



plik PDB

- Do zapisu danych w pliku PDB używa się znaków ASCII o graficznej reprezentacji czyli:

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890
` - = [] \ ; ' , . / ~ ! @ # \$ % ^ & * () _ + { } | : " < > ?
oraz spacji i znaku końca linii.

Dodatkowo:

- litery greckie zastępuje się ich postacią fonetyczną, np. alpha, beta, gamma itd.;
- strzałki reprezentowane są przez „-->” lub „<-- ”;
- przecinki, dwukropki i średniki służą do rozdzielania elementów list. (Jeśli znaki te mają mieć inny charakter uwalnia się je z pomocą ukośnika „\”.)

- W formacie PDB ściśle określona została również kolejność występowania rekordów, jak i konieczność pojawienia się niektórych z nich. W przypadku gdy obowiązkowe dane nie są dostępne, w rekordzie powinien pojawić się komunikat „NULL” . Gdy informacje wprowadzane do danego rekordu przekraczają długość wiersza, często następuje ich kontynuacja w kolejnych liniach, przy czym każda z nich musi zostać zidentyfikowana nazwą danego rekordu. Do rekordów obowiązkowych należą: HEADER, TITLE, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, REMARK 2, REMARK 3, CRYST1, ORIGX1, ORIGX2, ORIGX3, SCALE1, SCALE2, SCALE3, MASTER oraz END. Pozostałe mogą mieć obowiązek pojawienia się w pewnych specyficznych przypadkach.

Charakterystyka rekordów

- **Część tytułowa [Zawierająca opis eksperymentu oraz prezentowanej makromolekuły]**
- **Struktury pierwszorzędowe [Określa sekwencję reszt w każdym łańcuchu makrocząsteczki.]**
- **Heterogen [Opisuje reszty niestandardowe.]**
- **Struktury drugorzędowe**
- **Wiązania chemiczne**
- **Krystalografia i transformacja współrzędnych**
- **Współrzędne atomów**
- **Część podsumowująca**

Część tytułowa

HEADER Zawiera klasyfikację cząsteczki, datę wprowadzenia danych oraz kod PDB-ID.

OBSLTE Pojawia się w plikach wycofanych z dystrybucji. Udziela informacji o plikach które go zastąpiły.

TITLE Poza tytułem eksperymentu może zawierać informacje np. o rodzaju eksperymentu, mutacjach czy wprowadzeniu do pliku tylko danych na temat węgla alfa.

CAVEAT Ostrzega o błędach danych związanych z chiralnością.

COMPND Opisuje cząsteczkę przedstawioną w pliku, przez podanie min. nazwy, synonimów, listy łańcuchów składowych, metody otrzymania cząsteczki, mutacji.

SOURCE Identyfikuje biologiczne lub/i chemiczne pochodzenie opisywanej cząsteczki (np. klasyfikacja organizmu, z którego pochodzi, linii komórek użytych w eksperymencie, typ komórki, lokację wewnątrz komórki, identyfikację genu, plazmidu, systemu ekspresji.) Wykorzystuje nazwy formalne jak i zwyczajowe.

KEYWDS Zawiera zbiór terminów związanych z plikiem, jak również rozszerzenie treści wiersza HEADER. EXPDTA Prezentuje informacje o eksperymencie, takie jak typ wykorzystanego promieniowania, czy technikę spektroskopii lub modelowania. (np. ELEKTRON DIFFRACTION, ELECTRON MICROSKOPY, FLUORESCENCE TRANSFER, NMR)

AUTHOR Podaje imiona ludzi odpowiedzialnych za zawartość pliku.

REVDAT Zawiera historię wprowadzanych do pliku modyfikacji.

SPRSDE Lista kodów ID plików usuniętych uprzednio z dystrybucji, powiązanych z plikiem bieżącym.

JRNL Charakteryzuje główną publikację opisującą eksperyment, którego wyniki są podstawą danych zawartych w pliku. Pozostałe publikacje uwzględnione zostają w REMARK1.

