UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

WYDZIAŁ FIZYKI, ASTRONOMII I INFORMATYKI STOSOWANEJ
UNIWERSYTET JAGIELLOŃSKI

MECHANIZM TRANSFORMACJI AMYLOIDOWEJ BAZUJĄCY NA ZMIANIE ROZKŁADU HYDROFOBOWOŚCI

DAWID DUŁAK

ROZPRAWA DOKTORSKA NAPISANA POD KIERUNKIEM

PROF. DR HAB. IRENY ROTERMAN-KONIECZNEJ

ZAKŁAD TECHNOLOGII INFORMATYCZNYCH

KRAKÓW 2020

Oświadczenie

Ja niżej podpisany Dawid Dułak doktorant Wydziału Fizyki, Astronomii i Informatyki Stosowanej Uniwersytetu Jagiellońskiego oświadczam, że przedłożona przeze mnie rozprawa doktorska pt. „*Mechanizm transformacji amyloidowej bazujący na zmianie rozkładu hydrofobowości*" jest oryginalna i przedstawia wyniki badań wykonanych przeze mnie osobiście, pod kierunkiem prof. dr hab. Ireny Roterman-Koniecznej. Pracę napisałem samodzielnie.

Oświadczam, że moja rozprawa doktorska została opracowana zgodnie z Ustawą o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dziennik Ustaw 1994 nr 24 poz. 83 wraz z późniejszymi zmianami).

Jestem świadom, że niezgodność niniejszego oświadczenia z prawdą ujawniona w dowolnym czasie, niezależnie od skutków prawnych wynikających z ww. ustawy, może spowodować unieważnienie stopnia nabytego na podstawie tej rozprawy.

Kraków, dnia 31.07.2020 ……

podpis doktoranta

Dziękuję każdemu kto w jakikolwiek sposób przyczynił się do powstania tej pracy.

Przede wszystkim dziękuję Pani Profesor Irenie Roterman-Koniecznej, za jej ogromną cierpliwość, poszerzanie horyzontów wiedzy oraz liczne rozmowy pozwalające mi dobrze zrozumieć naturę wykonywanych badań. Ta praca nie ujrzałaby światła dziennego bez Jej wsparcia na każdym etapie jej tworzenia. Serdeczność mi okazaną trudno wyrazić słowami.

Dziękuje Pani Małgorzacie Gadzała której rezultaty pracy były podstawą moich badań oraz zespołowi CyfroNet, za wsparcie techniczne oraz udostępnienie zasobów, dzięki którym badania te mogły być przeprowadzone.

Dziękuję wszystkim których poznałem podczas prowadzonych badań i pisania pracy doktorskiej: dzięki Wam poznałem świat wcześniej dla mnie zakryty, a Wasz „ogień" rozpalił we mnie chęć dalszego poszukiwania odpowiedzi na pytania zadziwiającego świata ludzi i zwierząt.

Dziękuję mojej Żonie za wiarę we mnie do samego końca oraz przestrzeń na badania i przygotowanie tej pracy w niełatwym okresie wielu osobistych i rodzinnych zmagań.

Wreszcie dziękuję moim Rodzicom - bez nich nie byłbym tym, kim jestem dzisiaj.

Dawid Dułak

Spis treści

1. Wprowadzenie

Proces fałdowania białek jest szeroko opisywany w literaturze, ale mechanizm jego działania nie jest dobrze rozpoznany, pomimo wielu lat badań i prób zrozumienia tego zjawiska biologicznego zarówno eksperymentalnie jak i poprzez symulacje komputerowe [1]. Jest to istotnie fascynujący fenomen, iż pomimo tylu możliwości, dany łańcuch polipeptydowy zawsze fałduje się w jeden i ten sam sposób. Istnieje jednak grupa struktur, które fałdują się niepoprawnie, tzn. z jakiegoś powodu zaczynają fałdować się inaczej niż zazwyczaj dzieje się to dla danego łańcucha polipeptydowego [2].

Celem tej pracy było zbadanie fałdowania się właśnie takich „niepoprawnych" struktur, aby – w porównaniu do białek formujących się poprawnie – wyciągnąć wnioski na temat samego procesu fałdowania i tym samym zbliżyć się do rozwiązania tego jakże ciekawego problemu.

W tym celu zastosowano model uwzględniający obecność środowiska wodnego w postaci zewnętrznego pola siłowego wyrażonego za pomocą funkcji 3D Gaussa [3]. Obecność pola o tej postaci sugeruje ukierunkowanie procesu fałdowani do skupienia reszt hydrofobowych w centrum z równoczesną ekspozycją reszt polarnych na powierzchni. Takich globularnych form nie osiągają łańcuchy obecne w amyloidach pojawiających się w warunkach fizjologicznych.

Celem wykonanych symulacji była identyfikacja specyfiki łańcuchów polipeptydowych budujących formy fibrylarne. Do tego celu wykorzystano specjalistyczne oprogramowanie ocenianie w ogólnoświatowym projekcie CASP (*Critical Assessment of Protein Structure Prediction*; https://www.predictioncenter.org/) [4], jako wiarygodne w dziedzinie przewidywania struktury białek - łańcucha

polipeptydowego o zadanej sekwencji aminokwasowej. Dodatkowo, symulacje przeprowadzono z wykorzystaniem oprogramowania bazującego na ukierunkowanym mechanizmie fałdowania, poprzez wprowadzenie zewnętrznego pola siłowego w postaci funkcji 3D Gaussa. Wg. modelu środowisko takie pochodzi od otoczenia wodnego. W miejsce reprezentacji tego otoczenia za pomocą określonej liczby indywidualnych molekuł wody, model wprowadza pole o charakterze ciągłym.

W świetle przedstawionych badań wykorzystując analizę porównawczą wszystkich wymienionych narzędzi wykazano, że łańcuchy polipeptydowe budujące formy amyloidowe, wykazują specyfikę wykluczającą generowanie globularnych form rozpuszczalnych w wodzie, na korzyść form nisko upakowanych, o potencjalnych możliwościach wchodzenia w interakcję typu wielocząsteczkowego - wielołańcuchowego fibryla.

2. Materiały i Metody

2.1. MATERIAŁY

W pracy korzystano z danych pochodzących z PDB (Protein Data Bank; (https://www.rcsb.org/) [5]. Struktury form amyloidowych dostępnych w momencie opracowywania niniejszej pracy podano w Tab. 1.

PDB ID	Nazwa	Jednostka chorobowa	Protofibryl	Superfibryl	Ref.
5O3L	Tau	Alzheimer	2	Tak	[6]
2MPZ	Aβ(15-40)	Mutacja Iowa	3	Tak	[7]
2MXU	Aβ(11-42)	Alzheimer	1	Nie	[8]
2MVX	Aβ(1-40)	Mutacja Osaka	2	Tak	[9]
2N0A	α-Synukleina	Ciała Lewy'ego	1	Nie	[10]

TAB. 1 ZESTAWIENIE BIAŁEK AMYLOIDOWYCH STANOWIĄCYCH OBIEKT ANALIZY. PODANO TEŻ ICH KRÓTKĄ CHARAKTERYSTYKĘ.

Praca została wykonana z wykorzystaniem Infrastruktury PLGrid (Cyfronet AGH).

2.2. METODY

Symulację fałdowania łańcuchów polipeptydowych budujących formy amyloidowe przeprowadzono w oparciu o dostępne powszechnie specjalistyczne oprogramowanie w postaci pakietu I-TASSER [11] i Robetta [12]. Są to narzędzia

specjalistycznie przygotowane do symulacji fałdowania łańcuchów polipeptydowych o zadanej sekwencji aminokwasowej. Programy te, co wspomniano już wcześniej, są oceniane jako najbardziej wiarygodne w projekcie CASP - *Critical Assessment of Protein Structure Prediction* [4].

2.2.1.1. ODDZIAŁYWANIA NIEWIĄŻĄCE

Powyższe pakiety wykorzystują charakterystykę zjawiska fałdowania białka jako procesu spontanicznego i opierają się na poszukiwaniu optymalnej struktury łańcucha ufałdowanego na drodze optymalizacji (minimalizacji) energii. Minimalizacji poddane są wszystkie uwzględniane w modelach oddziaływania niewiążące. Są nimi:

A. **oddziaływania elektrostatyczne** (Coulombowskie) – są to siły ładunek-ładunek występujące w odniesieniu do każdej pary atomów. Poza oddziaływaniami w ramach łańcucha polipeptydowego, szczególnie ważne są w tym wypadku siły między-molekularne w oddziaływaniach między białkiem i ligandem lub innym białkiem podczas wystąpienia kompleksu wielobiałkowego. Ich siła jest zależna odwrotnie proporcjonalnie od odległości pomiędzy danymi parami atomów. Energia oddziaływania mieści się pomiędzy -20 a -40 kJ/mol i jest wyznaczana wzorem:

$$E = \frac{q_1 q_2}{4\pi\varepsilon_0 r_{1,2}}$$

q – ŁADUNEK; r – ODLEGŁOŚĆ; ε_0 – STAŁA DIELEKTYCZNA

B. **oddziaływania typu dipol – dipol** (jeden z typów oddziaływań van der Waalsa) – jest to wiązanie, które tworzy się pomiędzy układami atomów o niezerowym momencie dipolowym. Na dwóch biegunach posiadają one przeciwstawne ładunki elektryczne, oddziałujące podobnie jak jony. Jako że rozważamy tu ładunki cząstkowe, jest to wiązanie znacznie słabsze od oddziaływań elektrostatycznych. Chociaż siły tego typu wiązania są bardzo słabe porównując

do pozostałych kształtujących konformację białka, to ich bardzo duża liczba sprawia, że stanowią istotny element fałdowania białek (nawet w polipeptydach o dużej liczbie aminokwasów). Układy dipolowe są obecne w ramach wiązania peptydowego oraz w niektórych łańcuchach bocznych aminokwasów. Jego siła jest zależna odwrotnie proporcjonalnie od sześcianu odległości. Energia oddziaływania mieści się pomiędzy -5 a -20 kJ/mol i jest wyznaczana wzorem:

$$E = \frac{\mu_1 \mu_2}{4\pi\varepsilon_0 r_{1,2}^3}$$

μ - MOMENT DIPOLOWY; r – ODLEGŁOŚĆ; ε_0 – STAŁA DIELEKTYCZNA

C. **wiązania wodorowe** – są to słabe oddziaływanie pomiędzy atomem wodoru połączonym wiązaniem kowalencyjnym z atomem elektroujemnym (np. atomem azotu lub atomem tlenu), a drugim atomem o ładunku ujemnym np. atomem azotu lub atomem tlenu innej molekuły. Energia tego wiązania w dużej mierze zależy od odległości między donorem a akceptorem (która jest krótsza niż 3,5Å) i od ich wzajemnej orientacji. Mieści się ona (tj. energia wiązania) pomiędzy -4 a -30 kJ/mol. Obecność wiązań wodorowych jest bardzo znaczna ze względu na możliwość tego typu oddziaływania w ramach wiązań peptydowych, co jest np. decydujące w konstrukcji formy heliakalnej czy beta-strukturowej.

D. **potencjał torsyjny związany ze swobodnymi rotacjami** – jest to energia potencjalna związana ze zmianą kątów torsyjnych (dwuściennych) dla każdego aminokwasu w łańcuchu polipeptydowym. Kąt torsyjny znajduje się w miejscu wiązania C-C_α (kąt ψ) oraz C_α-N (kąt ϕ). Potencjał torsyjny uwzględniany jest też dla innych swobodnych rotacji, występujących w łańcuchach bocznych i oznaczanych λ.

Parametryzację obecnych w danym pakiecie oddziaływań dostarcza Autor danego pakietu, co czyni zastosowane pola siłowe zrożnicowane, dając możliwość analizy porównawczej przyjętej parametryzacji.

2.2.1.2. ODDZIAŁYWANIA HYDROFOBOWE

Naturalnym środowiskiem dla białek i ich aktywności biologicznej jest środowisko wodne. Uwzględnienie obecności i wpływu tego środowiska na proces fałdowania się białek wydaje się mieć krytyczne znaczenie, jako że efekt hydrofobowy jest jedną z kluczowych sił uczestniczących w fałdowaniu się białek, powodując umiejscowienie elementów hydrofobowych wewnątrz polipeptydu i ekspozycję elementów hydrofilnych na jego powierzchni [13].

Pakiety dostępne powszechnie (wspomniany pakiet I-TASSER i Robetta) uwzględniają obecność wody w postaci określonej liczby indywidualnych molekuł wody (wynikającej z gęstości wody). Interakcja cząsteczek wody z aminokwasami obecnymi w łańcuchu - jak się zakłada - skutkuje ekspozycją reszt hydrofilnych na powierzchni białka. Jest to jednak procedura bardzo wymagająca z punktu widzenia czasu obliczeń jak i potrzebnych do wykorzystania zasobów pamięci.

W modelu zastosowanym w obecnej pracy wykorzystano tzw. *fuzzy oil drop model* wywodzący się z analizy struktur miceli, szczegółowo opisany w [14].

Warto przy tej okazji zauważyć, że każdy z aminokwasów w łańcuchu polipeptydowym ma przypisaną sobie hydrofobowość własną. Jest dostępnych wiele skali hydrofobowości opartych na pomiarach eksperymentalnych oraz wyznaczonych teoretycznie. Do celów naszych badań stosujemy własną skalę hydrofobowości [15], zbliżoną do istniejących - np. Eisenberga [16] czy Doolittle [17] - na tyle, że jest to nieistotna różnica w kontekście opisywanych działań.

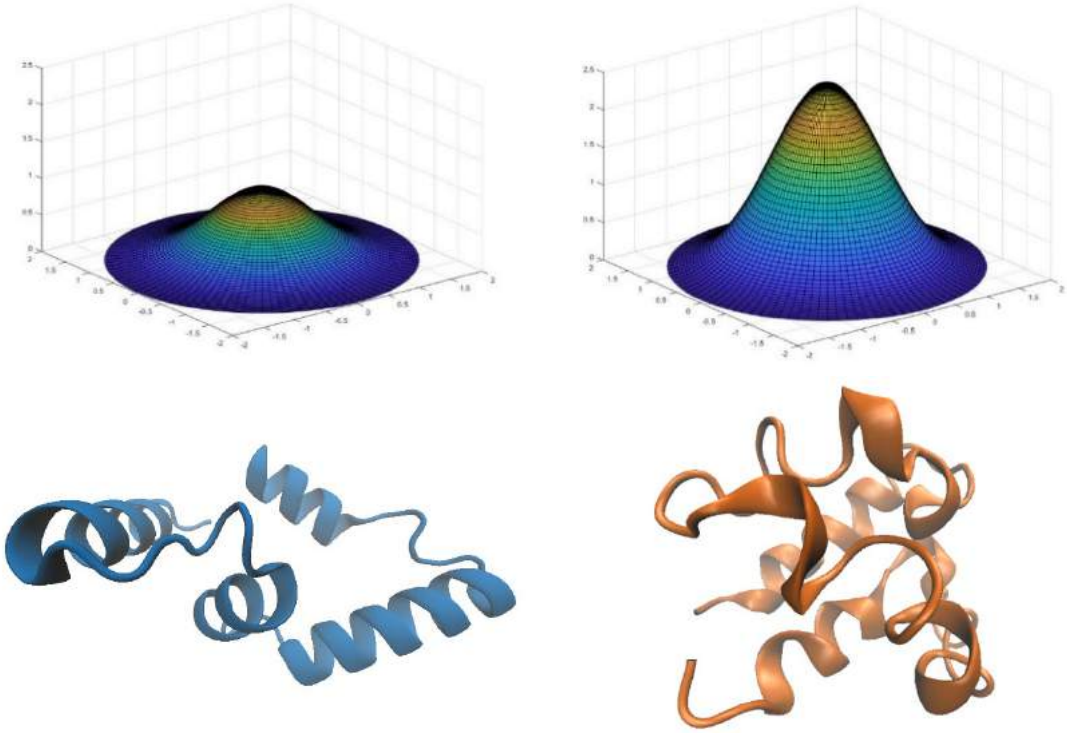


FIG. 1 HYDROFOBOWOŚĆ BIAŁKA PRZEDSTAWIONA W FUNKCJI 3D GAUSSA. DLA ŁATWIEJSZEJ WIZUALIZACJI ZREDUKOWANO DO PREZENTACJI 2D GAUSSA.

W środowisku wodnym cząsteczki bipolarne w sposób spontaniczny generują strukturę micelarną. Jest to efekt ekspozycji reszt polarnych z równoczesną izolacją części hydrofobowych molekuły w centralnej części miceli. Jeśli w środowisku wodnym znajduje się więcej niż jedna cząsteczka bipolarna, to generowana jest tzw. ko-micela inkorporująca w swoim składzie cząsteczki bipolarne o zróżnicowanej proporcji części polarnej do niepolarnej.

Rozkład hydrofobowości w takim układzie można opisać za pomocą funkcji 3D Gaussa, gdzie jej maksimum jest zlokalizowane w centrum miceli - a w naszym przypadku także białka - a wartości bliskie zeru na powierzchni co obrazuje Fig. 1 [18]. W przypadku miceli utworzonej przez cząsteczki bipolarne, parametry σ (sigma) dla wszystkich trzech

osi układu współrzędnych przyjmują taką samą wartość zależną od rozmiaru cząsteczki monomerycznej.

Analizując aminokwasy można stwierdzić, że 20 cząsteczek reprezentuje układ 20 form bipolarnych, o zróżnicowanych proporcjach części polarnej do niepolarnej. Układ taki w sposób naturalny dąży do utworzenia formy miceli jako efekt wpływu środowiska wodnego. Aminokwasy w łańcuchu polipeptydowym mają jednak znaczące ograniczenie w postaci narzuconego sąsiedztwa poprzez zdeterminowany układ wiązań peptydowych. W przeciwieństwie do klasycznej miceli w przypadku łańcucha polipeptydowego swobodne przemieszczanie się nie jest możliwe. Dlatego odtworzenie miceli przez łańcuch polipeptydowy jest ograniczone lub wręcz w niektórych przypadkach niemożliwe. Tendencja to utworzenia układu micelarnego w możliwie osiągalnym stopniu jest jednak obecna.

Dlatego w modelu o nazwie *fuzzy oil drop* (model rozmytej kropli oliwy) zastosowano funkcję 3D Gaussa jako wyraz obecności środowiska wodnego ukierunkowującego proces fałdowania łańcucha w stronę generowania jądra hydrofobowego z równoczesną tendencją ekspozycji reszt polarnych na powierzchni. Jak powiedziano wcześniej, stanu idealnej zgodności w formie pełnego pokrycia powierzchni grupami polarnymi oraz całkowitego ukrycia reszt hydrofobowych nie da się uzyskać. Stopień niezgodności jest jednak zróżnicowany. Zidentyfikowano na przykład białka o idealnie zgodnym rozkładzie hydrofobowości z funkcja 3D Gaussa. Są to białka *antifreeze* typu III [19]. Okazało się też, że zdecydowana większość domen (ponad 90 %) wykazuje uporządkowanie hydrofobowości zgodne z rozkładem 3D Gaussa [20]. Zidentyfikowano także takie białka, które wykazują lokalną niezgodność w postaci ekspozycji reszt hydrofobowych na powierzchni. Wykazano, że ekspozycja ta wykorzystywana jest do tworzenia kompleksów białko-białko [21]. Lokalna niezgodność w formie lokalnego deficytu hydrofobowości związana jest z obecnością zagłębienia

(*cavity*), gdzie bardzo często znajduje się grupa czynna enzymu lub kompleksowany jest ligand [22].

2.2.1.3. OPIS PROCEDURY - FUZZY OIL DROP MODEL

Model, którego szczegółowy opis znajdziemy tu [23] wygląda następująco. Molekuła białka zlokalizowana jest w układzie współrzędnych tak, aby środek geometryczny białka pokrywał się w punkcie (0, 0, 0). Następnie molekuła jest orientowana tak, aby najdłuższa oś (największa odległość między atomami) pokrywała się z jedną z osi (np. osią X-ów). Dalej wybieramy dwa atomy, których lokalizacja jest najbardziej dystalna wg. odległości między punktami, stanowiącymi rzuty położeń atomów na płaszczyźnie YZ. Molekułę obraca się tak, aby odcinek tak wyznaczony pokrywał się z kolejną osią (powiedzmy Y-ów).

Dla tak zorientowanej molekuły wyznacza się jej rozmiar, poprzez wyliczenie maksymalnej rozpiętości wartości współrzędnych X, co pozwala wyznaczyć wartość parametru σ_X funkcji Gaussa jako 1/6 odległości między najbardziej dystalnymi punktami (zgodnie z regułą trzech sigm). Podobnie wyznacza się wartości parametrów σ_Y i σ_Z.

W następnym etapie określa się pozycje tzw. efektywnych atomów reprezentujących kolejne aminokwasy, czyli uśrednioną pozycję atomów wchodzących w skład aminokwasu. Dla tych punktów wyznaczana jest wartość funkcji 3D Gaussa wg. wzoru poniżej:

$$\widetilde{H}t_j = \frac{1}{\widetilde{H}t_{sum}} \exp\left(\frac{-\left(x_j - \bar{x}\right)^2}{2\sigma_x^2}\right) \exp\left(\frac{-\left(y_j - \bar{y}\right)^2}{2\sigma_y^2}\right) \exp\left(\frac{-\left(z_j - \bar{z}\right)^2}{2\sigma_z^2}\right)$$

$\bar{x}, \bar{y}, \bar{z}$ - POZYCJE CENTRUM MOLEKUŁY; σ_X, σ_Y I σ_Z - 1/6 NAJWIĘKSZYCH ODLEGŁOŚCI MIĘDZY ATOMAMI WZDŁUŻ ODPOWIEDNIEJ OSI

Przy orientacji podanej powyżej wartości pozycji centrum molekuły są równe 0. Wyrażenie przez funkcję eksponencjalną jest wyrażeniem normalizującym, gdzie

odwrotność sumy wszystkich wartości uzyskanych dla kolejnych pozycji aminokwasów powoduje, że suma wszystkich wartości znormalizowanych jest równa 1.

Rzeczywisty poziom hydrofobowości obserwowanej jest wynikiem oddziaływań między-aminokwasowych i jest zależny od hydrofobowości własnej oddziaływujących aminokwasów oraz odległości między nimi. Wyznaczany jest wg. wzoru poniżej:

$$Ho_j = \frac{1}{Ho_{sum}} \sum_i \begin{cases} \left(H_i^r + H_j^r\right)\left(1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right), & r_{ij} \le c, \\ 0, & r_{ij} > c \end{cases}$$

H_i^r - HYDROFOBOWOŚĆ WŁASNA; r - ODLEGŁOŚĆ MIĘDZY ATOMAMI EFEKTYWNYMI; c - POZIOM ODCIĘCIA

Poziom odcięcia – wg. propozycji Levitt'a – ustalamy na 9Å [24]. Wyrażenie na pierwszej pozycji jest czynnikiem normalizującym.

Profile uzyskane przez wartości T_i i O_i wizualizują podobieństwo i różnice pomiędzy poziomem oczekiwanym (T_i) oraz obserwowanym (O_i). Ilościowa ocena stopnia podobieństwa dokonywana jest za pomocą entropii dywergencji zaproponowanej przez Kullback'a-Leibler'a [25].

$$D_{KL}(p|p^0) = \sum_{i=1}^{N} p_i \log_2(p_i/p_i^0)$$

$\boldsymbol{p_i^0}$ – WARTOŚĆ W ROZKŁADZIE REFERENCYJNYM; $\boldsymbol{p_i}$ – POZIOM W ROZKŁADZIE OBSERWOWANYM

W przypadku omawianego modelu, rolę rozkładu referencyjnego pełni rozkład wyidealizowany, wyrażony funkcją Gaussa, co prowadzi do następującej postaci:

$$O|T = \sum_{i=1}^{N} O_i \log_2(O_i/T_i)$$

T_i – WARTOŚĆ W ROZKŁADZIE TEORETYCZNYM; O_i – WARTOŚĆ W ROZKŁADZIE OBSERWOWANYM

Uzyskana wartośc wyznaczona za pomocą wskazanego równania jest wartością entropii, a więc nie może być interpretowana bezpośrednio. Dlatego wprowadzamy drugi rozkład referencyjny jakim jest rozkład jednostajny, gdzie każdy aminokwas wyposażony jest w parametr wyrażający poziom hydrofobowości jendnakowy dla każdego aminokwasu. Poziom R_i przypisany każdenu aminokwasowi jest równy 1/N gdzie N oznacza liczbę aminokwsów w łańcuchu.

Aby określic status rozkładu omawianego białka należy porównać wartość D_{KL} dla relacji O-T z relacją O-R. Relacja O-T < O-R wyraża obecność centrycznego jądra hydrofobowego.

$$O|R = \sum_{i=1}^{N} O_i \log_2(O_i/R_i)$$

R_i – WARTOŚĆ W ROZKŁADZIE JEDNOSTAJNYM; O_i – WARTOŚĆ W ROZKŁADZIE OBSERWOWANYM

Aby nie operowć dwoma wartościami wprowadzono parametr RD zdefiniowany jak podano ponizej:

$$RD = \frac{O|T}{O|T + O|R}$$

Wartość RD (Relative Distance) poniżej 0.5 oznacza obecnośc jądra hydrofobowego w danej strukturze. Graficznie omawianą ocenę wyraża Fig. 2 poniżej [26].

Na pokazanym rysunku Fig. 2 wartość RD jest mniejsza niż 0.5 co oznacza, że rozkład środkowy (obserwowany) wykazuje stopień podobieństwa wobec rozkładu wyidealizowanego (strona lewa), wystarczający dla przyjęcia rozkładu centrycznego. Rozkład obserwowany nie ma w tej analizie statusu zbliżonego do rozkładu jednostajnego (rozkład prawy).

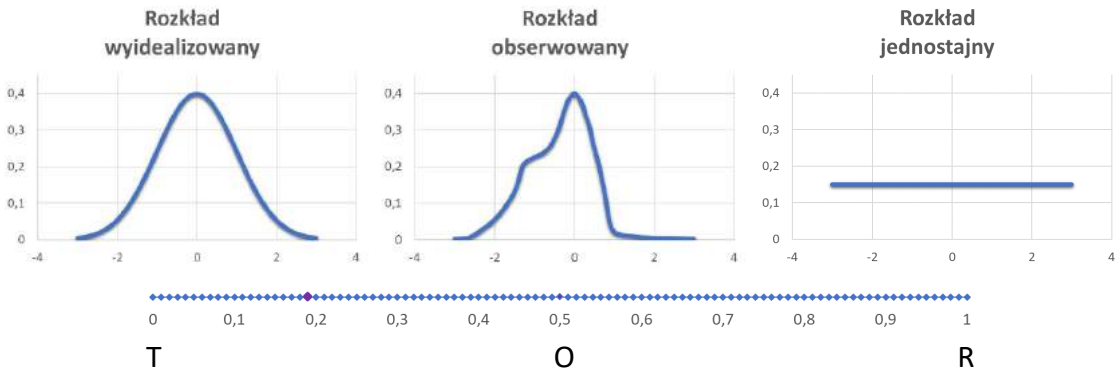


FIG. 2 WIZUALIZACJA MODELU DLA 1D GAUSSA. WARTOŚĆ RD JEST RÓWNA 0.194.

W symulacji procesu fałdowania wg. *fuzzy oil drop model* zachodzi on nie tylko wg. procedury minimalizacji energii wewnętrznej, ale także minimalizacji różnic między rozkładem T i O, a więc minimalizacji wartości RD.

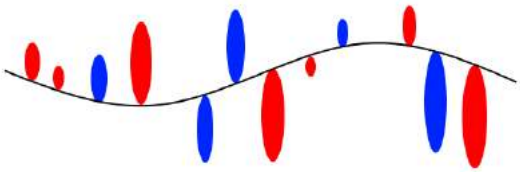


FIG. 3 HYDROFOBOWOŚĆ WŁASNA POSZCZEGÓLNYCH AMINOKWASÓW W ROZFAŁDOWANYM ŁAŃCUCHU. NA NIEBIESKO ZAZNACZONO AMINOKWASY HYDROFILNE, NA CZERWONO AMINOKWASY HYDROFOBOWE.

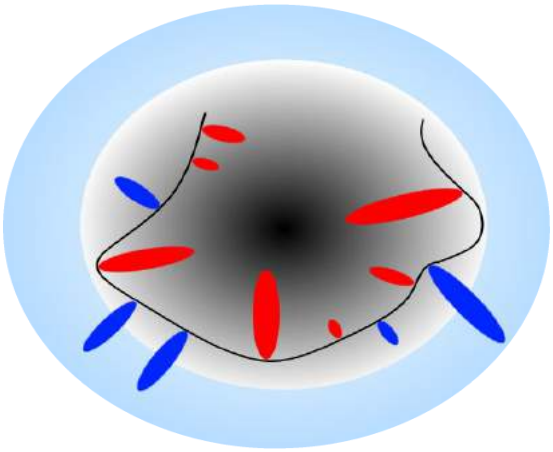


FIG. 4 HYDROFOBOWOŚĆ WŁASNA POSZCZEGÓLNYCH AMINOKWASÓW I ICH UŁOŻENIE WZGLĘDEM CENTRUM BIAŁKA W SFAŁDOWANYM ŁAŃCUCHU. NA NIEBIESKO ZAZNACZONO AMINOKWASY HYDROFILNE, NA CZERWONO AMINOKWASY HYDROFOBOWE.

Proces fałdowania przebiega ze zmniejszaniem rozmiaru elipsoidy 3D Gaussa, co obrazuje Fig. 5 poniżej, a symulację tej jego fazy w *fuzzy oil drop model* nazywamy *Late Stage* [27].

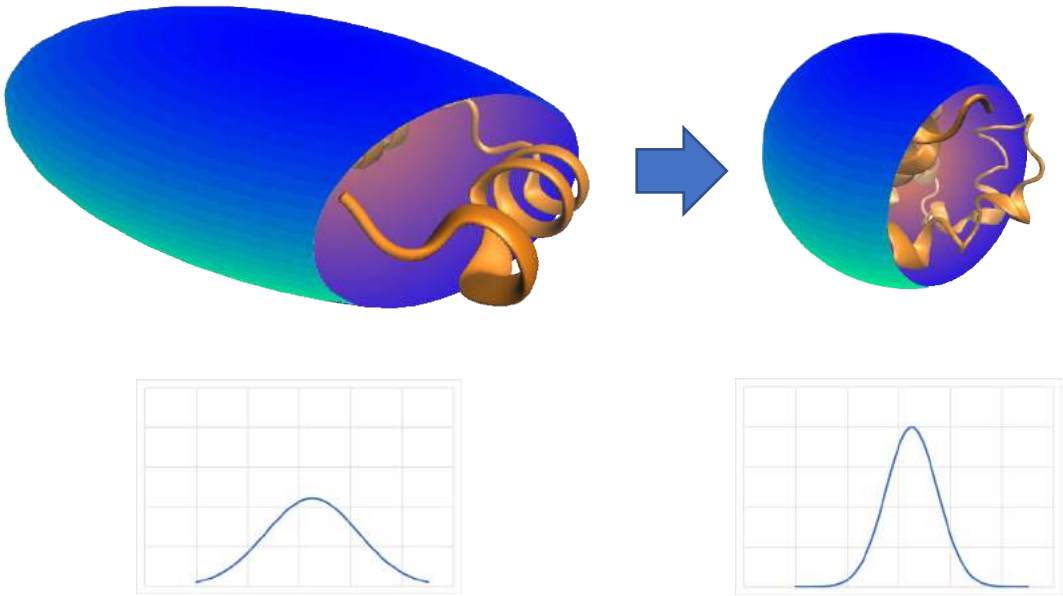


FIG. 5 FUNKCJA GAUSSA W POSTACI WYKRESU 1D DLA ROZFAŁDOWANEGO (LEWA STRONA) I SFAŁDOWANEGO (PRAWA STRONA) BIAŁKA.

Na Fig. 5 powyżej widoczne jest zwiększanie stopnia upakowania prowadząc do podwyższenia poziomu hydrofobowości w centrum.

2.2.1.4. PRZYKŁADY DLA BIAŁEK NIEAMYLOIDOWYCH

Przykładem białka o rozkładzie hydrofobowości zgodnym z rozkładem oczekiwanym jest białko zapobiegające zamrażaniu wody (1KDE). Status tego białka opisany parametrem RD (RD = 0.404) wykazuje zgodność rozkładu O z rozkładem T co sugeruje rozkład hydrofobowości odtwarzający układ micelarny (Fig. 6).

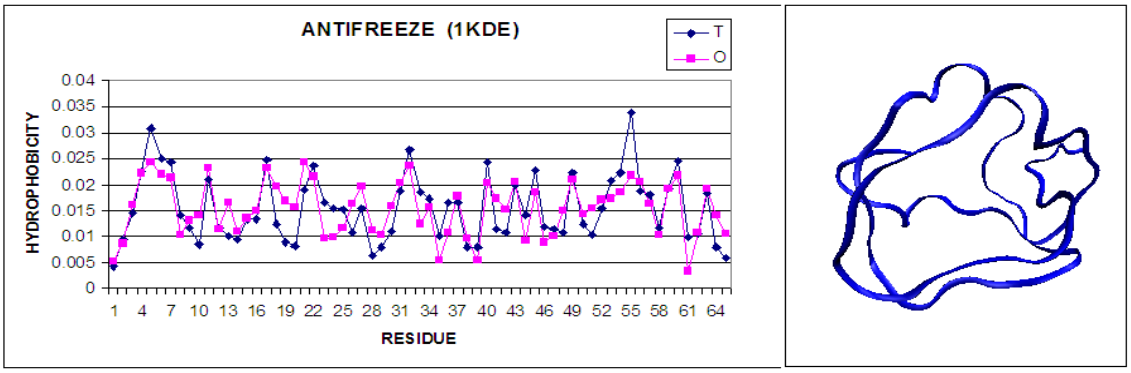


FIG. 6 ROZKŁAD T I O DLA BIAŁKA ANTIFREEZE (1KDE) WYKAZUJĄCY WYSOKA ZGODNOŚĆ ROZKŁADU T I O. PO PRAWEJ STRONIE WIDOCZNA REGULARNA SFERYCZNA, GLOBULARNA BUDOWA TEGO BIAŁKA.

DANE DO WYKRESU UZYSKANO ZA ZGODĄ OD DR MATEUSZA BANACHA - ZAKŁAD BIOINFORMATYKI I TELEMEDYCYNY COLLEGIUM MEDICUM UNIWERSYTET JAGIELLOŃSKI

Nieregularność w postaci ekspozycji reszt hydrofobowych na powierzchni uwidacznia się w białku - kapsyd HIV (1Y7Q). Struktura dostępna w PDB to homo-dimer. Poniżej przedstawiono rozkłady T_i i O_i dla tego białka wskazując reszty zaangażowane w kompleksowanie drugiej molekuły białka (Fig. 7).

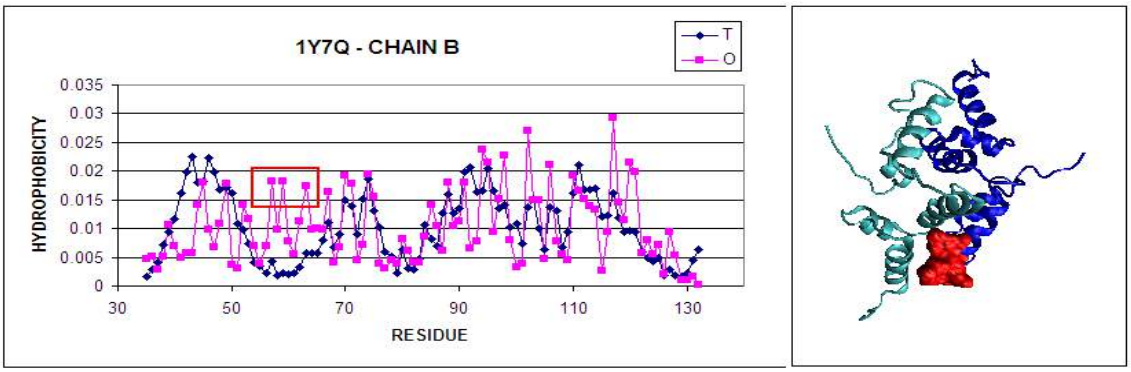


FIG. 7 PROFILE T I O DLA POJEDYNCZEGO ŁAŃCUCHA W DIMERZE KAPSYDU HIV. ZAZNACZONY POZYCJE - RESZTY WYKAZUJĄCE LOKALNY NADMIAR HYDROFOBOWOŚCI W MIEJSCU GDZIE OCZEKIWANA JEST WYSOKA POLARNOŚĆ. RYSUNEK PO PRAWEJ POKAZUJE PREZENTACJĘ 3D Z WYRÓŻNIONYMI FRAGMENTAMI ŁAŃCUCHA. DIMER: MONOMER TURKUSOWY I GRANATOWY.

DANE DO WYKRESU UZYSKANO ZA ZGODĄ OD DR MATEUSZA BANACHA - ZAKŁAD BIOINFORMATYKI I TELEMEDYCYNY COLLEGIUM MEDICUM UNIWERSYTET JAGIELLOŃSKI

Status pojedynczego łańcucha wyraża się wartością RD = 0.514. Wartość powyżej 0.5 sugeruje obecność nieuporządkowania zgodnego z *fuzzy oil drop model*. Jak jest to widoczne na profilach, ta wysoka wartość jest efektem zbyt wysokiej hydrofobowości w obszarze, gdzie oczekiwana jest polarność. Jak się okazuje, reszty te są zaangażowane w kompleksowanie drugiego łańcucha. Wyeliminowanie tych reszt z obliczeń wartości RD powoduje zmniejszenie do 0.432. Oznacza to, że czynnikiem powodującym wysoką wartość RD są właśnie wymienione reszty. Wysoka ekspozycja na powierzchni wskazuje potencjalne miejsce kompleksowania drugiego łańcucha.

Przykładem białka wykazującym lokalny deficyt hydrofobowości jest lizozym, gdzie na wykresie pokazano status reszt budujących zagłębienie (*cavity*) kompleksujące ligand.

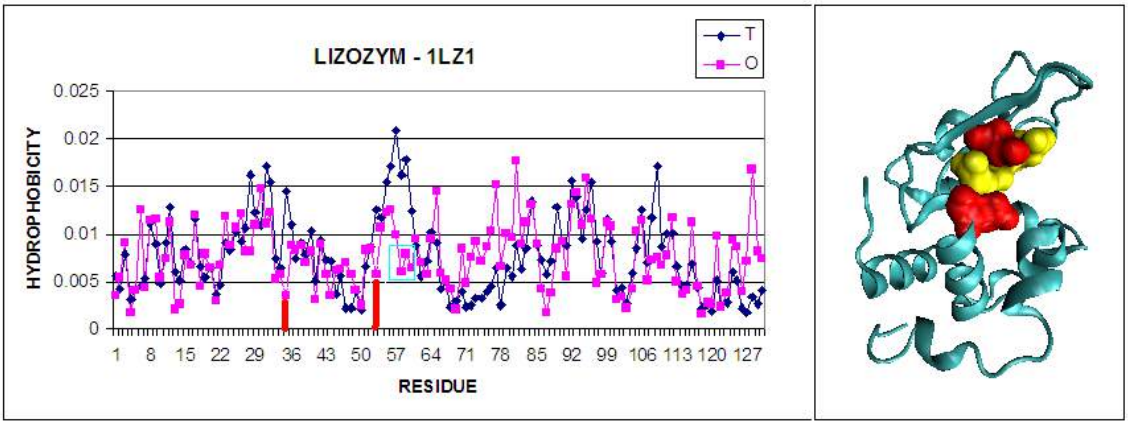


FIG. 8 PROFILE T I O DLA ŁAŃCUCHA LIZOZYMU (1LZ1). ZAZNACZONO RESZTY KATALITYCZNE - CZERWONE ORAZ INNE, KTÓRYCH USUNIĘCIE SKUTKUJE UZYSKANIEM WARTOŚCI RD PONIŻEJ 0.5.

DANE DO WYKRESU UZYSKANO ZA ZGODĄ OD DR MATEUSZA BANACHA - ZAKŁAD BIOINFORMATYKI I TELEMEDYCYNY COLLEGIUM MEDICUM UNIWERSYTET JAGIELLOŃSKI

Przykładem identyfikacji zagłębienia (*cavity*) wiążącego ligand - w tym konkretnym przypadku – substrat, jest lizozym. Status kompletnego łańcucha wyraża się wartością RD = 0.529. Wyeliminowanie dwóch reszt katalitycznych: E35 i D53 powoduje

obniżenie wartości RD do 0.515, natomiast eliminacja reszt w bezpośrednim sąsiedztwie skutkuje osiągnięciem wartości RD poniżej 0.5 (0.492). Zaznaczyć należy, że są to reszty w bezpośrednim sąsiedztwie reszt katalitycznych i wchodzące w skład zagłębienia. Są to też reszty wykazujące wysoki lokalny deficyt hydrofobowości. Pokazuje to Fig. 8 [28].

2.2.1.5. PRZYKŁADY DLA BIAŁEK AMYLOIDOWYCH

Przykładem białek wykazujących globalną niezgodność rozkładów T_i i O_i są fibryle amyloidowe, czyli wielołańcuchowe kompleksy określane jako białka patologiczne. Przykładowe rozkłady T_i i O_i dla amyloidu pokazano na Fig. 9 poniżej:

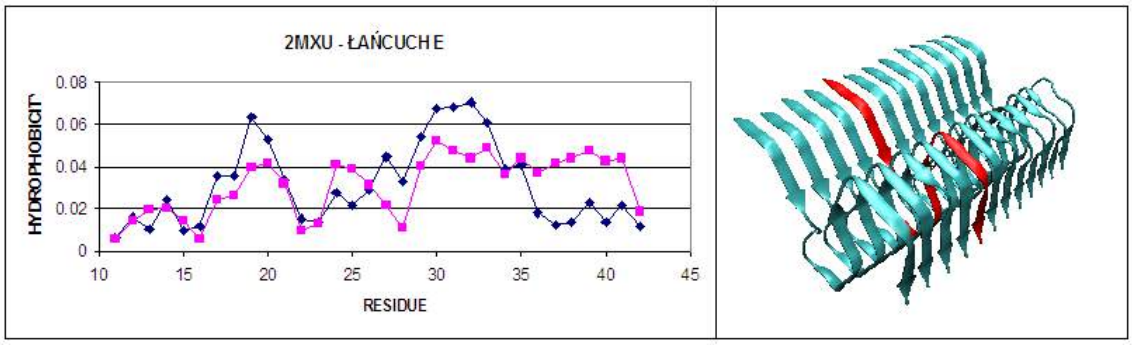


FIG. 9 PRZYKŁADOWY ROZKŁAD T_i I O_i AMYLOIDU Aβ(11-42) (2MXU). WIDOCZNA JEST NIEZGODNOŚĆ PRAWIE WZDŁUŻ CAŁEGO ŁAŃCUCHA.

DANE DO WYKRESU UZYSKANO ZA ZGODĄ OD DR MATEUSZA BANACHA - ZAKŁAD BIOINFORMATYKI I TELEMEDYCYNY COLLEGIUM MEDICUM UNIWERSYTET JAGIELLOŃSKI

Kolejny przykład dotyczy opisu formy amyloidowej, gdzie niezgodność rozkładu hydrofobowości dotyczy całego łańcucha wykazując okresowe zmiany (forma prawie sinusoidalna) wynikające z identycznego składu aminokwasowego indywidualnych łańcuchów. Przykładem jest tutaj amyloid Aβ(11-42) (2MXU). Status fibryla złożonego z 12 łańcuchów wyraża się wartością RD = 0.680. Profile T i O dla tego amyloidu wykazują niezgodność polegającą na braku koncentracji hydrofobowości w centralnej części

fibryla. Propagacja pasm o zgodnym i niezgodnym z oczekiwaniem poziomem hydrofobowości jest wyraźnie widoczna.

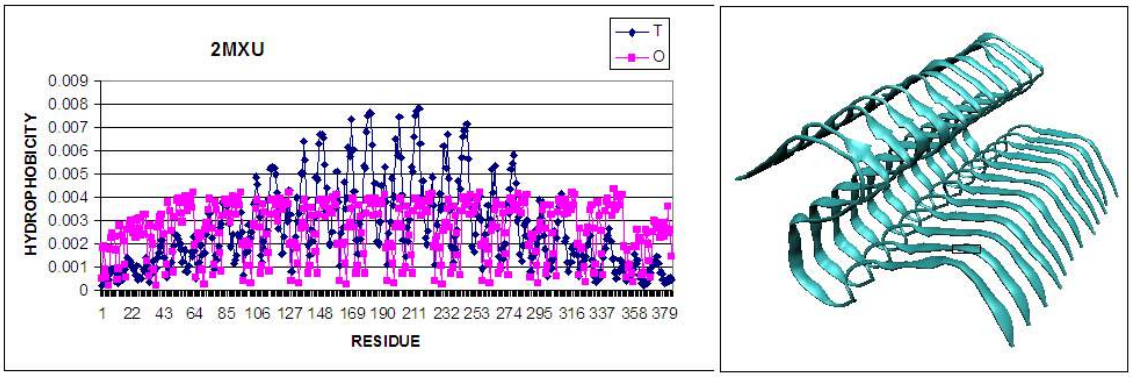


FIG. 10 ROZKŁAD T I O DLA AMYLOIDU. POKAZANE PROFILE OBRAZUJĄ STATUS AMYLOIDU Aβ(11-42) O RD = 0.680. PO PRAWEJ - PREZENTACJA 3D WYBRANEGO AMYLOIDU.

DANE DO WYKRESU UZYSKANO ZA ZGODĄ OD DR MATEUSZA BANACHA - ZAKŁAD BIOINFORMATYKI I TELEMEDYCYNY COLLEGIUM MEDICUM UNIWERSYTET JAGIELLOŃSKI

Ta niezgodność rozkładu hydrofobowości wynika z faktu, że w przypadku amyloidów poziom hydrofobowości przyjmuje formę pasm biegnących równolegle do osi fibryla. Ten pasmowy rozkład wynika z identycznej sekwencji kolejnych łańcuchów, gdzie dany aminokwas przyjmuje identyczną pozycję w kolejnych łańcuchach. Wśród tych pasów wyróżnia się pasy o poziomie zgodnym z oczekiwanym oraz niezgodnym z oczekiwaniem.

Problem omawiany w obecnej pracy to identyfikacja przyczyn, dla których dany łańcuch nie generuje struktury globularnej, która gwarantowałaby postać rozpuszczalną. Pytanie brzmi: czy istnieje specyfika sekwencji aminokwasowej, która wyklucza powstanie formy o centrycznym jądrze hydrofobowym?

W tym celu łańcuchy obecne w amyloidach dostępnych w PDB poddano fałdowaniu za pomocą programów I-TASSER oraz Roberta a także *fuzzy oil drop model*. Finalne struktury oceniono z punktu widzenia stopnia zgodności rozkładu

obserwowanego wobec rozkładu wyidealizowanego. Wysoki stopień takiej zgodności sugerowałby uzyskanie formy globularnej, a więc rozpuszczalnej.

2.2.1.6. WYKORZYSTANE OPROGRAMOWANIE

Programy I-TASSER [11] oraz Robetta [12] dostępne za pomocą serwerów umożliwiały wygenerowanie modeli struktur łańcuchów obecnych w omawianych amyloidach.

Oprogramowanie opracowane z uwzględnieniem *fuzzy oil drop model* wykorzystuje pakiet o nazwie GROMACS [29]. Za pomocą tego programu wykonywana była optymalizacja oddziaływań wewnętrznych (niewiążące). Dodatkowo naprzemiennie wprowadzono procedurę optymalizacji uporządkowania rozkładu hydrofobowości. Program ten dostępny jest w ramach PLGRID w Cyfronecie (http://plgrid.pl/) [30].

3. Wyniki

3.1. BIAŁKO TAU [5O3O]

Struktura TAU jest to amyloid sklasyfikowany na liście PDB jako m.in. 5O3O, 5O3L i 5O3T (te formy zostały poddane analizie) [31]. Ich specyficzność polega nie tylko na destrukcyjnym wpływie na mikrotubule, ale także na tworzeniu tzw. złogów w przestrzeni międzykomórkowej [32]. Istnieją tak naprawdę dwie formy amyloidów TAU: TAU o prostym filamencie i TAU o filamencie w postaci sparowanej helisy. Warto wspomnieć, że w swojej natywnej formie amyloid przyjmuje konformację, która jest wysoce odporna na temperaturę i środowisko kwasowe [33].

FIG. 11 AMYLOID 5O3O.

RYSUNEK POCHODZI ZE STRONY [HTTP://WWW.RCSB.ORG/](http://www.rcsb.org/)

W tym miejscu warto zwrócić uwagę na budowę amyloidów: jako cała jednostka składa się on z liniowo propagowanych, identycznych elementów. Każdy z elementów

to jeden lub wiele zwojów ułożonych względem siebie tak, aby podobny element mógł łatwo dołączyć się istniejącej już struktury. Zwoje zaś to łańcuchy polipeptydowe złożone z konkretnych sekwencji aminokwasów.

Wszystkie wyżej wymienione formy amyloidu poddane analizie wykazują duże różnice pomiędzy profilami teoretycznym (T) i obserwowanym (O) zarówno dla modelu T-O-R jak i dla modelu T-O-H. Zapis T-O-H oznacza wartość RD dla układu, w którym rozkład wynikający z ciągu wartości hydrofobowości własnej aminokwasu stanowi rozkład referencyjny zamiast rozkładu R. Oznacza to brak hydrofobowego jądra u każdej z tych struktur. Dodatkowo wysokie wartości RD wskazują na to, że proces fałdowania powyższych amyloidów jest kształtowany przez wewnętrzną hydrofobowość poszczególnych elementów łańcucha, a nie poprzez pole zewnętrzne, jak się zakłada w *fuzzy oil drop model* [34].

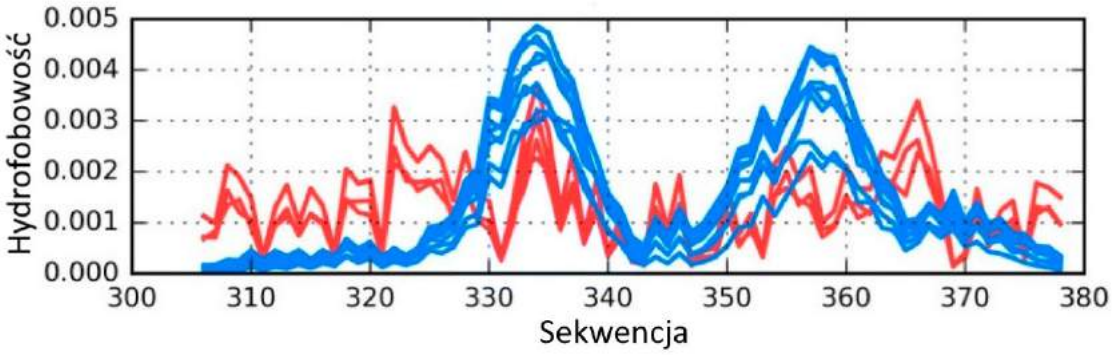


FIG. 12 ROZKŁAD HYDROFOBOWOŚCI W SUPERFIBRYLU TAU ZGODNIE Z SEKWENCJĄ RESZT W POSZCZEGÓLNYCH ŁAŃCUCHACH. NA NIEBIESKO ZAZNACZONO PRZEBIEG T, A NA CZERWONO PRZEBIEG O.

RYSUNEK Z ZAŁĄCZONEJ PRACY "FILAMENTOUS AGGREGATES OF TAU PROTEINS FULFIL STANDARD AMYLOID CRITERIA PROVIDED BY THE FUZZY OIL DROP (FOD) MODEL".

Wizualizacja jednostki amyloidu (np. 5O3O), pokazuje dla teoretycznego rozkładu dwa lokalne maksima oraz to, że fragmenty hydrofilne są eksponowane na powierzchni polipeptydu. W przypadku wizualizacji struktury obserwowanej nie

dostrzegamy tak jednoznacznie lokalnych maksimów w miejscach, gdzie występują one w teoretycznej wizualizacji. Tu (struktura obserwowana) mamy wiele lokalnych maksimów i minimów, które odpowiadają rozkładowi hydrofobowości wzdłuż całego pojedynczego zwoju. Cała struktura jest widoczna jako nałożone linie na grafie, gdyż kolejne propagacje zwojów mają bardzo podobne wizualizacje jedna do drugiej. Charakterystycznym zjawiskiem dla amyloidów są liczne lokalne maksima tam, gdzie w modelu teoretycznym obserwujemy niską hydrofobowość i na odwrót (co oczywiście nie daje jednoznacznego stwierdzenia, że mamy do czynienia z amyloidami przy odwróconym rozumowaniu).

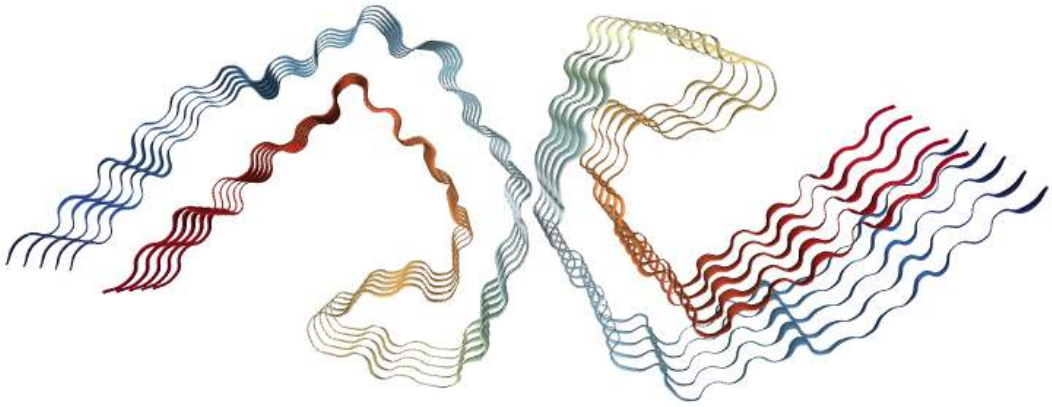


FIG. 13 AMYLOID 5O3L.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

Analizując pojedyncze zwoje amyloidu jako odseparowanej struktury (tj. dla której rozkład Gausa jest konstruowany indywidualnie a nie w kontekście całego amyloidu) dostajemy wyniki podobne do wyników w kontekście całego, rozpropagowanego polipeptydu (powyższy akapit), a mianowicie: wysokie wartości RD (zarówno dla T-O-R jak i dla T-O-H) oraz HvO dla bardzo niskich (wręcz ujemnych) wartości HvT i TvO. Ponownie wskazuje to na propagowanie się zwoju poprzez

wewnętrzną hydrofobowość pojedynczych elementów łańcucha oraz brak hydrofobowego jądra struktury. Wizualizując otrzymane wyniki, ponownie można zauważyć liczne lokalne maksima tam, gdzie spodziewana jest niska hydrofobowość wg. modelu teoretycznego.

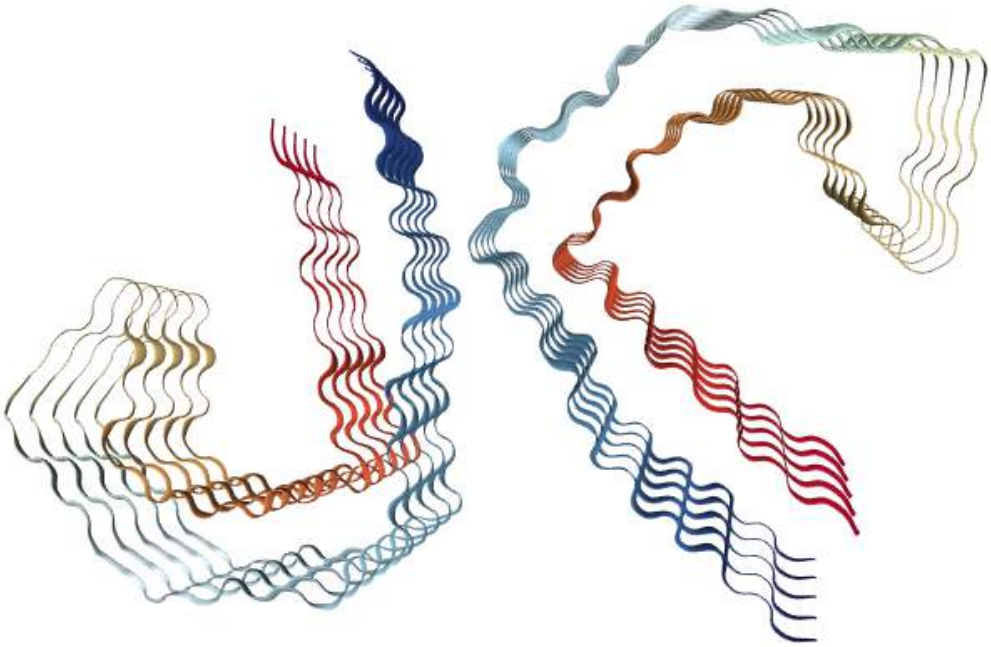


FIG. 14 AMYLOID 5O3T.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

W przypadku pojedynczych łańcuchów (traktowanych jako niezależne struktury i tak też analizowanych) wyniki są podobne do tych przedstawionych dla zwojów i całej jednostki amyloidowej, jednak nie wskazują jednoznacznie na amyloidową formę struktury. Analizowany łańcuch dzielimy na fragmenty, dla których różnica między modelem teoretycznym a obserwowanym jest szczególnie widoczna, w celu znalezienia przyczyn tych różnic. Pojedyncze fragmenty łańcucha różnią się w szczegółach, ale ogólnie są bardzo podobne niezależnie od ich struktury, tj. są bardzo niezgodne z modelem teoretycznym.

