

**Einführung in die
Einführung in die Angewandte Bioinformatik:
Struktur von Proteinen und RNA
25.06.2009**

Prof. Dr. Sven Rahmann

Beschreibung und Vorhersage der Struktur von RNA und Proteinen

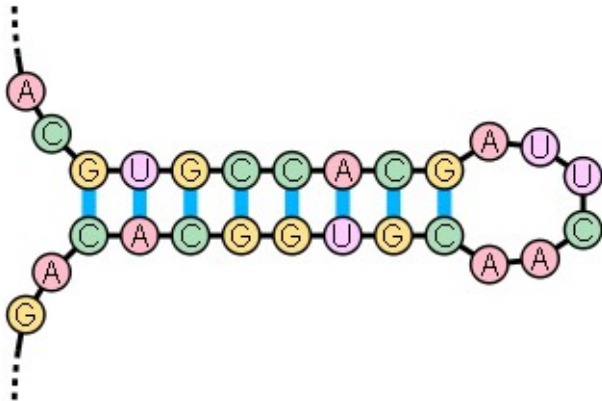
- RNA-Sekundärstruktur
- Protein-Sekundärstruktur
- Protein-Tertiärstruktur

Wichtige Hinweise

- Die Vorlesung am 02.07.2009 fällt aus.
- Sie erhalten auf der Webseite für den 02.07.2009 ein Klausurübungsblatt.
- Nutzen Sie die Zeit, um sich damit auf die Klausur am 23.07. vorzubereiten.
- Üben Sie unter realistischen Bedingungen.
- Sie erhalten am 09.07. Musterlösungen; Teilaufgaben werden besprochen.

RNA-Sekundärstruktur

- Einzelsträngige RNA bildet mit sich selbst stabile Strukturen durch RNA-Basenpaarung (A-U, G-C, sowie G-U).
- Mehr Basenpaare = mehr Stabilität (weniger freie Energie)



RNA-Sekundärstruktur-Elemente

A: Haarnadelschleife (hairpin loop)

B: Stapel (stacking region)

C: Ausbeulung (bulge loop)

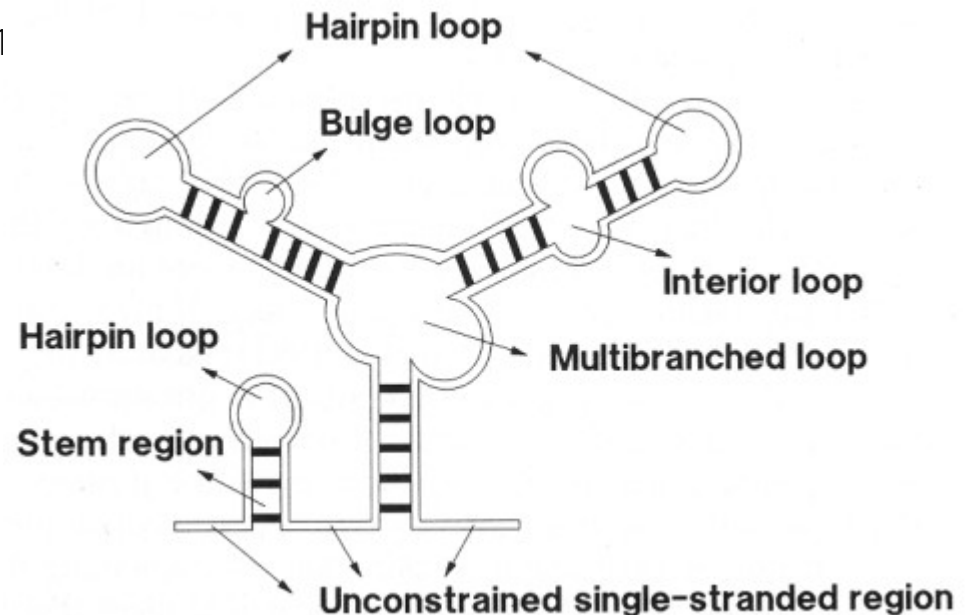
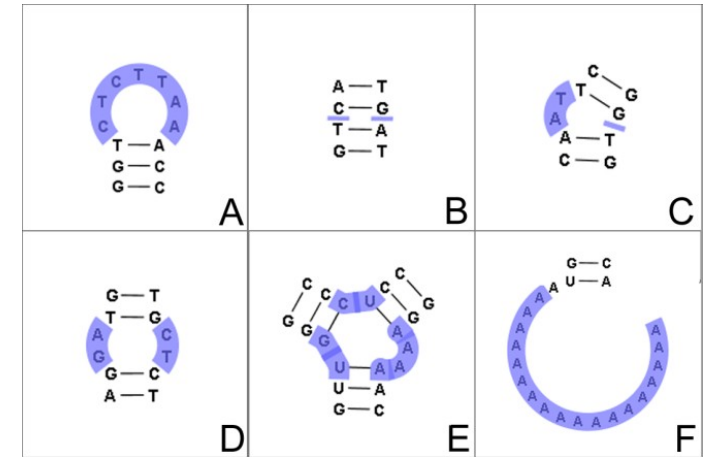
D: innere Schleife (interior loop)

E: Verzweigung (multiloop)

F: äußere Schleife (exterior loop)

Quelle: <http://www.clcbio.com/index.php?id=11>

Quelle unten: Chemie, Leiden University, NL.



RNA-Sekundärstruktur-Vorhersage

Gegeben: RNA-Sequenz

Gesucht: MFE-Struktur (minimum free energy),
d.h. Menge kompatibler Basenpaare.

