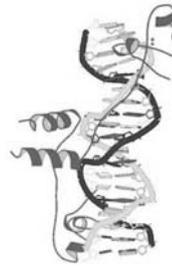


```
>1D66:A|PDBID|CHAIN|SEQUENCE
MKLLSSIEQACDICRLKLLKCSKEKPKCAKCLKN
NWECRYSPKTKRSPLTRAHLTEVESRLERLEF
>1D66:B|PDBID|CHAIN|SEQUENCE
MKLLSSIEQACDICRLKLLKCSKEKPKCAKCLKN
NWECRYSPKTKRSPLTRAHLTEVESRLERLEF
>1D66:D|PDBID|CHAIN|SEQUENCE
CCGGAGGACAGTCCTCCGG
>1D66:E|PDBID|CHAIN|SEQUENCE
CCGGAGGACTGTCCTCCGG
```



# Databases for Protein Structure

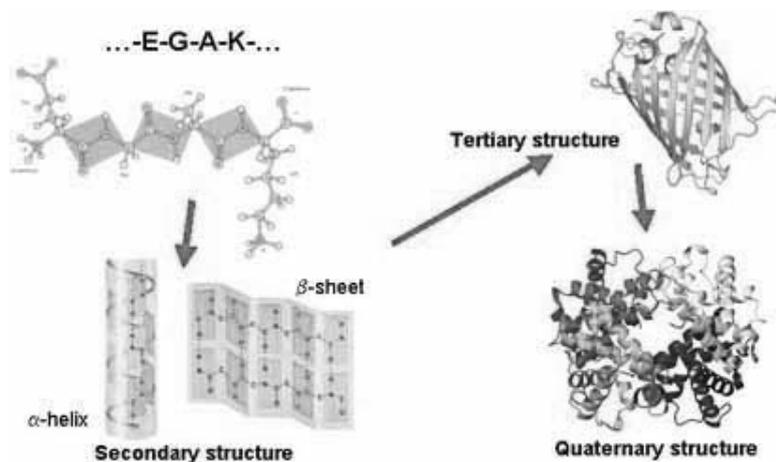
呂平江

國立清華大學

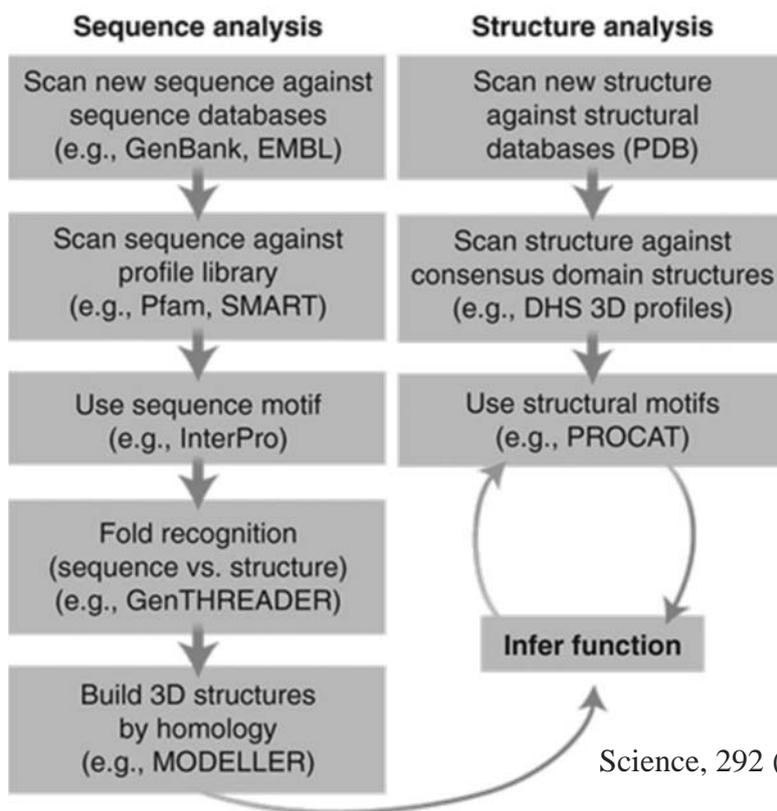
生命科學系/生物資訊與結構生物研究所

2012/06/27

## From Sequence to Structure



# From sequence and structure to function



## Molecular Biology Database Collection

~ 1380 databases

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
  - Small molecules
  - Carbohydrates
  - Nucleic acid structure
  - Protein structure
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases
- Cell biology

**Structure  
Databases**

- Structure Databases

- Small molecules

- [ACANT - Amino Acid - Nucleotide interaction database](#)
- [ChEBI - Chemical Entities of Biological Interest](#)
- [ChemBank](#)
- [ChemDB](#)
- [CSD - Cambridge Structural Database](#)
- [DrugBank](#)
- [Het-PDB Navi](#)
- [HIC-Up](#)
- [Klotho](#)
- [LIGAND](#)
- [PDB-Ligand](#)
- [PubChem](#)
- [R.E.DD.B.](#)
- [SuperDrug](#)
- [SuperNatural](#)

- Carbohydrates

- [BCSDB/Glycoscience](#)
- [CCSD - Complex Carbohydrate Structure Database \(CarbBank\)](#)
- [CSS - Carbohydrate Structure Suite](#)
- [Glycan](#)
- [Glycoconjugate Data Bank](#)
- [GlycoMapsDB](#)
- [GlycoSuiteDB](#)
- [Monosaccharide Browser](#)
- [SWEET-DB](#)

- Structure Databases

- Nucleic acid structure

- [Greglist](#)
- [GRSDB](#)
- [ITS2](#)
- [MeRNA](#)
- [NCIR - Non-Canonical Interactions in RNA](#)
- [NDB](#)
- [NTDB](#)
- [QuadBase](#)
- [Rfam](#)
- [RNA FRABASE](#)
- [RNA SSTRAND](#)
- [RNABase](#)
- [RNAJunction](#)
- [SARS-CoV RNA SSS](#)
- [SCOR - Structural Classification Of RNA](#)
- [Vir-Mir db](#)

- Protein structure

- [3D-Genomics](#)
- [3DID - 3D interacting domains](#)
- [ArchDB](#)
- [ASTRAL](#)
- [AutoPSI](#)
- [BANMOKI](#)
- [BioMagResBank](#)
- [CADB - Conformational Angles DataBase of Proteins](#)
- [CATH](#)
- [CE](#)
- [CoC Central](#)
- [ColiSNP](#)

- Structure Databases

- Protein structure

- [Columba](#)
- [CSA - Catalytic Site Atlas](#)
- [Dali database](#)
- [DBAli](#)
- [Decoys-R-Us](#)
- [DisProt - Database of Protein Disorder](#)
- [DMAPS](#)
- [Dockground](#)
- [DomIns - Database of Domain Insertions](#)
- [DSDBASE - Disulfide Database](#)
- [DSMM - a Database of Simulated Molecular Motions](#)
- [E-MSD - EBI-Macromolecular Structure Database](#)
- [eF-site - Electrostatic surface of Functional site](#)
- [EzCatDB](#)
- [FireDB](#)
- [FSN](#)
- [Gene3D](#)
- [Genomic Threading Database](#)
- [GTOP - Genomes To Protein structures](#)
- [HOMSTRAD - Homologous Structure Alignment Database](#)
- [HotSprint](#)
- [IMB Jena Image Library](#)
- [IMGT/3Dstructure-DB](#)
- [IMOTdb](#)
- [MALISAM](#)
- [LPFC](#)
- [MegaMotifbase](#)
- [MMDB](#)
- [ModBase](#)
- [MolMovDB - Database of Macromolecular Movements](#)
- [PASS2](#)