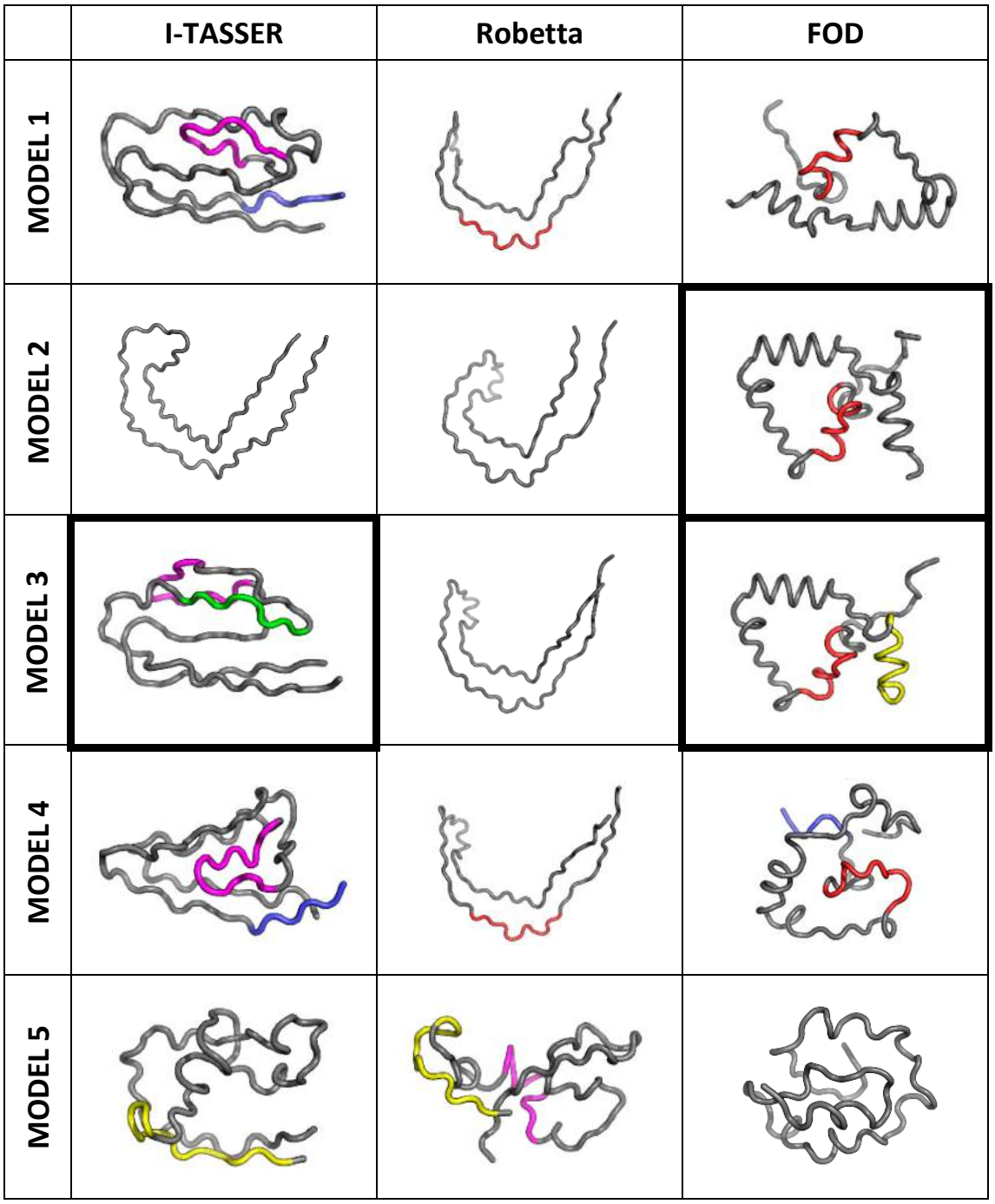


FIG. 15 REPREZENTACJA 3D STRUKTUR OTRZYMANYCH PRZY UŻYCIU I-TASSER, ROBETTA I *FUZZY OIL DROP* (FOD) *MODEL*. MODELE O RD <0.5 ZOSTAŁY OZNACZONE POGRUBIONĄ RAMKĄ.

Otrzymane wyniki (tj. przebieg teoretyczny) bazują na fałdowaniu *in silico* sekwencji aminokwasów zgodnej z TAU. Fałdowanie było przeprowadzone przy pomocy programów: I-TASSER, Robetta i odpowiadającego modelowi FOD. Na pięć wyników otrzymany przez I- TASSER aż 4 dawały wizualnie formę globularną, przy ostatecznie jednym wyniku z jądrem hydrofobowym; na pięć wyników otrzymanych z użyciem programu Robetta tylko jeden formował się w globulę, jednak bez hydrofobowego jądra; dla około 500 wyników FOD przeważały formy wizualnie zbliżone do globuli jednak tylko dwie z nich posiadały hydrofobowe jądro. Dalsza analiza otrzymanych wyników pozwoliła wyodrębnić fragment, który dla żadnego z modelów nie przyjmuje założeń teoretycznych, co potwierdzają wartości liczbowe, w szczególności bardzo wysokie RD. W związku z powyższym zidentyfikowaliśmy wyodrębniony fragment jako potencjalnie amyloidogenny, którego sekwencja nie przyjmuje zcentralizowanego rozmieszczenia hydrofobowości.

Sekwencję łańcucha TAU (wg. sekwencji obecnej w 5O3O) potraktowano jako INPUT dla wymienionych programów: I-TASSER, Robetta oraz FOD (*fuzzy oil drop model*). Uzyskane wyniki - modele finalne ocenione z punktu widzenia struktury jądra hydrofobowego podaje Tab. 2.

	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
I-TASSER	0.519	0.635	0.478	0.605	0.513
Robetta	0.745	0.656	0.682	0.748	0.566
FOD	0.503	0.364	0.369	0.648	0.644

TAB. 2 WARTOŚCI RD UZYSKANE PRZY POMOCY I-TASSER, ROBETTA I FOD. WARTOŚCI ZAZNACZONE NA ZIELONO SĄ ZGODNE Z MODELEM FOD (RD < 0.5).

Interpretacja wyników podanych w Tab. 2 wskazuje na znikome możliwości uzyskiwania struktury wykazującej centryczną strukturę jądra hydrofobowego.

Powyższe wyniki sugerują ponadto, że specyfika sekwencji aminokwasowej obecnej w amyloidzie TAU nie posiada możliwości generowania struktury globularnej o rozkładzie zgodnym z rozkładem teoretycznym wykazując tym samym, że konieczność poszukiwanie innego rozwiązania w przypadku tego białka jest oczywista.

Pełna analiza amyloidu TAU jest dostępna w Załącznik nr. 1.

3.2. STRUKTURA Aβ(15-40) [2MPZ]

Struktura 2MPZ jest to amyloid Aβ zawierający mutację zwaną „mutacją Iowa" (D23N) [Załącznik nr. 2]. Wyniki zarówno I-TASSER jaki i oprogramowania bazującego na FOD dostarczyły pięciu różnorodnych struktur: rozpoczynając od formy globularnej kończąc na formie bliskiej zupełnemu rozfałdowaniu pomiędzy którymi znajdziemy formy pośrednie.

Analiza superfibrylu struktury 2MPZ obrazuje brak hydrofobowego jądra i dużą rozbieżność od modelu *fuzzy oil drop*, zastąpioną symetryczną zgodnością identycznych fragmentów łańcucha. Dalsza analiza wskazuje, że wewnętrzna hydrofobowość ma relatywnie ograniczony wpływ na konformację superfibrylu. Hydrofobowość centralnych części struktury jest niemal identyczna i zmienia się wraz z kolejnymi peptydami będącymi w dalszej odległości od centrum łańcucha. Analiza poszczególnych protofibryli ukazuje rozkład charakterystyczny dla struktur amyloidowych, czyli bez centrycznej hydrofobowości tj. bez hydrofobowości skoncentrowanej w centralnej części protofibryla. Zagłębiając się dalej w strukturę badamy pojedyncze łańcuchy protofibryla, aby wykryć element potencjalnie odpowiedzialny za amyloidową postać peptydu. Pojedyncze łańcuchy w kontekście całego protofibryla wykazują (dla niektórych fragmentów) całkowite przeciwieństwo wartościom obserwowanym i teoretycznym, zapobiegając formowaniu się postaci globularnej peptydu. W końcu analiza części łańcucha protofibryla w oderwaniu od niego samego pozwala na zidentyfikowanie fragmentu wysoce niezgodnego z rozkładem teoretycznym. Dalsza

analiza pozwala sklasyfikować tenże fragment (22-28) jako reprezentatywny dla amyloidogenezy.

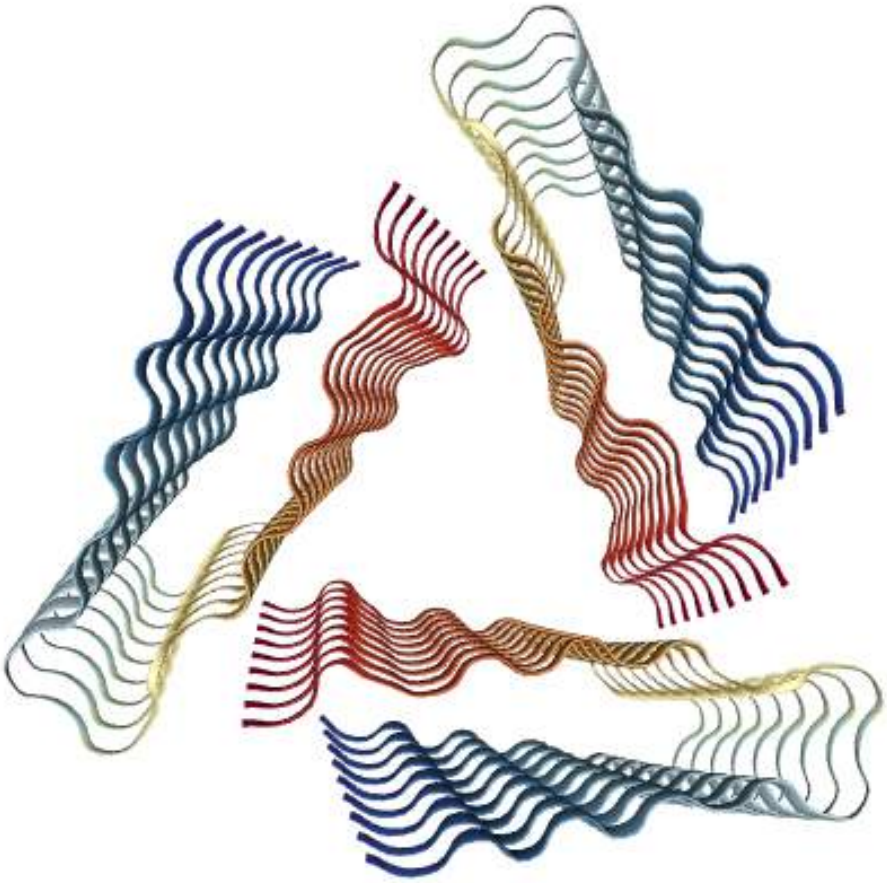


FIG. 16 AMYLOID 2MPZ.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

Porównując wszystkie 10 struktur otrzymanych z użyciem programów I-TASSER i FOD widzimy, że niektóre z wygenerowanych struktur przyjmują konformację globularną, dodatkowo zgodną z dystrybucją 3D Gausa, co oznacza, że struktury te mają hydrofobowe jądra. Jest tak zarówno w wypadku FOD gdzie algorytm zawiera optymalizację pod kątem działania środowiska wodnego na strukturę, jak i w wypadku

I-TASSER gdzie podejście jest inne. W wyżej wymienionych wypadkach struktur globularnych elementy hydrofilne są eksponowane podczas gdy elementy hydrofobowe znajdują się bliżej jądra struktury, jednakże nie wykazują one cech rozpuszczalności. Każda kolejna wersja wygenerowanych białek odbiega od form globularnych przyjmując coraz bardziej formę amyloidową.

Skupiając się na konkretnych fragmentach łańcucha zidentyfikowanych jako potencjalne „źródła" przyjmowania przez białko formy amyloidowej dokonywana jest analiza tychże fragmentów pod kątem amyloidogenezy. Na tą chwilę nie jesteśmy w stanie jednoznacznie stwierdzić które części łańcucha są odpowiedzialne za przyjmowanie formy patologicznej, ale kilka z nich określamy jako wysoce prawdopodobne przy inicjacji amyloidu.

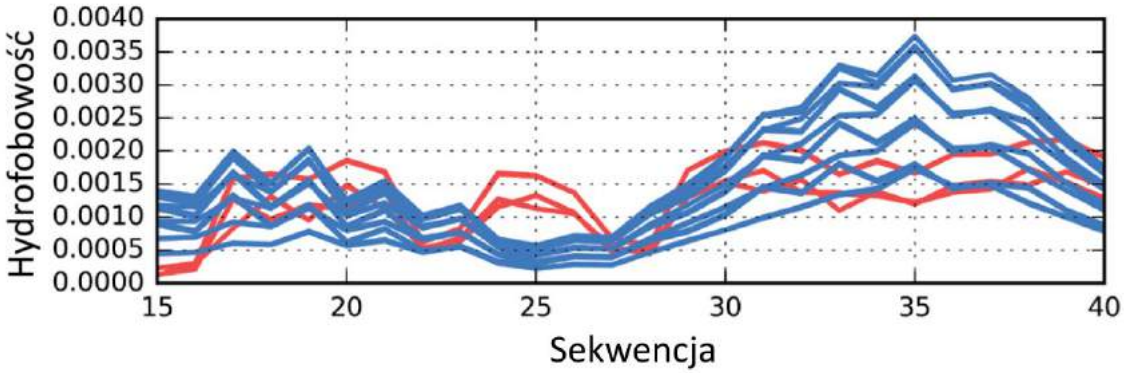


FIG. 17 ROZKŁAD HYDROFOBOWOŚCI W SUPERFIBRYLU Aβ(15-40) ZGODNIE Z SEKWENCJĄ RESZT W POSZCZEGÓLNYCH ŁAŃCUCHACH. NA NIEBIESKO ZAZNACZONO PRZEBIEG T, A NA CZERWONO PRZEBIEG O.

RYSUNEK Z ZAŁĄCZONEJ PRACY "STRUCTURAL ANALYSIS OF THE Aβ(15-40) AMYLOID FIBRIL BASED ON HYDROPHOBICITY DISTRIBUTION".

Ciekawą obserwacją jest fakt formowania się amyloidu z trzech pojedynczych włókien (fibryli), tworzących wysoce symetryczną formę. Model FOD pozwala zrozumieć tę strukturę: mianowicie przedstawia się ona w tej postaci, gdyż struktura – zgodnie z modelem – formuje się poprzez kooperację pomiędzy włóknami, a nie wewnętrzną

hydrofobowością. W tym wypadku interfejs rozciąga się na całej długości amyloidu będąc w zgodności z modelem FOD z drugiej strony przyjmując formę patologiczną.

Analiza amyloidu 2MPZ pozwala nam wysnuć wniosek, że – bazując na modelu FOD – struktura przyjmuje formę patologiczną jako rezultat zmian w zewnętrznym polu oddziaływań, które w normalnych warunkach spowodowałoby wygenerowanie się formy polipeptydu zachowującego formę globularną. Naturalnym środowiskiem fragmentu peptydu jest membrana w związku z czym struktura pozbawiona swojego *chaperonu*, dodatkowo nie otoczona całkowicie środowiskiem wodnym przyjmuje formę niezgodną z centryczną hydrofobowością jądra. Z tego też powodu struktura „skupia" się na interakcji pomiędzy pojedynczymi elementami, bardziej niż na interakcji z polem zewnętrznym.

Jeśli podsumować struktury wygenerowane przez I-TASSER – które także pokazują szereg wyników od wysoce zgodnych z modelem po formy amyloidowe – potwierdzają one powyższe założenia, a mianowicie iż pole zewnętrzne bierze aktywny udział w formowanie się polipeptydu, w związku z czym w zależności od jego wpływu dostajemy bardziej lub mniej globularną formę finalną.

	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5
I-TASSER	0.426	0.466	0.558	0.620	0.870
Robetta	-	-	-	-	-
FOD	0.188	0.215	0.218	0.571	0.591

TAB. 3 WARTOŚCI RD UZYSKANE PRZY POMOCY I-TASSER, ROBETTA I FOD. WARTOŚCI ZAZNACZONE NA ZIELONO SĄ ZGODNE Z MODELEM FOD (RD < 0.5). POMIMO SPOREJ LICZBY MODELI ZGODNYCH Z FOD, TAK NAPRAWDĘ REPREZENTUJĄ ONE KILKA REZULTATÓW NA KILKASET SYMULACJI.

Wyniki pokazane w Tab. 3 wskazują na wyraźne preferowanie form strukturalnych pozbawionych centrycznego jądra hydrofobowego. Pakiet Roberta nie

dostarczył rozwiązania dla omawianego łańcucha. Zaznaczyć należy, że wśród kilkuset modeli dostarczonych przez model FOD jedynie trzy podane w Tab. 3 wykazują obecność jądra hydrofobowego.

Pełna analiza amyloidu 2MPZ jest dostępna w Załącznik nr. 2.

3.3. STRUKTURA Aβ(11-42) [2MXU]

Kolejną przeanalizowaną strukturą jest fragment białka Aβ(1-42) bez jego odcinka N-końcowego, czyli amyloid Aβ(11-42). Raz jeszcze przyjrzymy się najpierw strukturze jako pojedynczemu włóknu, następnie wybranemu łańcuchowi w kontekście całego włókna a na koniec wybranemu łańcuchowi jako indywidualnej jednostce.

FIG. 18 AMYLOID 2MXU.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

Struktura polipeptydu został wygenerowana *in silico* z użyciem zarówno I-TASSER, Robetta jak i FOD. Dla każdego z narzędzi otrzymaliśmy 5 wyników które poddaliśmy analizie pod kątem hydrofobowego jądra w kontekście całego włókna jak również wybranego, pojedynczego łańcucha tejże struktury.

Struktura – poddana analizie jako włókno – wykazuje typowe rozbieżności pomiędzy T i O, z obserwowanymi wysokimi wartościami na powierzchni kompleksu (co jest w sprzeczności z oczekiwaniami). Dodatkowo widoczny wzorzec sinusoidalny wskazuje na identyczne ukształtowanie się poszczególnych łańcuchów ze zmiennymi wartościami – na przemian wysokimi i niskimi – hydrofobowości eksponowanej z struktury. Wyżej wymienione rozbieżności wyraźnie pokazują brak hydrofobowego jądra w obserwowanej strukturze.

W wypadku pojedynczego łańcucha – który analizujemy w związku z tym, że trudno nam określić właściwości włókna biorąc pod uwagę skończoną liczbę łańcuchów, gdy ten może się propagować w nieskończoność – w kontekście całego włókna znajdujemy obszary, gdzie T – można przyjąć - pokrywa się z O a także z H, ale uwagę zwracają tak naprawdę ogromne różnice dla pozostałych fragmentów. Aby zidentyfikować te fragmenty, wyliczamy współczynniki korelacji HvT, TvO i HvO dla poszczególnych odcinków 5 aminokwasowych. Analiza ta pozwala wyodrębnić fragment (pozycja 5 do 11) zgodny z modelem Gaussowskim jak również zupełnie mu przeciwstawne (pozycja 11 do 16; pozycja 24 do 18). Pojawiają się także fragmenty posiadające lokalne niezgodności (pozycja 18 do 23). Dla fragmentów łańcucha niezgodnych z modelem o centrycznej hydrofobowości uwagę zwraca obecność nieoczekiwanych lokalnych maksimów tam, gdzie powinny znajdować się lokalne minima i na odwrót. Ma to szczególny charakter w obecności lizyny, co wskazuje, że jej pozycja ma kluczowe znaczenie w kontekście centralizacji hydrofobowości struktur polipeptydowych.

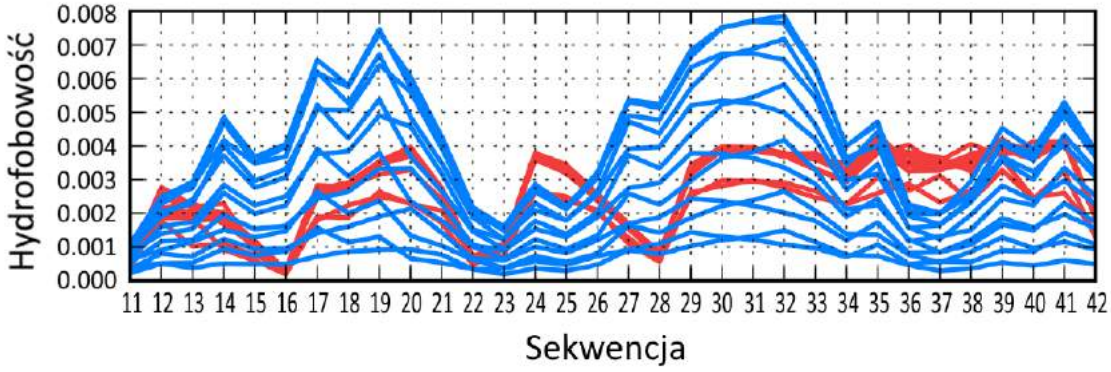


FIG. 19 ROZKŁAD HYDROFOBOWOŚCI W FIBRYLU Aβ(11-42) ZGODNIE Z SEKWENCJĄ RESZT W POSZCZEGÓLNYCH ŁAŃCUCHACH. NA NIEBIESKO ZAZNACZONO PRZEBIEG T, A NA CZERWONO PRZEBIEG O.

RYSUNEK Z ZAŁĄCZONEJ PRACY "STRUCTURAL ANALYSIS OF THE Aβ(11–42) AMYLOID FIBRIL BASED ON HYDROPHOBICITY DISTRIBUTION".

Analiza wyników modelowania za pomocą I-TASSER, Robetta i FOD dla 2MXU w postaci komparatywnego porównania całego łańcucha jak i zadanych jego fragmentów wraz z wynikami obserwowanymi dla pojedynczego łańcucha w kontekście całego włókna jak i bez takiego kontekstu wykazują co następuje. Głównym kandydatem dla transformacji amyloidowej w analizowanym amyloidzie są dwa fragmenty: 11-16 i 24-28. W wygenerowanych strukturach, najbardziej zbliżonych do amyloidu (I-TASSER wynik 4 i Robetta wynik 4) obszary te nie spełniają warunku wysokiej wartości dla RD(T-O-H). Wysokie wartości tego parametru wymagają kompleksu wielołańcuchowego, jako że nie występują raczej dla pojedynczych łańcuchów. Dlatego wyniki te są poprawne tylko przyjmując model FOD jako metody identyfikacji jądra hydrofobowego i jego transformacji.

Podsumowując, rezultaty badań wykazują, że Aβ(11-42) może przyjąć teoretycznie różne formy strukturalne, w tym formy o charakterze globularnym z hydrofobowym jądrem. Dalsze wnioski pozwalają nam stwierdzić, że pewne fragmenty struktury możemy wskazywać jako potencjalne źródło zmian amyloidowych, przy

założeniu, że warunki zewnętrzne będą czynnikiem startowym takiej przemiany. Najwyższą niezgodność względem modelu teoretycznego wykazuje pozycja lizyny, gdzie bardzo wysokie lokalne minimum dla O jest skorelowane z bardzo wysokim lokalnym maksimum dla T. Kiedy fragmenty amyloidu agregują się w sposób linearny, to właśnie na pozycji lizyny wchodzą ze sobą w kontakt, tworząc wiązkę hydrofilną. To pokazuje, że siły hydrofobowe przeciwstawiają się interakcjom elektrostatycznym i że niezgodności w zewnętrznym polu mogą powodować układanie się struktury względem hydrofobowości wewnętrznej. Ta hipoteza potwierdza zgodność modelu FOD, co oznacza, że osłabienie zewnętrznego pola (jakim w tym wypadku jest środowisko wodne) prowadzi do fałdowania się struktury względem hydrofobowości wewnętrznej a nie w kierunku hydrofobowego jądra, co w rezultacie prowadzi do powstania struktur amyloidowych.

	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5
I-TASSER	0.376	0.428	0.487	0.715	0.768
Robetta	0.256	0.283	0.320	0.457	0.660
FOD	0.234	0.240	0.242	0.555	0.559

TAB. 4 WARTOŚCI RD UZYSKANE PRZY POMOCY I-TASSER, ROBETTA I FOD. WARTOŚCI ZAZNACZONE NA ZIELONO SĄ ZGODNE Z MODELEM FOD (RD < 0.5). POMIMO SPOREJ LICZBY MODELI ZGODNYCH Z FOD, TAK NAPRAWDĘ REPREZENTUJĄ ONE KILKA REZULTATÓW NA KILKASET SYMULACJI.

Pełna analiza amyloidu 2MXU jest dostępna w Załącznik nr. 3.

3.4. STRUKTURA Aβ(1-40) [2MVX]

Struktura Aβ(1-40) to polipeptyd z tak zwaną mutacją Osaka znany pod kodową nazwą 2MVX w bazie PDB. Trzy programy: I-TASSER, Robetta i FOD zostały wyselekcjonowane do wygenerowania teoretycznych struktur bazujących na ciągu

aminokwasów 2MVX. Rezultat pokazał, że przy pewnych warunkach, struktura ta może przyjąć formę globularną oraz że środowisko pełni w tym procesie kluczową rolę. Dla każdego z powyższych narzędzi wygenerowaliśmy kilka struktur (po pięć dla I- TASSER i FOD oraz 10 dla Robetta) bazujących na sekwencji 2MVX a następnie porównaliśmy pod kątem całego fibrylu jak i poszczególnych łańcuchów ze strukturą obserwowaną.

Jak widać na rysunku (Fig. 20) struktura 2MVX składa się z dwóch propagujących się włókien wykazujących symetrię C2 – każde z włókien przypomina wypłaszczone „C" wsunięte jedno w drugie. W szczególności widoczny jest brak koncentryczności hydrofobowości w centralnej części tej struktury. Hydrofobowość jest równomiernie rozłożona wzdłuż głównej osi poszczególnych włókien. Taki rozkład naprzemiennej propagacji wysokich i niskich obszarów hydrofobowości jest charakterystyczny dla amyloidów.

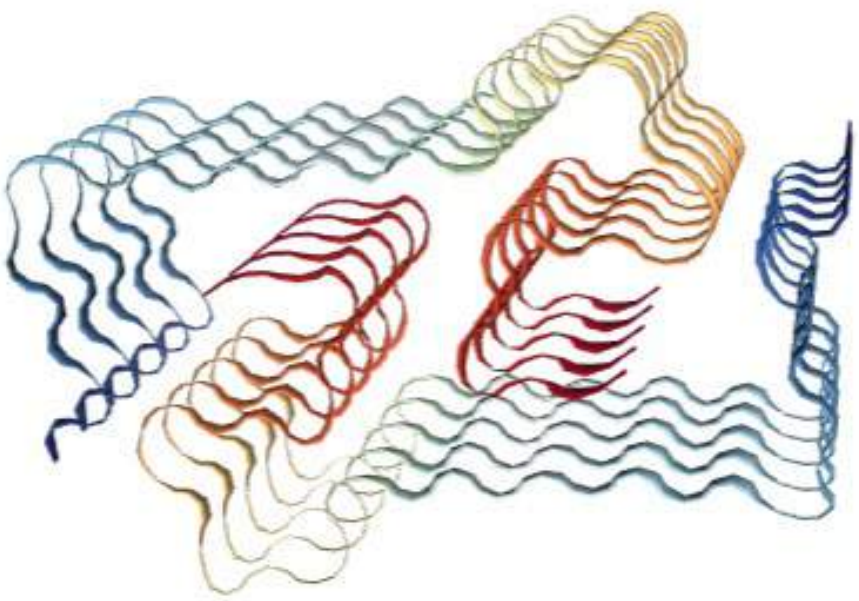


FIG. 20 AMYLOID 2MVX.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

Analiza wyników dla FOD sugeruje, że struktura jako całość nie posiada jądra hydrofobowego, aczkolwiek pojedyncze włókno mogłoby wytworzyć takie jądro, gdyby struktura nie podążała w fałdowaniu za wewnętrznymi właściwościami hydrofobowymi poszczególnych elementów włókna (co jest widoczne porównując teoretyczny model z obserwowanym jak również wartości O i H). Dalsza analiza pozwala nam na zidentyfikowanie fragmentów i pozycji szczególnie różnych w modelu teoretycznym w porównaniu do struktury obserwowanej – te obszary identyfikujemy jako potencjalnie amyloidogenne. Podobne rezultaty otrzymujemy analizując pojedynczy łańcuch w kontekście całej struktury.

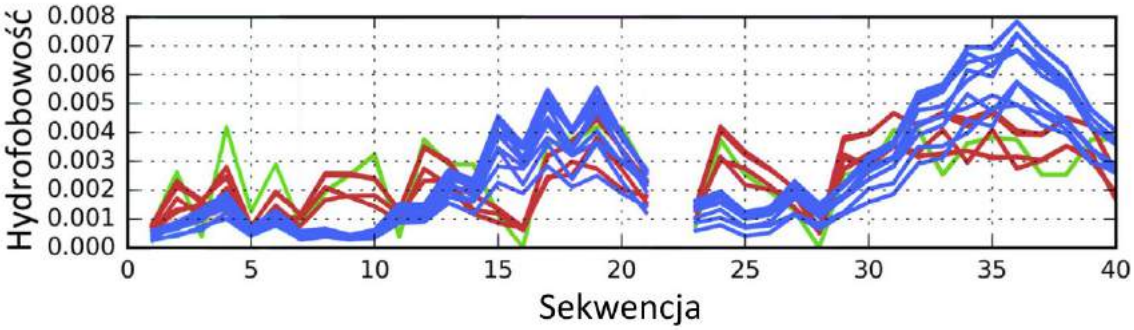


FIG. 21 ROZKŁAD HYDROFOBOWOŚCI W SUPERFIBRYLU Aβ(1-40) ZGODNIE Z SEKWENCJĄ RESZT W POSZCZEGÓLNYCH ŁAŃCUCHACH. NA NIEBIESKO ZAZNACZONO PRZEBIEG T, NA CZERWONO O A NA ZIELONO H.

RYSUNEK Z ZAŁĄCZONEJ PRACY "ANALYSIS OF ALTERNATIVE CONFORMATIONS OF THE Aβ(1-40) AMYLOID PROTEIN".

Szczegółowe wyniki i analiza w punktach styku pojedynczych włókien pozwalają stwierdzić, że choć samo, pojedyncze włókno jest zdominowane w fazie fałdowania przez wewnętrzne właściwości hydrofobowe jego komponentów to już cały superfibryl (kompleks) formuje się jako rezultat interakcji pomiędzy pojedynczymi włóknami biorąc pod uwagę obecność środowiska zewnętrznego, czyli wody.

Jak już wcześniej wspomniano, wysoka wartość HvO i negatywne wartości współczynników korelacji HvT i TvO pozwalają nam zidentyfikować potencjalne obszary amyloidogenne danej struktury. Jeśli dodatkowo wartość RD jest powyżej 0.5, możemy podejrzewać, że dany fragment struktury jest zdeterminowany przez wewnętrzną hydrofobowość jego elementów. Część wyników dla wygenerowanych struktur przez FOD i Robetta daje rezultaty zgodne z rozkładem teoretycznym, który mówi o możliwości wytworzeniu się hydrofobowego jądra w wygenerowanej strukturze. W większości jednak wypadków wartość RD dla T-O-H modelu referencyjnego nie jest zbyt wysoka, co oznacza formowanie się polipeptydu w kierunku hydrofobowych właściwości wewnętrznych elementów struktury. Co ciekawe to głównie modele wygenerowane przy pomocy Robetta są wysoce zgodne z monocentrycznym rozkładem hydrofobowości, podczas gdy aplikacja ta nie traktuje środowiska zewnętrznego jako pole a raczej pojedyncze cząsteczki wody.

Podsumowując: Robetta wygenerowała cztery modele z RD poniżej 0.5 co sugeruje obecność jądra hydrofobowego, podczas gdy I-TASSER nie wygenerował ani jednego takiego modelu. Struktury wygenerowane przez FOD promują internalizację elementów hydrofobowych, aczkolwiek, finalnie znacznie odbiegają od form globuli.

	MODEL 1	**MODEL 2**	**MODEL 3**	**MODEL 4**	**MODEL 5**
I-TASSER	0.602	0.623	0.634	0.642	0.702
Robetta	0.434	0.458	0.465	0.465	0.508
FOD	0.250	0.266	0.279	0.629	0.631

TAB. 5 WARTOŚCI RD UZYSKANE PRZY POMOCY I-TASSER, ROBETTA I FOD. WARTOŚCI ZAZNACZONE NA ZIELONO SĄ ZGODNE Z MODELEM FOD (RD < 0.5). POMIMO SPOREJ LICZBY MODELI ZGODNYCH Z FOD, TAK NAPRAWDĘ REPREZENTUJĄ ONE KILKA REZULTATÓW NA KILKASET SYMULACJI.

Wnioski wynikające z analizy wyników podanych w Tab. 5 są podobne w swojej wymowie do podanych w poprzednich przykładach.

Pełna analiza amyloidu 2MVX jest dostępna w Załącznik nr. 4.

3.5. STRUKTURA α-SYNUKLEINA [2N0A]

α-Synukleina występuje głównie w komórkach mózgowych (łącząc się z pęcherzykami presynaptycznymi) [35]. Łańcuch α-Synukleiny składa się z fragmentu N-końca (1-60), NAC (nieamyloidowy komponent beta; 61-95) i silnie hydrofilnego fragmentu C-końca (96-140) [36]. Aktywność α-Synukleiny jest na tą chwilę niestety słabo zrozumiana, aczkolwiek niepoprawne fałdowanie się tej struktury występuje w chorobie Parkinsona. Forma amyloidowa (w bazie PDB reprezentowana pod nazwą 2N0A) składa się z centralnego włókna (30-100), podczas gdy zarówno N-koniec jak i C-koniec przyjmują losową konformację zwoju. Różnica pomiędzy α-Synukleiną a pozostałymi amyloidami przejawia się w długich fragmentach łańcuch które nie są częścią samego włókna. Analiza struktury jako całego polipeptydu jak również poszczególnych jego fragmentów została przeprowadzona przy użyciu narzędzi I-TASSER, Robetta i FOD.

α-Synukleina w formie powiązanej z micelą (1XQ8) została poddana badaniom przy użyciu narzędzia realizującego model FOD. Struktura ta - jak się zakłada - spełnia warunki formy natywnej ze względu na jej stan skompleksowania z micelą. Analiza został przeprowadzona zarówno dla całego łańcucha jak i jego fragmentu (1-95) bez nieustrukturalizowanego luźnego C-końca (96-140). Analiza dla całego łańcucha wykazuje fragmenty, dla których rozkład teoretyczny różni się od obserwowanego, z tego też powodu łańcuchowi brakuje jednoznacznie widocznego jądra hydrofobowego. Jednak, jeśli wyeliminujemy te fragmenty z analizy (zakładając, że ich charakter ma postać kompleksu w białku), dostajemy strukturę z jądrem hydrofobowym. Jeśli zredukujemy łańcuch o C-koniec widzimy wyraźnie tendencję do

ustawiania się elementów hydrofilnych na powierzchni i elementów hydrofobowych wewnątrz struktury jednak wciąż pojawiają się fragmenty, dla których przebieg teoretyczny i obserwowany jest znacznie różny. Raz jeszcze, usunięcie tych fragmentów łańcucha daje wyniki które wykazują zcentralizowane, hydrofobowe jądro.

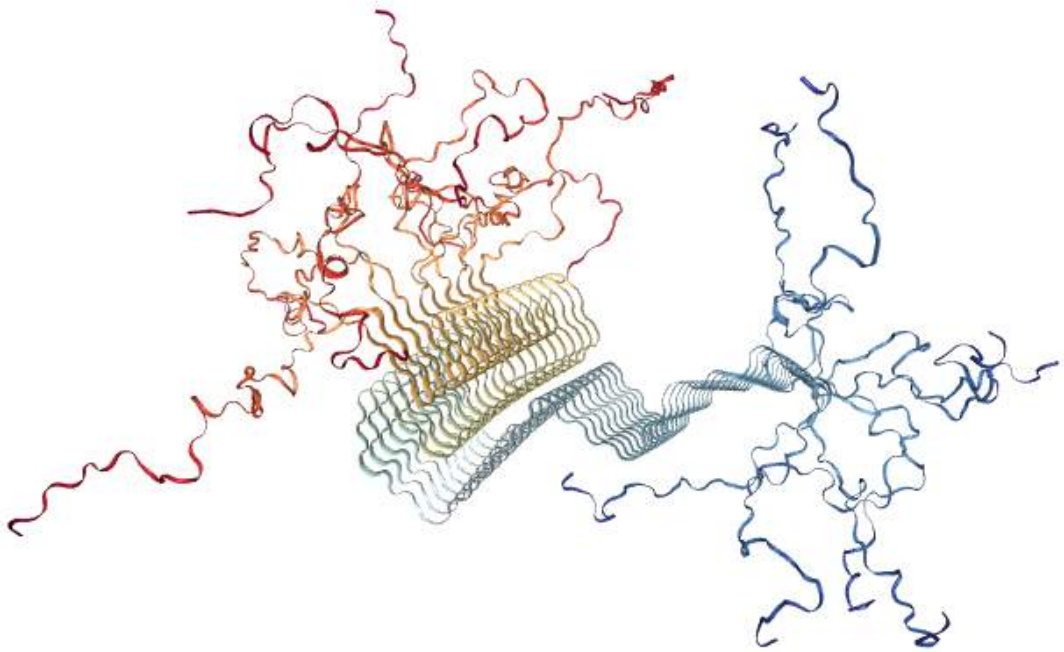


FIG. 22 STRUKTURA 2N0A.

RYSUNEK POCHODZI ZE STRONY HTTP://WWW.RCSB.ORG/

Przejdźmy teraz do analizy α-Synukleiny w jej formie amyloidowej (2N0A). Wyniki dla całego łańcucha wykazują, iż rozkład teoretyczny posiada dwa, dobrze widoczne maksima podczas gdy rozkład obserwowany posiada wiele lokalnych maksimów, najprawdopodobniej z powodu interakcji z innymi elementami tego samego łańcucha będącymi w najbliższym sąsiedztwie badanych fragmentów. Dotyczy to zarówno całego łańcucha jak i fragmentu amyloidowego, który – jak wcześniej zostało

to już wspomniane - były analizowany oddzielnie. Interesującą obserwacją jest dystrybucja hydrofobowości, która wykazuje posiadanie hydrofobowego jądra zarówno patrząc na strukturę całościowo jak i wyodrębniając amyloidowy fragment (w tym drugim wypadku wymaga to usunięcia 4 aminokwasów z analizy). Zaskakująca jest też wysoka zgodność modelu teoretycznego z obserwowanym pod kątem rozkładu hydrofobowości we fragmencie amyloidowym.

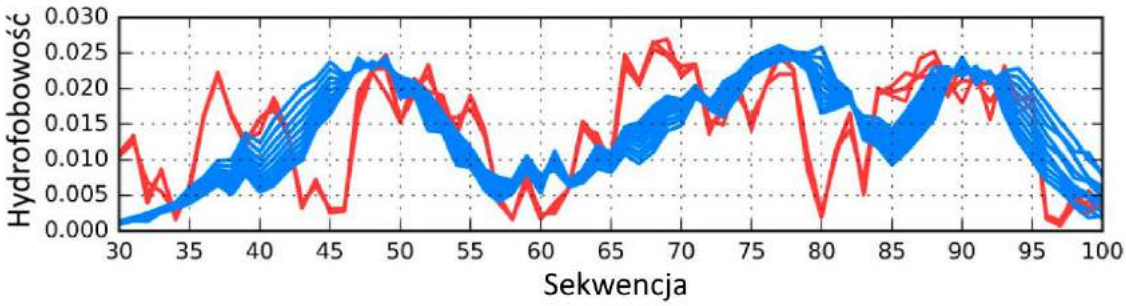


FIG. 23 ROZKŁAD HYDROFOBOWOŚCI W FIBRYLU α-SYNUKLEINY ZGODNIE Z SEKWENCJĄ RESZT W POSZCZEGÓLNYCH ŁAŃCUCHACH. NA NIEBIESKO ZAZNACZONO PRZEBIEG T, A NA CZERWONO PRZEBIEG O.

RYSUNEK Z ZAŁĄCZONEJ PRACY "ALTERNATIVE STRUCTURES OF α-SYNUCLEIN".

To ostatnie można wytłumaczyć w następujący sposób: model zakłada, że globularne jądro powstaje poprzez interakcję pomiędzy aminokwasami a środowiskiem wodnym. Elementy hydrofobowe „chowają" się do środka struktury podczas gdy elementy hydrofilne eksponują się na jej powierzchni. Jednak, gdy właściwości środowiska się zmienią, kontrolę nad formowaniem się łańcucha przejmie wewnętrzna hydrofobowość pojedynczych aminokwasów. Forma 1XQ8 α-Synukleiny jest relatywnie zgodna z modelem FOD – tak się dzieje dzięki obecności stałego „opiekuna" (w tym wypadku komórka nerwowa) który nie pozwala przybrać innej formy danej strukturze. W momencie, gdy brakuje komórki nerwowej, która stabilizowałaby formę łańcucha, ten przechodzi zmiany konformacji prowadzącej do powstania struktury amyloidowej.

Analiza pojedynczego łańcucha jako części całej struktury wykazuje zbliżony rozkład hydrofobowości jak dla pozostałych łańcuchów, podczas gdy potraktowanie tegoż łańcucha jako osobnej struktury wykazuje dodatkowe niskie i wysokie wartości hydrofobowości niezgodne z modelem teoretycznym.

Porównanie wyników generacji modelów 3D dla całej sekwencji pod kątem posiadania przez nie jąder hydrofobowych pokazuje, że żaden nie uzyskał takiego wyniku (tj. został sfałdowany w sposób, który wykazuje obecność jądra hydrofobowego). Biorąc pod uwagę fakt, że metodą FOD generujemy około 500 różnych struktur a sama metoda faworyzuje tworzenie się jądra hydrofobowego, wynika stąd, że α-Synukleina tak naprawdę nie jest w stanie przyjąć rozpuszczalnej, globularnej formy białka.

Jeśli weźmiemy pod uwagę tylko początkowy fragment struktury (aminokwasy od 1 do 30) czy też końcowy fragment (aminokwasy od 100 do 140), są ona w stanie przyjąć formę globularną, co wykazują rezultaty generacji przy pomocy założonych narzędzi. Tak samo jest w wypadku środkowego fragmentu łańcucha (aminokwasy od 30 do 100) aczkolwiek w tym wypadku osiągnięcie formy globularnej udało się w bardzo małej liczbie przepadków, a w niektórych wypadkach wymagało dodatkowych założeń.

	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5
I-TASSER	0.592	-	-	-	-
Robetta	0.574	0.585	0.651	0.676	-
FOD	0.525	0.558	0.567	0.575	0.583

TAB. 6 WARTOŚCI RD UZYSKANE PRZY POMOCY I-TASSER, ROBETTA I FOD DLA CAŁEGO ŁAŃCUCHA α-SYNUKLEINY (1-140). WARTOŚCI ZAZNACZONE NA ZIELONO SĄ ZGODNE Z MODELEM FOD (RD < 0.5).

Podsumowując przeprowadzone badania należy zauważyć, iż najbardziej prawdopodobnym fragmentem inicjującym zmianę amyloidową w wypadku tej struktury wydaj się być ten pomiędzy 68 a 82 aminokwasem. Fragment ten wykazuje

silne amyloidowe preferencje konformacyjne i jest istotny w ustabilizowaniu formy amyloidowej całego łańcucha.

Sam fragment amyloidowy pokazuje nam, że przyjmowanie tejże formy wynikać może z mutacji, ale także - co wynika z powyższych badań - ze zmian w środowisku, które jest bardzo mało rozpoznane pod kątem biologicznym.

Nasuwa się jeszcze jedna ważna hipoteza, a mianowicie jeżeli łańcuch nie może przyjąć formy globularnej (tu poprzez obecność długich, nieuporządkowanych fragmentów terminalnych, które promują izolację środkowej części łańcucha), stara się przyjąć jak najkorzystniejszą formę, którą okazuje się być formą fibrylarną, gdzie eksponuje się wewnętrzna hydrofobowość poszczególnych aminokwasów.

Pełna analiza amyloidu 2N0A jest dostępna w Załącznik nr. 5.

4. Dyskusja

Cechą charakterystyczną dostępnych w bazie PDB struktur amyloidowych jest płaska struktura pojedynczego łańcucha. Z punktu widzenia *fuzzy oil drop model* oznacza to, że jeden z parametrów sigma (powiedzmy σ_Z) zbliża się do zera. Dlatego w świetle ostatnich badań zespołu odpowiednia reprezentacja dla rozkładu hydrofobowości dla pojedynczego łańcucha wyraża się funkcja 2D Gaussa.

Jeśli natomiast przyjąć strukturę fibryla, to rozkład w tej formie strukturalnej wyrażony może być za pomocą funkcji 3D Gauss dla σ_Z zdążającej do nieskończoności. Obrazuje to Fig. 24 poniżej.

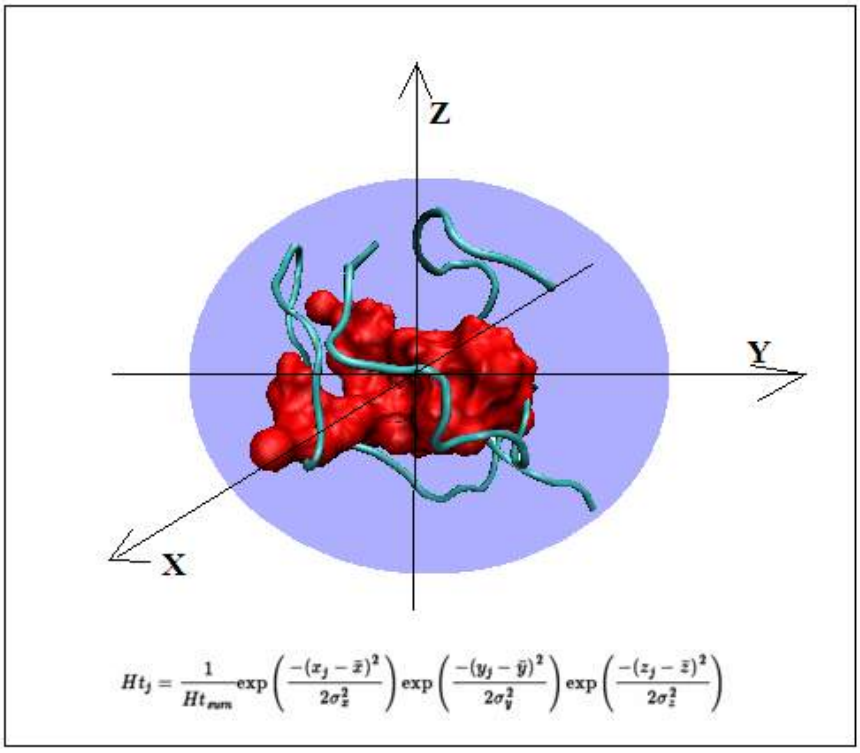


FIG. 24 MODEL DLA WYRAŻENIA FORMY GLOBULARNEJ BIAŁKA (FUNKCJA 3D GAUSSA). NA CZERWONO ZAZNACZONO RESZTY WCHODZĄCE W SKŁAD JĄDRA HYDROFOBOWEGO.

Forma globularna białka o rozkładzie zbliżonym do rozkładu typu 3D Gaussa wyraża postać białka rozpuszczalnego. Lokalne niezgodności sugerują jedynie rodzaj potencjalnej aktywności biologicznej.

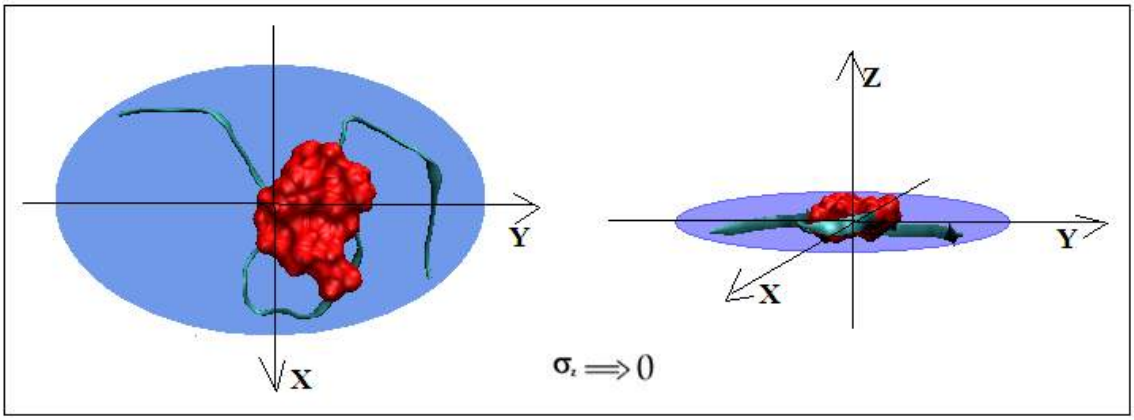


FIG. 25 PŁASKA STRUKTURA POJEDYNCZEGO ŁAŃCUCHA PRZY SIGMA ZBLIŻAJĄCEJ SIĘ DO ZERA. NA CZERWONO ZAZNACZONO RESZTY WCHODZĄCE W SKŁAD JĄDRA HYDROFOBOWEGO.

Konstrukcja struktury o rozkładzie hydrofobowości wyrażonej za pomocą 2D Gaussa określa tendencję do centralizacji reszt hydrofobowych w centralnej części płaskiej struktury. Ekspozycja hydrofobowości w tej sytuacji jest znacząca. Niekorzystny efekt entropowy można zmniejszyć kompleksując drugi płaski łańcuch o podobnym rozkładzie. Nie rozwiązuje to jednak problemu, ponieważ terminalne łańcuchy stale wykazują kontakt wysokiej hydrofobowości ze środowiskiem wodnym. Tłumaczy to fakt lawinowego przyłączania kolejnych łańcuchów prowadzących do tworzenia fibryli o teoretycznie nieograniczonej długości.

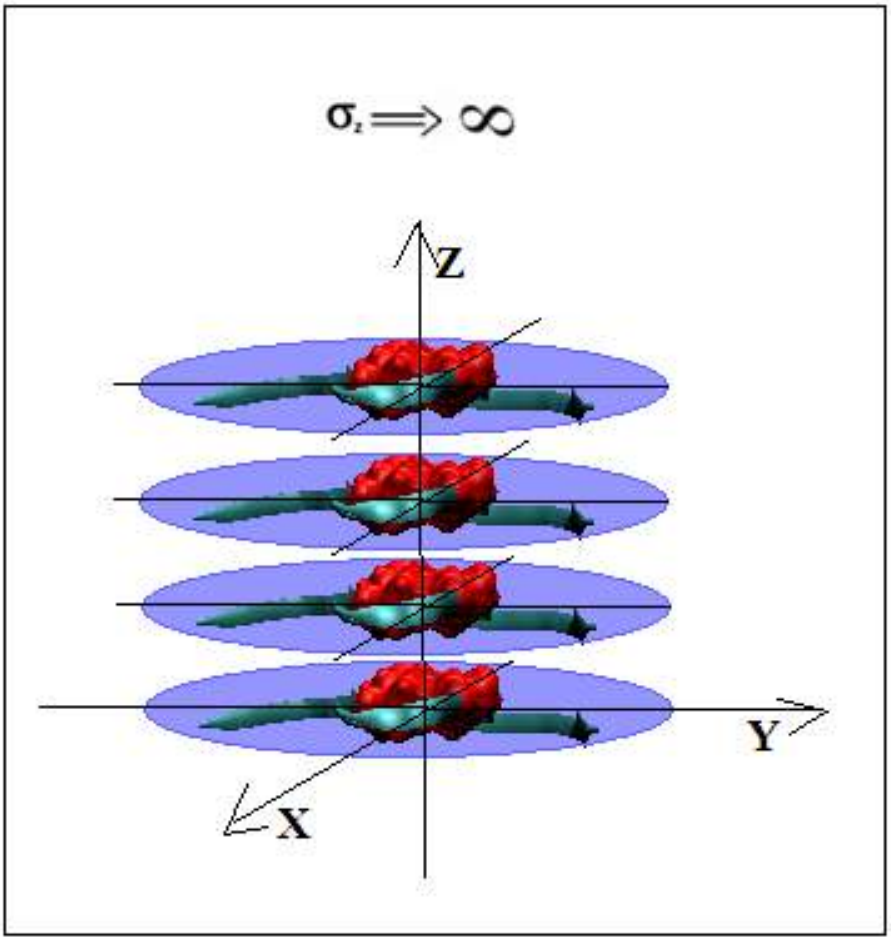


FIG. 26 STRUKTURA FIBRYLA PRZY WIELOKROTNIE ZWIĘKSZONEJ LICZBIE POJEDYNCZYCH PŁASKICH ŁAŃCUCHÓW SKUTKUJE WZROSTEM PARAMETRU σ_Z FUNKCJI 3D GAUSSA. NA CZERWONO ZAZNACZONO RESZTY WCHODZĄCE W SKŁAD JĄDRA HYDROFOBOWEGO.

W ten sam sposób tłumaczy się fakt tworzenia superfibryli, gdzie układ dwóch czy trzech łańcuchów prowadzi do stanu koncentracji hydrofobowości w centralnej części superfibryla, co pokazano w [37].

5. Wnioski

Analiza porównawcza w załączonych pracach potwierdza hipotezę dotyczącą właściwości strukturalnych amyloidu. Zgodnie z tą hipotezą amyloid charakteryzuje się obecnością naprzemiennych pasm o zmiennej hydrofobowości. Wydaje się, że propagacja liniowa - którą można uznać za sprzeczną z pojawieniem się scentralizowanego jądra hydrofobowego (co widać w białkach globularnych) - jest charakterystyczną właściwością amyloidów.

Jedna z interpretacji analizowanych wyników sugeruje wpływ środowiska zewnętrznego na niepoprawne fałdowanie się białek poprzez brak działania w celu utworzenia jądra hydrofobowego. Szczegółowa analiza pokazuje, że siły hydrofobowe mają większy wpływ na fałdowanie niż interakcje elektrostatyczne i dysproporcje w zewnętrznym polu siłowym mogą produkować konformacje bazujące na wewnętrznej hydrofobowości elementów łańcucha. Potwierdza to poprawność metody *fuzzy oil drop model*, która generuje formy globularne biorąc pod uwagę obecność rozpuszczalnika jako zewnętrznego pola siły, ukierunkowując elementy hydrofobowe do środka struktury białkowej i eksponując elementy hydrofilne na zewnątrz tejże struktury. Proces zakłada, że struktura białka jest ściśle zależna od właściwości rozpuszczalnika, w związku z czym zmiany w jego właściwościach mogą powodować zmiany w konformacji otrzymanych łańcuchów polipeptydowych, tak jak jest to eksperymentalnie zaobserwowane [Załącznik nr. 5].

Nasze obserwacje dowodzą, iż zmiany w środowisku zewnętrznym są wystarczające, aby rozfałdować białko (np. dodając do środowiska wodnego mocznik lub wytrząsając). Wstrząsanie zwiększa napowietrzenie wody, a co za tym idzie zmienia właściwości środowiska, co w rezultacie prowadzi do zmian konformacyjnych białka. Co ciekawe zaprzestanie wstrząsania lub też oddializowanie mocznika powoduje, że białko

wraca do swojej pierwotnej funkcji. A więc rola właściwie zorganizowanego środowiska spełnia swoją funkcję ukierunkowującą.

Podsumowując otrzymane wyniki, warto zadać pytanie czemu polipeptyd zdolny do wygenerowania formy globularnej, przyjmuje konformację fibryla? Jeśli odwołamy się do modelu *fuzzy oil drop* odpowiedzią jest zmiana w zewnętrznym polu siły (środowisku), który w normalnych warunkach prowadziłby łańcuch polipeptydowy do przyjęcia formy globularnej.

W jednej z prac [Załącznik nr. 3] autorzy sugerują, że transformacja amyloidowa może być skutkiem niewystarczającego wpływu sił zewnętrznych (wody) na proces fałdowania, co jest zgodne z wynikami badań dla poszczególnych amyloidów. Zatem osłabienie zewnętrznego pola siły (środowiska wodnego) w rezultacie pozwala na dominację wewnętrznej hydrofobowości elementów łańcucha, a co za tym idzie zapobiega procesowi fałdowania w kierunku monocentrycznego jądra hydrofobowego.

