

**Doktori (PhD) értekezés tézisei**

**AZ ANYATEJ MOLEKULÁRIS  
ÖSSZETÉTELÉNEK DINAMIKUS MODELLEZÉSE  
A KÖRNYEZETI TÉNYEZŐK FÜGGVÉNYÉBEN**

Készítette:

**Vámosiné Pacza Tünde**

doktorjelölt

Témavezető:

Dr. Baranyi József, PhD



DEBRECENI EGYETEM

TÁPLÁLKOZÁS - ÉS ÉLELMISZERTUDOMÁNYI DOKTORI ISKOLA

Debrecen, 2025



## 1. A DOKTORI ÉRTEKEZÉS ELŐZMÉNYEI ÉS CÉLKITŰZÉSEI

A táplálkozás egészségre gyakorolt hatását már több ezer éve felismerték, azonban az eddig létrehozott tápanyag adatbázisok nem tartalmazzák és elemzik az élelmiszerekben található molekulák jelentős részét, melyek közül pedig soknak jól dokumentált egészségügyi hatásai vannak (*Hooton és mtsai., 2020*).

Az egyik legfontosabb és legösszetettebb táplálék -mellyel először találkozunk emberi életünk során - az anyatej, mely nélkülözhetetlen az újszülött növekedéséhez és fejlődéséhez közvetlenül a születés után, és pótolhatatlan táplálékforrás a csecsemő túléléséhez (*Rossum és mtsai., 2005; Agostoni és mtsai., 2009*), és a táplálkozási és bioaktív komponensek komplex biológiai rendszere, ahol az összetevők állandó kölcsönhatásban vannak egymással. (*Christian és mtsai., 2021; Samuel és mtsai., 2020*). Az elmúlt évtizedekben széles körű kutatási erőfeszítések történtek az anyatej összetevőit és az azok jelenlétét, illetve az összetevők mennyiségét befolyásoló tényezők megismerésére (*Ballard & Morrow, 2013; Perrella és mtsai., 2021; Sánchez és mtsai., 2021; Carr és mtsai., 2021; De Weerth és mtsai., 2022; Samuel és mtsai., 2020*), azonban az adatok előállítása és gyűjtése, a statisztikai és modellezési módszerek, valamint az eredmények értelmezése ritkán összpontosít az „anya-tej-csecsemő” rendszer dinamikus jellegére.

Legjobb tudásunk szerint még nem készült számszerű, statisztikai értékelés az anyatej-összetétel heterogenitásának okairól és azok rangsorolásáról. Ennek egyik oka az anyatej komponensekre vonatkozó longitudinális adatok hiánya (*Christian és mtsai., 2021; Shenhav & Azad, 2022*), különösen az egyéni anyákra vonatkoztatva. Kutatásunk egyik célja az volt, hogy rávilágítson arra, hogyan lehet ezt a tudáshiányt betölteni, egy adatbázis ( MilkyBase adatbázis (*Pacza és mtsai., 2022*)) létrehozásával, mely ajánlott sablonként szolgál; egy adatbázis ontológiával és dinamikus modellezésre alkalmas adatokkal.

Jelen kutatás középpontjában az anyatej molekuláris szintű összetételének feltérképezése és időbeli változásainak matematikai modellezése állt. Ehhez elsődleges cél a tudományos kutatások publikált adataiból egy adatbázis létrehozása volt, majd a létrehozott adatbázis felhasználásával az anyatej összetétel időbeli változásának modellezése.

A disszertáció első része az anyatej biokémiai összetételét tartalmazó, MilkyBase nevű adatbázist írja le. Az adatokat tudományos publikációkból részben gépi tanulással, részben manuálisan válogattuk és digitalizáltuk. Elsődleges célunk egy olyan ontológia definiálása volt, amely annak feltárását segíti, hogy az anyatej összetétele hogyan függ különböző tényezőktől. Másodlagos cél, hogy egy ilyen ontológia használható legyen másfajta élelmiszerek kutatásában is, ahol a felhasználók saját adataikat is hasonló formátumban tárolhatják, illetve publikálják.

A disszertáció második része a létrehozott MilkyBase adatbázis adatainak felhasználásával bemutatja, hogy az anyatej kutatásban hogyan alkalmazhatóak az élelmiszer-mikrobiológiában már elterjedten használt prediktív modellezés módszerei.

A cél az volt, hogy a modell segítségével pontosabb képet alkothassunk az anyatej komponenseinek időbeli és egyéntől függő változásáról, az azt leginkább befolyásoló tényezőkről, ezzel is segítve a további anyatej összetevőket célzó kísérletek megtervezését. A módszer bemutatásán túl a dolgozat utolsó részében rámutatok a "prediktív anyatejkutatás" lehetséges korlátaira és felhasználási lehetőségeire is.

## 2. ANYAG ÉS MÓDSZER

A kutatás során alkalmazott módszereket 3 fő szakaszra lehet felosztani. Először az anyatej összetevők rögzítésére alkalmas adatbázis ontológiáját definiáltuk, és töltöttük fel adatokkal a publikációk alapján, majd az így kapott alap adatbázist bővítettük irányított kereséssel, végül a kapott adatokat elemeztük és modelleztük.

Mivel célunk egy olyan ontológia létrehozása volt, amely a táplálkozási szakemberek és kutatók számára, valamint az iparban és a szabályozásban is jól használható eszköz lehet, ezért az ontológia definiálásakor számos kompromisszumot kellett kötni, hogy megtaláljuk az egyensúlyt Big Data a négy fő alappillére - *méret, sebesség, változatosság* és *hitelesség* - között.

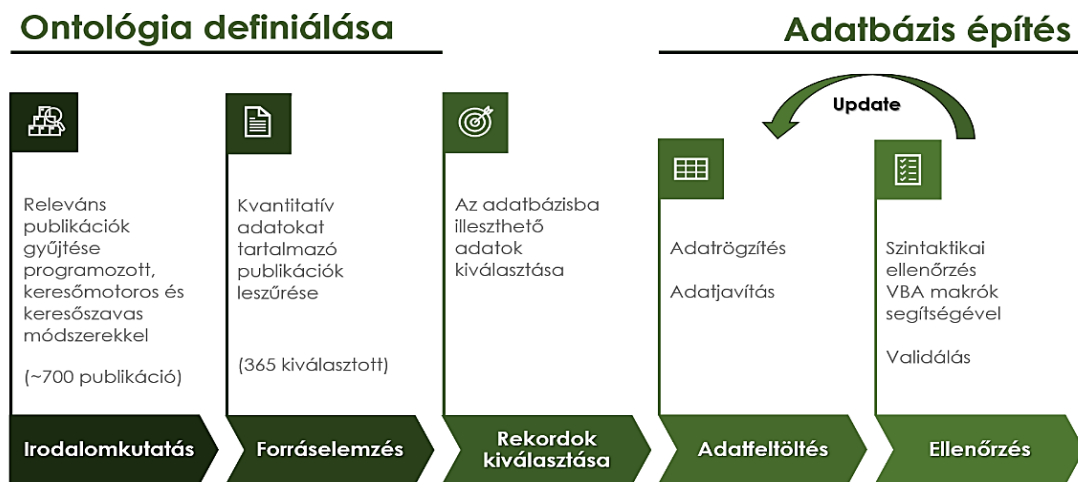
Ezen irányelvek alapján a választás az adatok adatbázisba rögzítéséhez a Microsoft Excel nagy teljesítményű táblázatkezelő programra esett. Ez a legelterjedtebb és legismertebb felhasználói program, amely képes táblázatok összekapcsolására, és emellett adatmegjelenítési és elemzési eszközöket is kínál. Mindemellett funkciói kibővíthetők a Visual Basic for Applications (VBA) objektum orientált programozási nyelv segítségével. Az általunk létrehozott VBA segédprogramok lehetővé teszik a rekordok rögzítése során a bemeneti ellenőrzést, valamint segítik az adatelemzést is (például a saját és mások hasonló publikált mérés eredményeinek összehasonlítását), ösztönzve ezzel az adatkurátorokat a releváns adataik adatbázisunkba való rögzítésére. Ez az úgynevezett wiki-filozófia realizációja, azaz a tudás hozzáadása a közöshöz, amely potenciálisan jóval nagyobb adatbázishoz vezethet.

### 2.1. A MILKYBASE ONTOLÓGIA LÉTREHOZÁSA

Adatbázis építési folyamatunk (1.ábra. ) során először az adatbázis „vázát” jelentő ontológiát definiáltuk, majd a felállított ontológia alapján létrehoztuk az tervezett adatbázist.

Az ontológia definiálásához elengedhetetlen volt az ontológiában rögzítendő adatok felmérése és a legfontosabb rögzítendő adatszoportok meghatározása. Ehhez a publikációk keresőmotoros és keresőszavas gyűjtése (**Irodalomkutatás**) után kapott tudományos cikkeket áttekintettük, értékeltük és kiválasztottuk az adatbázisba rögzíthetőket (**Forráselemzés**). A fő kiválasztási kritérium az volt, hogy az adatbázisba rögzítendő publikációnak mennyiségi adatokat kell tartalmaznia az anyatej összetevőiről;

lehetőleg úgynevezett dinamikus adatokat, amelyek az összetevők időbeli változását mutatják. Ezeket ideális esetben a publikációkban olyan táblázatokba rendezve találhattuk meg, amelyeknek adatait könnyebb volt beilleszteni az Excel táblázatunk lapjaira.