- Structure Databases

- Protein structure

- [PDB](#)
- [PDB-REPRDB](#)
- [PDBselect](#)
- [PDBsum](#)
- [PDB\\_TM](#)
- [PepConfDB](#)
- [PFD - Protein Folding Database](#)
- [Phospho3D](#)
- [PIDD](#)
- [PMDB - Protein Model Database](#)
- [Structure Superposition Database](#)
- [ProSAS](#)
- [PROTCOM](#)
- [PRTAD](#)
- [RESID](#)
- [S4: Structure-based Sequence Alignments of SCOP Superfamilies](#)
- [SCOP - Structural Classification Of Proteins](#)
- [SCOPPI](#)
- [SitesBase](#)
- [SNAPPI](#)
- [SSToSS - Sequence-Structural Templates of Single-member Superfamilies](#)
- [STINGreport](#)
- [SUPERFAMILY](#)
- [SURFACE](#)
- [SWISS-MODEL Repository](#)
- [TargetDB](#)
- [TMBETA-GENOME](#)
- [TOPOFIT-DB](#)
- [TOPS - Topology Of Protein Structures](#)

# Protein Data Bank (PDB)

- ❑ <http://www.pdb.org/>
- ❑ Structure data determined by X-ray crystallography and NMR
- ❑ The data include the atom coordinate, reference, sequence, secondary structure, disulfide bond .....etc.



The number of protein structure and the last update date

An Information Portal to Biological Macromolecular Structures  
A MEMBER OF THE PDB  
Wednesday Jun 12, 2012 at 5 PM PDT there are 82347 Structures | PDB Statistics

Search | All Categories: e.g., PDB ID, molecule name, author

All Categories Author Macromolecule Sequence

**Biological Macromolecular Resource**

Full Description

↓ Featured Molecules

Structural View of Biology List View of Archive By: Protein Synthesis

**Molecule of the Month**  
**Sliding Clamps**  
In our genome, it takes thousands of DNA nucleotides to encode the information for each protein, and even more to encode the information for the regulatory information. So, when a cell needs to regulate a long, long stretch of DNA, it has to manage long, long stretches of DNA. As you might imagine, this is not easy in the chaotic environment of the cell.

Full Article

**Protein Structure Initiative Featured System**  
**Anthrax Stealth Siderophores**  
Researchers at MCSI have reconstructed the structure of an essential synthetic enzyme in the pathway.

Full Article | Archive | PSI Structural Biology Knowledgebase

↓ Explore Archive

Organism Taxonomy

Exp. Method X-ray Resolution

Organism

- Homo sapiens (20160)
- Escherichia coli (4603)
- Mus musculus (3541)

**PDB Statistics**  
82347 Structures  
Last Update: Jun 12, 2012

**PDB Statistics**  
79851 Structures  
Last Update: Mar 06, 2012

**PDB Statistics**  
79697 Structures  
Last Update: Feb 28, 2012

**PDB Statistics**  
72550 Structures  
Last Update: Apr 19, 2011

**SMILES**

- Has exact structure C1c1cccc1
- Has sub-structure C1c1cccc1
- Is very similar (95%) with C1c1cccc1
- Is similar (70%) with C1c1cccc1
- Super-structure of C1c1cccc1

Easily perform substructure, exact structure, or similar structure searches. more

- **Molecule of the Month** reaches 150
- **Tour the PDB with Drill-down Pie Charts**

All Categories Author Macromolecule Sequence Ligand  
Search | All Categories: e.g., PDB ID, molecule name, author

PDB Statistics

PDB Current Holdings Breakdown

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	67466	1365	3409	2	72242
NMR	8276	988	186	7	9457
ELECTRON MICROSCOPY	293	22	120	0	435
HYBRID	44	3	2	1	50
other	141	4	5	13	163
Total	76220	2382	3722	23	82347

61642 structures in the PDB have a structure factor file.

6764 structures in the PDB have an NMR restraint file.

526 structures in the PDB have a chemical shifts file.

- DNA Only
- RNA Only
- Protein Nucleic Acid Complexes
- Growth of Unique Protein Classifications Per Year
- As Folds Defined By SCOP
- As Topologies Defined By CATH
- As Superfamilies Defined By SCOP
- As Superfamilies Defined By CATH

Use Search Unreleased to search and view entries that are currently being processed or are awaiting release.

Statistics are for experimentally-determined structures.

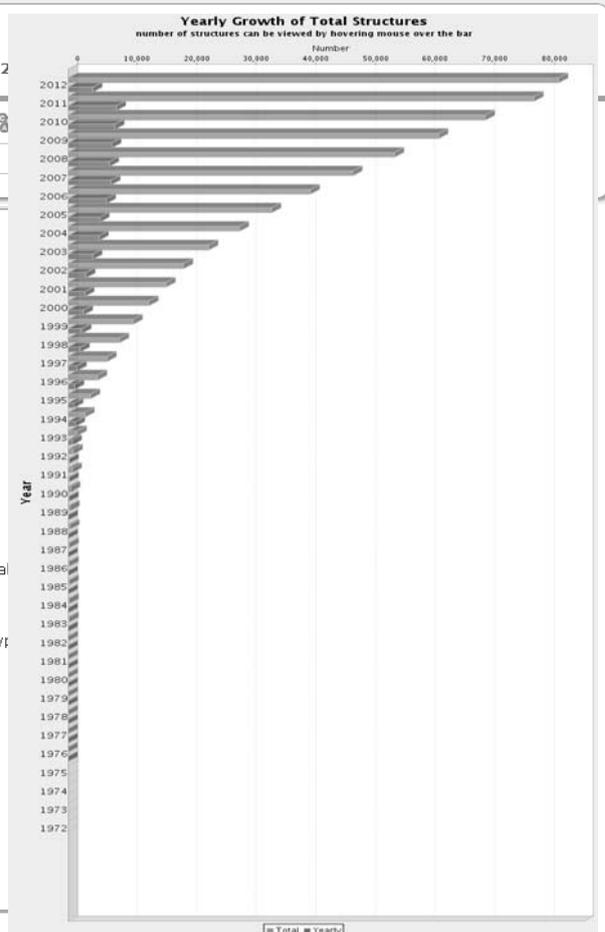
All Categories Author Macromolecule Sequence Ligand  
Search | All Categories: e.g., PDB ID, molecule name, author

PDB Statistics

- Content Distribution
  - Summary Table of Released Entries
  - Status of Unreleased Entries
  - Proteins solved by multiple experimental methods
  - Redundancy based on sequence similarity
  - By Resolution
  - By Space Group
  - By Natural Source Organism
  - By Gene Source Scientific Organism
  - By Top 100 Journals
  - By Structural Genomics Centers
  - By Structure Molecular Weight
  - By Enzyme Classification
- Content Growth
  - Growth of Released Structures Per Year
  - Growth of Released Structures Per Year by Experimental Method
    - X-ray
    - NMR
    - Electron Microscopy
  - Growth of Released Structures Per Year By Molecular Type
    - Protein Only
    - DNA Only
    - RNA Only
    - Protein Nucleic Acid Complexes
  - Growth Of Unique Protein Classifications Per Year
    - As Folds Defined By SCOP
    - As Topologies Defined By CATH
    - As Superfamilies Defined By SCOP
    - As Superfamilies Defined By CATH

Use Search Unreleased to search and view entries that are currently being processed or are awaiting release.