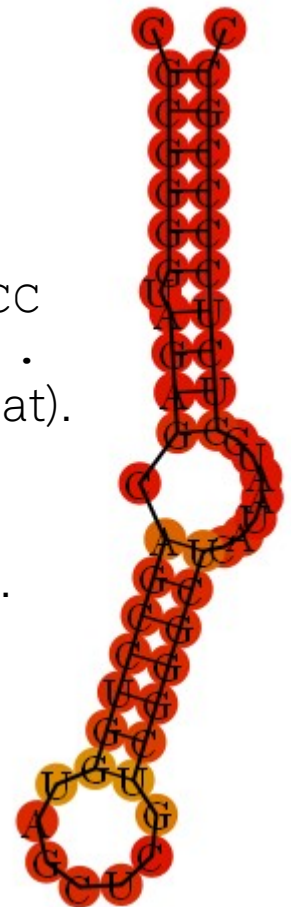
Beispiel:

```
CGCGGGGUAGAGCAGCCUGGUAGCUCGUCGGGCUCAUAAUCCUCUCCCCGCC  
.(((((((.((((.(((((((.....)))))))).(((((((.....)))))))).
```

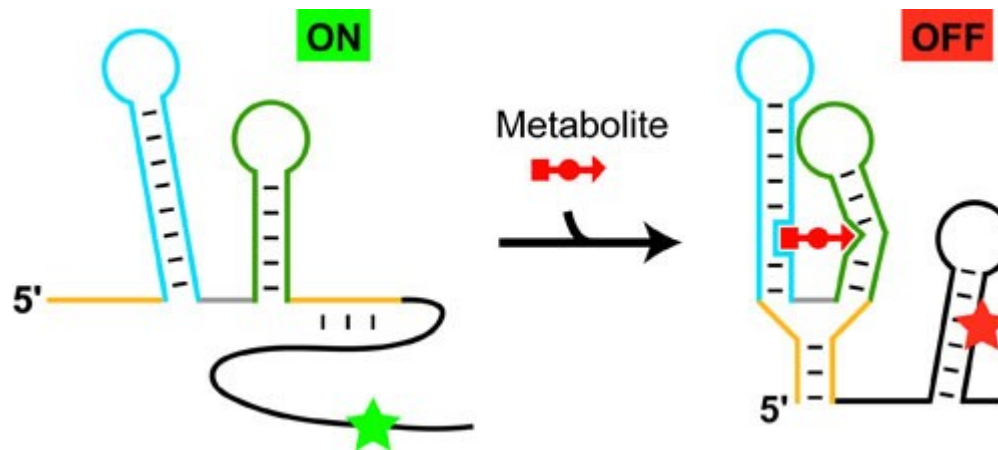
Zueinander gehörende Klammern bilden Basenpaare (Vienna-Format).

Nicht immer gibt es nur eine energetisch günstige Struktur.

Daher auch: Alternativen anschauen („thermodynamic ensembles“).



Riboswitches



Quelle: Serganov, A., Polonskaia, A., Phan, A.-T., Breaker, R.R., Patel, D.J.
Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.
Nature 441, 1167-1171 (2006).

Software zur RNA-Sekundärstruktur-Vorhersage

RNAfold (Vienna RNA package):

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>

UNAFold (vormals mfold, Zuker-Gruppe):

<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>

Beispielsequenz:

CGCGGGGUAGAGCAGCCUGGUAGCUCGUCGGGCUCAUAAUCCUCUCCCCGCC

Allgemeines zur Proteinstruktur

Funktion eines Proteins wird durch **Interaktions-Möglichkeiten** mit anderen Molekülen bestimmt,
z.B. mit anderen Proteinen, Metaboliten, DNA, RNA, ...
Interaktionen ergeben sich aus der 3D-Struktur des Proteins.



Bild: Mantelprotein-Dimer eines Virus bindet an eine RNA-Haarnadelschleife des viruseigenen RNA-Replicase-Gens und reprimiert so dessen Expression.
Quelle: Uppsala Universitet

Beispiele für Proteinfunktionen

- Ionenkanäle (Membranproteine)
- Enzyme (Katalyse chemischer Reaktionen)
- Transport von Molekülen (z.B. Hämoglobin transportiert Sauerstoff)
- Struktur (z.B. Crystallin, Kollagen)
- DNA-Transkription
- Protein-Abbau (Ubiquitin)
- ...



Bild: Crystallin, Strukturprotein der Augenlinse
symmetrischer Aufbau, kristalline Anordnung;
deswegen ist die Linse durchsichtig.

Die vier Strukturebenen eines Proteins

Primärstruktur: Aminosäure-Sequenz

Sekundärstruktur: Einteilung in

- regelmäßig gefaltete Bereiche (Helizes, Faltblätter),
- stabile Schleifen
- unstrukturierte Bereiche dazwischen

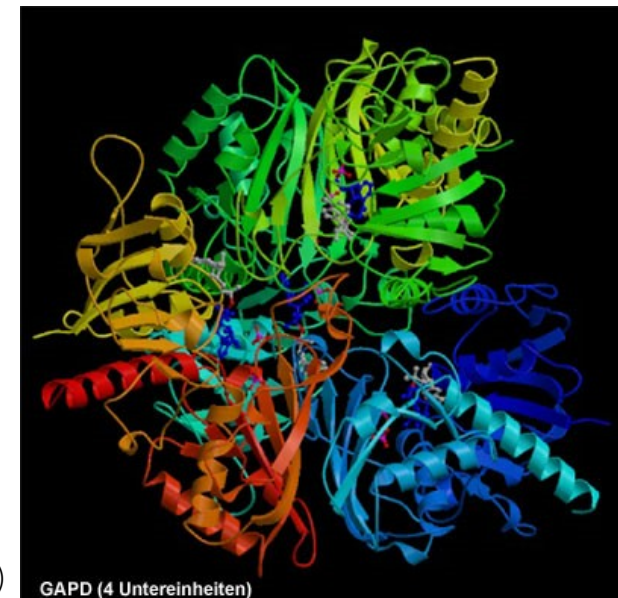
Tertiärstruktur: räumliche (3D) Faltung des Proteins, bestimmt durch die Koordinaten aller Atome

Quartärstruktur: Aufbau des Protein(komplexe)s aus mehreren Untereinheiten.

Quartärstruktur von GAPD (Glycerinaldehydphosphat-Dehydrogenase)



Visualisierung der Tertiärstruktur von Ubiquitin. Sekundärstruktur durch Cartoons hervorgehoben.



GAPD (4 Untereinheiten)

Struktur-Stabilität bei Proteinen

Warum bilden Proteine Sekundär- und Tertiärstrukturen?
Weil diese stabil sind (Energieminimierung).

Stabilität der Sekundärstruktur:

Wasserstoffbrückenbindungen des Polypeptid-Backbones

Stabilität der Tertiärstruktur:

Disulfidbindungen

Wasserstoffbrücken

ionische Wechselwirkungen

hydrophobe Wechselwirkungen

Van-der-Waals-Kräfte

Struktur-Vorhersage für Proteine

Primärstruktur:

- Aminosäure-Sequenz lässt leicht erhalten
(systematisches Übersetzen der mRNA-Sequenzen).

