

C2115

Praktický úvod do superpočítání

II. lekce

Petr Kulhánek

kulhanek@chemi.muni.cz

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta
Masarykova univerzita, Kamenice 5, CZ-62500 Brno

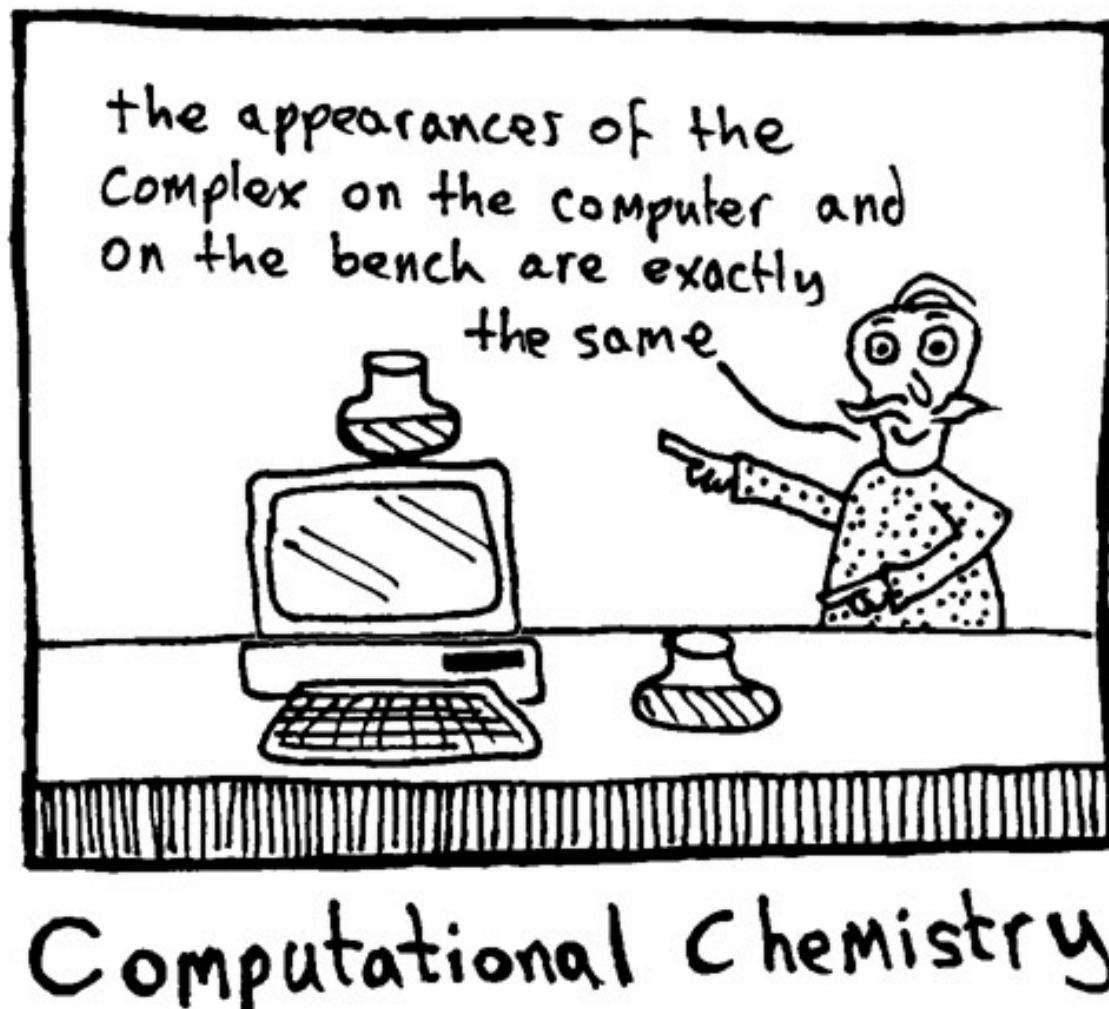
- **Skupina výpočetní chemie**
přehled řešených projektů
- **Modelové problémy a systémy do cvičení**
násobení matic, numerická integrace, QM a MD výpočty

Skupina výpočetní chemie

přehled řešených projektů

vedoucí skupiny: prof. RNDr. Jaroslav Koča, DrSc.

Výpočetní chemie



<http://www.ninger.com/images/comp.jpg>

Výpočetní chemie

Výpočetní chemie (Computational Chemistry, počítačová chemie)

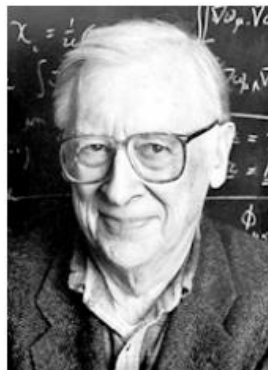
je odvětví chemie, které využívá **počítačů** při řešení **chemických problémů**. Používá výsledků **teoretické chemie** implementované do výkonných **počítačových programů** určených k výpočtům struktury, vlastností a reaktivity molekul a pevných látek. I když její výsledky doplňují informace získané chemickými experimenty, v určitých případech může předpovědět doposud nepozorované chemické jevy. Výpočetní chemie je široce používaná v **návrhu nových léčiv a materiálů**.

www.wikipedia.org

Nobelova cena za chemii 1998/2013



Walter Kohn



John A. Pople



© Harvard University
Martin Karplus



Photo: © S. Fisch
Michael Levitt



Photo: Wikimedia
Commons
Arieh Warshel

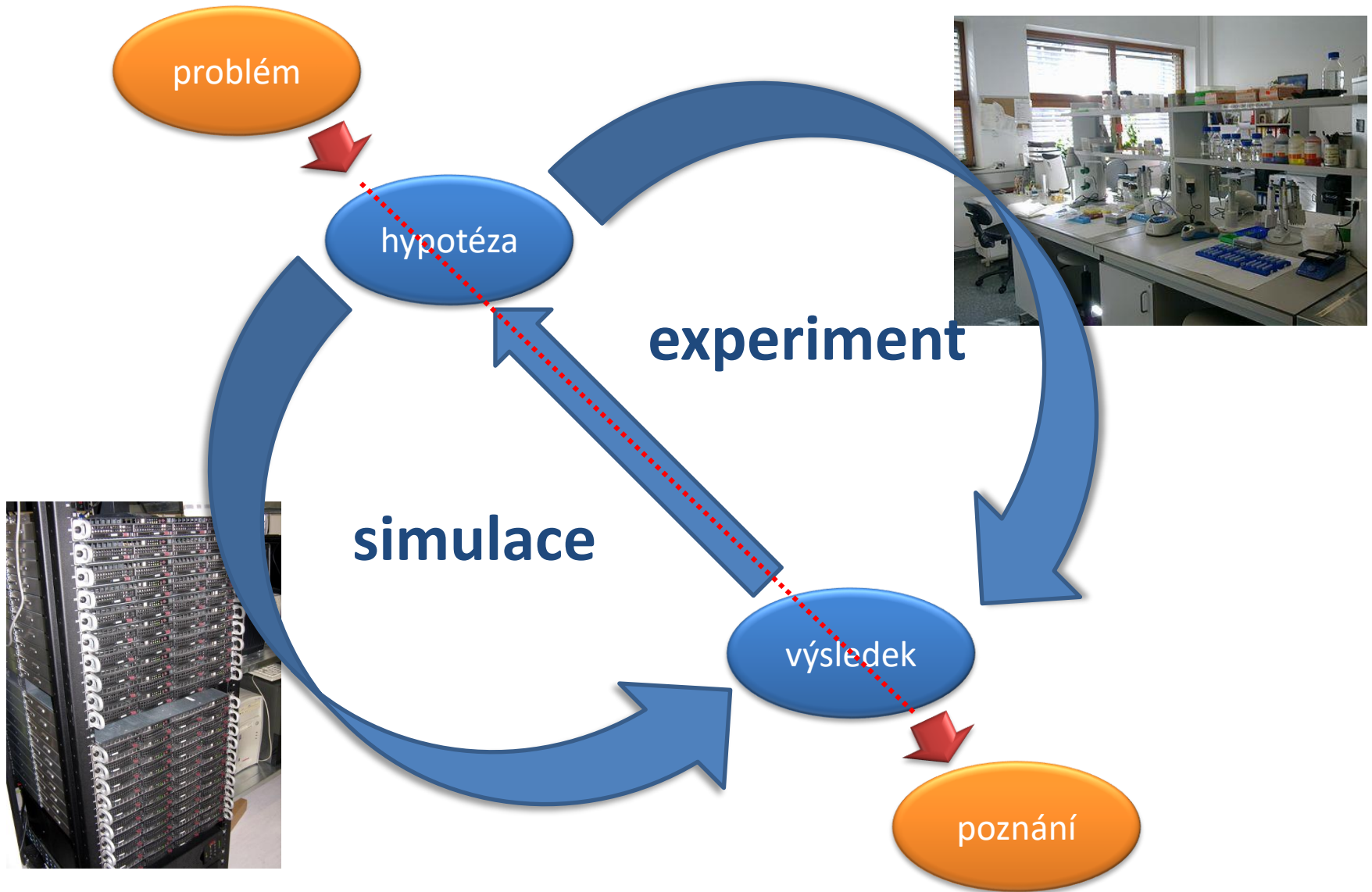
The Nobel Prize in Chemistry 1998 was divided equally between

Walter Kohn "for his development of the **density-functional theory**" and
John A. Pople "for his development of **computational methods in quantum chemistry**"