Statistics are for experimentally-determined structures.



RCSB **PDB** PROTEIN DATA BANK **PDB-101** An Information Portal to Biological Macromolecular Structures  
 As of Tuesday Jun 12, 2012 at 5 PM PDT there are 82347 Structures | PDB Statistics

**Search** | All Categories: e.g., PDB ID, molecule name, author

All Categories  Author  Macromolecule  Sequence  Ligand

News & Publications  
 Usage/Reference Policies  
 Deposition Policies  
 Website FAQ  
 Deposition FAQ  
 Contact Us  
 About Us  
 Careers  
 External Links  
 Sitemap  
 New Website Features

All Deposit Services  
 Electron Microscopy  
 X-ray | NMR  
 Validation Server  
 BioSync  
 Beamlines/Facilities  
 Related Tools

Download Files  
 Compare Structures  
 File Formats  
 Services: RESTful | SOAP  
 Widgets

**Molecule of the Month**  
**Sliding Clamps**  
 In our genome, it takes thousands of DNA nucleotides to encode the information for each protein, and even more to store all the regulatory information. So, when a cell needs to copy this information, it has to manage long, long stretches of DNA. As you might imagine, this is not easy in the chaotic environment of the cell.

**Protein Structure Initiative Featured System**  
**Anthrax Stealth Siderophores**  
 Researchers at MCSG have reconstructed the biosynthetic pathway for the stealth siderophore petrobactin, and determined the structure of an essential synthetic enzyme in the pathway.

**Explore Archive**

Latest Release  
 New Structure Papers  
 Search Unreleased Entries

Improved Ligand Summary Report  
 Latest features released:

Weekly | Quarterly | Yearly  
 2012-06-12  
**Top Bar SMILES String Searching**

Easily perform substructure, exact structure, or similar structure searches.

• Molecule of the Month reaches 150  
 • Tour the PDB with Drill-down Pie Charts

# Advanced Search Interface

RCSB **PDB** PROTEIN DATA BANK **PDB-101** An Information Portal to Biological Macromolecular Structures  
 As of Tuesday Jun 12, 2012 at 5 PM PDT there are 82347 Structures | PDB Statistics

**Search** | All Categories: e.g., PDB ID, molecule name, author

All Categories  Author  Macromolecule  Sequence  Ligand

**Advanced Search Interface**

**Choose a Query Type:**  
**ID(s) and Keywords**  
 PDB ID(s)  
 PubMed ID(s)  
 UniProtKB Accession Number(s)  
 Text Search  
 mmCIF Keyword Search (Classification)  
 Pfam ID(s)  
**Structure Annotation**  
 Structure Title  
 Structure Description  
 Macromolecule Name  
**Deposition**  
 Author Name  
 Deposit Date  
 Release Date  
 Revision Date  
 Latest Released Structures  
 Latest Modified Structures  
 Structural Genomics Project



# DNA recognition by GAL4: structure of a protein–DNA complex

Ronen Marmorstein, Michael Carey<sup>†</sup>, Mark Ptashne & Stephen C. Harrison<sup>†</sup>

Harvard University, Department of Biochemistry and Molecular Biology, and <sup>†</sup>Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

A specific DNA complex of the 65-residue, N-terminal fragment of the yeast transcriptional activator, GAL4, has been analysed at 2.7 Å resolution by X-ray crystallography. The protein binds as a dimer to a symmetrical 17-base-pair sequence. A small, Zn<sup>2+</sup>-containing domain recognizes a conserved CCG triplet at each end of the site through direct contacts with the major groove. A short coiled-coil dimerization element imposes 2-fold symmetry. A segment of extended polypeptide chain links the metal-binding module to the dimerization element and specifies the length of the site. The relatively open structure of the complex would allow another protein to bind coordinately with GAL4.

THE yeast protein GAL4 activates transcription of genes required for catabolism of galactose and melibiose<sup>1–3</sup>. The DNA sequences recognized by GAL4 are 17 base pairs (bp) in length<sup>4–6</sup>, and each site binds a dimer of the protein<sup>7</sup>. Four such sites, similar but not identical in sequence, are found in the upstream activating sequence (UAS<sub>G</sub>) that mediates GAL4 activation of the GAL1 and GAL10 genes, for example<sup>8</sup>.

Functions have been assigned to various parts of the 881-amino-acid GAL4 protein (Fig. 1a), including DNA binding

(1–65) is a monomer in the absence of DNA. The open features of the complex, in which a long stretch of DNA at the centre of the 17-bp site is accessible in the major groove, suggest that another protein may be able to bind coordinately with GAL4.

## Structure determination

Crystals in space group *P*<sub>4</sub><sub>3</sub><sub>2</sub><sub>1</sub><sub>2</sub> were prepared as described in the legend to Table 1. The structure of a Cd<sup>2+</sup>-containing complex was determined and refined, because the crystals were of better quality than the isomorphous crystals containing Zn<sup>2+</sup>. Isomorphous derivatives were obtained either by replacing Cd<sup>2+</sup> with Zn<sup>2+</sup> or Hg<sup>2+</sup>, or by preparing duplex DNA in which 5-iodo-uridine was substituted for thymidine in selected positions (Fig. 1; Table 1).

The structure of the cadmium-containing complex was initially determined to 3.2 Å by multiple isomorphous replacement (MIR) using phase information from one Hg<sup>2+</sup> and four 5-iodo-uridine derivatives (Table 1). Locations of the heavy atom derivations confirmed that there was one complete protein–DNA complex per asymmetric unit, and that the protein bound the consensus DNA site as a homodimer. The initial MIR map showed clear density for B-form DNA, and the highest peaks in the map confirmed earlier spectroscopic experiments indicating that each protein monomer bound two closely spaced metal ions<sup>10</sup>. But the protein chain could not be traced. The map was improved by non-crystallographic averaging about a dyad relating the two protein–DNA half-sites<sup>19</sup>. The initial dyad was calculated using heavy-atom positions. Base pairs with ideal B-DNA geometry were built into the twofold averaged map using the model-building program FRODO<sup>20</sup>. The DNA model

## View Structure: 1D66





A MEMBER OF THE 

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jun 12, 2012 at 5 PM PDT there are 82347 Structures | PDB Statistics | 

Search | All Categories:

**MyPDB**

Login to your Account  
Register a New Account

**Home**

News & Publications  
Usage/Reference Policies  
Deposition Policies  
Website FAQ  
Deposition FAQ  
Contact Us  
About Us  
Careers  
External Links  
Sitemap  
New Website Features

**Deposition**

All Deposit Services  
Electron Microscopy  
X-ray | NMR  
Validation Server  
BioSync  
Beamlines/Facilities  
Related Tools

**Tools**

Download Files  
Compare Structures  
File Formats  
Services: RESTful | SOAP  
Widgets

**PDB-101**

Structural View of Biology  
Understanding PDB Data  
Molecule of the Month  
Educational Resources

**Help**

Launch Help System

Summary
Sequence
Annotations
Seq. Similarity
3D Similarity
Literature
Biol. & Chem.
Methods
Geometry
Links

**DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX** 1D66

DOI:10.2210/pdb1d66/pdb NDB ID: PDT003

**Primary Citation**

DNA recognition by GAL4: structure of a protein–DNA complex.  
Marmorstein, R., Carey, M., Ptashne, M., Harrison, S.C.,  
Journal: (1992) Nature 356: 408–414  
PubMed: 1557122  
DOI: 10.1038/356408a0  
Search Related Articles in PubMed

**PubMed Abstract:**  
A specific DNA complex of the 65-residue, N-terminal fragment of the yeast transcriptional activator, GAL4, has been analysed at 2.7 Å resolution by X-ray crystallography. The protein binds as a dimer to a symmetrical 17-base-pair sequence. A small, Zn(2+)-containing domain... [ Read More & Search PubMed Abstracts ]

**Molecular Description**

**Classification:** Transcription/dna  
**Structure Weight:** 27737.04

<b>Molecule:</b> DNA (5'-D(CP*CP*GP*GP*AP*GP*GP*AP*GP*TP*CP*CP*TP*CP*C P*GP*G)-3')	<b>Polymer:</b> 1	<b>Type:</b> dna	<b>Length:</b> 19
<b>Chains:</b> D			
<b>Molecule:</b> DNA (5'-D(CP*CP*GP*GP*AP*GP*GP*AP*GP*TP*GP*TP*CP*CP*TP*CP*C P*GP*G)-3')	<b>Polymer:</b> 2	<b>Type:</b> dna	<b>Length:</b> 19
<b>Chains:</b> E			
<b>Molecule:</b> PROTEIN (GAL4)	<b>Polymer:</b> 3	<b>Type:</b> protein	<b>Length:</b> 66
<b>Chains:</b> A, B			
<b>Organism:</b> Saccharomyces cerevisiae			
<b>UniProtKB:</b> P04386			

**Biological Assembly**



**More Images...**

Other Viewers

Biological assembly 1 assigned by authors

**MyPDB Personal Annotations**

To save personal annotations, please login to your MyPDB account.

**Deposition Summary**

**Authors:** Marmorstein, R., Carey,

# Sequence / Structure Details

Summary **Sequence** Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

**DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX** **1D66** [Display Files](#) [Download Files](#) [Share this Page](#)

**Sequence Display**

The sequence display provides a graphical representation of the UniProtKB, PDB - ATOM and PDB - SEQRES sequences. Different 3rd party annotations can be graphically mapped on the sequence and displayed in the Jmol viewer.

The structure **1D66** has in total **4** chains. Out of these **3** are sequence-unique.

Currently viewing **unique chains** only. [show all chains](#)

**Chain A : PROTEIN (GAL4)**

FASTA | [Sequence & DSSP](#) | [Image](#)  
 Polymer 3  
 Length: 66 residues  
 Chain Type: polypeptide(L)  
 Reference: [UniProtKB P04386](#)

**Annotations**

Domain Assignment: **SCOP**  
 [hide] [reference]  
 Secondary Structure: **DSSP**  
 [hide] [reference]  
 Structural Feature: **Protein Modification** 0252: Metal coordination, CD  
 [hide] [reference] [reference]

*Add Annotations*  
 Select  
 010661 Gal4: 41 residues [\[?\]](#)  
 010662 CD2-Gal4: 16 residues [\[?\]](#)  
 34% helical (3 helices; 23 residues)

SCOP:   
 DSSP:   
 Protein Modification:   
 PDB: MKLLSSIEQACDLCRLKLLKCSKEKPKCAKCLKNWECRYSPKTKRSPLTRAHLTEVESR  
 PDB: N 10 20 30 40 50 60

SCOP:   
 DSSP:   
 Protein Modification:   
 PDB: LERLEF  
 PDB: 61

**Protein Modification Legend**  
 Metal coordination, CD

**Display Parameters**  
 Identical chains: B | [show all chains](#)  
 show polypeptide chains only  
 Currently displayed: **SEQRES** sequence.  
 Display external (UniProtKB) sequence  
 Mouse over an annotation to see more details.

# Biology and Chemistry Report

Summary Sequence Annotations Seq. Similarity 3D Similarity Literature **Biol. & Chem.** Methods Geometry Links

**DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX** **1D66** [Display Files](#) [Download Files](#) [Share this Page](#)

**Biology and Chemistry Report**

**Structure Details** [Hide](#)

**Structure Keywords**

Keywords TRANSCRIPTION/DNA  
 Text PROTEIN-DNA COMPLEX, DOUBLE HELIX, TRANSCRIPTION/DNA COMPLEX

**Polymeric Molecules**

**Chain D**

Description DNA (5'-D(\*CP\*CP\*GP\*GP\*AP\*GP\*GP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*C P\*GP\*G)-3')  
 Nonstandard Linkage no  
 Nonstandard Monomers no  
 Polymer Type polydeoxyribonucleotide  
 Formula Weight 5831.8  
 Source Method synthetic

**Chain E**

Description DNA (5'-D(\*CP\*CP\*GP\*GP\*AP\*GP\*GP\*AP\*CP\*AP\*TP\*GP\*TP\*CP\*CP\*TP\*CP\*C P\*GP\*G)-3')  
 Nonstandard Linkage no  
 Nonstandard Monomers no  
 Polymer Type polydeoxyribonucleotide  
 Formula Weight 5822.8  
 Source Method synthetic

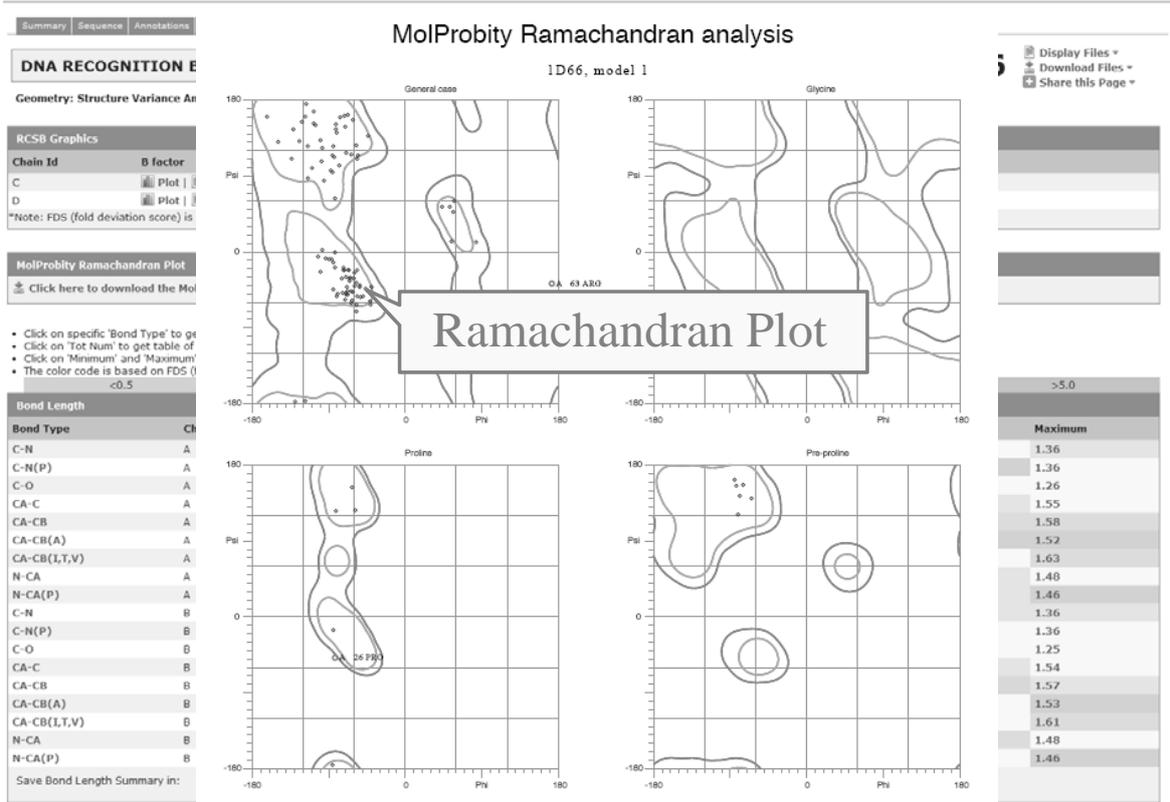
**Chain A,B**

Description PROTEIN (GAL4)  
 Nonstandard Linkage no  
 Nonstandard Monomers no  
 Polymer Type polypeptide(L)  
 Formula Weight 7816.4  
 Source Method genetically manipulated