Sekundärstruktur:

- Vorhersage aus Primärstruktur möglich, aber unsicher.
- Aus Tertiärstruktur ergibt sich zweifelsfrei Sekundärstruktur.

Tertiärstruktur:

- ab initio Vorhersage aus Primärstruktur (und Sekundärstruktur): schwierig
- homologiebasierte Vorhersage aus bekannter verwandter Tertiärstruktur
- experimentelle Bestimmung

Wichtige Protein - Ressourcen

UniProt: Umfassende Protein-Datenbank (Swiss-Prot, TrEMBL),
mit Links zu spezialisierten externen Datenbanken
<http://www.uniprot.org>

Zusatz: MCSIS - spezielle Protein-Datenbanken (z.B. GPCRDB)

ExPASy: Expert Protein Analysis System [<http://www.expasy.org>],
Sammlung von Tools zur Protein-Analyse, gut mit UniProt verlinkt

PDB: Protein-Strukturen (insbesondere 3D-Strukturen)
<http://pdb.org> oder <http://www.pdb.org/pdb/home/home.do>
Internationaler Auftritt: wwPDB unter <http://wwpdb.org>

SCOP, CATH: Strukturelle Klassifikation von Proteinen, Proteindomänen.

Erinnerung: Universal Protein Resource (UniProt)

<http://www.uniprot.org>

UniProt besteht aus vier Datenbanken

- Protein Knowledgebase (UniProtKB) – bestehend aus
 - TrEMBL – Translated EMBL (übersetzte EMBL-Nukleotideinträge)
 - Swiss-Prot (von Hand annotierte Protein-Datenbank)
- Sequenz-Cluster (UniRef) – repräsentative Sequenzen für Proteinfamilien
- Sequenz-Archiv (UniParc) – Historie der Proteinsequenzen
- Proteinsequenzen aus Metagenomprojekten (UniMES) (neu)

Molecular Class-Specific Information System (MCSIS) project



Available MCSIS

- **The GPCRDB:** a Molecular-Specific Information System for G Protein-Coupled Receptors (created in 1994)
 - [The GPCRDB](#) at the CMBI, the Netherlands
- **The NucleaRDB:** a Molecular-Specific Information System for Nuclear Receptors (created in April 2000)
 - [The NucleaRDB](#) at the CMBI, the Netherlands
 - Mirror site at UCSF, USA (no longer available)
- **The PrionDB:** a Molecular-Specific Information System for Prion proteins (created July, 21 2003)
 - [The PrionDB](#) at the CMBI, the Netherlands
 - Mirror site at UCSF, USA (no longer available)
- **The KChannelDB:** a Molecular-Specific Information System for potassium channels (created July, 25 2003)
 - [The KChannelDB](#) at the CMBI, the Netherlands
 - Mirror site at UCSF, USA (no longer available)
- **The GPCRIPDB:** a Molecular-Specific Information System for GPCR Interacting Partners (G proteins & GPCRs)
 - [The GPCRIPDB](#) at the CMBI, the Netherlands

jeweils auf eine Proteinfamilie zugeschnittene Datenbanken, enthalten Expertenwissen, meist von Hand gepflegt:

wenige, aber qualitativ hochwertige Einträge

z.B. <http://www.gpcrdb.org>

Erinnerung: ExPASy – Tools zur Analyse von Proteinen

Beispiele

- ProtParam: diverse (theoretische) Eigenschaften eines Proteins.
- Topology prediction / TMHMM: Entscheidet, ob Transmembranprotein vorliegt.
- Sekundär- und Tertiärstrukturvorhersage

[\[Databases\]](#) [\[Tools & Software\]](#) [\[Education & Services\]](#) [\[Links\]](#)
[\[Announcements\]](#) [\[Mirror Sites\]](#) [\[Job openings\]](#)

Databases

- **UniProt Knowledgebase (Swiss-Prot and TrEMBL)** - Protein knowledgebase
- **PROSITE** - Protein families and domains
- **SWISS-2DPAGE** - Two-dimensional polyacrylamide gel electrophoresis
- **World-2DPAGE Repository** - A public standards-compliant repository for gel-based proteomics data published in the literature
- **MIAPEGelDB** - A public repository for MIAPE Gel electrophoresis documents **new**
- **ENZYME** - Enzyme nomenclature
- **UniPathway** - Metabolic pathways **new**
- **SWISS-MODEL Repository** - Automatically generated protein models
- [Links to many other molecular biology databases](#)

Tools and software packages

- **Proteomics and sequence analysis tools**
 - Identification and characterization (Aldente, FindMod, Popitam, Phenyx, pI/Mw, **ProtParam...**)
 - DNA -> Protein
 - Similarity searches (BLAST...)
 - Pattern and profile searches (ScanProsite...)
 - Post-translational modification and **Topology prediction**
 - **Primary structure analysis**
 - **Secondary and tertiary structure tools** (Swiss-PdbViewer...)
 - Alignment and Phylogenetic analysis
- **Melanie / ImageMaster** - Software for 2-D PAGE analysis
- **MSight** - Mass Spectrometry Imager
- **Roche Applied Science's Biochemical Pathways**

Education and services

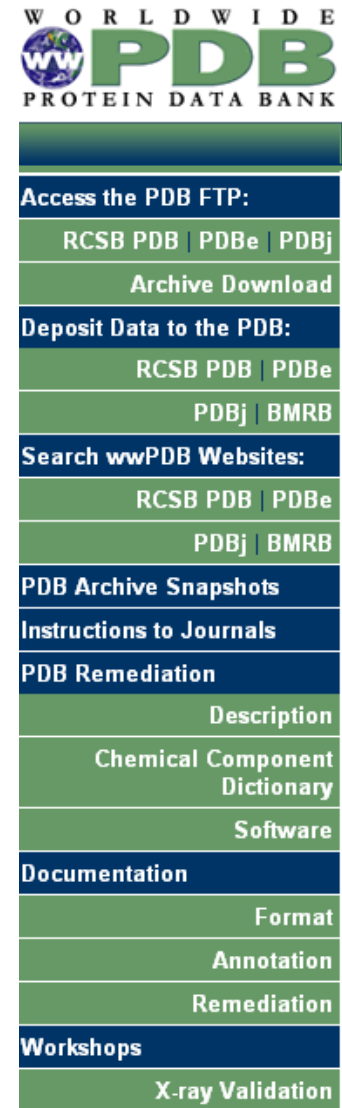
- **The ExPASy FTP server**
- **Swiss-Shop** - automatically obtain (by email) new sequence entries relevant to your field(s) of interest
- Popular Science
 - **Protein Spotlight**
 - **Protéines à la «Une»**
- **e-Proxemis** - Bioinformatics Learning Portal for Proteomics
- **Bioinformatics Core Facility for Proteomics**
- **Proteomics Core Facility** (formerly known as SWISS-2DSERVICE)
- **SPS' Digest** - the Swiss Proteomics Society selection of proteomics articles
- **Master's degree in Proteomics and Bioinformatics**
- **Swiss-Quiz**

