

BANK MATTS2021: WYKORZYSTANIE KLASTROWANIA DO INTERPRETACJI GĘSTOŚCI ELEKTRONOWEJ TYPÓW ATOMOWYCH

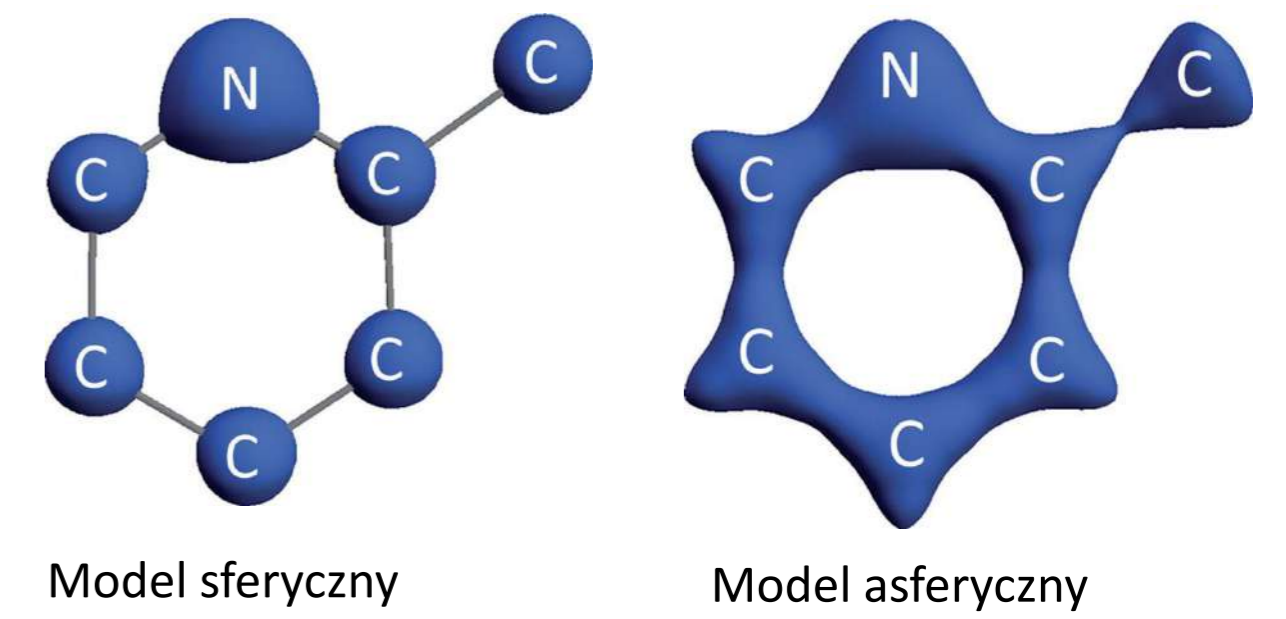
Paulina M. Rybicka, Marta Kulik, Michał L. Chodkiewicz, Paulina M. Dominiak

Centrum Nauk Biologiczno-Chemicznych, Uniwersytet Warszawski
ul. Żwirki i Wigury 101, 02-089, Warszawa