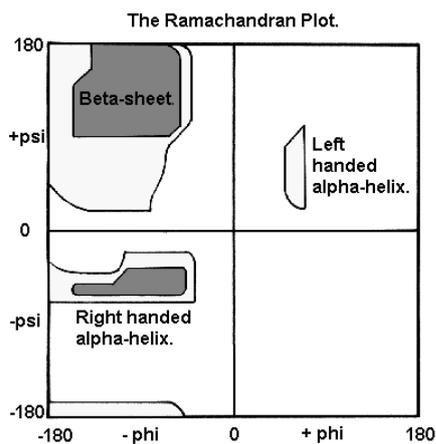
**Ligands and Prosthetic Groups**

ID	Name	Chemical Formula	Weight	Ligand Structure
CD	CADMIUM ION	Cd	112.41	<a href="#">View</a>

# Geometry



# Ramachandran plot



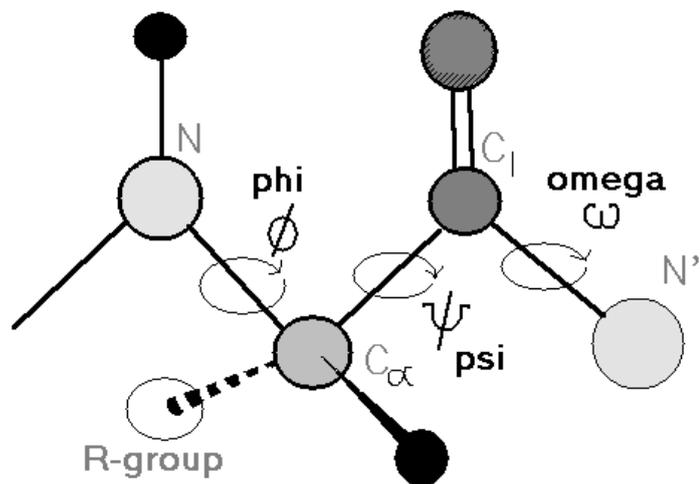
**$\beta$ -strand:**

$$-180 < \phi < -60$$

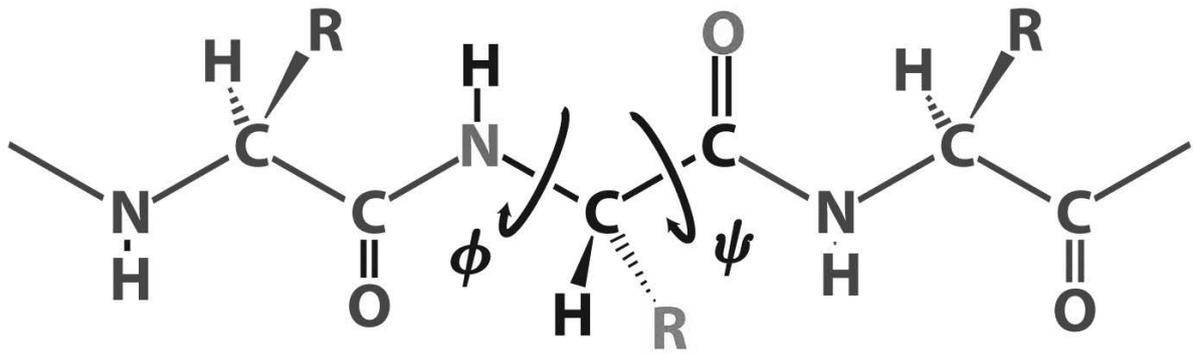
$$180 > \psi > 60$$

**$\alpha$ -helix:**

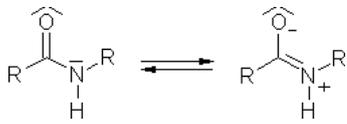
$$\phi: \sim -60$$



$\Phi$  (phi,  $C_{\alpha}$ -N bond) vs.  $\Psi$  (psi,  $C_{\alpha}$ -C(O) bond)

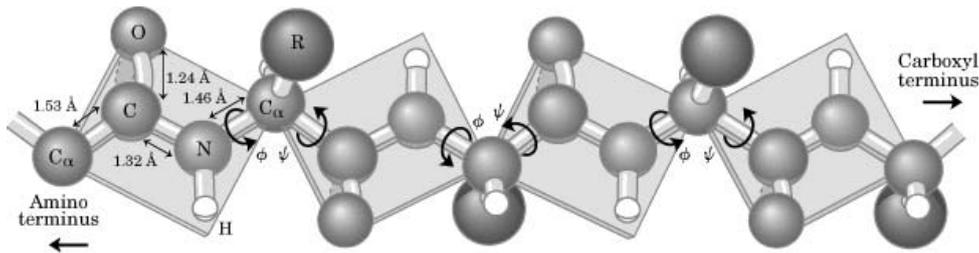


**Figure 2.22a**  
*Biochemistry, Seventh Edition*  
 © 2012 W. H. Freeman and Company



"partial" double bond character

- Peptide bond N–C–O atoms and atoms attached to them lie all in the same plane
  - Peptide bond is planar !
- Only 2 bonds can freely rotate
  - C $\alpha$ –N bond and C $\alpha$ –C(O) bond



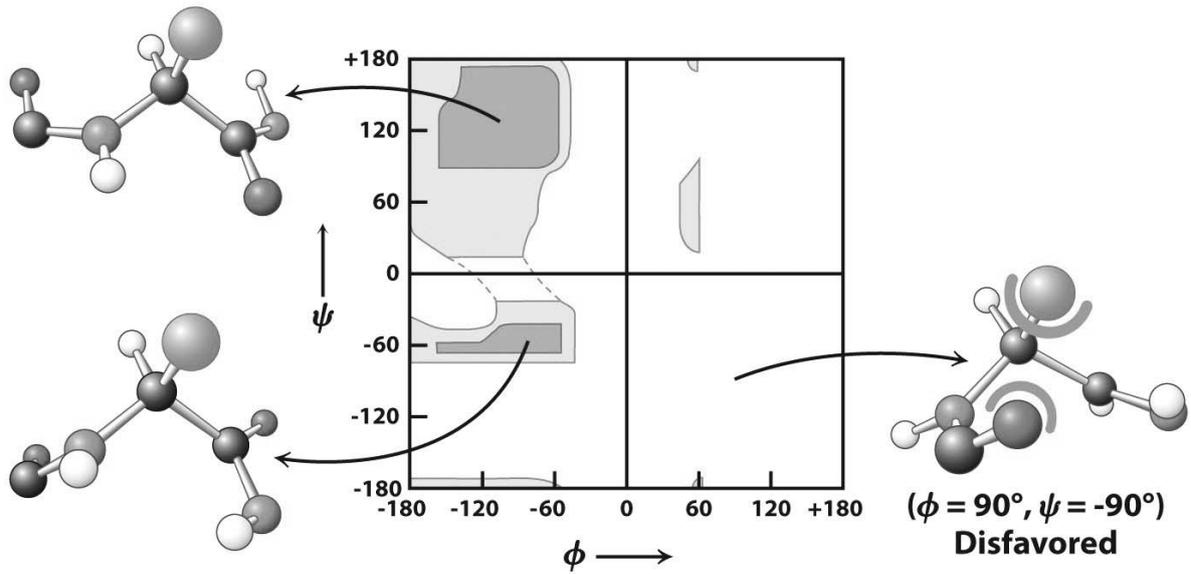
**Limit amount of free rotations possible (high torsion barriers)**