1.ábra. Az adatbázisépítés folyamata

Ezt a kiválasztott publikációk részletes elemzése követte, melynek célja az volt, hogy azonosítsuk a rögzítésre alkalmas adatkészleteket (**Rekordok kiválasztása**). Mivel az ontológia építés egyik fő célja a lehető legszélesebb körű információ összegyűjtése volt, nemcsak az anyatej összetételére vonatkozó adatokat választottuk ki rögzítésre, hanem az arra vonatkozó adatokat is, hogy milyen körülmények között generálódtak az anyatej összetevőkre mért adatok.

Az ontológia definiálását követően, az adatbázis adatokkal való feltöltése során (**Adatfeltöltés**) folyamatosan validáltuk a rögzítendő és már rögzített adatokat (**Ellenőrzés, hibajavítás**). Ehhez egyrészt a megfelelő szintaxis ellenőrzésére „ellenőrző makrókat” fejlesztettünk ki, másrészt szemantikai validációt alkalmaztunk a publikációkban található anomáliák azonosítására. A minőségellenőrzés (szemantika ellenőrzése) sokkal nehezebb feladat, mint a szintaktikai ellenőrzés, hiszen elvégzése emberi intelligenciát igényel. Az adatbázis szintaktikai ellenőrzését MS Excel VBA makrók segítik. A makrókat tartalmazó MBmacros.xlsm fájl bármely felhasználó számára elérhető a Figshare adattárban. Az adatok rögzítésekor „Syntax check” validáló makró

rendszeresen futtatva elérhető, hogy a rögzített adatok a mezők előírt formátumának megfelelően legyenek rögzítve, elősegítve ezzel az adatelemzést.

A szemantikai ellenőrzés során sokszor találtunk kétértelműséget vagy ellentmondást a szerzők által használt terminológiában is. Ezekben az esetekben legjobb tudásunk szerint definiáltuk és számszerűsítettük a publikált adatokat.

## 2.2. AZ ALAP MILKYBASE ADATBÁZIS KIBŐVÍTÉSE IRÁNYÍTOTT KERESÉSSEL

A kutatás második szakaszában az alap MilkyBase adatbázist a pontosabb adatkiértékelés érdekében további adatokkal egészítettük ki. Az adatbázis bővítésekor, az irodalomkutatás során az anyatej kutatások regionális helyzetére koncentráltunk, és két kiemelt régió adataival bővítettük, Ázsiával és Dél-Amerikával.

Az irodalomfeltárás során a SciELO – Scientific Electronic Library Online (SciELO) a Scopus, PubMed, Web of Science (WOS) online publikációs adatbázisokban végeztünk kereséseket.

Az adatbázisbővítésre irányuló irodalomkutatás során talált 190 publikációból végül 23-at találtunk alkalmasnak adatbázisunk bővítésére. Mivel nagy hangsúlyt fektettünk a dinamikus adatok felkutatására, az adatbázisbővítésünk során a dinamikus adataink száma közel másfél szeresére nőtt. (2. ábra)

MilkyBase version	Publikáció	Rekord (Master lap)			Összetevő	DynVal
		Összesen	Statikus	Dinamikus		
Core (1.0.)	140	840	205	635	752	7666
Adatbázis bővítés	23	54	11	43	71	2965
2.0.	163	894	216	678	823	10631

2. ábra. A MilkyBase adatbázisban található rekordok az adatbővítés után

## 2.3. AZ ANYATEJ ÖSSZETEVŐK IDŐBELI VÁLTOZÁSÁNAK ELEMZÉSE ÉS MODELLEZÉSE

Az ontológia létrehozásán túl, az adatbázisban összegyűjtött adatok alapján egy olyan modell elkészítése volt a cél, mely leírja az anyatej komponensek időbeli

változásának pályáját az emberi élet kritikus első négy hónapjában, amikor a csecsemők tápanyag igényeinek az ideális esetben elsősorban az anyatejből származik.

Elsődleges modellünk a következőképp írható le:

A fókuszingtervállum a szülést követő első négy hónap. Anyatej kutatásra specializálódott kutató kollégáinkkal való konzultációk során arra jutottunk, hogy az anyatej komponensek  $y(t)$  koncentrációjának időbeli változását két fázissal írjuk le;

1. először egy gyors kezdeti lineáris fázis a  $[0, \lambda]$  időintervallumban (kolosztrum),
2. majd egy exponenciálisan konvergens fázis, amelyet egy szaturációs modell ír le:

Ezzel a következő **kétfázisú modellhez** jutottunk:

$$(1.) \quad y(t) = \begin{cases} y_0 + a \cdot t & (0 \leq t < \lambda) \\ y_\lambda \cdot e^{-r \cdot (t-\lambda)} + y_{End}(1 - e^{-r \cdot (t-\lambda)}) & (\lambda \leq t) \end{cases}$$

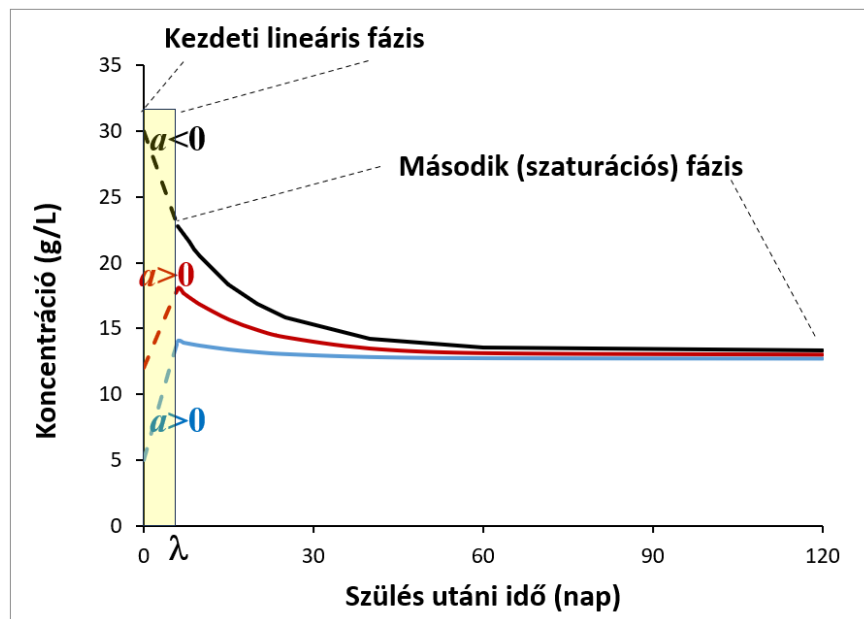
$$ahol \ y_\lambda = y_0 + a \cdot \lambda, \quad 0 \leq r, \quad 0 \leq \lambda.$$

Jelölések:

- $y(t)$  : egy anyatej komponens koncentrációja a szüléstől mért idő függvényében
- $t$  : a szülés után eltelt idő ( $t=0$  a szülés időpontja)
- $y_0$  : egy anyatej komponens koncentrációja a szülés időpontjában (kezdeti koncentráció)
- $a$  : az anyatej komponens koncentrációjának változási gyorsasága a kezdeti szakaszban
- $y_\lambda = y(\lambda)$  : az anyatej komponens koncentrációja a kezdeti fázis végén
- $r$  : szaturációs ráta
- $\lambda$  : a kezdeti (kolosztrum) fázis hossza
- $y_{End}$  : az anyatej komponens koncentrációja a vizsgált időszak végén (végső koncentráció)

A kezdeti fázis paraméterei  $y_0$  és  $a$ , csakúgy, mint a második fázis paraméterei  $r$ , és az  $y_{End}$  végső koncentrációs szint, számos elsősorban az anyát jellemző tényezőtől függ.

A második, stacionárius állapothoz konvergáló fázist **szaturációs modellnek** nevezzük.



**3. ábra.** Általános kétfázisú szaturációs modell. Szaggatott vonal: kezdeti fázis (kolosztrum). Folyamatos vonal: második (szaturációs) fázis.