E-mail: p.rybicka@uw.edu.pl



Model Multipolowy (MM) wykorzystuje podejście **asferyczne** do opisu gęstości elektronowej i może być używany do interpretacji danych z dyfrakcji rentgenowskiej w sposób dokładniejszy niż przy użyciu przybliżenia sferycznego. Model asferyczny, w przeciwieństwie do sferycznego, **opisuje deformację gęstości elektronowej związaną z powstawaniem wiązań kowalencyjnych i obecności wolnych par elektronowych**. Tworzenie banku danych do opisu gęstości elektronowej w MM opiera się o teorię przenaszalności parametrów gęstościowych. Jeśli w wielu różnych cząsteczkach znajdują się atomy tego samego pierwiastka o identycznym otoczeniu chemicznym, to ich parametry gęstościowe mogą zostać uśrednione i tworzona jest definicja typu atomowego, przechowywana w banku. Bank danych MATTS (Multipolar Atom Types from Theory and Statistical clustering) gromadzi parametry MM specyficzne dla typów atomowych w białkach, kwasach nukleinowych i cząsteczkach organicznych. Do tej pory nie było jednak w pełni zrozumiałe, w jaki sposób gęstość elektronowa poszczególnych atomów reaguje na ich otoczenie i jakie czynniki opisują gęstość elektronową w cząsteczkach w MM. Bank danych MATTS zastąpił University at Buffalo DataBank (UBDB). W chwili obecnej bank danych MATTS2021 zawiera 651 typów atomowych sklasyfikowanych według rodzaju pierwiastka centralnego i zawiera informacje o rozkładzie gęstości elektronowej, strukturze i sąsiedztwie.