Die Protein-Tertiärstruktur-Datenbank PDB

- PDB: Protein-Strukturen, <http://pdb.org>
- RCSB: Konsortium hinter PDB in den USA,
Research Collaborative in Structural Bioinformatics
- wwPDB: Internationaler Auftritt (RCSB PDB, PDBe, PDBj),
zusätzlich BMRB
(Biological Magnetic Resonance Bank,
Daten aus NMR-Spektroskopie)

Wichtigste Inhalte:

- 3D-Koordinaten der Atome, die ein Protein bilden
- daraus abgeleitete Informationen



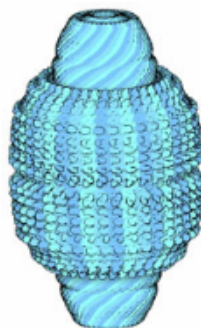
- Home
- Getting Started
- Structural Genomics
- BioSync
- Electron Microscopy
- Download Files
- Deposit and Validate
- Dictionaries & File Formats
- Software Tools
- General Education
- General Information
- Acknowledgements
- Frequently Asked Questions

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the [wwPDB](#), the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

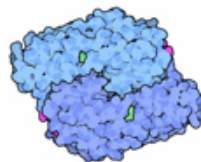
Molecule of the Month: Vaults



Our cells are filled with compartments, each performing a specific function. Some of these compartments, such as mitochondria and lysosomes, are very large and enclose many different molecular machines. Other intracellular compartments are smaller, such as the transport vesicles that shuttle proteins from site to site inside the cell. Most of these compartments, including mitochondria, lysosomes and transport vesicles, are surrounded by membranes. However, in special cases, cells build smaller compartments surrounded by a protein shell. In our own cells, vaults are a spectacular example of these protein-enclosed compartments.

[Read more ...](#) [Previous Features](#)

PSI Featured Molecule: Hda and DNA Replication



When cells divide, they need to ensure that each daughter cell gets one copy of each chromosome. Bacteria contain one big circle of DNA, so they start replication in one place, then copy the DNA both ways around until it finishes on the other side. PSI Researchers have solved the first atomic structure at how bacteria use the Hda protein to initiate replication at this origin only once for each generation of the cell.

[Read more from PSI SGKB](#) [Previous Features](#)

New user? Try the browser [compatibility check](#), information on [Getting Started](#), and see [What's New](#) on the site.

News

- [Complete News](#)
- [Newsletter](#)
- [Discussion Forum](#)
- [Job Listings](#)

16-June-2009

How does an HPUB structure get released?

An HPUB status indicates that a structure will be released when the primary reference is published. When it's confirmed that the corresponding article is available, the structure is included in the weekly update of the PDB.

The [wwPDB](#) receives publication dates and citation information directly from a few journals. For most articles, however, the [wwPDB](#) searches PubMed and scans the literature for publication information. Citations emailed to deposit@wwpdb.org are also greatly appreciated.

[More >>](#)

Data Snapshots

Time-stamped yearly snapshots of the PDB archive are available via FTP at: <ftp://snapshots.wwpdb.org> The snapshots provide readily identifiable data sets for research on the PDB archive.

Benutzung von PDB: Beispiel CRGD_BOVIN (1elp)

Einstieg:

- Schnell bei bekannter PDB accession number: direkt suchen
- Stichwortsuche in PDB
- Bequem: Link von UniProt zu PDB folgen

Möglichkeiten:

- PDB-Datei anschauen
(Literatur, Primärstruktur,
Sekundärstrukturannotation,
Atomkoordinaten = Tertiärstruktur)
- PDB-Datei (Tertiärstruktur) visualisieren
- PDB-Datei herunterladen
- Daten zum experimentellen Protokoll ansehen
- Links auf die Strukturdatenbanken SCOP, CATH
- Link auf die Domänendatenbank Pfam

Home Search Structure

Summary Derived Data Sequence Seq. Similarity Literature Biol. & Chem. Methods Geometry Links

GAMMA-D CRYSTALLIN STRUCTURE AT 1.95 A RESOLUTION

DOI:10.2210/pdb1elp/pdb

1elp

- Display Files ▾
- Download Files ▾
- Print this Page

Primary Citation

Structure of bovine eye lens gammaD (gammaIIIb)-crystallin at 1.95 A.
Chirgadze, Y.N., Driessen, H.P., Wright, G., Slingsby, C., Hay, R.E., Lindley, P.F.
(1996) Acta Crystallogr., Sect. D 52: 712-721

PubMed: 15299634

[Search Related Articles in PubMed](#)

PubMed Abstract:

The crystal structure of bovine lens gammaIIIb-crystallin at 2.5 A resolution previously reported was interpreted using a consensus sequence derived from related vertebrate sequences on the assumption that gammaIIIb-crystallin derived from the gammaC-crystallin gene. It has recently been shown that ...
[\[Read More & Search PubMed Abstracts \]](#)

Molecular Description

Hide

Classification: **Eye Lens Protein**
Structure Weight: 41556.80

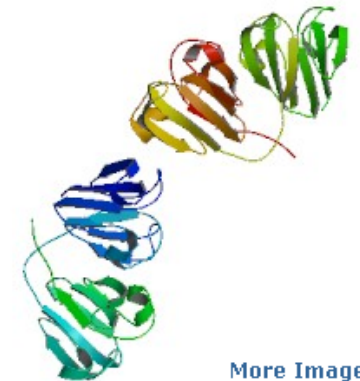
Molecule: GAMMA-D CRYSTALLIN
Polymer: 1 Type: polypeptide(L) Length: 173
Chains: A, B

Source

Hide

Polymer: 1 Scientific Name: **Bos taurus** Common Name: Cattle

Biological Molecule



[More Images...](#)

3-D Viewers: ?

