 Dec 2014
- [Meeting this Friday on the vibrancy of Eterna's community](#)
04 Dec 2014

© 2000 Randy Glasbergen.
www.glasbergen.com



"THE COMPUTER SAYS I NEED TO UPGRADE MY BRAIN
TO BE COMPATIBLE WITH ITS NEW SOFTWARE."