Specified by the **torsion angles**  $\Phi$  (phi, C $\alpha$ –N bond) and  $\Psi$  (psi, C $\alpha$ –C(O) bond)

Possible  $\Phi$  and  $\Psi$  values are **constrained** by the structure of adjacent amino acid residues

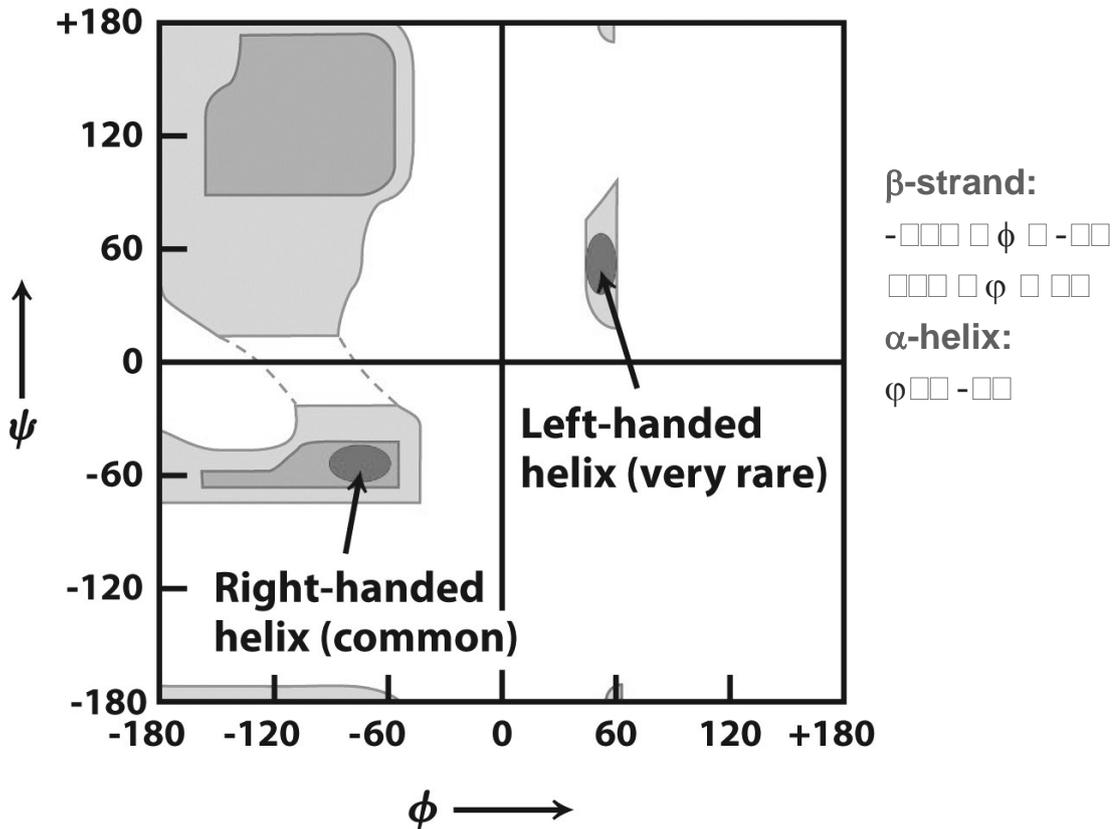
繞N–C $\alpha$ 鍵旋轉的角度稱為phi( $\psi$ )，而繞C $\alpha$ –C'鍵旋轉的角度則稱為psi( $\Psi$ )。因此，每一胺基酸的phi( $\psi$ ) & psi( $\Psi$ )兩個角度決定主鏈原子的型態。

# Ramachandran plot

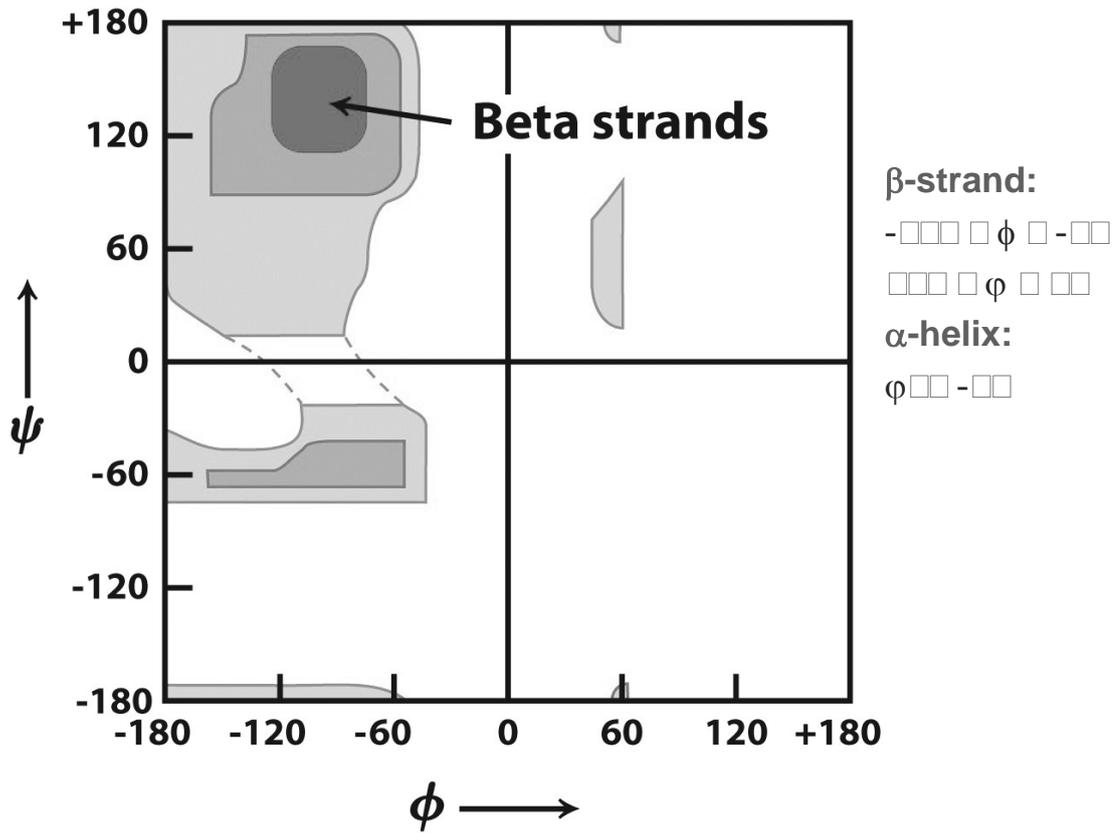


**Figure 2.23**  
*Biochemistry, Seventh Edition*  
© 2012 W. H. Freeman and Company

- □h□□s allowed  $\Phi$  and  $\Psi$  angles
- **White areas** □ sterically disallowed □□□□□□ □i□□s
- □□□□i □ s□□□□□□□s □□□□□□□□hi□ □□□□□□ □□i□□s