Jmol Protein Workshop
SimpleViewer Other Viewers ▾

Oligomeric State:
DIMERIC

- 1ELP
- Download Files
- FASTA Sequence
- Display Files
- Display Molecule
- Structural Reports
- External Links
- Structure Analysis
- Help
- Queries
- Reset View

Inhalt eines PDB-Eintrags

Summary:

- Literatur (Hauptartikel mit Strukturaufklärung)
- Moleküldarstellung (3D-Struktur, externe Betrachter)
- Molekülbeschreibung (z.B. Funktion)
- Quelle (Organismus, Taxonomie-Link)
- Übersicht Abgeleiteter Daten (SCOP, CATH; s. später)
- Geschichte des PDB-Eintrags
- Details zum Experiment

Weitere Tabs:

- Details zu den Summary-Punkten
- Beispiel: 1elp

Gesamter Inhalt: siehe PDB-Dateiformat (folgende Folien).

```

HEADER          EYE LENS PROTEIN                               20-DEC-95
TITLE          GAMMA-D CRYSTALLIN STRUCTURE AT 1.95 A RESOLUTION
COMPND         MOL_ID: 1;
COMPND         2 MOLECULE: GAMMA-D CRYSTALLIN;
COMPND         3 CHAIN: A, B
SOURCE        MOL_ID: 1;
SOURCE        2 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE        3 ORGANISM_COMMON: CALF;
SOURCE        4 ORGAN: EYE;
SOURCE        5 TISSUE: EYE LENS
KEYWDS        EYE LENS PROTEIN
EXPDTA        X-RAY DIFFRACTION
AUTHOR        YU.N.CHIRGADZE,H.P.C.DRIESSEN,G.WRIGHT,C.SLINGSBY,R.E.HAY,
AUTHOR        2 P.F.LINDLEY
REVDAT        2   17-MAR-99 1ELP      1          JRNL    REMARK
REVDAT        1   10-JUN-96 1ELP      0
JRNL          AUTH   Y.N.CHIRGADZE,H.P.DRIESSEN,G.WRIGHT,C.SLINGSBY,
JRNL          AUTH 2 R.E.HAY,P.F.LINDLEY
JRNL          TITL   STRUCTURE OF BOVINE EYE LENS GAMMAD
JRNL          TITL 2 (GAMMAIIIB)-CRYSTALLIN AT 1.95 A.
JRNL          REF   ACTA CRYSTALLOGR.,SECT.D          V.   52    712 1996
JRNL          REFN  ASTM ABCRE6   DK ISSN 0907-4449
REMARK        1
REMARK        1 REFERENCE 1
REMARK        1 AUTH   YU.CHIRGADZE,N.NEVSKAYA,E.VERNOSLOVA,S.NIKONOV,
REMARK        1 AUTH 2 YU.SERGEEV,E.BRAZHNIKOV,N.FOMENKOVA,V.LUNIN,
REMARK        1 AUTH 3 A.URZHUMTSEV

```

1ELP

PDB Identifier
(accession number)

```

...
DBREF 1ELP A      1    174 UNP      P08209    CRGD_BOVIN      1
DBREF 1ELP B      1    174 UNP      P08209    CRGD_BOVIN      1
SEQADV 1ELP ILE A   75  UNP      P08209    VAL      75 CONFLICT
SEQADV 1ELP ILE B   75  UNP      P08209    VAL      75 CONFLICT
SEQRES  1 A  173  GLY LYS ILE THR PHE TYR GLU ASP ARG GLY PHE
SEQRES  2 A  173  ARG HIS TYR GLU CYS SER SER ASP HIS SER ASN
SEQRES  3 A  173  PRO TYR LEU GLY ARG CYS ASN SER VAL ARG VAL ASP SER
...
SEQRES 13 B  173  GLY ALA MET ASN ALA LYS VAL GLY SER LEU ARG ARG VAL
SEQRES 14 B  173  ILE ASP ILE TYR
FORMUL  3 HOH *218(H2 O)
HELIX  1  1 ARG A    9 PHE A   11  5
HELIX  2  2 ASP A   64 MET A   69  5
HELIX  3  3 SER A  111 PHE A  116  1
HELIX  4  4 ARG A  153 GLY A  158  5
HELIX  5  5 ARG B    9 PHE B   11  5
HELIX  6  6 ASP B   64 MET B   69  5
HELIX  7  7 SER B  111 HIS B  117  1
HELIX  8  8 ARG B  153 GLY B  158  5
SHEET  1  A 4 GLN A  12  CYS A  18  0
SHEET  2  A 4 LYS A   2  ASP A   8 -1  N  ILE A   3  O  CYS A  18
SHEET  3  A 4 SER A  34  SER A  39 -1  O  SER A  34  N  TYR A   6
SHEET  4  A 4 GLY A  60  TYR A  62 -1  O  GLY A  60  N  VAL A  37
SHEET  1  B 3 GLN A  54  LEU A  57  0
SHEET  2  B 3 TRP A  42  TYR A  45 -1  O  TRP A  42  N
SHEET  3  B 3 SER A  77  LEU A  80 -1  O  SER A  77  N
...

```

Verweis auf Uniprot und Konflikte zur dortigen Sequenz

Primärstruktur

Positionen (Start, Ende) der Helices, Helix-Typ, Helix-Länge

Positionen (Start, Ende) der Faltblätter

```

...
SHEET      3      H 3 SER B 166  ARG B 169 -1  O  SER B 166  N  TYR B 134
CRYST1    57.810   70.030  117.250  90.00  90.00  90.00 P 21 21 21    8
ORIGX1      1.000000  0.000000  0.000000  0.000000
ORIGX2      0.000000  1.000000  0.000000  0.000000
ORIGX3      0.000000  0.000000  1.000000  0.000000
SCALE1      0.017298  0.000000  0.000000  0.000000
SCALE2      0.000000  0.014280  0.000000  0.000000
SCALE3      0.000000  0.000000  0.008529  0.000000
ATOM       1  N  GLY A  1      2.897  8.141  16.970  1.00 31.82
ATOM       2  CA GLY A  1      3.515  7.459  18.105  1.00 18.21
ATOM       3  C  GLY A  1      4.346  8.490  18.842  1.00 26.49
ATOM       4  O  GLY A  1      4.594  9.607  18.366  1.00 25.32

```

Koordinaten-
transformationen

Atom-Name,
Aminosäure,
Kette (A, B),
Aminosäure-Nr.