Niektóre z otrzymanych struktur wykazują zdolność do przyłączania kompleksów z innymi, identycznie sfałdowanymi łańcuchami. Taka struktura może zawierać fragmenty które możemy zidentyfikować jako źródła wzorca konformacji idącego w kierunku hydrofobowości wewnętrznej poszczególnych elementów łańcucha. Dotychczasowo zaproponowany model zakłada traktowanie struktury globularnej białka jako rezultat zewnętrznego pola siły formy 3D Gaussa. Jednakże środowisko reprezentowane w postacie funkcji 2D Gaussa promuję formę strukturalną obserwowaną w amyloidach.

Można zatem stwierdzić, że omawiane sekwencje w łańcuchach obecnych w amyloidach wykazują niskie możliwości wygenerowania struktury globularnej. Zaznaczyć należy, że białka omawiane tutaj w warunkach fizjologicznych pozostają w kompleksach z odpowiednimi organellami tkanki nerwowej. Bazując na założeniach *fuzzy oil drop model*, białko przygotowane do kompleksowania innego obiektu (innego

białka, błony czy innej organelli) powinien mieć wyeksponowaną odpowiednią kompatybilną do obiektu docelowego powierzchnię nie polarną. Skutkuje to predyspozycją do pozostawania w kompleksie z molekułą docelową. Dlatego można *a priori* zakładać, że pełna forma globularna z powierzchnią pokrytą grupami polarnymi nie powinna być możliwa. Gdyby bowiem taka forma globularna była osiągalna, to zapewne pojawiałyby się trudności w odnajdowaniu specyficznego obiektu docelowego.

W sytuacji przewagi hydrofobowości wobec hydrofilności jedyna forma, jaka może być osiągnięta to struktura o centrycznym jądrze, ale w reprezentacji 2D Gauss [37].

6. Bibliografia

1. Dill, K.A.; MacCallum, J.L. The protein-folding problem. 50 years on. Science 2012, 338, 1042-1046.
2. Anfinsen, C.B. The formation and stabilization of protein structure. The Biochemical Journal 1972, 128 (4), 737–49.
3. Banach, M.; Konieczny, L.; Roterman, I. The Amyloid as a Ribbon-Like Micelle in Contrast to Spherical Micelles Represented by Globular Proteins. Molecules 2019, 24, 4395.
4. CASP13 proceedings: Proteins: Structure, Function, and Bioinformatics. 2019, 87 (12), 1007-1388.
5. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. Nucleic Acids Research 2000, 28, 235-242.
6. Bloom, G.S. Amyloid-β and tau: The trigger and bullet in Alzheimer disease pathogenesis. JAMA Neurol. 2014, 71, 505–508.
7. Sgourakis, N.G.; Yau, W.M.; Qiang, W. Modeling an in-register, parallel “iowa” aβ fibril structure using solid-state NMR data from labeled samples with rosetta. Structure 2015, 23, 216–227.
8. Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y. A beta (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer’s disease. Nat Struct Mol Biol 2015, 22 (6), 499–505.
9. Schütz, A.K.; Vagt, T.; Huber, M.; Ovchinnikova, O.Y.; Cadalbert, R.; Wall, J.; Meier, B.H. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. Angewandte Chemie International Edition 2014, 54 (1), 331-5.
10. Tuttle, M.D.; Comellas, G.; Nieuwkoop, A.J.; Covell, D.J.; Berthold, D.A.; Kloepper, K.D.; Courtney, J.M.; Kim, J.K.; Barclay, A.M.; Kendall, A.; et al. Solid-state NMR structure of a pathogenic fibril of full-length human-synuclein. Nat. Struct. Mol. Boil. 2016, 23, 409–415.
11. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat. Protoc. 2010, 5, 725–738.
12. Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004, 32, W526–W531.
13. Hartley, G. Aqueous solutions of paraffin-chain salts. Paris: Hermann & Cie. 1936.
14. Emsley, P.; Charles, I.G.; Fairweather, N.F.; Isaacs, N.W. Structure of Bordetella pertussis virulence factor P.69 pertactin. Nature 1996, 381, 90–92.

15. Bryliński, M.; Konieczny, L.; Roterman, I. Fuzzy-oil-drop hydrophobic force field – a model to represent late-stage folding (In Silico) of lysozyme. J. Biomol. Struct. Dyn. 2006, 23, 519–528.
16. Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 1984, 179, 125–142.
17. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 1982, 157, 105–132.
18. Konieczny L.; Bryliński M.; Roterman I. Gauss-function-Based model of hydrophobicity density in proteins. In Silico Biol. 2006, 6 (1-2), 15-22.
19. Banach, M.; Konieczny, L.; Roterman, I. Why do antifreeze proteins require a solenoid? Biochimie 2018, 144, 74–84.
20. Sałapa, K.; Kalinowska, B.; Jadczyk, T.; Roterman, I. Measurement of Hydrophobicity Distribution in Proteins - Non-redundant Protein Data Bank. Bio-Algorithms and Med-System 2012, 8, 327-338.
21. Dygut, J.; Kalinowska, B.; Banach, M.; Piwowar, M.; Konieczny, L.; Roterman, I. Structural Interface Forms and Their Involvement in Stabilization of Multidomain Proteins or Protein Complexes. Int. J. Mol. Sci. 2016, 17, 1741.
22. Banach, M.; Konieczny, L.; Roterman, I. Ligand binding cavity encoded as a local hydrophobicity deficiency. In: Roterman-Konieczna, I. editor. From Globular Proteins to Amyloids. Elsevier 2020, 91-93.
23. Gadzała, M.; Dułak, D.; Kalinowska, B.; Baster, Z.; Bryliński, M.; Konieczny, L.; Banach, M.; Roterman, I. The aqueous environment as an active participant in the protein folding process. J. Mol. Graph. Model. 2019, 87, 227-239.
24. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 1976, 104, 59-107.
25. Kullback, S.; Leibler, R.A. On information and sufficiency. Ann. Math. Stat. 1951, 22, 79–86.
26. Kalinowska, B.; Banach, M.; Konieczny, L.; Roterman, I. Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins. Entropy 2015, 17, 1477–1507.
27. Banach, M.; Konieczny, L.; Roterman, I. The late-stage intermediate. In: Roterman-Konieczna, I. editor. Protein Folding in Silico. Elsevier 2012, 21-37.
28. Banach, M.; Konieczny, L.; Roterman, I. Can the structure of hydrophobic core determine the complexation site? In: Roterman-Konieczna, I. editor. Identification of ligand binding site and protein-protein interaction area. Berlin: Springer 2013, 41–54.

29. van Gunsteren, W. F., Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. Angew. Chem. Int. Ed. Engl. 1990, 29,992–1023.
30. PLGrid User Manual. Available online: https://docs.cyfronet.pl/pages/viewpage.action?pageId=23626756
31. Fitzpatrick, A.W.P.; Falcon, B.; He, S.; Murzin, A.G.; Murshudov, G.; Garringer, H.J.; Crowther, R.A.; Ghetti, B.; Goedert, M.; Scheres, S.H.W. Cryo-EM structures of tau filaments from Alzheimer's disease. Nature 2017, 547, 185–190.
32. Maccioni, R.B.; Muñoz, J.P.; Barbeito, L. The Molecular Bases of Alzheimer's Disease and Other Neurodegenerative Disorders. Arch. Med. Res. 2001, 32, 367-381.
33. Cleveland, D.W.; Hwo, S.Y.; Kirschner, M.W. Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. J. Mol. Biol. 1977, 116, 227–247.
34. Roterman, I.; Banach, M.; Kalinowska, B.; Konieczny, L. Influence of the Aqueous Environment on Protein Structure—A Plausible Hypothesis Concerning the Mechanism of Amyloidogenesis. Entropy 2016, 18, 351.
35. Bonini, N.M.; Giasson, B.I. Snaring the Function of α-Synuclein. Cell 2005, 123, 359-361.
36. Cremades, N.; Chen, S.; Dobson, C. Structural Characteristics of α-Synuclein Oligomers. Int. Rev. Cell. Mol. Biol. 2017, 329, 79-143.
37. Fabian, P.; Banach, M.; Stąpor, K.; Konieczny, L.; Ptak-Kaczor, M.; Roterman, I. The Structure of Amyloid Versus the Structure of Globular Proteins. Int. J. Mol. Sci. 2020, 21, 4683.

Załącznik nr. 1

Filamentous Aggregates of Tau Proteins Fulfil Standard Amyloid Criteria Provided by the Fuzzy Oil Drop (FOD) Model

DAWID DUŁAK, MAŁGORZATA GADZAŁA, MATEUSZ BANACH, MAGDALENA PTAK,
ZDZISŁAW WIŚNIOWSKI, LESZEK KONIECZNY, IRENA ROTERMAN

International Journal of
Molecular Sciences

MDPI

Article

Filamentous Aggregates of Tau Proteins Fulfil Standard Amyloid Criteria Provided by the Fuzzy Oil Drop (FOD) Model

Dawid Dułak [1], Małgorzata Gadzała [2,†], Mateusz Banach [3], Magdalena Ptak [3,4], Zdzisław Wiśniowski [3], Leszek Konieczny [5] and Irena Roterman [3,*]

1 ABB Business Services Sp. z o.o. ul. Żegańska 1, 04-713 Warszawa, Poland; dawid.dulak@gmail.com
2 ACK—Cyfronet AGH, Nawojki 11, 30-950 Kraków, Poland; m.k.gadzala@gmail.com
3 Department of Bioinformatics and Telemedicine, Medical College, Jagiellonian University, Łazarza 16, 31-530 Kraków, Poland; mateusz.banach@uj.edu.pl (M.B.); magdalena.ptak@uj.edu.pl (M.P.); mywisnio@cyf-kr.edu.pl (Z.W.)
4 Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland
5 Chair of Medical Biochemistry, Medical College, Jagiellonian University, Kopernika 7, 31-034 Kraków, Poland; mbkoniec@cyf-kr.edu.pl
* Correspondence: myroterm@cyf-kr.edu.pl
† Currently address: Schibsted Tech Polska Sp. z o. o. ul. Armii Krajowej 28, 30-150 Kraków, Poland.

Received: 14 July 2018; Accepted: 20 September 2018; Published: 25 September 2018

Abstract: Abnormal filamentous aggregates that are formed by tangled tau protein turn out to be classic amyloid fibrils, meeting all the criteria defined under the fuzzy oil drop model in the context of amyloid characterization. The model recognizes amyloids as linear structures where local hydrophobicity minima and maxima propagate in an alternating manner along the fibril's long axis. This distribution of hydrophobicity differs greatly from the classic monocentric hydrophobic core observed in globular proteins. Rather than becoming a globule, the amyloid instead forms a ribbonlike (or cylindrical) structure.

Keywords: tau amyloid; Alzheimer's disease; tauopathy

1. Introduction

The origin of amyloid transformation has attracted scientific attention for more than 35 years—at least since being acknowledged as the cause of various neurodegenerative disorders [1]. The coexistence and mutual relations between Aβ amyloids and tau tangles, resulting in the damage and destruction of synapses, is believed to provoke behavioral changes that are associated with cognitive impairment [2,3]. The appearance of amyloid fibrils is the consequence of plasticity of proteins, which can adopt different conformational states [4]. The proteins of high content of intrinsically disordered structural forms seem to be the candidates ready for partially folded state which may transform to disordered aggregates with low packing [5,6].

Emergence of fibrillary structures is also thought of as the result of involvement of intrinsically disordered proteins, especially at early phases of the folding process [7].

Reaching the form of highly packed structuralised aggregates that are based mainly on β-structural forms opens the possibility for the unlimited elongation of highly packed ordered amyloid form [8]. The presence of Beta-structural form (cross Beta) allows for the propagation due to the possible H-bonds system to be organised on both sites.

The specificity of tau amyloidosis, as evidenced by abnormal phosphorylation, results not only in disorganization of microtubulin, but also in the appearance of intracellular tau filaments, referred to as neurofibrillary tangles [9]. Tauopathies are defined as clinically, morphologically, and biochemically heterogeneous neurodegenerative diseases that are characterized by the deposition of abnormal tau protein in the brain [10].

Tau amyloid fibrils are regarded as peculiar due to the existence of two distinct superfibrillary forms: straight filament and paired helical filament [11]. Two individual protofilaments may form different structures depending on their mutual arrangement in the dimer. The authors of [11] refer to the conformation of the protofilament as "C-shaped". The dimer (superfibril), when analyzed under cryo-electron microscopy (cryo-EM) imaging [12], resembles two arched C-shaped structures that are bound back to back. In one form, the structure is symmetrical (with the same residues in both unit molecules involved in complexation), while the other form lacks such symmetry (in this case, the central fragment of one chain makes contact with a fragment which is somewhat closer to the N terminus of the adjacent chain).

According to the analysis shown in [11], the dimerization of protofilaments occurs by way of hydrogen bonds forming between adjacent fragments. In its native form, tau in complex with a microtubule adopts a conformation referred to as "natively unfolded". This conformation is highly resistant to spikes in temperature and acidity [13].

The presented here analysis focuses on microtubule-associated tau neurofibrillary tangle protein, paired helical filaments (consisting of two individual fibrils), individual protofibrils, as well as individual chains.

Identification and characterization of amyloid structures, as shown in this work, bases on comparative analysis of the structure of globular and fibrillary proteins. The proposed fibrillarization model for globular structures is also discussed in [14,15]. At the core of the fuzzy oil drop (FOD) model lies the concept of applying a three-dimensional (3D) Gaussian form to express the idealized distribution of hydrophobicity in a globular protein. Such distribution has a distinct peak at the geometric center of the globule and then falls off along with distance from the center, reaching almost zero at the surface (i.e., at a distance of 3σ from the center). We have identified proteins whose structure closely corresponds to this theoretical distribution [16]. Any local deviations are usually associated with the protein's intended function: local excess of hydrophobicity, if occurring on the surface, usually marks a complexation interface, while the hydrophobicity deficiencies tend to correspond to binding cavities, which are capable of housing ligands (or substrates, in the case of enzymes) [17,18]. The universality and ubiquity of hydrophobic cores is attested to by a study of a large number of proteins with varying secondary and supersecondary structural characteristics [19]. A specific type of discordance vs. the idealized (monocentric) distribution of hydrophobicity is observed in the case of proteins which contain so-called solenoid fragments, including some antifreeze proteins [20]. Such fragments deviate from a centralized core in favor of a distribution comprising alternating bands of high and low hydrophobicity, propagating along solenoids long axis [20]. However, in addition to solenoids, biologically active proteins also include fragments whose purpose is to restrict unchecked propagation of such linear sequences (thereby preventing dimerization or polymerization), as well as to ensure solubility. The special "caps" are identified in such proteins [20]. Structures that strongly deviate from the monocentric distribution and lack suitable "caps" are prone to amyloid filaments [14,15]. In the context of the FOD model, such structures can be likened to ribbonlike micelles, capable of unrestricted propagation [21].

A thorough description of the FOD model can be found in [22].

In the FOD model, the emergent structure of the hydrophobic core is though to result from interactions between the protein and the aqueous solvent, which directs hydrophobic residues towards the center of the globule and favors exposure of hydrophilic residues on the surface. As already indicated, local discordances are often associated with the presence of external factors. The fact that amyloid forms do not require mutations to emerge suggests that misfolding is not caused by

factors internal to the polypeptide itself. One shall mention the mutation-related amyloidosis [23], however the prion-based amyloid transformation does not require the presence of mutation, as it is the discussed case.

Shaking is known to promote amyloidogenesis—and it can hardly be called a chemical factor. Many factors—including chemical ones—were identified to support the amyloidosis transformation [24]. However other factors than environmental (shaking in particular) are not the object of our analysis. One shall also take into account that the folding as well the misfolding processes take place in macromolecular crowding conditions, however the immanent presence of water makes the water environment of high importance [25].

Perhaps shaking disrupts the structure of the solvent in such a way as to prevent it from guiding "natural" conformational changes within the protein chain. Alternatively, shaking is notable for aerating the solvent. The resulting increase in the area of the liquid/gas interface may produce structural changes within the solvent itself.

In addition to analysis of the tau amyloid, as listed in Protein Data Bank [26], this work proposes an in silico experiment, which involves determining alternative structures that the tau amyloid sequence may attain (using specialized protein folding software, such as Robetta [27,28] and I-Tasser [29,30]), and performing folding simulations based on the FOD model. It turns out that the sequence is indeed capable of producing a globular form with a single, monocentric hydrophobic core. Subjecting globular structures to FOD characterization enables us to track changes that result in amyloidogenesis. The work focuses on three distinct structures: (1) the superfibril (seeking the causes behind its structural variability); (2) the protofibril (identifying the characteristic properties of amyloid structures); and (3) a single chain participating in the fibril. Our research is based on observations rooted in the FOD model, specifically, the linear propagation of hydrophobicity in amyloids (which prevents a shared hydrophobic core from forming). As discussed in [14,15], the presence of alternating bands of high and low hydrophobicity can be regarded as one of the principal indicators of amyloid transformation.

2. Results

2.1. Abbreviations Used

FOD—Fuzzy Oil Drop model

RD—Relative Distance—The divergence entropy introduced by Kullback and Leibler (described in Methods) used to express the distance between two compared profiles is of entropy category thus it requires an introduction of reference distribution. This is why the distance between T-O (T theoretical—idealized distribution and O-observed distribution) measured by divergence entropy is compared with the O-R (O-observed, R-uniform distribution deprived on any form of hydrophobicity concentration), also measured by divergence entropy. The parameter expressing the relative distance O | T versus (O | T + O | R) measures the closeness of O distribution versus T distribution in respect to O versus R distribution. The RD parameter becomes polypeptide chain length independent. It makes possible comparison of different proteins.

RD (T-O-R)—RD parameter calculated for two reference distributions T-theoretical and R-uniform

RD (T-O-H)—RD parameter calculated in respect to reference distribution called H-distribution based on intrinsic hydrophobicity of amino acids present in particular polypeptide chain fragment

HvT—correlation coefficient expressing the relation between H-intrinsic hydrophobicity of amino acids versus the T-theoretical (expected) hydrophobicity for the idealized status of the residue

HvO—correlation coefficient expressing relation between H-intrinsic hydrophobicity of amino acid versus is status as observed in particular protein

TvO—correlation coefficient expressing relation between T-idealized hydrophobicity and O-observed in protein under consideration

phf-tau—paired helical filament-tau

phf-tauO—paired helical filament-tau—as it is available in 5O3O phf-tau in symmetrical form of superfibril
phf-tauT—paired helical filament-tau—as it is available in 5O3T phf-tau in asymmetrical form of superfibril
phf-tauL—paired helical filament-tau—as it is available in 5O3L phf-tau in form of superfibril similar to phf-tauO
IT-#—identification of the model constructed using I-Tasser program with number 1–5 since 5 models were constructed using this program
ROB-#—identification of the model constructed using Robetta program with number 1–5 since 5 models were constructed using this program
FOD-#—identification of the model constructed using FOD model with number 1–5 since 5 models were constructed using this program
Tau (267–312)—fragment of tau peptide—protein under PDB ID 2MZ7
Tpp—tau phosphothreonine peptide—protein under PDB ID 1I8H
Tau (306–311A)—fragment of tau to identify the structure available in PDB as 2ON9
Tau (306–310)—fragment of tau to identify the structure available in PDB as 3Q9G
Tau (306–311B)—fragment of tau to identify the structure available in PDB as 3OVL
Tau (305–311)—fragment of tau to identify the structure available in PDB as 4E0M
Tau (623–628)—fragment of tau to identify the structure available in PDB as 4NP8
Tau (306–311C)—fragment of tau to identify the structure available in PDB as 5K7N
F-actin—actin, alpha skeletal muscle as available in PDB as 3J8I
PDB—Protein Data Bank
CASP—Critical Assessment Protein Structure Prediction
BLAST—Basic Local Alignment Search Tool
PSI-BLAST—Position-Specific Iterated BLAST
MSA—Multiple Sequence Alignment

2.2. Superfibril

This analysis concerns the amyloid form that is listed in PDB as 5O3O, 5O3L, and 5O3T (pronase-treated paired helical filament in Alzheimer's disease brain neurofibrillary tangle protein, paired helical filament-tau, phf-tau, *Homo sapiens*). Fragment: residues 623–695 of tau protein (306–378 according to PDB numbering) Chains A, C, E, G, I, along with their counterparts (B, D, F, H, J) make up the proto-fibrils [11]. In order to characterize individual chains in the context of the superfibril and proto-fibrils, we have singled out chains E and F. These two chains are located in the central part of the fibril and can be regarded as representative of an arbitrarily long structure. This selection also minimizes edge effects caused by the finite width of the complex.

Properties of Superfibrils and Interfaces—What Is the Source of Different Isoforms of Tau Filaments?

Table 1 presents the status of tau amyloid structures in terms of RD values, revealing large discordances between T and O profiles in both models (T-O-R and T-O-H). This means that the distribution of hydrophobicity does not involve a central hydrophobic core. Further analysis will reveal that the amyloids are dominated by a pattern that consists of alternating bands of high and low hydrophobicity. High values of RD further indicate that the folding process is driven by the intrinsic properties of each residue rather than by a global force field—this is also typical for amyloids [14,15]. Regarding the hydrophobicity profile correlation coefficients, HvT and TvO lag behind HvO. This is also due to the absence of a central hydrophobic core, which is replaced by linear propagation of narrow "bands". In further sections we will specifically describe locations that exhibit these properties.

Visual comparison of T and O (Figure 1) highlights the major differences between these distributions. It should be noted that the chart consists of many overlapping profiles, which means that the distribution of local minima and maxima is replicated in each adjacent chain, resulting in a set of narrow bands, as suggested above.

The FOD model may also be used to predict the properties of shared hydrophobic cores in protein complexes [31]. In order to properly characterize a given complex, it is important to assess the status of its interface. With regard to proto-fibrils, the distribution that was observed in phf-tauT differs from those exhibited by the remaining structures. However the difference is limited only to the structure of interface, which is discussed in this paper. In phf-tauO and phf-tauL the status is similar and it suggests that the superfibril emerges as a result of factors consistent with the FOD model, i.e., under the influence of the aqueous solvent. This interpretation is supported by the high values of all correlation coefficients. We may conclude that the interface is shaped by all factors which determine the structure of the complex itself, with major involvement of water.

The picture changes, however, when dealing with phf-tauT. Its high value of RD (T-O-R), coupled with negative values of HvT and TvO coefficients and a high value of the HvO coefficient, suggest that, in this case, the solvent does not play a significant role in complexation.

It should be noted that the status of the interface is computed by taking into account all interface residues in the entire fibril (following protein-protein contact distance criteria of PDBsum [32]). When all three correlation coefficients adopt strongly positive values, we may assume that the structure of the interface represents a compromise between all three hydrophobicity profiles (observed, intrinsic and theoretical). In contrast, negative values of HvT and TvO are understood to mean that the interface folds "in spite of" the FOD model and in consequence in spite of environmental effects that act upon the protofibril complex.

Table 1. Relative Distance (RD) values and correlation coefficients computed for the superfibril, proto-fibrils and individual chains treated as components of the superfibril.

Phf-tau	RD		Correlation Coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
phf-tauO	0.745	0.687	0.012	0.259	0.675
phf-tauL	0.731	0.669	0.013	0.296	0.646
phf-tauT	0.724	0.641	0.008	0.301	0.716
Inter-fibril interface					
phf-tauO	0.401	0.550	0.666	0.754	0.950
phf-tauL	0.368	0.587	0.696	0.772	0.929
phf-tauT	0.527	0.386	−0.164	−0.003	0.850
Chains in superfibril					
Phf-tauO					
Chain E	0.741	0.700	0.011	0.237	0.731
Chain F	0.761	0.722	0.010	0.224	0.731
Chains E and F	0.752	0.711	0.013	0.230	0.731
Phf-tauL					
Chain E	0.728	0.675	0.017	0.274	0.696
Chain F	0.747	0.698	0.012	0.265	0.696
Chains E and F	0.741	0.690	0.014	0.268	0.606
Phf-tauT					
Chain E	0.732	0.666	0.034	0.400	0.769
Chain F	0.728	0.666	−0.018	0.136	0.785
Chain E and F	0.730	0.666	0.008	0.273	0.777

The characterization concerns chains E and F, which are located centrally and therefore representative of an unrestricted fibril. The calculated values are typical for amyloid forms, and include high values of RD and HvO, along with very low (sometimes even negative) values of HvT and TvO. In phf-tauT, both individual chains, as well as the interface fragment, are shaped by intrinsic hydrophobicity rather than by the external environment, which would favor the formation of a monocentric hydrophobic core.

As the status of individual chains (in the context of the superfibril) is largely similar in all structures, we will limit their presentation to Phf-tau in form, as observed in phf-tauO and phf-tauT (Table 2).

Table 2. Status of individual chains treated as components of the superfibril. The presentation of phf-tauL is omitted since due to its similarity to phf-tauO.

Chain	RD				Correlation Coefficient					
	T-O-R		T-O-H		HvT		TvO		HvO	
	Phf-tauO	Phf-tauT	Phf-tauO	Phf-tauT	Phf-tauO	Phf-tauT	Phf-tauO	Phf-tauT	Phf-tauO	Phf-tauT
A	0.780	0.801	0.696	0.681	0.010	0.041	0.154	0.441	0.669	0.710
B	0.787	0.739	0.711	0.600	0.031	−0.022	0.303	0.115	0.636	0.723
C	0.756	0.766	0.715	0.705	0.014	0.038	0.202	0.394	0.731	0.772
D	0.779	0.728	0.741	0.665	0.022	−0.020	0.248	0.114	0.730	0.788
E	0.741	0.732	0.700	0.666	0.016	0.034	0.237	0.400	0.731	0.769
F	0.761	0.728	0.722	0.666	0.010	−0.010	0.224	0.136	0.731	0.785
G	0.733	0.700	0.690	0.628	0.016	0.028	0.265	0.394	0.731	0.770
H	0.745	0.733	0.704	0.670	0.003	−0.017	0.192	0.157	0.732	0.786
I	0.727	0.660	0.613	0.522	0.015	0.022	0.343	0.405	0.652	0.686
J	0.750	0.738	0.634	0.618	−0.019	−0.014	0.137	0.192	0.691	0.707

When analyzing individual chains as components of the superfibril, we arrive at similar RD values. In contrast, when the same chains are analyzed as components of individual proto-fibrils (phf-tauT), their values differ due to differences in the orientation of each proto-fibril. Negative values of correlation coefficients for HvT and TvO with a high value of correlation coefficient for HvO relation suggest that phf-tauT is a typical amyloid form.

Figure 1 provides a graphical representation of the superfibril and both chains (E and F) treated as components of the superfibril.

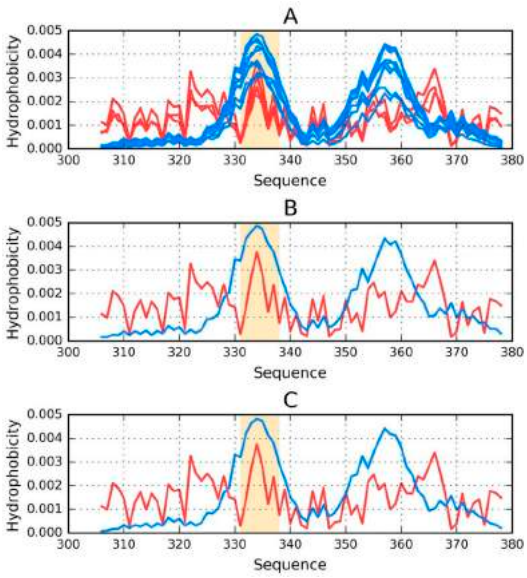


Figure 1. Theoretical (T—blue) and observed (O—red) distributions of hydrophobicity in: (**A**) the phf-tauO superfibril; (**B**,**C**) chains E and F as components of the phf-tauO superfibril. Orange highlighting marks residues which comprise the inter-fibril interface area (331–338).

The status of phf-tauO superfibril is visualized in Figure 1A, which shows that the theoretical distribution involves two local maxima, along with hydrophilic fragments that are exposed on the surface. Neither maximum is evident in the observed distribution, however O includes other local maxima, located in areas where low hydrophobicity is expected. It should be noted that each of these local maxima (as well as minima) represents an entire band stretching along the fibril's long axis. The overlap is due to the repeating pattern that is present in each individual chain, with only the outlying chains exhibiting slightly lower hydrophobicity. On the other hand, the differences between the theoretical distributions are readily apparent since this distribution predicts that hydrophobicity should decrease along with distance from the center. The degree of discordance between T and O can be analyzed by comparing theoretical charts with the observed distributions for chains E and F (which are centrally located and therefore representative for the entire fibril—see Figure 1B,C).

The interface fragment appears to be consistent with the FOD model. Given the central location of the interface, a hydrophobicity peak is expected and—to a certain extent—present in the actual complex. Comparing O with T reveals that two outlying residues exhibit relatively low hydrophobicity, while the central section corresponds to a major spike. Consequently, we rate the interface fragment as being accordant with the model.

A characteristic feature of amyloids is the presence of numerous local maxima in areas where low hydrophobicity (and vice versa) is predicted by the theoretical model. However, it is important to remember that, unlike globular proteins (which may also exhibit this phenomenon), the complexed chains form here bands which stretch along the entire long axis of the fibril. These observations are confirmed by analysis of T and O for chains E and F (treated as components of the superfibril). The discordance between T and O distribution in most of proteins is of local character.

2.3. Properties of Proto-Fibrils

Table 3 presents the hydrophobicity parameters for each proto-fibril. In this case, each proto-fibril is treated as a distinct structural unit. This means that a separate Gaussian is constructed for each proto-fibril (in the preceding section, a shared Gaussian form was computed for the entire superfibril). Results are indicative of an amyloid form: high values of RD and HvO along with very low (even negative) values of HvT and TvO. The correlation coefficients reveal that the structure is dominated by the intrinsic properties of its component residues—an observation that is supported by the observed high values of RD in both models (T-O-R and T-O-H). Thus, the observed distribution is more closely aligned with R (or H) rather than T.

The status of chains E and F, treated as components of their respective proto-fibrils, confirms that they adopt amyloid-like forms, although this effect is less pronounced than in the case of the superfibril (lower values of T-O-H RD and HvO—see Table 3).

Table 3. Status of individual chains treated as components of their respective proto-fibrils. The table lists only values obtained for the proto-fibril (chains A, C, E, G and I) (Differences with regard to the other proto-fibril are negligible.

PDB ID	RD		Correlation Coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
Phf-tauO	0.661	0.607	−0.012	0.089	0.772
Chain E	0.679	0.430	−0.027	0.095	0.548
Chain F	0.679	0.439	−0.027	0.095	0.548
Phf-tauL	0.664	0.595	−0.022	0.082	0.767
Chain E	0.674	0.410	−0.039	0.091	0.545
Chain F	0.674	0.410	−0.039	0.091	0.545
Phf-tauT	0.673	0.602	−0.017	0.118	0.773
Chain E	0.683	0.415	−0.033	0.098	0.550
Chain F	0.683	0.415	−0.033	0.096	0.551

Figure 2 provides a visual representation of these results, showing T and O distributions for one proto-fibril (chains A, C, E, G and I) and the status of the E chain within this structural unit.

An asymmetrical distribution of local maxima is observed in the proto-fibril as a result of significant displacement of the system's central point as compared to the superfibril. Numerous local maxima are present in areas where low hydrophobicity is expected. The involvement of a local maximum in the interface fragment indicates that complexation of protofibrils is generated as the effect of the influence of environment (according to FOD model).

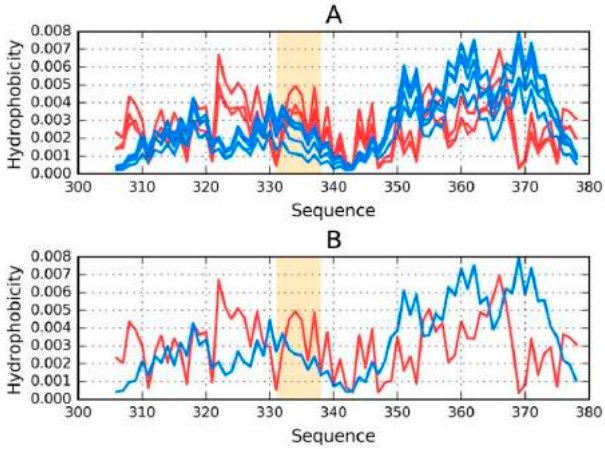


Figure 2. Theoretical (T—blue) and observed (O—red) distributions: (**A**) as calculated for the proto-fibril (chains A, C, E, G and I) listed under phf-tauO; (**B**) for the E chain treated as a component of the proto-fibril. Orange highlighting marks residues which comprise the inter-fibril interface (331–338).

2.4. Properties of Individual Chains Treated as Distinct Structural Units

Our analysis also covers individual chains that are treated as distinct structural units, with a separate Gaussian being plotted for each chain (under the assumption that each chain folds in separation from other chains). To determine the causes of the discordance between the observed and theoretical distributions, we have singled out fragments for which this discordance is particularly evident. Note that we are not dealing with isolated deviations—in many areas, both distributions strongly oppose each other, indicating that the chain does not produce a globule and is likely insoluble due to the lack of a polar surface.

Figure 3 illustrates the status of chains E and F treated as individual structural units.

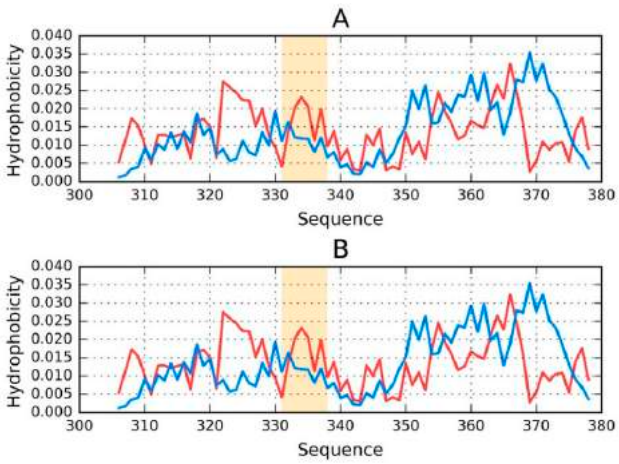


Figure 3. Theoretical (T—blue) and observed (O—red) distributions for phf-tauO: (**A**) chain E treated as a distinct structural unit; and (**B**) chain F treated as a distinct structural unit. Orange highlighting marks residues which comprise the inter-fibril interface (331–338).

It is clear that even when analyzed as distinct units, the discussed chains still diverge from the theoretical distribution of hydrophobicity (Figure 3). No C-terminal maxima (predicted by T) are present in the observed distributions.

Regarding phf-tauL, both of the distributions are similar to those calculated for phf-tauO, with only the interface being somewhat different. The observed distribution, while discordant, does not resemble an amyloid (which would appear as a sinusoidal pattern consisting of similar local maxima).

Taking into account the discussed distributions, it is easy to pinpoint fragments where O deviates from T (see Figure 4).

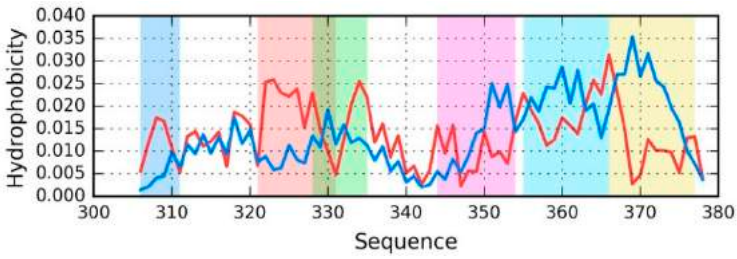


Figure 4. Theoretical (T—blue) and observed (O—red) distribution for the chain E from phf-tauO (the distribution is similar in all structural units). Fragments where O significantly deviates from T are highlighted by different colors (see Table 4) that match the colors of three-dimensional (3D) presentations on Figure 5.

Figure 4 provides a sample set of distributions (T and O) for a single chain—E from phf-tauO. As already noted, the chains differ in detail, while the overall pattern remains largely identical, regardless of the structural unit in question (superfibril, protofibril, or individual chain). The highlighted fragments have been singled out on the basis of visual inspection, supplemented with correlation coefficients computed for successive five-residue segments. Fragments for which HvT and TvO are negative while HvO assumes that a large value will be subjected to further analysis.

Table 4 summarizes the results obtained for all structural units in phf-tauO.

As shown in Table 4, the status of selected fragments is quite similar, regardless of the structural unit in question—in all cases, these fragments are strongly discordant vs. the theoretical distribution.

Table 4. RD values in both models (T-O-R and T-O-H), along with HvT, TvO, and HvO correlation coefficients for chains E and F analyzed as part of the superfibril, as part of a protofibril and on their own. Figure 4 illustrates the division of the chain into individual fragments.

Phf-tauO					
Chain E—Superfibril					
Fragment	RD		Correlation Coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
1–6	0.521	0.405	0.207	0.270	0.855
16–26	0.820	0.771	−0.126	−0.491	0.874
23–30	0.496	0.495	−0.350	0.435	0.580
39–49	0.575	0.544	0.040	0.147	0.971
50–61	0.851	0.802	−0.106	−0.782	0.446
61–72	0.563	0.491	−0.157	−0.024	0.700
CHAIN	0.741	0.700	0.016	0.237	0.731

Table 4. *Cont.*

Chain F—Superfibril					
Fragment	**RD**		**Correlation Coefficient**		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.627	0.512	0.038	0.117	0.875
16–26	0.822	0.773	−0.129	−0.495	0.875
23–30	0.496	0.495	−0.318	0.465	0.580
39–49	0.563	0.533	0.054	0.153	0.971
50–61	0.861	0.814	−0.112	−0.720	0.444
61–72	0.583	0.512	−0.153	0.006	0.699
CHAIN	0.761	0.722	0.010	0.224	0.731
Chain E—Protofibril					
Fragment	**RD**		**Correlation Coefficient**		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.668	0.553	−0.006	0.064	0.875
16–26	0.615	0.557	0.288	−0.141	0.875
23–30	0.571	0.573	0.146	−0.151	0.810
39–49	0.657	0.623	−0.139	−0.110	0.971
50–61	0.700	0.622	0.016	−0.558	0.443
61–72	0.613	0.540	−0.221	−0.200	0.700
CHAIN	0.661	0.607	−0.012	0.089	0.772
Chain F—Protofibril					
Fragment	**RD**		**Correlation Coefficient**		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.668	0.553	−0.005	0.064	0.875
16–26	0.615	0.558	0.286	−0.144	0.875
23–30	0.572	0.573	0.143	−0.155	0.809
39–49	0.657	0.623	−0.139	−0.110	0.971
50–61	0.700	0.621	0.016	−0.558	0.442
61–72	0.612	0.539	−0.221	−0.200	0.698
CHAIN	0.661	0.607	−0.012	0.089	0.772
Chain E—Individual					
Fragment	**RD**		**Correlation Coefficient**		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.741	0.372	−0.093	−0.037	0.543
16–26	0.676	0.433	0.286	−0.387	0.696
23–30	0.582	0.387	0.117	−0.228	0.620
39–49	0.687	0.388	−0.114	0.017	0.888
50–61	0.700	0.489	0.057	−0.537	0.116
61–72	0.656	0.398	−0.237	−0.240	0.507
CHAIN	0.679	0.430	−0.027	0.095	0.548
Chain F—Individual					
Fragment	**RD**		**Correlation Coefficient**		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.741	0.372	−0.093	−0.037	0.543
16–26	0.676	0.433	0.286	−0.387	0.696
23–30	0.582	0.387	0.117	−0.228	0.620
39–49	0.687	0.388	−0.114	0.017	0.888
50–61	0.700	0.489	0.057	−0.537	0.115
61–72	0.656	0.397	−0.233	−0.240	0.507
CHAIN	0.679	0.430	−0.027	0.095	0.548

2.5. Comparative Analysis Involving Theoretical Models

As previously noted, we have carried out an in silico experiment that consisted of predicting the conformation of a tau protein whose sequence matches the discussed amyloids. Our analysis concerned the entire molecule as well as fragments that are highlighted in Figure 4.

Figure 5 presents 3D models of tau polypeptides obtained using software packaged described in the Materials and Methods section. Visual inspection reveals the possible emergence of globular forms: I-Tasser produces four such structures (out of five input cases), while Robetta produces one (out of five). The tendency of the FOD model to produce globular forms should come as no surprise given the model's propensity to direct hydrophobic residues towards the center of the molecule (due to interactions with the aqueous solvent).

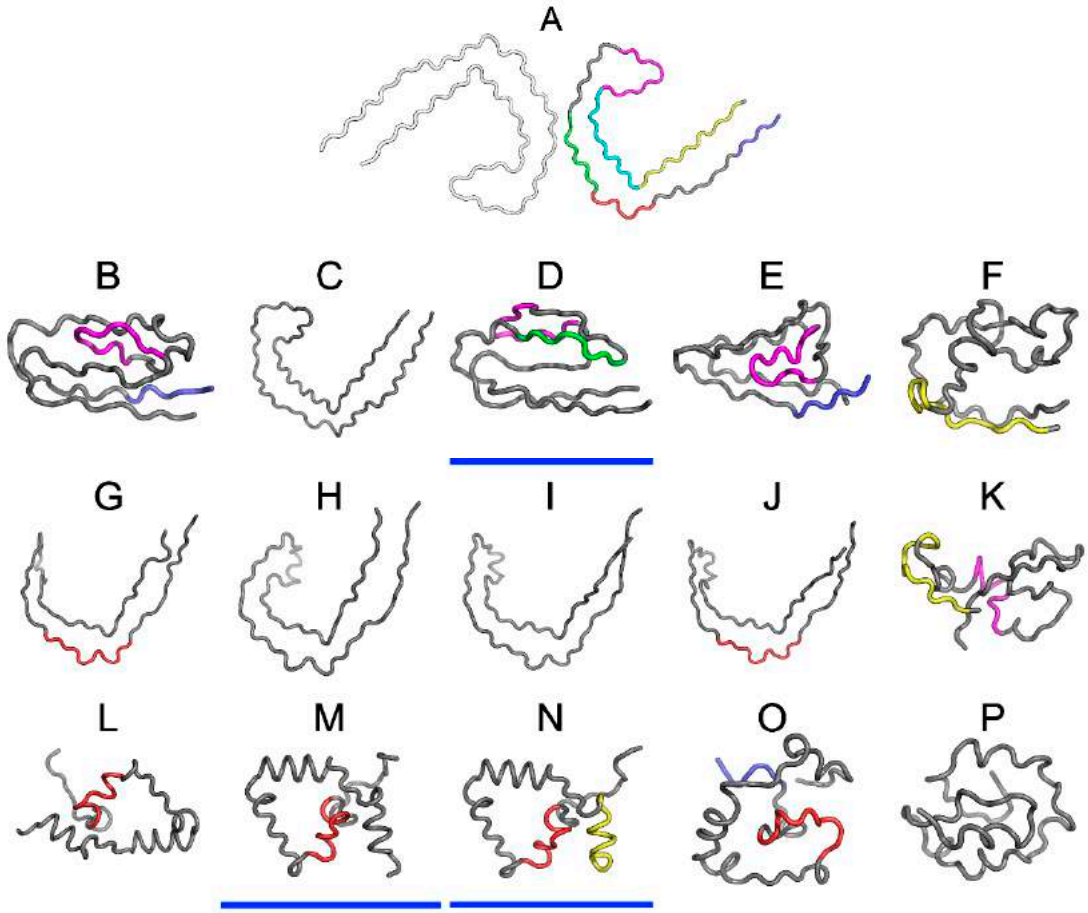


Figure 5. 3D presentation of structures obtained using I-Tasser upper row (**B**–**F**), Robetta central row (**G**–**K**) and the fuzzy oil drop (FOD) model bottom row (**L**–**P**). The top structure (**A**) is the structure phf-tauO as appears in PDB database. The colors used correspond to highlights on Figure 4. Models whose status is consistent with FOD predictions (RD < 0.5) have been additionally underlined in blue (**D**,**M**,**N**).

One of the models that was produced by I-Tasser appears to involve a hydrophobic core (in the sense of the FOD model—cf. underlined structures in Figure 5). None of the models produced by Robetta satisfies this criterion. Regarding the FOD model, despite its natural tendency to generate hydrophobic cores, only two among 500 structures analyzed in the course of the study contain a hydrophobic core (i.e., satisfy the RD < 0.5 condition).

Analysis of numerical values that are listed in Table 5, along with visual inspection of 3D forms reveals that some of these fragments adopt structures consistent with the Gaussian distribution. Of particular note is the fragment marked in red on Figure 5 (residues 50–61, numbered 356–357

according to PDB), which does not conform to theoretical predictions in any model. In interpreting this fact, we may refer to the dominant role of the fragment whose sequence does not adapt to the centralized distribution of hydrophobicity. We can speculate that this fragment (GSLDNITHVPGG) is therefore the most amyloidogenic sequence in the set under consideration. This suggestion is supported by the data shown in Table 5, particularly RD values that are either highest or second highest in the entire set.

Table 5. RD values obtained using I-Tasser, Robetta and FOD. In addition to RD values (for T-O-R and T-O-H) the table also lists HvT, TvO and HvO correlation coefficients. The division into fragments is consistent with the one shown in Figure 4. Values in bold distinguish the status accordant with FOD model (RD < 0.5).

I-Tasser					
IT-1	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.411	0.142	0.032	0.600	0.469
16–26	0.509	0.200	0.308	0.659	0.706
23–30	0.641	0.357	0.308	0.219	0.415
39–49	0.494	0.283	0.501	0.306	0.841
50–61	0.614	0.618	0.169	0.196	0.466
61–72	0.627	0.232	0.089	0.176	0.559
CHAIN	0.519	0.266	0.200	0.473	0.574
IT-2	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.825	0.138	−0.016	0.551	0.242
16–26	0.607	0.279	0.372	0.039	0.730
23–30	0.503	0.236	0.336	0.226	0.651
39–49	0.765	0.313	−0.230	−0.358	0.849
50–61	0.667	0.548	0.027	−0.416	0.122
61–72	0.605	0.367	−0.486	−0.133	0.475
CHAIN	0.635	0.353	−0.028	0.172	0.519
IT-3	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.729	0.258	−0.438	−0.109	0.416
16–26	0.558	0.179	0.381	0.567	0.830
23–30	0.408	0.173	0.607	0.635	0.805
39–49	0.369	0.191	0.282	0.659	0.828
50–61	0.701	0.702	0.102	−0.196	0.587
61–72	0.545	0.285	0.212	0.162	0.638
CHAIN	**0.478**	0.262	0.270	0.524	0.684
IT-4	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.407	0.183	0.216	0.702	0.554
16–26	0.817	0.230	0.191	0.125	0.730
23–30	0.746	0.468	0.253	−0.453	0.345
39–49	0.432	0.140	0.190	0.494	0.818
50–61	0.630	0.585	0.426	0.094	0.367
61–72	0.601	0.350	0.412	0.193	0.781
CHAIN	0.605	0.315	0.229	0.252	0.581

Table 5. *Cont.*

IT-5	RD		Correlation Coefficient		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.545	0.138	−0.024	0.198	0.590
16–26	0.537	0.232	0.321	0.313	0.840
23–30	0.663	0.323	0.254	0.040	0.670
39–49	0.504	0.162	0.444	0.426	0.848
50–61	0.723	0.862	0.194	0.208	0.872
61–72	0.477	0.257	0.501	0.524	0.858
CHAIN	0.513	0.327	0.260	0.418	0.659
Robetta					
ROB-1	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.627	0.474	0.010	0.165	0.633
16–26	0.473	0.380	0.687	0.400	0.808
23–30	0.566	0.619	0.619	0.157	0.830
39–49	0.606	0.381	0.042	0.275	0.889
50–61	0.650	0.495	0.070	−0.105	0.376
61–72	0.675	0.405	−0.27	−0.182	0.594
CHAIN	0.745	0.568	−0.001	0.109	0.683
ROB-2	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.535	0.345	−0.300	0.218	0.690
16–26	0.601	0.518	0.448	0.004	0.809
23–30	0.621	0.675	0.388	−0.147	0.824
39–49	0.839	0.292	0.014	0.036	0.904
50–61	0.681	0.531	0.083	−0.297	0.404
61–72	0.699	0.364	−0.250	0.017	0.552
CHAIN	0.656	0.422	0.032	0.249	0.611
ROB-3	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.600	0.382	−0.495	−0.012	0.623
16–26	0.523	0.410	0.576	0.144	0.795
23–30	0.572	0.510	0.463	−0.038	0.821
39–49	0.842	0.473	−0.173	−0.189	0.955
50–61	0.646	0.512	0.127	−0.168	0.394
61–72	0.698	0.396	−0.182	0.226	0.576
CHAIN	0.682	0.444	0.074	0.351	0.650
ROB-4	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.624	0.519	−0.013	0.141	0.766
16–26	0.441	0.329	0.714	0.464	0.806
23–30	0.593	0.557	0.618	0.171	0.831
39–49	0.782	0.453	−0.120	−0.072	0.939
50–61	0.658	0.575	0.160	−0.187	0.344
61–72	0.678	0.346	−0.187	0.016	0.568
CHAIN	0.748	0.542	0.022	0.154	0.657

Table 5. *Cont.*

ROB-5	RD		Correlation Coefficient		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.545	0.111	−0.045	0.571	0.558
16–26	0.701	0.341	0.201	0.166	0.916
23–30	0.748	0.314	0.251	−0.284	0.666
39–49	0.372	0.220	0.255	0.676	0.809
50–61	0.643	0.712	0.258	0.398	0.640
61–72	0.432	0.155	0.513	0.587	0.794
CHAIN	0.566	0.351	0.272	0.375	0.733
FOD					
FOD-1	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.774	0.288	−0.555	−0.040	0.023
16–26	0.346	0.116	0.238	0.643	0.503
23–30	0.808	0.041	0.510	−0.319	0.339
39–49	0.602	0.187	−0.264	0.267	0.606
50–61	0.733	0.714	0.369	−0.291	0.457
61–72	0.568	0.217	−0.269	0.279	0.200
CHAIN	0.503	0.253	0.126	0.563	0.390
FOD-2	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.502	0.017	−0.687	0.723	−0.013
16–26	0.315	0.110	0.165	0.770	0.527
23–30	0.675	0.261	0.249	0.105	0.521
39–49	0.543	0.271	−0.175	0.258	0.693
50–61	0.593	0.698	0.375	0.092	0.706
61–72	0.522	0.191	−0.211	0.268	0.268
CHAIN	**0.364**	0.191	0.093	0.665	0.410
FOD-3	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.605	0.104	−0.373	0.440	0.386
16–26	0.325	0.107	0.054	0.700	0.552
23–30	0.580	0.194	0.328	0.127	0.574
39–49	0.575	0.187	−0.171	0.313	0.689
50–61	0.676	0.682	0.331	−0.282	0.554
61–72	0.338	0.080	0.135	0.603	0.395
CHAIN	**0.369**	0.175	0.152	0.635	0.414
FOD-4	**RD**		**Correlation Coefficient**		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.492	0.137	−0.349	0.443	0.231
16–26	0.387	0.136	0.224	0.601	0.546
23–30	0.705	0.470	0.301	−0.414	0.290
39–49	0.589	0.140	−0.211	0.453	0.410
50–61	0.840	0.744	−0.151	−0.366	0.252
61–72	0.556	0.319	0.038	0.092	0.357
CHAIN	0.648	0.346	0.059	0.150	0.395

Table 5. *Cont.*

FOD-5	RD		Correlation Coefficient		
Fragment	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1–6	0.620	0.075	−0.363	0.484	0.004
16–26	0.513	0.121	0.025	0.576	0.446
23–30	0.668	0.247	−0.113	0.031	0.070
39–49	0.806	0.205	−0.18	0.220	0.646
50–61	0.807	0.588	−0.458	−0.326	0.033
61–72	0.638	0.344	−0.154	−0.022	0.348
CHAIN	0.644	0.313	−0.049	0.200	0.357

2.6. Other Fragments of Tau Proteins

In order for our comparative analysis to be as comprehensive as possible, we also include tau proteins (or fragments of tau amyloids) that PDB lists as being capable of adopting non-amyloid conformations. The Tpp (1I8H) represents the 541–553 fragment of the previously described tau protein, in complex with a microtubule—specifically, ww domain complexed with human tau phosphothreonine peptide microtubule-associated protein tau. In this complex, the A chain comprises residues 541–553 (so-called phf-tau), while chain B represents the ww domain (6–44) [33].

The Tpp sequence does not fully match phf-tau (and the others), however we have included it in our analysis due to functional similarities.

Results shown in Table 6 and on Figure 6 reveal that the Tpp complex does not conform to the FOD model. The status of chain A (tau), when analyzed on its own, is also discordant. On the other hand, the same chain conforms to the model when analyzed as part of the complex. This means that chain B creates suitable conditions for chain A to produce a shared hydrophobic core that is consistent with the 3D Gaussian.

Table 6. Parameters describing the 541–553 fragment (chain A) of the tau protein Tpp in complex with the ww domain (residues 6–44).

Tpp	RD		Correlation Coefficient		
	T-O-R	**T-O-H**	**HvT**	**TvO**	**HvO**
Complex	0.680	0.546	0.171	0.235	0.700
Chain A (tau) in complex	0.452	0.152	0.365	0.730	0.756
Chain A(tau) individual	0.630	0.240	0.080	0.348	0.753

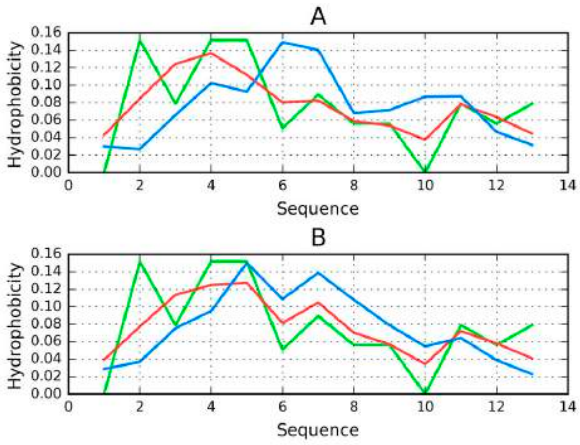


Figure 6. Distribution of hydrophobicity in chain A in the Tpp in complex (T—blue, O—red, H—green): (**A**) when treated as a distinct structural unit; (**B**) when treated as a component of the complex. Good alignment between O, T, and H can be observed in the latter case.

Considering the individual fragments of Tpp, it turns out that the fragment 27–29 of chain B is the most discordant, and that eliminating this fragment from calculations lowers the RD value of the complex. This may indicate the conformational alignment between chain B and chain A (given that the presence of chain A disrupts the distribution of hydrophobicity in chain B).

Analysis of our results indicates the need of a "chaperone", which chain A requires to reach together a conformation consistent with the FOD model.

Tau (267–312) is another protein related to the discussed tau structure. Its sequence matches the short fragment at 306–312 in tau as it is present in phf-tauO. According to [34] this fragment (267–312) of tau protein is bound to microtubules.

The status of the tau chain in tau (267–312) reveals strong discordance with regard to the theoretical model (see Figure 7), with RD = 0.680 (T-O-R) and 0.527 (T-O-H). Correlation coefficients are −0.049, −0.042, and 0.673 for HvT, TvO, and HvO, respectively. These values indicate that the structure of the chain is dominated by the conformational tendencies of individual residues rather than by the external hydrophobic force field. Interaction with microtubules is likely to be the driving force behind conformational adaptation. The structure of the entire complex is not known, however information regarding the interaction of individual residues with the microtubule might explain the discordance that was observed throughout the chain. Only the helical fragment at 295–299 appears accordant with the FOD model (with RD values and correlation coefficients of 0.264, 0.170, 0.876, 0.921, and 0.965, respectively, showing good alignment between the theoretical and observed distribution).

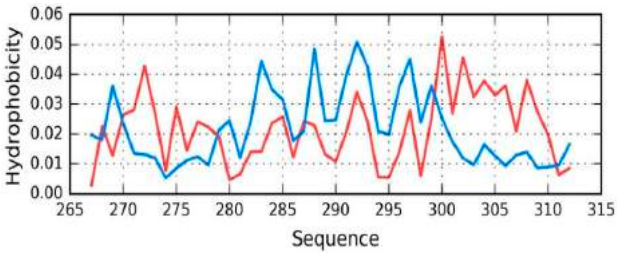


Figure 7. T (blue) and O (red) hydrophobicity distributions in Tau (267–312) (2MZ7), revealing overall strong discordance.

2.7. Peptides

To complete our study of tau-derived structures that are listed in PDB we also need to consider peptides capable of amyloid transformation. The possible mechanism driving this process, discussed in [14,15], remains applicable in the case under consideration.

Peptides that match the tau protein sequence are mostly related to the fragment at 306–311—the short N-terminal fragment of the tau protein present in phf-tauO, phf-tauL, and phf-tauT. A short peptide which does not produce a globular form should not, in principle, be analyzed using the FOD model. Nevertheless, for the sake of completeness, we will present RD (T-O-R) values calculated for such peptides—see Table 7.

Table 7. Properties of peptides which match the N-terminal fragment of the tau protein. The table lists PDB ID, sequences (indicating which fragments are identical to the tau chain) and RD (T-O-R) values. Positions + aa inform about positions of amino acids present in the chain but these residues are not the object of analysis.