A legalsó, kék görbe alig tér el az  $r=0$  esettől, amely egy olyan **kétfázisú függvényt** eredményezne, ahol a kezdeti, kolosztrum ideje alatti lineáris változást egy állandó ( $y_{End}=y_\lambda$ ) koncentráció követ:

$$(2.) \quad y(t) = \begin{cases} y_0 + a \cdot t & (0 \leq t < \lambda) \\ y_\lambda & (\lambda \leq t) \end{cases}$$

$$\text{ahol } y_\lambda = y_0 + a \cdot \lambda, \quad r \geq 0, \quad \lambda \geq 0.$$

A legfelső az  $\lambda=0$ ,  $a=0$  esethez hasonlít, amely az **egyfázisú egyszerű szaturációs** modellt képviseli, három paraméterrel.

$$(3.) \quad y(t) = y_0 \cdot e^{-r \cdot t} + y_{End}(1 - e^{-r \cdot t}) \quad (0 \leq t, r \geq 0)$$

Azt, hogy az adathalmaz leírásához szükséges-e a teljes kétfázisú modell négy paraméterrel ( $y_0$ ,  $a$ ,  $r$ ,  $y_{End}$ ), vagy bármelyik paraméter fix értéknek vehető a modell

dimenziójának csökkentése érdekében, F-próba döntötte el. Fontos megemlíteni, hogy  $\lambda$  értékészletét binárisnak tekintettük, azaz vagy 6 (alapértelmezett), vagy 0 értéket kapott. Az utóbbi esetben az első fázis beágyazódik a szaturációs modellbe, és az eredmény egy egyfázisú, egyszerű szaturációs modell lesz, három paraméterrel. Szintén F-próba döntötte el, hogy elegendő-e az egyfázisú egyszerű szaturációs modell az adott adatsorra való illeszkedéshez, vagy a  $\lambda = 6$  eset  $a$  meredekséggel a szignifikáns.

A másodlagos modellel számszerűsíthető a különböző tényezők (az anyai előzmények és egyéb jellemzők) hatása az elsődleges modell illesztett paramétereire (pl.  $y_{End}$ ). Hasonlóképpen, az F-próbát használtuk annak eldöntésére, hogy egy vagy két paraméter azonosnak tekinthető-e egy adatkészlet-pár vonatkozásában.

A számításokat egy Visual Basic nyelven írt Microsoft Excel Add-In-ben felépített nemlineáris regressziós algoritmussal végeztük, a standard Levenberg-Marquardt-módszer (*Levenberg, 1944*) (*Marquardt, 1963*) implementálásával. Az Excel integrált Data Analysis Add-In programját használtuk a lineáris regresszió és ANOVA számolások elvégzésére, 5%-os szignifikancia-szinttel.

### 3. EREDMÉNYEK

A kutatás eredményeként létrehoztuk a MilkyBase adatbázist, mely a tej összetételére vonatkozó rekordokat tárolja összekapcsolt Excel táblázatokban, majd ezeket az adatokat felhasználva matematikai modellezést végeztünk.

A létrehozott adatbázis egy olyan Excel munkafüzet, amely az adatok rögzítését a nem számítástudományos érdeklődésű táplálkozástudományi szakemberek számára is egyszerűvé teszi. A mezők hierarchikus szervezése biztosítja az adatok elemzésére szolgáló statisztikai és grafikai módszerek programozhatóságát. Ezt szolgálja az is, hogy az adatbázis és leírása, valamint a hozzá készített Excel makrók, mind szabadon és díjmentesen hozzáférhetőek és szükség szerint módosíthatók.

#### 3.1. MILKYBASE ADATBÁZIS ÚJDONSÁGAI

Az adatbázisunk fő újdonsága az ontológia, amely az anyatej összetételének mérési körülményeinek hatására, az anya/gyermek különböző jellemzőire, valamint a környezeti és előzményi körülményekre, adott válaszként tekinti az összetételre vonatkozó adatokat, valamint ezen adatok dinamikájára és bizonytalansági jellemzőire összpontosít, amelyek a magyarázó és válaszváltozóknak kerülnek beírásra.

Az újítások a következőképp foglalhatók össze:

##### 3.1.1. A MilkyBase ontológia definiálása a magyarázó változók (azaz a feltételek) és a válasz változók (azaz az összetevők) szem előtt tartásával.

Adatbázis-építési elvünk alapja, hogy egy-egy rekordot úgy tekintünk, mint a különböző magyarázó körülmények - amelyek mellett a megfigyelések történtek - leképezését az anyatej összetételére, mint válaszváltozóra. Azokat a változókat, melyek az adatok keletkezésének körülményeit írják le *Magyarázó változóknak*, míg az anyatej komponenseinek mért adatait a *Válaszváltozóknak* tekintjük. A magyarázó változók az anya, gyermek, illetve a születés részleteit (anya kora, diétája, gyerek súlya, terhesség hányadik hetében született, stb.) írják le, és az anyatej úgynevezett „történelmét” (a földrajzi helyzet, mérési és tárolási módszerek stb.), míg a válaszváltozók az anyatej összetevők (nutriensek, bioaktív komponensek stb.) mennyiségi adatai.

E válaszváltozók rögzített értékei lehetnek úgynevezett "kiterjesztett numerikus" formátumúak, illetve dinamikus, időbeli változást követők, ahol az összetevők időbeli

változását egy táblázatban tároljuk, és a táblázatra irányító mutató a változó beviteli értéke. A magyarázó változók numerikus értékei (feltétel mezők) szintén rögzíthetők a válaszváltozókhoz hasonló módon, de nem feltétlenül csak numerikus értékeket tartalmazhatnak. Lehetnek Boolean értékek vagy kategória listák is.

### **3.1.2. Dinamikus változók használata, melyekre " a bejegyzés egy táblázatra irányító mutató" technikát alkalmazzuk.**

Az időfüggő magyarázó és válaszváltozókat [idő, érték] adattáblák reprezentálják, míg az adatbázisban fő lapjának a vonatkozó bejegyzése csak egy mutató erre a táblára. Az ezen időfüggő adatokra illeszkedő pályák (trajektóriák) származtatott paraméterei, mint például a sebesség vagy az állandósult állapot szintje, a pályák lehetséges skaláris reprezentánsai. Vagyis egy változó időbeli profilját néhány kulcsparaméter (elsődleges modell) írja le, míg e paraméterek változása a választ befolyásoló körülmények (feltételek) függvényében a másodlagos modellekkel írható le. Ez az adatstruktúra tette lehetővé az anyatej összetevők prediktív modellezését is.

Az egyetlen rekordba rögzített többváltozós dinamikus válaszok megkönnyítik az időben változó adatok vizualizációját és ezzel azok összehasonlítását, elemzését, például az adatokban lévő minták és kiugró értékek felismerését; az adathiányok vagy az esetleges hibák azonosítását.

### **3.1.3. A kiterjesztett numerikus változók lehetőséget adnak a bizonytalanságok mérésére is**

A mérési bizonytalanságok adatbázisba való rögzítésének elősegítésére bevezettük az adatbázis "*numerikus mező*" fogalmának kibővített definícióját. Alapértelmezett formája egy közönséges valós szám, a leíró statisztikák által kalkulálható érték, rendelkezésre álló releváns adatok bizonyos középértéke, melyet ezen adatok szóródásának mértékével is el lehet látni. Ez általában vagy a szórás, vagy a minimum-maximum tartomány. Egy további, pontosvevesszővel elválasztott második részben rögzíthető a valós átlag előrejelzése (vagy becslése). Ez utóbbi becslült érték is megadható bizonytalansági számszerúsítéssel, ami általában vagy a becslés standard hibája, vagy annak 95%-os konfidenciaintervalluma

Ha a rögzített adatok intervallumokat képviselnek, akkor a későbbi adatelemzések során sztochasztikus intervallum elemzés használható, amely sokkal erőteljesebb, mint a

determinisztikus értékekkel végzett számítások, mivel nem csak a mennyiségi következtetéseket lehet levonni, hanem a következtetések megbízhatósága is számszerűsíthető.

#### **3.1.4. Adatstruktúra (fa struktúra)**

Adatbázisunkban Összetevő („*Component*”) alatt egy molekulát vagy molekulacsoportot értünk. Összetevő a **C20:4n-6** („Arachidic acid”) molekula, és a "zsírsav" is, azonban míg az első egy molekula összetevő, addig a második egy molekulacsoport, melynek eleme az említett **C20:4n-6** molekula is. Az ilyen típusú csoportosítás egy hierarchikus fa struktúrát követ. Így nem csak egy adott molekula koncentrációja, hanem bármely molekulacsoport az anyatej gyökér melletti szintről bármely összetevő koncentrációja rögzíthető.

A MilkyBase adatbázisban nemcsak az összetevők alkotnak hierarchikus faszerkezetet, hanem az egész adatbázisra jellemző ez a struktúra.

#### **3.1.5. Az adatok direkt és indirekt (származtatott) formában is rögzíthetők**

Az összetevők mért koncentrációin kívül számos szerző csak átváltott vagy származtatott értékeket publikál egy-egy komponens esetén, azaz vagy mért koncentrációval (g/L) vagy a komponenset tartalmazó csoporthoz viszonyított arányként adnak meg.

Az ilyen esetek egyértelmű kezelése érdekében a komponensre vonatkozó (mért adat) számértéket direkt (közvetlen), míg az arányra vonatkozó értéket indirekt (származtatott) válasznak nevezzük.

Adatbázisunk lehetőséget nyújt mindezen formátumok rögzítésére, tehát egy változó neve tartalmazhatja a ":" karaktert, hogy a lehető legközelebb álljon a biokémiai jelöléseikhez, valamint a "/" és "+" speciális karaktereket is, mint az indirekt változók jelzőkódjait.

A hierarchikus osztályozás és az adatok indirekt formában való rögzíthetősége lehetővé teszi, hogy nem csak egy adott molekula koncentrációja, hanem az molekulát tartalmazó csoportok is elemezhetőek.