Atom-Koordinaten

Protein-Sekundärstruktur

Einteilung der Primärstruktur (Sequenz)
in Bereiche gleicher regelmäßiger Geometrie, insbes.

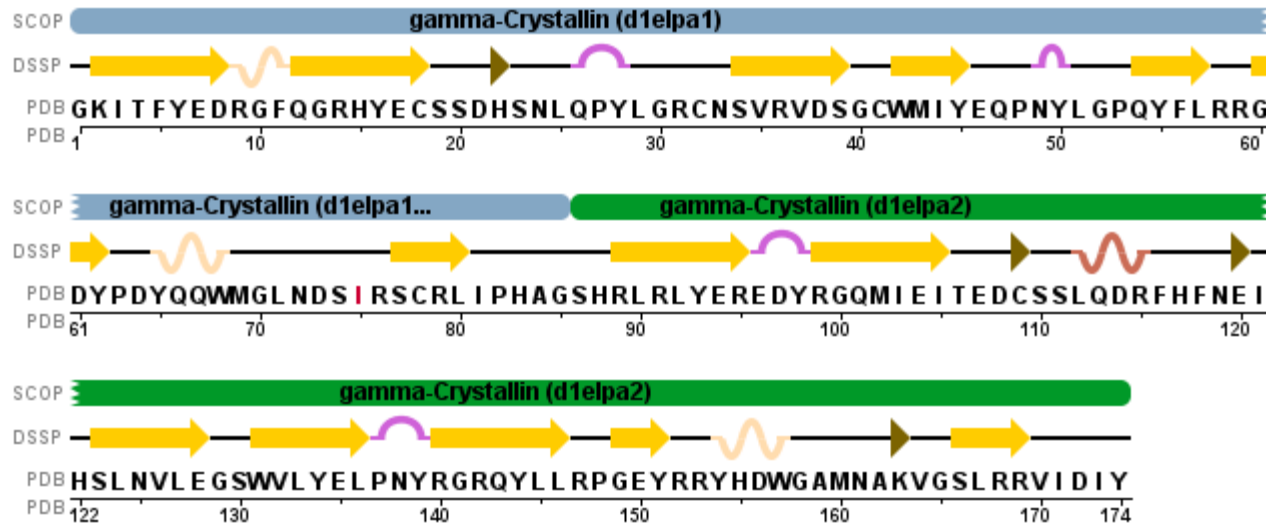
- Helix (H)
- Faltblatt (shEet)
- Rest (random Coil)

Beispiel

UniProt Accession Number P08209 (CRGD_BOVIN) ist Gamma-crystallin D (Rind).
Sekundärstruktur und 3D-Struktur sind bekannt.

Protein-Sekundärstruktur in PDB

Visualisierung der Sekundärstruktur im Sequence-Tab in PDB aus DSSP: Dictionary of Secondary Structure of Proteins (Teil von PDB), erkennt Struktur aus Atomkoordinaten.



Protein-Sekundärstruktur-Vorhersage über ExPASy

Häufig ist die Struktur nicht bekannt.

Man kann versuchen, sie vorherzusagen:

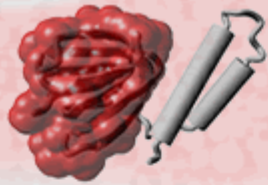
- statistische Sequenzeigenschaften,
- Homologie zu Proteinen mit bekannter Struktur.

Vorhersage über ExPASy mit SSpro (3 und 8 Klassen)

Scratch Protein Predictor (<http://scratch.proteomics.ics.uci.edu/>).

Secondary structure prediction

- [AGADIR](#) - An algorithm to predict the helical content of peptides
- [APSSP](#) - Advanced Protein Secondary Structure Prediction Server
- [GOR](#) - Garnier et al, 1996
- [HNN](#) - Hierarchical Neural Network method (Guermeur, 1997)
- [HTMSRAP](#) - Helical TransMembrane Segment Rotational Angle Prediction **new**
- [Jpred](#) - A consensus method for protein secondary structure prediction at University of Dundee
- [JUFO](#) - Protein secondary structure prediction from sequence (neural network)
- [nnPredict](#) - University of California at San Francisco (UCSF)
- [Porter](#) - University College Dublin
- [PredictProtein](#) - PHDsec, PHDacc, PHDhtm, PHDtopology, PHDthreader, MaxHom, EvalSec from Columbia University
- [Prof](#) - Cascaded Multiple Classifiers for Secondary Structure Prediction
- [PSA](#) - BioMolecular Engineering Research Center (BMERC) / Boston
- [PSIpred](#) - Various protein structure prediction methods at Brunel University
- [SOPMA](#) - Geourjon and Deléage, 1995
- [SSpro](#) - Secondary structure prediction using bidirectional recurrent neural networks at University of California
- [DLP-SVM](#) - Domain linker prediction using SVM at Tokyo University of Agriculture and Technology



SCRATCH

Protein Predictor

Email

Name Of Query (Optional)

Protein Sequence

(plain sequence, no headers,
spaces and newlines will be
ignored)

ACCpro: Solvent Accessibility (25%)

SSpro: Secondary Structure (3 Class)

DISpro: Disorder

CONpro: Contact Number

CMAPro: Contact Map

3Dpro: Tertiary Structure

SOLpro: Solubility upon Overexpression

ACCpro20: Solvent Accessibility (20 Class)

SSpro8: Secondary Structure (8 Class)