Peptide	Sequence	Position	RD
Tau (306–311)	VQIVYK	306–311	0.164
Tau (306–310)	VQIVYK + LA	306–311 + 2 aa	0.332
Tau (306–311B)	VQIVYK	306–311	0.160
Tau (305–311)	AS + VQIVYK + AEFYK	2 aa + 306–311 + 5 aa	0.707
Tau (623–628)	VQIVYK	306–311	0.174
Tau (306–311C)	VQIVYK	306–311	0.170

The values shown in Table 7 only reveal the type of hydrophobicity distribution, with no assessment of the hydrophobic core structure. Low values indicate that hydrophobic residues are located in the central part of the chain, surrounded by N- and C-terminal hydrophilic residues. As shown, the short VQIVYK fragment, despite including an outlying Val residue, is a good match for the centralized hydrophobic core structure. When additional neighboring residues (adjacent to the 306–311 fragment) are included in analysis, the value of RD increases significantly.

Peptides that are identified as capable of amyloidogenesis appear to adopt amyloid-like conformations themselves. As discussed in [14,15], the distribution of hydrophobicity in a peptide may—regardless of its accordance with the theoretical distribution—give rise to amyloid formation, as long as the environment favors linear complexation of additional peptides, with alternating bands of high and low hydrophobicity emerging along the axis of the fibril. This is visualized in Figure 8, which compares two fringe cases (in terms of RD values).

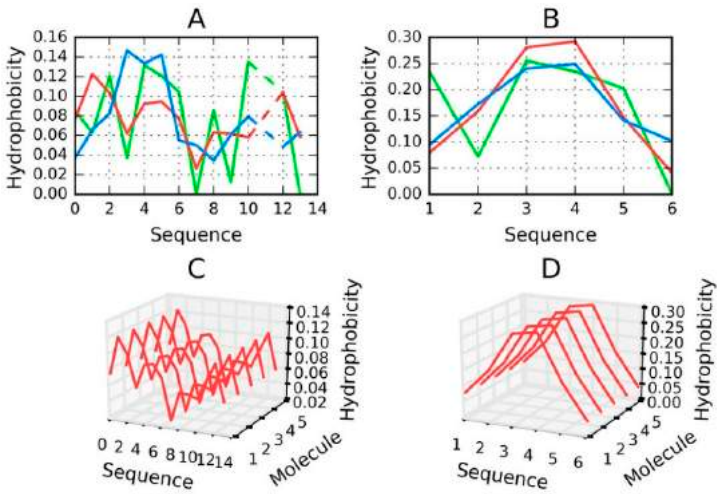


Figure 8. H (green), T (blue) and O (red) distributions for selected peptides: (**A**) Tau (305–311), (**B**) Tau (306–311B); (**C**) for (**A**) and (**D**) for (**B**)—pseudo-3D view, presenting the observed hydrophobicity of a theorized structure of fibrillary tangles formed by linear propagation of the corresponding peptides. Dashed lines on (**A**) represent deleted residue number 11 Tau (305–311) according to data available in PDB.

Figure 8 evidences the appearance of structures that are characterized by linear propagation of hydrophobicity peaks/troughs, which is a precondition of amyloid formation. As highlighted by to-date observations and interpretations, the environment must "support" the creation of such forms. It is thought that under natural conditions the structure of water does not favor the formation of amyloid fibrils.

The value of RD computed for a short peptide implies how many local maxima are present. A low value indicates that hydrophobicity is concentrated in the central part of the peptide (e.g., Tau (306–311B)), while a high value suggests the presence of numerous local maxima (e.g., Tau (305–311)).

2.8. Is It Possible to Differentiate between the Amyloid Fibril and the Fibrillary Structure Present in the Microfilament?

The defining property of amyloids is their fibrillary nature. This phenomenon, however, is not restricted to amyloids. Many biologically active fibrillary proteins exist, often serving as biological scaffolds—this includes polymer microfilaments, such as F-actin (filamentous actin). An example structure of this protein is listed in PDB under ID 3J8I [35].

Analysis of protein 3J8I reveals an alternative approach to a fibrillary structural formation that relies on different mechanisms than those, which drive amyloidogenesis. The PDB structure comprises five monomeric units arranged into a linear complex. Each monomer is a single-domain chain, 375 aa

in length, with varying secondary folds: three beta sheets (11 beta strands in total) and 17 helices. The monomers are spatially arranged in a shape of a helix, forming an elongated fibril. Our analysis focuses on the F chain, which is placed in the center of this fibril. We believe that this chain best represents any inner subunit of a long fibril, with adjacent neighbors on either side.

In analyzing the F chain we apply a twofold approach: first, we treat the chain as an independent structure (constructing a 3D Gaussian capsule calculated specifically for that chain), and subsequently we analyze it as part of the complex (with a broader Gaussian encapsulating the entire complex). The former approach enables us to determine the status of the chain itself, while the latter provides clues regarding its role in the formation of a fibril.

The same observations that we relied on when an analyzing the tau amyloid (high values of RD in both configurations; negative HvT and TvO correlation coefficients and strongly positive HvO) will be sought in our study of F-actin to determine whether the conditions that give rise to amyloid fibrils also apply in the presented case.

The above mentioned set of parameters shows that T not only deviates from O, but can, in some respects, be viewed as its polar opposite. This is taken as evidence that the given conformation is driven by intrinsic hydrophobicity of individual residues.

Table 8 presents FOD parameters that describe F-actin (as listed in PDB under ID 3J8I), which were derived from T and O distributions plotted in Figure 9. It appears that the entire complex, as well as both versions of the F chain, deviate from the theoretical distribution, with no monocentric hydrophobic core being observed in either case. Note that Table 8 only lists the status of selected fragments—those, whose amyloid-like conformation may be important in light of the current discussion regarding identification of amyloid forms.

Results show that local amyloid-like properties may be attributed (in varying degrees) to the beta sheet. Such localized amyloid-like folds can indeed be found in many biologically active proteins (e.g., antifreeze proteins that contain solenoid fragments) [36]. Likewise, the beta sheet found in the lysozyme may also be regarded as amyloid-like [37]. (Note that this particular beta sheet plays an important role given its proximity to the active site—it even contributes one of the catalytic residues of the lysozyme). It seems that the presence of a similar structure in actin is not a unique phenomenon, especially given the structure of its immediate neighborhood. As it turns out, local amyloid-like folds in biologically proteins are typically bracketed by "stop" signals (or "caps"), which prevent unchecked linear propagation. They do so by ensuring that the structure, as a whole, conforms to the theoretical distribution of hydrophobicity and mediating entropically advantageous contact with water. This is highlighted in Table 8 with "stop" annotations.

In the scope of our analysis we also computed FOD correlation coefficients for successive fragments of the input chain while using a 5 aa moving frame. This reveals the exact placement of residues which exhibit amyloid-like characteristics. Eliminating such residues lowers the value of RD (although not below 0.5—see the "No neg CC" annotations in Table 8). Visual inspection of both profiles (theoretical—T and observed—O) reveals residues that contribute to the discordance (these are highlighted in Figure 9 and marked in Figure 10, which presents the protein's 3D structure). It is worth noting that these residues together comprise only 10% of the chain. Eliminating visually inspected residues brings RD down below 0.5, which means that the remainder of the chain conforms to the monocentric distribution of hydrophobicity.

Table 8. Parameters describing the microfilament structure as present in F-actin. β-sheet-1—sheet with starting numbers 8–11, β-sheet-2—sheet with starting numbers 150–154, β-sheet-3—sheet with starting numbers 70–72. Stop sign—fragment "stopping" linear propagation: *—fragment 351–374, **—fragment 325–331, ***—fragment 173–176, No neg CC—status of the chain with residues representing negative correlation coefficients (CC) for HvT and TvO with high HvO, No selected—status of the chain with residues identified as discordant under visual analysis of the T and O profile—shown in Figures 9 and 10.

F-actin	RD		Correlation Coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
Complex	0.783	0.704	0.058	0.081	0.722
Chain F in complex	0.614	0.512	0.111	0.155	0.730
β-sheet-1 *	0.743	0.465	−0.163	−0.411	0.723
β-sheet-2 **	0.593	0.415	−0.08	−0.036	0.635
β-sheet-3 ***	0.532	0.593	−0.398	−0.253	0.939
Stop sign β1	0.445	0.346	0.504	0.420	0.867
No neg CC	0.605	0.467	0.121	0.179	0.717
No selected	0.488	0.391	0.312	0.425	0.748
C-term	0.445	0.346	0.504	0.420	0.867
Chain F individual	0.641	0.544	0.140	0.382	0.709
P-P	0.754	0.594	0.062	0.136	0.687
No P-P	0.635	0.543	0.158	0.357	0.737
β-sheet-1 *	0.774	0.503	−0.161	−0.393	0.721
β-sheet-2 **	0.556	0.402	−0.137	−0.367	0.489
β-sheet-3 ***	0.438	0.507	0.221	0.460	0.940
Stop sign β1	0.492	0.376	0.610	0.414	0.866
No neg CC	0.615	0.502	0.209	0.402	0.714
No selected	0.487	0.400	0.338	0.528	0.729
C-term	0.492	0.376	0.610	0.414	0.866

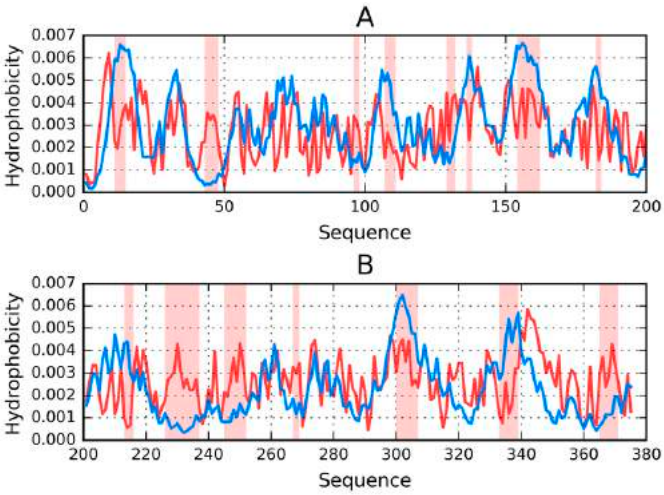


Figure 9. T (blue) and O (red) profiles for the F chain from F-actin (PDB ID 3J8I), divided into two parts for visibility: 1–200 (**A**) and 201–375 (**B**). Highlighted positions mark residues that cause discordance between those distributions (on the basis of visual inspection). The remainder of the chain is regarded as accordant with the theoretical distribution. It likely contributes to the protein's structural stability—under the assumption that a well-ordered hydrophobic core and the presence of disulfide bonds both play a role in stabilizing tertiary conformations.

Figure 10. 3D presentation of F chain from F-actin (PDB ID 3J8I). Beta sheets are displayed in different shades of yellow. Red fragments distinguish residues that cause discordance between T and O distributions. These fragments correspond to highlighted parts of hydrophobicity profiles presented in Figure 9.

The amyloid-like beta sheet-1 is characterized by the linear propagation of alternating bands that differ in terms of hydrophobicity. This effect manifests itself as a strong discordance between T and O profiles, where—in some cases—the observed distribution appears to be a polar opposite of the theoretical distribution.

Linear propagation can be observed by studying the status of successive fragments that comprise the beta sheet. It is therefore interesting to speculate about the participation of such beta sheets in formation of a complex with a clearly fibrillary nature. The sheets in question are dispersed and do not form a continuous band of alternating hydrophobicity. Consequently, they cannot be regarded as a structural scaffold for the complex. What is more, the beta sheets that are contributed by different chains are not in contact (as shown in Figure 11). The presence of "stop" fragments, also shown in Figure 11, which arrest linear propagation, suggests that amyloid-like conditions are intended to remain local and not dominate the structure. Similar "caps" can be found in many other proteins, which include amyloid-like fragments [38] and prevent the unrestricted elongation of such structures.

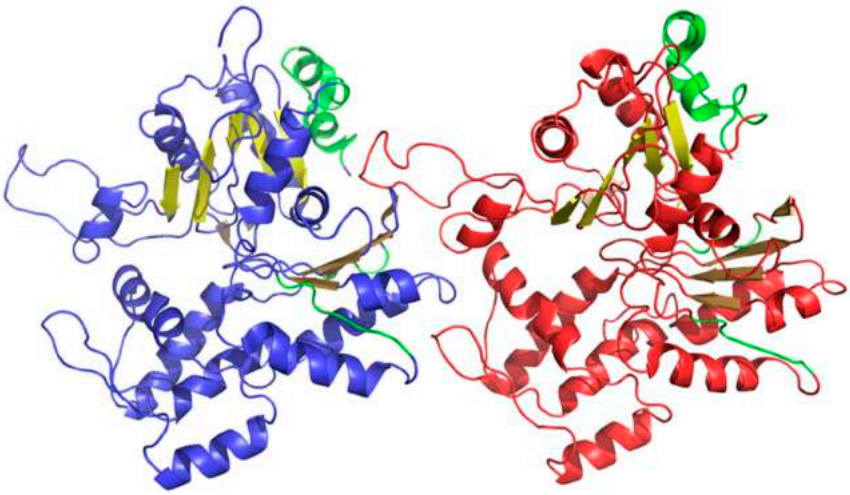


Figure 11. 3D presentation of F chain (blue) and H chain (red) from F-actin (PDB ID 3J8I). Beta sheets are displayed in different shades of yellow (same three shades in each chain). "Stop" fragments are shown in green.

In summarizing our comparative analysis of fibrillary structures, it should be noted that these structures owe their existence to different mechanisms. An amyloid emerges as a consequence of linear propagation of alternating bands of high and low hydrophobicity, whereas globular proteins form complexes via nonbinding interactions (including salt bridges and hydrogen bonds). In the latter case, even when amyloid-like fragments can be found in the proteins' structures, they are dispersed and protected by "stoppers", which prevent them from interacting with one another to form complexes.

In effect, we can state that the structure referred to as a "fibril" might be produced in various ways. Linear propagation of hydrophobicity bands is the prerequisite of amyloid formation (as well as a useful criterion for identifying amyloids), whereas other fibrillary structures (such as F-actin) are formed through nonbinding interactions (including salt bridges and hydrogen bonds). Thus, even though the end result (elongated fibril) is similar, the underlying mechanisms differ. Against this background we propose that the criteria listed in this book differentiate amyloids and enable their identification. We also show that the presence of beta folds is not required (e.g., as evidenced by the tau amyloid). Instead, amyloids may form whenever the folding process is driven by the intrinsic properties of individual residues, as confirmed by the parameters that are studied in this work.

3. Discussion

The comparative analysis of proteins associated with amyloid tau confirms the previously stated hypothesis concerning the structural properties of the amyloid. According to this hypothesis, the amyloid is characterized by the presence of alternating bands of variable hydrophobicity. It seems that linear propagation—which can be regarded as contrary to the emergence of a centralized hydrophobic core (as seen in globular proteins)—is a characteristic property of amyloids. A similar phenomenon can be observed in Aβ amyloids [39,40]. The network of hydrogen bonds that is discussed in numerous studies [11,41] favors this type of conformation and is thought to be associated with the linear properties of beta folds. In the tau amyloid, however, β-strands play a much smaller role than in other known amyloids. This suggests that while hydrogen bonds are important, their role is not necessarily linked to β-structures.

Hydrophobicity is capable of binding together proximate charged residues, however, electrostatic interactions should, in principle, prevent such clustering. Under such conditions only hydrophobic forces can result in the observed arrangement. Thus, a conformation that is driven by intrinsic hydrophobicity (and does not generate a central hydrophobic core) may be regarded as both the cause and the mechanism of amyloid transformation.

The FOD model recognizes several possible forms for the tau superfibril. This diversity is likely caused by interactions between the solvent and the emerging amyloid. We suggest that, while phf-tauO and phf-tauL emerge as the effect of the influence of surrounding water, in phf-tauT, the structure is driven by the specific band-like arrangement of hydrophobicity in the amyloid itself.

The tau protein, whose task is to mediate interaction with microtubules, must align itself to the complexed object. When the protein is subjected to folding on its own, in an independent manner, it may adopt a globular conformation and remain soluble. An open question is why the same protein undergoes complexation in a form which does not resemble a globule. As shown, a chain that is sequentially identical to the amyloid fragment of the tau chain cannot produce a globular structure. In this context, microtubules may be viewed as a "chaperone", which ensures that the protein adopts its intended conformation, required for biological activity.

Conclusions that are related to the process of amyloidogenesis and the role of the FOD model in explaining this process, all point to the need for further research into the properties of the aqueous solvent. While we possess good knowledge of the properties of ice, the corresponding "normal" (or physiological) condition of liquid water is poorly understood—for example, we are still unsure of why the density of water peaks at 4 degrees C. This may explain the recent uptake in investigations that aim to explain such phenomena [41–46]. We believe that these studies may also cast a new light on the process of amyloidogenesis, which—in all likelihood—is associated with the (heretofore unknown)

influence of the force field exerted by the surrounding water. This field should be modeled as a continuum rather than (as is common practice in modern molecular dynamics packages) as a collection of distinct molecules. The FOD model provides a good baseline for such research.

The analysis allows for distinguishing of critical short sequences especially resistant to adopt the conformation accordant with the expected uni-centric hydrophobic distribution. This phenomenon is also observed in other amyloids, especially Aβ(1–42) amyloid [39].

4. Materials and Methods

4.1. Data

The analysis concerns tau protein amyloids listed in PDB as capable of forming highly ordered superfibrils. In addition, we also consider selected fragments of the tau protein, including short peptides. Table 9 gives the full list of structures subjected to analysis.

Table 9. Set of proteins subjected to analysis, along with an indication of chain length and complexation capabilities. The rightmost column provides references.

PDB ID	Characteristics	Length	Complex	Reference
	Tau—amyloid			
5O3O phf-tauO	Microtubule-associated protein tau	73 aa	10 chains	[11]
5O3L phf-tauL	Microtubule-associated protein tau	73 aa	10 chains	[11]
5O3T phf-tauT	Microtubule-associated protein tau	73 aa	10 chains	[11]
	Tau—non-amyloid			
2MZ7 Tau (267–312)	Tau (267–312) bound to microtubules	46 aa		[34]
1I8H Tpp	Pin1 ww domain complexed with human tau phosphothreonine peptide	14 aa	Complex	[33]
	Peptides			
2ON9 Tau (306–311A)	Amyloid forming peptide VQIVYK from the repeat region of tau (in tau 306-311)	6 aa		[47]
3Q9G Tau (306–310)	VQIVY segment from Alzheimer's tau displayed on 42-membered macrocycle scaffold Cyclic pseudo-peptide vqiv(4bf)(orn)(hao)kl(orn)	5 aa		[48]
3OVL Tau (306–311B)	Microtubule-associated protein. VQIVYK (residues 306–311)	6 aa		[49]
4E0M Tau (305–311)	SVQIVYK segment from human tau (305–311) displayed on 54-membered macrocycle scaffold (form i)	7 aa		[50]
4NP8 Tau (623–628)	Structure of an amyloid forming peptide VQIVYK from the second repeat region of tau (alternate polymorph) (in tau 623–628)	7 aa		[51]
5K7N Tau (306–311C)	tau VQIVYK peptide	6 aa		[52]
	Fibrilar form as appears in microfilament			
F-actin	Actin, alpha skeletal muscle	375 aa		[53]

Table 8 includes tau superfibrils (phf-tauL, phf-tauO, phf-tauT), smaller structural units (including individual chains—tau (267–312)), as well as complexes with other proteins (Tpp). We also consider individual peptides that are widely characterized as capable of forming amyloid structures.

All of the above structures are subjected to FOD characterization in the context of the superfibril, the protofibril and the individual chain. Our analysis further extends to peptides whose composition is similar or identical to PDB sequences. The status of such molecules is determined by computing their RD coefficients. It should be noted that seeking proper hydrophobic cores in very short peptides (<15 aa) makes little sense—such peptides are characterized while using FOD criteria only in order to provide a coherent platform for comparative studies. The FOD model provides useful information regarding the relationship between each residue's intrinsic hydrophobicity and its placement in a fully folded chain.

4.2. Folding of Peptides—Components of Amyloid Structures

Peptide sequences which form parts of the tau amyloid (e.g., 306–378, as listed under phf-tauO) have been subjected to folding simulations while using Robetta [26,28] and I-Tasser [29,30], as well as to simulations based on the FOD model [53]. This operation can be regarded as an in silico experiment whose aim is to provide alternatives to structures generated by specialized 3D structure prediction software. Our goal is to identify theoretical opportunities for alternative folds (unlike those listed under phf-tauO and similar entries). The globular forms that are generated by the FOD model may provide clues regarding the discordance between the theoretical distribution of hydrophobicity and the actual location of hydrophobicity maxima/minima. A ranking list of the resulting structures may be composed in order to identify factors that increase similarities between the theorized conformation and the corresponding amyloid form.

Robetta is a software package that is aimed at the modeling and analysis of protein structures [27,28]. It is a strong performer in successive editions of the CASP challenge, which focuses on predicting the 3D conformations of input residue sequences [54]. Robetta works in the following manner: the user is asked to input a sequence of amino acids comprising a given protein chain. This sequence is then subdivided into fragments (called domains) while using the "Ginzu" hierarchical scanning algorithm. The algorithm recognises fragments homologous to sequences for which the preferred secondary conformation has been established on the basis of experimental studies. Such homologous areas are detected by (in the order of accuracy) BLAST, PSI-BLAST [55], FFAS03 [56], and 3D-Jury [57] taking as input the sequences produced in the preceding step. The identified domains are modeled by applying a comparative modeling protocol, while all other chain fragments are treated as linkers (if they consist of fewer than 50 residues) or are assigned to structural families as defined in the Pfam-A database [58] using HMMER [59]. Fragments and sequences that have not been recognized as putative domains are analyzed via MSA of the full-length target derived from a PSI-BLAST search against the NCBI non-redundant (NR) protein sequence database [60]. Putative domains identified through Pfam-A and MSA are modeled using de novo structure prediction. Finally, following assembly, side chains are modeled by applying Monte Carlo algorithms [61]. The description is based on [60].

I-Tasser (Iterative Threading ASSEmbly Refinement) is a software package which can predict the structure of a protein given its sequence. In this application, prediction bases on querying PDB for templates using the multiple threading approach. I-Tasser is a strong contender in CASP challenges, topping the ranking in editions 7 through 12 [62–64].

The user submits a sequence of amino acids, which is then compared (by LOMETS [65]) to template proteins with similar structural characteristics. An optimal template is then selected and overlapping fragments are assembled into an output model while using replica-exchange Monte Carlo simulations [66], while differing fragments are modeled ab initio. If LOMETS is unable to identify a suitable template, the entire structure is subjected to ab initio modeling. The next step involves a search for low energy states (using SPICKER [67]) in the resulting chain via clustering simulation decoys. This is followed by the reassembly of the template protein starting with SPICER cluster centroid,

however this time the simulation is guided by spatial constraints that are provided by TM-align on the basis of LOMETS templates and PDB data. The purpose of the second iteration is to remove steric clashes as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are clustered and the lowest energy structures are selected [68]. The final step involves the construction of a detailed model from the available structures via optimization of the hydrogen bond network using REMO. Further information can be found in [68].

Robetta computations were carried out while using the publicly available service [27]. I-Tasser computations were carried out using [68]. FOD model computations were carried out using PL-Grid platform on "Cyfronet" Computer Center AGH Krakow infrastructure [69], a detailed description of which can be found in [70].

The FOD model involves two intermediate folding stages: the early-stage intermediate [71–74] and the late-stage intermediate [72–74]. The initial step, which is meant to generate a starting structure for further optimization, is omitted since the conformation listed under 5O3O is taken as the starting structure (treated as early-stage in this case). This chain is then immersed in an aqueous solvent, whose effects are modeled using a 3D Gaussian (as an external force field). In line with the FOD model, hydrophobic residues tend to congregate at the center of the protein body while hydrophilic residues are exposed on its surface. The process produces a prominent hydrophobic core that is encapsulated by a hydrophilic "shell" (with near-zero values of hydrophobicity on the surface). Optimization of hydrophobic interactions and optimization of nonbinding internal interactions is carried out in an alternating fashion, with each step being repeated several times.

Nonbonding interactions are optimized using Gromacs 4.6.5 software suite (Groningen, The Netherlands) [75], available on the PL-Grid infrastructure at ACK Cyfronet AGH Kraków [69]. FOD-based optimization aims to minimize differences between the idealized (3D Gauss function) and observed distribution of hydrophobicity in the target protein. The workflow interleaves both procedures in order to converge on the final conformation.

Folding simulations that rely on the FOD model are relevant since they acknowledge the effects exerted by the aqueous solvent, and treat them as a global phenomenon (i.e., external force field producing a molecule-wide hydrophobic core).

4.3. Comparative Analysis

All tertiary conformations that were produced by the modeling algorithms, as well as structures that are listed in PDB, were analyzed with regard to the status of their hydrophobic cores, which is described by the RD (relative distance) coefficient. Comparing RD values brings the information about the degree of disorder in respect to ideal distribution. In consequence, the approach to amyloid form can be assessed. RD expresses the degree of order present in the protein's hydrophobic core and indirectly indicates whether the protein is globular or not. Generating a globular structure with a prominent hydrophobic core (hydrophobicity peaking at the center of the molecule and decreasing along with distance from the center, becoming very low on the surface) suggests that the given amyloid peptide may, under certain circumstances, adopt a globular conformation. The ranking of protein structures, sorted in the order of decreasing globularity, reveals changes which cause proteins to forfeit their centralized hydrophobic cores and that may—in extreme cases—produce amyloid forms.

The RD coefficient can be computed for two independent cases: T-O-R and T-O-H. The former case expressing the relative distance between the observed distribution (O) and two boundary distributions: theoretical (T) which is given by the 3D Gaussian, and uniform (R, random) where each residue is ascribed a hydrophobicity value of 1/N (N being the number of residues in the input chain). R-distribution represents the case of uniform (absence of any local hydrophobicity concentration) distribution, which is the opposite one versus the centralized distribution. In the latter case the uniform distribution is replaced by a distribution corresponding to the intrinsic hydrophobicity of each residue in the input chain (H). Comparing both values reveals factors that guide the folding process (this is particularly true in the T-O-H case). A high value of RD (T-O-H) indicates that folding

is dominated by the intrinsic properties of each residue with no regard to cooperative generation of a shared hydrophobic core. When this type of distribution is repeated in successive fragment of the polypeptide, the result is a linear sequence of alternating bands of high and low hydrophobicity. This, in turn, enables the unrestricted elongation of the fibril. An interpretation of this phenomenon (referred to as "ladders") can also be found in [76].

A comparative assessment of T-O-R and T-O-H coefficients in fibrillary/amyloid structures as well as the structures that are produced by various folding algorithms may enable us to identify the "seeds" of linear propagation. FOD criteria have previously been used to assess the distribution of hydrophobicity in structures published by the CASP project [77].

It is hard to compare the interpretation based on FOD with other methods due to the fact that the hydrophobic interaction is underestimated in the discussion concerning amyloid transformation. However, some aspects of intrinsically disordered proteins that were extensively investigated [78] remain in agreement with the results of the analysis of these proteins in respect to the FOD model [79]. The development of techniques as cryo-electron microscopy [80] as well as solid state NMR [81] makes the availability of amyloid structures possible. Structuralization of water is recently in focus of attention, especially the ordering of water on surface [82] what remains in close relation to our model

Author Contributions: Conceptualization, I.R. and L.K.; Methodology, D.D. and M.G.; Software, M.B. and M.G.; Validation, D.D. and M.G.; Formal Analysis, I.R.; Investigation, I.R.; Resources, I.R. and M.B.; Writing-Original Draft Preparation, I.R.; Writing-Review & Editing, I.R. and M.B.; Visualization, M.B.; Supervision, L.K.; Project Administration, Z.W; Funding Acquisition, M.P.

Funding: This research was funded by Jagiellonian University—Medical College grant number [K/ZDS/006363 and K/ZDS/006366] and PLGrid Infrastructure—Cyfronet AGH, University of Science and Technology, 30-059 Krakow, Al. Mickiewicza 30, Poland.

Acknowledgments: We are indebted to Anna Śmietańska for technical work and to Piotr Nowakowski for translation services.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wilcock, G.K.; Esiri, M.M. Plaques, tangles and dementia. A quantitative study. *J. Neurol. Sci.* **1982**, *56*, 343–356. [CrossRef]
2. Bloom, G.S. Amyloid-β and tau: The trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol.* **2014**, *71*, 505–508. [CrossRef] [PubMed]
3. Rafii, M.S.; Lukic, A.S.; Andrews, R.D.; Brewer, J.; Rissman, R.A.; Strother, S.C.; Wernick, M.N.; Pennington, C.; Mobley, W.C.; Ness, S.; et al. PET Imaging of Tau Pathology and Relationship to Amyloid, Longitudinal MRI, and Cognitive Change in Down Syndrome: Results from the Down Syndrome Biomarker Initiative (DSBI). *J. Alzheimers Dis.* **2017**, *60*, 439–450. [CrossRef] [PubMed]
4. Chiti, F.; Dobson, C.M. Protein Misfolding, Functional Amyloid, and Human Disease. *Ann. Rev. Biochem.* **2006**, *75*, 333–366. [CrossRef] [PubMed]
5. Uversky, V.N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 932–951. [CrossRef] [PubMed]
6. Dobson, C.M. Protein folding and misfolding. *Nature* **2003**, *426*, 884–890. [CrossRef] [PubMed]
7. Fichou, Y.; Eschmann, N.A.; Keller, T.J.; Han, S. Conformation-based assay of tau protein aggregation. *Methods Cell Biol.* **2017**, *141*, 89–112. [CrossRef] [PubMed]
8. Eakin, C.M.; Berman, A.J.; Miranker, A.D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat. Struct. Mol. Biol.* **2006**, *13*, 202–208. [CrossRef] [PubMed]
9. Maccioni, R.B.; Muñoz, J.P.; Barbeito, L. The Molecular Bases of Alzheimer's Disease and Other Neurodegenerative Disorders. *Arch. Med. Res.* **2001**, *32*, 367–381. [CrossRef]
10. Kovacs, G.G. Neuropathology of tauopathies: Principles and practice. *Neuropathol. Appl. Neurobiol.* **2015**, *41*, 3–23. [CrossRef] [PubMed]

11. Fitzpatrick, A.W.P.; Falcon, B.; He, S.; Murzin, A.G.; Murshudov, G.; Garringer, H.J.; Crowther, R.A.; Ghetti, B.; Goedert, M.; Scheres, S.H.W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547*, 185–190. [CrossRef] [PubMed]
12. Bai, X.C.; McMullan, G.; Scheres, S.H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57. [CrossRef] [PubMed]
13. Cleveland, D.W.; Hwo, S.Y.; Kirschner, M.W. Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *J. Mol. Biol.* **1977**, *116*, 227–247. [CrossRef]
14. Roterman, I.; Banach, M.; Kalinowska, B.; Konieczny, L. Influence of the Aqueous Environment on Protein Structure—A Plausible Hypothesis Concerning the Mechanism of Amyloidogenesis. *Entropy* **2016**, *18*, 351. [CrossRef]
15. Roterman, I.; Banach, M.; Konieczny, L. Application of the Fuzzy Oil Drop Model Describes Amyloid as a Ribbonlike Micelle. *Entropy* **2017**, *19*, 167. [CrossRef]
16. Banach, M.; Konieczny, L.; Roterman, I. The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. *J. Theor. Biol.* **2014**, *359*, 6–17. [CrossRef] [PubMed]
17. Banach, M.; Konieczny, L.; Roterman, I. Ligand-binding-site recognition. In *Protein Folding In Silico*; Roterman-Konieczna, I., Ed.; Woodhead Publishing (Currently Elsevier): Oxford/Cambridge, UK; Philadelphia, PA, USA; New Dehli, India, 2012; pp. 79–93, ISBN 978-1-907568-17-6.
18. Banach, M.; Konieczny, L.; Roterman, I. Use of the fuzzy oil drop model to identify the complexation area in protein homodimers. In *Protein Folding In Silico*; Roterman-Konieczna, I., Ed.; Woodhead Publishing (Currently Elsevier): Oxford/Cambridge, UK; Philadelphia, PA, USA; New Dehli, India, 2012; pp. 95–122, ISBN 978-1-907568-17-6.
19. Kalinowska, B.; Banach, M.; Wiśniowski, Z.; Konieczny, L.; Roterman, I. Is the hydrophobic core a universal structural element in proteins? *J. Mol. Model.* **2017**, *23*, 205. [CrossRef] [PubMed]
20. Banach, M.; Konieczny, L.; Roterman, I. Why do antifreeze proteins require a solenoid? *Biochimie* **2018**, *144*, 74–84. [CrossRef] [PubMed]
21. Banach, M.; Konieczny, L.; Roterman, I. Secondary and super-secondary structure of proteins in light of the structure of hydrophbic core. In *Supersecondary Structure of Proteins*; Kister, A., Ed.; Springer: Cham, Switzerland, 2018, in press.
22. Kalinowska, B.; Banach, M.; Konieczny, L.; Roterman, I. Hydrophobic Core in DNA Interacting Proteins. *Entropy* **2015**, *17*, 1477–1507. [CrossRef]
23. Condello, C.; Lemmin, T.; Stöhr, J.; Nick, M.; Wu, Y.; Maxwell, A.M.; Watts, J.C.; Caro, C.D.; Oehler, A.; Keene, C.D.; et al. Structural heterogeneity and intersubject variability of Aβ in familial and sporadic Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E782–E791. [CrossRef] [PubMed]
24. Serpell, L.C. Alzheimer's amyloid fibrils: Structure and assembly. *Biochim. Biophys. Acta* **2000**, *1502*, 16–30. [CrossRef]
25. Trovato, F.; Fumagalli, G. Molecular simulations of cellular processes. *Biophys. Rev.* **2017**, *9*, 941–958. [CrossRef] [PubMed]
26. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
27. Ovchinnikov, S.; Park, H.; Kim, D.E.; DiMaio, F.; Baker, D. Protein structure prediction using Rosetta in CASP12. *Proteins* **2018**, *86*, 113–121. [CrossRef] [PubMed]
28. Robetta: Full-Chain Protein Structure Prediction Server. Available online: http://robetta.bakerlab.org (accessed on 5 July 2018).
29. Zhang, Y. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins* **2009**, *77*, 100–113. [CrossRef] [PubMed]
30. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [CrossRef] [PubMed]
31. Dygut, J.; Kalinowska, B.; Banach, M.; Piwowar, M.; Konieczny, L.; Roterman, I. Structural Interface Forms and Their Involvement in Stabilization of Multidomain Proteins or Protein Complexes. *Int. J. Mol. Sci.* **2016**, *17*, 1741. [CrossRef] [PubMed]
32. De Beer, T.A.P.; Berka, K.; Thornton, J.M.; Laskowski, R.A. PDBsum additions. *Nucleic Acids Res.* **2014**, *42*, D292–D296. [CrossRef] [PubMed]

33. Wintjens, R.; Wieruszeski, J.M.; Drobecq, H.; Rousselot-Pailley, P.; Buée, L.; Lippens, G.; Landrieu, I. 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. *J. Biol. Chem.* **2001**, *276*, 25150–25156. [CrossRef] [PubMed]
34. Kadavath, H.; Jaremko, M.; Jaremko, Ł.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. Folding of the Tau Protein on Microtubules. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 10347–10351. [CrossRef] [PubMed]
35. Galkin, V.E.; Orlova, A.; Vos, M.R.; Schröder, G.F.; Egelman, E.H. Near-atomic resolution for one state of F-actin. *Structure* **2015**, *23*, 173–182. [CrossRef] [PubMed]
36. Roterman, I.; Banach, M.; Konieczny, L. Antifreeze proteins. *Bioinformation* **2017**, *13*, 400–401. [CrossRef] [PubMed]
37. Banach, M.; Kalinowska, B.; Konieczny, L.; Roterman, I. Possible Mechanism of Amyloidogenesis of V Domains. In *Self-Assembled Molecules—New Kind of Protein Ligands—Supramolecular Ligands*; Roterman, I., Konieczny, L., Eds.; Springer: Cham, Switzerland, 2018; pp. 77–100, ISBN 978-3-319-65638-0.
38. Roterman, I.; Banach, M.; Konieczny, L. Propagation of Fibrillar Structural Forms in Proteins Stopped by Naturally Occurring Short Polypeptide Chain Fragments. *Pharmaceuticals* **2017**, *10*, 89. [CrossRef] [PubMed]
39. Roterman, I.; Dułak, D.; Gadzała, M.; Banach, M.; Konieczny, L. Structural analysis of the Aβ(11-42) amyloid fibril based hydrophobicity distribution. Unpublished work. 2018.
40. Dułak, D.; Gadzała, M.; Banach, M.; Konieczny, L.; Roterman, I. Analysis of the structure of Aβ(15-40) amyloid. *Acta Biochim. Pol.* **2018**, submitted.
41. Chiti, F.; Dobson, C.M. Protein misfolding, amyloid formation and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* **2017**, *86*, 27–68. [CrossRef] [PubMed]
42. Kim, K.H.; Späh, A.; Pathak, H.; Perakis, F.; Mariedahl, D.; Amann-Winkel, K.; Sellberg, J.A.; Lee, J.H.; Kim, S.; Park, J.; et al. Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science* **2017**, *358*, 1589–1593. [CrossRef] [PubMed]
43. Gallo, P.; Stanley, H.E. Supercooled water reveals its secrets. *Science* **2017**, *358*, 1543–1544. [CrossRef] [PubMed]
44. Palmer, J.C.; Martelli, F.; Liu, Y.; Car, R.; Panagiotopoulos, A.Z.; Debenedetti, P.G. Metastable liquid-liquid transition in a molecular model of water. *Nature* **2014**, *510*, 385–388. [CrossRef] [PubMed]
45. Tanaka, H. A self-consistent phase diagram for supercooled water. *Nature* **1996**, *380*, 328–330. [CrossRef]
46. Poole, P.H.; Sciortino, F.; Essmann, U.; Stanley, H.E. Phase behaviour of metastable water. *Nature* **1992**, *360*, 324–328. [CrossRef]
47. Sawaya, M.R.; Sambashivan, S.; Nelson, R.; Ivanova, M.I.; Sievers, S.A.; Apostol, M.I.; Thompson, M.J.; Balbirnie, M.; Wiltzius, J.J.; McFarlane, H.T.; et al. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453–457. [CrossRef] [PubMed]
48. Liu, C.; Sawaya, M.R.; Cheng, P.N.; Zheng, J.; Nowick, J.S.; Eisenberg, D. Characteristics of amyloid-related oligomers revealed by crystal structures of macrocyclic β-sheet mimics. *J. Am. Chem. Soc.* **2011**, *133*, 6736–6744. [CrossRef] [PubMed]
49. Landau, M.; Sawaya, M.R.; Faull, K.F.; Laganowsky, A.; Jiang, L.; Sievers, S.A.; Liu, J.; Barrio, J.R.; Eisenberg, D. Towards a pharmacophore for amyloid. *PLoS Biol.* **2011**, *9*, e1001080. [CrossRef] [PubMed]
50. Liu, C.; Zhao, M.; Jiang, L.; Cheng, P.N.; Park, J.; Sawaya, M.R.; Pensalfini, A.; Gou, D.; Berk, A.J.; Glabe, C.G.; et al. Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20913–20918. [CrossRef] [PubMed]
51. Wiltzius, J.J.; Landau, M.; Nelson, R.; Sawaya, M.R.; Apostol, M.I.; Goldschmidt, L.; Soriaga, A.B.; Cascio, D.; Rajashankar, K.; Eisenberg, D. Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **2009**, *16*, 973–978. [CrossRef] [PubMed]
52. De la Cruz, M.J.; Hattne, J.; Shi, D.; Seidler, P.; Rodriguez, J.; Reyes, F.E.; Sawaya, M.R.; Cascio, D.; Weiss, S.C.; Kim, S.K.; et al. Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **2017**, *14*, 399–402. [CrossRef] [PubMed]
53. Konieczny, L.; Bryliński, M.; Roterman, I. Gauss-function-Based model of hydrophobicity density in proteins. *In Silico Biol.* **2006**, *6*, 15–22. [PubMed]
54. Simons, K.T.; Bonneau, R.; Ruczinski, I.; Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **1999**, *37* (Suppl. 3), 171–176. [CrossRef]

55. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]
56. Rychlewski, L.; Jaroszewski, L.; Li, W.; Godzik, A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.* **2000**, *9*, 232–241. [CrossRef] [PubMed]
57. Ginalski, K.; Elofsson, A.; Fischer, D.; Rychlewski, L. 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics* **2003**, *19*, 1015–1018. [CrossRef] [PubMed]
58. Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S.R.; Griffiths-Jones, S.; Howe, K.L.; Marshall, M.; Sonnhammer, E.L. The Pfam protein families database. *Nucleic Acids Res.* **2002**, *30*, 276–280. [CrossRef] [PubMed]
59. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [CrossRef] [PubMed]
60. Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526–W531. [CrossRef] [PubMed]
61. Kuhlman, B.; Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10383–10388. [CrossRef] [PubMed]
62. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40. [CrossRef] [PubMed]
63. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [CrossRef] [PubMed]
64. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [CrossRef] [PubMed]
65. Wu, S.; Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* **2007**, *35*, 3375–3382. [CrossRef] [PubMed]
66. Swendsen, R.H.; Wang, J.S. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609. [CrossRef] [PubMed]
67. Zhang, Y.; Skolnick, J. SPICKER: A clustering approach to identify near-native protein folds. *J. Comput. Chem.* **2004**, *25*, 865–871. [CrossRef] [PubMed]
68. About I-TASSER Server. Available online: https://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html (accessed on 5 July 2018).
69. PLGrid User Manual. Available online: https://docs.cyfronet.pl/pages/viewpage.action?pageId=23626756https://docs.cyfronet.pl/display/PLGDoc/User+manual (accessed on 5 July 2018).
70. Roterman-Konieczna, I. *Protein Folding—Protein Folding Versus Protein Structure Prediction*; Woodehead Publishing (Currently Elsevier): Oxford/Cambridge, UK; Philadelphia, PA, USA; New Dehli, India, 2012; ISBN 978-1-907568-17-6.
71. Roterman, I.; Konieczny, L.; Banach, M.; Jurkowski, W. Intermediates in the protein folding process: A computational model. *Int. J. Mol. Sci.* **2011**, *12*, 4850–4860. [CrossRef] [PubMed]
72. Kalinowska, B.; Krzykalski, A.; Roterman, I. Contingency Table Browser—Prediction of early stage protein structure. *Bioinformation* **2015**, *11*, 486–488. [CrossRef] [PubMed]
73. Jurkowski, W.; Bryliński, M.; Konieczny, L.; Wiíniowski, Z.; Roterman, I. Conformational subspace in simulation of early-stage protein folding. *Proteins* **2004**, *55*, 115–127. [CrossRef] [PubMed]
74. Kalinowska, B.; Fabian, P.; Stąpor, K.; Roterman, I. Statistical dictionaries for hypothetical in silico model of the early-stage intermediate in protein folding. *J. Comput. Aided Mol. Des.* **2015**, *29*, 609–618. [CrossRef] [PubMed]
75. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef] [PubMed]
76. Biancalana, M.; Makabe, K.; Koide, S. Minimalist design of water-soluble cross-β architecture. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3469–3474. [CrossRef] [PubMed]
77. Gadzała, M.; Kalinowska, B.; Banach, M.; Konieczny, L.; Roterman, I. Determining protein similarity by comparing hydrophobic core structure. *Heliyon* **2017**, *3*, e00235. [CrossRef] [PubMed]
78. Uversky, V.N. Amyloidogenesis of natively unfolded proteins. *Curr. Alzheimer Res.* **2008**, *5*, 260–287. [CrossRef] [PubMed]

79. Kalinowska, B.; Banach, M.; Konieczny, L.; Marchewka, D.; Roterman, I. Intrinsically disordered proteins—Relation to general model expressing the active role of the water environment. *Adv. Protein Chem. Struct. Biol.* **2014**, *94*, 315–346. [CrossRef] [PubMed]
80. Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R.B.G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; et al. Fibril structure of amyloid-β(1-42) by cryo-electron microscopy. *Science* **2017**, *358*, 116–119. [CrossRef] [PubMed]
81. Tycko, R. Indirect detection in solid state NMR: An illustrious history and a bright future. *J. Magn. Reson.* **2018**, *288*, 122–123. [CrossRef] [PubMed]
82. Macias-Romero, C.; Nahalka, I.; Okur, H.I.; Roke, S. Optical imaging of surface chemistry and dynamics in confinement. *Science* **2017**, *357*, 784–788. [CrossRef] [PubMed]

Załącznik nr. 2

Structural analysis of the Aβ(15-40) amyloid fibril based on hydrophobicity distribution

DAWID DUŁAK, MATEUSZ BANACH, MAŁGORZATA GADZAŁA,

LESZEK KONIECZNY, IRENA ROTERMAN

ACTA ABP BIOCHIMICA POLONICA

Vol. 65, No 4/2018
595–604
https://doi.org/10.18388/abp.2018_2647

Regular paper

Structural analysis of the Aβ(15-40) amyloid fibril based on hydrophobicity distribution

Dawid Dułak[1], Mateusz Banach[2], Małgorzata Gadzała[3], Leszek Konieczny[4] and Irena Roterman[2,5]✉

[1]ABB Business Services Sp. z o.o., Warszawa, Poland; [2]Departament of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Kraków, Poland; [3]ACK – Cyfronet AGH, Kraków, Poland, currently: Schibsted Tech Polska Sp. z o. o., Kraków, Poland; [4]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kraków, Poland; [5]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Kraków, Poland

The Aβ42 amyloid is the causative factor behind various neurodegenerative processes. It forms elongated fibrils which cause structural devastation in brain tissue. The structure of an amyloid seems to be a contradiction of protein folding principles. Our work focuses on the Aβ(15-40) amyloid containing the D23N mutation (also known as the "Iowa mutation"), upon which an *in silico* experiment is based. Models generated using I-Tasser software as well as the fuzzy oil drop model – regarded as alternatives to the amyloid conformation – are compared in terms of their respective distributions of hydrophobicity (i.e. the existence of a hydrophobic core). In this process, fuzzy oil drop model parameters are applied in assessing the propensity of selected fragments for undergoing amyloid transformation.

Key words: amyloidosis, fibril formation, Aβ42, Aβ(15-40)

Received: 22 July, 2018; revised: 12 September, 2018; accepted: 09 October, 2018; available on-line: 21 November, 2018

✉e-mail: myroterm@cyf-kr.edu.pl

Abbreviations: 3D Gauss, 3-dimensional Gauss function; CASP, Critical Assessment of protein Structure Prediction; EM, electron microscopy; FOD, fuzzy oil drop; HvO, correlation between H – intrinsic hydrophobicity *versus* O – observed distribution; HvT, correlation coefficient for relation intrinsic hydrophobicity H versus theoretical distribution (T); PDB, Protein Data Bank; RD, relative distance; ssNMR, solid state nuclear magnetic resonance; T-O-R, symbol to express the relation of observed (O) distribution versus T – theoretical and R, random; T-O-H – T, theoretical , O – observed, H – intrinsic hydrophobicity – relation between three distributions; T-O-R – T, theoretical , O – observed, R – unified distribution – relation between three distributions; TvO, correlation coefficient for relation T – theoretical *versus* O – observed distribution

INTRODUCTION

Attempts at protein structure prediction (as well as at identifying the underlying mechanisms of protein folding) have a long history (Dill *et al.,* 2012). The CASP initiative (http://predictioncenter.org) aims to develop various structure prediction methods which exemplify two competing approaches: comparative modeling (Monzon *et al.,* 2017; Zhou *et al.,* 2010; Bystroff *et al.,* 2004) and *ab initio* (new fold) models (Saunders *et al.,* 2002; Liwo *et al.,* 2005; Moult *et al.,* 2016). The former group focuses on sequential similarities under the assumption that identical sequences tend to produce similar secondary folds. This, in addition to acknowledgment of evolutionary factors, enables researchers to narrow down the search to a set of homologous proteins, which, in turn, enables the given structure to be assigned to a specific branch of the evolutionary tree. Homology-based comparative modeling provides clues regarding the properties of peptides which share sequential similarities and biological function (and which are therefore likely to adopt similar conformations). In contrast, *ab initio* (new fold) methods rely entirely on theoretical speculation with no need for comparative analysis. This approach assumes that if the model is correct, it should produce an accurate fold for any arbitrary sequence, with no need to study its evolutionary counterparts.

In to-date editions of the CASP comparative modeling techniques have generally yielded better results, although the search for *ab initio* models is by no means over (Moult *et al.,* 2016).

The longstanding dogma stating that similar sequences should produce similar structures was shaken by the discovery of chameleon sequences – with a length of up to 11 aa – which adopt drastically different conformations depending on their host protein: for example, 1D2E (β-fold) and 2C78 (helix). An entire database of chameleon fragments has since been compiled (Li *et al.,* 2015; Ghozlane *et al.,* 2009).

Existing protein structure prediction methods have not yet yielded satisfactory results, as attested to by leading experts who perceive the need for new solutions to augment their research efforts (Khoury *et al.,* 2005).

Somewhat paradoxically, misfolded proteins – including amyloids – may provide important clues regarding the folding process itself, and explain structural rearrangements which sometimes occur within fully folded proteins (Lührs *et al.,* 2005; Fu *et al.,* 2015). In the process of amyloidogenesis, structural changes may occur despite the lack of mutations. It appears that sequentially identical polypeptides may, under most circumstances, fold in an appropriate manner, producing a biologically active protein, while in some cases they may adopt conformations which differ from their respective native forms. In many cases, these alternative conformations manifest as elongated, fibrillary structures capable of unrestricted growth. In such cases the protein in question forfeits its biological function (whatever it may be), becomes insoluble and resists proteolytic enzymes.

Ever since their discovery, amyloids have been closely linked to a specific type of beta fold referred to as cross-beta (Xu, 2009). Studies which focus on structural properties of amino acids show that while some residues are more likely to form helical folds, others are more commonly found in beta fragments (Ghozlane *et al.,* 2009; Kister, 2015). Based on this observation, conformational rearrangement mechanisms are proposed that involve replication of beta folds and (in the presence

of a hydrogen bond network) may support unrestricted linear propagation. This phenomenon is regarded as the principal driving force behind amyloidogenesis (Eisenberg *et al.,* 2017).

The structure of amyloid fibrils eluded analysis for a long time (they do not crystallize, which precludes X-ray imaging, and are insoluble, preventing the application of classic NMR techniques). Recently, however, solid-state NMR (ssNMR) has been successfully applied in amyloid fibril studies (Tycko, 2011).

The presented analysis focuses on a specific type of amyloid structure referred to as Aβ(15-40) (Sgourakis *et al.,* 2015), itself part of the Aβ42 amyloid (Younkin, 1998). We apply concepts derived from the hydrophobic core theory to characterize this structure – which has already been shown to exhibit linear propagation of alternating bands of high and low hydrophobicity (Roterman *et al.,* 2017; Roterman *et al.,* 2016). Such propagation progresses along the long axis of the amyloid and is notable for lacking an arrestor which would otherwise prevent arbitrary elongation of the resulting structure. Based on these observations we propose a structural rearrangement mechanism which may cause a globular protein to transform into an elongated fibril (Roterman *et al.*, 2016; Roterman *et al.,* 2017).

Globular proteins are characterized by the presence of a monocentric hydrophobic core of hydrophobicity distribution following the 3 D Gauss function. The distribution in proteins includes local deformations. Local hydrophobicity deficiencies usually correspond to ligand binding cavities (Banach *et al.,* 2012a), while excess hydrophobicity – if found on the surface – typically marks a complexation site (Banach *et al.,* 2012b). The fuzzy oil drop model expresses the structure of the protein's hydrophobic core and is applicable to a wide variety of structurally diverse proteins (Kalinowska *et al.,* 2017).

The global discordance introduces the alternative distribution form which in amyloids adopts the linear form producing fibrils (Roterman *et al.,* 2016; Roterman *et al.,* 2017). The presence of a monocentric hydrophobic core promotes solubility – conversely, if the core is replaced by a linear pattern of alternating bands the structure becomes susceptible to unrestricted elongation. This effect may be achieved e.g. by introducing highly hydrophilic residues into the central part of the protein body. When such residues adopt a cooperative (repetitive) pattern, linear ordering becomes a strong possibility (Kalinowska *et al.,* 2017; Banach *et al.,* 2018).

The fuzzy oil drop model may be applied to assess the degree of deformations in the protein's hydrophobic core – up to and including amyloid-like conformations, as discussed in (Roterman *et al.,* 2016; Roterman *et al.,* 2017). A detailed description of the model can be found in (Kalinowska *et al.,* 2015).

This work provides a comparative analysis of Aβ(15-40) peptides capable of adopting non-amyloid conformations. In order to generate a set of starting structures we applied the I-Tasser software (Zhang, 2008); a strong contender in recent editions of the CASP challenge (Yang *et al.,* 2015). We also generated alternative structures based on fuzzy oil drop-driven folding algorithms, which – in addition to optimizing internal force fields – account for the presence of the aqueous solvent, favoring generation of a monocentric hydrophobic core (Konieczny *et al.,* 2006). Each program yielded 5 alternative structures (in accordance with CASP rules) for a total of 10 structures. These structures were compared with one another, and with the reference Aβ(15-40) amyloid. The comparative study was based on fuzzy oil drop criteria, as described in the Materials and Methods section.

The presented analysis should be viewed as a follow-up to our study of the properties of amyloid aggregates (particularly Aβ(15-40)), which acknowledge all structural forms tagged in PDB as amyloid seeds.

MATERIALS AND METHODS

The structure of Aβ(15-40). The atomic model of the Aβ amyloid which includes the so-called "Iowa mutation" (D23N) is listed in PDB under ID 2MPZ (Sgourakis *et al.,* 2015). Its structure has been determined using solid-state NMR and EM, and further confirmed by Roberta modeling (http://robetta.bakerlab.org). The 15-20 fragment of the classic Aβ42 amyloid, as presented in PDB, will be referred to as Aβ(15-20).

The listed structure appears in the form of a superfibril consisting of three individual protofibrils in a triangular configuration. Our analysis concerns the entire superfibril, an individual protofibril as well as an isolated chain (component of the protofibril). Each of these structures is characterized using the same set of parameters, facilitating meaningful comparisons regardless of the composition and size of the structure in question.

Alternatives to the ssNMR structure. In addition to the structure listed under 2MPZ we also generated alternative models for the Aβ(15-40) sequence, using I-Tasser software (https://zhanglab.ccmb.med.umich.edu/I-TASSER), which is highly ranked in the CASP competition (http://predictioncenter.org; https://www.dnastar.com/blog/structural-biology/novafold-and--i-tasser-a-winning-combination-for-protein-structure-prediction-and-analysis). While Roberta (Kim *et al.,* 2004) provides similarly accurate results, it could not be applied in the presented case due to its minimum chain length limitation, which the Aβ(15-40) fragment does not satisfy.

The input for I-Tasser computations was provided by the Aβ(15-40) sequence as listed in 2MPZ, i.e. inclusive of the D23N mutation. I-Tasser program produced five models for the target sequence. All of them are taken for analysis in this paper.

The second software package used to model the presented peptide is a tool based on the fuzzy oil drop model (Kalinowska *et al.,* 2015; Konieczny *et al.,* 2006). The model approaches the protein folding problem by introducing an external force field which accounts for the presence of water. However, rather than model the solvent as a collection of individual molecules, the FOD model treats it as a continuum, mathematically expressed by a 3D Gaussian. The effect of the environment is to direct hydrophobic residues towards the center of the emerging molecule. Folding is therefore assumed to occur inside a suitably defined 3D Gaussian capsule, where the distribution of hydrophobicity is subjected to optimization using the Gaussian as a reference. The models produced on the basis of FOD model visualize the alternative folding scenario with aim-oriented hydrophobic core generation.

The described software operates on the supercomputing resources provided by the Academic Computing Centre Cyfronet AGH (as part of the PL-Grid infrastructure) and was used to generate five alternative conformations for the Aβ(15-40) peptide. During these computations, internal force fields were optimized using the Gromacs software package (Berendsen *et al.,* 1995; http://www.gromacs.org), which is also available at Cyfronet.

FOD model delivered many structures (about 500). The models taken for analysis are those representing the extreme status: the highest and the lowest approach in respect to 3D Gauss function representing the distribution of hydrophobicity.

Comparative analysis of the obtained structures and of ssNMR Aβ(15-40). According to the amyloidosis model presented in (Roterman *et al.*, 2016; Roterman *et al.,* 2017), it is assumed that interaction between the polypeptide and the aqueous solvent plays a crucial role in ensuring correct folding and functioning of proteins. Anomalies which result in misfolded proteins may – under the fuzzy oil drop criteria – be attributed to unusual interactions between the protein and its environment (i.e. the external force field). This is why our comparative analysis bases on the fuzzy oil drop model, which provides a measure of the structural ordering of the protein's hydrophobic core. Similarities between the observed (O) and theoretical (T – given by the 3D Gaussian) distribution are quantified using Kullback-Leibler's divergence entropy formula (Kullback *et al.,* 1951). The resulting parameter provides a way to compare a variety of diverse structural forms, including alternative conformations of a specific sequence, ranked according to their accordance with the theoretical hydrophobic core structure.

The aforementioned parameter, denoted RD (Relative Distance) expresses the distance between O and two reference distributions treated as boundary cases. The first of these is the aforementioned theoretical distribution (T), while the other one, referred to as uniform (or random) ascribes a hydrophobicity value of 1/N to each residue (N being the number of residues comprising the chain). When RD<0.5, the observed distribution is regarded as more closely aligned with T, indicating the presence of a hydrophobic core. In the opposite case – RD≥0.5 – the protein is thought to lack a monocentric core. The entire model is also referred to as RD(T-O-R), which means that O (the observed distribution) is compared against T (perfect 3D Gaussian) and R (no concentration of hydrophobicity at any point within the protein body).