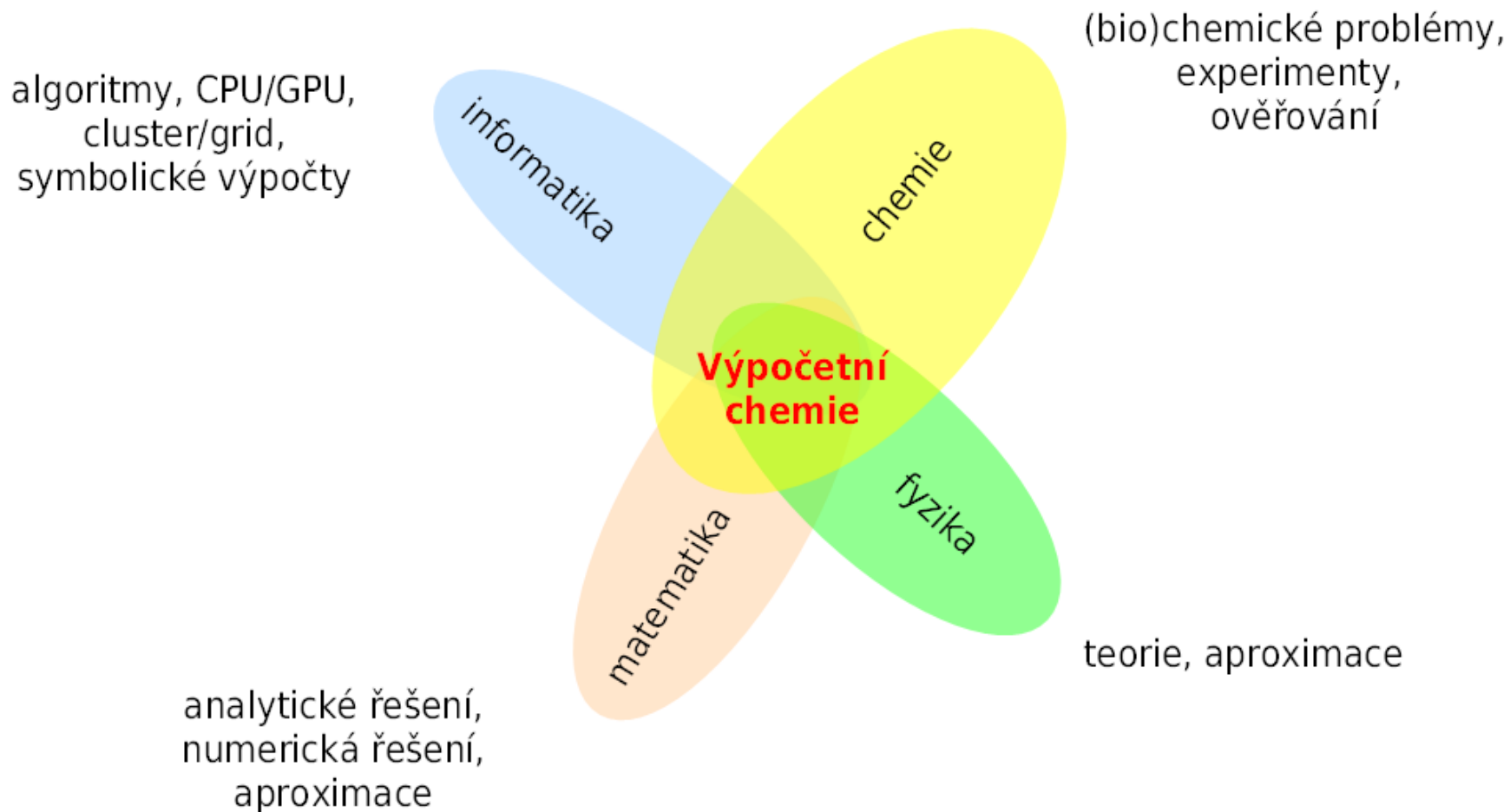
Development of Multiscale Models for Complex Chemical Systems

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/
http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/

Experiment vs výpočetní chemie

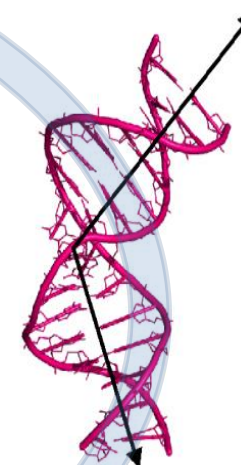
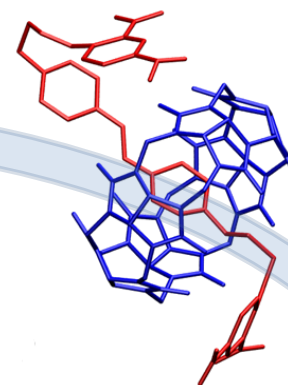
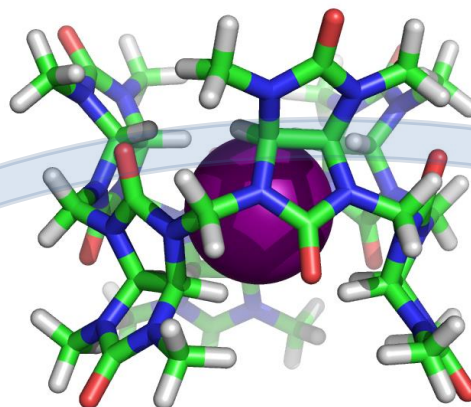
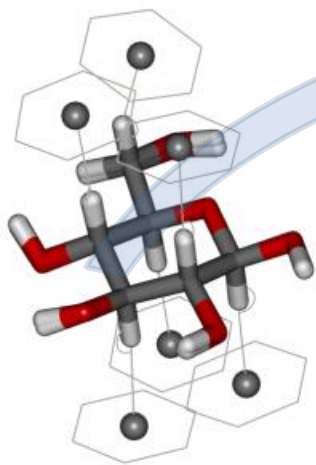


Interdisciplinární obor



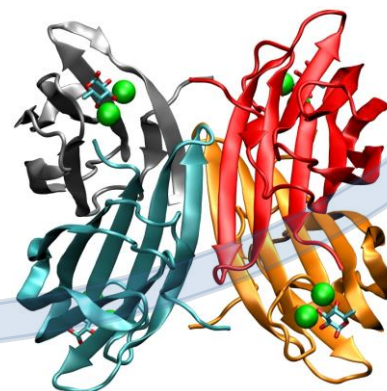
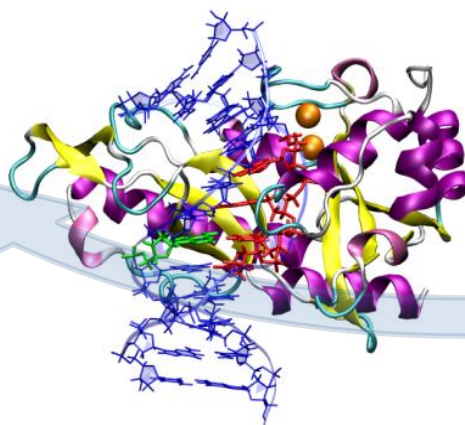
Co studujeme ...

"malé" komplexy



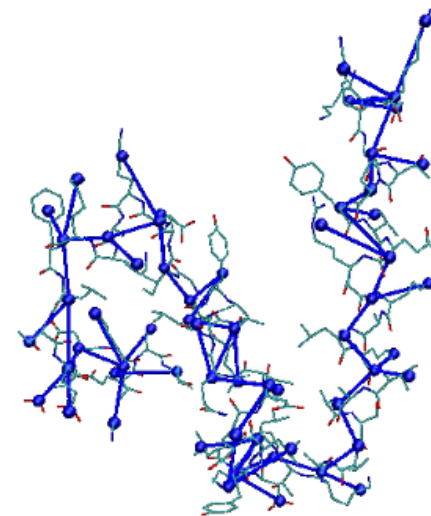
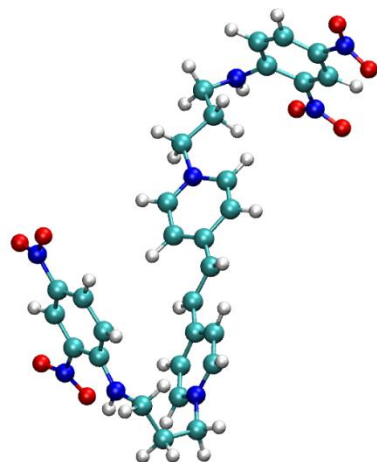
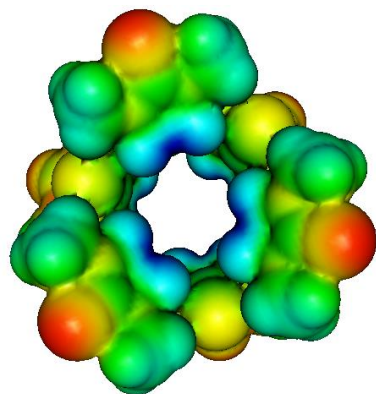
atomové rozlišení