**Figure 2.26**  
*Biochemistry, Seventh Edition*  
© 2012 W. H. Freeman and Company



**Figure 2.29**  
*Biochemistry, Seventh Edition*  
 © 2012 W. H. Freeman and Company

## Geometry

Summary | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | **Geometry** | Links

**DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX** **1D66** [Display Files](#) [Download Files](#) [Share this Page](#)

Geometry: Structure Variance Analysis Results

Bond Angle	Chain Id	Tot Num	Cal Ave	Cal StdDev	Std Val	Std StdDev	Minimum	Maximum
<b>Dihedral Angle</b>								
Dihedral Angle	Chain Id	Tot Num	Cal Ave	Cal StdDev	Std Val	Std StdDev	Minimum	Maximum
Chi1 g(+)	A	22	-73.90	17.794	-66.7	15.0	-117.20	-22.80
Chi1 g(-)	A	12	56.52	17.006	64.1	15.7	30.00	84.80
Chi1 trans	A	20	193.05	20.939	183.6	16.8	167.50	236.20
Omega	A	56	176.15	20.302	180	5.8	29.10	188.30
Phi	A	30	-66.85	66.876	-65.3	11.9	-132.20	172.60
Phi helix	A	23	-67.03	15.678	-65.3	11.9	-100.70	-41.90
Phi(P)	A	3	-68.67	10.253	-65.4	11.2	-83.00	-59.60
Psi	A	33	93.14	78.415	-39.4	11.3	-176.50	176.10
Psi helix	A	23	-38.39	19.054	-39.4	11.3	-70.80	3.90
Chi1 g(+)	B	25	-74.08	18.861	-66.7	15.0	-112.50	-36.50
Chi1 g(-)	B	14	63.87	20.734	64.1	15.7	35.80	113.10
Chi1 trans	B	15	196.39	16.726	183.6	16.8	173.60	229.00
Omega	B	56	176.54	23.547	180	5.8	6.00	196.40
Phi	B	29	-79.58	52.756	-65.3	11.9	-163.30	57.00
Phi helix	B	24	-67.62	14.287	-65.3	11.9	-103.40	-35.80
Phi(P)	B	3	-84.83	1.443	-65.4	11.2	-86.00	-82.80
Psi	B	32	91.02	87.488	-39.4	11.3	-177.60	167.20
Psi helix	B	24	-37.06	16.801	-39.4	11.3	-60.40	-0.40

Save Dihedral Angle Summary in:  CSV (Excel) Format

Bond Angle	Chain Id	Tot Num	Cal Ave	Cal StdDev	Std Val	Std StdDev	Minimum	Maximum
N-CA-C	B	54	109.75	4.621	111.2	2.8	97.25	120.74
N-CA-C(P)	B	3	115.34	2.966	111.8	2.5	111.14	117.47
N-CA-CB	B	46	109.83	2.540	110.5	1.7	104.36	116.35
N-CA-CB(A)	B	3	110.38	3.615	110.4	1.5	107.50	115.48
N-CA-CB(I,T,V)	B	5	109.94	3.726	111.5	1.7	104.85	114.36
N-CA-CB(P)	B	3	102.17	0.739	103.0	1.1	101.27	103.08
O-C-N	B	53	121.77	2.344	123.0	1.6	115.70	126.06
O-C-N(P)	B	3	121.96	1.735	122.0	1.4	119.51	123.31

Save Bond Angle Summary in:  CSV (Excel) Format

# Download/Display File

**1D66** Display Files ▾  
Download Files ▾

FASTA Sequence  
**PDB File (Text)**  
PDB File (gz)  
mmCIF File  
mmCIF File (gz)  
PDBML/XML File  
PDBML/XML File (gz)  
Biological Assembly (gz) (A)

**View in Jmol** SimpleViewer  
Other Viewers ▾ Protein Workshop

Biological assembly assigned by authors

```

HEADER  TRANSCRIPTION REGULATION          06-MAR-92  1D66  1D66  2
COMPND  GAL4 (RESIDUES 1 - 65) COMPLEX WITH 19MER DNA          1D66  3
SOURCE  (SACCHAROMYCES $CEREVISIAE) OVEREXPRESSED IN (ESCHERICHIA  1D66  4
SOURCE  2 $COLD)          1D66  5
AUTHOR  R.MARMORSTEIN,S.HARRISON          1D66  6
REVDAT  1 15-NOV-93 1D66  7
JRNL    1 15-NOV-93 1D66  8
JRNL    6 9
JRNL    19
JRNL    21
REMARK  4 THERE ARE TWO DNA CHAINS WHICH HAVE BEEN ASSIGNED CHAIN  1D66 24
REMARK  4 INDICATORS *D* AND *E*. THERE ARE TWO PROTEIN CHAINS  1D66 25
REMARK  4 WHICH HAVE BEEN ASSIGNED CHAIN INDICATORS *A* AND *B*.  1D66 26
REMARK  4 EACH PROTEIN - DNA COMPLEX CONTAINS FOUR BOUND CD IONS.  1D66 27
REMARK  5          1D66 28
REMARK  5 THE PROTEIN CONTAINS THE N-TERMINAL 65 RESIDUES OF GAL4  1D66 29
REMARK  5 PLUS A C-TERMINAL PHE DERIVED FROM THE CLONING CONSTRUCT.  1D66 30
REMARK  6          1D66 31
REMARK  6 RESIDUES LEU A 19 - LYS A 27 AND LEU B 19 - LYS B 27 FORM  1D66 32
REMARK  6 TIGHT TURNS WHICH CONNECT HELICES. RESIDUES TRP A 39 -  1D66 33
REMARK  6 LEU A 49 AND TRP B 39 - LEU B 49 FORM EXTENDED CHAINS  1D66 34
REMARK  6 WHICH CONNECT HELICES.          1D66 35

```

\*Use WordPad to view the text file

# Download/Display File

**1D66** Display Files ▾  
Download Files ▾

FASTA Sequence  
**PDB File (Text)**  
PDB File (gz)  
mmCIF File  
mmCIF File (gz)  
PDBML/XML File  
PDBML/XML File (gz)  
Biological Assembly (gz) (A)