REMARK n Prezentuje szczegóły dotyczące eksperymentu, adnotacje, komentarze i inne informacje nie zawarte w pozostałych rekordach. Przykładowo: REMARK 1 – lista ważnych publikacji powiązanych z przedstawioną cząsteczką; REMARK 2 – najlepsza rozdzielczość wykorzystana do budowy modelu; REMARK 3 – programy zostały użyte do zwiększenia dokładności danych krystalograficznych; REMARK 4 – wersja formatu PDB użytą do wygenerowania danego pliku.

Struktury pierwszorzędowe

DBREF Zawiera odnośniki do innych reprezentacji opisanych sekwencji.

SEQADV Identyfikuje konflikty pomiędzy informacjami o sekwencjach w rekordzie SEQRES a innymi bazami danych, uwzględnionymi w rekordzie DBREF.

SEQRES Zawiera sekwencję aminokwasów oraz kwasów nukleinowych w każdym łańcuchu opisywanej cząsteczki.

MODRES Opisuje modyfikacje białka lub kwasu nukleinowego, określa jak zmieniono nazwy użyte w pliku w stosunku do nazw standardowych.

Heterogen

HET Opisuje niestandardowe reszty, takie jak grupy prostetyczne, inhibitory czy jony, których współrzędne zostały określone (min. hetID, identyfikator łańcucha, numer sekwencji, numer HETATM opisującego daną grupę)

HETNAM Podaje nazwy chemiczne związków wraz z kodami hetID.

HETSYN Podaje istniejące synonimy związków określonych w HETNAM.

FORMUL Podaje formułę chemiczną oraz ładunek(?) grup niestandardowych.

Struktury drugorzędowe

HELIX Identyfikuje pozycję helisy alfa w cząsteczce, przyporządkowuje im nazwy oraz numery. Zaznaczone zostają początkowe i końcowe reszty, oraz całkowitą długość helisy.

SHEET Identyfikuje harmonijki beta w sposób analogiczny jak helisy.

TURN Identyfikuje struktury „turn” (jak wyżej).

Wiązania chemiczne

SSBOND Identyfikuje wszystkie mostki dwusiarczkowe w białkach i polipeptydach, poprzez oznaczenie reszt połączonych tymi wiązaniami.

LINK Opisuje połączenia pomiędzy resztami nie uwzględnionymi w strukturach pierwszorzędowych.

CISPEP Charakteryzuje proliny i inne białka znajdujące się w konformacji cis.

CONNECT Prezentuje wiązania pomiędzy atomami, których współrzędne zostały podane. Rekordy te umieszcza się poniżej części określającej współrzędne atomowe.

Krystalografia i transformacja współrzędnych

CRYST1 Opisuje parametry komórki elementarnej, grupę przestrzenną i wartość Z.

ORIGXn Prezentuje transformację współrzędnych prostokątnych na współrzędne ukośne. (n = 1, 2 lub 3)

SCALEn Prezentuje transformację współrzędnych prostokątnych na ułamkowe współrzędne krystalografii. Niestandardowy system współrzędnych powinien być objaśniony w rekordzie REMARK. (n = 1, 2 lub 3)

MTRIXn Prezentuje transformację symetrii nieokreślonej krystalograficznie.

TVECT Prezentuje przesunięcie wektora dla struktur powiązanych kowalencyjnie.

Współrzędne atomów

MODEL Zawiera numer porządkowy modelu w przypadku struktur złożonych.

ATOM Prezentuje współrzędne atomów reszt standardowych.

SIGATM Odchylenie standardowe parametrów atomowych które pojawiają się w rekordach ATOM I HETATM.

ANISOU Anizotropowy czynnik temperaturowy.

SIGUIJ Odchylenie standardowe anizotropowego czynnika temperaturowego skalowanego przez 10^{**4} .

TER Wskazuje koniec listy rekordów ATOM i HETATOM dla danego łańcucha.