### 3.2. AZ ANYATEJ MOLEKULÁRIS ÖSSZETÉTELÉBEN LÉVŐ MINTÁZATOK FELISMERÉSE (PREDIKTÍV ANYATEJ KOMPONENS MODELLEZÉS)

A "prediktív anyatej-modell készítés", azaz az anyatej molekuláris összetételében lévő mintázatok felismerése, érdekében két részre koncentráltunk:

1. Először egy olyan **elsődleges modell** létrehozása, amely képes leírni az anyatej komponenseinek időbeli pályáit,
2. majd annak bemutatása, hogy az **elsődleges modell paraméterei hogyan függenek olyan tényezőktől**, mint például a földrajzi elhelyezkedés.

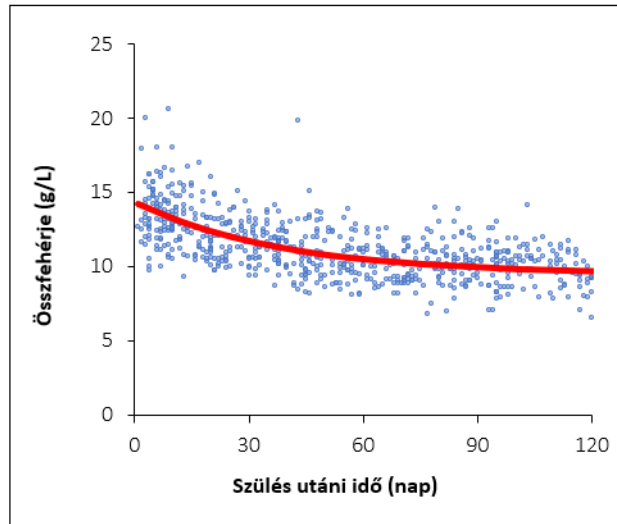
#### 3.2.1. Elsődleges modell elkészítése

Ahhoz, hogy az elsődleges modell felépítéséhez, paramétereinek meghatározásához ötleteket kaphassunk, egy-egy anya esetében mért részletes egyedi adatokra volt szükségünk az anyatej összetevők időbeli változásával kapcsolatosan. Ideális esetben több száz ilyen idő függvényében leírt pályát tudtunk volna összegyűjteni a publikált szakirodalomból, de sajnos azt tapasztaltuk, hogy az ilyen mélységű részletes adathalmaz nagyon ritka.

Egy ilyen induló elemzésre alkalmas adathalmazt John és munkatársainak publikációjában (*John és mtsai., 2019*) találtunk.

**Először az egyedi anyákra vonatkozó mérések** adatait vizsgáltuk meg, azokra az esetekre összpontosítva, ahol a lehető legtöbb mérés történt egy-egy anya esetében. Megfigyeléseink azt mutatták, hogy ezek az egyéni pályák, egyszerű ereszkedő görbék, azaz mindegyik lineárisan csökken a megfigyelési időintervallum alatt, és az egyéni pályák leírhatóak monoton konvergens csökkenő matematikai függvénnyel (szaturációs modell).

Ezt követően, **a publikációban található összes anyára megadott egyéni mérési adatot** elemeztük. A **4. ábra** az anyatejben mért teljes fehérjekoncentrációt mutatja a szüléstől eltelt idő függvényében, egyedi anyákra mérve, a szülést követő első 120 napban. Mivel a 177 anya vonatkozásában összesen 545 mérés történt, így az egy anya által adott átlagos mintaszám körülbelül 3 volt. Ez az adatszám nem elég az egyéni pályák azonosítása, azonban az átlagos koncentrációk az idő függvényében jól illeszthetők egy szaturációs modellel.



**4. ábra.** Az anyatej fehérje koncentrációjának időbeli változása a (John és mtsai., 2019) publikáció összes közölt mérési adata alapján.

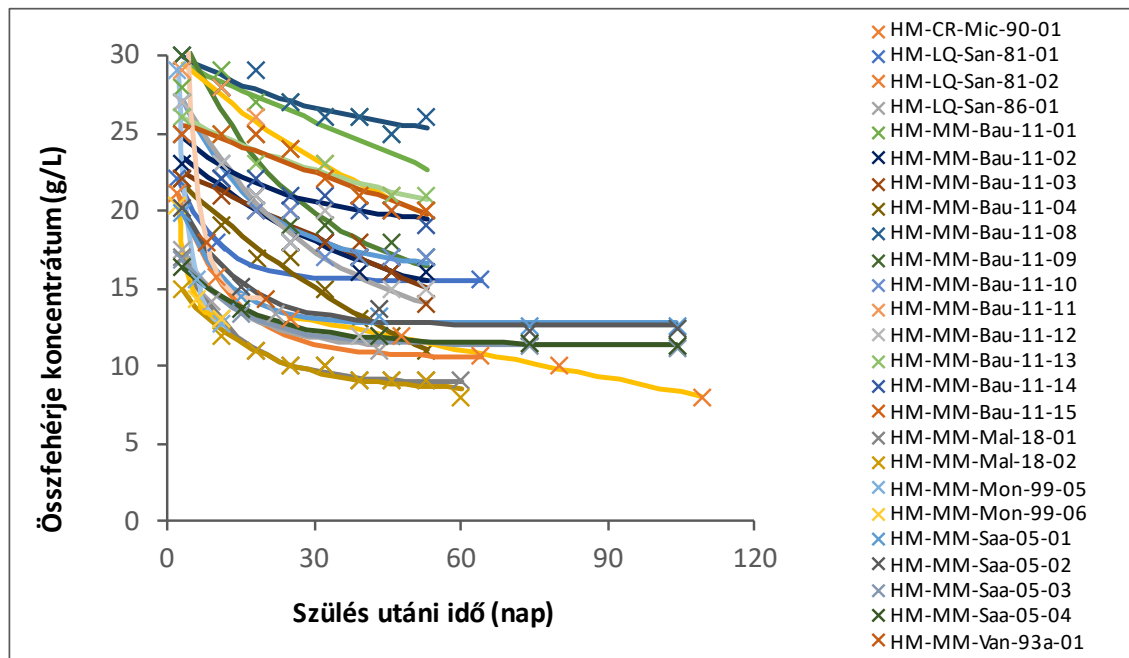
Mivel az exponenciális konvergencia sebessége itt egy populáción alapul, ezért ezt *populációs konvergencia sebességnek* nevezzük, szemben az egyes anyák *egyéni konvergencia sebességével*. Mivel a modell nem lineáris, ezért a populációs konvergencia sebesség nem az egyéni konvergencia sebességek számtani átlaga, hanem azok eloszlásának függvénye, és az egyszerű szaturációs függvény még mindig jól alkalmazható modell a tipikus (átlagos) viselkedésükre.

Az illesztett modell a következőképpen írható le:

$$y(t) = 14.245 \cdot e^{-0.031 \cdot t} + 9.691 \cdot (1 - e^{-0.031 \cdot t}) \quad (0 \leq t)$$

Ez, a mikrobiológiai „felezési idők” -höz hasonlóan, a következőképpen interpretálható: az exponenciális konvergencia sebessége 0,031/nap, ami körülbelül 22 napos "feleződési időnek" felel meg; azaz a végső stacionárius szintig hátralévő távolság nagyjából 3 hetente feleződik. A sebesség becslésének relatív hibája 15%, az illeszkedés standard hibája 1,5 g/L, míg a végső telítettségi szint  $9,691 \pm 0,19$  g/L-re becsült. Tekintettel arra, hogy a vizsgálat kohorszát a lehető leghomogénebbnek tekinthetjük, a kapott eredmények azt mutatják, hogy **nagyobb változatosság** mutatható ki a teljes fehérjekoncentrációban az **egyes anyák közötti biológiai különbségeknek köszönhetően** (úgynevezett keresztmetszeti eltérésnek), és jóval **kisebb mértékű az idő függvényében történő** (longitudinális) eltérés.

Hipotézisünk további vizsgálatához a MilkyBase egy olyan adathalmazát használtuk, melyben a kohorsz Európából származott. Az 5. ábra az anyatej teljes fehérje koncentrációjának pályáit mutatja a születéstől eltelt idő függvényében, 7 publikáció közölt adatai alapján. Az egy-egy pályát meghatározó adatpontok az adott kohorszra vonatkozó mérések adataiból származtatott átlagos fehérjekoncentrációk.