DOMpro: Domains

DIpro: Disulfide Bonds

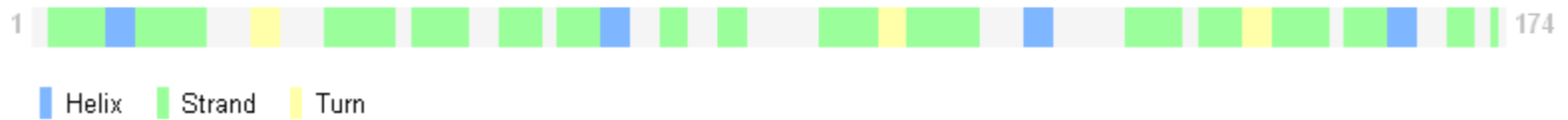
SVMcon: New SVM Contact Map

COBEpro: Continuous B-cell Epitopes

Reset

Submit Query

Secondary structure



Sequenz (CRGD_BOVIN):

MGKITFYEDRGRFQGRHYECSSDHSNLQPYLGRCNVSRVDSGCWMIYEQPNYLGPQYFLRRGDYPDYQQWMGLNDSVRSCL
IPHAGSHRLRLYEREDYRGQMIEITEDCSSLQDRFHFNEIHSLNVLEGSWVLYELPNYRGRQYLLRPGEYRRYHDWGAMNA
KVGSLRRVIDIY

SSPro (3 Ausgabeklassen) – vorhergesagte Sekundärstruktur (s. auch Abbildung)

CCEEEEEEECCCCCEEEEECCCCCCCCCHHCCCCCEEEEECEEEEEEECCCCCEEEEECCCCCHHHHHHCCCCCCCCCCC
CCCCCEEEEECCCCCEEEEECCCHHHHCCCCCEEEEECEEEEEEECCCCCEEEEECCCCCCCCCCCCCCCCCCC
CHHCHHHHHHCC

SSPro8 (8 Ausgabeklassen) – vorhergesagte Sekundärstruktur

CCEEEEEECTTCCEEEEEECTTCTTCTTCTTCEEEEEETSEEEEEECTTCTTEEEEEETTCCHHHHHHTCCSHEEECCC
CCTTSCCEEEEEETTCTTEEEEEETTCHHHHHHTCCCEEEEEEEETEEEEEEECTTCCEEEEEECTTCCCCCTTTCCCH
TEEEEEEEHECC

Legende (SSPro3 liefert nur H,E,C):

H: alpha-helix	G: 3-10-helix	l: pi-helix
E: extended strand	B: beta-bridge	
C: „random coil“	T: turn	S: bend

Strukturelle Klassifikation von Protein(domän)en: CATH

<http://www.cathdb.info>

Die CATH-Datenbank klassifiziert Proteindomänen in 4 Ebenen:

- Class
- Architecture
- Topology
- Homologous Superfamily

(Mischung aus automatischer und manueller Klassifikation.)

CATH: Protein Structure Classification

Introduction

CATH is a hierarchical classification of protein domain structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T) and Homologous superfamily (H). The boundaries and assignments for each protein domain are determined using a combination of automated and manual procedures which include computational techniques, empirical and statistical evidence, literature review and expert analysis. [more...](#)



Links for Researchers

- [Search CATH by ID/sequence/text](#)
- [Browse CATH from the top of the hierarchy](#)
- [Download CATH data](#)
- [Introduction to CATH](#)
- [Frequently Asked Questions](#)
- [Other research projects in the Orengo Group](#)
- [CATH contributors](#)

Links for Developers

- [Linking to CATH](#)
- [What has changed in v3.2.0?](#)
- [Webservices @ CATH](#)
- [CATH blog](#)

CATH References

[The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies.](#)
Alison L. Cuff, Ian Sililloe, Tony Lewis, Oliver C. Redfern, Richard Garratt, Janet Thornton, Christine A. Orengo (2008) Nucleic Acids Research
CATH: A Hierarchic Classification of Protein Domain Structures.
Orengo,C.A., Michie,A.D., Jones,D.T., Swindells,M.B., Thornton,J.M. (1997). Structure, 5, 1093-1108

New Data Available

We have added new links to the download section:

- HMM libraries
- CATH Domain PDB libraries

[...more](#)

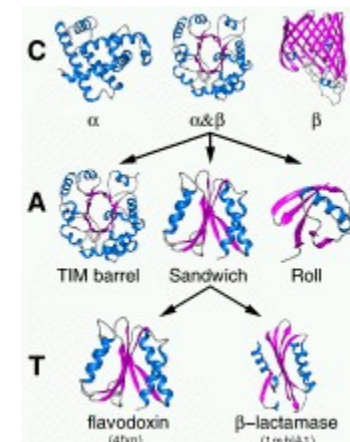
CATH v3.2.0 Released

The latest release of CATH (version 3.2.0) was fixed on July 14, 2008 and contains 114215 assigned domains.

[...more](#)

New Look CATH

The release of CATH v3.2.0 brings with it shiny new web pages. We hope you find these pages clearer and more intuitive to use. If you can't find the data that you are looking for then please do [let us know](#). [...more](#)



Strukturelle Klassifikation von Protein(domän)en: CATH

<http://www.cathdb.info>

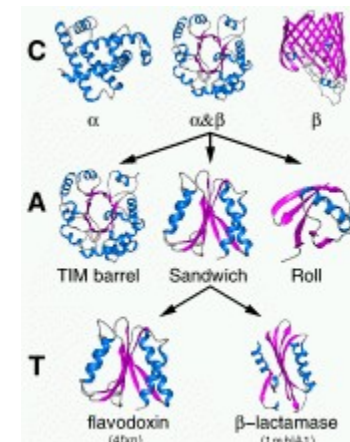
Class: aus Sekundärstrukturgehalt automatisch bestimmt:
mainly-alpha, mainly-beta, alpha-beta, few.

Architecture: aus Orientierung der Sekundärstrukturen,
unabhängig von ihrer Abfolge, manuell bestimmt;
z.B. four-layer-sandwich.

Topology: aus Abfolge der Sekundärstrukturelemente
und den Verbindungen dazwischen manuell bestimmt.

Homologous superfamilies:

Proteine mit sehr ähnlicher Struktur und Funktion



Strukturelle Klassifikation von Protein(domän)en: SCOP

<http://scop.mrc-lmb.cam.ac.uk/scop/>

SCOP-Datenbank beschreibt strukturelle und evolutionäre Beziehungen zwischen Proteindomänen bekannter Struktur.

SCOP enthält alle bekannten Faltungen.

SCOP liefert zu jedem Protein strukturell verwandte Proteine.