HETATM Współrzędne atomów reszt niestandardowych. Rekordy te są używane też dla cząsteczek wody w grupach HET.

ENDMDL Rekord ten jest sprzężony z rekordem MODEL w celu pogrupowania indywidualnych struktur.

Część podsumowująca

MASTER Zawiera listę liczb linii poświęconych poszczególnym rekordom. (kolejno: REMARK, HET, HELIX, SHEET, TURN, SITE, ORIGXn+SCALEn+MTRIX, ATOM+HETATM, TER, CONECT, SEQRES)

END Oznacza ostatni wiersz pliku.

- Plik PDB składa się z wierszy rozdzielonych znakami końca linii.
- Pole wpisu zawiera 80 pionowych kolumn, z których każda odpowiada jednemu znakowi graficznemu.

ATOM lines in PDB files

```

column          1          2          3          4          5          6          7          8
number 1234567890123456789012345678901234567890123456789012345678901234567890

```

```

ATOM 1 N GLY A 3 17.119 0.186 36.320 1.00 64.10 N
ATOM 2 CA GLY A 3 16.944 -0.800 35.208 1.00 63.46 C
ATOM 3 C GLY A 3 16.818 -0.087 33.851 1.00 61.22 C
ATOM 4 O GLY A 3 15.721 0.337 33.463 1.00 62.81 O
ATOM 5 N PRO A 4 17.944 0.077 33.129 1.00 57.39 N
ATOM 6 CA PRO A 4 17.950 0.742 31.815 1.00 53.27 C
ATOM 7 C PRO A 4 18.005 -0.247 30.629 1.00 49.78 C
ATOM 8 O PRO A 4 19.086 -0.678 30.218 1.00 48.17 O
ATOM 9 CB PRO A 4 19.191 1.613 31.898 1.00 54.33 C
ATOM 10 CG PRO A 4 20.161 0.686 32.625 1.00 55.45 C
ATOM 11 CD PRO A 4 19.305 0.019 33.701 1.00 55.83 C

```

```

field id 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

```

field id	definition	length	format	range	string slicing (Python)
1	"ATOM " or "HETATM"	6	%-6s	01-06	[0:6]
2	atom serial number	5	%5d	07-11	[6:11]
3	atom name	4	%4s	13-16	[12:16]
4	alternate location indicator	1	%1s	17	[16:17]
5	residue name	3	%3s	18-20	[17:20]
6	chain identifier	1	%1s	22	[21:22]
7	residue sequence number	4	%4d	23-26	[22:26]
8	code for insertion of residues	1	%1s	27	[26:27]
9	orthogonal coordinates for X (in Angstroms)	8	%8.3f	31-38	[30:38]
10	orthogonal coordinates for Y (in Angstroms)	8	%8.3f	39-46	[38:46]
11	orthogonal coordinates for Z (in Angstroms)	8	%8.3f	47-54	[46:54]
12	occupancy	6	%6.3f	55-60	[54:60]
13	temperature factor	6	%6.3f	61-66	[60:66]
14	element symbol	2	%2s	77-78	[76:78]
15	charge on the atom	2	%2s	79-80	[78:80]

Atomic Coordinates: PDB Format

		Amino Acid		Chain name	Sequence Number	-----Coordinates-----			
	Element					X	Y	Z	(etc.)
ATOM	1	N	ASP L	1	1	4.060	7.307	5.186	...
ATOM	2	CA	ASP L	1	1	4.042	7.776	6.553	...
ATOM	3	C	ASP L	1	1	2.668	8.426	6.644	...
ATOM	4	O	ASP L	1	1	1.987	8.438	5.606	...
ATOM	5	CB	ASP L	1	1	5.090	8.827	6.797	...
ATOM	6	CG	ASP L	1	1	6.338	8.761	5.929	...
ATOM	7	OD1	ASP L	1	1	6.576	9.758	5.241	...
ATOM	8	OD2	ASP L	1	1	7.065	7.759	5.948	...

\\
Element position within amino acid