biomolekulární
systémy



vybrané systémy studované skupinou výpočetní chemie

Úrovně teorie



Kvantová mechanika

Molekulová mechanika

***Coarse-grained* mechanika**

atomové rozlišení

bead resolution

reaktivita

konformační pohyby

pohyb domén, folding

až 1'000 atomů *

až 1'000'000 atomů *

až 1'000'000 beads *

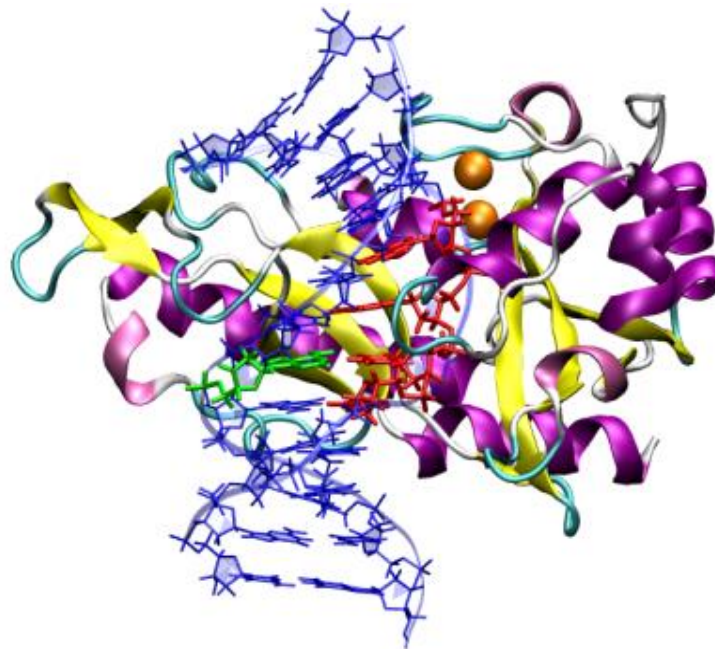
až 100 ps *

až 1 μ s *

až ms *

Projekty

Studium (bio)molekulárních systémů



Kvantově chemické výpočty

časově nezávislá Schrödingerova rovnice

$$\hat{H}\psi_k(\mathbf{r}) = E_k\psi_k(\mathbf{r})$$

Formální škálování	Metody			
	HF	CI metody	MP metody	CC metody
$N^4 \rightarrow N^2 \rightarrow N^1$	HF, DFT			
N^5			MP2	CC2 (iterativní)
N^6		CISD	MP3, MP4(SDQ)	CCSD (iterativní)
N^7			MP4	CCSD(T), CC3 (iterativní)
N^8		CISDT	MP5	CCSDT
N^9			MP6	
N^{10}		CISDTQ	MP7	CCSDTQ (iterativní)

Škálování, časová náročnost: http://en.wikipedia.org/wiki/Time_complexity

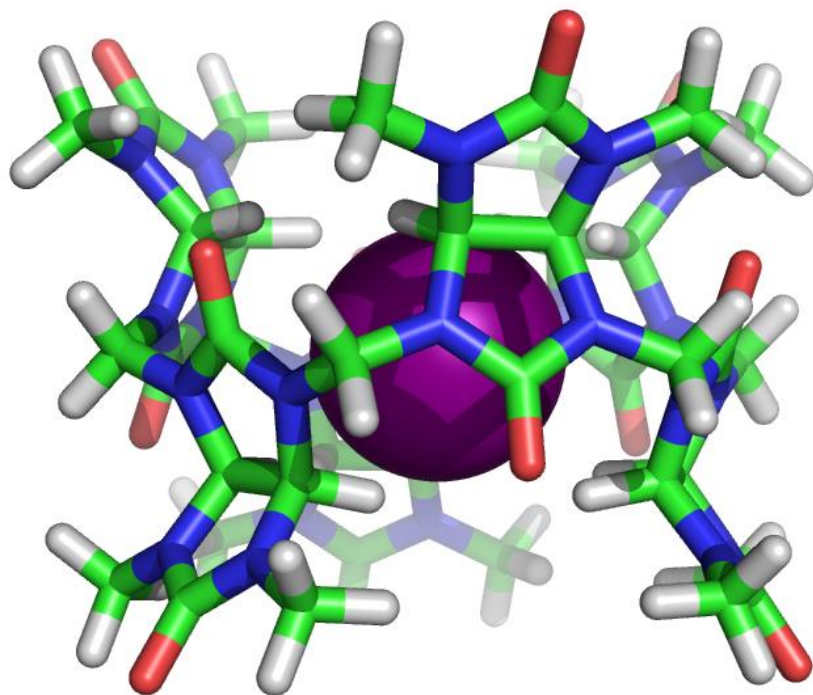
HF - Hartreeho–Fockova metoda, DFT - teorie funkcionálu hustoty,

CI - metody konfigurační interakce, MP - Møllerova–Plessetova poruchová teorie,

CC - metoda vázaných klastrů, N - počet bázeových funkcí

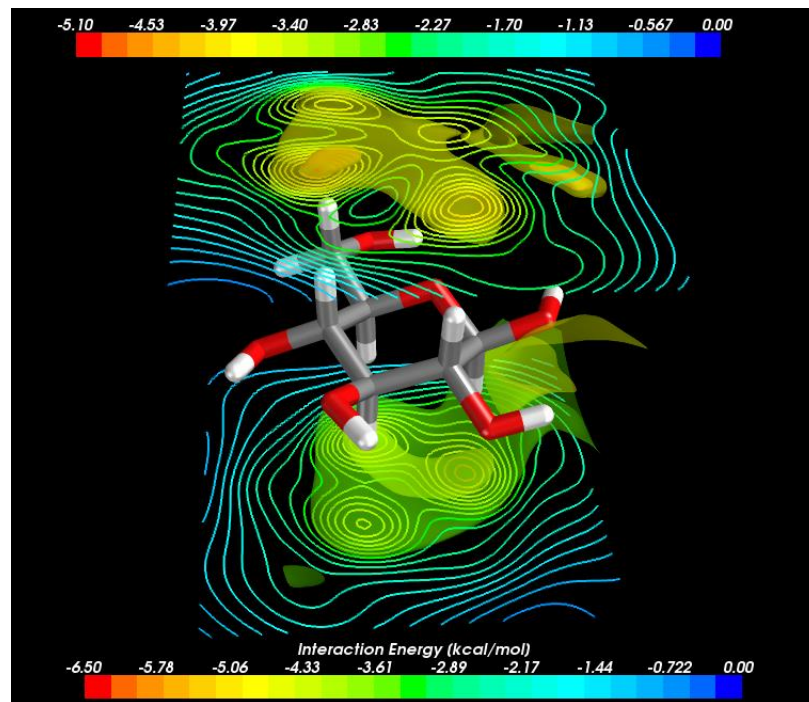
Jensen, F. Introduction to computational chemistry; 2nd ed.; John Wiley & Sons: Chichester, England; Hoboken, NJ, 2007.

Kvantově chemické výpočty



supramolekulární komplexy

vazebné schopnosti sacharidů



Molekulová mechanika

Schrödingerova rovnice => kvantově mechanický pohled

$$H_a \Psi(r) = E(R) \Psi(r)$$

aproximace využívající klasickou fyziku
neuvažuje se explicitní pohyb elektronů
(pohyb je implicitně zahrnut v empirických parametrech)

$$E(R) = \underbrace{E_{bonds} + E_{angles} + E_{torsions}}_{\text{vazebné příspěvky}} + \underbrace{E_{el} + E_{vdw}}_{\text{nevazebné příspěvky}}$$

Klasická fyzika => mechanický pohled

vazebné příspěvky

nevazebné příspěvky

Formální škálování: $N^2 \rightarrow N \log_2 N$

N - počet atomů

Molekulová dynamika

$$-\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}} = \mathbf{F} \quad \boxed{\mathbf{F}_i = m_i \mathbf{a}_i} \quad \mathbf{a}_i = \frac{d^2 \mathbf{r}_i}{dt^2}$$

II. Newtonův pohybový zákon (zákon síly)

$$-\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}} = m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$

soustava diferenciálních rovnic druhého řádu
vyžaduje numerické řešení



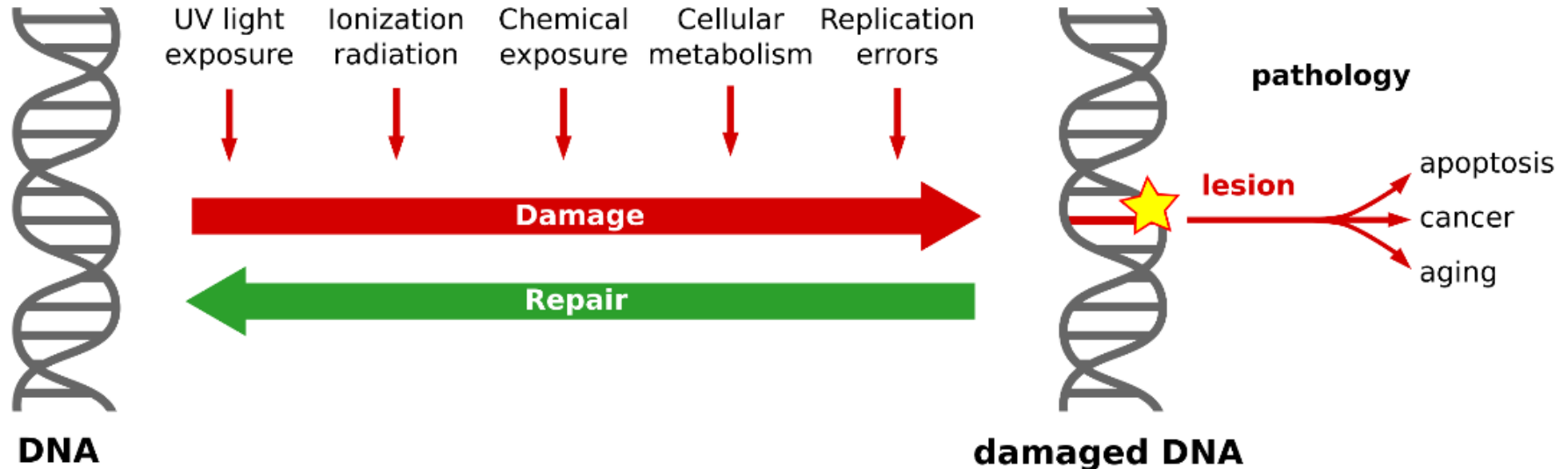
diskretizace molekulárního pohybu v krátkých
časových intervalech

dáno nejrychleším pohybem (vibrace vazeb) \longrightarrow **1 fs** typický integrační krok