**View in Jmol** SimpleViewer  
Other Viewers ▾ Protein Workshop

Biological assembly assigned by authors

```

HET CD  41  1  CADMIUM          1D66 68
HET CD  42  1  CADMIUM          1D66 69
FORMUL  5  CD  4(CD1)          1D66 70
FORMUL  6  HOH  *51(H2 O1)     1D66 71
HELIX   1  H1A CYS A  11  LYS A  18  1  RESIDUE 18 HAS POSITIVE PHI  1D66 72
HELIX   2  H2A CYS A  28  ASN A  35  1  RESIDUE 35 HAS POSITIVE PHI  1D66 73
HELIX   3  H3A THR A  50  LEU A  64  1          1D66 74
HELIX   4  H1B CYS B  11  LYS B  18  1  RESIDUE 18 HAS POSITIVE PHI  1D66 75
HELIX   5  H2B CYS B  28  ASN B  35  1  RESIDUE 35 HAS POSITIVE PHI  1D66 76
HELIX   6  H3B THR B  50  LEU B  64  1          1D66 77
CRYST1  80.850  80.850  73.700  90.00  90.00  90.00 P 43 21 2  8 1D66 78
ORIGX1  1.000000  0.000000  0.000000  0.00000  1D66 79
ORIGX2  0.000000  1.000000  0.000000  0.00000  1D66 80
ORIGX3  0.000000  0.000000  1.000000  0.00000  1D66 81
SCALE1  0.012369  0.000000  0.000000  0.00000  1D66 82
SCALE2  0.000000  0.012369  0.000000  0.00000  1D66 83
SCALE3  0.000000  0.000000  0.013569  0.00000  1D66 84
MTRIX1  1  0.969990  0.014680 -0.242700  7.19246  1  1D66 85
MTRIX2  1  0.014290 -0.999900 -0.003900  83.38941  1  1D66 86
MTRIX3  1 -0.242710 -0.000190 0.970100  2.87497  1  1D66 87
ATOM    1  O5*  C D  1  23.081  73.401  36.511  1.00  44.77  1D66 88
ATOM    2  C5*  C D  1  24.340  73.259  35.792  1.00  46.46  1D66 89
ATOM    3  C4*  C D  1  24.267  72.789  34.262  1.00  42.04  1D66 90
ATOM    4  O4*  C D  1  25.550  72.957  33.595  1.00  41.08  1D66 91
ATOM    5  C3*  C D  1  23.957  71.289  34.142  1.00  38.19  1D66 92
ATOM    6  O3*  C D  1  23.249  71.081  32.947  1.00  33.45  1D66 93
ATOM    7  C2*  C D  1  25.339  70.690  33.983  1.00  35.90  1D66 94
ATOM    8  C1*  C D  1  26.031  71.694  33.078  1.00  39.17  1D66 95
ATOM    9  N1  C D  1  27.530  71.609  33.190  1.00  38.42  1D66 96
ATOM   10  O2  C D  1  28.318  71.429  32.033  1.00  32.78  1D66 97
ATOM   11  C2  C D  1  27.833  71.357  30.908  1.00  30.98  1D66 98
ATOM   12  N3  C D  1  29.661  71.362  32.174  1.00  28.51  1D66 99
ATOM   13  C4  C D  1  30.215  71.469  33.389  1.00  30.53  1D66 100
ATOM   14  N4  C D  1  31.535  71.390  33.519  1.00  28.65  1D66 101

```

# PDB File Title Section

HEAD	First line of the entry, contains PDB ID code, classification, and date of deposition.	HELIX	Identification of helical substructures.
COMPND	Description of macromolecular contents of the entry.	CRYST1	Unit cell parameters, space group, and Z.
SOURCE	Biological source of macromolecules in the entry.	ORIGXn	Transformation from orthogonal coordinates to the submitted coordinates (n = 1, 2, or 3). 由直角(orthogonal)座標系，轉換到 submitted座標系，座標系之間的轉換
AUTHOR	List of contributors.	SCALEn	Transformation from orthogonal coordinates to fractional crystallographic coordinates (n = 1, 2, or 3).由直角座標系，轉換到晶圖(s <sub>1</sub> s <sub>2</sub> s <sub>3</sub> phi)座標系，座標系之間的轉換值。
REVDAT	Revision date and related information.	MTRIXn	Transformations expressing non-crystallographic symmetry (n = 1, 2, or 3). There may be multiple sets of these records. 非晶圖對稱的轉換
JRNL	Literature citation that defines the coordinate set.	ATOM	Atomic coordinate records for standard groups.
REMARK	General remarks, some are structured and some are free form.	HETATM	Atomic coordinate records for heterogens.
SEQRES	Primary sequence of backbone residues.	TER	Chain terminator.
FORMUL	Chemical formula of non-standard groups.	END	Last record in the file.

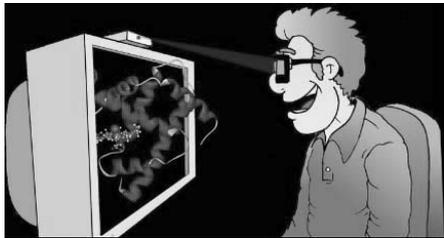
The screenshot shows the PDB website interface for entry 1P66. The title is "DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX". The page is divided into several sections:

- External Links:** A list of links to related databases and resources, including Protein Databank in Europe (PDBe), Protein Data Bank Japan (wwPDB Partner) (PDBJ), PSI Structural Biology Knowledgebase (PSI/KB), Protein Interfaces, Surfaces and Assemblies (PISA), Molecular Modeling Database (NCBI/Entrez) (MMDB), PDBeSum, Jena Library, PDBWiki, Proteopedia, OCA Browser (OCA), and PDB\_SEDO.
- STRUCTURE FEATURES:** A list of features related to the structure, including Homology derived Secondary Structure of Proteins (HSSP), Analysis of Ligand-Protein Contacts (LPC), Analysis of Interatomic Contacts of Structural Units (CSU), Computed Atlas of Surface Topography of proteins (CASTp), and Gaussian Network Model (GNM).
- LIGAND FEATURES:** A list of features related to ligands, including BindingDB, Ligand-Expo, Chem BLAST, PubChem, and DrugBank.
- STRUCTURE CLASSIFICATION AND COMPARISON:** A list of classification and comparison methods, including Structural Classification of Proteins (SCOP), Protein Structure Classification (CATH), Vector Alignment Search Tool (VAST), Flexible structure Alignment by Chaining Aligned fragment pairs allowing Twists (FATCAT), DALI, and SUPERFAMILY.
- SECONDARY STRUCTURE:** A list of secondary structure assignments, including Secondary Structure Assignments (DSSP).
- EXPERIMENTAL DATA:** A section for experimental data, currently empty.
- BIOLOGICAL DETAILS:** A list of biological details, including CSA and IEDB.
- PATHWAYS:** A list of pathways, including METACYC.
- PROTEIN MOTIONS:** A list of protein motions, including Molecular Movements Database (MMD).
- STEREOCHEMICAL QUALITY:** A list of stereochemical quality checks, including WHAT\_CHECK (WHAT IF) and PROCHECK.

An arrow points to the "Links" section, which is highlighted in the image.

# Molecular Graphics

- MDL Chime
- RasMol
- PyMOL
- Cn3D
- Swiss PDB viewer
- MOLMOL
- MolScript
- Raster3D
- GRASP
- WebLabViewer



<http://alpha.life.nthu.edu.tw/>

A screenshot of the NTHU Bioinformatics website. The page is titled 'Download' and lists various molecular graphics software. The list includes: 1. RasMol 2.7.5 (新版), 2. RasMol Help File (rasmol.hlp) (Command mode), 3. RasMol Help File (raswin.hlp) (Windows HELP file), 4. MDL Chime SP6, 5. Swiss-pdb viewer, 6. Cn3D, 7. Pymol, 8. 蛋白質結構下載\_1D66. The page also features a search bar, a 'LATEST NEWS' section, and a 'POPULAR' section. A large arrow points from the left towards the 'Download' section, and a smaller arrow points from the 'Download' section towards the 'PyMol' entry.

**Thanks for your attentions.**