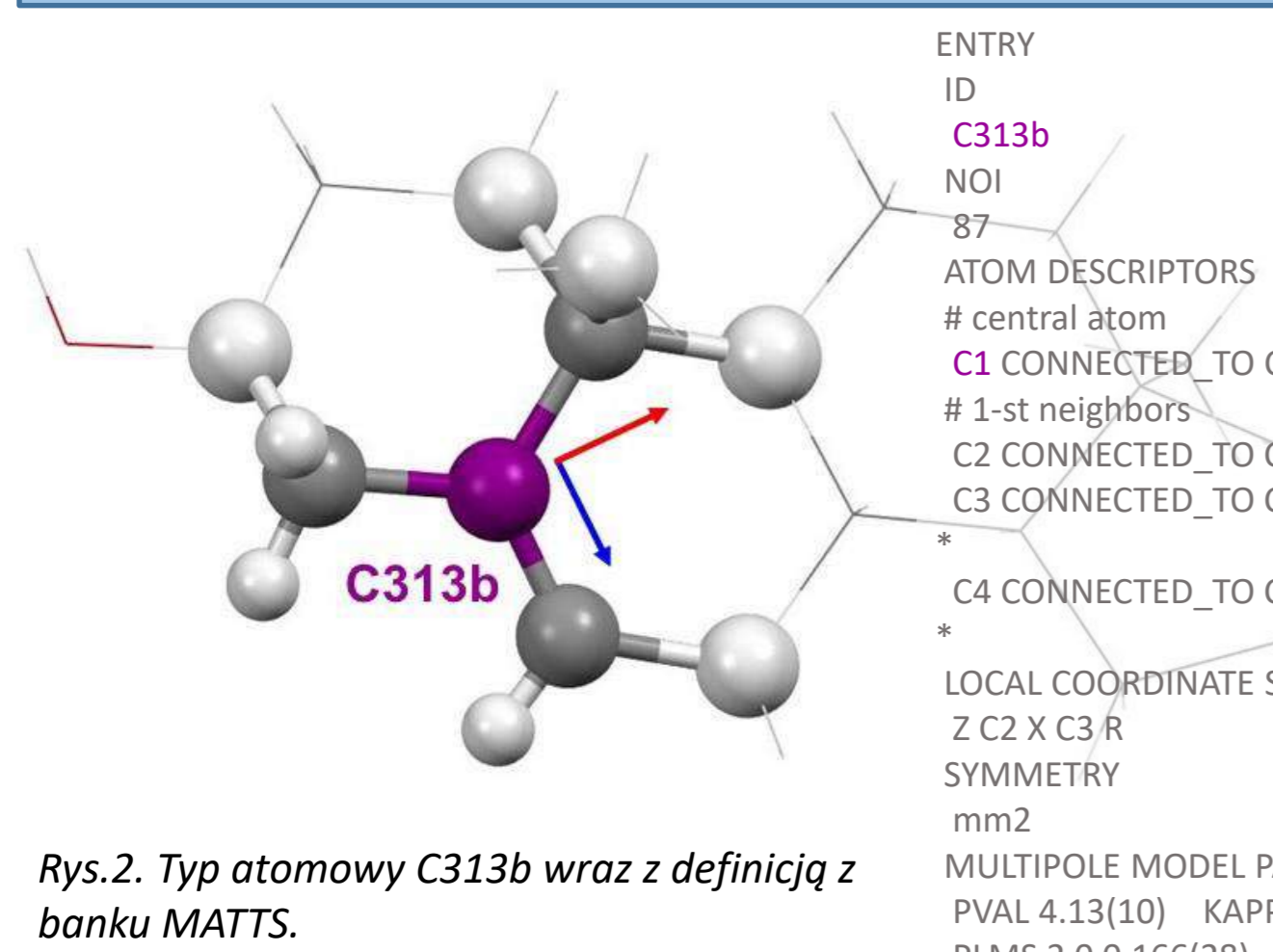


Rys.1. Porównanie modeli gęstości elektronowej.

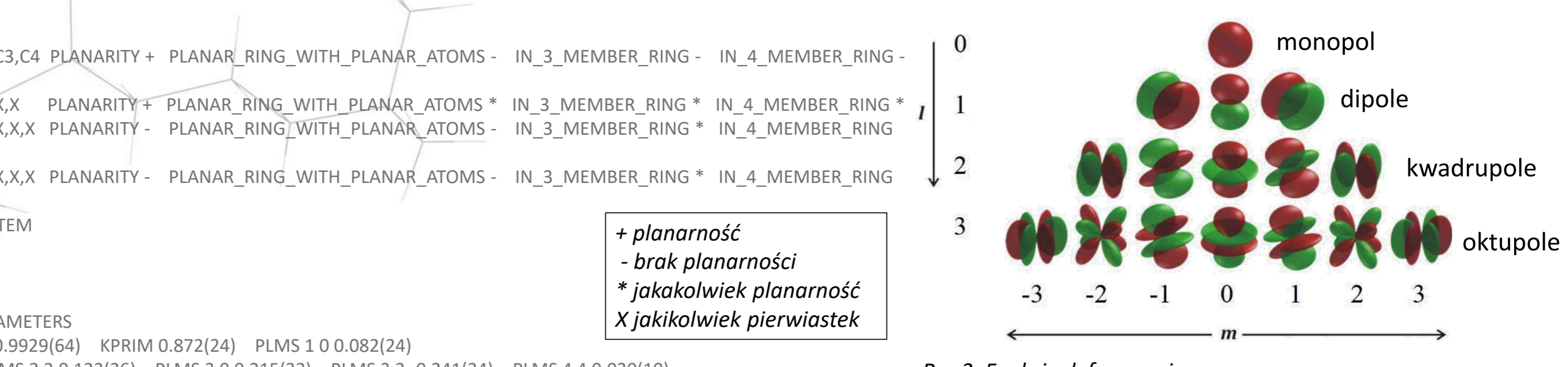
Celem projektu było pełne wykorzystanie informacji zawartych w banku MATTS poprzez znalezienie podobieństw i różnic między typami atomowymi, analizę pełnego obrazu gęstości elektronowej uzyskanego przez możliwe lokalne układy współrzędnych oraz zilustrowanie powiązań między typami atomów na podstawie topologia lub/i parametrów gęstościowych.

W MM, znanym również jako **formalizm Hansena-Coppensa**, podstawowe równanie opisujące gęstość elektronową zawiera trzy składniki: **sferyczny jądrowy, sferyczny walencyjny oraz deformacyjny walencyjny**. Deformacyjna gęstość elektronowa jest opisywana przez **populacje funkcji deformacyjnych $P_{lm\pm}$**

$$\rho_{atom}(r, \theta, \phi) = P_{core}\rho_{core}(r) + P_{val}\kappa^3\rho_{val}(\kappa r) + \sum_{l=0}^{l_{max}} \kappa^l R_l(\kappa^l r) \sum_{m=-l}^l P_{lm\pm} Y_{lm\pm}(\theta, \phi)$$



Rys.2. Typ atomowy C313b wraz z definicją z banku MATTS.