5. ábra. Az anyatej összfehérje koncentrációjának 24 trajektóriája Európai mintákból származó adatok alapján. ((Pacza és mtsai., 2022)-ben tárolt publikált adatok)

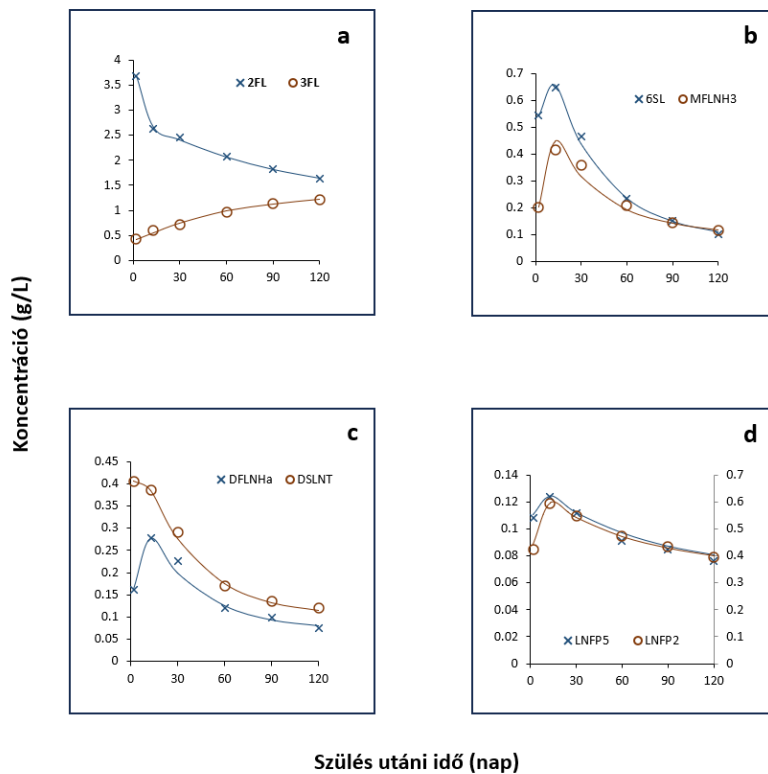
Az adatpontok átlagos koncentrációk, jellemzően több mint 10 anya adataiból. A teljes variáció főként olyan körülményeknek köszönhető, mint a terhességi kor, a szülés módja, a földrajzi elhelyezkedés stb. A trajektóriák illesztésére a háromparaméteres egyszerű szaturációs modellt használtuk.

A kapott értékek alapján megállapíthattuk, hogy két **véletlenszerűen kiválasztott anya** tejfehérje-koncentrációja közötti különbség **a szülés után azonos időpontban mérve** várhatóan 3-4-szer **nagyobb** lenne, mint **egy ugyanazon anyától származó két véletlenszerűen kiválasztott időpontban** vett minta – pl. egy kolosztrum és egy érett anyatej minta - fehérjekoncentrációja közötti különbség.

Az 4. ábra és az 5. ábra azt mutatta, hogy az egyszerű szaturációs modell az egyszerű szaturációs modell elég rugalmas ahhoz, hogy mind egyéni, mind populációs szinten illeszkedjen a fehérjepályákhoz, és az illeszkedése populációs

szinten meglehetősen erős (mindhárom paramétert 20%-nál kisebb relatív hibával becsültük).

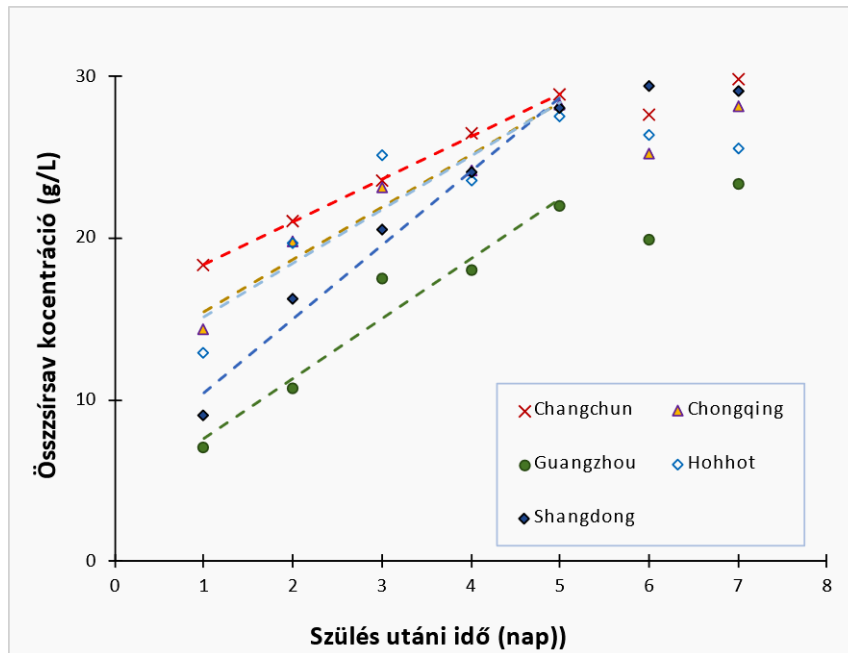
Hasonlóképpen megvizsgáltuk, hogy a molekulák **egy csoportjára**, például a fehérjékre, zsírsavakra, oligoszacharidokra, ásványi anyagokra vagy vitaminokra **azonosított mintázat ugyanúgy érvényes-e az adott csoport egyes molekuláira is**. Samuel és munkatársai (*Samuel és mtsai., 2022*) publikációjában található be ilyen specifikus molekulákra vonatkozó adatok, melyekre populációs pályákat illesztettünk (**6. ábra**) és láhattuk, hogy a pályák jó része az egyszerű szaturációs modellt követte, de néhány csak egy gyors kezdeti, lineáris szakasza után.



**6. ábra.** Anyatej oligoszacharidjainak mért koncentrációinak trajektóriái.

Az publikációkban nagyon kevés adat állt rendelkezésre annak bizonyítására, hogy a szaturációs fázist megelőző kezdeti lineáris időszak valóban általánosan létezik. A MilkyBase adatbázisban található egyetlen ilyen rekord Liu és munkatársaié (*Liu és mtsai., 2019*), mely meggyőzően bizonyította (lásd **7. ábra**), hogy az anyatej összzsírsav koncentrációja lineáris növekedett a kezdeti kolosztrum időszakban, mielőtt a pálya második egyszerű szaturációs fázisba lépett. Az illesztett meredekségek átlaga

3,5 g/L/nap volt a szülést követő első 6 napon, nagyjából. 12%-os relatív standard hibával. Ezek alapján azt mondhatjuk, hogy az anyatej zsírtartalma nagyjából megduplázódik az első 6 napban.



7. ábra. Az anyatej teljes zsírtartalmának változása (Liu és mtsai., 2019) alapján

Feltételezhetjük tehát, hogy egy általános modell esetében a szaturációs fázist először egy kezdeti lineáris „felszálló” időszak előzi meg. Az ilyen általános modellt **kétfázisú szaturációs modellnek** nevezzük, melynek 5 paramétere van, szemben az (kezdeti lineáris szakasz nélküli) **egyszerű szaturációs modellel**, amely csak 3 paraméterrel rendelkezik. Ha a két fázis azonos tendenciát mutat, nehéz szétválasztani őket, hiszen a rövid lineáris fázis jól beágyazható az egyszerű szaturációs modellbe. Azonban, ha a két fázis eltérő trenddel rendelkezik, nagy biztonsággal feltételezhetjük a kezdeti lineáris fázis létezését. Ennek fiziológiai magyarázata lehet, hogy az ilyen eltérő tendenciát mutató molekulák azok lehetnek, amelyek termelésében a csecsemő születése hirtelen változáshoz vezet.

A kétfázisú modellt a Samuel és munkatársai (Samuel és mtsai., 2022) adataira illesztettük (6. ábra). A modell illesztése során, először a kolosztrum időszak utáni adatokra egy egyfázisú, háromparaméteres szaturációs modellt illesztettünk nemlineáris regresszióval, majd a kolosztrumból származó egyetlen pontot kombináltuk az első kolosztrum utáni pont illesztett értékével.

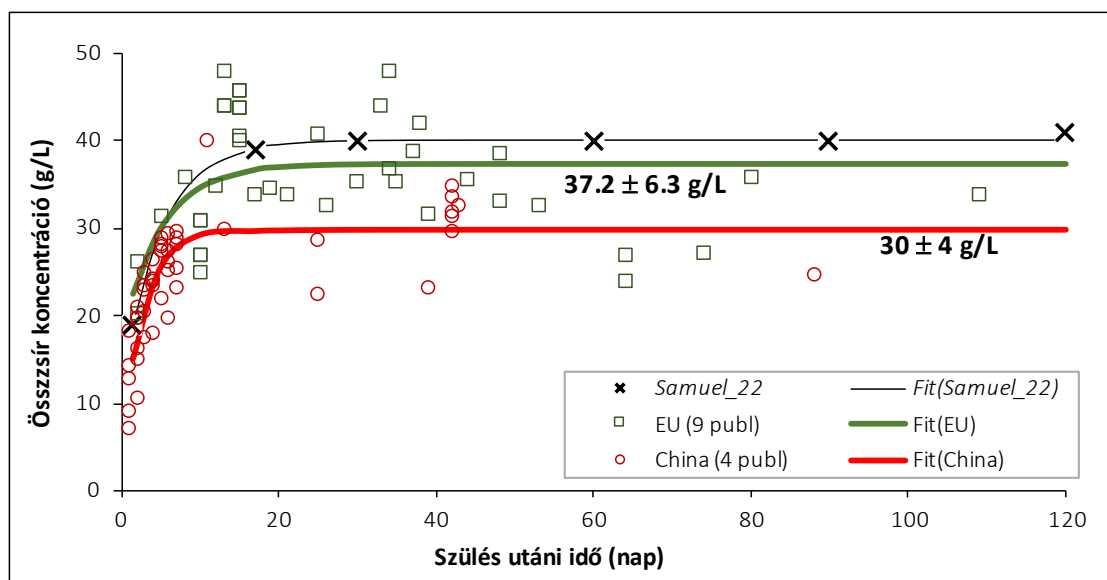
### 3.2.2. Másodlagos modell (Földrajzi különbségek)

Míg anyatej komponensek elsődleges modelljei a komponensek koncentrációjának az időbeli változását, azaz trajektóriáját írják le többé-kevésbé állandó körülmények között, addig a másodlagos modellek e körülményeknek az elsődleges modell paramétereire gyakorolt hatását jellemzik. A fő ok, amiért a két típusú modell nem összevonható, mechanisztikus: a körülmények azt befolyásolják, hogy egy komponens mennyisége hogyan változik az idővel (milyen gyorsan nő vagy csökken, azaz a változás sebességét), nem pedig közvetlenül magának a komponensnek a szintjét. A komplexitás redukálására az elsődleges modell paraméterkészlete használható, mellyel az egész időfüggő pálya helyettesíthető.