As shown in (Roterman *et al.,* 2016; Roterman *et al.,* 2017), amyloids exhibit a peculiar distribution of hydrophobicity which in no way resembles the monocentric core model. In order to better analyze such structures, we introduce another variant of RD, designated RD(T-O-H), where the "random" distribution is replaced with a distribution reflecting the intrinsic hydrophobicity of each residue in the input chain (denoted H). When O is more closely aligned with H than with T, we may claim that the folding process is dominated by the individual preferences of residues with no cooperative tendency to generate a shared (protein-wide) hydrophobic core. For the same reasons this type of conformation may be regarded as "selfish" – indeed, no common "policy" emerges to define the common centric hydrophobic core – as the result of cooperative participation of all residues.

The observed distribution is a result of hydrophobic interactions between neighboring residues. The force of such interactions depends on the mutual separation between residues, as well as on their intrinsic hydrophobicity, as discussed in (Levitt, 1976). In contrast, the theoretical distribution (T) can be used to derive the expected values of hydrophobicity at any point within the protein body – including at the locations of effective atoms (averaged-out positions of all atoms comprising a given residue).

In order to fully characterize the presented structural forms, we also computed three correlation coefficients (for each structure): HvT, TvO and HvO. High values of these coefficients, particularly in the case of HvO, suggest a strong influence of intrinsic hydrophobicity upon the final distribution observed in the molecule.

All the above parameters were computed separately for the whole complex (superfibril), for an individual protofibril and for an individual chain (subunit of the fibril). In each case it is assumed that the introduced parameters can perform the comparative analysis of all discussed structural forms. In addition, selected chain fragments can be studied to assess their contribution to the final structure – these computations may be useful in determining which fragments cause the fibril to emerge in the first place. Low HvT and TvO coefficients coupled with high values of HvO coefficients may be regarded as an indicator of amyloidogenic potential, where the structure is more closely aligned with H than with T. If high HvO is accompanied by negative HvT and TvO, we may conclude that the given fragment is an amyloid seed. It means that the structure is generated "against" the rules producing the globular form of protein. Such conclusions can be reinforced by high values of RD (T-O-R and particularly T-O-H).

The analysis presented in this work bases on interpreting the above parameters.

All models delivered by I-Tasser program (five of them) are taken as objects of analysis.

Structural analysis using FOD model. The characteristics of superfibril is based on the construction of 3D Gauss function encapsulating the complete superfibril. The status of protofibril is defined based on 3D Gauss function. The individual chain characteristics are concluded from the application of 3D Gauss function constructed for individual chain.

The status of chain as part of the selected form of fibril is defined via considering the chain as part of the object, which in our case is the protofibril. The reference distribution in this case is the distribution as it appears in 3D Gauss function defined for protofibril.

RESULTS

The five structures generated with I-Tasser (all structures produced by this program are present in this analysis – according to CASP rules each participant is allowed to deliver 5 models for each target) represent the full spectrum – from globules to near-unfolded forms. The same is true for the structures obtained using the FOD model. The summary of results is given in Table 1.

Structure of the Aβ(15-40) superfibril

The Aβ(15-40) structure listed in PDB, when analyzed from the fuzzy oil model perspective, may be characterized as highly discordant with respect to the monocentric distribution of hydrophobicity. This is visualized in Fig. 1A, where the expected hydrophobicity peak in the central part of the molecule is not replicated by the actual protein. Instead, we face a sinusoidal sequence of alternating peaks and troughs, resulting from symmetrical alignment of identical chain fragments. The observed flattening of peaks in the N- and C-terminal section is not caused by alignment with the monocentric core model, but rather by the fact that the listed structure consists of a finite number of polypeptides.

Structural parameters calculated for the superfibril are as follows: RD(T-O-R)=0.578; RD(T-O-H)=0.494;

Table 1. Characteristics of all the models discussed in the paper (left column).
In the main text, Af is used for description of Aβ(15-40) chain in protofibril, and Ai for status of Aβ(15-40) chain treated as individual structural unit. Parameters based on fuzzy oil drop model expressing RD(T-O-R) and RD(T-O-H) and correlation coefficients for relation are named HvT, TvO and HvO. The parameter values that suggest the status of amyloid character are given in bold. Parameters are given both for the complete chain of Aβ(15-40) and for the chain fragments.

	FRAGMENT 15-40					FRAGMENT 14-20					FRAGMENT 21-28					FRAGMENT 28-31					FRAGMENT 34-40				
FORM	RD		Corr. Coeff.			RD		Corr. Coeff.			RD		Corr. Coeff.			RD		Corr. Coeff.			RD		Corr. Coeff.		
	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O--H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O--R	T-O--H	HvT	TvO	HvO	T-O-R	T-O--H	HvT	TvO	HvO
F1	0.188	0.122	0.158	0.918	0.295	0.198	0.129	0.560	0.962	0.660	0.300	0.143	-0.080	0.819	0.474	0.136	0.159	0.595	0.988	0.665	0.227	0.161	-0.443	0.948	-0.463
F2	0.215	0.073	0.208	0.856	0.254	0.232	0.058	0.527	0.825	0.543	0.224	0.315	0.798	0.922	0.903	0.124	0.033	0.278	0.971	0.324	0.304	0.187	-0.458	0.728	-0.685
F3	0.218	0.179	0.208	0.912	0.279	0.253	0.209	0.608	0.905	0.662	0.219	0.356	0.477	0.916	0.596	0.175	0.172	0.575	0.995	0.527	0.258	0.155	-0.503	0.798	-0.684
I1	0.426	0.309	0.375	0.629	0.634	0.240	0.240	0.880	0.960	0.888	0.412	0.587	0.333	0.646	0.834	0.406	0.117	0.283	0.652	0.821	0.764	0.595	0.471	0.473	-0.503
I2	0.466	0.289	0.285	0.551	0.497	0.288	0.160	0.415	0.812	0.815	0.426	0.610	0.174	0.635	0.849	0.204	0.077	0.895	0.925	0.946	0.731	0.672	0.234	0.611	0.138
I3	0.558	0.416	0.255	0.438	0.717	0.371	0.290	0.737	0.616	0.876	0.401	0.506	0.259	0.608	0.892	0.447	0.284	0.353	0.529	0.965	0.915	0.782	0.505	0.227	-0.513
CHAIN S IN FIBRYL Af	**0.567**	**0.564**	0.360	0.345	0.839	0.331	0.350	0.751	0.686	0.962	**0.748**	**0.802**	**-0.248**	**-0.172**	**0.917**	0.284	0.229	0.598	0.844	0.924	0.892	0.730	0.255	0.291	-0.339
F4	0.571	0.544	0.010	0.525	0.480	0.347	0.442	0.697	0.748	0.776	0.530	0.870	0.159	0.296	0.904	0.303	0.270	0.522	0.938	0.686	0.874	0.765	0.317	-0.125	-0.680
F5	0.591	0.553	-0.020	0.447	0.553	0.572	0.738	-0.163	-0.065	0.922	0.382	0.541	0.168	0.711	0.783	0.098	0.075	0.628	0.991	0.716	0.794	0.706	0.404	-0.101	-0.571
I4	0.620	0.387	0.183	0.442	0.552	0.363	0.177	0.523	0.659	0.856	0.295	0.180	0.494	0.844	0.755	0.400	0.171	0.055	0.581	0.690	0.834	0.802	0.171	0.316	0.068
CHAINS INDIVID Ai	0.626	0.466	0.355	0.351	0.615	0.289	0.182	0.788	0.720	0.890	**0.865**	**0.716**	**-0.260**	**-0.222**	**0.838**	0.097	0.038	0.661	0.960	0.826	0.874	0.754	0.246	0.168	-0.504
I5	0.870	0.573	-0.046	-0.006	0.460	0.829	0.188	-0.099	0.089	0.877	0.563	0.240	0.821	0.630	0.559	0.614	0.110	0.486	0.296	0.850	0.942	0.872	0.330	-0.559	-0.266

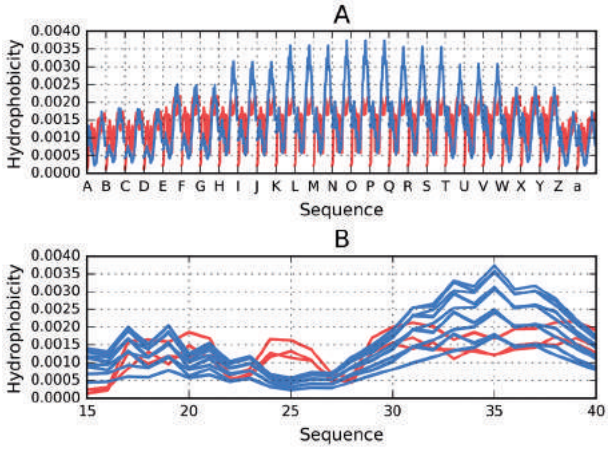


Figure 1. Distribution of hydrophobicity in the Aβ(15-40) superfibril: T – blue, O – red.
(**A**) profile for consecutive chains as they appear in superfibril. (**B**) overlapped profiles according to residual sequence in all chains.

HvT=0.394; TvO=0.554; HvO=0.790. The RD(T-O-R) value in excess of 0.5 suggests that no central hydrophobic core is present in this structure. On the other hand, RD(T-O-H)<0.5 indicates relatively limited influence of intrinsic hydrophobicity upon the conformation of the superfibril.

We can observe the expected reduction in hydrophobicity in outlying peptides, along with an increase in hydrophobicity in the central chains (Fig. 1B). Eliminating the outlying (edge) chains reveals two distinct hydrophobicity profiles (Fig. 2A). When considering only the chains labeled G through U, a common pattern emerges (Fig. 2B), while the central chains (M, N and O) in each protofibril all share a nearly-identical distribution (Fig. 2C), which diverges from the theoretical profile.

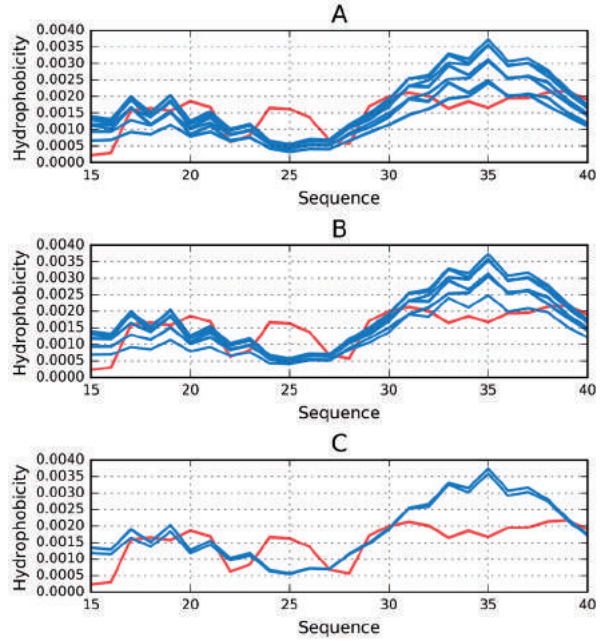


Figure 2. Distribution of hydrophobicity in the Aβ(15-40) superfibril (T – blue, O – red) presented as overlapped profiles according to residual sequence.
(**A**) following elimination of outlying chains in each protofibril: A, a, Z and B, C, D. (**B**) for the internal chains of each protofibril: G – U. (**C**) for the central chains in each protofibril: M, N and O.

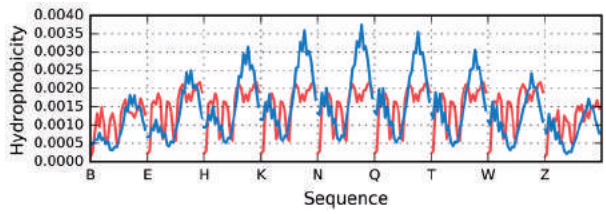


Figure 3. Theoretical (blue) and observed (red) distributions calculated for the Aβ(15-40) protofibril.
Chains are displayed in order as they appear in the protofibril.

Structure of the Aβ(15-40) protofibril

Figure 3 illustrates the distributions of hydrophobicity (T and O) in the Aβ(15-40) protofibril which consists of chains B, E, H, K, N, Q, T, W and Z. In this case, the shape of the Gaussian capsule was adjusted to encapsulate only this protofibril. The observed distribution is characteristic of amyloid structures, with no concentration of hydrophobicity observed in the central part of the fibril. A sinusoidal arrangement of alternating maxima and minima is evident instead.

Structural parameters calculated for the protofibril are as follows: RD(T-O-R)=0.615; RD(T-O-H)=0.600. Both values indicate that the distribution deviates from a monocentric model. HvT, TvO and HvO values are 0.262, 0.412 and 0.788 respectively.

Status of the chain S analyzed as a component of the protofibril

In order to identify potential amyloid seeds, we analyzed chain S as a component of the protofibril.

Distribution of hydrophobicity in chain S as part of protofibril (Fig. 4A) reveals fragments where O not only diverges from T but may even be regarded as its polar opposite. These fragments are also characterized by good alignment between H and O, which shows that their conformation is driven by the properties of individual residues. Correlation coefficients calculated for a 5 aa moving frame further confirm that for certain fragments O opposes T, preventing the formation of a globular

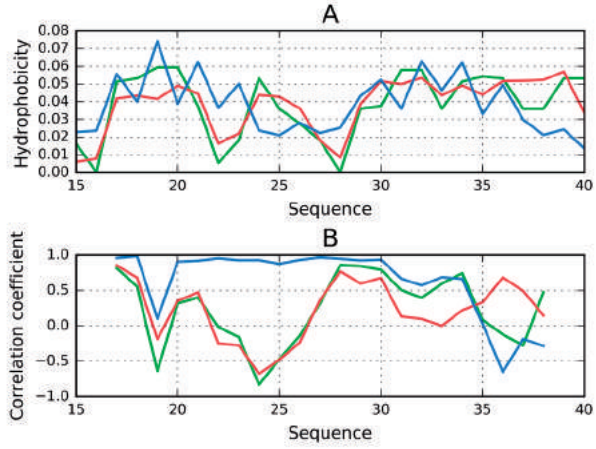


Figure 4. Chain S analyzed as a component of the protofibril.
(**A**) T (blue), O (red) and H (green) hydrophobicity distributions. (**B**) correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i.e. 20 corresponds to residue 20 in the 18-19-20-21-22 frame).

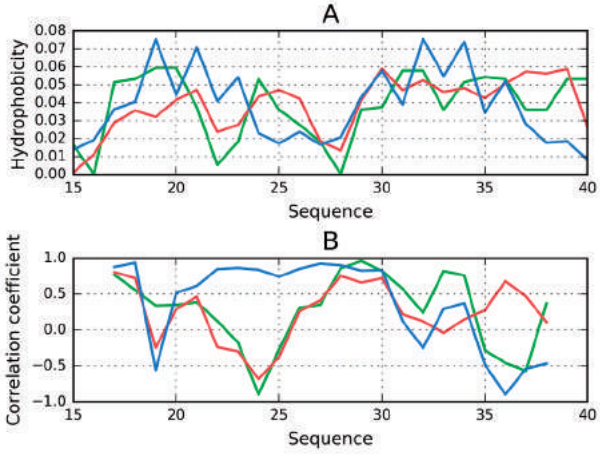


Figure 5. Chain S analyzed as an individual molecule.
(**A**) T (blue), O (red) and H (green) hydrophobicity distributions. (**B**) correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i.e. 20 corresponds to residue 20 in the 18-19-20-21-22 frame).

structure. As shown in Fig. 4B, the fragment at 21-26 may be regarded as an amyloid seed, with a very high value of HvO and negative values of HvT and TvO.

As a result, fragments 15-21, 22-26, 27-31 and 34-40 have been singled out for individual analysis concerning their adherence to the hydrophobic core model in structures generated by I-Tasser and FOD software.

Status of the chain S analyzed as an individual molecule

Chain S treated as an individual molecule characterized by FOD-based parameters can be described as follows: RD(T-O-R)=0.626, RD(T-O-H)=0.467, correlation coefficients: HvT=0.355, TvO=0.351, HvO=0.615. The structure is dominated by intrinsic hydrophobicity despite RD(T-O-H) value below 0.5, which does not support this observation. It is the complete set of parameters that reveals that this chain is not any result of uni-centric tendency in folding process.

The visualization of the hydrophobicity distributions shown in Fig. 5A indicates the same polypeptide fragments of chain S as highly discordant versus the uni-centric construction of the hydrophobic core.

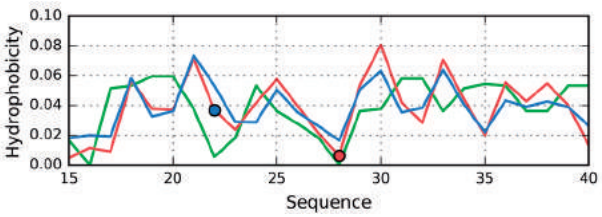


Figure 7. Intrinsic (H – green), theoretical (T – blue) and observed (O – red) distributions of hydrophobicity for structure F1 generated using the fuzzy oil drop model.
The two circles mark observed hydrophobicity values for residues 22Glu (blue) and 28Lys (red).

The profiles shown in Fig. 5B resemble the form of chain S treated as part of protofibril.

In summary, the fragments 15-21, 22-26, 27-31 and 34-40 have been singled out for individual analysis concerning their adherence to the hydrophobic core model in structures generated by I-Tasser and FOD software. The detailed analysis of chain fragments is given for all discussed structural forms in Table 1.

Analysis of the above results is that the fragment 22-28 represents the status discordant in respect to expected hydrophobicity distribution. The local maximum observed for this fragment is highly discordant with idealized distribution. The high values for RD and negative values of correlation coefficients for this fragment – as interpreted using the fuzzy oil drop model – classifies this fragment as representing the amyloid status. The status of this fragment is especially traced in all models discussed in this paper. This is why this fragment is distinguished in all 3D presentations of models discussed in this paper (white fragment in the presentations). The positions of residues 22Glu and 28Lys are shown in all discussed structural forms. However, the other fragments in received models appear to represent the status recognized by fuzzy oil drop model as amyloidogenetic.

Comparative analysis of conformations adopted by the Aβ(15-40) sequence

Table 1 provides a list of parameters which characterize all 10 structures generated using I-Tasser and FOD

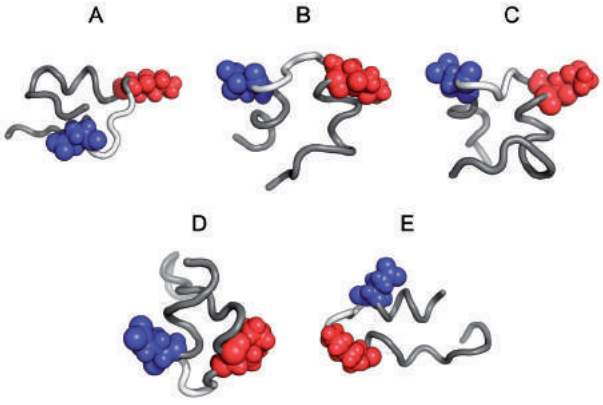


Figure 6. 3D views of models exhibiting RD(T-O-R) value <0.5. 22Glu is shown in blue and 28Lys in red. Green backbone section corresponds to residues 23-26.
(**A**) Model F1. (**B**) Model F2. (**C**) Model F3. (**D**) Model I1. (**E**) Model I2.

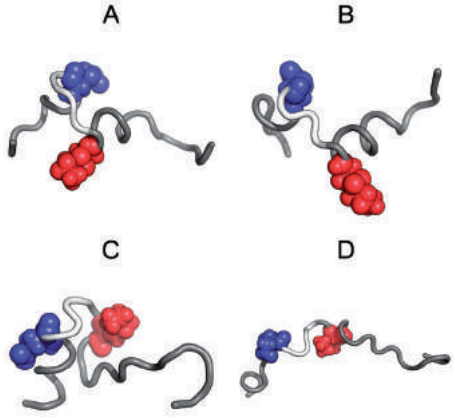


Figure 8. 3D views of models exhibiting RD(T-O-R) value >0.5. 22Glu is shown in blue and 28Lys in red. Green backbone section corresponds to residues 23-26.
(**A**) Model F4. (**B**) Model F5. (**C**) Model I4. (**D**) Model I5.

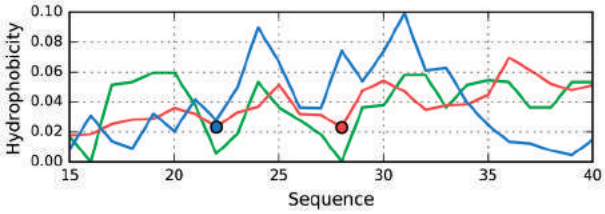


Figure 9. Intrinsic (H – green), theoretical (T – blue) and observed (O – red) distributions of hydrophobicity for a structure I5 generated using I-Tasser.
The two circles mark hydrophobicity values observed for residues 22Glu (blue) and 28Lys (red).

software while acknowledging the status of the 15-40 fragment as a component of the protofibril. The structures were sorted in the order of increasing values of the RD(T-O-R) coefficient.

It is immediately apparent that the presented polypeptide may, in fact, adopt a globular conformation, which corresponds to low values of RD(T-O-R). These low values indicate good agreement between the observed distribution and the 3D Gaussian form, which in turn suggests the presence of a hydrophobic core – as in the case of globular proteins. Figure 6 presents 3D views of some of the structures provided by I-Tasser and FOD.

Somewhat expectedly, structures which possess well-defined hydrophobic cores have been generated by an algorithm which involves optimization of hydrophobic interactions (models F1-F3), however this list also includes two I-Tasser models (I1 and I2), as revealed in Fig. 6.

Structural analysis of models characterized by low RD(T-O-R) reveals "proper" exposure of hydrophilic residues on the protein surface. This observation is further supported by plotting the full distribution of hydrophobicity in a representative structure F1 – Fig. 7.

The hydrophilic residue 28Lys occupies location on the surface, as predicted by the model (Fig. 7). Its sequential neighborhood, characterized by progressive increases of hydrophobicity, is directed towards the central part of the molecule (cf. centrally placed local maxima). We may also observe good alignment between T, O and H for the fragment at 24-30.

Despite "correct" location of the two selected hydrophilic residues (Fig. 8) (exposure on surface in F4, F5, I4 and I5 models), these proteins exhibit a major deviation from the theoretical distributions (Table 1). Figure 8 visualizes the possible different structural forms predicted for Aβ(15-40), where a globular form is obtained, however, these structures do not satisfy the condition of the order expected for soluble proteins. The solubility – according to fuzzy oil drop model – requires the presence of external layer of low hydrophobicity, which is not the case.

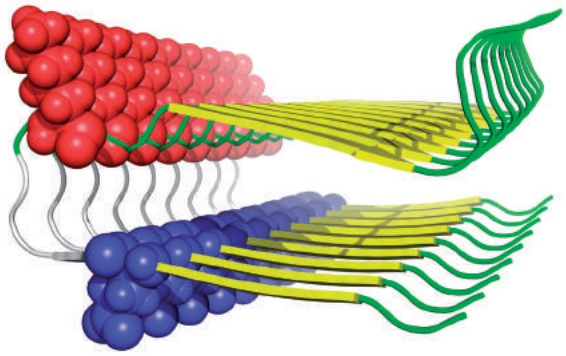


Figure 10. Location of selected residues (22Glu – blue; 28Lys – red) in the protofibril. Green backbone section corresponds to residues 23-26.

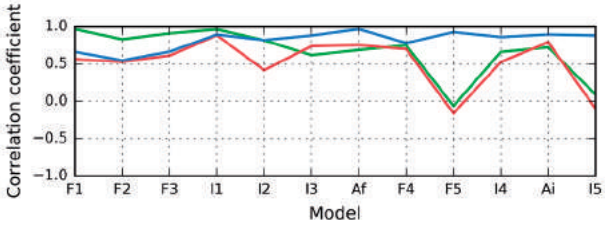


Figure 11. Correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for residues 15-20 in the successive structures as listed in Table 1.
High values of HvO coupled with negative values of HvT and TvO identify these fragments as amyloid-like in F5 and in I5.

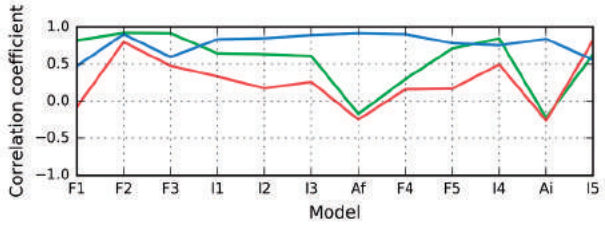


Figure 12. Correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for residues 21-26 in the successive structures as listed in Table 1.
High values of HvO coupled with negative values of HvT and TvO identify these fragments as amyloid-like in Af (chain in protofibril) and in Ai (chain as individual unit).

Model I5 provides an example of a discordant structure (Fig. 9). While residue 22Glu is correctly positioned, Lys28 is found in an area where a local maximum is expected, generating a distribution which opposes the theoretical model. These conditions may be regarded as conductive to amyloid transformation, with the 26-31

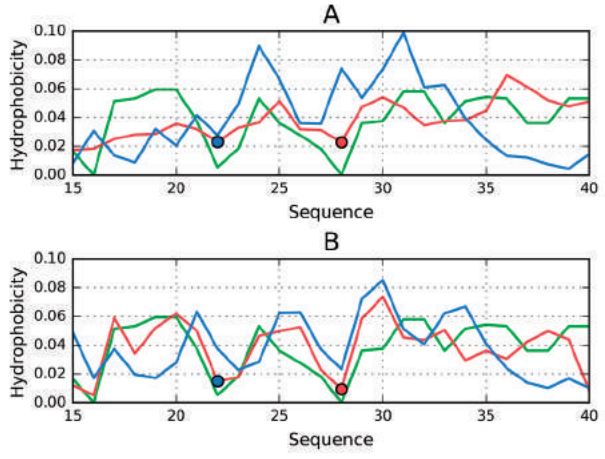


Figure 13. Theoretical (T – blue), intrinsic (H – green) and observed (O – red) distributions of hydrophobicity in models which meet amyloid seed criteria regarding the Aβ(15-40) fragments treated as part of the protofibril. The two circles mark observed hydrophobicity values for residues 22Glu (blue) and 28Lys (red).
(**A**) Model I5. (**B**) Model F5.

fragment, which is expected to form a part of the hydrophobic core, disrupted by 28Lys residue. 3D visualization (Fig. 8) reveals that this residue is located close to the central part of the molecule (more so than residue 22Glu, which is clearly exposed on the surface).

Figure 10 highlights the location of residues 22Glu and 28Lys in the protofibril, enabling visual comparison.

Figure 4A (which illustrates the distribution of hydrophobicity in chains of the protofibril) shows that residue 22Glu is locally discordant, being low in hydrophobicity, whereas the model expects this residue to be found in a highly hydrophobic environment (towards the center of the molecule). The location of residue 28Lys, as shown on Fig. 10, is consistent with the model – this residue is exposed on the surface and takes no part in interactions with the neighboring protofibril. It should be noted that residue 22Glu, while appearing exposed, is in fact internalized, since the chain is adjacent to another polypeptide where residue 22Glu is located close by. Thus, when considering the protofibril as a whole, these locations should be occupied by more hydrophobic residues than Glu.

Table 1 presents the parameters of the entire set of input structures. If we assume that high values of RD (in either model), negative HvT and TvO coefficients, and high values of the HvO coefficients are all indicative of an amyloid-like conformation, then the diagrams shown in Fig. 11 and 12 clearly identify the respective fragments as amyloid seeds.

In the final structures returned by the programs, the amino acids at position 15-20 (6 aa fragment) contains amyloid components in both I5 and F5 models, at least according to the fuzzy oil drop model (Fig. 11, Table 1).

The 21-26 fragment satisfies the criteria under which a given peptide may be regarded as amyloid-like, both in chain S in fibril (Af) and in chain S treated as individual (Ai) (Fig. 12). None of the other structures follow this pattern, which means that in their case the analyzed fragment is not an amyloid seed (although the structures F4 and I3 come close). The remaining fragments do not meet the stated criteria and are not regarded as amyloid seeds.

Figure 13 reveals a discordance between distributions in the 15-20 fragment as it is observed in I5 and F5, which, considering the presented analysis, suggests the presence of a seed of amyloidogenecity.

Structure of the superfibril interface

An additional question related to amyloid structures concerns the formation of superfibrils. Perhaps the most interesting example is the tau amyloid, which is capable of adopting two distinct conformations, differentiated by a mutual alignment of participating protofibrils (Eisenberg *et al.,* 2017).

In case of 2MPZ we observe an arrangement of three individual fibrils, adopting a highly symmetrical (equilateral triangle) form. The fuzzy oil drop model provides an explanation of the mechanisms responsible for this structure. If we calculate RD value for the residues which form part of the stabilizing interface, it turns out that the interface is consistent with the theoretical model of hydrophobicity. This implies that if the entire superfibril is encapsulated in a suitably shaped 3D Gaussian capsule, the placement of residues which comprise the interface is consistent with the fuzzy oil drop model (specifically, the following values were obtained: RD(T-O-R)=0.364; RD(T-O-H)=0.123; HvT=0.459; TvO=0.749; HvO=0.847). This set of values suggests high consistency with the model – in particular, the low value of RD(T-O-H) indicates that the interface emerges as a result of cooperation between individual units and not through intrinsic hydrophobicity. Notably, the interface area stretches along the entire complex, parallel to the fibril. It meshes with the linear aggregation logic that governs the propagation of amyloid fibrils, remaining consistent with the fuzzy oil drop model at the same time. This selective adherence to the model explains the obtained low values of RD for the superfibril, listed in the Results section.

CONCLUSIONS

Summarizing the presented results, it may be useful to inquire why a polypeptide capable of adopting a globular conformation forms an elongated fibril instead. If we base our analysis on the fuzzy oil drop model, the answer is that the fibril emerges as a result of changes in the external (environmental) force field, which, under ordinary conditions, would guide the folding process to produce a globular protein (note that a vast majority of protein domains is consistent with the fuzzy oil drop model and includes a prominent hydrophobic core (Sałapa *et al.,* 2012). The natural environment for Aβ(1-42) fragment is membrane environment. This is why the peptide deprived of its permanent chaperone adopts the unusual structural form. The water environment in its standard order shall be able to direct the folding toward the centric hydrophobic core generation. The final structure of proteins reflects a consensus between internal (inter-atomic) and external forces, the latter being exerted by the aqueous solvent. It seems that the linear propagation of bands of variable hydrophobicity is caused by the alignment between the actual distribution of hydrophobicity and the intrinsic properties of individual residues, with a limited influence of the environment.

The role of the environment is reflected by the properties of superfibril, which – as suggested by its RD values – emerges as a result of the interactions between protofibrils and the aqueous solvent (Brumshtein *et al.,* 2014).

Structures generated using I-Tasser (ranging from highly accordant globules to structures strongly deviating from the theoretical model) suggests that the polypeptide – guided only by its sequence (and, as reflected in the I-Tasser algorithms – by internal interactions between constituent atoms) – may adopt highly variable final conformations. This is further confirmed by the results of the CASP challenge, in which the force field (i.e. the same algorithm) – applied by a particular participant - produced a variety of results when applied to a specific protein. Applying a force field, which directs the folding process towards the generation of a hydrophobic core, also fails to produce a uniform answer – the diversity of the resulting forms (especially regarding the properties of their hydrophobic cores) suggests that the interplay between internal and environmental forces is not accurately captured by either model.

We have discussed other forms of Aβ(1-42) (paper in preparation) and tau-amyloid (Dułak *et al.,* 2018) to analysis similar to the one presented above and we obtained results that are consistent with previous observations in respect to identification of the polypeptide chain fragments which, according to the analysis, seem to play the role of seed for amyloid transformation. The 22Glu-28Lys fragment seems to be the main candidate for amyloid transformation in Aβ(1-42). The results of

presented analyses appear consistent – it turns out that in addition to proteins highly consistent with the fuzzy oil drop model (Banach *et al.,* 2014; Dygut *et al.,* 2016) we can also identify proteins in which the hydrophobic core is locally deformed (Kalinowska *et al.,* 2017), either by local hydrophobicity deficiency (corresponding to the ligand or substrate binding cavity (Banach *et al.,* 2012a) or local excess of hydrophobicity (corresponding to the complexation interface, in the case of proteins which have a quaternary conformation (Banach *et al.,* 2012b). Of particular note are the antifreeze proteins which exhibit broad structural variability, likely associated with their biological role – i.e. disrupting the aqueous environment and thereby preventing formation of ice crystals (Banach *et al.,* 2018).

The central tenet of the fuzzy oil drop model is that the aqueous solvent generates an external force field, affecting the protein in a continuous fashion (rather than as a collection of individual water molecules). The effects can be observed in the case of membranes, which self-organize to produce flat structures consisting of identical pieces. It appears that environmental influence is critical for the formation of amyloid structures, since proper folding depends upon "normal" structuralization of water. This view is based on the observation that environmental changes alone are sufficient to transform the protein from a globule to an amyloid fibril, as observed when shaking samples. Notably, shaking increases the phase transition surface area, leading to increased aeration of the solvent. This non-chemical process alters the structure of the environment, potentially impacting the conformation of solvated proteins. There is ongoing research on the subject, which may explain the peculiar effects of the environment upon the properties of proteins (and upon life in general) (Kim *et al.,* 2017), including the observed levitation of water molecules on top of hydrophobic surfaces (Schutzius *et al.,* 2015).

Acknowledgements

The authors wish to thank Piotr Nowakowski and Anna Śmietańska for editorial assistance.

Acknowledgements of Finnacial Support

This research was financially supported by Jagiellonian University Collegium Medicum under grant no. K/ZDS/006363. The presented studies were carried out in part using the PLGrid infrastructure at Cyfronet AGH, University of Science and Technology, 30-059 Kraków, al. Mickiewicza 30, Poland.

REFERENCES

Banach M, Konieczny L, Roterman I (2012a) Ligand-binding site recognition. In *Protein folding in silico – Protein folding versus protein structure prediction*, Roterman-Konieczna I ed, Woodhead Publishing Oxford Cambridge Philadelphia New Delhi (Currently Elsevier), pp 79–94. doi: 10.1533/9781908818256.79

Banach M, Konieczny L, Roterman I (2012b) Use of the "fuzzy oil drop" model to identify the complexation area in protein homodimers. In *Protein folding in silico – Protein folding versus protein structure prediction*, Roterman-Konieczna I ed, Woodhead Publishing Oxford Cambridge Philadelphia New Delhi (Currently Elsevier), pp 95–122. doi: 10.1533/9781908818256.95

Banach M, Kalinowska B, Konieczny L, Roterman I (2018) Possible mechanism of amyloidogenesis of V domains. In *Self-assembled molecules – new kind of protein ligands – Supramolecular ligands*, Roterman I, Konieczny L eds, Springer Cham, pp 77–100. doi: 10.1007/978-3-319-65639-7

Banach M, Konieczny L, Roterman I (2014) The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. *J Theor Biol* **359**: 6–17. doi: 10.1016/j.jtbi.2014.05.007

Banach M, Konieczny L, Roterman I (2018) Why do antifreeze proteins require a solenoid? *Biochimie* **144**: 74–84. doi: 10.1016/j.biochi.2017.10.011

Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* **91**: 43–56. doi: 10.1016/0010-4655(95)00042-e

Brumshtein B, Esswein SR, Landau M, Ryan CM, Whitelegge JP, Phillips ML, Cascio D, Sawaya MR, Eisenberg DS (2014) Formation of amyloid fibers by monomeric light chain variable domains. *J Biol Chem* **289**: 27513–27525. doi: 10.1074/jbc.M114.585638

Bystroff C, Shao Y (2004) Modeling protein folding pathways. In *Practical Bioinformatics*, Bujnicki J ed, pp. 97–122. Springer Berlin Heidelberg. doi: 10.1007/978-3-540-74268-5_5

Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* **338**: 1042–1046. doi: 10.1126/science.1219021

Dulak D, Gadzala M, Banach M, Ptak M, Wiśniowski Z, Konieczny L, Roterman I (2018) Filamentous aggregates of tau proteins fulfil standard amyloid criteria provided by the fuzzy oil drop (FOD) model. *Int J Mol Sci* **19**. pii: E2910. doi: 10.3390/ijms19102910

Dygut J, Kalinowska B, Banach M, Piwowar M, Konieczny L, Roterman I (2016) Structural interface forms and their involvement in stabilization of multidomain proteins or protein complexes. *Int J Mol Sci* **17**: 1741. doi: 10.3390/ijms17101741

Eisenberg DS, Sawaya MR (2017) Taming tangled tau. *Nature* **547**: 170–171. doi: 10.1038/nature23094

Fu Z, Aucoin D, Davis J, Van Nostrand WE, Smith SO (2015) Mechanism of Nucleated Conformational Conversion of Aβ42. *Biochemistry* **54**: 4197–41207. doi: 10.1021/acs.biochem.5b00467

Ghozlane A, Joseph AP, Bornot A, de Brevern AG (2009) Analysis of protein chameleon sequence characteristics. *Bioinformation* **3**: 367–369. doi: 10.6026/97320630003367

Kalinowska B, Banach M, Wiśniowski Z, Konieczny L, Roterman I (2017) Is the hydrophobic core a universal structural element in proteins? *J Mol Model* **23**: 205. doi: 10.1007/s00894-017-3367-z

Kalinowska B, Banach M, Konieczny L, Roterman I (2015) Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins. *Entropy* **17**: 1477–1507. doi: 10.3390/e17031477

Khoury GA, Liwo A, Khatib F, Zhou H, Chopra G, Bacardit J, Bortot LO, Faccioli RA, Deng X, He Y, Krupa P, Li J, Mozolewska MA, Sieradzan AK, Smadbeck J, Wirecki T, Cooper S, Flatten J, Xu K, Baker D, Cheng J, Delbem AC, Floudas CA, Keasar C, Levitt M, Popović Z, Scheraga HA, Skolnick J, Crivelli SN; Foldit Players (2014) WeFold: a coopetition for protein structure prediction. *Proteins* **82**: 1850–1868. doi: 10.1002/prot.24538

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**(Web Server), W526–W531. doi: 10.1093/nar/gkh468

Kim KH, Späh A, Pathak H, Perakis F, Mariedahl D, Amann-Winkel K, Sellberg JA, Lee JH, Kim S, Park J, Nam KH, Katayama T, Nilsson A (2017) Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science* **358**: 1589–1593. doi: 10.1126/science.aap8269

Kister A (2015) Amino acid distribution rules predict protein fold: protein grammar for beta-strand sandwich-like structures. *Biomolecules* **5**: 41–59. doi: 10.3390/biom5010041

Konieczny L, Brylinski M, Roterman I (2006) Gauss-function-Based model of hydrophobicity density in proteins. *In Silico Biol* **6**: 15–22

Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* **22** 79–86. doi: 10.1214/aoms/1177729694

Levitt MA (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* **104** 59–107. doi: 10.1016/0022-2836(76)90004-8

Li W, Kinch LN, Karplus PA, Grishin NV (2015) ChSeq: A database of chameleon sequences. *Protein Sci* **24**: 1075–1086. doi: 10.1002/pro.2689

Liwo A, Khalili M, Scheraga HA (2005) Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc Natl Acad Sci U S A* **102**: 2362–2367. doi: 10.1073/pnas.0408885102

Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Döbeli H, Schubert D, Riek R (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A* **102**: 17342–17347. doi: 10.1073/pnas.0506723102

Macias-Romero C, Nahalka I, Okur HI, Roke S (2017) Optical imaging of surface chemistry and dynamics in confinement. *Science* **357**: 784–788. doi: 10.1126/science.aal4346

Monzon AM, Zea DJ, Marino-Buslje C, Parisi G (2017) Homology modeling in a dynamical world. *Protein Sci* **26**: 2195–2206. doi: 10.1002/pro.3274

Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins* **84**: 4–14. doi: 10.1002/prot.25064

Roterman I, Banach M, Konieczny L (2017) Application of the fuzzy oil drop model describes amyloid as a ribbonlike micelle. *Entropy* **19**: 167. doi: 10.3390/e19040167

Roterman I, Banach M, Kalinowska B, Konieczny L (2016) Influence of the aqueous environment on protein structure – a plausible hypothesis concerning the mechanism of amyloidogenesis. *Entropy* **18**: 351. doi: 10.3390/e18100351

Sałapa K, Kalinowska B, Jadczyk T, Roterman I (2012) Measurement of hydrophobicity distribution in proteins – non-redundant protein Data Bank. *Bio-Algorithms and Med-Systems* **8**: 327–338. doi: 10.2478/bams-2012-0023

Saunders JA, Gibson KD, Scheraga HA (2001) Ab initio folding of multiple-chain proteins. *Biocomputing 2002* 601–612. doi: 10.1142/9789812799623_0056

Schutzius TM, Jung S, Maitra T, Graeber G, Köhme M, Poulikakos D (2015) Spontaneous droplet trampolining on rigid superhydrophobic surfaces. *Nature* **527**: 82–85. doi: 10.1038/nature15738

Sgourakis NG, Yau WM, Qiang W (2015) Modeling an in-register, parallel "iowa" aβ fibril structure using solid-state NMR data from labeled samples with rosetta. *Structure* **23**: 216–227. doi: 10.1016/j.str.2014.10.022

Tycko R (2011) Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* **62**: 279–299. doi: 10.1146/annurev-physchem-032210-103539

Xu S (2009) Cross-beta-sheet structure in amyloid fiber formation. *J Phys Chem B* **113**: 12447–12455. doi: 10.1021/jp903106x

Yang J, Zhang Y (2015) Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinformatics* **52**: 5.8.1–5.815. doi: 10.1002/0471250953.bi0508s52

Younkin SG (1998) The role of A beta 42 in Alzheimer's disease. *J Physiol Paris* **92**: 289–292. doi: 10.1016/s0928-4257(98)80035-1

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40. doi: 10.1186/1471-2105-9-40

Zhou H, Skolnick J (2010) Improving threading algorithms for remote homology modeling by combining fragment and template comparisons. *Proteins* **78**: 2041–2048. doi: 10.1002/prot.22717

Załącznik nr. 3

Structural analysis of the Aβ(11–42) amyloid fibril based on hydrophobicity distribution

IRENA ROTERMAN, DAWID DUŁAK, MAŁGORZATA GADZAŁA,

MATEUSZ BANACH, LESZEK KONIECZNY

Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-019-00209-9

Structural analysis of the Aβ(11–42) amyloid fibril based on hydrophobicity distribution

Irena Roterman[1] · Dawid Dułak[2] · Małgorzata Gadzała[3,4] · Mateusz Banach[1] · Leszek Konieczny[5]

Received: 6 August 2018 / Accepted: 31 May 2019

Abstract

The structure of the Aβ(11–42) amyloid available in PDB makes possible the molecular analysis of the specificity of amyloid formation. This molecule (PDB ID 2MVX) is the object of analysis. This work presents the outcome of in silico experiments involving various alternative conformations of the Aβ(11–42) sequence, providing clues as to the amylodogenecity of its constituent fragments. The reference structure (PDB) has been compared with folds generated using I-Tasser and Robetta—the strongest contenders in the CASP challenge. Additionally, a polypeptide which matches the Aβ(11–42) sequence has been subjected to folding simulations based on the fuzzy oil drop model, which favors the production of a monocentric hydrophobic core. Computer simulations yielded 15 distinct structural forma (five per software package), which, when compared to the experimentally determined structure, allow us to study the role of structural elements which cause an otherwise globular protein to transform into an amyloid. The unusual positions of hydrophilic residues disrupting the expected hydrophobic core and propagating linearly along the long axis of fibril is recognized as the seed for amyloidogenic transformation in this polypeptide. This paper discusses the structure of the Aβ(11–42) amyloid fibril, listed in PDB under ID 2MXU (fragment od Aβ(1–42) amyloid).

Keywords Amyloidosis · Fibrillation · Aβ(11–42) · Aβ(1–42)

✉ Irena Roterman
myroterm@cyf-kr.edu.pl

Dawid Dułak
dawid.dulak@gmail.com

Małgorzata Gadzała
m.k.gadzala@gmail.com

Mateusz Banach
mateusz.banach@uj.edu.pl

Leszek Konieczny
mbkoniec@cyf-kr.edu.pl

1 Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Łazarza 16, 31-530 Kraków, Poland

2 ABB Business Services Sp. z o.o., Żegańska 1, 04-713 Warsaw, Poland

3 ACK – Cyfronet AGH, Nawojki 11, 30-150 Kraków, Poland

4 Present Address: Schibsted Tech Polska Sp. z o. o., Armii Krajowej 28, 30-150 Kraków, Poland

5 Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kopernika 7, 31-034 Kraków, Poland

Introduction

The volume of published papers which focus on amyloids is rapidly growing. Literature reviews [1] provide an up-to-date overview of current trends in protein misfolding research. This broad field encompasses various specific issues, such as the genetic underpinnings of amyloidogenesis and molecular studies [2], including protein folding simulations. Analysis of amyloid structures has to deal with the dynamics and flexibility of proteins, which must be capable of specific interactions with their intended ligands and substrates [3, 4]. The emergence of amyloids is linked to the peculiarities of the folding process, which still awaits a comprehensive theoretical description—despite many decades of research [5].

In order to study such phenomena, we require techniques which would enable us to track the intermediate phases of folding. While traditional NMR is a useful tool, it requires soluble molecules [6], and that presents a problem when studying the pathogenicity of amyloids [7–9]. Much progress has recently been made owing to introduction of solid-state NMR [10]. One example is the elucidation of the structure of Aβ(1–42), and particularly

Published online: 10 July 2019

of its Aβ(11–42) fragment, which is now listed in PDB [11]. This amyloid is the focus of the presented work. We subjected it to analysis from the point of view of hydrophobicity distribution. The presented work follows upon the results presented in [12], where we single out fragments exhibiting specific deviations from the theoretical ("idealized") distribution of hydrophobicity expected in a globular protein and mathematically defined by a 3D Gaussian [13–15].

The Gaussian distribution peaks at the center of an ellipsoid capsule. Its values decrease along with distance from the center, reaching nearly 0 on the surface. When the size of the capsule is adjusted to encapsulate the molecule in question, the corresponding Gaussian yields the theoretical (expected) values of hydrophobicity at any point within the protein body. This theoretical distribution is subsequently compared with the observed distribution of hydrophobicity, which depends on inter-residual interactions (themselves dependent on the mutual separation and intrinsic hydrophobicity of interacting residues). Differences between both distributions manifest themselves as either excess hydrophobicity or hydrophobicity deficiencies in specific areas of the protein body. The former—if present on the surface—mark protein complexation interfaces [16], while the latter are typically associated with ligand binding cavities [17] which enable the protein to perform its biological function [18].

To-date studies of amyloids point to a specific disagreement between the observed distribution of hydrophobicity and the theoretical (monocentric) Gaussian. In place of a central core, we are faced with linear propagation of repetitive patterns stretching along the fibril's main axis. Several fragments of the Aβ(1–42) polypeptide have been identified as amyloid seeds—this includes the fragments at 11–16, 16–22 and 22–28 [12]. Consequently, our analysis will focus on these fragments in particular.

The Aβ(11–42) sequence has been used as input for protein folding simulation toolkits, including I-Tasser [19–22] and Robetta [23, 24]—both highly ranked in the CASP competition [25, 26]. Structures produced by both packages were assessed from the point of view of hydrophobicity distribution. In this respect, they proved to be highly diverse, ranging from near-globular to amyloid-like. In addition, we also generated several reference structures using software based on the fuzzy oil drop model (FOD), augmenting optimization of nonbonding interactions with alignment with an external (Gaussian) hydrophobic force field.

Each of the presented toolkits produced five structures, enabling us to compile a ranking list sorted by increasing differences between the observed structure and the theoretical Gaussian. This, in turn, shows how progressive deviations from the monocentric hydrophobic core model eventually cause the polypeptide chain to transform into a fibril, marked by alternating bands of high and low hydrophobicity. This analysis can be related to the reported polymorphism of amyloid structures [27].

This is why this paper can be treated as in silico experiment.

Materials and methods

Protein under consideration

Our analysis focused on the Aβ(11–42) amyloid listed in PDB under ID 2MXU [11]. It represents the Aβ(1–42) protein devoid of its N-terminal fragment. The PDB structure comprises 12 separate polypeptides arranged into a fibril. This structural form is treated as reference one for all models delivered by mentioned programs. The status of complete fibrils, status of the chain as part of the fibril and one selected chain treated as individual structural unit are taken as the reference objects for those generated in silico.

Folding of Aβ(11–42)

The Aβ(11–42) polypeptide was subjected to in silico folding simulations using two state-of-the-art protein folding toolkits: I-Tasser [19–22] and Robetta [23, 24]. Both packages consistently obtain high marks in the CASP [25] challenge and have been singled out as the most reliable computational tools currently available for this purpose [22, 23]. The diversity of the resulting structures has its roots in the reported polymorphism of amyloid forms [27].

In addition to the above, our study set was extended with five structures generated by a custom toolkit based on the fuzzy oil drop model, which acknowledges the influence of the aqueous solvent upon the chain in question. This influence is modeled as an external force field with a Gaussian distribution, promoting internalization of hydrophobic residues along with exposure of hydrophilic residues on the surface [13–15]. As the application requires a starting structure (referred to as the Early Stage intermediate) we used the 2MXU structure as its input.

All resulting folds (15 in total; five per application—pursuant to CASP criteria), along with the structure available in PDB, were subject to analysis of the shape of their hydrophobic cores in two separate contexts: as a composite fibril and as an individual chain from the 2MXU file.

FOD model was applied for protein folding for selected targets available in CASP6 (2004) [24]. We submitted models for 23 targets with the highest DTT_TS value equal to 41.98.

Fuzzy oil drop model (FOD)

The status of the hydrophobic core is expressed using the RD value which bases on the fuzzy oil drop model. This coefficient is calculated separately for two distinct variants: RD(T-O-R), which compares the observed distribution (O) with two reference distributions: theoretical (T) and uniform (R), and RD(T-O-H), where the uniform distribution is replaced by a distribution which reflects the distribution expressed by intrinsic hydrophobicity of each residue. The definition of these parameters will be given below in the next part of Materials and Methods. In addition, three correlation coefficients of the above hydrophobicity distributions are computed: H versus T, T versus O and H versus O called as HvT, TvO and HvO in this paper.

Since a detailed presentation of the fuzzy oil drop model can be found in [13–15] we will limit ourselves to a brief recapitulation of its core concepts.

The input molecule is encapsulated in an ellipsoid used to calibrate the 3D Gaussian function, which, in turn, yields theoretical values of hydrophobicity at arbitrary points in the protein body. In contrast, the observed distribution depends on inter-residual interactions (as described in [28]). Both distributions (T and O) are measured at specific points which correspond to the so-called effective atoms (averaged-out positions of all atoms comprising each residue). In addition, each residue is assumed to represent certain intrinsic hydrophobicity as listed in [19], which is also used in observed hydrophobicity calculations.

In order to meaningfully compare alternative distributions, we apply the so-called Kullback–Leibler divergence entropy formula [29]. Since the result produced by this formula is a measure of entropy, it cannot be interpreted on its own—instead, it requires a reference value. This is why, in addition to T, we introduce another reference distribution denoted R, which stands from random. It is a uniform distribution which assigns every residue under consideration a hydrophobicity of 1/N where is N the number of these residues. Under these assumptions the RD value expresses the "closeness" of O to either T or R. Since—as discussed in [12]—in amyloid structures the observed distribution is dominated by the intrinsic properties of each participating residue, we also define another type of reference, denoted H, which reflects the intrinsic hydrophobicity of each amino acid in the input chain. This results in two distinct values of RD: one for the T-O-R variant and one for the T-O-H variant. In the first case, a RD value greater or equal than 0.5 is taken as indication that protein's observed hydrophobicity profile does not follow the 3D Gaussian distribution, while in the other—that it explicitly follows the intrinsic model (H).

As already mentioned, our comparative analysis also relies on three distinct correlation coefficients, providing a pairwise comparison of all distributions: HvT, TvO and HvO (v stands from "versus"). Together, these coefficients express the influence of intrinsic hydrophobicity upon the structure of the hydrophobic core.

Globular proteins are usually closely aligned with T (this is particularly true for domains [30]), which means that hydrophobic residues are internalized while hydrophilic residues appear on the surface. Such conditions emerge as a result of "cooperation" between residues in an attempt to produce a common core. On the other hand, when residues act in a "selfish" manner, without cooperative tendencies, the result is a high value of HvO coupled with low (or even negative) values of both TvO and HvT. This suggests that the structure does not contain a monocentric core, and instead may exhibit other—in case of amyloid: linear propagation of repetitive patterns of hydrophobicity. Consequently, analysis of RD (together with the aforementioned correlation coefficients) may reveal progressive dilution of the hydrophobic core in favor of an entirely different structural pattern.

In order to further identify strongly amyloidogenic fragments, the above coefficients were also calculated for specific fragments of the input polypeptide.

Folding simulations were performed using computational resources provided by the Cyfronet AGH—Academic Computing Center within the PL-Grid infrastructure. In fuzzy oil drop simulations the optimization of nonbonding interactions was carried out using the GROMACS package (also provided by Cyfronet) [31, 32].

Results

Structure of the Aβ(11–42) fibril

The structure of the Aβ(11–42) fibril may be assessed on the basis of T and O hydrophobicity distributions, as illustrated in Fig. 1. The figure reveals typical discordance between both distributions, with the observed hydrophobicity remaining high on the surface of the complex (contrary to expectations). In addition, a characteristic sinusoidal pattern is observed—the hallmark of a complex comprising multiple identical subunits arranged in a linear fashion, with alternating bands of high and low hydrophobicity exposed on the surface.

RD and correlation coefficients characterizing the fibrillary form will be discussed further on, in conjunction with the analysis of structures generated by in silico folding models.

The distribution of T shown in Fig. 1 reveals the characteristic concentration of hydrophobicity in the central part of fibril which is not followed by O distribution which represents sinusoid-like distribution along the whole fibril. The red lines do not represent the different O distributions. Only

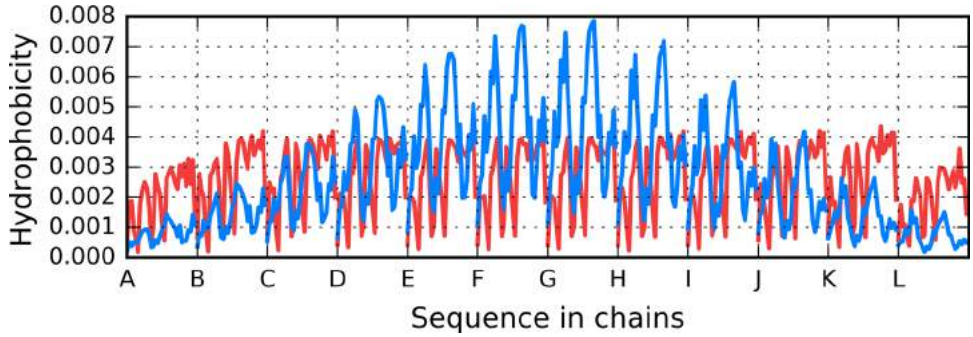


Fig. 1 Theoretical (T: blue) and observed (O: red) hydrophobicity distributions in the Aβ(11–42) fibril

two O profiles can bees in Fig. 1a. This is due to overlapping of almost identical profiles for central polypeptides. Elimination of border polypeptide chains (chains A and L) visualizes it very clear (Fig. 2b). The T distributions for these polypeptides still represent different form depending of the position of polypeptide chain under consideration.

Since our analysis involves a complex which consists of a finite number of peptides, it does not accurately reflect the theoretical capability for unrestricted propagation. For this reason, we have singled out chain F (the central one) as representative of the fibril's structure. This chain was subsequently analyzed from two perspectives: as a part of the fibril (with 3D Gaussian fitted to the whole complex) and as an individual molecule (with 3D Gaussian fitted only to this chain and with disregard of other chains during O profile calculations). Figures 3a and 4a illustrate the theoretical, observed and intrinsic distributions for chain F. Besides sections where T is somewhat aligned with O (as well as with H), there are areas where both distributions diverge notably. In order to identify these discordant sections, we calculated HvT, TvO and HvO correlation coefficients for individual 5 aa fragments using a moving frame approach, as shown in Fig. 3b and in Fig. 4b. These charts reveal fragments where all three coefficients adopt relatively high values. These fragments (residues 5 through 11) may be regarded as accordant with the 3D Gaussian model—they evidence the sort of cooperation which is required throughout the whole protein for a monocentric hydrophobic core to emerge.