Rys.3. Funkcje deformacyjne.

Każdy typ atomowy w banku MATTS jest wynikiem uśrednienia parametrów gęstości (κ , P_{val} , κ' , $P_{lm\pm}$) dla rodziny chemicznie równoważnych atomów. Geometria modelu oparta jest na danych eksperymentalnych z bazy danych **CSD (Cambridge Structural Database)**. Funkcję falową i gęstość elektronową cząsteczki modelowej w próżni uzyskuje się metodami kwantowo-mechanicznymi.

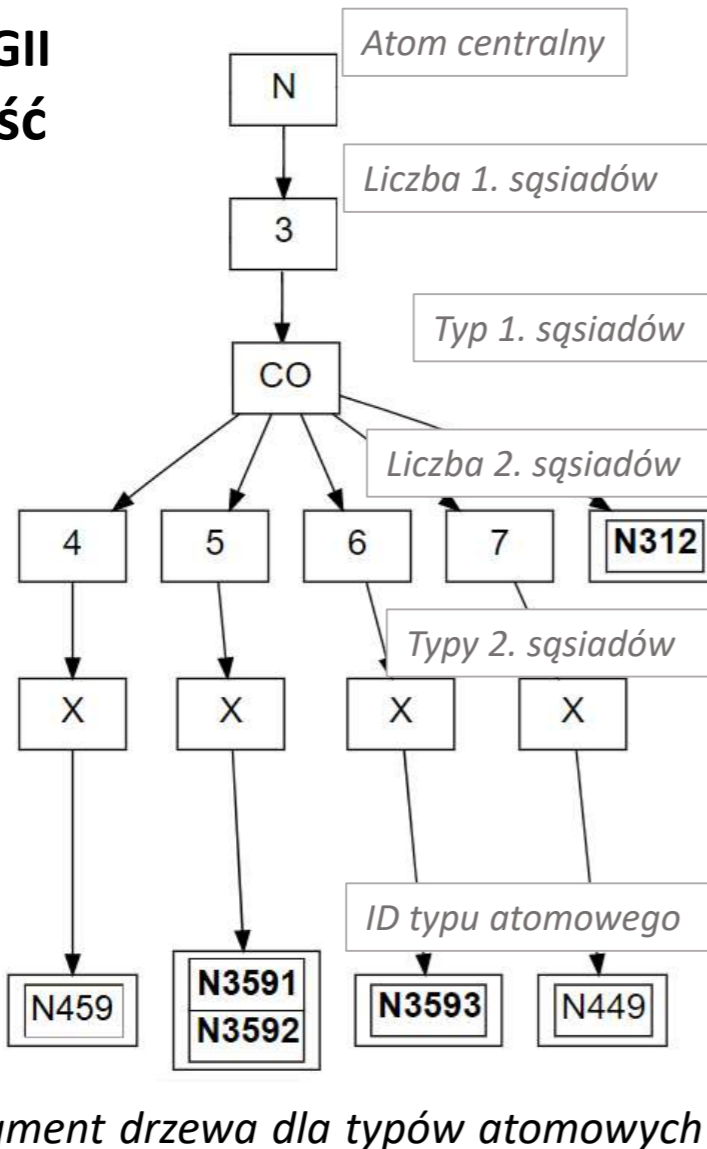
Definicja typu atomowego w banku MATTS zawiera informacje na temat:

Pierwiastka chemicznego opisywanego typu	Liczby i typu atomów połączonych z atomem centralnym (tzw. pierwsi sąsiedzi)	Planarności grupy (atom centralny i sąsiedzi)	Czy typ atomowy należy do płaskich pierścieni	Czy typ atomowy należy do pierścieni 3- lub 4- członowych	Lokalne symetrii	Parametrów gęstościowych (κ , P_{val} , κ' , $P_{lm\pm}$)
------------------------------------------	------------------------------------------------------------------------------	-----------------------------------------------	-----------------------------------------------	-----------------------------------------------------------	------------------	-----------------------------------------------------------------------------

1. KLASTROWANIE OPARTE NA TOPOLOGII

- Wykorzystuje **sąsiadów i planarność**
- Zwizualizowane w hierarchicznej strukturze („drzewa”)

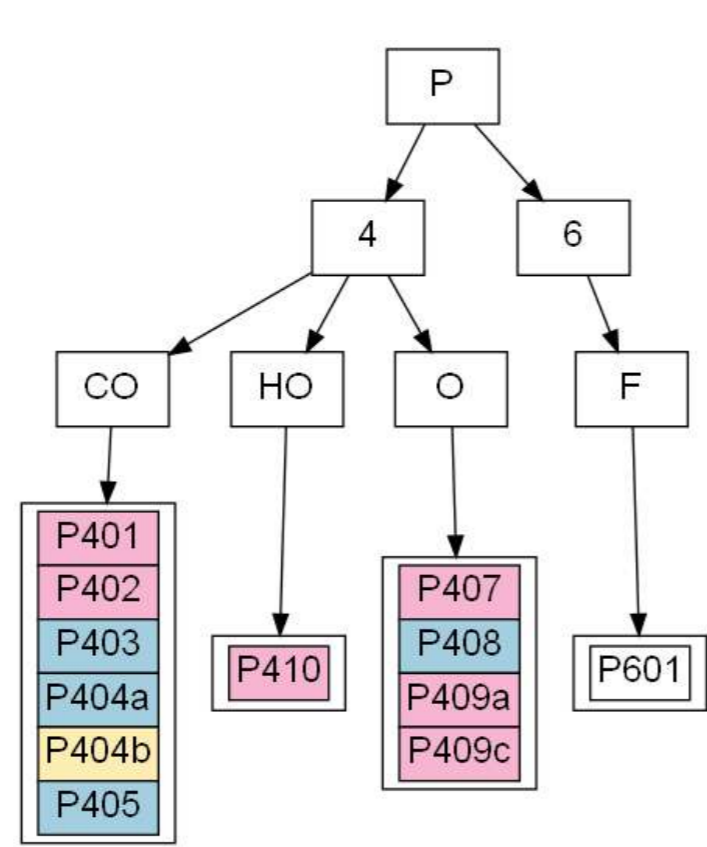
Hierarchiczna struktura drzewa zawiera informacje na temat liczby i typu pierwszych i drugich sąsiadów dla typów atomowych, planarności oraz przynależności do pierścieni. Klasyfikacja gromadzi **typy atomowe z identyczną liczbą i typem pierwszych i drugich sąsiadów**. Typy atomowe, dla których nie zdefiniowano drugich sąsiadów występują wyżej w strukturze drzewa.



Rys.4. Fragment drzewa dla typów atomowych azotu. Pogrubiona czcionka = planarność.

3. POŁĄCZENIE OBU METOD

Wyniki klastrowania opartego na topologii zostały zwizualizowane na nowym, krótszym drzewie z informacją tylko o liczbie i typie pierwszych sąsiadów. Takie podejście jest lepsze do określania relacji i czynników różnicujących pomiędzy typami atomowymi. Następnie **wyniki klastrowania opartego na gęstości elektronowej zostały wprowadzone do drzewa za pomocą kolorowego tła**.
→ Grupy typów atomowych identycznych pod względem topologii i gęstości
→ Unikalne typy atomowe
→ Połączenie między różnymi grupami
→ Możliwość uogólnienia niektórych typów atomowych