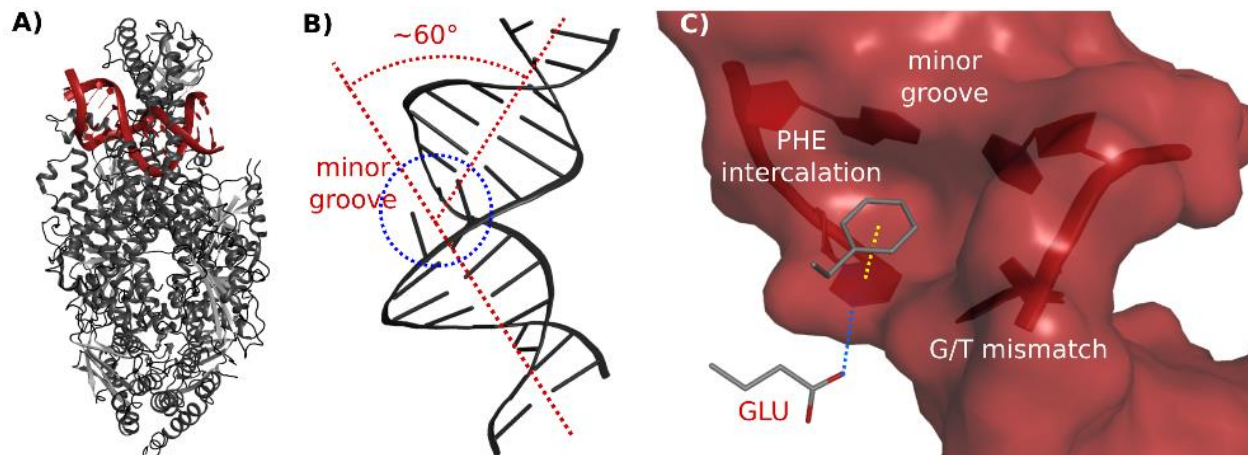
Nedokonalosti v integraci se odstraňují použitím **termostatů** a **barostatů**, které zároveň zajišťují požadované podmínky simulace.

Oprava poškozené DNA

DNA je vystavována celé řadě faktorů, které ji poškozují. Aby nedošlo k degradaci genetické informace, poškozená DNA je opravována celou řadou mechanismů, které pracují s různou účinností. Cílem projektu je pochopit způsob detekce poškození na molekulární úrovni s primárním zaměřením na mechanické vlastnosti poškozené DNA.

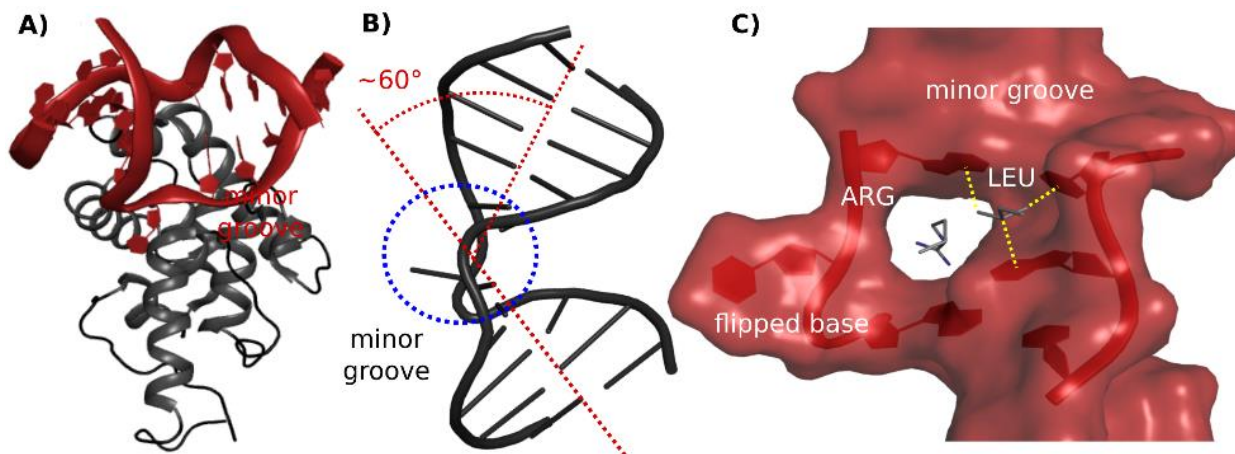


MMR a BER



MMR – Mismatch Repair MutS α

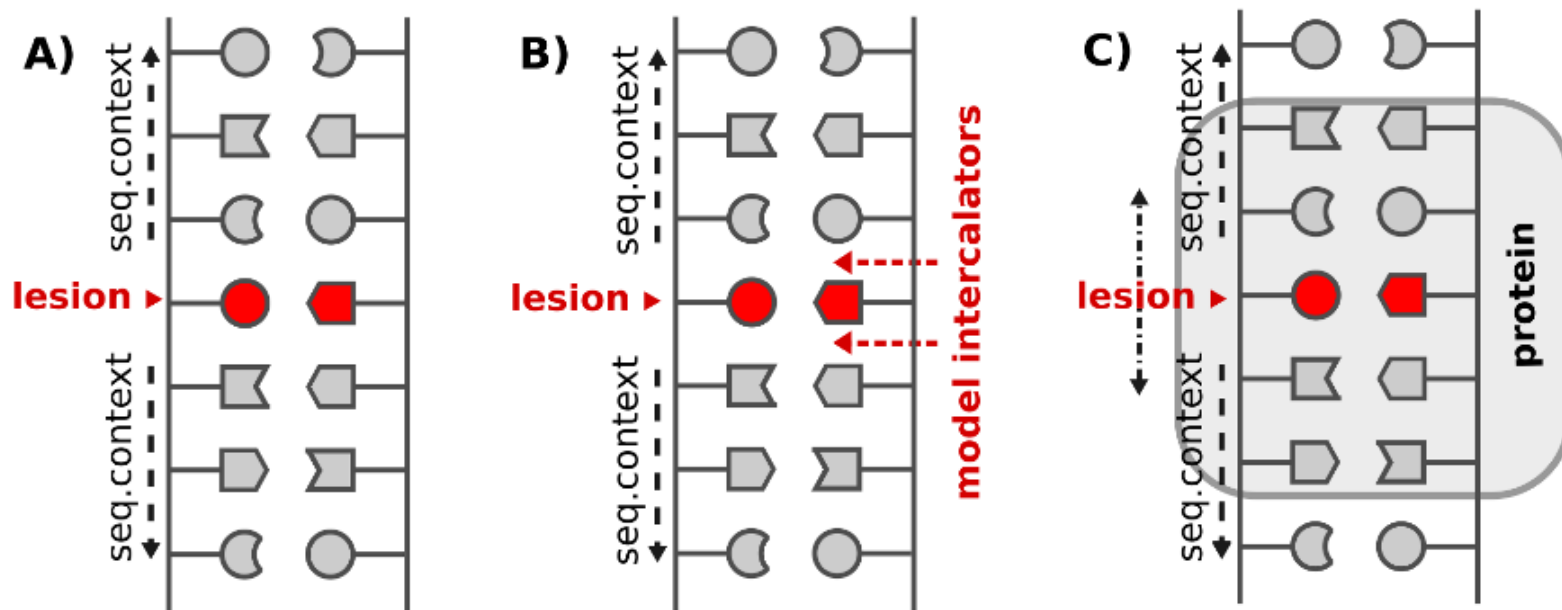
DNA je ohnutá a
interkalovaná



BER – Base Excision Repair MBD4 glycosylase domain

DNA je ohnutá a
interkalovaná

Studované modely



Spoluřešitelé (škoolitelé či konzultanti):

- Mgr. Kamila Réblová, Ph.D.
(Lékařská genomika - Centrum molekulární medicíny - Středoevropský technologický institut)
- Mgr. Naděžda Špačková, Ph.D.
(Ústav fyziky kondenzovaných látek - Fyzikální sekce - Přírodovědecká fakulta)

Metody:

- molekulová dynamika
- výpočty volných (Gibbsových) energií
- kvantově chemické výpočty