The remaining fragments exhibit discordance which, in extreme cases, may produce a conformation which is a polar opposite of theoretical predictions (negative values of TvO and HvT). In these cases, the observed distribution is driven by the "selfish" tendencies of each residue rather

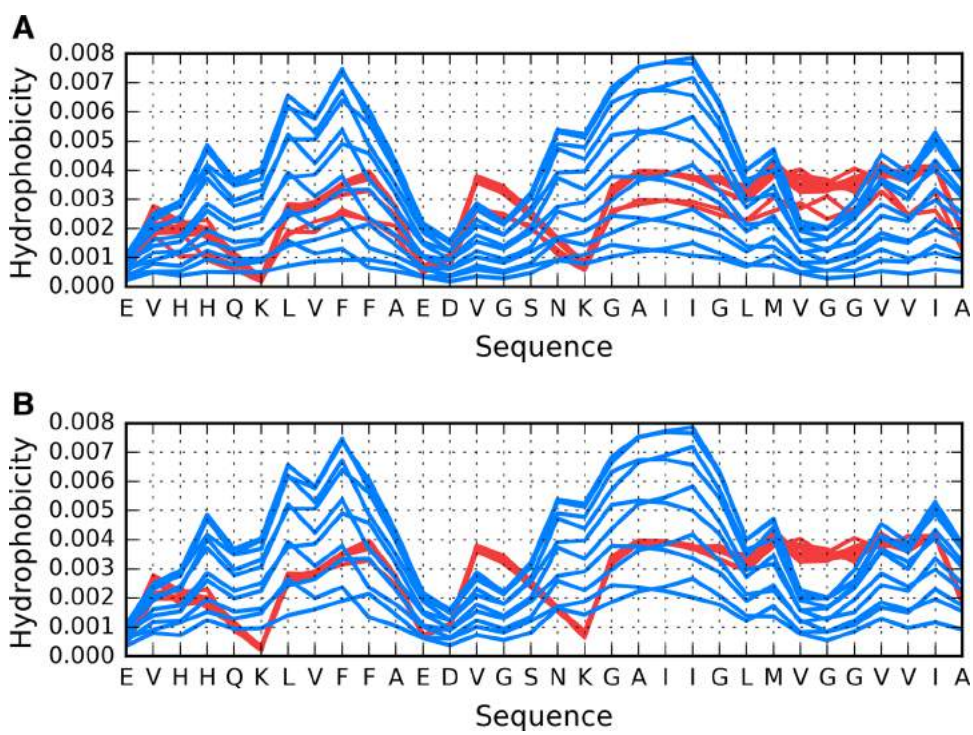


Fig. 2 Theoretical (T: blue) and observed (O: red) hydrophobicity distributions in the Aβ(11–42) fibril presented in an overlapped mode (with all chains sharing the X axis): **A** all chains present in the fibril; **B** without one outlying chain from each end of the structure

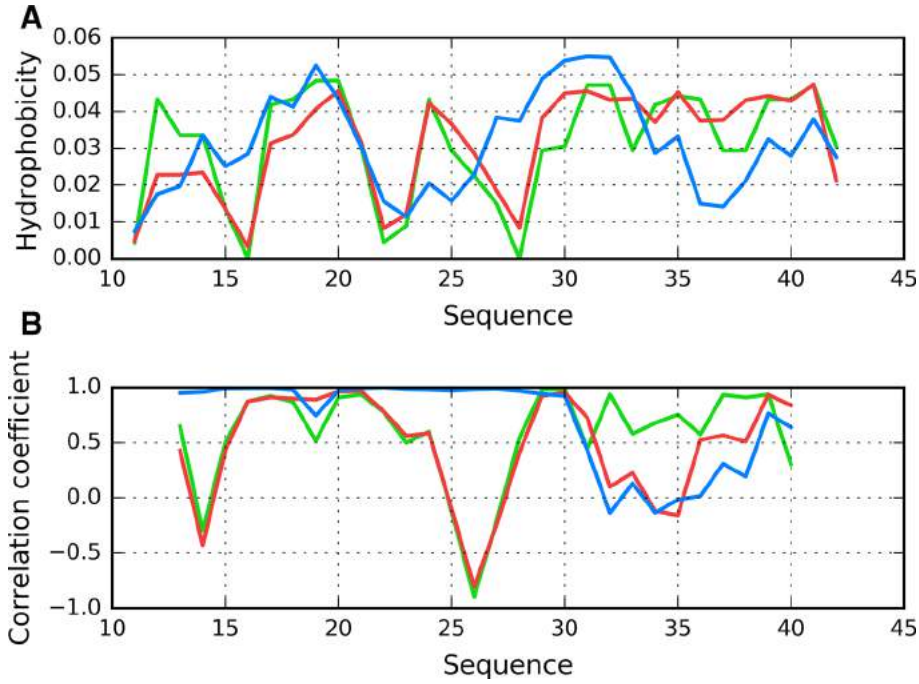


Fig. 3 Chain F analyzed as a component of the fibril: **A** T (blue), O (red) and H (green) hydrophobicity distributions; **B** correlation coefficients (HvO: blue, HvT: red, TvO: green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i. e. 20 corresponds to residue 20 in the 18–19–20–21–22 frame)

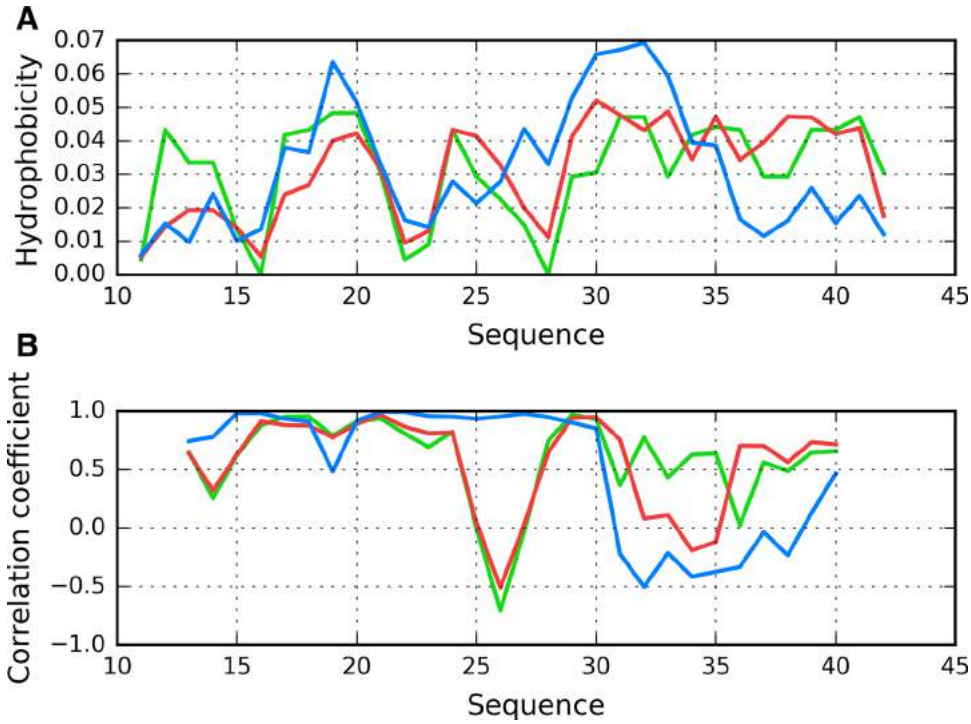


Fig. 4 Chain F analyzed as an individual molecule: **A** T (blue), O (red) and H (green) hydrophobicity distributions; **B** correlation coefficients (HvO: blue, HvT: red, TvO: green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i. e. 20 corresponds to residue 20 in the 18–19–20–21–22 frame)

than by any tendency to produce a common hydrophobic core. Accordingly, we singled out fragments 11–16 and 24–28 as dominated by intrinsic hydrophobicity, while the 18–23 fragment may be described as locally accordant. Other fragments, while not exhibiting a clear preference for either T or H, are included in our analysis to enable comparisons with structures produced by in silico tools. Their variable status (from globular to amyloid-like) will be discussed on the basis of RD and correlation coefficients.

Profiles observed for both views of chain F similarly reveal the presence of unexpected local maximum in the localization of expected local minimum in fragment 22–28. The discordant position of Lys16 and particularly Lys28 destroy the expected maximum introducing local minimum. Taking into account that this characteristics is continued

along the whole fibril. This is why the positions of lysines play the contradictory role in respect to centralization of hydrophobicity. Calculation for both forms reveals the unusual distribution which can be even treated as contradictory status in respect to usual characteristics observed in globular proteins.

Comparative analysis of protein structures generated by in silico folding models

Structures generated by I-Tasser are labeled "I", those generated by Robetta are labeled "R", while those produced by the fuzzy oil drop software are labeled "F". All structures are numbered, with their respective numbers affixed to the source label. Results are listed in order of increasing values of RD(T-O-R) for the whole molecule, meaning that F1 is the FOD output with lowest value, R1—Robetta's, and so on.

Table 1 shows the full range of RD and correlation coefficients calculated for the entire molecule and for selected fragments. The list also includes chain F seen as a part of the fibril (labeled Af—amyloid/fibril, to avoid confusion with FOD results) and as an individual molecule (Ai—amyloid/individual). Table 1 also shows the status of the 29–42 fragment, even though—while somewhat disordered—it does not resemble an amyloid seed [12].

According to Fig. 5, in models F1–R2, status of the 11–16 fragment is consistent with the 3D Gaussian distribution. This property, typical for globular proteins, fails to hold for model I1. In contrast, models R5 and I5 exhibit clear amyloid-like characteristics. Their HvO coefficients reach very high values, while the remaining coefficients are negative, indicating strong influence of intrinsic hydrophobicity. The set includes chain F analyzed as part of the complex (Af). Notably, under FOD classification criteria, model I4 exhibits the strongest amyloid affinity of the 11–16 fragment from among all analyzed structures. Its hydrophobicity profiles and 3D representation are shown on Fig. 6.

The summary shown in Fig. 7 shows that the 16–22 fragment in structure I3 exhibits amyloid-like characteristics, while structures F1–R4 are generally consistent with the theoretical distribution.

Structures shown in Fig. 8 visualize the hypothesis that the folding following intrinsical hydrophobicity directs the process toward the amlyoid-like structural forms. Two structural forms compared in Fig. 8 show that the 16–22 fragment may adopt a helical conformation, yet in an amyloid it becomes beta-like. In addition to that, these two models prove that the Aβ(11–42) sequence may, in fact, produce a globule which can be seen to break apart as it transformation into an uncoiled loop (Fig. 8b) on its way to the amyloid form.

As illustrated in Fig. 9, the fragment at 24–28 becomes amyloid-like in structures Af (chain F in fibril) and I4. Analysis of distribution charts in this figure shows that in models F1, R1, R2 and R3 the fragment retains globular characteristics, while in model R4 it also resembles an amyloid. Sample 3D structures are visualized in Fig. 10.

From among the analyzed structures, F1's chain exhibits a particularly low RD(T-O-R) value, as shown in Fig. 11, where globular form was received with very little discordance between T and O profiles. In contrast, the highest RD(T-O-R) value was recorded for model I5's chain, with its T and O profiles illustrated in Fig. 12. To allow further analysis and comparison with other models, a selection of them is shown in Figs. 13 and 14. They visualize what kind of wide variety of structures can be produced by highly specialized programs.

In summary, we can single out structure I5 as potentially susceptible to linear propagation via complexation of identically folded polypeptides. Structure I5 (random coil structural form) also exhibits a propensity for complexation, however its distribution of hydrophobicity does not reveal the characteristic sinusoidal pattern observed in amyloids [12].

One can conclude that the main candidate for amyloid transformation in the Aβ(11–42) are two fragments: 11–16 and 24–28. It can be seen in Table 1: the underlined sets of parameters denote models close to amyloid form: I4 and R4. These positions do not satisfy the condition of high value of RD(T-O-H). High value of this parameter requires a multichain complex as it is rather unusual for an isolated chain. This conclusion is correct on the condition of fuzzy oil drop acceptance as the method to trace the hydrophobic core identification and its transformation. This is why conclusion can be limited just to such a case.

Discussion

Summing up the presented results, we can state that the Aβ(11–42) polypeptide may theoretically adopt various structural forms, including tightly packed globules characterized by high solubility (good agreement between O and T distributions). The set of candidate structures produced by I-Tasser and Robetta is highly diverse—from coherent globules all the way to disordered folds. Of particular note are the structured generated using the FOD model, where the presence of an external force field (aqueous solvent) drives the folding process towards the generation of a monocentric hydrophobic core—even though, under certain (so far unknown) conditions, this model may also produce strongly discordant structures.

Some of the obtained structures, e.g. I5 with RD(T-O-R) = 0.768, appears capable of forming complexes with other identically folded chains. The structure in question includes

Table 1 Hydrophobicity-based parameters characterizing the structure of chain F from 2MXU in following forms: Af—when analyzed as a part of the fibril, Ai—when analyzed as an individual molecule, I#—as a result of simulation with I-Tasser, R#—as a result of simulation with Robetta, F#—as a result of simulation with FOD model

FORM	RD 11–42		CC 11–42			RD 11–16		CC 11–16			RD 17–23	
	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H
F1	0.234	0.175	0.440	0.850	0.531	0.335	0.059	0.188	0.653	0.807	0.138	0.119
F2	0.240	0.181	0.252	0.866	0.370	0.184	0.081	0.309	0.961	0.486	0.194	0.155
F3	0.242	0.156	0.139	0.854	0.415	0.249	0.050	0.380	0.835	0.773	0.169	0.083
R1	0.256	0.260	0.589	0.776	0.626	0.275	0.115	0.807	0.828	0.977	0.330	0.326
R2	0.283	0.260	0.659	0.777	0.676	0.182	0.137	0.751	0.941	0.810	0.354	0.329
R3	0.320	0.280	0.447	0.746	0.716	0.417	0.123	0.376	0.485	0.953	0.387	0.384
I1	0.376	0.240	0.258	0.696	0.687	0.346	0.035	−0.208	0.717	0.257	0.304	0.248
I2	0.428	0.287	0.272	0.612	0.709	0.757	0.172	−0.512	−0.111	−0.597	0.473	0.348
R4	0.457	0.371	0.166	0.477	0.651	0.385	0.167	0.226	0.628	0.871	0.187	0.191
I3	0.487	0.341	0.167	0.468	0.640	0.707	0.189	−0.481	0.015	0.531	0.757	0.509
Ai	0.536	0.519	0.408	0.567	0.698	0.488	0.220	0.494	0.500	0.831	0.121	0.230
F4	0.555	0.285	−0.073	0.412	0.329	0.534	0.064	−0.533	0.566	0.084	0.447	0.197
F5	0.559	0.226	0.154	0.391	0.260	0.533	0.054	−0.005	0.190	0.606	0.874	0.275
R5	0.660	0.521	0.172	0.254	0.641	*0.655*	*0.201*	*−0.474*	*−0.075*	*0.739*	0.675	0.458
Af	0.680	0.756	0.246	0.363	0.821	0.506	0.562	0.121	0.257	0.966	0.133	0.473
I4	0.715	0.392	−0.262	0.415	0.311	*0.903*	*0.176*	*−0.469*	*−0.452*	*0.788*	0.464	0.166
I5	0.768	0.610	−0.138	0.014	0.558	0.822	0.170	0.259	−0.118	0.680	*0.826*	*0.720*

FORM	CC 17–23			RD 24–28		CC 24–28			RD 29–42		CC 29–42		
	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO
F1	0.632	0.939	0.716	0.269	0.251	0.845	0.866	0.965	0.264	0.189	−0.030	0.824	−0.307
F2	0.489	0.897	0.579	0.429	0.251	−0.159	0.562	0.559	0.260	0.174	−0.243	0.880	−0.526
F3	0.092	0.908	0.440	0.353	0.179	0.521	0.717	0.912	0.168	0.123	−0.386	0.910	−0.306
R1	0.706	0.731	0.739	0.116	0.074	0.910	0.974	0.972	0.537	0.360	−0.248	0.600	−0.057
R2	0.608	0.653	0.863	0.367	0.091	0.566	0.718	0.937	0.433	0.431	0.675	0.650	0.373
R3	0.303	0.637	0.846	0.129	0.057	0.906	0.957	0.978	0.322	0.264	0.213	0.872	0.136
I1	0.416	0.808	0.826	0.530	0.092	−0.167	0.243	0.854	0.618	0.675	0.259	0.674	0.514
I2	0.284	0.433	0.741	0.217	0.047	0.525	0.828	0.867	0.622	0.648	0.313	0.688	0.506
R4	0.766	0.959	0.632	*0.705*	*0.192*	*−0.641*	*−0.633*	*0.978*	0.575	0.486	−0.199	0.536	0.085
I3	−0.282	−0.314	0.847	0.336	0.121	0.187	0.829	0.526	0.576	0.528	0.544	0.703	0.317
Ai	0.907	0.925	0.888	*0.669*	*0.382*	*−0.509*	*−0.706*	*0.954*	0.864	0.789	0.100	0.504	0.110
F4	0.388	0.453	0.288	0.435	0.237	−0.499	0.464	0.435	0.711	0.582	−0.062	0.433	−0.260
F5	0.155	−0.762	−0.047	0.606	0.135	−0.633	0.019	0.628	0.536	0.322	−0.337	0.566	−0.739
R5	0.065	−0.089	0.665	0.226	0.037	0.726	0.872	0.952	0.810	0.784	0.266	0.410	0.322
Af	0.969	0.913	0.968	*0.731*	*0.557*	*−0.803*	*−0.898*	*0.982*	0.845	0.776	0.141	0.384	0.404
I4	−0.879	0.444	−0.194	*0.702*	*0.133*	*−0.787*	*−0.840*	0.949	0.765	0.660	−0.045	0.705	−0.104
I5	*−0.750*	*−0.454*	0.662	0.330	0.045	0.175	0.781	0.719	0.871	0.800	0.013	−0.022	−0.093

Sequential numbers (1–5) after I, R and F sort models obtained with given program in order of increasing value of RD(T-O-R) calculated for the whole chain (residues 11–42). Parameters are given as as follows: RD for T-O-R and T-O-H variants, correlation coefficients (CC) for relations H versus T (HvT), T versus O (TvO) and H versus O (HvO). Id of structures given in bold distinguish fragment classified as amyloid; values italic are examples of forms close to amyloid form, understood as those expressing negative CC value of HvT and/or TvO with high CC value of HvO and with high RD values (in both variants)

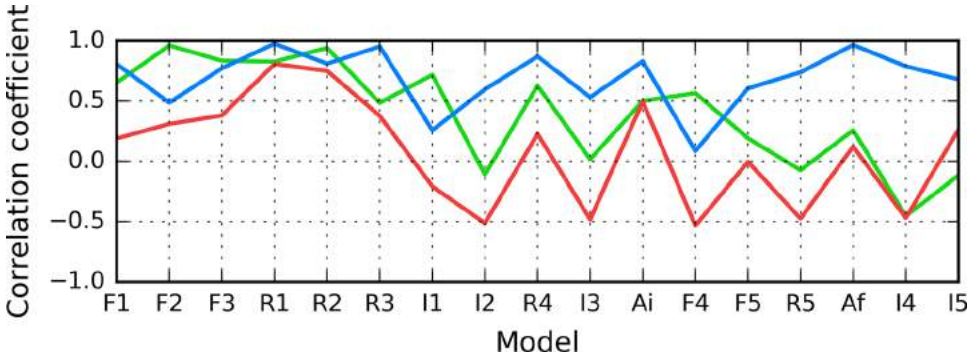


Fig. 5 HvO (blue), HvT (red), TvO (green) calculated for residues 11–16 in successive structures as listed in Table 1

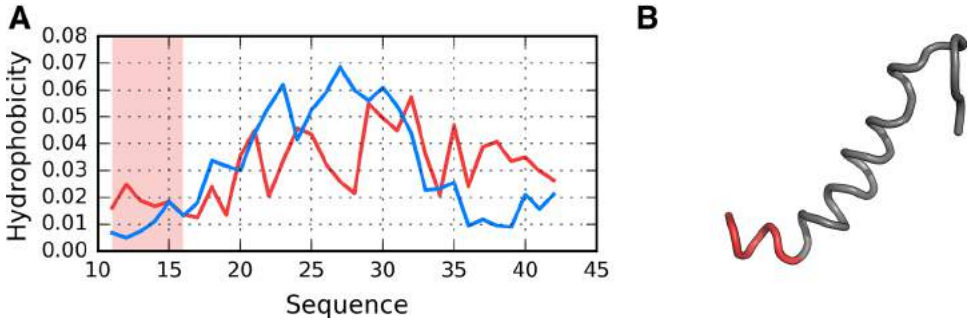


Fig. 6 Presentation of structure I4: **A** theoretical (T: blue) and observed (O: red) hydrophobicity profiles; **B** 3D view. Red highlight in **A** and red fragment in **B** correspond to 11–16 residue range

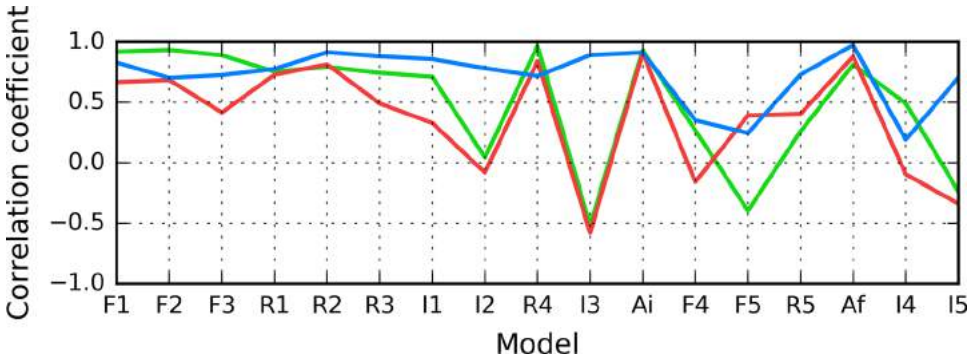


Fig. 7 HvO (blue), HvT (red), TvO (green) calculated for residues 16–22 in successive structures as listed in Table 1

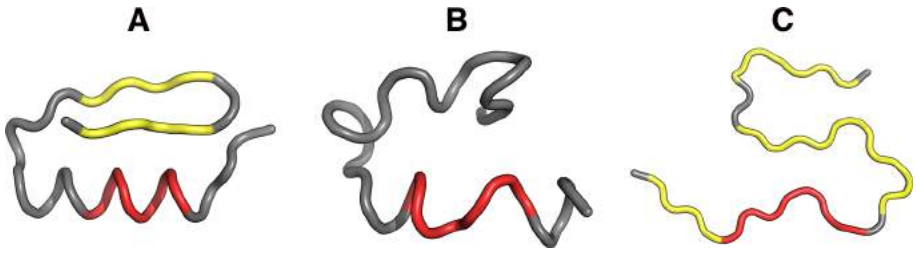


Fig. 8 3D view of structures R3 (**A**), I3 (**B**) and chain F (**C**) with 16–22 fragment colored red. Yellow fragments mark the locations of beta sheets (partially covered with red in **C**)

fragments which may be regarded as seeds of a conformational pattern guided by the intrinsic hydrophobicity of individual residues (negative correlation coefficients—see Table 1).

The interpretation provided in [33, 34] suggests that external conditions may support the misfolding not supporting the formation of centric hydrophobic core. The form of all hydrophobicity profiles as observed in amyloid forms analyzed here mark the positions of lysines as highly discordant in respect to what is expected by T distribution. The position of lysines introduces sharp local minimum in O in location where the local maximum is expected

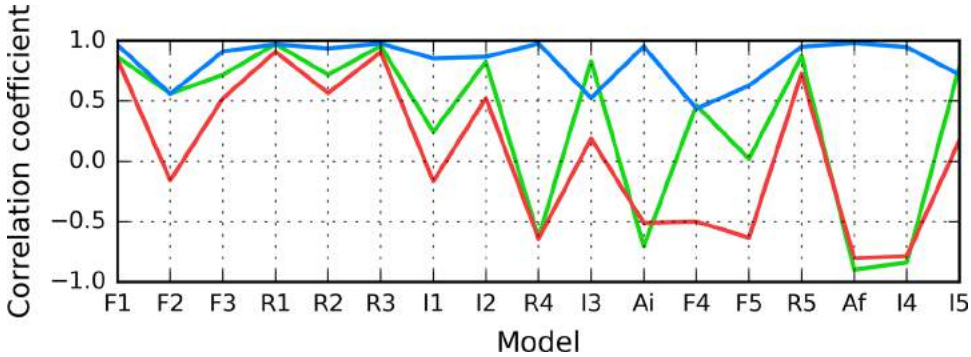


Fig. 9 HvO (blue), HvT (red), TvO (green) calculated for residues 24–28 in successive structures as listed in Table 1

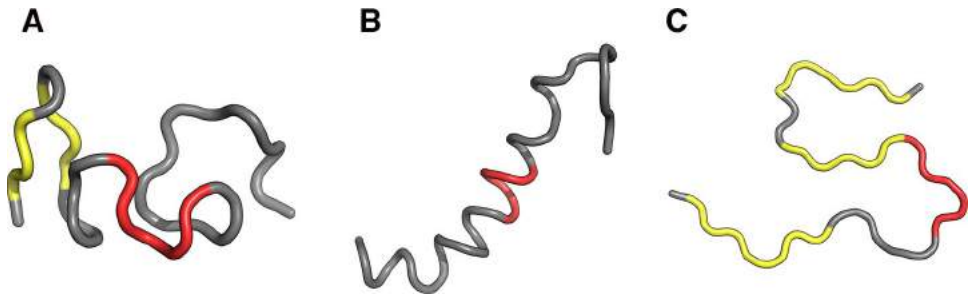


Fig. 10 3D view of structures R4 (**A**), I4 (**B**) and chain F (**C**) with 24–28 fragment colored red. Yellow fragments mark the locations of beta sheets

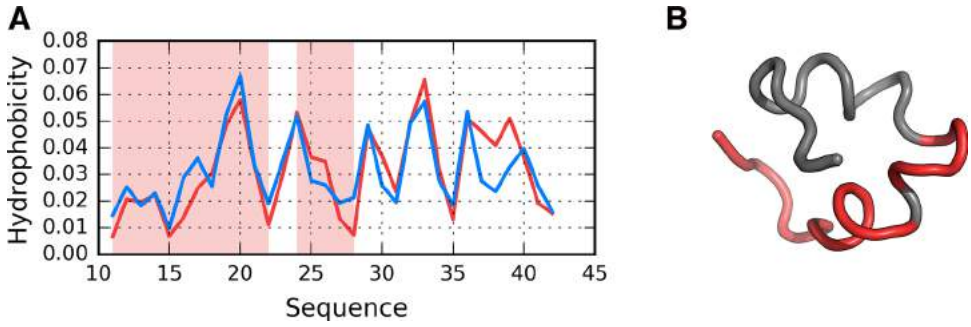


Fig. 11 Presentation of structure F1: **A** theoretical (T: blue) and observed (O: red) hydrophobicity profiles; **B** 3D view. Red highlights in **A** and red fragments in **B** correspond to 11–16, 16–22 and 24–28 residue ranges

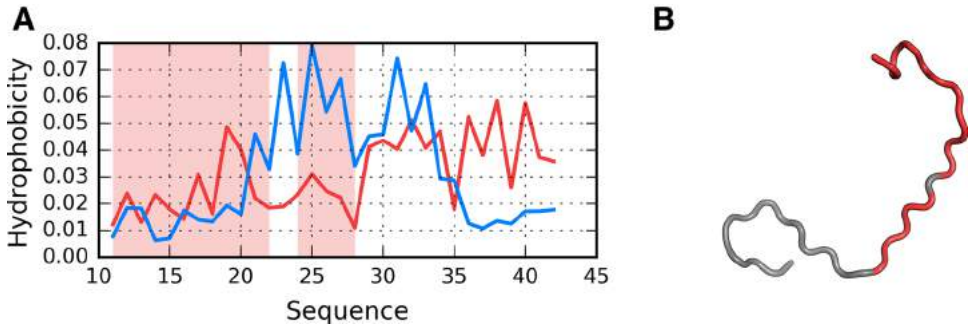


Fig. 12 Presentation of structure I5: **A** theoretical (T: blue) and observed (O: red) hydrophobicity profiles; **B** 3D view. Red highlights in **A** and red fragments in **B** correspond to 11–16, 16–22 and 24–28 residue ranges

(due to the lowest intrinsic hydrophobicity in the scale). Especially position Lys28 introduces significant discordance. In this context, it is surprising to note the emergence of hydrophilic bands formed by linearly arranged lysine residues, as such conformations should be deterred through optimization of electrostatic interactions—and yet they can be observed under experimental conditions. In fragments characterized by negative correlation coefficients the central position is frequently occupied by lysine. When multiple fragments aggregate in a linear manner,

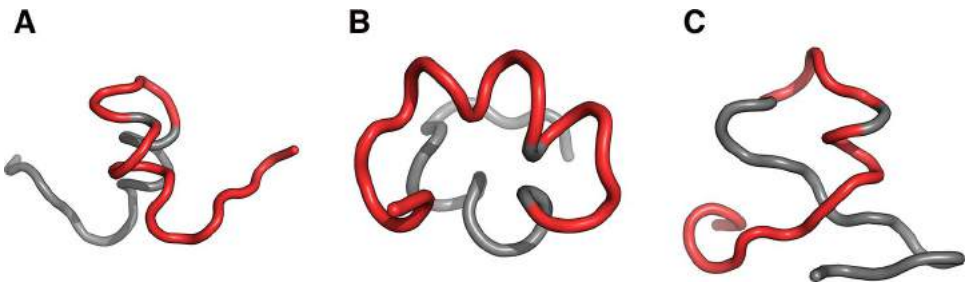


Fig. 13 3D presentation of selected structures exhibiting RD(T-O-R) values calculated for the whole chain below 0.4 (exhibiting good accordance with FOD model): F2 (**A**), F3 (**B**) and I1 (**C**) with fragments 11–16, 16–22 and 24–28 in red

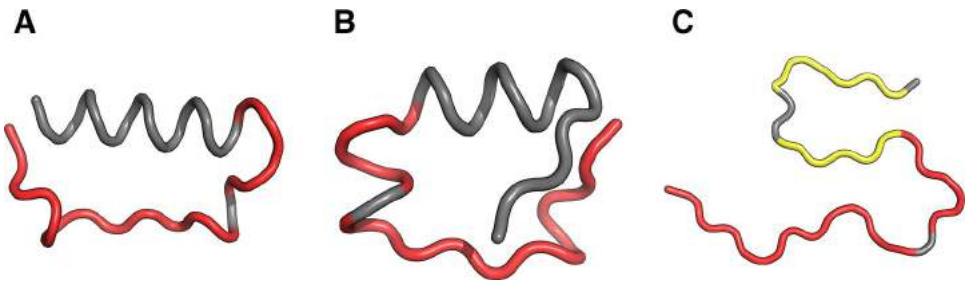


Fig. 14 3D presentation of structures accordant with the FOD model: R1 (**A**), R2 (**B**) contrasted with the amyloid structure–chain F (**C**)—with fragments 11–16, 16–22 and 24–28 in red. Yellow fragments mark the locations of beta sheets

these lysines come into close contact with one another, forming a hydrophilic band. This shows that hydrophobic forces override electrostatic interactions, and that disruptions in the external force field may produce a conformation which depends on the intrinsic hydrophobicity of participating residues. The above interpretation provides indirect confirmation of the correctness of the fuzzy oil drop model, which, under ordinary circumstances, leads to the emergence of a monocentric hydrophobic core encapsulated by a hydrophilic "shell". The importance of hydrophobic interactions for seeding amyloid transformation has already been noted in [35, 36]. In [36] the authors suggest that amyloid transformation may result from insufficient influence of the external force field (water) upon the folding process. This conclusion is consistent with the interpretation of results obtained using the fuzzy oil drop model, as presented in this paper.

In conclusion based on the presented model the weakening of standard external force field (water environment) prevents it from driving the forlding process toward centralization of hydrophobic residues and allows the intrinsic hydrophobicity to dominate.

The effect in this case is the micellarization. In particular, a ribbon-like micelle is preferred despite non-favorable interaction of charged residues arranged in electrostatically suboptimal form. The arguments for this conclusion are coming from in vitro experiments—especially shaking, which introduces air to water and—in consequence—much higher presence of inter-phase-related order of water molecules. This is why the structure of water in its standard form as well as under influence of external factors (not necessarily of chemical character) should be in focus of research oriented on amyloid transformation.

Acknowledgements The authors are indebted to Piotr Nowakowski for translation services, and to Anna Śmietańska for editorial work. This work was supported by the Jagiellonian University Collegium Medicum Grant No. K/ZDS/006363.

Author contributions I.R., and L.K.: Conceptualization; D.D., and M.G.: Methodology; M.B., and M.G.: Software; D.D., and M.G.: Validation; I.R.: Formal Analysis; I.R.: Investigation; I.R., and M.B.: Resources; I.R.: Writing-Original Draft Preparation; I.R., and M.B.: Writing-Review & Editing; M.B.: Visualization; L.K.: Supervision.

References

1. Chiti F, Dobson CM (2017) Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. Annu Rev Biochem 86(1):27–68. https://doi.org/10.1146/annurev-biochem-061516-045115
2. Eisenberg DS, Sawaya MR (2017) Structural studies of amyloid proteins at the molecular level. Annu Rev Biochem 86(1):69–95. https://doi.org/10.1146/annurev-biochem-061516-045104
3. Joseph AP, de Brevern AG (2014) From local structure to a global framework: recognition of protein folds. J R Soc Interface 11(95):20131147. https://doi.org/10.1098/rsif.2013.1147
4. Stirnemann G, Giganti D, Fernandez JM, Berne BJ (2013) Elasticity, structure, and relaxation of extended proteins under force. Proc Natl Acad Sci U S A 110(10):3847–3852. https://doi.org/10.1073/pnas.1300596110
5. Khoury GA, Liwo A, Khatib F, Zhou H, Chopra G, Bacardit J, Bortot LO, Faccioli RA, Deng X, He Y, Krupa P, Li J, Mozolewska MA, Sieradzan AK, Smadbeck J, Wirecki T, Cooper S,

Flatten J, Xu K, Baker D, Cheng J, Delbem AC, Floudas CA, Keasar C, Levitt M, Popović Z, Scheraga HA, Skolnick J, Crivelli SN, Foldit Players (2014) WeFold: a coopetition for protein structure prediction. Proteins 82(9):1850–1868. https://doi.org/10.1002/prot.24538
6. Hu K-N, Tycko R (2010) What can solid state NMR contribute to our understanding of protein folding? Biophys Chem 151(1–2):10–21. https://doi.org/10.1016/j.bpc.2010.05.009
7. Tay WM, Huang D, Rosenberry TL, Paravastu AK (2013) The Alzheimer's amyloid-β(1-42) peptide forms off-pathway oligomers and fibrils that are distinguished structurally by intermolecular organization. J Mol Biol 425(14):2494–2508. https://doi.org/10.1016/j.jmb.2013.04.003
8. Rasmussen J, Mahler J, Beschorner N, Kaeser SA, Häsler LM, Baumann F, Nyström S, Portelius E, Blennow K, Lashley T, Fox NC, Sepulveda-Falla D, Glatzel M, Oblak AL, Ghetti B, Nilsson KPR, Hammarström P, Staufenbiel M, Walker LC, Jucker M (2017) Amyloid polymorphisms constitute distinct clouds of conformational variants in different etiological subtypes of Alzheimer's disease. Proc Natl Acad Sci U S A 114(49):13018–13023. https://doi.org/10.1073/pnas.1713215114
9. Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. Cell 148(6):1188–1203. https://doi.org/10.1016/j.cell.2012.02.022
10. Tycko R (2016) Molecular structure of aggregated amyloid-β: insights from solid-state nuclear magnetic resonance. Cold Spring Harb Perspect Med 6(8):a024083. https://doi.org/10.1101/cshperspect.a024083
11. Xiao Y, Ma B, McElheny D, Parthasarathy S, Long F, Hoshi M, Nussinov R, Ishii Y (2015) A beta (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. Nat Struct Mol Biol 22(6):499–505. https://doi.org/10.1038/nsmb.2991
12. Dułak D, Banach M, Konieczny L, Roterman I (2018) Structural analysis of the aβ(15–40) amyloid fibril based on hydrophobicity distribution. Acta Biochim Pol 65(4):595–604. https://doi.org/10.18388/abp.2018_2647
13. Konieczny L, Bryliński M, Roterman I (2006) Gauss-function-based model of hydrophobicity density in proteins. Silico Biol 6(1–2):15–22
14. Kalinowska B, Banach M, Konieczny L, Roterman I (2015) Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins. Entropy 17(3):1477–1507. https://doi.org/10.3390/e17031477
15. Roterman I, Banach M, Konieczny L (2017) Application of the fuzzy oil drop model describes amyloid as a ribbonlike micelle. Entropy 19(4):167. https://doi.org/10.3390/e19040167
16. Banach M, Konieczny L, Roterman I. Use of the fuzzy oil drop model to identify the complexation area in protein homodimers. In: I. Roterman-Konieczna (Eds), Protein folding In silico, Woodhead Publishing (currently Elsevier), Oxford, Cambridge, Philadelphia, New Dehli, 2012, pp 95–122, https://doi.org/10.1533/9781908818256.95
17. Banach M, Konieczny L, Roterman I. Ligand-binding-site recognition. In: I. Roterman-Konieczna (Eds), Protein folding In silico, Woodhead Publishing (currently Elsevier), Oxford, Cambridge, Philadelphia, New Dehli, 2012, pp 79–93, https://doi.org/10.1533/9781908818256.79
18. Banach M, Konieczny L, Roterman I (2014) The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. J Theor Biol 359:6–17. https://doi.org/10.1016/j.jtbi.2014.05.007
19. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9(1):40. https://doi.org/10.1186/1471-2105-9-40
20. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738. https://doi.org/10.1038/nprot.2010.5
21. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12(1):7–8. https://doi.org/10.1038/nmeth.3213
22. https://zhanglab.ccmb.med.umich.edu/I-TASSER. Accessed 20 July 2018
23. http://robetta.org. Accessed 20 July 2018
24. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web Server):W526–W531. https://doi.org/10.1093/nar/gkh468
25. http://predictioncenter.org. Accessed 20 July 2018
26. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2013) Critical assessment of methods of protein structure prediction (CASP)—round x. Proteins 82:1–6. https://doi.org/10.1002/prot.24452
27. Qiang W, Kelley K, Tycko R (2013) Polymorph-specific kinetics and thermodynamics of β-amyloid fibril growth. J Am Chem Soc 135(18):6860–6871. https://doi.org/10.1021/ja311963f
28. Levitt MA (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104(1):59–107. https://doi.org/10.1016/0022-2836(76)90004-8
29. Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86. https://doi.org/10.1214/aoms/1177729694
30. Sałapa K, Kalinowska B, Jadczyk T, Roterman I (2012) Bio-algorithms and med-systems measurement of hydrophobicity distribution in proteins—non-redundant protein data Bank. BAMS. https://doi.org/10.2478/bams-2012-0023
31. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91(1–3):43–56. https://doi.org/10.1016/0010-4655(95)00042-e
32. http://www.gromacs.org. Accessed 20 July 2018
33. Roterman I, Banach M, Kalinowska B, Konieczny L (2016) Influence of the aqueous environment on protein structure—a plausible hypothesis concerning the mechanism of amyloidogenesis. Entropy 18(10):351. https://doi.org/10.3390/e18100351
34. Das AK, Rawat A, Bhowmik D, Pandit R, Huster D, Maiti S (2015) An early folding contact between Phe19 and Leu34 is critical for amyloid-β oligomer toxicity. ACS Chem Neurosci 6(8):1290–1295. https://doi.org/10.1021/acschemneuro.5b00074
35. Paravastu AK, Qahwash I, Leapman RD, Meredith SC, Tycko R (2009) Seeded growth of beta-amyloid fibrils from Alzheimer's brain-derived fibrils produces a distinct fibril structure. Proc Natl Acad Sci U S A 106(18):7443–7448. https://doi.org/10.1073/pnas.0812033106
36. Schmit JD, Ghosh K, Dill K (2011) What drives amyloid molecules to assemble into oligomers and fibrils? Biophys J 100(2):450–458. https://doi.org/10.1016/j.bpj.2010.11.041

Załącznik nr. 4

Analysis of alternative conformations of the Aβ(1-40) amyloid protein

DAWID DUŁAK, MAŁGORZATA GADZAŁA, MATEUSZ BANACH, IRENA ROTERMAN

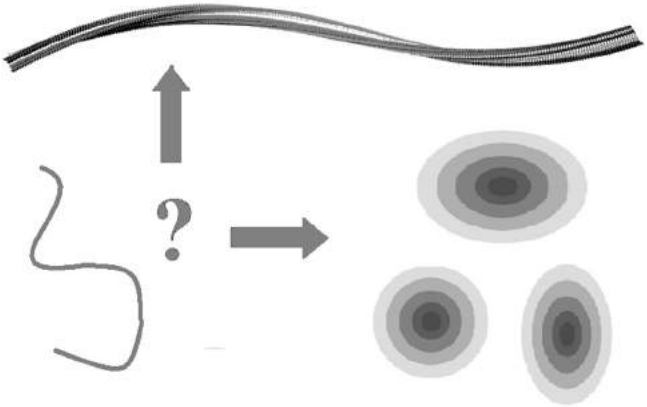
CHAPTER 10.B

Analysis of alternative conformations of the Aβ(1–40) amyloid protein

Dawid Dułak[1], Małgorzata Gadzała[2], Mateusz Banach[3], Irena Roterman[3]

[1]Department of Biophysics, Faculty of Physics, Astronomy and Applied Computer Sciencee Jagiellonian University, Krakow, Poland
[2]Cyfronet AGH Academic Computer Centre CYFRONET - University of Science and Technology in Cracow, Kraków, Poland; Currently - Schibsted Tech Polska Sp. z o. o., Kraków, Poland
[3]Department of Bioinformatics and Telemedicine, Jagiellonian University–Medical College, Krakow, Poland

Contents



From Globular Proteins to Amyloids
ISBN: 978-0-08-102981-7
https://doi.org/10.1016/B978-0-08-102981-7.00017-8

Making decision by the folding polypeptide: amyloid ? or globular ?

Alternative structural forms of polypeptide Aβ(1−40) polypeptide which includes the Osaka mutation

The amyloid structures like for example Aβ(1−40) [1] become available due to new technique ssNMR [2]. A polypeptide corresponding to the Aβ(1−40) sequence with the so-called Osaka mutation was subjected to an *in silico* experiment. Specialized folding software was used to generate five different folds per software package, in accordance with the rules of the CASP challenge. The experiment provided structures which could be regarded as alternative with respect to the experimentally observed conformation (PDB ID: 2MVX [1]). This, in turn, facilitated comparative analysis aimed at identifying fragments which support or deviate from the expected monocentric distribution of hydrophobicity. The presence of a centralized hydrophobic core provides the protein with solubility and thereby prevents unchecked complexation (potentially producing an amyloid). In addition to the above, the folding process was simulated in the presence of an external force field (FOD model) which mimics the active participation of aqueous solvent in folding process. The resulting structures provide evidence that − under the appropriate conditions − the Aβ(1−40) polypeptide may adopt a globular conformation, and suggest that the environment plays a critical role in this process.

In an attempt to identify the causes of amyloid transformation of sequences which include the Aβ(1−40) fragment, we have performed an *in silico* experiment to identify possible conformational preferences of this polypeptide. The experiment involved folding simulation tools currently regarded as the most accurate: I-Tasser [3−6] and Robetta [7−9] - two of the highest-scoring CASP challenge entrants [10]. Both are capable of predicting the conformation of chains with a given sequence, and each produces five distinct candidate structures, referred to as models. We subjected these models to comparative analysis set against the backdrop of the experimentally observed structure (obtained using ssNMR). In addition to the above, the input polypeptide was also subjected to simulations based on the fuzzy oil drop (FOD) model, which asserts the presence of an external force field representing the aqueous solvent and treated as a continuum − unlike other algorithms where the solvent is modeled as a collection of individual molecules [11,12]. In FOD, the solvent is mathematically

represented by a 3D Gaussian form, directing hydrophobic residues toward the center of the protein body while exposing hydrophilic residues on its surface. The resulting set of models enhances our ability to perform comparative analysis of amyloid structures with the outcomes of *in silico* folding simulations.

All computations using FOD were performed at the Academic Computing Center CYFRONET AGH using resources provided by the PL-Grid infrastructure.

The resulting set of 20 models (FOD: 5, I-Tasser: 5, Robetta: 10) is the subject of the presented analysis. Each model will be compared to the experimentally determined target structure – both in the context of the superfibril and as an individual chain.

In order to carry out comparative analysis, we begin by computing fuzzy oil drop coefficients for all models. This includes values of RD (relative distance – as defined by Kullback-Leibler divergence entropy) in two distinct reference models: T-O-R and T-O-H respectively [13].

The FOD model, presented in Chapter 1, suggests that the value of RD(T-O-H) is particularly important. As proposed in Refs. [14,15], which deals with the structure of the tau amyloid, amyloid seeds may be identified by looking for high values of RD (T-O-R as well as T-O-H), along with negative values of HvT and TvO and strongly positive values of HvO. These specific conditions indicate that the given fragment opposes the theoretical distribution of hydrophobicity, and is dominated by the intrinsic properties of its residues.

Values of RD and correlation coefficients may be calculated for the entire protein (polypeptide), but also for selected fragments. In the latter case, the process tells us whether the given fragment contributes to the creation of a hydrophobic core or opposes it

The structure listed in PDB under ID 2MVX consists of two protofibrils exhibiting C2 symmetry. Each protofibril resembles a flattened "C", and contacts the opposing protofibril at both tips, while exposing its backside to the environment. In order to determine the FOD status of the superfibril we compute its T and O distributions (Figs. 10.B.1 and 10.B.2). As shown in Fig. 10.B.1, these distributions are a poor match for each other. In particular, no concentration of hydrophobicity can be

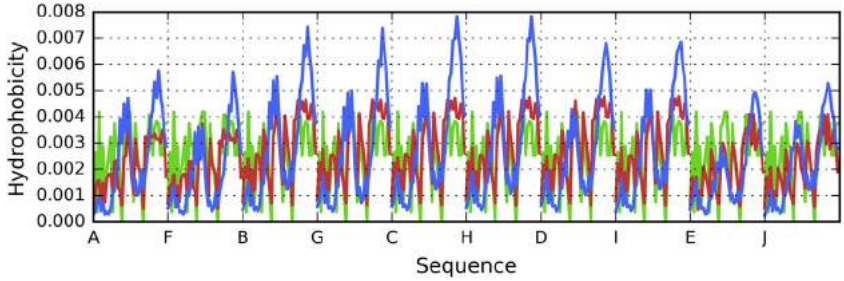


Fig. 10.B.1 Theoretical (T, blue), observed (O, red) and intrinsic (H, green) hydrophobicity distribution profiles for Aβ(1—40) (2MVX) superfibril.

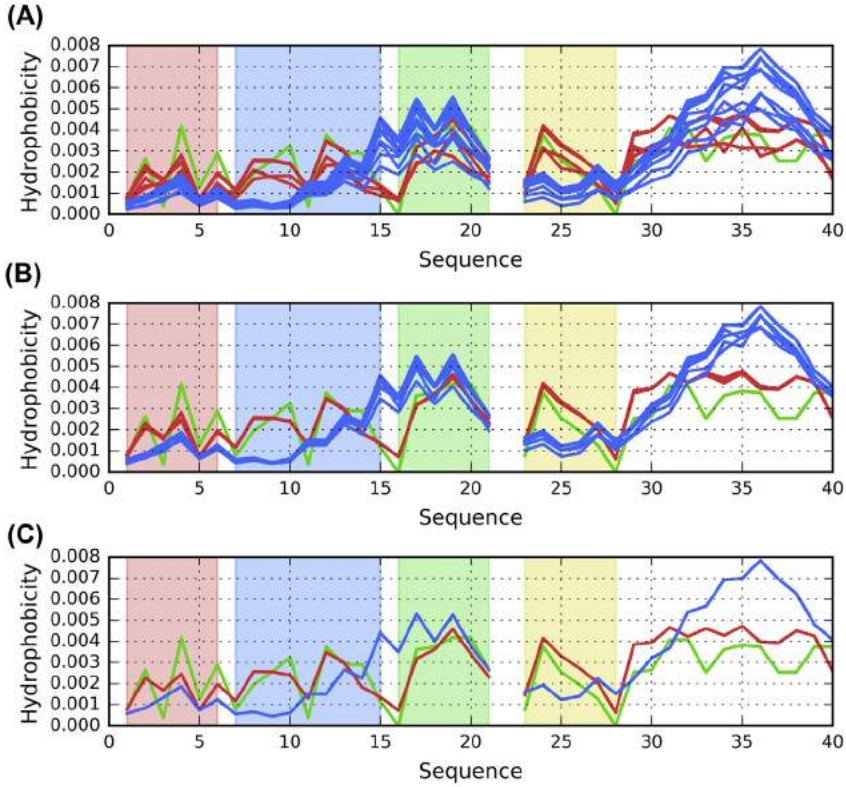


Fig. 10.B.2 Theoretical (T, blue), observed (O, red) and intrinsic (H, green) hydrophobicity distribution profiles for Aβ(1—40) (2MVX). (A) superfibril (chains A — J). (B) central part of the superfibril (following elimination of the outlying chains A, E, F and J). (C) individual (central) chain C treated as part of superfibril. Colored backgrounds highlight fragments analyzed in this study: red — 1-6, blue — 7-15, green — 16-21, yellow — 23-28. There is no residue number 22 in the sequence, however there is no gap in the chains within the PDB structure.

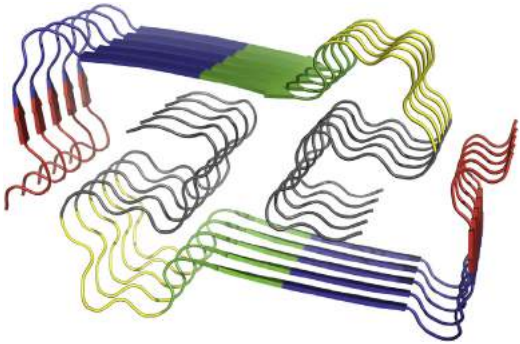


Fig. 10.B.3 3D presentation of Aβ(1–40) (2MVX) with fragments analyzed in this study highlighted by different colors: red – 1–6, blue – 7–15, green – 16–21, yellow – 23–28. Rest of each chain (29–40) is gray.

observed in the central part of the structure; instead, hydrophobicity is evenly distributed along the fibril's main axis. This type of distribution, where bands of high and low hydrophobicity propagate in an alternating fashion, resulting in a sinusoid pattern, is typical for amyloids. It follows from the repetitive nature of the input chain, as well as from conformational symmetries between each set of fragments making up the fibril (Fig. 10.B.3).

In the presented case, FOD parameters adopt the following values: RD (T-O-R) = 0.590; RD (T-O-H) = 0.592; HvT = 0.438; TvO = 0.673; HvO = 0.727. These results suggest that the structure as a whole does not contain a prominent hydrophobic core, and furthermore that the superfibril represents a consensus between the tendency to generate such a core and the intrinsic properties of each participating residue.

In contrast to Fig. 10.B.2A, profiles shown in Fig. 10.B.2B represents the distribution of hydrophobicity taking the entire superfibril as the structural unit for 3D Gauss construction.

Theoretical distribution, plotted in Fig. 10.B.2A and Fig. 10.B.2B, reveal the expected concentration of hydrophobicity in the central section of the fibril. The variability observed in Fig. 10.B.2B is due to the presence of edge chains, which lack an outlying neighbor and therefore exhibit slightly lower hydrophobicity than their centrally located counterparts (Fig. 10.B.2B). In contrast, the intrinsic distribution (H) distribution follows a sinusoidal pattern since all participating chains share identical sequences. The observed distribution (O) is also sinusoidal, consisting of alternating

minima and maxima. Fig. 10.B.2B also reveals the high similarity of repeating patterns for O and H.

Fig. 10.B.2B visualizes the theoretical and observed distributions for the central chain treated as part of superfibril. Comparing both plots reveals significant differences at positions 7–15, 16–21, 22–27 and 23–28. Rather than being limited to modest deviations, these differences point to an entirely different structural pattern, counteracting the tendency to form a shared hydrophobic core.

Similar discordance is observed when calculating hydrophobicity profiles for chain C treated as part of protofibril (centrally placed chain C – Fig. 10.B.2C). The abovementioned discordant fragments will be subjected to further analysis, comparing them with the structural properties of *in silico* models.

Plotting a 3D Gaussian for the entire complex (superfibril) also enables us to determine the status of interface fragments, i.e. those residues which remain in contact with adjacent protofibrils. The status of the interface (residues satisfying the distance criterion according to PDBsum [16]: 3, 4,15, 28, 29, 30, 37–40) expressed by FOD parameters is given by the following RD values: 0.432 and 0.387 (T-O-R and T-O-H respectively), while correlation coefficients are 0.378, 0.672 and 0.658 (HvT, TvO and HvO respectively). These values indicate good alignment between the observed distribution of hydrophobicity and the distribution predicted by the fuzzy oil drop model. Notably, TvO and HvO lead us to conclude that O is similarly aligned with T and H. We may therefore speculate that while each protofibril is dominated by the intrinsic hydrophobicity of its component residues, the entire complex (superfibril) forms as a result of interactions between protofibrils which acknowledge the presence of the aqueous environment.

Comparative analysis of protein structures generated using folding simulation software

Similarly to the analysis of the Aβ(15–40) (PDB ID: 2MPZ) [15], we will conduct a comparative study by seeking fragment of the chain whose properties suggest that they may act as amyloid seeds. In other words, the fragments of interest need to be characterized by the following: high value of HvO and negative values of both HvT and TvO. A negative correlation coefficient suggests that the given fragment not only deviates from the

reference distribution, but in fact may be regarded as a polar opposite thereof. When such values are accompanied by high RD (above 0.5), we may suspect that the conformation of the given fragment is determined by the intrinsic hydrophobicity of its residues.

As it was discussed formerly [15], to search the origin of the amyloidosis tendency of the polypeptide, the best programs predicting the structure for given amino acid sequence were used to construct alternative structural forms for this polypeptide: I-TASSER [3–5] and ROSETTA [6–9]. These two programs delived 5 alternative structural forms for given sequence (following the CASP project rule [10]). As was said earlier – Robetta delivered 10 models in this case. Additionally, the folding was performed using fuzzy oil drop model (Chapter 2). The FOD model based folding simulation delivers the structures formed by the active participation of water environment. In this study, structures belonging to the output generated by I-Tasser and Robetta are marked by letters I and R respectively. Names of structures generated by FOD start with F. Number following these IDs distinguish the individual models.

Table 10.B.1 lists the FOD coefficients (both values of RD as well as all three correlation coefficients – HvT, TvO and HvO) for each fragment under consideration: 1–6, 7–15, 16–21, 23–28. The study set consists of models obtained using Robetta (R1…R10), I-Tasser (I1…I5) and FOD-based simulations (F1…F5). It also contains results for a representative of native structure of Aβ(1–40) – chain C from 2MVX (as part of superfibril and as an individual unit).

The presented results suggest that structures labeled F1, F2, F3 as well as R1, R2, R3 and R4 are all consistent with the theoretical distribution, implying that the chain is capable of adopting a globular conformation. Table 10.B.1 also highlights forms which exhibit amyloid-like properties – evidenced by negative values of both HvT and TvO, a strongly positive value of HvO, and high values of both RD (T-O-R and T-O-H). All such structures are underscored in the table.

In most cases, however, the RD value for the T-O-H reference model is not particularly high. This may be explained by observing that we are dealing with standalone individual chains – in contrast to chains analyzed as part of an amyloid fibril. Under such conditions the dominant role of intrinsic hydrophobicity is not as evident as could be expected taking into account the structural forms obtained by FOD and Robetta: F1, F2, F3 as well as R1, R2, R3 and R4.

When summarizing the results presented in Table 10.B.1 (see Table 10.B.2 for a compact presentation), it is worth noting that the

Table.1 Fuzzy oil drop parameters for simulated model structures of 2MVX / Aβ(1-40) and their fragments (as denoted by the leftmost column), sorted in order of increasing RD(T-O-R) values for the whole chain. Columns "M" contains model names (which also designate methods used to obtain them): F1…F5 — FOD, I1...I5 — I-Tasser, R1...R10 — Robetta. "CC" stands for chain C from experimentally determined structure of 2MVX treated as part of the complex, while CS — as a standalone structure.