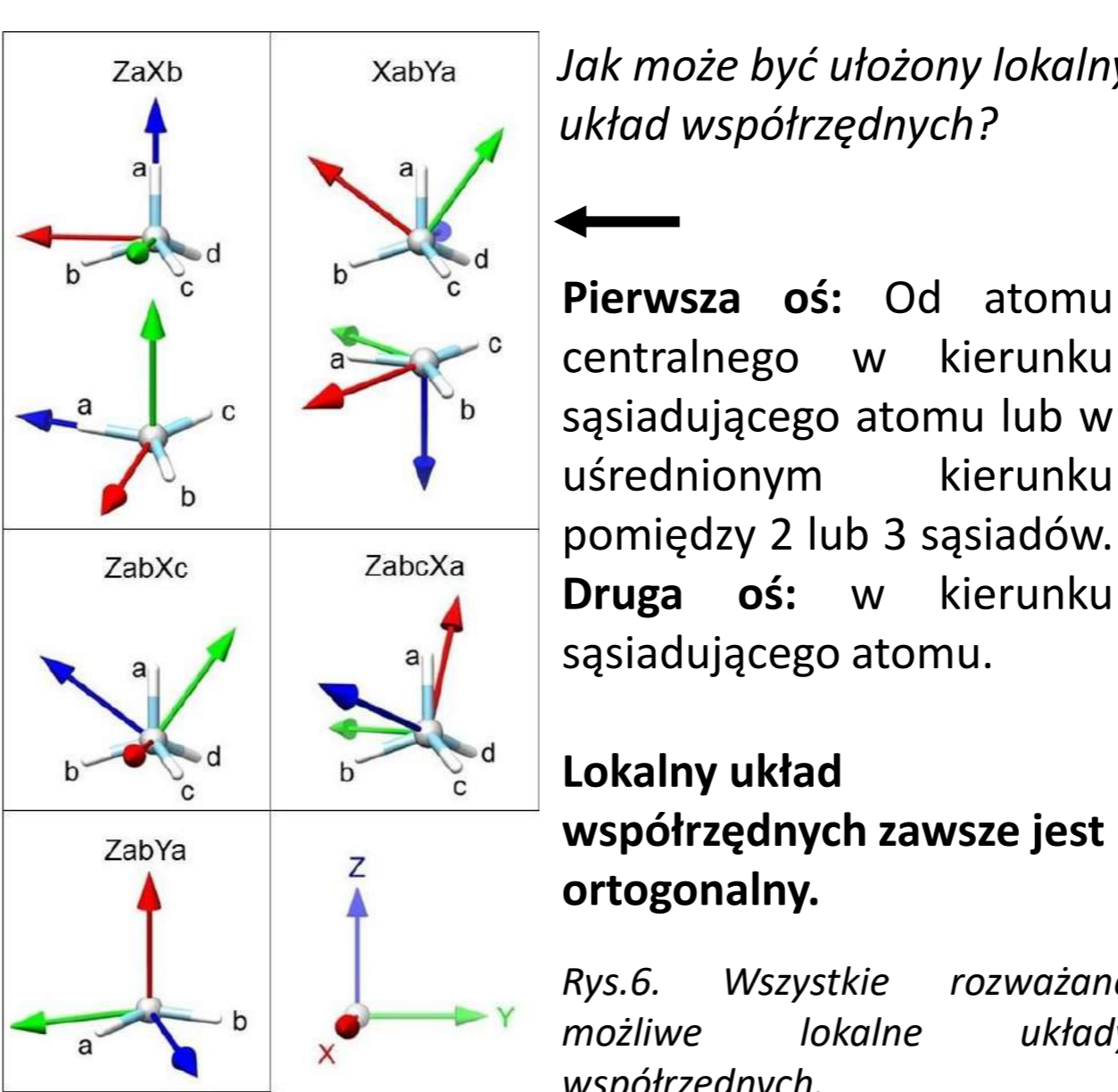


Rys.8. Fragment drzewa dla typów atomowych fosforu z zaznaczonymi grupami uzyskanymi z klastrowania opartego na gęstości elektronowej.