Az egyszerű szaturációs modellnek három paramétere van: a kezdeti és a végső koncentrációs szint, valamint a szinthez való exponenciális konvergencia sebessége. Az opcionális kezdeti kolosztrumbeli lineáris fázisnak két paramétere van: a vizsgált komponens koncentrációja a születés napján, és a koncentráció változás lineáris növekedésének (vagy csökkenésének) a sebessége.

Kutatásunk során azt vizsgáltuk, hogy az elsődleges modell harmadik paraméterét, a végső koncentrációs szintet (másodlagos paraméter) hogyan befolyásolja a földrajzi helyzet. Mivel a MilkyBase adatbázisban a földrajzi régió az egyik magyarázó változó, míg a különböző anyatej komponensek időbeli pályái dinamikus válaszváltozók, így könnyen vizsgálható volt a földrajzi helyzettől való függőség.

A Samuel és munkatársai (*Samuel és mtsai., 2022*) adataira illesztett két fázisú szaturációs görbéket (elsődleges modellek) vettük alapul az előrejelzéshez és az összehasonlításhoz. Ennek oka, hogy itt a mintákat nagyjából azonos időpontokban gyűjtötték az anyáktól, és a közzétett koncentrációk nagy minták átlagai, ezért standard hibáik kicsik. A MilkyBase adatai alapján létrehozott **8. ábra**, az anyatej - zsírkoncentrációi közötti különbséget mutatja az európai és kínai kohorszok esetében. Mindkét esetben a végső zsírkoncentráció nagyjából kétszerese a kiindulási értéknek, de az európai kohorszra jellemző (átlagos) zsírkoncentráció (ábrán zöld vonallal jelzett) a megfigyelési idő alatt végig 25-30%-kal magasabb volt, mint a kínai kohorszé (az ábrán piros vonallal jelzett).

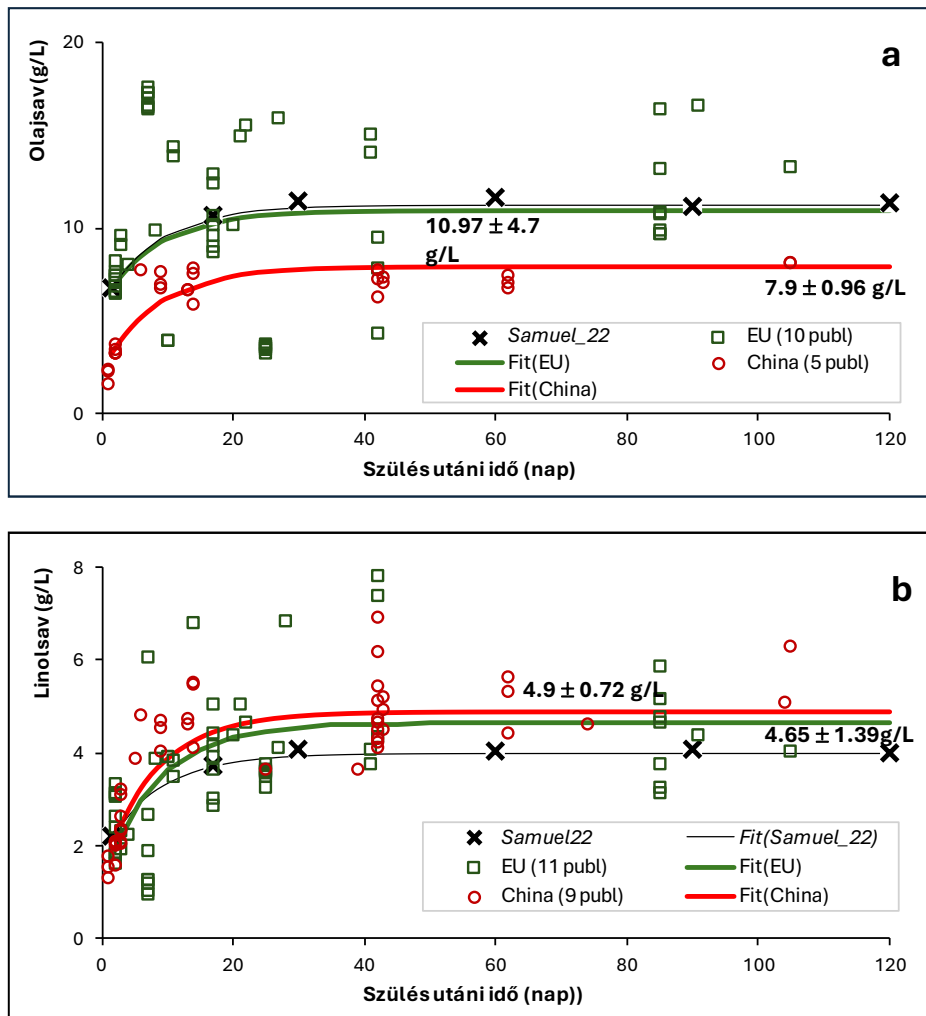


**8. ábra.** Az anyatej teljes zsírkoncentrációjának változása európai és kínai kohorszokban.

Az alapmodell (fekete folytonos vonal) olyan átlagos zsírkoncentrációkra illeszkedik, melyhez egy-egy időpontban több száz anya szolgáltatott mintát. Ezek az átlagos koncentrációk észrevehetően sima mintázatot követnek, még akkor is, ha az egyes egyénekenkénti pályák zajosabbak is. Az ANOVA teszt megerősítette, ami a grafikonon is jól ábrázolódik, hogy ez az egyedi anyák által okozott eltérés szignifikánsan nagyobb, mint a földrajzi elhelyezkedés (a kohorsz az EU-ból vagy Kínából származik-e) okozta eltérés.

Végül felmerül az a kérdés is, hogy ez **az összzsírsav koncentrációban látható különbség az egyes zsírsavmolekulákra is érvényes-e**. Erre a MilkyBase adatbázis molekulaszpecifikus adatainak felhasználásával kaphatunk választ.

A legnagyobb mennyiségben előforduló zsírsavmolekulák az olajsav (C18:1n-9) és a linolsav (C18:2n-6), ezek összege az anyatej teljes zsírsavtartalmának több mint felét adja. A **9. ábra** az olajsavra és a linolsavra vonatkozóan mutatja a MilkyBase rendelkezésre álló adatait, amin látható, hogy a kínai kohorszokból származó koncentrációkban általában alacsonyabb az olajsav szintje, mint az európai kohorszokban, azonban a linolsav esetében fordított a helyzet, különösen az első két hétben.



**9.ábra.** Az anyatej olajsav (C18:1 n-9) és a linolsav (C18:2 n-6) koncentrációjának a szülést követő 120 napban.

A MilkyBase-ben az eikozadiénsavra (C20:2n-6) és a dokozahexaénsavra (DHA; C22:6n-3) vonatkozóan rendelkezésre álló adatok is hasonló tendenciát mutatnak.

A földrajzi elhelyezkedés okozta különbség a teljes zsírsavkoncentrációk között tehát nem arányosan oszlik meg a különböző zsírsavmolekulák között.