		RD		Correlation Coefficient				RD		Correlation Coefficient				RD		Correlation Coefficient		
FRAGMENT	**M**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**	**M**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**	**M**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
1-40	F1	0.250	0.216	0.251	0.819	0.378	F2	0.266	0.235	0.246	0.815	0.372	F3	0.279	0.224	0.164	0.799	0.358
1—6		0.448	0.154	-0.319	0.543	-0.329		0.775	0.289	0.831	-0.506	-0.187		0.675	0.329	0.308	-0.041	0.061
07—15		0.264	0.201	0.428	0.831	0.403		0.389	0.310	0.148	0.561	0.359		0.392	0.306	0.365	0.557	0.372
16—21		0.306	0.195	0.188	0.820	0.613		0.448	0.399	0.005	0.552	0.702		0.342	0.292	0.077	0.802	0.593
23—28		0.243	0.234	0.430	0.907	0.676		0.243	0.150	0.166	0.804	0.594		0.321	0.110	-0.214	0.757	0.450
29—40		0.396	0.211	-0.273	0.727	-0.555		0.233	0.130	-0.322	0.846	-0.538		0.245	0.137	-0.417	0.829	-0.538
11—19		0.287	0.196	0.231	0.778	0.652		0.399	0.324	0.115	0.610	0.666		0.436	0.357	0.154	0.576	0.641
1—40	R1	0.434	0.375	0.502	0.578	0.704	R2	0.458	0.438	0.501	0.564	0.694	R3	0.465	0.457	0.459	0.531	0.703
1—6		0.356	0.328	0.674	0.805	0.822		0.203	0.206	0.632	0.892	0.721		0.168	0.139	0.788	0.904	0.880
07—15		0.486	0.874	0.444	0.535	0.945		0.539	0.488	0.215	0.341	0.617		0.717	0.450	-0.121	0.231	0.867
16—21		0.200	0.065	0.829	0.894	0.870		0.240	0.254	0.827	0.855	0.831		0.196	0.282	0.814	0.918	0.836
23—28		0.356	0.058	0.015	0.727	0.527		0.639	0.133	0.072	0.228	0.804		0.604	0.113	0.018	0.253	0.800
29—40		0.482	0.382	0.230	0.781	0.085		0.725	0.573	0.287	0.609	-0.005		0.747	0.595	0.287	0.619	0.033
11—19		0.404	0.277	0.650	0.697	0.929		0.212	0.281	0.786	0.894	0.817		0.139	0.225	0.782	0.957	0.850
1—40	R4	0.465	0.365	0.414	0.565	0.755	R5	0.508	0.506	0.411	0.465	0.791	R6	0.509	0.500	0.555	0.544	0.780
1—6		0.229	0.425	0.832	0.864	0.985		0.378	0.658	0.643	0.656	0.965		0.297	0.478	0.832	0.775	0.958
07—15		0.356	0.349	0.510	0.719	0.739		0.330	0.423	0.505	0.770	0.813		0.493	0.533	0.489	0.595	0.797
16—21		0.112	0.047	0.759	0.934	0.872		0.580	0.139	0.050	0.131	0.748		0.354	0.222	0.852	0.559	0.747
23—28		0.616	0.121	-0.335	0.283	0.479		0.396	0.547	0.340	0.671	0.920		0.592	0.260	0.031	0.251	0.862
29—40		0.770	0.568	-0.050	0.325	-0.119		0.753	0.723	0.206	0.371	0.376		0.749	0.695	0.192	0.267	0.273
11—19		0.117	0.097	0.895	0.975	0.877		0.331	0.159	0.457	0.726	0.869		0.187	0.154	0.856	0.821	0.837
1—40	R7	0.519	0.434	0.412	0.436	0.721	R8	0.532	0.427	0.370	0.515	0.681	I1	0.602	0.409	0.132	0.332	0.516
1—6		0.218	0.101	0.824	0.909	0.804		0.330	0.125	0.235	0.856	0.290		0.421	0.243	0.666	0.472	0.275
07—15		0.618	0.515	-0.075	0.229	0.869		0.746	0.668	-0.103	-0.154	0.913		0.806	0.348	-0.412	0.049	-0.475

16−21		0.255	0.132	0.746	0.815	0.771		0.265	0.155	0.833	0.803	0.840		0.652	0.385	-0.350	-0.591	0.859
23−28		0.422	0.078	0.231	0.509	0.857		0.727	0.190	-0.119	0.010	0.806		0.814	0.253	-0.866	-0.609	0.291
29−40		0.790	0.684	0.231	0.345	0.149		0.767	0.687	0.289	0.323	0.145		0.792	0.744	0.366	0.521	0.287
11−19		0.136	0.110	0.724	0.937	0.828		0.145	0.120	0.762	0.923	0.824		0.492	0.222	-0.173	0.363	0.526
1−40	CC	0.608	0.620	0.459	0.665	0.784	I2	0.623	0.508	0.147	0.218	0.540	F4	0.629	0.368	-0.157	0.312	0.238
1−6		0.332	0.207	0.621	0.739	0.770		0.352	0.157	-0.317	0.680	0.322		0.506	0.111	-0.406	0.593	-0.283
07−15		0.857	0.857	-0.113	-0.198	0.850		0.762	0.641	0.222	-0.250	0.433		0.623	0.499	0.152	0.727	0.531
16−21		0.407	0.282	0.485	0.622	0.946		0.677	0.602	-0.593	-0.905	0.792		0.581	0.358	-0.020	0.196	0.726
23−28		0.533	0.461	0.065	0.039	0.995		0.648	0.108	-0.643	-0.153	0.403		0.680	0.293	-0.958	-0.342	0.417
29−40		0.853	0.648	0.157	0.301	-0.005		0.758	0.617	0.140	0.344	-0.285		0.548	0.386	-0.260	0.607	-0.672
11−19		0.565	0.484	0.285	0.348	0.929		0.564	0.482	-0.110	-0.182	0.776		0.371	0.215	0.156	0.631	0.697
1−40	F5	0.631	0.433	-0.122	0.350	0.285	I3	0.634	0.372	0.134	0.401	0.552	CS	0.636	0.562	0.295	0.363	0.616
1−6		0.466	0.329	-0.101	0.530	0.251		0.576	0.325	-0.162	0.185	0.622		0.682	0.370	0.747	0.025	0.233
07−15		0.756	0.454	0.358	0.356	0.295		0.859	0.600	-0.325	-0.283	0.173		0.864	0.783	-0.062	-0.405	0.606
16−21		0.542	0.324	-0.767	0.145	0.397		0.499	0.162	0.330	0.492	0.876		0.456	0.171	0.199	0.407	0.844
23−28		0.821	0.443	-0.920	-0.903	0.754		0.790	0.122	-0.797	-0.148	0.015		0.421	0.230	0.494	0.508	0.994
29−40		0.659	0.506	0.040	0.587	-0.514		0.884	0.723	0.204	0.626	-0.139		0.797	0.690	0.159	0.611	-0.022
11−19		0.473	0.217	-0.144	0.618	0.211		0.407	0.177	0.169	0.624	0.627		0.568	0.362	0.213	0.258	0.809
1−40	I4	0.642	0.459	0.277	0.333	0.491	R9	0.650	0.601	0.372	0.413	0.757	I5	0.702	0.521	0.266	0.208	0.533
1−6		0.419	0.531	0.805	0.532	0.847		0.602	0.555	0.357	-0.157	0.684		0.740	0.566	0.403	0.381	0.917
07−15		0.886	0.502	-0.099	-0.447	0.200		0.685	0.892	0.317	0.155	0.909		0.893	0.673	-0.079	-0.836	0.183
16−21		0.722	0.467	0.124	-0.460	0.615		0.723	0.297	-0.393	-0.394	0.798		0.662	0.510	-0.006	-0.459	0.693
23−28		0.808	0.299	-0.760	-0.454	0.665		0.459	0.312	0.275	0.440	0.914		0.837	0.222	-0.844	-0.452	0.143
29−40		0.502	0.357	0.185	0.618	-0.392		0.737	0.644	0.119	0.454	0.169		0.728	0.462	0.473	0.410	-0.504
11−19		0.446	0.231	0.254	0.444	0.673		0.509	0.320	0.087	0.217	0.827		0.542	0.340	0.038	0.199	0.711
1−40	R10	0.718	0.612	0.142	0.025	0.680												
1−6		0.669	0.467	0.781	0.062	0.615												
07−15		0.782	0.630	-0.070	-0.390	0.701												
16−21		0.527	0.135	0.027	0.129	0.802												
23−28		0.269	0.080	0.676	0.810	0.965												
29−40		0.724	0.822	0.587	0.422	0.693												
11−19		0.515	0.160	-0.012	0.244	0.886												

Table 10.B.2 Status of suspected amyloid seeds in simulated model structures of Aβ(1−40), compared with experimental observations (2MVX structure). Numbers in the table correspond to names of models produced by each software package. The rightmost column presents the experimentally determined status for native chain C as part of the complex as a whole as well as its individual chains when treated as an individual unit.

Fragment	FOD	I-Tasser	Robetta	Aβ(1−40) (2MVX)
Accordant fragments (RD < 0.5)				
1−40	F1,F2,F3		R1,R2,R3,R4	
1−6	F1,F5	I1,I2,I4	R1,R2,R3,R4,R5,R6, R7,R8	Comp.
7−15	F1,F2,F3		R1,R4,R5,R6	
16−21	F1,F2,F3	I3	R1,R2,R3,R4,R6, R7,R8	Comp., Indiv.
23−28	F1,F2,F3		R1,R5,R7,R9,R10	Indiv.
29−40	F1,F2,F3		R1	
Discordant fragments (RD ≥ 0.5)				
1−40	F4,F5	I1,I2,I3,I4,I5	R5,R6,R7,R8,R9,R10	Comp., Indiv.
1−6	F2,F3,F4	I3,I5	R9,R10	Indiv.
7−15	F4,F5	I1,I2,I3,I4,I5	R2,R3,R7,R8,R9,R10	Comp., Indiv.
16−21	F4,F5	I1,I2,I4,I5	R5,R9,R10	
23−28	F4,F5	I1,I2,I3,I4,I5	R2,R3,R4,R6,R8	Comp.
29−40	F4,F5	I1,I2,I3,I4,I5	R2,R3,R4,R5,R6,R7, R8,R9,R10	Comp., Indiv.

fragments at 7−15, 16−21 and 22−27, 23−28 frequently deviate from the theoretical monocentric distribution in favor of experimentally determined properties (except for the 16−21, where such conditions are not observed).

When discussing the status of each fragment it should be noted that the evaluation criteria are supplied by the FOD model.

The 3D structures of generated models are shown in Fig. 10.B.4 (F1…F5, I1…F5, R1 … R5) and Fig. 10.B.5 (R6 … R10). Hydrophobicity profiles of selected structures accordant and discordant with the fuzzy oil drop model are shown in Fig. 10.B.6 and Fig. 10.B.7 respectively.

Analysis of results shown in Table 10.B.1 and Table 10.B.2 reveals − somewhat surprisingly − that the models produced by Robetta are largely consistent with the monocentric distribution of hydrophobicity. While the Robetta algorithm does not directly account for internalization of hydrophobic residues, it nevertheless proves that the presented chain may,

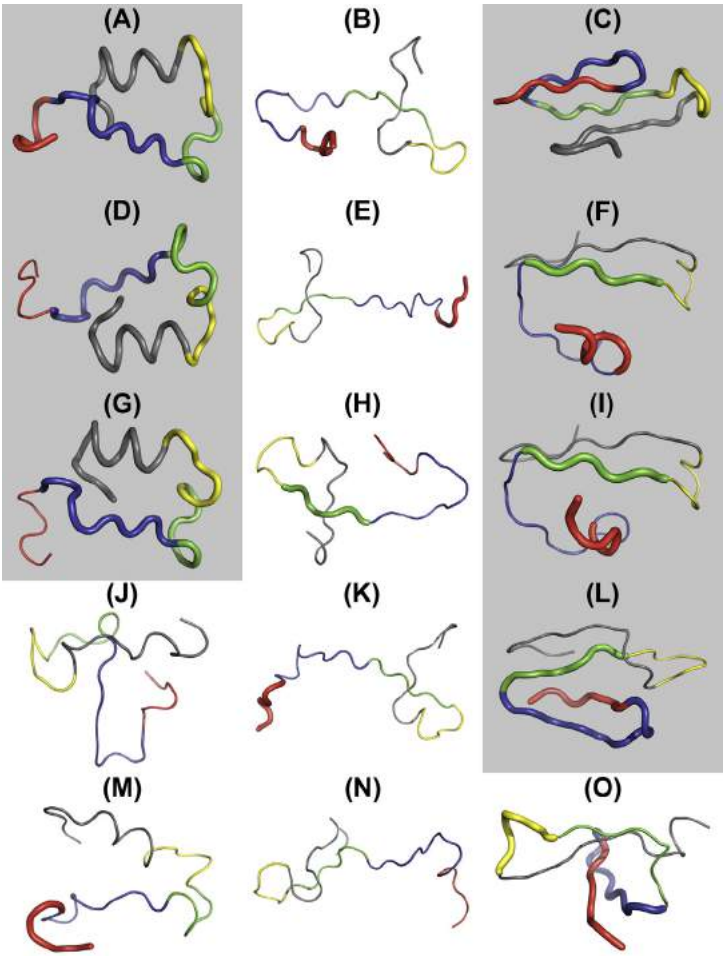


Fig. 10.B.4 3D presentation of simulated model structures of Aβ(1—40) (2MVX), part 1. A, D, G, J, M— F1…F5. B, E, H, K, N — I1—I5. C, F, I, L, O — R1 … R5. Fragments analyzed in this study are highlighted by colors: red — 1—6, blue — 7—15, green — 16—21, yellow — 23—28. Rest of each chain (29—40) is gray. Gray background denotes RD (T-O-R) < 0.5 for the whole chain (1—40). Status of each fragment of the structures is given by the backbone trace style: thick — accordant (RD < 0.5), thin — discordant (RD ≥ 0.5).

under certain circumstances, adopt a globular conformation. It is interesting to speculate why such phenomena are not observed *in vivo*.

In summary, it is worth noting that Robetta generated four models whose RD is lower than 0.5 (indicating the presence of a centralized

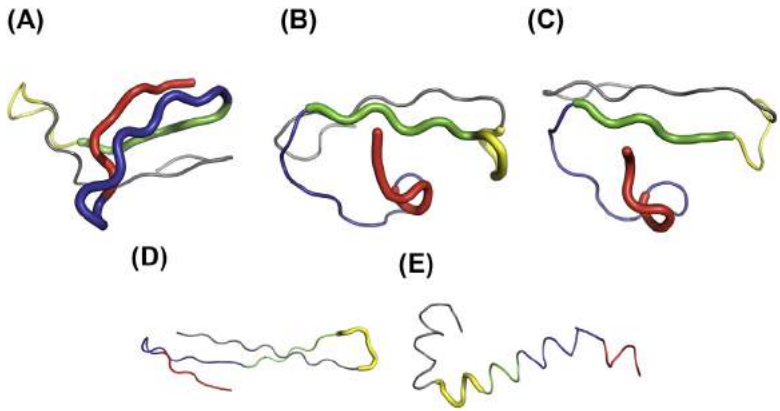


Fig. 10.B.5 3D presentation of simulated model structures of Aβ(1−40) (2MVX), part 2. A, B, C, D, E R6 … R10. Fragments analyzed in this study are highlighted by colors: red − 1−6, blue − 7−15, green − 16−21, yellow − 23−28. Rest of each chain (29−40) is gray. Gray background denotes RD (T-O-R) $<$ 0.5 for the whole (1−40) chain. Status of each fragment of the structures is given by the backbone trace style: thick − accordant (RD $<$ 0.5), thin − discordant (RD $\geq$ 0.5).

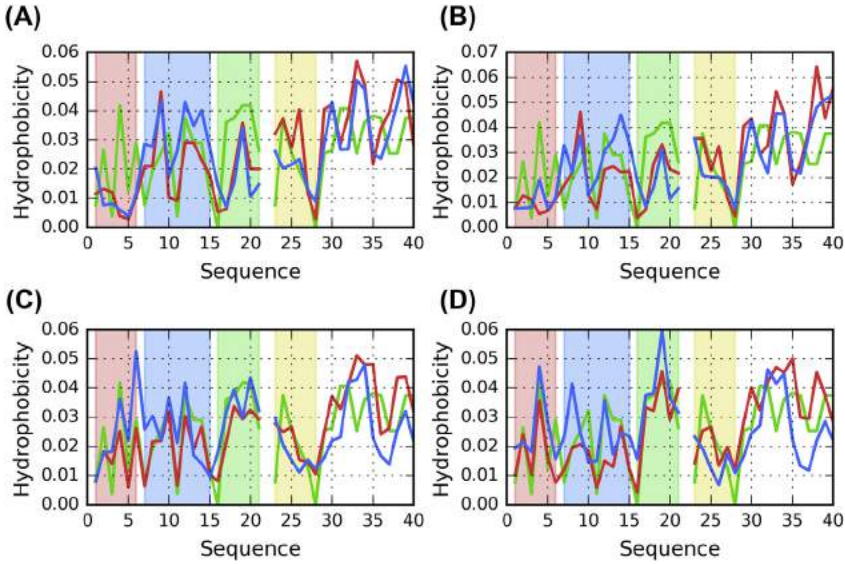


Fig. 10.B.6 Theoretical (T, blue), observed (O, red) and intrinsic (H, green) hydrophobicity distribution profiles for selected accordant simulated model structures of Aβ(1−40) (2MVX). (A) F1; (B) F2; (C) R1; (D) R2. Colored backgrounds highlight fragments analyzed in this study: red − 1−6, blue − 7−15, green − 16−21, yellow − 23−28.

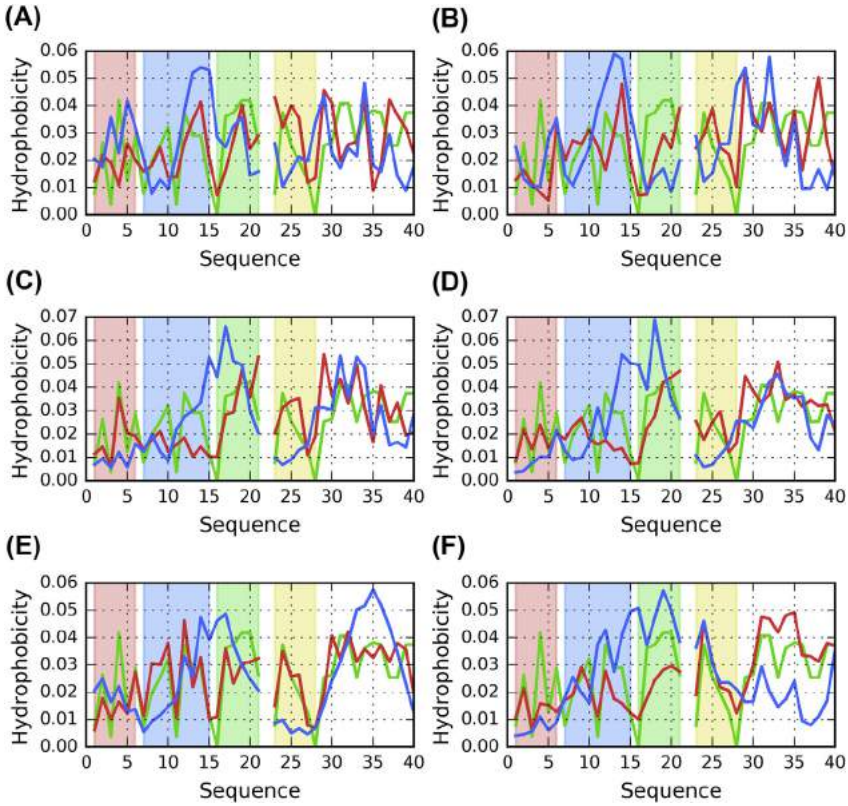


Fig. 10.B.7 Theoretical (T, blue), observed (O, red) and intrinsic (H, green) hydrophobicity distribution profiles for selected discordant simulated model structures of Aβ(1–40) (2MVX). Colored backgrounds highlight fragments analyzed in this study: red – 1–6, blue – 7–15, green – 16–21, yellow – 23–28.

hydrophobic core). In contrast, I-Tasser produced no such models. While I-Tasser models are generally more reminiscent of amyloid forms (Fig. 10.B.4), the greatest deviation from the monocentric core pattern is observed for one of the presented Robetta models. Regarding FOD, the computed models promote internalization of hydrophobic residues but are nevertheless quite divergent from globular forms [17,18].

References

[1] Schütz AK, Vagt T, Huber M, Ovchinnikova OY, Cadalbert R, Wall J, Meier BH. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. Angewandte Chemie International Edition 2014;54(1):331–5. https://doi.org/10.1002/anie.201408598.

[2] Tycko R. Molecular structure of aggregated amyloid-β: insights from solid-state nuclear magnetic resonance. Cold Spring Harbor Perspectives in Medicine 2016;6(8): a024083. https://doi.org/10.1101/cshperspect.a024083.
[3] Zhang Y. I-tasser server for protein 3D structure prediction. BMC Bioinformatics 2008;9(1):40. https://doi.org/10.1186/1471-2105-9-40.
[4] Roy A, Kucukural A, Zhang Y. I-tasser: a unified platform for automated protein structure and function prediction. Nature Protocols 2010;5(4):725–38. https://doi.org/10.1038/nprot.2010.5.
[5] Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-tasser Suite: protein structure and function prediction. Nature Methods 2015;12(1):7–8. https://doi.org/10.1038/nmeth.3213.
[6] https://zhanglab.ccmb.med.umich.edu/I-TASSER.
[7] http://robetta.org.
[8] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Research 2004;32(Web Server):W526–31. https://doi.org/10.1093/nar/gkh468.
[9] http://predictioncenter.org.
[10] Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP) – round x. Proteins: Structure, Function, and Bioinformatics 2013;82:1–6. https://doi.org/10.1002/prot.24452.
[11] Konieczny L, Brylinski M, Roterman I. Gauss-function-based model of hydrophobicity density in proteins. Silico Biology 2006;6(1–2):15–22.
[12] Kalinowska B, Banach M, Konieczny L, Roterman I. Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins. Entropy 2015;17(3):1477–507. https://doi.org/10.3390/e17031477.
[13] Kullback S, Leibler RA. On information and sufficiency. The Annals of Mathematical Statistics 1951;22(1):79–86. https://doi.org/10.1214/aoms/1177729694.
[14] Dułak D, Gadzała M, Banach M, Ptak M, Wiśniowski Z, Konieczny L, Roterman I. Filamentous aggregates of tau proteins fulfil standard amyloid criteria provided by the fuzzy oil drop (FOD) model. International Journal of Molecular Sciences 2018;19(10): 2910. https://doi.org/10.3390/ijms19102910.
[15] Dułak D, Banach M, Gadzała M, Konieczny L, Roterman I. Structural analysis of the Aβ(15–40) amyloid fibril based on hydrophobicity distribution. Acta Biochimica Polonica 2018. https://doi.org/10.18388/abp.2018_2647.
[16] Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. Protein Science 2017;27(1):129–34. https://doi.org/10.1002/pro.3289.
[17] Schmit JD, Ghosh K, Dill K. What drives amyloid molecules to assemble into oligomers and fibrils? Biophysical Journal 2011;100(2):450–8. https://doi.org/10.1016/j.bpj.2010.11.041.
[18] Banach M, Konieczny L, Roterman I. Fuzzy oil drop model application—from globular proteins to amyloids. Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes 2018:639–58. https://doi.org/10.1007/978-3-319-95843-9_19.

Załącznik nr. 5

Alternative Structures of α-Synuclein

DAWID DUŁAK, MAŁGORZATA GADZAŁA, MATEUSZ BANACH,

LESZEK KONIECZNY, IRENA ROTERMAN

molecules

MDPI

Article

Alternative Structures of α-Synuclein

Dawid Dułak [1], Małgorzata Gadzała [2,†], Mateusz Banach [3], Leszek Konieczny [4] and Irena Roterman [3,*]

1 Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland; dawid.dulak@gmail.com
2 ACK–Cyfronet AGH, Nawojki 11, 30-950 Krakow, Poland; m.k.gadzala@gmail.com
3 Department of Bioinformatics and Telemedicine, Jagiellonian University—Medical College, Łazarza 16, 31-530 Krakow, Poland; mateusz.banach@uj.edu.pl
4 Chair of Medical Biochemistry, Jagiellonian University–Medical College, Kopernika 7, 31-034 Kraków, Poland; mbkoniec@cyf-kr.edu.pl
* Correspondence: myroterm@cyf-kr.edu.pl
† Current address: Schibsted Tech Polska Sp. z o.o. ul. Armii Krajowej 28, 30-150 Kraków, Poland.

Academic Editor: Botond Penke
Received: 19 December 2019; Accepted: 17 January 2020; Published: 30 January 2020

Abstract: The object of our analysis is the structure of alpha-synuclein (ASyn), which, under in vivo conditions, associates with presynaptic vesicles. Misfolding of ASyn is known to be implicated in Parkinson's disease. The availability of structural information for both the micelle-bound and amyloid form of ASyn enables us to speculate on the specific mechanism of amyloid transformation. This analysis is all the more interesting given the fact that—Unlike in Aβ(1–42) amyloids—only the central fragment (30–100) of ASyn has a fibrillar structure, whereas, its N- and C-terminal fragments (1–30 and 100–140, respectively) are described as random coils. Our work addresses the following question: Can the ASyn chain—as well as the aforementioned individual fragments—adopt globular conformations? In order to provide an answer, we subjected the corresponding sequences to simulations carried out using Robetta and I-Tasser, both of which are regarded as accurate protein structure predictors. In addition, we also applied the fuzzy oil drop (FOD) model, which, in addition to optimizing the protein's internal free energy, acknowledges the presence of an external force field contributed by the aqueous solvent. This field directs hydrophobic residues to congregate near the center of the protein body while exposing hydrophilic residues on its surface. Comparative analysis of the obtained models suggests that fragments which do not participate in forming the amyloid fibril (i.e., 1–30 and 100–140) can indeed attain globular conformations. We also explain the influence of mutations observed in vivo upon the susceptibility of ASyn to undergo amyloid transformation. In particular, the 30–100 fragment (which adopts a fibrillar structure in PDB) is not predicted to produce a centralized hydrophobic core by any of the applied toolkits (Robetta, I-Tasser, and FOD). This means that in order to minimize the entropically disadvantageous contact between hydrophobic residues and the polar solvent, ASyn adopts the form of a ribbonlike micelle (rather than a spherical one). In other words, the ribbonlike micelle represents a synergy between the conformational preferences of the protein chain and the influence of its environment.

Keywords: misfolding; A-synuclein; amyloid; fibril; protein folding; hydrophobicity

1. Introduction

A-synuclein (referred to as ASyn in our work) is strongly expressed in brain tissue [1], particularly at the presynaptic termini [2], and in synaptic membranes [3]. It is also found on the tips of nerve cells (neurons) in specialized structures called presynaptic terminals [4]. The ASyn chain is sometimes divided into an N-terminal fragment (1–60), a NAC (non-amyloid beta component; 61–95) and a

strongly hydrophilic C-terminal fragment (96–140) [5]. Another characteristic property of ASyn is the presence of imperfect KTKEGV repeats starting at positions 10, 21, 32, 43, 58, 69, and 80, respectively. The first 25 residues of the N-terminal fragment are responsible for anchoring the protein in the lipid bilayer, whereas, residues 26–98 mediate affinity towards the membrane, depending on the composition of the membrane itself [6]. This affinity is likely related to the biological activity of ASyn, which, however, remains poorly understood (see References [7,8] for further information). The amyloid form of ASyn (PDB ID: 2N0A [9]) comprises a fibrillar central fragment (30–100), while the remaining N- and C-terminal fragments both adopt random coil conformations. The presented work provides an analysis of the entire polypeptide, as well as of each of its fragments. The presence of relatively long fragments which do not contribute to the fibrillar structure distinguishes ASyn from among other amyloids (particularly Aβ [10–13], and tau [14] amyloids where the amyloid-like structure encompasses the entire chain [9,15]).

Micelle-bound (1XQ8) [16] and fibrillar (2N0A) forms of ASyn were subjected to analysis based on the fuzzy oil drop model (FOD) [17–19]. The model assumes that the distribution of hydrophobicity in a globular molecule can be mathematically modeled as a 3D Gaussian, with hydrophobicity peaking at the center of the protein body and decreasing along with the distance from the center, reaching almost 0 on the surface and beyond it. Unlike globular proteins, amyloids follow an entirely different structural pattern, presenting alternating bands of high and low hydrophobicity along their main axis. The fuzzy oil drop model provides a reference, which enables us to determine whether—And to what extent—Any given protein follows the spherical [15] or ribbon-like (linear) pattern [15]. In addition, in the presented work, we also try to determine whether the ASyn chain may adopt a globular form, and thus, become soluble. In order to answer this question, we apply specialized software toolkits—Robetta [17] and I-Tasser [18], both noted for their accuracy in recent editions of the CASP challenge [19]. Additionally, we perform calculations based on the fuzzy oil drop model, enabling us to acknowledge the effects of an external force field (contributed by the aqueous solvent) in addition to internal force fields (atom-atom interactions). This external field directs hydrophobic residues to congregate at the center of the molecule, while promoting exposure of hydrophilic residues [20–22]. Consequently, it favors the formation of a globular protein. Taken together, the three software frameworks enable us to generate a diverse spectrum of models [23–26].

2. Results

2.1. Parameters Used for Structure Description

The detailed description of the model is given in Materials and Methods. The basic assumptions are given here to make the interpretation of the results easier.

The molecule under consideration is encapsulated in ellipsoid (3D Gauss function). The values of Gauss function in particular points are treated as idealized hydrophobicity level. It is the consequence of the assumption that the protein molecule follows more or less construction of the spherical micelle exposing polar groups on the surface and hiding the hydrophobic residues in the central part of protein body (hydrophobic core). The hydrophobicity distribution is treated as idealized one-called T in this paper. On the other hand, the observed hydrophobicity distribution-the result of hydrophobic interaction depends on the intrinsic hydrophobicity of interacting residues and on the distance between them. This interaction calculated according to Levitt's function [27] is called as O-observed one. Comparison of these two distributions allows the identification of similarities/differences between them, revealing the status of protein under consideration. The quantitative measurements of differences are possible using the Kullback-Leibler divergence entropy [28] called D_{KL}. Distance between O and T expressed by D_{KL} cannot be interpreted (entropy category). This is why the second reference distribution is introduced: The unified one, where each residue is attributed by equal *hydrophobicity level* = 1/N, where N is the number of residues. This distribution called R represents the status of protein molecule deprived of any form of hydrophobicity level differentiation-opposite to T distribution

representing the presence of centric hydrophobic core. This is why two D_{KL} values describe the status of each residue: For the relation of O distribution versus two reference distributions: T and R. The relation between values D_{KL} for status: O|T and O|R characterizes the O distribution. If D_{KL} for O|$T > O$|R it means that the O distribution is closer to R distribution. In the opposite situation, the presence of a hydrophobic core is assumed to be present in the protein under consideration. To avoid dealing with two parameters, the *RD* (Relative Distance) is introduced expressed as D_{KL} for O|T divided by the sum of D_{KL} for O|T and O|R. The RD value lower than 0.5 suggests the presence of a hydrophobic core. This is why this parameter is used to characterize the status of proteins and models (as calculated using special programs: I-TASSER [29], ROBETTA [30]) and FOD model [26] folding protein in the presence of external force field in the form of 3D Gauss function to express the influence of water environment. The models representing the status of $RD < 0.5$ are treated as globular with a hydrophobic core. The detailed description of the procedures is given in Materials and Methods (chapter 4).

The aim of the calculations presented in this paper is to check the specificity of amino acids sequence present in ASyn and particular its sequence in three fragments of the chain (1–30, 30–100 and 100–140) representing different structural forms (only fragment 30–100 in amyloid form). The commonly known assumption is that the amino acids sequence determines the 3D structure of the protein. The question is what specificity of sequence promotes the generation of the spherical micelle and what is the specificity of sequence promoting ribbon-like micelle, which is observed in amyloids [31,32].

Programs predicting 3D structure using different force field are assumed to produce different 3D structures of ASyn and its chain fragments. The analysis of these models is assumed to reveal the specific influence of amino acid sequences on 3D structure promoting the amyloid forms.

2.2. Structure of Human Micelle-Bound Alpha-Synuclein (1XQ8)

The structure of ASyn in its micelle-bound form (1XQ8) was subjected to FOD analysis in two approaches, the first of which covered the entire micelle-bound chain, while the second was limited to the dual helix system (1–95) and excluded the loose C-terminal random coil (96–140).

Results obtained for the helical hairpin are visualized in Figure 1A, which provides a comparison of two distributions (T and O), revealing fragments where the observed distribution clearly diverges from the theoretical model. According to the fuzzy oil drop model fragments characterized by a local excess of hydrophobicity may play a role in the complexation of external structures which also expose excess hydrophobicity on their surface (in the case of the presented structure, complexation involves the micelle). Judging by *RD* and correlation coefficients, this fragment of the ASyn chain lacks a well-defined hydrophobic core (refer to Table S1—Note that all tables are available in Supplementary Materials, indexed by "S" together with the table number)—Although eliminating residues which exhibit deviations from the theoretical distribution produces a fragment whose *RD* value is below 0.5. Figure 1B identifies locations where interaction with external molecules is likely to occur.

The complete micelle-bound form of Asyn (1XQ8) can be characterized (on the basis of fuzzy oil drop parameters) as lacking a hydrophobic core, which is a direct consequence of its notable lack of a tertiary conformation. If, however, we limit our analysis to the helical hairpin (1–95), we may observe a tendency for hydrophobic residues to congregate in the central part of the molecule and for hydrophilic residues to migrate to its surface. Eliminating residues which exhibit the greatest difference between T_i and O_i reveals fragments contributing to the formation of a centralized core. These fragments (highlighted in Figure 1A) are likely primed for interaction with the hydrophobic surface of the target molecule, and their identification bases on comparing the theoretical (T) and observed (O) distributions of hydrophobicity (Figure 1B).

Elimination of residues highlighted in Figure 1A,B reduces the value of *RD* to below 0.5, indicating good agreement with the Gaussian distribution.

The status of the 1–95 fragment of micelle-bound ASyn (Table S1) appears to indicate the lack of a centralized core; however, eliminating residues highlighted in Figure 1 as representing a local

deviation from T (in the form of excess or insufficient hydrophobicity) produces a structure which is a good match for the centralized model.

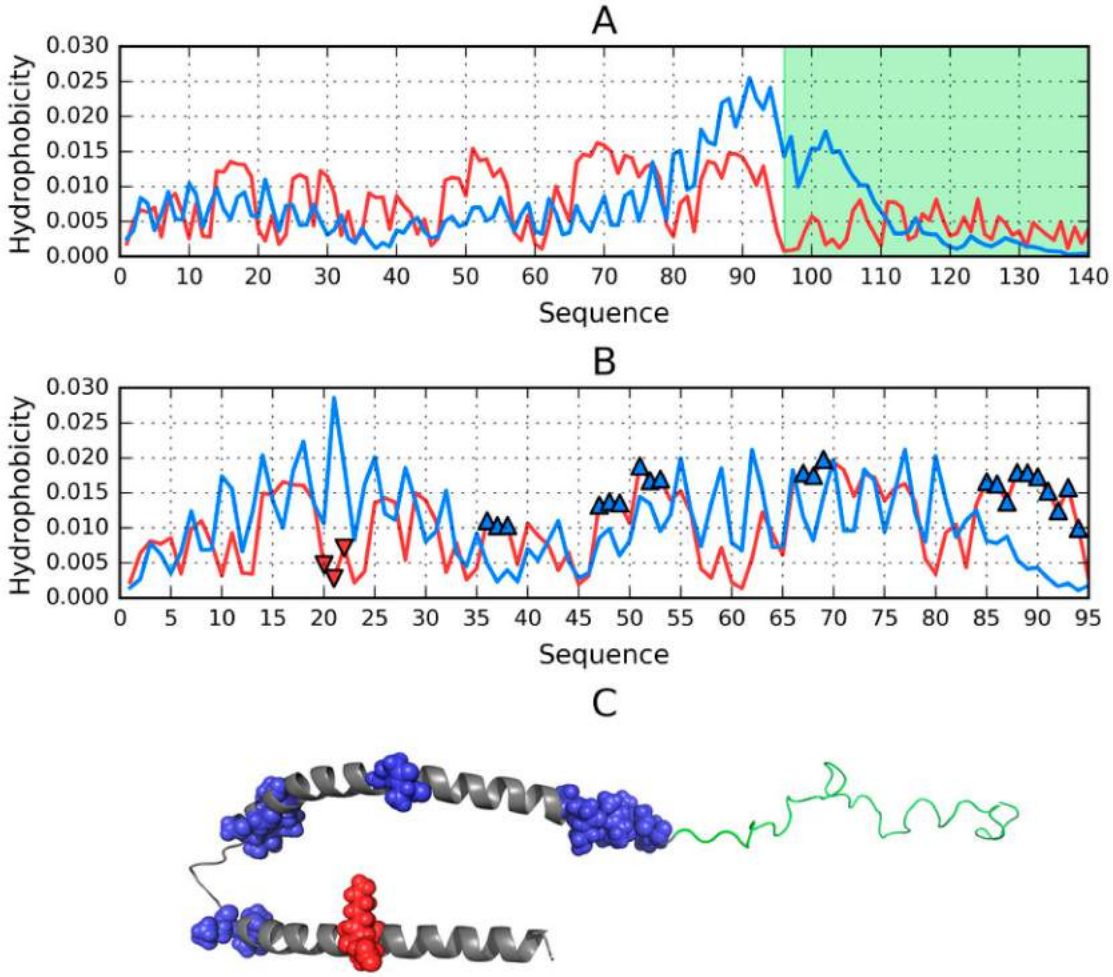


Figure 1. Structure of ASyn in its micelle-bound form (1XQ8); (**A**)—Theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles calculated for the complete chain (1–140). Random coil C-terminal fragment (96–140) is highlighted in green, (**B**)—Theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles calculated for the 1–95 fragment treated as an individual unit. Blue markers indicate residues No. 36–38, 47–49, 51–53, 67–69, 85–94 which exhibit local excess of hydrophobicity, while red markers correspond to local hydrophobicity deficiency (residues No. 20–22); (**C**)—3D presentation with color-coding corresponding to figures A and B.

Given the extensive literature devoted to the properties of specific ASyn fragments, Table S1 also lists the status of these fragments—This information will come in handy when discussing their conformation in the amyloid form of ASyn. Additionally, we have computed the status of certain repetitive fragments and other fragments which appear in previously published studies.

Repetitive fragments are generally found to deviate from the theoretical distribution—In fact, values listed in Table S1 indicate that some of these fragments exhibit amyloid-like properties. The status of fragments which adopt beta folds in the amyloid is also quite similar to an amyloid structure, particularly in the case of the 70–78 fragment. Helical fragments (in micelle-bound ASyn) exhibit similar properties to those identified in amyloid structures (high *RD* and significant correlation bias, with negative values of HvT and TvO and strongly positive values of HvO).

The 25–35 fragment, implicated in the onset of Parkinson's disease [6], exhibits amyloid-like properties in micelle-bound ASyn, as do two other fragments—1–25 and 26–98 (which, according to

Reference [5], are responsible for anchoring to the lipid bilayer). Other fragments have been singled out, due to their status in the amyloid protein (see the "Amyloid" tag in Table S1), as discussed further below.

2.3. Structure of the Amyloid Form of ASyn (2N0A)

This part of our analysis focuses on the amyloid form of ASyn, listed in PDB under ID 2N0A and visualized in Figure 2.

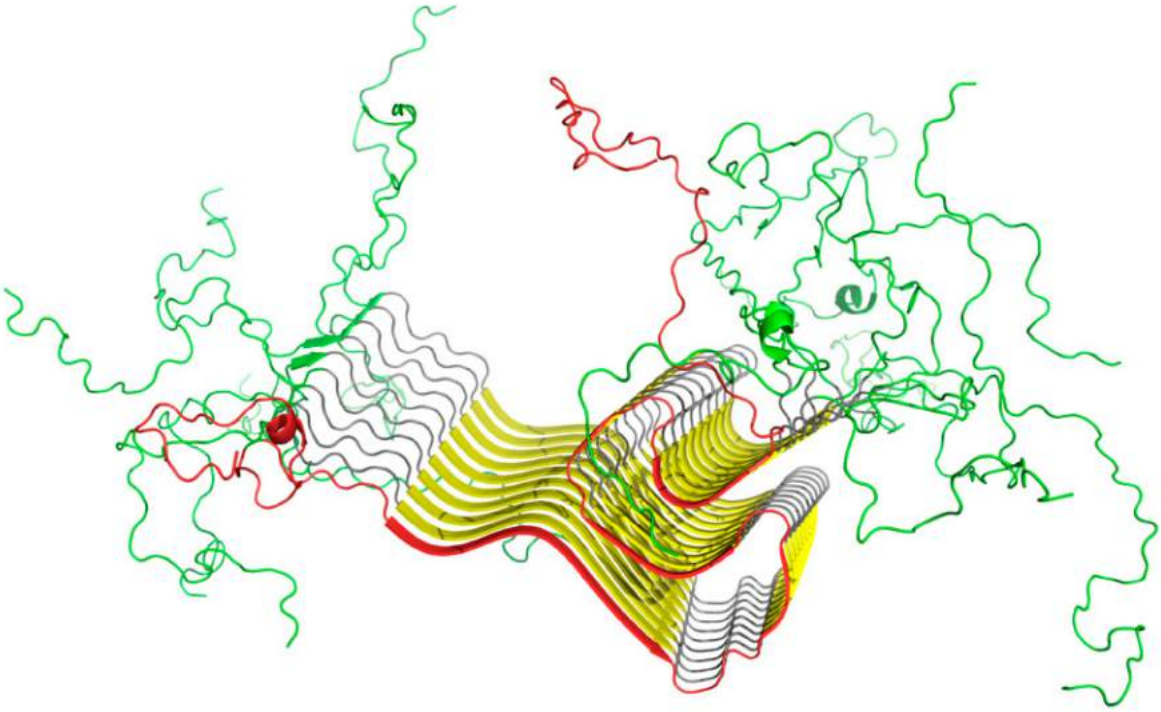


Figure 2. 3D structure of the amyloid form of ASyn, with a clearly distinguished amyloid-like section (30–100). Chain A has been marked red, while random coil fragments (1–29 and 101–140) are shown in green.

Figure 3A presents T and O distributions for the complete amyloid form of Asyn. The theoretical distribution contains two distinct hydrophobicity maxima; however, the observed distribution does not replicate this pattern and instead exposes numerous local maxima, including in the N- and C-terminal fragments, both of which are disordered (Figure 3). The presence of such local maxima in these fragments is likely due to close-range interactions with neighbors belonging to the same chain. This phenomenon differs from the emergence of alternating "bands" of hydrophobicity, although it might suggest that these terminal fragments may also be susceptible to producing amyloid-like structures. This brings up the following question—Why does the fibrillar conformation of ASyn not extend to its N-terminal fragment?

When the fibrillar fragment is treated as part of a larger complex (Figure 3B) or considered on its own (by restricting the 3D Gaussian capsule to that fragment alone–see Figure 3C), the resulting distribution of hydrophobicity is invariably found to contain numerous local maxima, similar to those presented in Figure 3A,B In order to avoid potentially misleading interpretations of the presence of alternating hydrophobicity bands in the N- and C-terminal areas, the presented computations take into account only cross-chain interactions (Figure 4). Here, local maxima resulting from the fibrillar conformation of the 30–100 fragment are clearly visible.

The values listed in Table S2 reveal the unique status of the amyloid, which–when considered in its entirety—Exhibits a distribution of hydrophobicity consistent with the presence of a central hydrophobic core (this observation holds even when the analysis is restricted to cross-chain interactions). The properties of the fibrillar section are also quite well aligned with the monocentric core model. Indeed, eliminating residues 80–83 brings the value of *RD* down to less than 0.5.

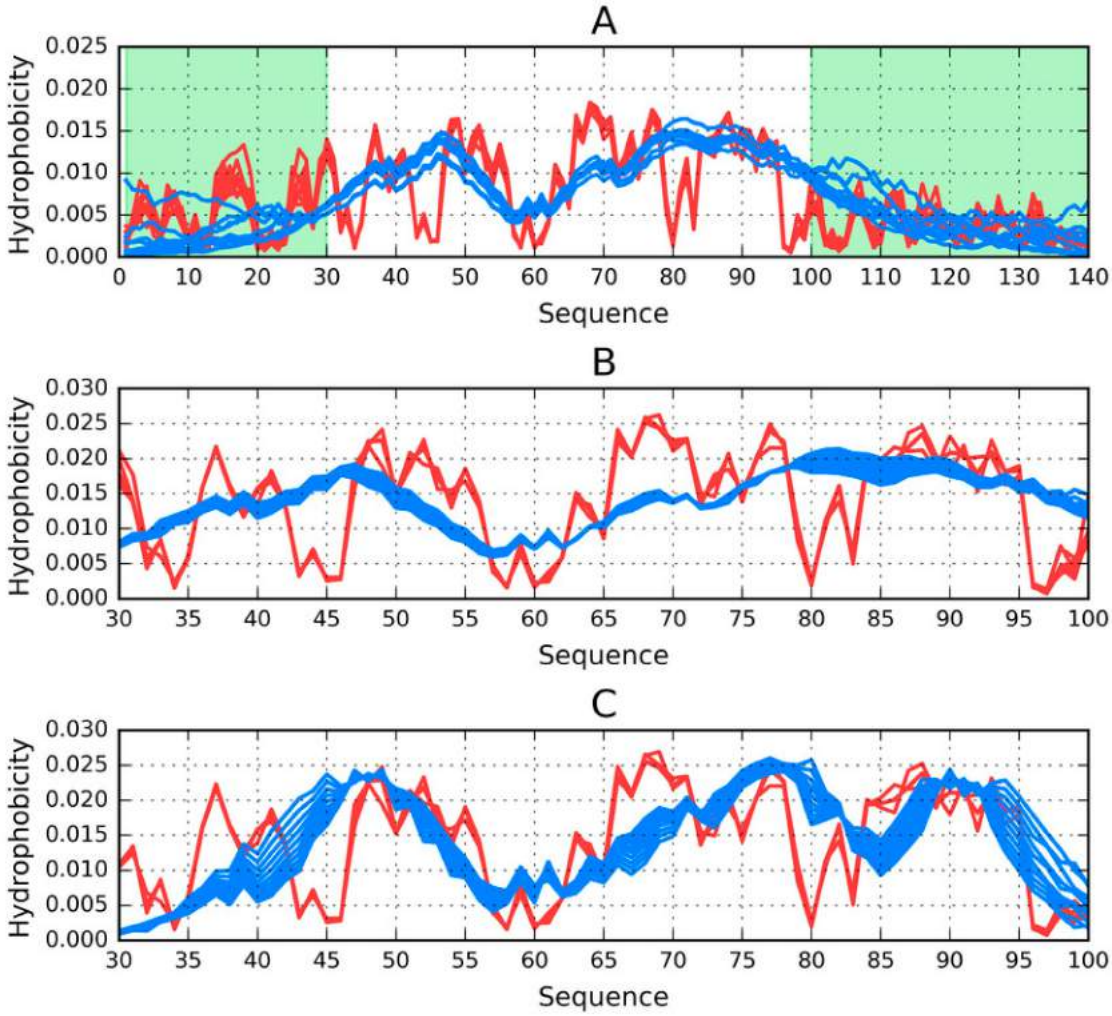


Figure 3. Theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for the ASyn amyloid (2N0A). Each chart shows two profiles for every chain from the complex (10 in total); (**A**)—Calculations performed for the entire complex (1–140). Random coil fragments (1–29, 101–140) are highlighted in green; (**B**)—Calculations performed for the amyloid fragment (30–100) treated as part of the complex; (**C**)—Calculations performed for the amyloid fragment (30–100) treated as an individual molecule.

Table S2 provides quantitative information for the presented complex and its constituent parts, as discussed above.

Assessment of the status of residues 1–30 and 100–140 falls out of the scope of the fuzzy oil drop model, due to the lack of cross-chain interactions. In this specific case, the observed distribution is the result of mutual interactions between residues belonging to each individual chain, and does not reflect the role performed by these fragments in the analyzed complex.

There is good agreement between the theoretical and observed distribution of hydrophobicity in the fibrillar fragment, which may be regarded as surprising. On the basis of the fuzzy oil drop model, we can propose the following explanation: The model assumes that a globular protein emerges as a

result of interactions between its residues and the aqueous solvent. This process causes migration of hydrophobic residues towards the center of the emerging globule, along with exposure of hydrophobic residues on its surface. If, however, the properties of the environment are altered so that the environment is not capable of acting upon the protein chain, forces associated with the intrinsic hydrophobicity of individual residues take over. By "properties of the environment" we refer to a continuous external force field generated by a specific arrangement of water molecules surrounding the protein (note, however, that the structural properties of water in its liquid phase remain poorly understood).

In its micelle-bound form (1XQ8) ASyn is relatively well aligned with the predictions of the fuzzy oil drop model; however, it can only retain stability in the presence of a "permanent chaperone". In vivo, the role of this chaperone typically falls to nerve cells, while the structure listed in PDB is stabilized by a micelle. When deprived of contact with its target molecule, ASyn undergoes significant conformational changes, leading to the emergence of an amyloid structure.

It should also be noted that the fibrillar part of ASyn is surrounded not by water, but by a buffer zone occupied by the randomly coiled terminal fragments, as shown in Table S3 and Figure 2.

2.4. Models Generated by Specialized Software

Our discussion of models produced by specialized software toolkits begins with a study of the structure of individual chains. We singled out the E chain for in-depth analysis, due to its central location in the sample fibril, which makes it the best available match for a fibril of arbitrary length. For the purposes of our analysis, the chain was treated as part of the larger complex, as well as a standalone structure. In the former case, we assessed its alignment with the overall distribution of hydrophobicity in the amyloid fibril, whereas, in the latter case our focus was on alignment with the theoretical distribution of hydrophobicity (T) given by the 3D Gaussian.

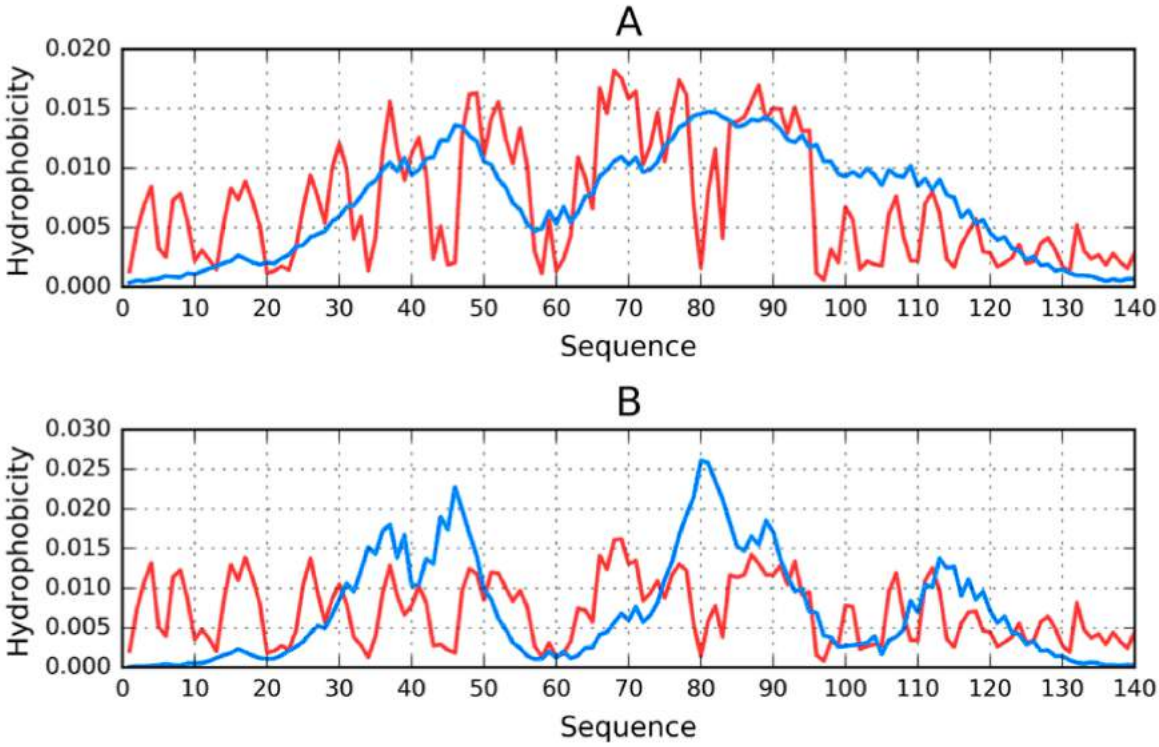


Figure 4. Theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for chain E (central) from the ASyn amyloid (2N0A); (**A**)—Treated as part of the complex; (**B**)—Treated as a standalone structure.

When the E chain is considered as part of the complex, its distribution approximates that which is evident in the complex as a whole (Figure 4A). On the other hand, calculating the distribution of hydrophobicity for the standalone chain reveals—In addition to similar gaps between local peaks at 43–46 and 79–83—An additional gap at 43–46. It should be noted that all such gaps occur within repetitive fragments and correspond to relatively high polarity, revealing strong discordance vs. the theoretical distribution. In addition, the 66–71 fragment (VGGAVV) represents a local excess of hydrophobicity.

The above properties, illustrated in Figure 4B, provide a set of references for the analysis of models generated by our 3D protein structure modeling software.

2.5. Analysis of Models for the 1–140 Fragment of ASyn

In order to facilitate comparative analysis, we computed 3D structural models for the entire sequence of ASyn (1–140). When comparing results, we focus on parameters which indicate alignment (or lack thereof) with the globular hydrophobic core model, revealing that none of the obtained models carries the properties of a spherical micelle.

Table S4 summarizes the results of structure prediction studies carried out for the ASyn (1–140) polypeptide. It appears that this polypeptide is incapable of adopting a distribution of hydrophobicity which would correspond to a spherical micelle, even though some models (mostly those produced by FOD) suggest a globular conformation. In this case, high values of correlation coefficients reveal an amyloid-like pattern with a strong bias towards intrinsic hydrophobicity.

I-Tasser produced a single model with balanced values of TvO and HvO; however, for this model, the value of *RD* (T-O-R) remains high. Robetta generated five distinct models, all of which are characterized by $RD > 0.5$ (except for T-O-H), with balanced values of TvO and HvO. The FOD toolkit produced 500 distinct models. From among these, models with the highest and lowest values of *RD* were selected for further analysis.

Figure 5 reveals strong agreement between T and O in the N- and C-terminal fragments. Major differences concern the expected hydrophobic core, which, in the I-Tasser model, appears to begin at residue 50, while in the PDB structure its beginning is located at residue 70. The Robetta model reveals greater involvement of the N-terminal fragment in shaping the protein's hydrophobic core, which is not evident in the PDB structure. Neither of these structures can be characterized as globular—The reason behind generating various models is to determine whether the ASyn sequence is at all capable of producing a globular fold. Such structures are (for obvious reasons) produced by FOD simulations; however, their status does not correspond to the properties of spherical micelles (in particular, *RD* (T-O-R) is greater than 0.5). Further analysis of distribution profiles highlights the causes of this phenomenon (Figure 3).

Eliminating the 95–102 fragment yields the desired value of *RD* (T-O-R), producing a structure which is quite similar to the target. Notably, FOD computations generate a large variety of models (500 in total), and it should be noted that FOD generally favors the formation of a hydrophobic core resembling a spherical micelle. We may, therefore, conclude that the ASyn sequence is, indeed, incapable of producing a soluble protein. It is also worth noting that the corresponding fragment of the micelle-bound ASyn structure (derived from PDB) exhibits similar discordance. This fragment has been highlighted in Figure 5A (1–140) and Figure 5B (1–140), as well as in Figure 1, and appears to be the causative factor determining the presented conformational properties of ASyn. It retains strong discordance in Robetta models, while I-Tasser deals with it by exposing it on the surface.

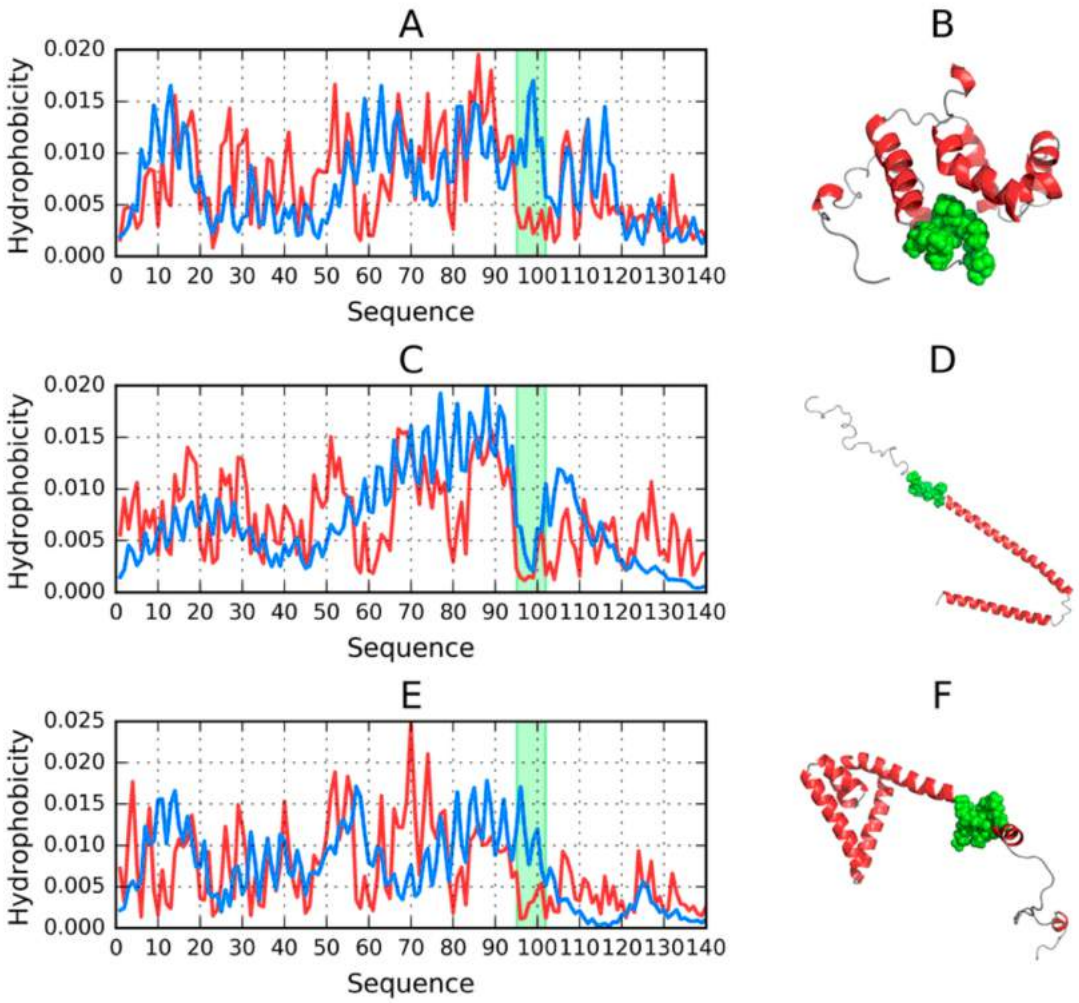


Figure 5. 3D visualizations along with theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for models of the 1–140 fragment produced by each software package and characterized by the lowest RD. The 95–102 fragment is highlighted in green; (**A**,**B**)—FOD_1_(202); (**C**,**D**)—ITASSER_1_(1); (**E**,**F**)—ROBETTA_1_(2).