2. KLASTROWANIE OPARTE NA GĘSTOŚCI ELEKTRONOWEJ

- Wykorzystuje **parametry gęstości elektronowej (κ , P_{val} , κ' , $P_{lm\pm}$)**
- Metoda: Density-Based Spatial Clustering for Application with Noise (DBSCAN)
- Od dużych klastrow do mniejszych klastrow zawierających najbardziej podobne do siebie typy atomowe
- Analiza statystyczna wyników w celu wyciągnięcia dalszych wniosków

Typy atomów zdeponowane w banku MATTS są zdefiniowane tylko w jednym z wielu dostępnych lokalnych układów współrzędnych. Ponieważ **parametry $P_{lm\pm}$ zmieniają się w różnych systemach**, konieczne było rozważenie wszystkich dostępnych możliwości.



Jak może być ułożony lokalny układ współrzędnych?

Pierwsza oś: Od atomu centralnego w kierunku sąsiadującego atomu lub w uśrednionym kierunku pomiędzy 2 lub 3 sąsiadów.
Druga oś: w kierunku sąsiadującego atomu.

Lokalny układ współrzędnych zawsze jest ortogonalny.

Rys.6. Wszystkie rozważane możliwe lokalne układy współrzędnych.

Jakie czynniki decydują o ułożeniu lokalnego układu współrzędnych?

- **Hybrydyzacja**
- **Liczba 1. sąsiadów**
- **Planarność**

	4 sąsiadów, brak planarności	3 sąsiadów, brak planarności	2 sąsiadów, współliniowość	1 sąsiad, współliniowość
sp3	43m mm2 m 3m	m 1 ZabXc 3m ZabcXa	mm2 m 1 ZabYa	3m
sp2		3 sąsiadów, planarność δm2 mm2 m XabYa	mm2 m XabYa	mm2 ZabXc'
sp1			2 sąsiadów, niewspółliniowe cyl	1 sąsiad, niewspółliniowe cyl ZabXb'