Hierarchische Clusterung auf drei Ebenen:

Familie (family):

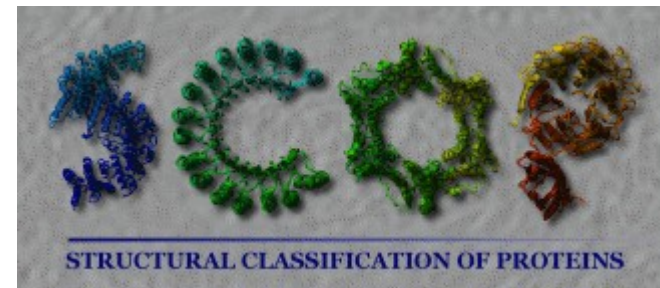
eindeutige evolutionäre Verwandtschaft, Sequenzähnlichkeit,
Proteine in einer Familie bestehen aus denselben Domänen.

Superfamilie (superfamily):

Anzeichen für gemeinsame Vorfahren, geringe Sequenzähnlichkeit.

Faltung (fold; ca. 1000 verschiedene):

Anzeichen für strukturelle Ähnlichkeit
(könnten auf Zufall beruhen).



Bestimmung und Vorhersage der Proteinstruktur

Experimentelle Bestimmung der Proteinstruktur

- Methoden: X-ray crystallography, NMR
- zeitraubend und teuer.
- nicht *in vivo*, sondern mit isolierten Proteinen

Strukturvorhersage: ab initio

- Grundsätzlich überhaupt möglich? Sequenz bestimmt Struktur?
- Modellierung beinhaltet viele Unsicherheiten (was beachten?)
- Simulation der Faltung des Moleküls im Computer (energiebasiert)
- zeitraubend und schwierig

Strukturvorhersage: homologiebasiert

- Angenommen, Struktur eines nah verwandten Proteins ist schon bekannt
- Kann große Teile der 3D-Struktur des backbones übertragen
- Nachteil: Man benötigt viele bekannte Strukturen

Ab initio - Strukturvorhersage

Idee 1

- Simulation auf Atom-/ Molekül-Ebene.
- Simuliere natürliche Zellumgebung (oder Wasser).
- Falte darin das denaturierte Protein, bis es einen stabilen Zustand erreicht.
- Prinzip: Verändere die Koordinaten der Atome ein wenig;
entscheide, ob die neue Konformation angenommen oder verworfen wird,
bevorzuge dabei stabilere Zustände.
- Entscheidung ist randomisiert.
- Problem: Woher kommt das Energiemodell?
- Problem: Verfahren nicht deterministisch und sehr zeitaufwändig.
- Problem: lokale Energieminima verhindern,
dass man die stabilste Konformation findet.

Ab initio - Strukturvorhersage

Idee 2: Zusammenstückeln

- Man betrachtet kurze Peptidstücke.
- Die Bindungswinkel zwischen den Atomen sind nicht frei.
- Bestimmte Peptidstücke nehmen bestimmte Konformationen ein.
- Daraus bekommt man eine Vorstellung von der Lage des Backbone.
- Dies kann man als Startwert für die Moleküldynamik nehmen.

Homologiebasierte - Strukturvorhersage

Grundidee und Annahme:

Es gibt zum unbekanntem Protein verwandte Proteine, deren Struktur schon aufgeklärt ist.

Bei hinreichend hoher Sequenzähnlichkeit überträgt man die Struktur so weit wie möglich und optimiert dann die Stabilität.

Threading (Fädeln):

Man betrachtet **jede** bekannte Struktur und prüft, ob sich die vorliegende Sequenz so falten lässt.

Wenn ja, optimiert man von dort aus weiter.

Critical Assessment of Structure Prediction (CASP)

Protein Structure Prediction Center

Menu

- [Home](#)
- [FORCASP Forum](#)
- [PC Login](#)
- [PC Registration](#)
- ▼ [CASP Experiments](#)
 - ▼ [CASP8 \(2008\)](#)
 - [Home](#)
 - [My CASP8 profile](#)
 - ▶ [CASP8 Registration](#)
 - ▶ [Targets](#)
 - ▶ [Predictions](#)
 - [CASP8 in numbers](#)
 - [CASP7 \(2006\)](#)
 - [CASP6 \(2004\)](#)
 - [CASP5 \(2002\)](#)
 - [CASP4 \(2000\)](#)
 - [CASP3 \(1998\)](#)
 - [CASP2 \(1996\)](#)
 - [CASP1 \(1994\)](#)
 - ▶ [CASP Initiatives](#)
 - ▶ [Outreach](#)
 - [Local Services](#)
 - ▶ [Downloads](#)
 - [Links](#)
 - [Feedback](#)
 - [FAQ](#)
 - [People](#)

Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. In addition to support of the CASP meetings our goal is to promote an evaluation of prediction methods on a continuing basis.

CASP experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused. The organizers are thankful to [CASP assessors](#) for their valuable contribution to this field.

There have been seven previous CASP experiments.

[CASP1 \(1994\)](#) | [CASP2 \(1996\)](#) | [CASP3 \(1998\)](#) | [CASP4 \(2000\)](#) | [CASP5 \(2002\)](#) | [CASP6 \(2004\)](#) | [CASP7 \(2006\)](#) | [CASP8 \(2008\)](#)

[Proceedings](#)

Click on the logo below to proceed to the main page of the latest CASP experiment.

CASP8

FORCASP
"no more dead trees"

[Discussion Forum](#)
[Old Discussion Forum](#)

Message Board

T0465 - switching to human/server target
[We are switching T0465 over to human/server target. In order to provide enough time for human-regime modeling, we are adding 10 extra days to the prediction window for this target so that it will expi ...](#)

No targets on July 4
[Next week we will have no targets released on Friday. Next target release date after the holiday - July 7. ...](#)

Extended timeframe groups
[We are launching an additional service designed to help test methods requiring more time than the usual 3 weeks allowable in CASP8. Registering an additional "extended timeframe" group will provide a ...](#)

Capri

[Prediction of docking interactions](#)

Live Bench

Automated benchmarking of prediction servers (will be again available shortly)

CAFASP

[Assessment of automated structure prediction](#)

EVA

[Automatic evaluation of prediction servers](#)

<http://predictioncenter.org>