2.6. Structure of the N-Terminal Fragment in FOD, I-Tasser and Robetta Models (1–30 aa)

In the ASyn amyloid form listed in PDB (2N0A) the 1–30 fragment does not belong to the fibril. Instead, it is characterized as a random coil, with a disordered, nonrepetitive structure. It is, therefore, interesting to speculate whether, under favorable conditions, this fragment may produce a globular fold. To address this issue, we carried out using FOD, I-Tasser and Robetta.

Table S5 provides a summary of results, revealing strong variability of models produced for the N-terminal fragment of the ASyn chain. This fragment was selected for analysis, due to its disordered structure in the ASyn amyloid (2N0A). Results can be described as highly variable. The FOD model produced a distribution consistent with the theoretical model, suggesting the presence of a globular form with a centrally located hydrophobic core. I-Tasser models also hint at the possibility of generating this type of structure, although they provide two alternative structures. Regarding Robetta, all of its models diverge from the theoretical distribution with a clear preference for helical folds dominating the entire fragment. Helical folds are also evident in I-Tasser and FOD models, although in their case the presence of twists results in globular conformations (particularly in FOD models). Both FOD and I-Tasser structures are dominated by the TvO correlation coefficient, while for other models, the HvO coefficient prevails.

Figure 6 provides a visualization of profiles and 3D structures of selected models.

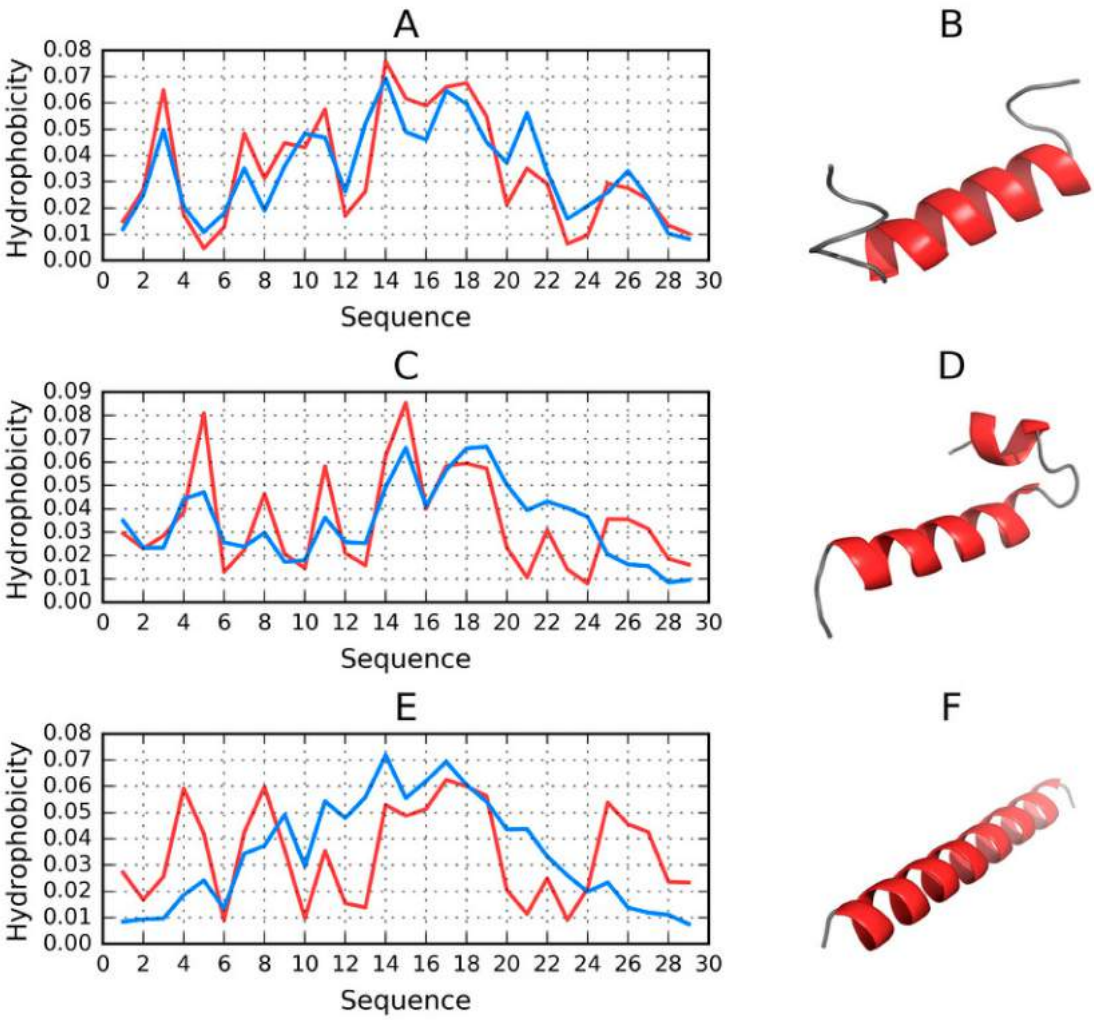


Figure 6. 3D visualizations along with theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for models involving the 1–30 fragment, produced by each software package and characterized by the lowest RD. (**A**,**B**)—FOD_1_(053); (**C**,**D**)—ITASSER_1_(3); (**E**,**F**)—ROBETTA_1_(4).

One conclusion which can be drawn from the analysis of the 1–30 sequence is that it admits a globular conformation with a centralized hydrophobic core (in the case of the FOD model, some structures selected from among the 500 output models exhibit this status).

2.7. Structure of the 30–100 Fragment of the ASyn Polypeptide According to FOD, I-Tasser and Robetta

This fragment should be regarded as particularly important, since it represents the fibrillar core of the ASyn amyloid. Consequently, it is interesting to speculate whether it can produce a spherical micelle.

The summary presented in Table S6 suggests that only two models approximate the spherical micelle. Their status is also visualized in Figure 7.

FOD_1 satisfies $RD < 0.5$, although its 72–77 fragment deviates from the theoretical distribution of hydrophobicity (by being overly hydrophobic). Eliminating this fragment results in much better alignment between both profiles, with RD (T-O-R) = 0.394.

Similarly, the structure of the I-Tasser model exhibits a local excess of hydrophobicity in its 79–94 fragment. Eliminating this fragment brings RD down to 0.468, suggesting good alignment with the spherical micelle pattern for the majority of the proposed structure.

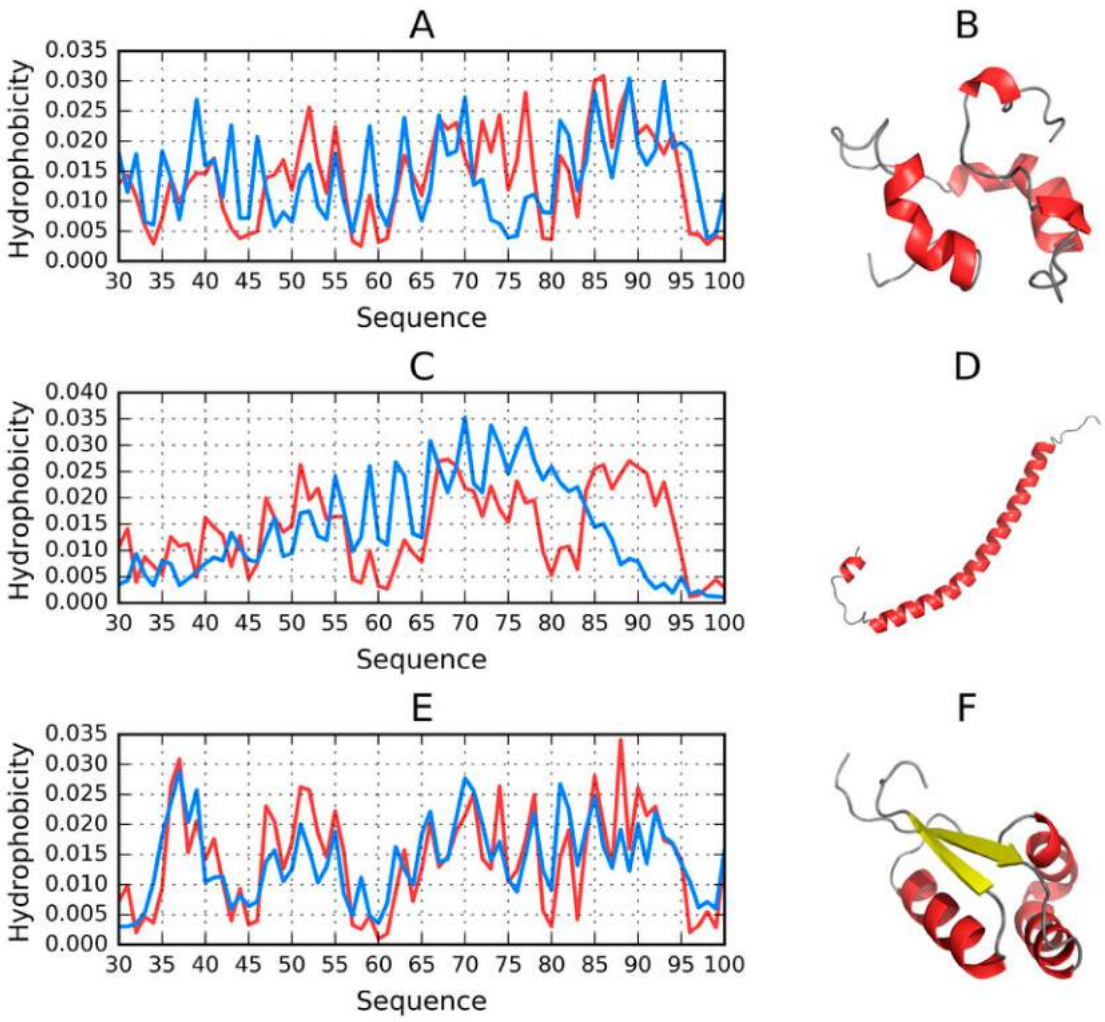


Figure 7. 3D visualizations along with theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for models of the 30–100 fragment, produced by each software package and characterized by the lowest RD. (**A**,**B**)—FOD_1_(289); (**C**,**D**)—ITASSER_1_(1); (**E**,**F**)—ROBETTA_1_(2).

Robetta also generates a conformation largely consistent with the spherical pattern. It is, in fact, the only toolkit which predicts the presence of beta folds in the analyzed chains (in models 1 and 2). The remaining models (3–5) are purely helical. The model visualized in Figure 7 represents the best match for the spherical pattern from among all analyzed models.

Summarizing our study of the conformational capabilities of the 30–100 fragment of ASyn, we need to note that this sequence is capable of adopting conformations which correspond to the spherical micelle pattern—Even though the vast majority of proposed models represent other patterns.

2.8. Comparative Analysis of Models Obtained for the 100–140 Fragment of ASyn

The C-terminal fragment of the ASyn amyloid (100–140) (2N0A) has been experimentally determined to adopt a highly disordered conformation, usually described as a random coil, with no obvious secondary folds. It is, therefore, interesting to speculate whether this fragment is at all capable of achieving an orderly secondary structure. In order to answer this question, we calculated a series of models using FOD, I-Tasser and Robetta.

Summarizing the parameters listed in Table S7 we may conclude that the C-terminal fragment (100–140)—which becomes a random coil in the protein's micelle-bound form—May adopt a globular conformation. From among the models produced by Robetta, only two structures do not conform to this pattern (*RD* (*T-O-R*) ≥ 0.5).

Most of the generated models are dominated by short helices and random coil fragments. Again, Robetta provides an exception, with a notable presence of beta folds in two models for which $RD < 0.5$.

The fact that the C-terminal fragment readily adopts globular conformations is all the more surprising given its high polarity and sparsity of hydrophobic residues.

All structures which satisfy $RD < 0.5$ are also characterized by balanced values of correlation coefficients, with no significant bias.

In conclusion, it appears that the 100–140 fragment of ASyn is capable of adopting a globular conformation with a prominent hydrophobic core, approximating a spherical micelle. Of course, this does not imply that the fragment retains this capability in the context of a larger structure—As evidenced by the lack of globular models for the complete chain of ASyn (1–140) (Figure 8).

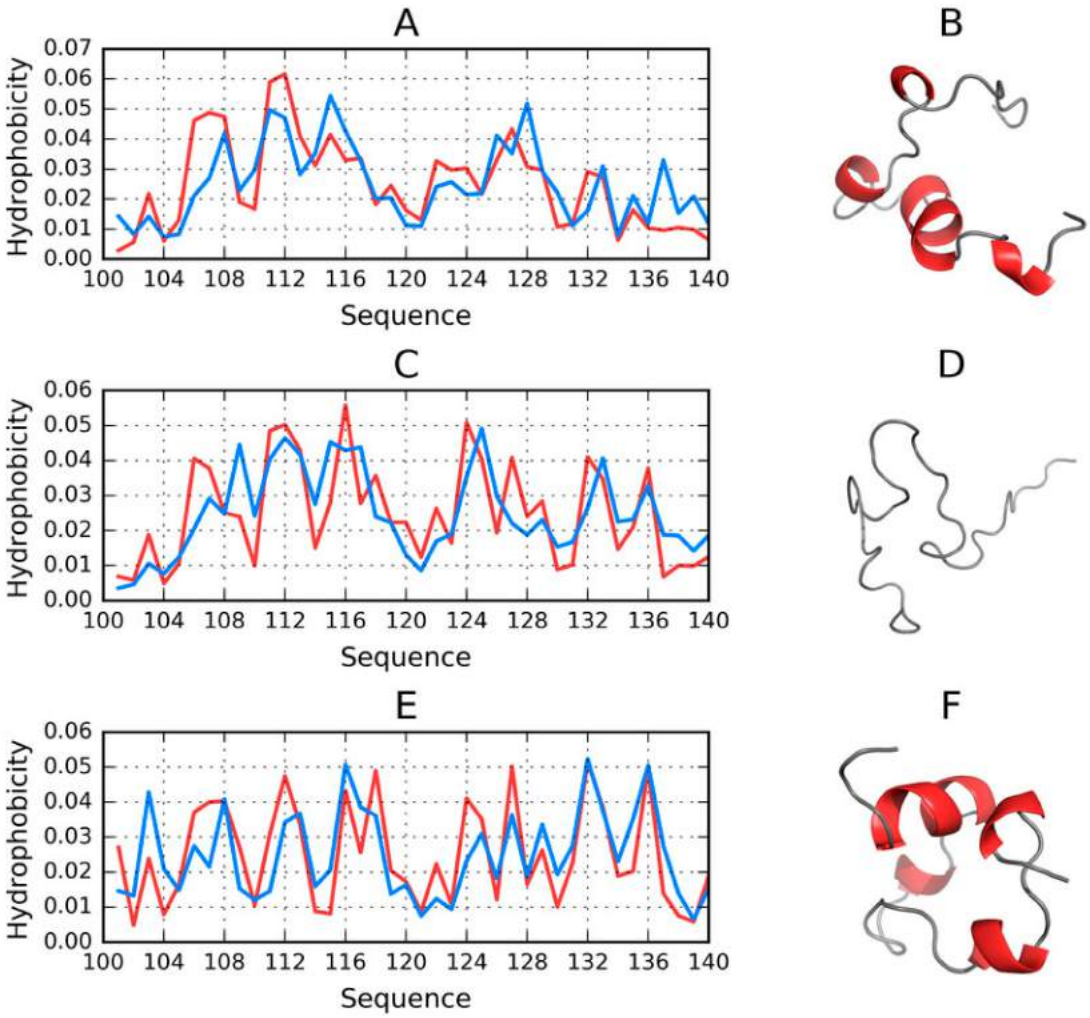


Figure 8. 3D visualizations, theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for models of 100–140 fragment, produced by each software package and characterized by the lowest RD. (**A**,**B**)—FOD_1_(002); (**C**,**D**)—ITASSER_1_(5); (**E**,**F**)—ROBETTA_1_(1).

2.9. Effect of Mutations on Amyloid Transformation

The effect of experimentally observed mutations on the amyloid transformation of ASyn may be studied on the basis of the fuzzy oil drop model, which provides a way to predict local conformational changes resulting from changes in the underlying distribution of hydrophobicity. This is visualized in Figure 9, which highlights several mutation loci (A53T, E46K, A30P and H50Q) [33]. In all of these cases, the intrinsic hydrophobicity of the substituent residue is lower than that of the original residue (according to any acknowledged intrinsic hydrophobicity scale).

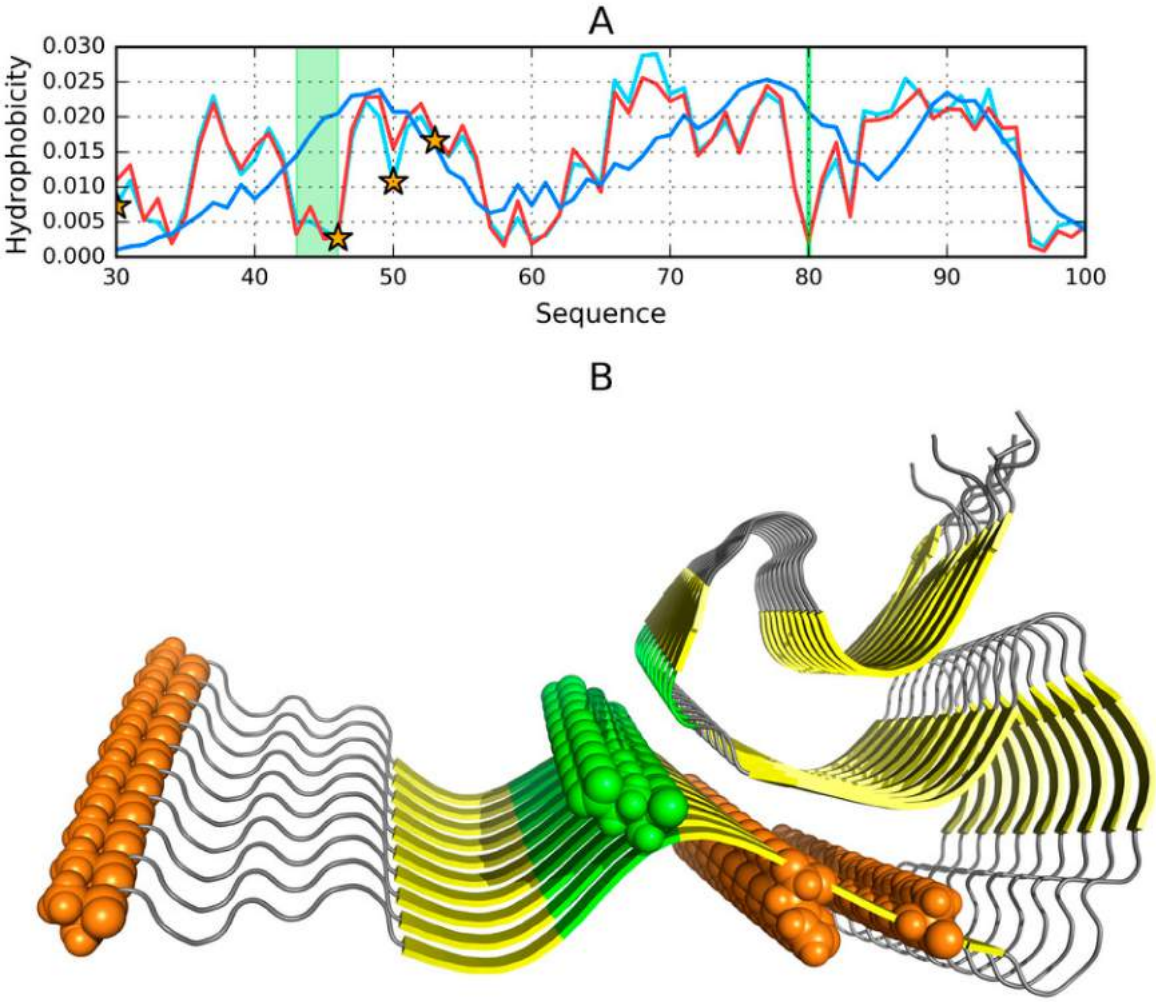


Figure 9. Presentation of the fibrillar part (30–100) of ASyn (2N0A); (**A**)—Theoretical (T, blue) and observed (O, red) hydrophobicity density distribution profiles for chain E. The teal chart represents the observed profile for the mutated sequence; (**B**)—3D visualization; Orange stars in Figure 9A, and orange/green spheres in Figure 9B correspond to loci of mutations which lower intrinsic hydrophobicity: A30P, E46K, H50Q and A53T. Green fragments correspond to residues No. 43–46 and 80, which–when eliminated from FOD computations—Produce a distribution consistent with the theoretical model (*RD* lowered from 0.506 to 0.490).

The diagrams are shown in Figure 9A, when confronted with the corresponding locations indicated in Figure 9B, reveal the specificity of the mutated residues.

The effect of A30P is described as neutral with regard to the protein's susceptibility to undergo amyloid transformation. From the point of view of the fuzzy oil drop model, residue 30, which is located in the fibril's outer layer, may admit a local reduction in hydrophobicity [34]. Residues 46, 50 and 53 are also found in the outer layer. Reducing their hydrophobicity may, therefore, mediate entropically favorable interactions with the aqueous solvent, stabilizing the ribbonlike micelle.

Analysis of Figure 9 suggests favorable conditions for the emergence of a ribbonlike micelle, with an internal hydrophobic core propagating along the fibril's axis.

2.10. Status of Selected Fragments Identified in Other Publications as Linked to Amyloid Transformation in Parkinson's Disease

The selected fragments are implicated in the onset of Parkinson's disease [5]. The status of this fragment in the micelle-bound form of ASyn (1XQ8) is clearly amyloid-like, whereas, in the true

amyloid (2N0A) it adopts a peculiar conformation, while retaining $RD < 0.5$. The fragment itself is located at the junction of the random coil and fibrillar sections. We have previously demonstrated that the 1–30 fragment may fold as a globule, while the 30–100 fragment lacks this capability. Thus, the 25–35 fragment may potentially play the role of an amyloid seed.

Fragments tagged as amyloid-like in Tables S2 and S3 can be identified by searching for a distribution of hydrophobicity which stands in opposition to the globular pattern. Their status also mostly reflects conditions which are likely to be encountered in amyloids. While investigating the causes of amyloid transformation, it is useful to refer to the sequences of such fragments. This work is currently ongoing and will be the subject of a separate publication.

Table S8 provides a summary of results obtained for fragments which other authors identify as amyloidogenic. It is clear that the structural properties of the generated models vary widely. Values listed in boldface correspond to classic amyloid-like conditions, which are understood as a combination of high values of RD, low (potentially negative) values of HvT and TvO and strongly positive values of HvO. Taken together, these parameters indicate that the conformation of the given fragment is driven by the intrinsic properties of its component residues.

The aim of this presentation is to highlight the structural variability of amyloidogenic fragments, while establishing that they generally diverge from the centralized hydrophobic core model. Given the latter, it should come as no surprise that these fragments are capable of producing an alternative (fibrillar) structural pattern.

3. Discussion

Amyloid nucleation of ASyn is often linked to the 68–82 fragment, which may initiate the early assembly of ASyn [34] Assessment of the role which this fragment plays in the stabilization of the ASyn amyloid is facilitated by the characteristics of the 61–95 fragment. Of note is the hydrophobic seed emerging in the 68–82 area, consistent with that fragment's status, shown in Figure 9A, and indicating an atypical, amyloid-like distribution of hydrophobicity.

The status of the 1–30 and 101–140 fragments remains puzzling. The presence of identical sequences in adjacent chains may create favorable conditions for complexation and generation of variably hydrophobic bands, similar to what can be observed in the central section (30–100). The 1–30 fragment was only capable of adopting a centric fold in FOD-based calculations, while Robetta did not produce any corresponding globular structures. Regarding the 30–100 fragment, only two of the obtained models predict a status consistent with a soluble protein, while in the case of the 100–140 fragment most models indicate the capability for producing a globular form with a coherent hydrophobic core. Considering the relative hydrophilicity of participating residues, this observation suggests that the terminal fragment "seeks" an alternative (non-fibrillar) conformation and may attain it under certain conditions.

Intensive studies concerning the ability of ASyn and Tau to undergo amyloid transformation often link their properties to the presence of intrinsically disordered fragments and polyperolin-like forms (Ramachandran map), regarded as likely amyloid seeds [35]. In general, the tertiary and quaternary structure of proteins stems from cooperative interaction between individual residues and the aqueous solvent, whose presence is an essential prerequisite of life. This cooperation typically results in the formation of a hydrophobic core overlaid by hydrophilic residues, roughly accordant with the 3D Gaussian distribution of hydrophobicity. Analysis of individual protein domains confirms that almost invariably comply with this pattern [36], while any local discordances are usually associated with biological function [32]. Cooperative tendencies, which produce a 3D Gaussian distribution of hydrophobicity in a globular protein, may instead result in a 2D Gaussian distribution when an amyloid unit chain is considered [31].

The comparative analysis of ASyn amyloids and tau proteins presented in Reference [37] is also of interest to us. The FOD model underscores the role of the environment, and of the solvent in particular. The role of membranes as a factor promoting amyloidogenesis is discussed in Reference [38,39].

The link between mutations and amyloidogenesis is undisputed [40]. Nevertheless, the environment—Especially an altered environment (in terms of pH and ionic strength)—Is frequently indicated as a factor which promotes the amyloid transformation of ASyn, including its mutated forms (A30P, E46K, G51D and A53T) in α-synuclein fibrils [41].

Analogies with prions may provide a lead in the search for the causes of amyloid transformation [42,43].

The search for new drugs focuses on dopamine (Dopa), amphotericin-B (Amph), epigallocatechingallate (EGCG), and quinacrinedihydrochloride (Quin) as factors affecting the oligomerization of ASyn [44,45].

Co-immunoprecipitation is also considered as a potential therapeutic technique [46]. Selection of the 30–100 fragment based on analysis of the ASyn amyloid structure remains consistent with the outcome of experimental studies which single out the NAC fragment (60–100) as particularly prone to amyloid transformation [47] An open question is why the fragments at 1–30 and 100–140, despite repeating the same sequence in all unit chains, do not participate in propagation of the fibril. Based on the analysis presented in this paper, we suggest that their tendency to generate a globular structure with a prominent hydrophobic core may preclude fibrillization. Strong hydrophilicity of residues comprising the 100–140 fragment enables penetration of water; however, the presence of the adjacent fibril (at 30–100) prevents the formation of a globule. Consequently, the random coil remains the only possible alternative. We should also note the lack of fibrillar properties in the 1–30 fragment, which, given its structural properties, might be suspected as being capable of producing a fibril. The ongoing analysis focuses on the properties of sequences which comprise known amyloids, as well as each fragment of the ASyn polypeptide (with a publication currently in preparation).

The fuzzy oil drop model describes and expects a spherical, centralized hydrophobic core. ASyn proves that a centralized hydrophobic core may also be present in a fibrillar structure. Summarizing the presented results, we may propose that when the chain is unable to "resolve" to a spherical micelle, with all of its hydrophobic residues isolated in the central part and all hydrophilic residues exposed on the surface, it instead adopts a ribbon-like micellar conformation. This structure is characterized by advantageous entropic effects, including the isolation of hydrophobic residues within a central (in the sense of a horizontal cross-section) band stretching along the fibril's axis. The ASyn amyloid is, in many respects, unique, especially when compared to Aβ(1–42) amyloids. As a result, its analysis reveals interesting aspects of the amyloid transformation process. For example, it turns out that the presence of long disordered N- and C-terminal fragments promote isolation of the central fibril, which contains the aforementioned bandlike hydrophobic core. Additionally, we reveal a common mechanism, observed in both globular and fibrillar structures, where the polypeptide chain attempts to isolate its hydrophobic residues from direct contact with the solvent. We are also currently involved in performing a comparative analysis of various amyloid structures listed in PDB, in search for general amyloid formation mechanisms which apply regardless of specific sequential properties [31].

As it is shown in this paper, the comparative analysis reveals no preference for globular forms for the amino acids sequence as it appears in ASyn. The models represent the structures expressed by $RD > 0.5$ (see Tables in Supplementary Materials). It suggests that the sequence specificity directs the folding process toward other than globular forms. It is also observed in other amyloids, the structure of which is available in Protein Data Bank [39]. The generalization of rules directing the folding process toward ribbon-like micelles is the object of currently conducted analysis.

4. Materials and Methods

4.1. Data

The structure of the amyloid form of ASyn (PDB ID: 2N0A) consists of 10 chains, where—Unlike other amyloids listed in PDB [10–14,48]—Only a portion of each chain adopts a fibrillar form (in contrast, the entire chains of Aβ (1–42) amyloids participate in the formation of a fibril. In the case of ASyn, both the N-terminal fragment (1–30) and the C-terminal fragment (100–140) are random

coils, while the 31–100 fragment can be characterized as a typical amyloid fibril in all 10 chains. This structural diversity, along with its underlying causes, make ASyn an interesting study subject.

Based on the above observations and on the published, experimentally determined structure of the ASyn amyloid (PDB ID: 2N0A), we distinguish three fragments: 1–30 (random coil), 31–100 (fibril) and 101–140 (random coil). Notably, PDBSUM [49] identifies the following beta folds in ASyn: 38–55, 60–67, 70–78, 81–84 and 88–97. PDB also provides structural information for a single chain of ASyn (PDB ID: 1XQ8). This micelle-bound form comprises two helical fragments (2–38 and 44–93) linked by a hairpin with a tight bend at 39–43. The C-terminal fragment (identified as 94–140) is, again, described as a random coil. We will use this structure as the reference for the amyloid form of ASyn, as well as for models produced by our software.

4.2. Obtaining Alternative ASyn Polypeptide Models

In order to determine whether ASyn is capable of adopting other conformations than those previously listed, we have carried out using I-Tasser (University of Michigan, Ann Arbor, MI, USA) [18] and (Robetta Department of Biochemistry, University of Washington, Seattle, WA, USA) [17]—Specialized protein structure prediction toolkits, both of which rank among the best in the CASP (Critical Assessment of Structure Prediction) challenge [19]. They both operate upon independently defined force fields and apply different algorithms (e.g., for energy minimization)—Thus, they can be expected to produce alternative structural forms for the ASyn sequence. In line with the CASP challenge rules, Roberta was used to producing five models, while I-Tasser generated between 1 and 5 models depending on the length of the input chain.

All calculations were carried out using online servers [29,30]. In addition, a separate calculation was performed using a software toolkit based on the FOD model (Jagiellonian University—Medical College, Krakow, Poland) [20–22], which, in addition to internal free energy optimization, also optimizes interactions with an external force field representing the aqueous solvent. This force field is mathematically defined as a 3D Gaussian and its presence results in internalization of hydrophobic residues, with the attendant exposure of hydrophilic residues on the protein surface. The presented calculations were conducted using the Gromacs (Rijksuniversiteit Groningen, Groningen, Netherlands) package [50], with optimization of internal and external force fields carried out in an interleaved fashion. Output models were ranked by their values of the *RD* (Relative Distance—See the section titled "Fuzzy oil drop model—Protein folding with the preferential generation of a hydrophobic core"), allowing us to select models which most closely approximate a spherical micelle.

According to the fuzzy oil drop model, the presence of the aqueous solvent favors the generation of globular structures containing hydrophobic cores—Therefore, the model can be used to determine whether a given polypeptide is capable of achieving a globular confirmation. FOD application requires the user to provide an input (starting) structure. In our case, we used the structure of an individual polypeptide belonging to the presented fibril (2N0A). The goal of these calculations was to identify alternative folding patterns which may manifest themselves under altered environmental conditions. Protein structure prediction was carried out for the entire ASyn chain (1–140), as well as for the previously identified fragments (1–30, 30–100 and 100–140).

FOD based computations were performed at the Academic Computing Centre CYFRONET AGH using resources provided by the PL-Grid (University of Science and Technology, Kraków, Poland) infrastructure [51].

4.3. Fuzzy Oil Drop Model—Protein Folding with Preferential Generation of a Hydrophobic Core

The base model has been thoroughly described in numerous publications [20–22]. At its core rests the assumption that a globular protein contains a hydrophobic core, which (in its "idealized" or "theoretical" version) can be mathematically modeled as a 3D Gaussian superimposed upon the protein body. Thus, each effective atom (averaged-out positions of all atoms comprising a single residue) has a theoretical hydrophobicity value (T_i) given by the Gaussian. In addition, each residue also carries

the so-called observed hydrophobicity (O_i), which is dependent on its hydrophobic interactions with adjacent residues. Such interactions depend on the separation between residues and on their intrinsic hydrophobicity, which is defined according to the scale proposed in Reference [52].

The following algorithm is applied to compute the theoretical and observed hydrophobicity in a protein molecule:

The molecule is oriented in such a way that A—Its geometric center coincides with the origin of the coordinate system; B—The longest axis of the molecule coincides with the *X* axis; C—The line connecting the two most distal atoms (projected on the YX plane) corresponds to the *Y* axis.

For each axis the greatest separation between any two atoms is computed and subsequently multiplied by 6, yielding values of σ_x σ_y σ_z, respectively.

For each residue, the position of its effective atom is calculated by averaging out the positions of all atoms which belong to the given residue.

The value of the 3D Gaussian at the point corresponding to the effective atom is taken as theoretical hydrophobicity (again, for the given residue). This can be mathematically expressed by the following formula:

$$\widetilde{H}t_j = \frac{1}{\widetilde{H}t_{sum}} \exp\left(\frac{-\left(x_j - \overline{x}\right)^2}{2\sigma_x^2}\right) \exp\left(\frac{-\left(y_j - \overline{y}\right)^2}{2\sigma_y^2}\right) \exp\left(\frac{-\left(z_j - \overline{z}\right)^2}{2\sigma_z^2}\right) \quad (1)$$

where $\widetilde{H}t_j$ describes the theoretical hydrophobic density (hence, the *t* index) at point *j*.

Hydrophobic interactions between residues are calculated based on Levitt's formula [27], which depends on intrinsic hydrophobicity (according to some predefined scale—See Reference [22]) and on the separation between interacting residues:

$$\widetilde{H}o_j = \frac{1}{\widetilde{H}o_{sum}} \sum_{i=1}^{N} \left(H_i^r + H_j^r\right) \begin{cases} \left[1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right] & \text{for} \\ 0 \text{ for } r_{ij} \leq c \end{cases} \quad (2)$$

where *N* is the number of amino acids in the protein, $\widetilde{H}_i^r$ expresses the hydrophobic parameter of the *i*-th residue, while r_{ij} expresses the distance between two interacting residues (*j*-th "effective side chain" and *i*-th "effective side chain"). C is the cutoff distance—Assumed equal to 9 Å.

All values of Ti and Oi are subjected to normalization to ensure that they add up to 1.

Plotting the values of Ti and Oi for successive amino acids in the polypeptide chain reveals differences between both distributions, as well as fragments where they remain closely aligned.

In order to quantitatively express the differences, divergence entropy [28] is computed according to the following formula:

$$D_{KL}\left(p \middle| p^0\right) = \sum_{i=1}^{N} p_i \log_2\left(p_i / p_i^0\right) \quad (3)$$

where p_i—Observed probability, p_i^0—Reference probability, N—Number of residues in the chain

Given that Equation (3) yields a measure of entropy, the resulting value cannot be interpreted on its own. To make meaningful observations, another reference distribution is required. In our case, this second "boundary" distribution is called R and denotes a case where all residues carry identical hydrophobicity (which is similar to the distribution of electrostatic interactions in many proteins [52]). Accordingly, the value of Ri for each residue is 1/N, N being the number of residues in the chain.

Comparing D_{KL} for the O-T relation and for the O-R relation tells us whether the observed distribution more closely approximates the 3D Gaussian form (O-T < O-R) or the uniform pattern

(O-T > O-R). In order to avoid having to deal with two distinct values, we compress them into a single parameter, referred to as Relative Distance, as follows:

$$RD = \frac{O|T}{O|T + O|R} \tag{4}$$

As noted above, $RD < 0.5$ is interpreted as accordance between O and T, indicating the presence of a hydrophobic core. In all other cases, we assume that the protein body lacks a clearly defined core.

The concept of *RD*—Relative Distance also enables us to assess the status of a specific fragment of the polypeptide chain (following normalization of Ti, Oi and Ri values). Eliminating residues for which Ti strongly deviates from Oi furthermore tells us which part of the polypeptide chain gives rise to the hydrophobic core.

The presented analysis may also be carried out for a case where the uniform distribution ® is replaced by a distribution reflecting the intrinsic hydrophobicity of each residue (denoted H). In this case, $RD < 0.5$ indicates the presence of a hydrophobic core, while $RD > 0.5$ suggests that the observed conformation of the protein is determined mostly by the intrinsic ("selfish") properties of its constituent residues, which override the previously described "cooperative" interactions with the solvent.

Figure 10 provides a graphical depiction of the presented model.

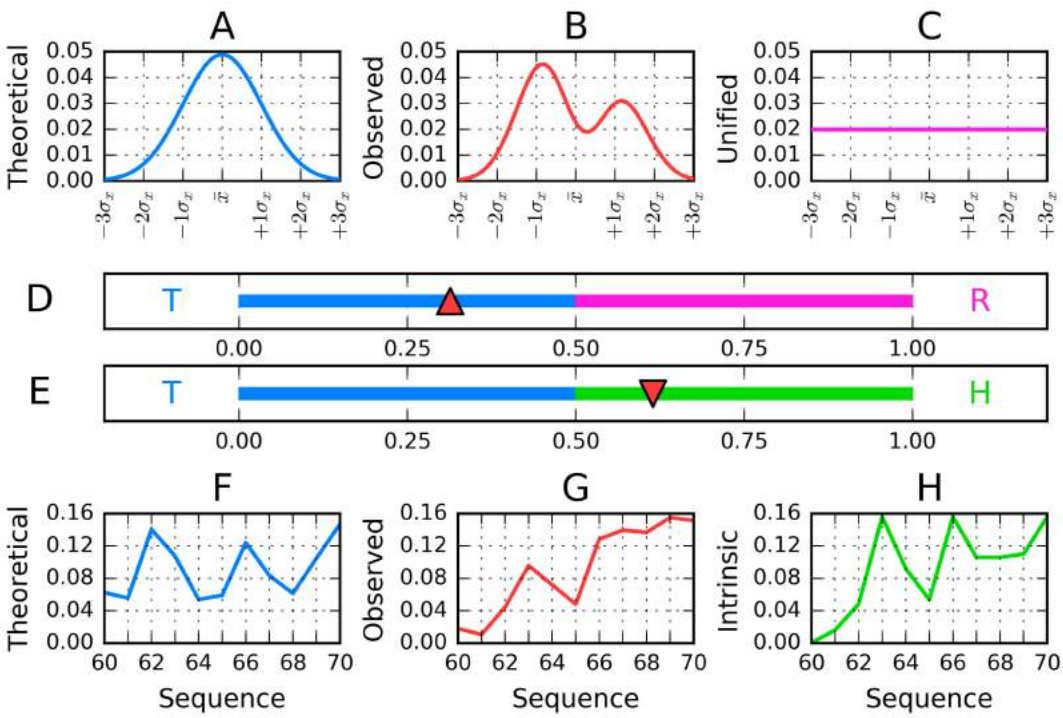


Figure 10. Visualization of the presented model, reduced to a single dimension for the sake of clarity: (**A**)—Gaussian distribution superimposed onto the protein molecule (T); (**B**)—Observed distribution in the molecule under consideration (O) (**C**)—Uniform distribution (R) (**D**)—Part of the Gaussian distribution plotted for the selected fragment ($T_{i–f}$) (**E**)—Observed distribution plotted for the selected fragment ($O_{i–f}$) (**F**)—Intrinsic distribution plotted for the selected fragment ($H_{i–f}$) (**G**)— *RD* scale for the example illustrated in Figures (**A**–**C**,**H**)—*RD* scale for the example illustrated in Figures (**D**–**F**) ("f" indicates "fragment").

The polypeptide chain folding procedure, in addition to optimizing internal free energy, also minimizes the difference between T_i and O_i, directing hydrophobic residues towards the interior of the protein body, while exposing hydrophilic residues on its surface. The degree of similarity between both distributions is quantitatively expressed by a parameter referred to as Relative Distance (RD), computed in accordance with the Kullback-Leibler divergence entropy (D_{KL}) formula [53]. However,

in order to properly interpret the value of RD, another reference distribution is required in addition to T and O. For this reason, we introduce a distribution referred to as R, which assigns a uniform value of hydrophobicity (equal to 1/N) to each residue, where N is the number of residues in the chain. Computing *RD* for this trio of distributions (T-O-R) tells us whether the observed distribution approximates the theoretical distribution and includes a hydrophobic core ($RD < 0.5$) or is more closely aligned with the uniform distribution ($RD \geq 0.5$). In this sense, R is treated as a "polar opposite" of T since it lacks any concentration of hydrophobicity anywhere in the protein body. In addition to the above, it is also interesting to calculate RD for a different pair of reference distributions, where R is replaced by a distribution reflecting the intrinsic hydrophobicity of each residue, denoted H. Altogether, the presented procedure yields two separate values of RD—One for the T-O-R variant and one for the T-O-H variant. These values can then be calculated either for entire chains or for selected fragments (treated as distinct structural units).

The structural properties of each input chain are also described by a set of correlation coefficients—HvT, TvO and HvO, which express pairwise differences between various distributions of hydrophobicity. In light of the presented analysis, globular proteins containing prominent hydrophobic cores should be characterized by low values of *RD* (far below 0.5) and balanced values of correlation coefficients, whereas, in amyloid proteins *RD* should remain high and the values of correlation coefficients should vary significantly (high HvO and low—Or even negative—TvO and HvT). Such conditions suggest that the given structure lacks a centralized hydrophobic core and that its conformation is determined by the individual properties (intrinsic hydrophobicity) of each participating residue. The amyloid does not "align" to the aqueous solvent and instead exhibits a linear pattern [23–26], where bands of high and low hydrophobicity alternate along the fibril's axis. Unlike in globular proteins where the synergy between various environmental factors can be observed, amyloids are solely dependent on the intrinsic properties of their residues. This phenomenon is further discussed in Reference [22].

The models produced by Robetta, I-Tasser and FOD for the entire ASyn chain (1–140) and for its fragments (1–30, 30–100 and 100–140) have been subjected to comparative analysis based on the values of *RD* (T-O-R and T-O-H), as well as the aforementioned correlation coefficients (HvT, TvO and HvO). Similarly, to previous analyses, we identify a set of criteria which suggest susceptibility to producing amyloid structures: High values of *RD* and HvO coupled with negative values of HvT and TvO. Our comparative analysis highlights relations between individual models, as well as with regard to the structures listed in PDB (2N0A and 1XQ8).

5. Conclusions

Analysis of ASyn models suggests a weak preference for adopting globular conformations, with notably different properties exhibited by each fragment (1–30, 30–100 and 100–140). The FOD model generates globular structures by taking into account the active involvement of the solvent as an external force field, guiding hydrophobic residues towards the center of the protein body and exposing hydrophilic residues on its surface. This process assumes that the structure of the protein is tightly dependent on the properties of the solvent, and that—Consequently—Changes in the solvent's properties may affect the conformations attained by polypeptide chains, as indeed experimentally observed [53]. This paper represents the part of the complex analysis of amyloid structures [23–25] searching for a common mechanism of amyloid formation. The model so far proposed in Reference [31] is to treat the globular structure of proteins as the result of the influence of the external force field of 3D Gauss form. However, the environment represented by 2D Gauss function promotes the structural forms observed in amyloids. The simulation of the folding process in the presence of the external force field of these two categories is currently conducted by the group.

Supplementary Materials: The following are available online.

Author Contributions: Conceptualization, I.R. and L.K.; methodology, I.R.; software, D.D.; M.G.; M.B.; validation, I.R., and L.K.; formal analysis, I.R.; investigation, I.R.; resources, D.D.; M.G.; M.B.; data curation, D.D.; M.G.; M.B.; writing—Original draft preparation, I.R.; writing—Review and editing, I.R., and M.B.; visualization, M.B.; supervision, L.K.; project administration, I.R.; funding acquisition, I.R. and M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jagiellonian University—Medical College, grant numbers N41/DBS/000211 and N41/DBS/000208.

Acknowledgments: The authors wish to thank Piotr Nowakowski, Anna Śmietańska and Zdzisław Wiśniowski for technical assistance. Acknowledgements are also due to ACK Cyfronet AGH University of Science and Technology in Kraków.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviation

ASyn	α-synuclein

References

1. Bonini, N.M.; Giasson, B.I. Snaring the Function of α-Synuclein. *Cell* **2005**, *123*, 359–361. [CrossRef]
2. Iwai, A.; Masliah, E.; Yoshimoto, M.; Ge, N.; Flanagan, L.; De Silva, H.R.; Kittel, A.; Saitoh, T. The precursor protein of non-Aβ component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. *Neuron* **1995**, *14*, 467–475. [CrossRef]
3. Visanji, N.P.; Wislet-Gendebien, S.; Oschipok, L.W.; Zhang, G.; Aubert, I.; Fraser, P.E.; Tandon, A. Effect of Ser-129 Phosphorylation on Interaction of α-Synuclein with Synaptic and Cellular Membranes*. *J. Boil. Chem.* **2011**, *286*, 35863–35873. [CrossRef] [PubMed]
4. Lykkebo, S.; Jensen, P.H. Alpha-Synuclein and Presynaptic Function: Implications for Parkinson's Disease. *NeuroMol. Med.* **2002**, *2*, 115–130. [CrossRef]
5. Cremades, N.; Chen, S.; Dobson, C. Structural Characteristics of α-Synuclein Oligomers. *Int. Rev. Cell. Mol. Biol.* **2017**, *329*, 79–143. [CrossRef]
6. Fusco, G.; De Simone, A.; Gopinath, T.; Vostrikov, V.; Vendruscolo, M.; Dobson, C.M.; Veglia, G.; Tata, G. Direct observation of the three regions in α-synuclein that determine its membrane-bound behaviour. *Nat. Commun.* **2014**, *5*, 3827. [CrossRef]
7. Burré, J.; Sharma, M.; Tsetsenis, T.; Buchman, V.; Etherton, M.R.; Südhof, T.C. Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* **2010**, *329*, 1663–1667. [CrossRef]
8. Nemani, V.M.; Lu, W.; Berge, V.; Nakamura, K.; Onoa, B.; Lee, M.K.; Chaudhry, F.A.; Nicoll, R.A.; Edwards, R.H. Increased expression of alpha-synuclein reduces neurotransmitter release by inhibiting synaptic vesicle reclustering after endocytosis. *Neuron* **2010**, *65*, 66–79. [CrossRef]
9. Tuttle, M.D.; Comellas, G.; Nieuwkoop, A.J.; Covell, D.J.; Berthold, D.A.; Kloepper, K.D.; Courtney, J.M.; Kim, J.K.; Barclay, A.M.; Kendall, A.; et al. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat. Struct. Mol. Boil.* **2016**, *23*, 409–415. [CrossRef]
10. Sgourakis, N.G.; Yau, W.-M.; Qiang, W. Modeling an In-Register, Parallel "Iowa" Aβ Fibril Structure Using Solid-State NMR Data from Labeled Samples with Rosetta. *Structure* **2015**, *23*, 216–227. [CrossRef]
11. Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y. Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Boil.* **2015**, *22*, 499–505. [CrossRef] [PubMed]
12. Schütz, A.K.; Vagt, T.; Huber, M.; Ovchinnikova, O.Y.; Cadalbert, R.; Wall, J.; Güntert, P.; Böckmann, A.; Glockshuber, R.; Meier, B.H. Atomic-Resolution Three-Dimensional Structure of Amyloid β Fibrils Bearing the Osaka Mutation**. *Angew. Chem. Int. Ed.* **2014**, *54*, 331–335. [CrossRef] [PubMed]
13. Colvin, M.T.; Silvers, R.; Ni, Q.Z.; Can, T.V.; Sergeyev, I.V.; Rosay, M.; Donovan, K.J.; Michael, B.; Wall, J.S.; Linse, S.; et al. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.* **2016**, *138*, 9663–9674. [CrossRef] [PubMed]

14. Fitzpatrick, A.W.P.; Falcon, B.; He, S.; Murzin, A.G.; Murshudov, G.; Garringer, H.J.; Crowther, R.A.; Ghetti, B.; Goedert, M.; Scheres, S.H.W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547*, 185–190. [CrossRef] [PubMed]
15. Banach, N.; Roterman, I. Amyloid as a ribbon-like micelle. In *From Globular Proteins to Amyloids*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 177–192.
16. Ulmer, T.S.; Bax, A.; Cole, N.B.; Nussbaum, R.L. Structure and Dynamics of Micelle-bound Human -Synuclein. *J. Boil. Chem.* **2004**, *280*, 9595–9603. [CrossRef]
17. Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526–W531. [CrossRef]
18. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [CrossRef]
19. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)-round x. *Proteins: Struct. Funct. Bioinform.* **2013**, *82*, 1–6. [CrossRef]
20. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss-function-Based model of hydrophobicity density in proteins. *Silico Boil.* **2006**, *6*, 15–22.
21. Gadzała, M.; Dułak, D.; Kalinowska, B.; Baster, Z.; Bryliński, M.; Konieczny, L.; Banach, M.; Roterman, I. The aqueous environment as an active participant in the protein folding process. *J. Mol. Graph. Model.* **2019**, *87*, 227–239. [CrossRef]
22. Kalinowska, B.; Banach, M.; Konieczny, L.; Roterman, I. Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins. *Entropy* **2015**, *17*, 1477–1507. [CrossRef]
23. Dułak, D.; Banach, M.; Gadzała, M.; Konieczny, L.; Roterman, I. Structural analysis of the Aβ(15-40) amyloid fibril based on hydrophobicity distribution. *Acta Biochim. Pol.* **2018**. [CrossRef] [PubMed]
24. Dułak, D.; Gadzała, M.; Banach, M.; Ptak, M.; Wiśniowski, Z.; Konieczny, L.; Roterman, I. Filamentous Aggregates of Tau Proteins Fulfil Standard Amyloid Criteria Provided by the Fuzzy Oil Drop (FOD) Model. *Int. J. Mol. Sci.* **2018**, *19*, 2910. [CrossRef] [PubMed]
25. Dułak, D.; Gadzała, M.; Stapor, K.; Fabian, P.; Konieczny, L.; Roterman, I. Holding with active paticipation of water. In *From Globular Proteins to Amyloids*; Elsevier: Amsterdam, The Netherlands, 2020.
26. Banach, M.; Konieczny, L.; Roterman, I. Fuzzy Oil Drop Model Application—From Globular Proteins to Amyloids. In *Springer Series on Bio- and Neurosystems*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2018; pp. 639–658.
27. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Boil.* **1976**, *104*, 59–107. [CrossRef]
28. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
29. Available online: https://zhanglab.ccmb.med.umich.edu/I-TASSER/ (accessed on 15 December 2018).
30. Available online: http://robetta.org/submit.jsp (accessed on 15 December 2018).
31. Banach, M.; Konieczny, L.; Roterman, I. The Amyloid as a Ribbon-Like Micelle in Contrast to Spherical Micelles Represented by Globular Proteins. *Molecules* **2019**, *24*, 4395. [CrossRef] [PubMed]
32. Konieczny, L.; Roterman, I. Protein is ane intelligent micelle. In *From Globular Proteins to Amyloids*; Roterman-Konieczna, I., Ed.; Elsevier: Amsterdam, The Netherlands; Oxford, UK; Cambridge, MA, USA, 2020; pp. 241–252.
33. Zibaee, S.; Jakes, R.; Serpell, L.C.; Davletov, B.; Crowther, R.A.; Goedert, M. Mutation E46K increases phospholipids binding and assembly into filaments of human alpha-synuclein. *FEBS Lett.* **2004**, *576*, 363–368.
34. Xu, L.; Bhattacharya, S.; Thompson, D. The fold preference and thermodynamic stability of-synuclein fibrils is encoded in the non-amyloid-component region. *Phys. Chem. Chem. Phys.* **2018**, *20*, 4502–4512. [CrossRef]
35. Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M.R.; Biernat, J.; Becker, S.; Mandelkow, E.; et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution from NMR and Small Angle Scattering. *Structure* **2014**, *22*, 238–249. [CrossRef]
36. Sałapa, K.; Kalinowska, B.; Jadczyk, T.; Roterman, I. Measurement of Hydrophobicity Distribution in Proteins (2012) Non-redundant Protein Data Bank. *BAMS* **2012**, *8*, 327–334.
37. Pech, U.; Verstreken, P. α-Synuclein and Tau: Mitochondrial Kill Switches. *Neuron* **2018**, *97*, 3–4. [CrossRef] [PubMed]

38. Terakawa, M.S.; Lee, Y.-H.; Kinoshita, M.; Lin, Y.; Sugiki, T.; Fukui, N.; Ikenoue, T.; Kawata, Y.; Goto, Y. Membrane-induced initial structure of α-synuclein control its amyloidogenesis on model membranes. *BBA-Biomembranes.* **2018**, *1860*, 757–766. [CrossRef] [PubMed]
39. Fusco, G.; Chen, S.W.; Williamson, P.T.F.; Cascella, R.; Perni, M.; Jarvis, J.A.; Cecchi, C.; Vendruscolo, M.; Chiti, F.; Cremades, N.; et al. Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science* **2017**, *358*, 1440–1443. [CrossRef] [PubMed]
40. Pupyshev, A.B.; Korolenko, T.A.; Akopyan, A.A.; Amstislavskaya, T.G.; Tikhonova, M.A. Suppression of autophagy in the brain of transgenic mice with overexpression of A53T-mutant α-synuclein as an early event at synucleinopathy progression. *Neurosci. Lett.* **2018**, *672*, 140–144. [CrossRef]
41. Flynn, J.D.; McGlinchey, R.P.; Walker, R.L.; Lee, J.C. Structural features of α-synuclein amyloid fibrils revealed by Raman spectroscopy. *J. Boil. Chem.* **2017**, *293*, 767–776. [CrossRef]
42. Tamgüney, G.; Korczyn, A.D. A critical review of the prion hypothesis of human synucleinopathies. *Cell Tissue Res.* **2017**, *373*, 213–220. [CrossRef]
43. Bengoa-Vergniory, N.; Roberts, R.F.; Wade-Martins, R.; Alegre-Abarrategui, J. Alpha-synuclein oligomers: A new hope. *Acta Neuropathol.* **2017**, *134*, 819–838. [CrossRef]
44. Jha, N.N.; Kumar, R.; Panigrahi, R.; Navalkar, A.; Ghosh, D.; Sahay, S.; Mondal, M.; Kumar, A.; Maji, S.K. Comparison of α-Synuclein Fibril Inhibition by Four Different Amyloid Inhibitors. *ACS Chem. Neurosci.* **2017**, *8*, 2722–2733. [CrossRef]
45. Avila, C.L.; Chaves, S.; Socias, S.B.; Vera-Pingitore, E.; González-Lizárraga, F.; Vera, C.; Ploper, D.; Chehín, R. Lessons learned from protein aggregation: Toward technological and biomedical applications. *Biophys. Rev.* **2017**, *9*, 501–515. [CrossRef]
46. Lek, S.; Vargas-Medrano, J.; Villanueva, E.; Marcus, B.S.; Godfrey, W.; Perez, R.G. Recombinant α- β- and γ-Synucleins Stimulate Protein Phosphatase 2A Catalytic Subunit Activity in Cell Free Assays. *J. Vis. Exp.* **2017**, *126*, e55361.
47. Ono, K. The Oligomer Hypothesis in α-Synucleinopathy. *Neurochem. Res.* **2017**, *42*, 3362–3371. [CrossRef] [PubMed]
48. Berman, H.M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
49. Laskowski, R.A.; Jabłońska, J.; Pravda, L.; Vařeková, R.S.; Thornton, J.M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **2017**, *27*, 129–134. [CrossRef] [PubMed]
50. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]
51. PLGrid User Manual. Available online: https://docs.cyfronet.pl/display/PLGDoc/User+manual (accessed on 5 July 2018).
52. Ptak-Kaczor, M.; Banach, M.; Konieczny, L.; Roterman, I. Internal force field in selected proteins. *Acta Biochim. Pol.* **2019**. [CrossRef]
53. Serpell, L.C. Alzheimer's amyloid fibrils: Structure and assembly. *BBA-Biomembranes* **2000**, *1502*, 16–30. [CrossRef]

Sample Availability: Samples of the compounds are not available from the authors.