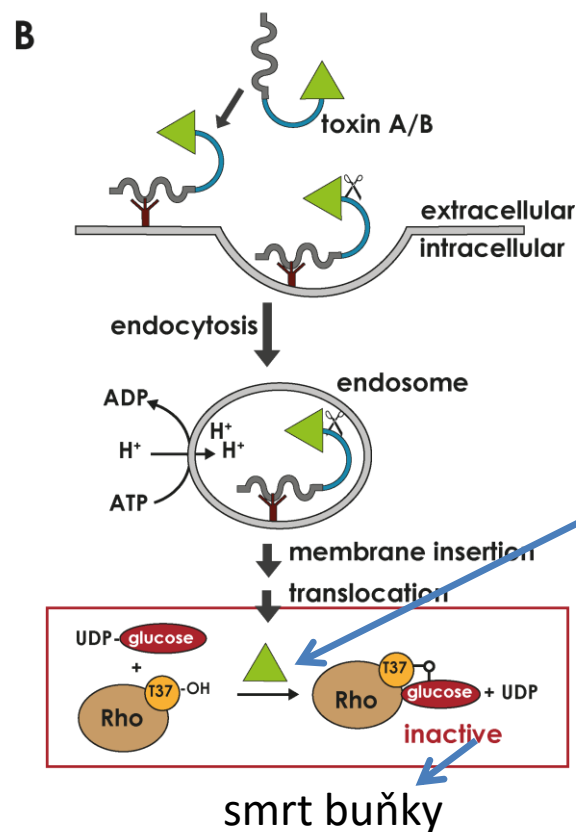
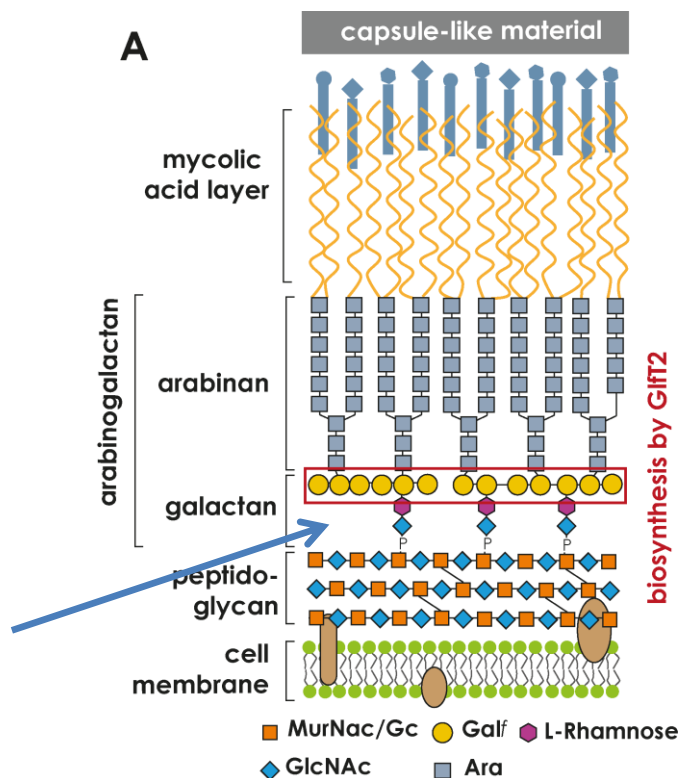
Glykosyltransferázy

Glykosyltransferázy jsou enzymy, které **katalyzují přenos aktivovaného cukerného zbytku** na oligosacharidy, proteiny či jiné biomolekuly. Jsou důležité v post-translační modifikaci proteinů, regulaci, či vytváření strukturní podpory.

Mycobacterium tuberculosis
(patogenní bakterie)

Clostridium difficile
(patogenní bakterie)

Motivace: inhibitor syntézy důležité složky membrány -> **antibiotikum**



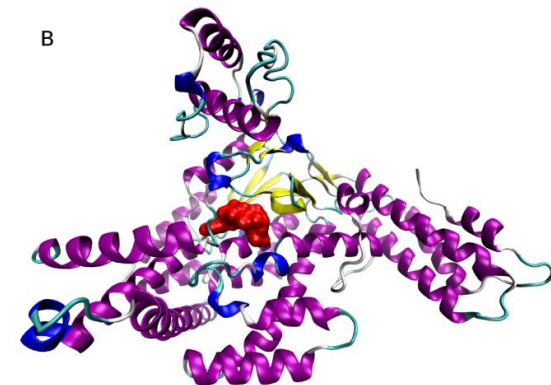
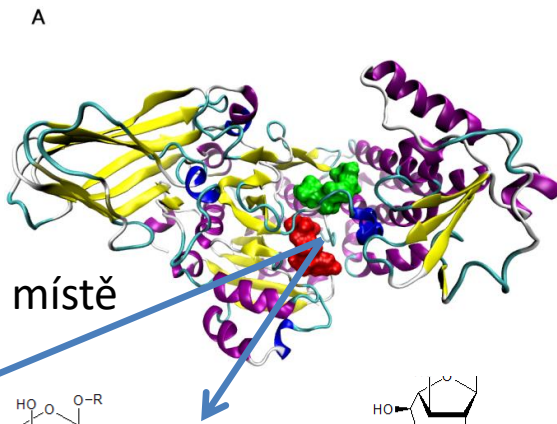
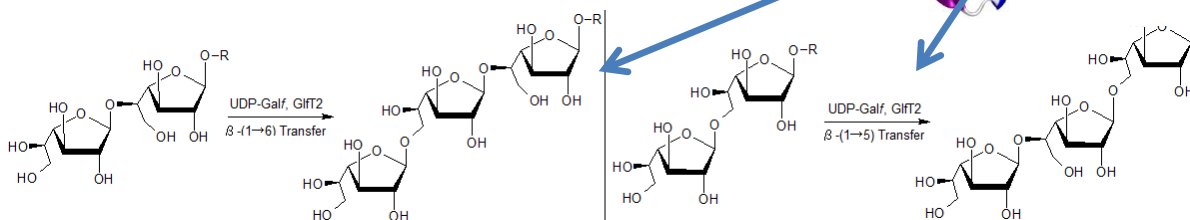
Motivace: inhibitor glykosyltransferázové aktivity toxinu -> **protijed**

Reakční mechanismy (QM/MM)

A) Glycosyltransferáza GifT2

B) Katalytická doména TcdB

dvě různé reakce v jednom aktivním místě

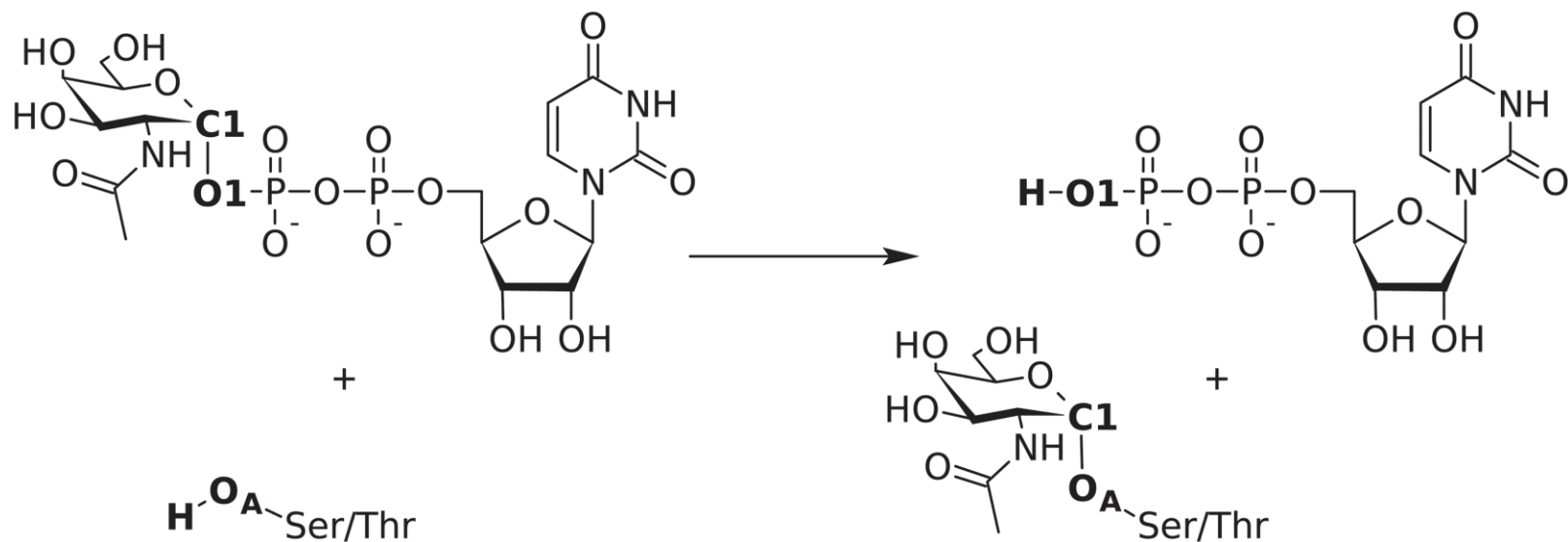


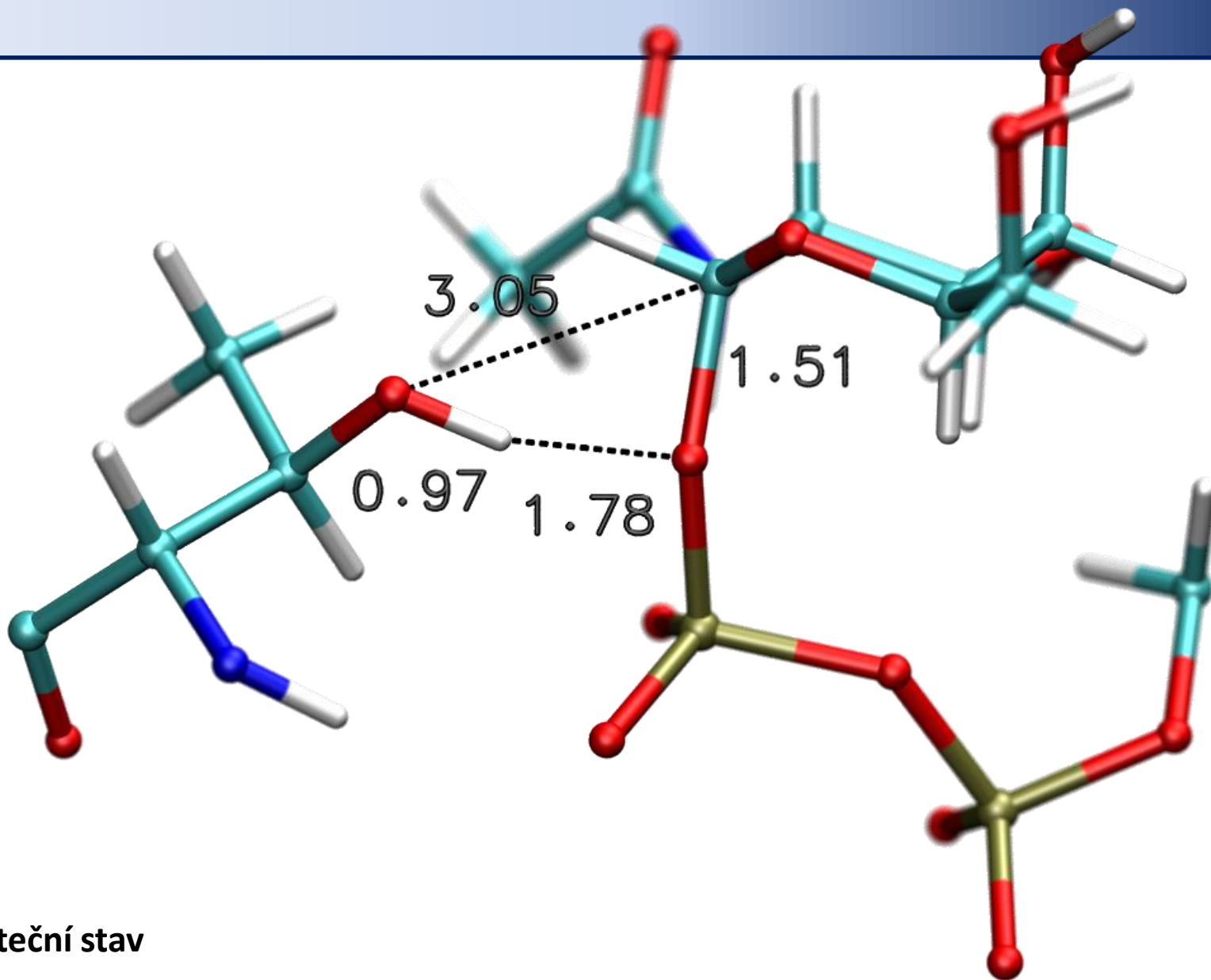
Školitelé či konzultanti:

- prof. RNDr. Jaroslav Koča, DrSc.
(Výpočetní chemie - Centrum strukturní biologie - Středoevropský technologický institut)
- Mgr. Stanislav Kozmon, Ph.D.
(Ústav chemie, Slovenská akademie věd)
- RNDr. Petr Kulhánek, Ph.D.
(Výpočetní chemie - Centrum strukturní biologie - Středoevropský technologický institut)
- Ing. Igor Tvaroška, DrSc.
(Ústav chemie, Slovenská akademie věd)

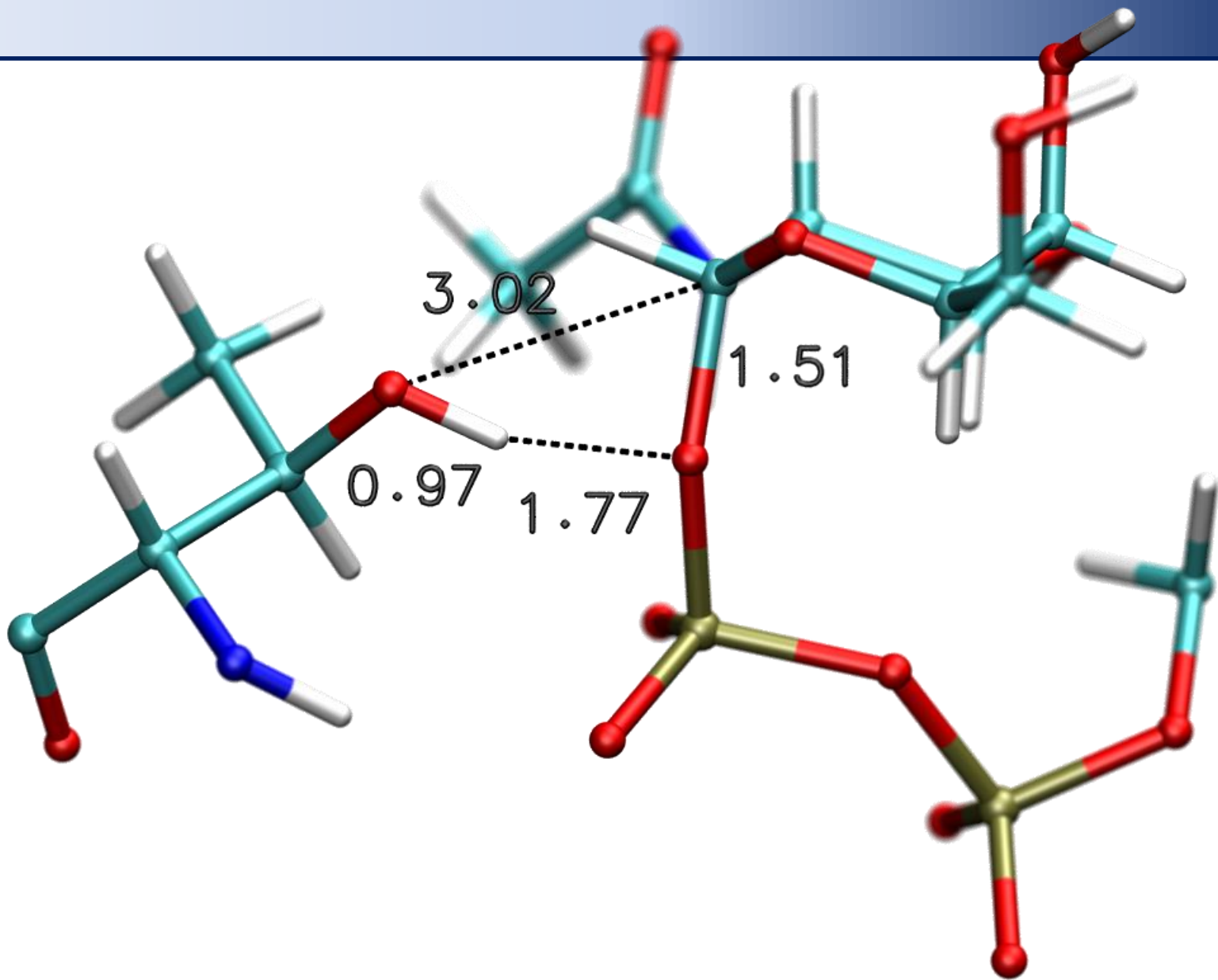
Příklad ppGalNAcT2

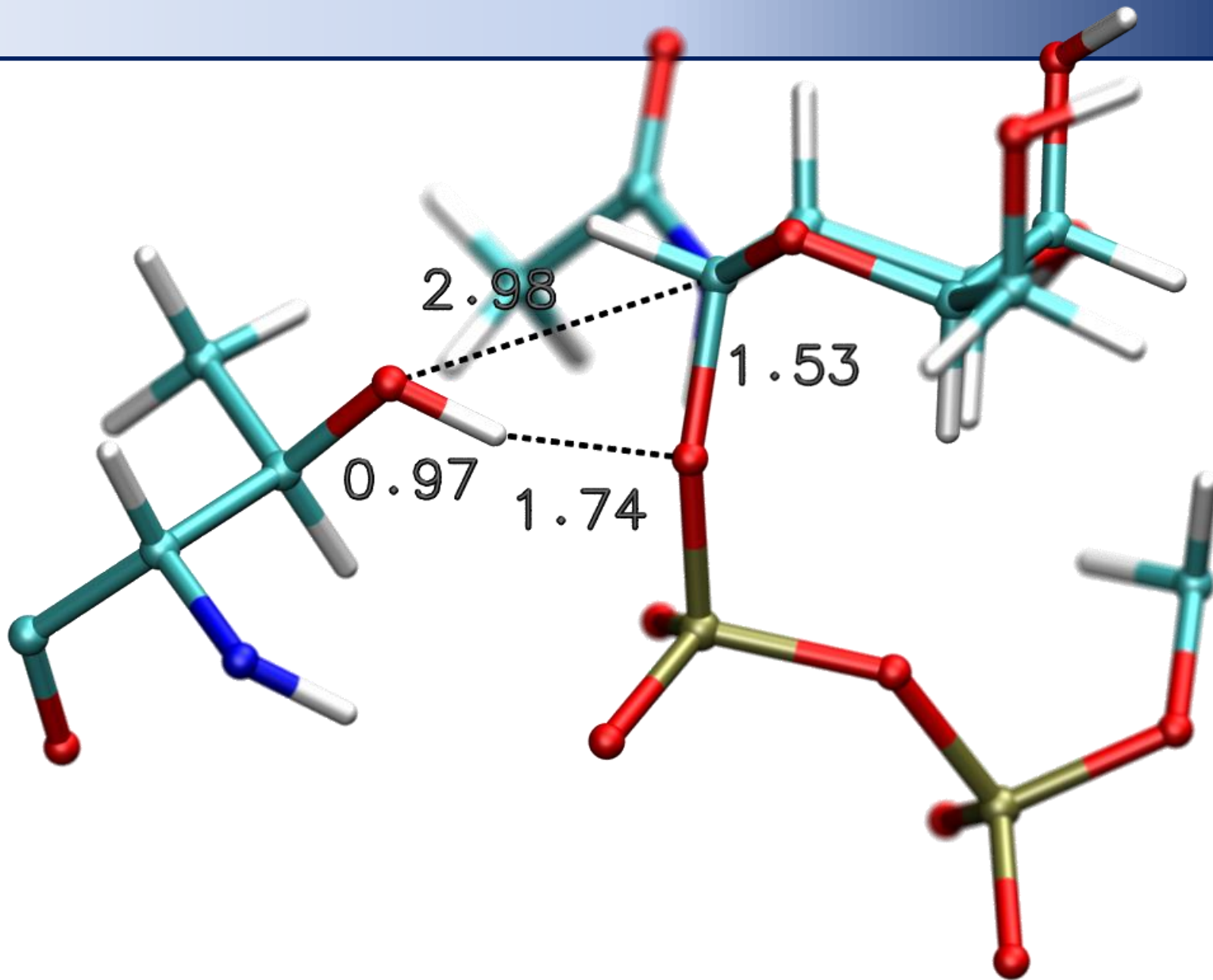