## 4. ÚJ TUDOMÁNYOS EREDMÉNYEK

### A MilkyBase adatbázis és újdonságai

1. A kutatás eredményeként – az anyatej kutatások területén először- **létrehoztunk egy adatbázist -a MilkyBase adatbázis-** ,az anyatej összetevőiről tudományosan publikált adatok alapján, mely **az anyatej összetételére vonatkozó** rekordokat tárolja és rendszerezi. Az adatbázis létrehozásakor a következő innovációkat vezettük be:
  - a. **Megalkottuk az ontológiát**, amely **az anyatej összetételére ható tényezőkre adott válaszként** tekinti az összetételre vonatkozó adatokat, amelyek a **magyarázó és válaszváltozóknban** kerülnek beírásra. A rekordok „ok-okozati” hatást reprezentálnak, azaz különböző magyarázó körülmények - amelyek mellett a megfigyelések történtek - leképezését az anyatej összetételére, mint válaszváltozóra.
  - b. Az adatbázis szerkezetét úgy alakítottuk ki, hogy alkalmas legyen az anyatej komponensek és összetétel időbeli változásának a követésére, azaz nem csak statikus, hanem **dinamikus (időfüggő) állapotok is rögzíthetők legyenek**. Ehhez a változók időbeli változását egy dedikált táblával reprezentáljuk, amelyek a pályák („időpont, mért érték”) értékpárjait tartalmazza és az ezekre hivatkozó mutatók biztosítják, hogy az időfüggőség az adott mezők természetes attribútuma legyen.
  - c. Bevezettük a **kiterjesztett numerikus változókat**, melyek lehetőséget adnak a bizonytalanságok feljegyzésére és mérésére, valamint az adatok **direkt és indirekt (származtatott) formában** való rögzíthetőségét, elősegítve ezzel a szélesebb körű adatgyűjtést és elemzést.
  - d. Az adatbázis szerkezetét úgy definiáltuk, hogy az elemek csoportosítása **hierarchikus fa struktúrát** követ. A létrehozott struktúra lehetővé teszi az anyatej összetevőit tartalmazó gráf létrehozását és támogatja többszintű elemzéseket.

### Az anyatej molekuláris összetételében lévő mintázatok felismerése

Az anyatejkutatások során először alkalmaztunk matematikai modellezést az anyatej összetevők jellemzésére, melynek során következő eredményeket kaptuk.

2. Az elsődleges modellezés során létrehoztunk egy **egyfázisú egyszerű szaturációs modellt**

$$y(t) = y_0 \cdot e^{-r \cdot t} + y_{End}(1 - e^{-r \cdot t}) \quad (0 \leq t, r \geq 0)$$

A modell három paraméterrel rendelkezik, az  $y_0$  kezdeti koncentráció és az  $y_{End}$  végső koncentráció, valamint a végső koncentrációhoz való exponenciális konvergenciának a sebessége ( $r$ ).

3. A MilkyBase adatbázisban rögzített adatok alapján az **anyatej teljes fehérje koncentrációját** vizsgálva az egyfázisú egyszerű szaturációs modell segítségével következő eredményekre jutottunk:

- a. Megállapítottuk, hogy a létrehozott egyszerű szaturációs modell illeszkedése populációs szinten meglehetősen **erős** (mindhárom paramétert 20%-nál kisebb relatív hibával becsültük), valamint elég **rugalmas** ahhoz, hogy mind egyéni, mind populációs szinten illeszkedjen a **fehérjepályákhoz**.
- b. Megállapítottuk, hogy **nagyobb változatosság** mutatható ki a teljes **fehérjekoncentrációban az egyes anyák közötti biológiai különbségeknek köszönhetően** (úgynevezett keresztmetszeti eltérésnek- szórás kb.5 g/L ), és jóval **kisebb** mértékű az **idő függvényében** történő (longitudinális eltérés- szórás kb. 1-2 g/L) eltérés.

4. Az anyatej összetevők időbeli pályáinak leírására megalkottunk egy **általános kétfázisú szaturációs modellt** (melynek speciális esete az egyfázisú egyszerű szaturációs modell), ahol a szaturációs fázist először egy kezdeti lineáris időszak előzi meg

$$y(t) = \begin{cases} y_0 + a \cdot t & (0 \leq t < \lambda) \\ y_\lambda \cdot e^{-r \cdot (t-\lambda)} + y_{End}(1 - e^{-r \cdot (t-\lambda)}) & (\lambda \leq t) \end{cases}$$

$$\text{ahol } y_\lambda = y_0 + a \cdot \lambda, \quad 0 \leq r, \quad 0 \leq \lambda.$$

A modellnek 5 paramétere van, a kezdeti koncentráció ( $y_0$ ) és végső koncentráció ( $y_{End}$ ), az anyatej komponens koncentrációjának változási gyorsasága a kezdeti szakaszban ( $a$ ), a szaturációs ráta ( $r$ ) és a kezdeti (kolosztrum) fázis időtartama ( $\lambda$ ).

Az opcionális kezdeti – kolosztrumbeli - lineáris fázisnak két paramétere van: a vizsgált komponens koncentrációja a születés napján, és a koncentráció változás sebessége. Ezek a fázisok igazolják a hagyományos anyatej időbeli változás felosztást (kolosztrum – átmeneti – érett).

5. A kétfázisú modellt a MilkyBase adatbázisban rögzített **egyedi molekulák** adataira illetve a következő eredményeket kaptuk:
  - a. A kétfázisú modell jó illeszkedést mutatott a vizsgált zsírsav molekulák, oligoszacharidok valamint ásványi anyagok esetében is.
  - b. Azt tapasztaltuk, hogy a C18:0, C18:1n9, C18:3n3 zsírsav molekulák trajektóriái a végső koncentrációs szintig gyorsan növekednek, a C20:2n6, C22:1n9 zsírsav molekulák, valamint a cink és szelénium mindvégig csökkenő tendenciát mutatnak. Továbbá a C22:6n3, C16:1n7 és C20:0 zsírsav molekulák, valamint a foszfor koncentrációja az anyatejben először lineáris növekszik, majd egy exponenciálisan csökkenő pályát követve konvergál a végső szinthez.
6. Másodlagos modellezés segítségével az anyatej összzsírsav tartalmának végső koncentráció szintjére gyakorolt földrajzi hatás vizsgálatokor a következő eredményeket kaptuk:
  - a. Európa és Kína tekintetében az anyatej **teljes zsírtartalma esetén is** az egyes anyák közötti **sztochasztikus biológiai különbségből származó keresztmetszeti eltérés** (kb. 15 g/L), **nagyobb, mint** a **hosszanti** avagy a földrajzi eredetű eltérés (az európai kohorszok esetében 6,3 g/L, míg a kínai kohorszok esetében ez 4 g/L).
  - b. Megállapítottuk, hogy a teljes zsírsavkoncentrációk közötti különbség, melyet a földrajzi elhelyezkedés okoz, a nem arányosan oszlik meg a különböző zsírsavmolekulák között.

## 5. AZ EREDMÉNYEK GYAKORLATI HASZNOSÍTHATÓSÁGA

Elsődleges célunk egy olyan ontológia definiálása volt, amely annak feltárását segíti, hogy az anyatej összetétele hogyan függ különböző tényezőktől. Másodlagos cél, hogy egy ilyen ontológia használható legyen másfajta élelmiszerek kutatásában is, ahol a felhasználók saját adataikat is hasonló formátumban tárolhatják.

A létrehozott ontológia és az alkalmazott modellezési módszerek gyakorlatban is alkalmazható eredményei a következők:

1. A kutatók és a táplálkozási szakemberek a MilkyBase segítségével különböző tényezők alapján azonosíthatják a tej összetételének mintázatait. Ezek a minták segíthetnek annak megértését, hogy az anyai körülmények, a szülés körülményei vagy egyéb környezeti hatások hogyan befolyásolhatják az anyatej összetételét.
2. Az adatbázis segíthet feltárni a „kevésbé kutatott” területeket, irányt mutatva a további kutatási projekteknek.
3. A MilkyBase olyan platformként szolgál, ahol a felhasználók szabványosított formátumban bevihetik saját adataikat, elősegítve ezzel az élelmiszer-összetételre vonatkozó adatok **formátumainak a szabványosítását**, az adatbázisok kompatibilissé tételét és teljes potenciáljának kihasználását.
4. A MilkyBase mezőinek hierarchikus szervezése megkönnyíti az adatkészleteken belüli és azok közötti összehasonlítást, a statisztikai módszerek alkalmazását az adatok elemzésére, lehetőséget nyújtva az adathalmazon belül összefüggések és korrelációk feltárására. A könnyű kezelhetőség a kutatókat arra ösztönzi, hogy az általuk létrehozott adatokat ebben a formában rögzítsék, **elősegítve az ellenőrzést, együttműködést és adatmegosztást**.
5. Az anyatej összetevők matematikai modellezésének bevezetése megteremti a lehetőségét a bennük rejlő mintázatok objektív alapon való felismerésére. Az előrejelzések, nagy előnyt jelenthetnek a kísérleti tervezésre és az adatok értelmezésére, valamint a kutatási és innovációs területek kiválasztására. A terület a gépi tanulási technikák integrálásával és az összetett és nagy adathalmazok kezelésével fejlődhet tovább. A jövő kihívásai közé tartozik a pontosabb és hatékonyabb számítási modellek kifejlesztése, amelyek képesek kezelni a tudományos problémák növekvő összetettségét és méretét. A mesterséges intelligencia térhódításával lehetőséget nyújt arra, hogy ezek a programok könnyebben hasznosíthatóak a méréseket, a felhalmozott közös tudást, segítve ezzel a döntéshozatalt is.
6. Továbbá - modellünk eredményei alapján - javasoljuk, hogy az anyatej-kutatások során a kutatási tervek nem-ekvidisztáns mintavételi időpontokat tartalmazzanak, a

szülést követő első két hétben gyakoribb mintavételezéssel. Ez a kutatói megközelítés hatékonyabban tudná jellemezni a kolosztrum időszak gyors változásait és pontosabb képet kaphatnánk az anyatej időbeli dinamikájáról.