* dla grupy 1p: ZaXx, dla grup 1x i 2x: ZaXany_orthogonal

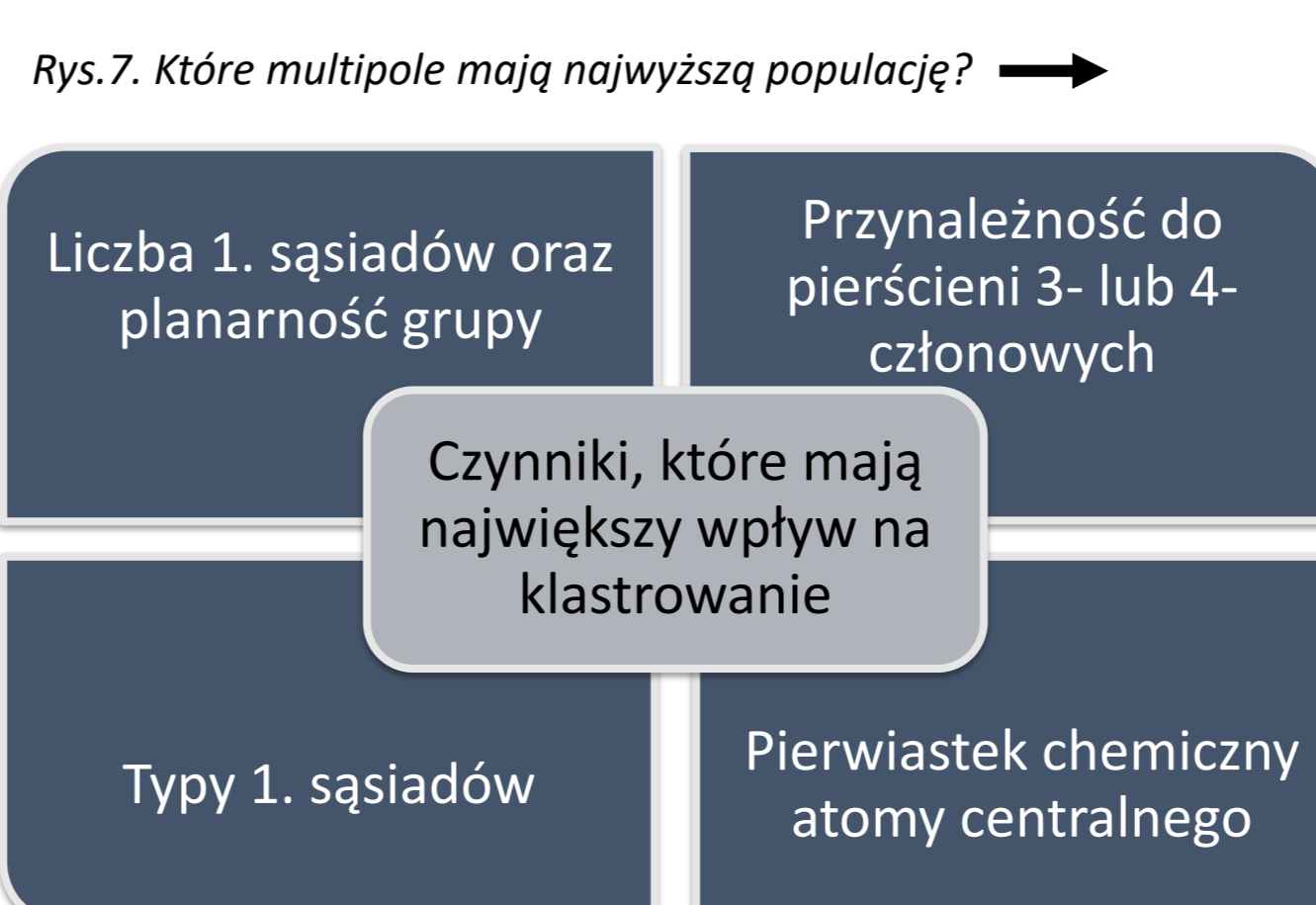
Rys.5. Możliwe symetrie i lokalne układy współrzędnych. Zielony – minimalny zestaw lokalnych układów współrzędnych pozwalający na zobaczenie wszystkich możliwych symetrii. Fioletowy – potrzebne aby klastrować i porównać różne grupy ze sobą.

WNIOSKI

- Analiza klastrow jest obiecującym narzędziem do zrozumienia relacji między typami atomowymi i automatyzacji procesu definiowania typów atomowych dla przyszłego rozwoju banku MATTS
- Tendencje typów atomowych do grupowania się mogą pomóc w zdefiniowaniu ogólnych typów, które zwiększą rozpoznawanie atomów w cząsteczkach
- Określając „dominujące” multipole można znaleźć najwyższą symetrię i hybrydyzację typu atomowego

WIĘCEJ O BANKU MATTS

- Jha, K. K. et al.; Multipolar Atom Types from Theory and Statistical Clustering (MATTS) Data Bank: Restructurization and Extension of UBDB. *J. Chem. Inf. Model.* **2022**. <https://doi.org/10.1021/acs.jcim.2c00144>.
- Rybicka, P. M. et al.; Multipolar Atom Types from Theory and Statistical Clustering (MATTS) Data Bank: Impact of Surrounding Atoms on Electron Density from Cluster Analysis. *J. Chem. Inf. Model.* **2022**. <https://doi.org/10.1021/acs.jcim.2c00145>.



BIBLIOGRAFIA

Hansen, N. K. (1978). *Acta Cryst.* A34, 909–921.
Kumar, P. (2019). *Acta Cryst.* A75, 398-408

PODZIĘKOWANIA

Praca wykonana w ramach grantu NCN UMO-2017/27/B/ST4/02721.