## 6. IRODALOMJEGYZÉK (FELHASZNÁLT IRODALOM)

1. Agostoni, C. - Braegger, C. - Decsi, T. - Kolacek, S. - Koletzko, B. - Michaelsen, K. F. - Mihatsch, W. - Moreno, L. A. - Puntis, J. - Shamir, R. - Szajewska, H. - Turck, D. - van Goudoever, J. and Nutrition, E. C. o.: 2009. Breast-feeding: A Commentary by the ESPGHAN Committee on Nutrition. *Journal of Pediatric Gastroenterology and Nutrition*. 49. (1). 112-125. 10.1097/MPG.0b013e31819f1e05
2. Ballard, O. and Morrow, A. L.: 2013. Human Milk Composition. *Pediatric Clinics of North America*. 60. (1). 49-74. 10.1016/j.pcl.2012.10.002
3. Carr, L. E. - Virmani, M. D. - Rosa, F. - Munblit, D. - Matazel, K. S. - Elolimy, A. A. and Yeruva, L.: 2021. Role of Human Milk Bioactives on Infants' Gut and Immune Health. *Front Immunol*. 12. 604080. <https://doi.org/10.3389/fimmu.2021.604080>
4. Christian, P. - Smith, E. R. - Lee, S. E. - Vargas, A. J. - Bremer, A. A. and Raiten, D. J.: 2021. The need to study human milk as a biological system. *The American Journal of Clinical Nutrition*. 113. (5). 1063-1072. <https://doi.org/10.1093/ajcn/nqab075>
5. De Weerth, C. - Aatsinki, A.-K. - Azad, M. B. - Bartol, F. F. - Bode, L. - Collado, M. C. - Dettmer, A. M. - Field, C. J. - Guilfoyle, M. - Hinde, K. - Korosi, A. - Lustermans, H. - Mohd Shukri, N. H. - Moore, S. E. - Pundir, S. - Rodriguez, J. M. - Slupsky, C. M. - Turner, S. - Van Goudoever, J. B. - Ziomkiewicz, A. and Beijers, R.: 2022. Human milk: From complex tailored nutrition to bioactive impact on child cognition and behavior. *Critical Reviews in Food Science and Nutrition*. 63. (26). 1-38. 10.1080/10408398.2022.2053058
6. Hooton, F. - Menichetti, G. and Barabási, A.-L.: 2020. Exploring food contents in scientific literature with FoodMine. *Scientific Reports*. 10. (1). 16191. 10.1038/s41598-020-73105-0
7. John, A. - Sun, R. - Maillart, L. - Schaefer, A. - Hamilton Spence, E. and Perrin, M. T.: 2019. Macronutrient variability in human milk from donors to a milk bank: Implications for feeding preterm infants. *PLOS ONE*. 14. (1). e0210610. 10.1371/journal.pone.0210610
8. Levenberg, K. J. Q. o. A. M.: 1944. A METHOD FOR THE SOLUTION OF CERTAIN NON – LINEAR PROBLEMS IN LEAST SQUARES. 2. 164-168.
9. Liu, Y. - Liu, X. and Wang, L.: 2019. The investigation of fatty acid composition of breast milk and its relationship with dietary fatty acid intake in 5 regions of China. *Medicine*. 98. (24). e15855. 10.1097/MD.00000000000015855

10. *Marquardt, D. W.*: 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. 11. (2). 431-441. 10.1137/0111030
11. *Pacza, T. - Martins, M. L. - Rockaya, M. - Müller, K. - Chatterjee, A. - Barabási, A.-L. and Baranyi, J.*: 2022. MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions. <https://figshare.com/s/c44b92932fc1a5785cd3>
12. *Perrella, S. - Gridneva, Z. - Lai, C. T. - Stinson, L. - George, A. - Bilston-John, S. and Geddes, D.*: 2021. Human milk composition promotes optimal infant growth, development and health. *Seminars in Perinatology*. 45. (2). 151380. <https://doi.org/10.1016/j.semperi.2020.151380>
13. *Rossum, C. - Büchner, F. and Hoekstra, J.*: 2005. Quantification of health effects of breastfeeding - Review of the literature and model simulation. *Annals of Nutrition and Metabolism*. 51.
14. *Samuel, T. M. - Zhou, Q. - Giuffrida, F. - Munblit, D. - Verhasselt, V. and Thakkar, S. K.*: 2020. Nutritional and Non-nutritional Composition of Human Milk Is Modulated by Maternal, Infant, and Methodological Factors. *Frontiers in Nutrition*. 7. 576133. <https://doi.org/10.3389/fnut.2020.576133>
15. *Samuel, T. M. - Thielecke, F. - Lavalley, L. - Chen, C. - Fogel, P. - Giuffrida, F. - Dubascoux, S. - Martinez-Costa, C. - Haaland, K. - Marchini, G. - Agosti, M. - Rakza, T. - Costeira, M. J. - Picaud, J. C. - Billeaud, C. and Thakkar, S. K.*: 2022. Mode of Neonatal Delivery Influences the Nutrient Composition of Human Milk: Results From a Multicenter European Cohort of Lactating Women. *Front Nutr*. 9. 834394. 10.3389/fnut.2022.834394
16. *Sánchez, C. - Franco, L. - Regal, P. - Lamas, A. - Cepeda, A. and Fente, C.*: 2021. Breast Milk: A Source of Functional Compounds with Potential Application in Nutrition and Therapy. *Nutrients*. 13. (3). 1026. 10.3390/nu13031026
17. *Shenhav, L. and Azad, M. B.*: 2022. Using Community Ecology Theory and Computational Microbiome Methods To Study Human Milk as a Biological System. *mSystems*. 7. (1). e01132-01121. 10.1128/msystems.01132-21

## 7. PUBLIKÁCIÓK AZ ÉRTEKEZÉS TÉMAKÖRÉBEN



**DEBRECENI  
EGYETEM**

**DEBRECENI EGYETEM  
EGYETEMI ÉS NEMZETI KÖNYVTÁR**

H-4002 Debrecen, Egyetem tér 1, Pf.: 400  
Tel.: 52/410-443, e-mail: publikaciok@lib.unideb.hu

Nyilvántartási szám: DEENK/567/2024.PL  
Tárgy: PhD Publikációs Lista

Jelölt: Pacza Tünde

Doktori Iskola: Táplálkozás- és Élelmiszertudományi Doktori Iskola. Élelmiszertudományi doktori program

MTMT azonosító: 10084787

### A PhD értekezés alapjául szolgáló közlemények

#### Idegen nyelvű tudományos közlemények külföldi folyóiratban (4)

1. Baranyi, J., Csorba, S., Farkas, Z., **Pacza, T.**, Józwiak, Á.: Internal dynamics of patent reference networks using the Bray-Curtis dissimilarity measure.  
*J Big Data*. 11 (1), 1-10, 2024. EISSN: 2196-1115.  
DOI: <http://dx.doi.org/10.1186/s40537-024-00883-z>  
IF: 8.6 (2023)
2. Baranyi, J.\*, **Pacza, T.\***, Martins, M. L., Thakkar, S. K., Samuel, T. M.: Modelling the temporal trajectories of human milk components.  
*BMC Pregnancy Childbirth*. 24 (1), 1-13, 2024. EISSN: 1471-2393.  
DOI: <http://dx.doi.org/10.1186/s12884-024-06896-z>  
\*Megosztott első szerzős közlemény.  
IF: 2.8 (2023)
3. Martins, M. L., **Pacza, T.**, Müller, K. E., Baranyi, J.: A computational approach to nutrition science reveals the dynamics of the protein content of human milk.  
*Innovative Food Science & Emerging Technologies*. 82, 1-5, 2022. ISSN: 1466-8564.  
DOI: <http://dx.doi.org/10.1016/j.ifset.2022.103167>  
IF: 6.6
4. **Pacza, T.**, Martins, M. L., Rockaya, M., Müller, K. E., Chatterjee, A., Barabási, A. L., Baranyi, J.: MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions.  
*Sci Data*. 9 (1), 1-7, 2022. EISSN: 2052-4463.  
DOI: <http://dx.doi.org/10.1038/s41597-022-01663-1>  
IF: 9.8





Idegen nyelvű absztrakt kiadványok (1)

5. **Pacza, T.**, Martins, M. L., Müller, K. E., Baranyi, J.: MilkyBase- A Database for Molecular-Level Mapping of the Composition of the Human Milk.  
In: International Milk Genomic Consortium : IMGC HYBRID Symposium 2023, IMGC, Cork, 1, 2023.

**További közlemények**

Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

6. Mposula, Z., **Pacza, T.**, Szepesi, J., Máthé, E.: Lifestyle and socio-economic inequalities in diabetes prevalence in Madadeni Township, South Africa.  
*Acta Med. Sociol.* 14 (37), 5-21, 2023. ISSN: 2062-0284.  
DOI: <http://dx.doi.org/10.19055/ams.2023.12/15/1>

**A közlő folyóiratok összesített impakt faktora: 27,8**

**A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre): 27,8**

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudományometriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2024.11.14.