N-acetylgalaktosamin

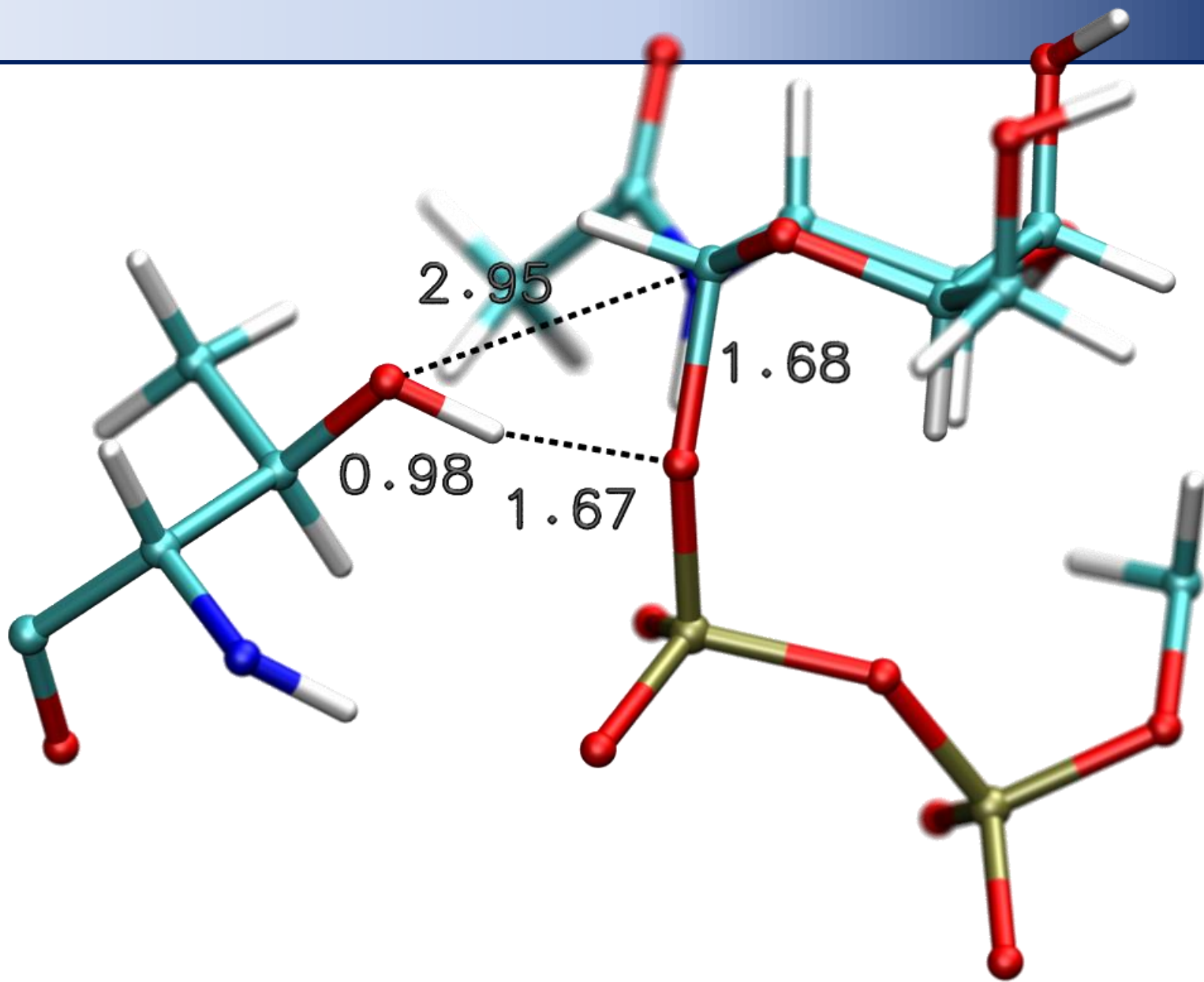


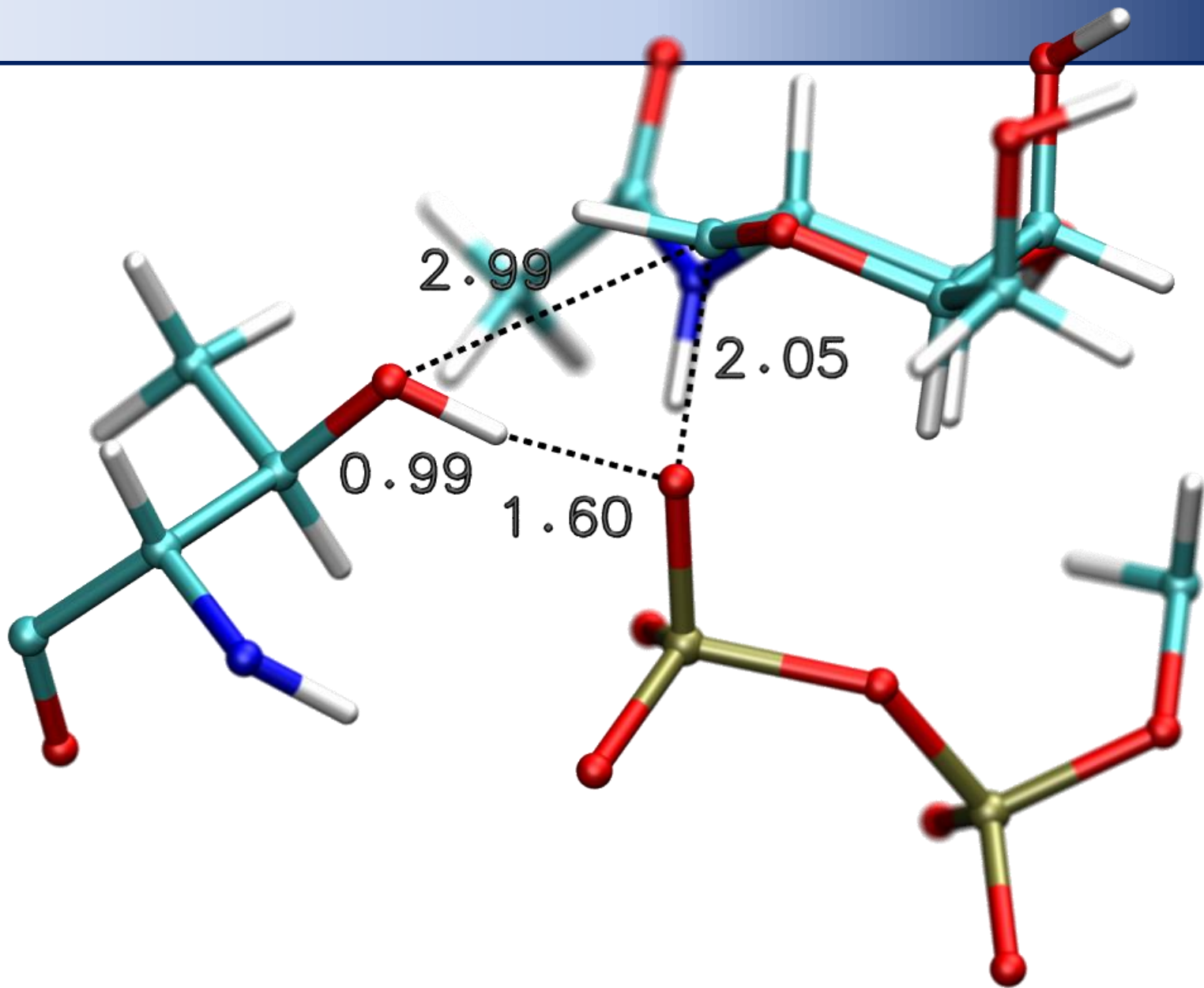


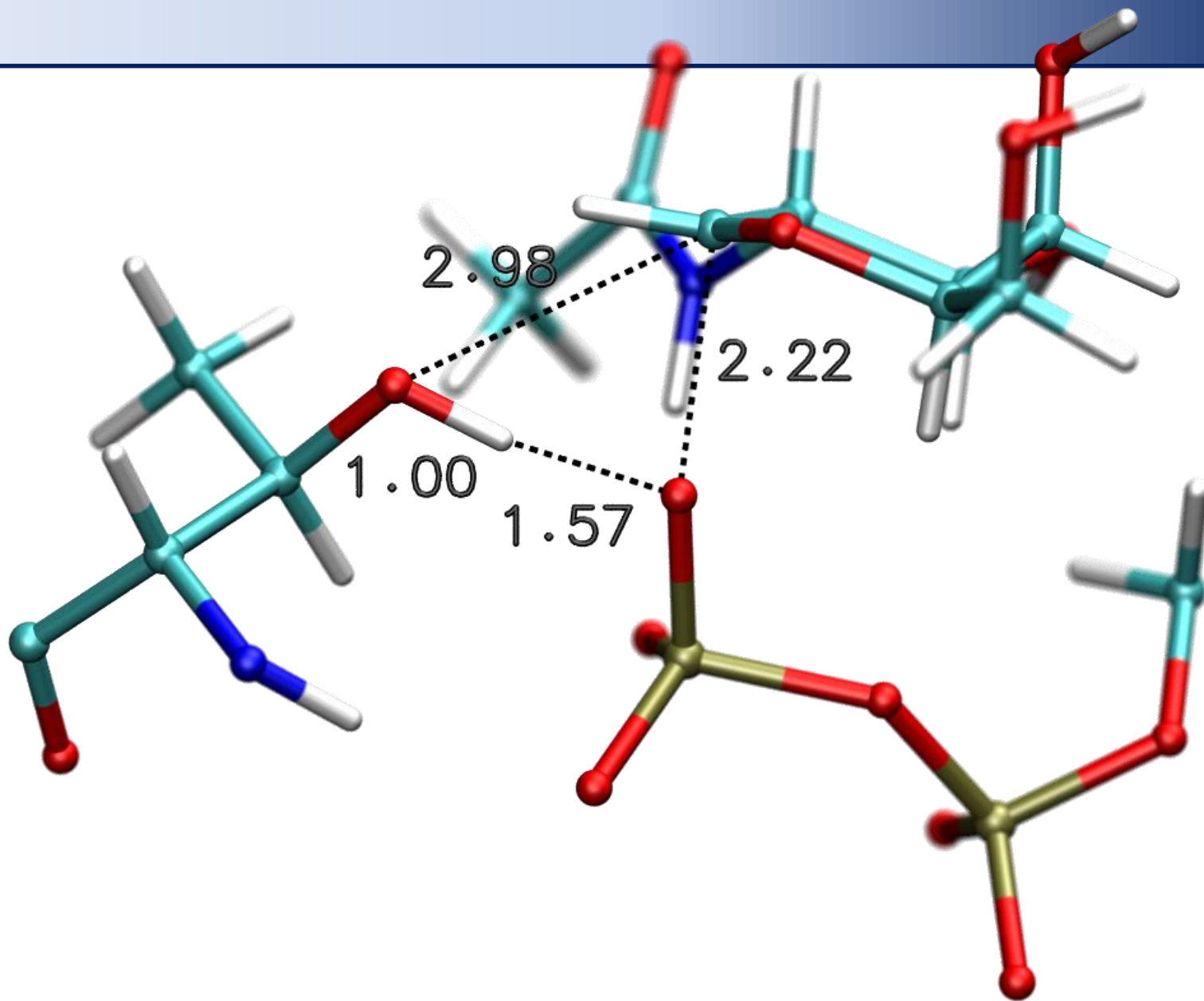
počáteční stav

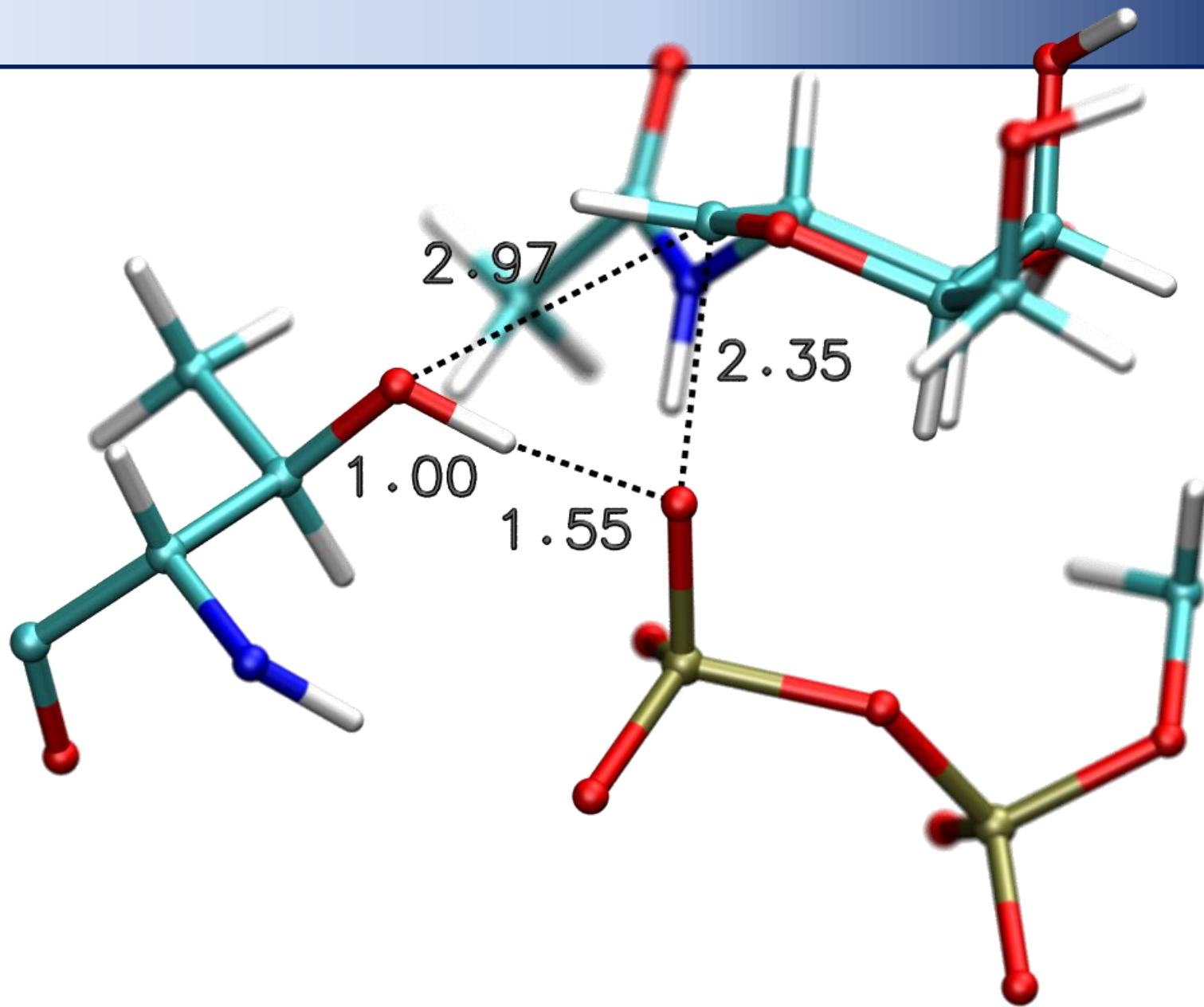


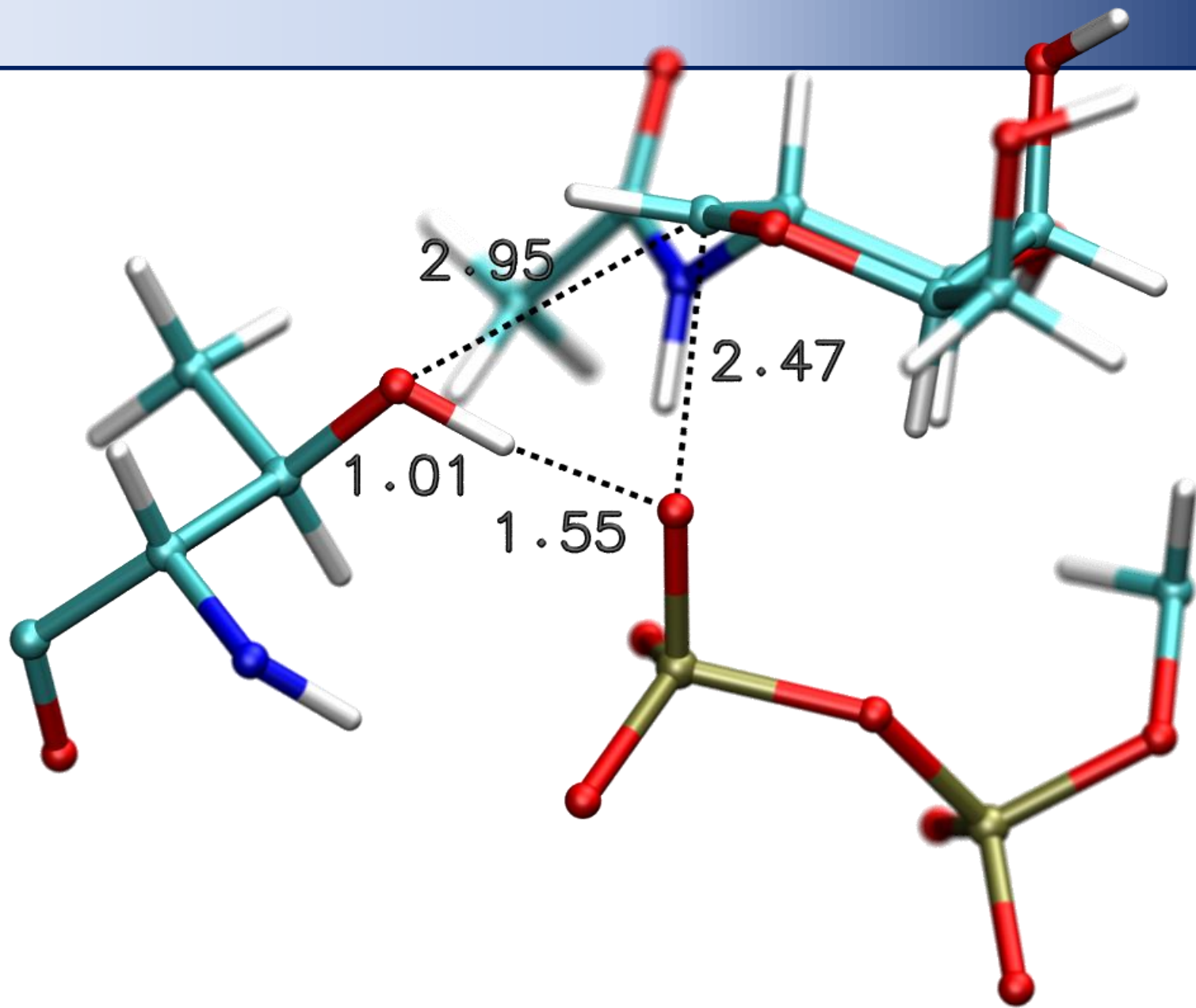


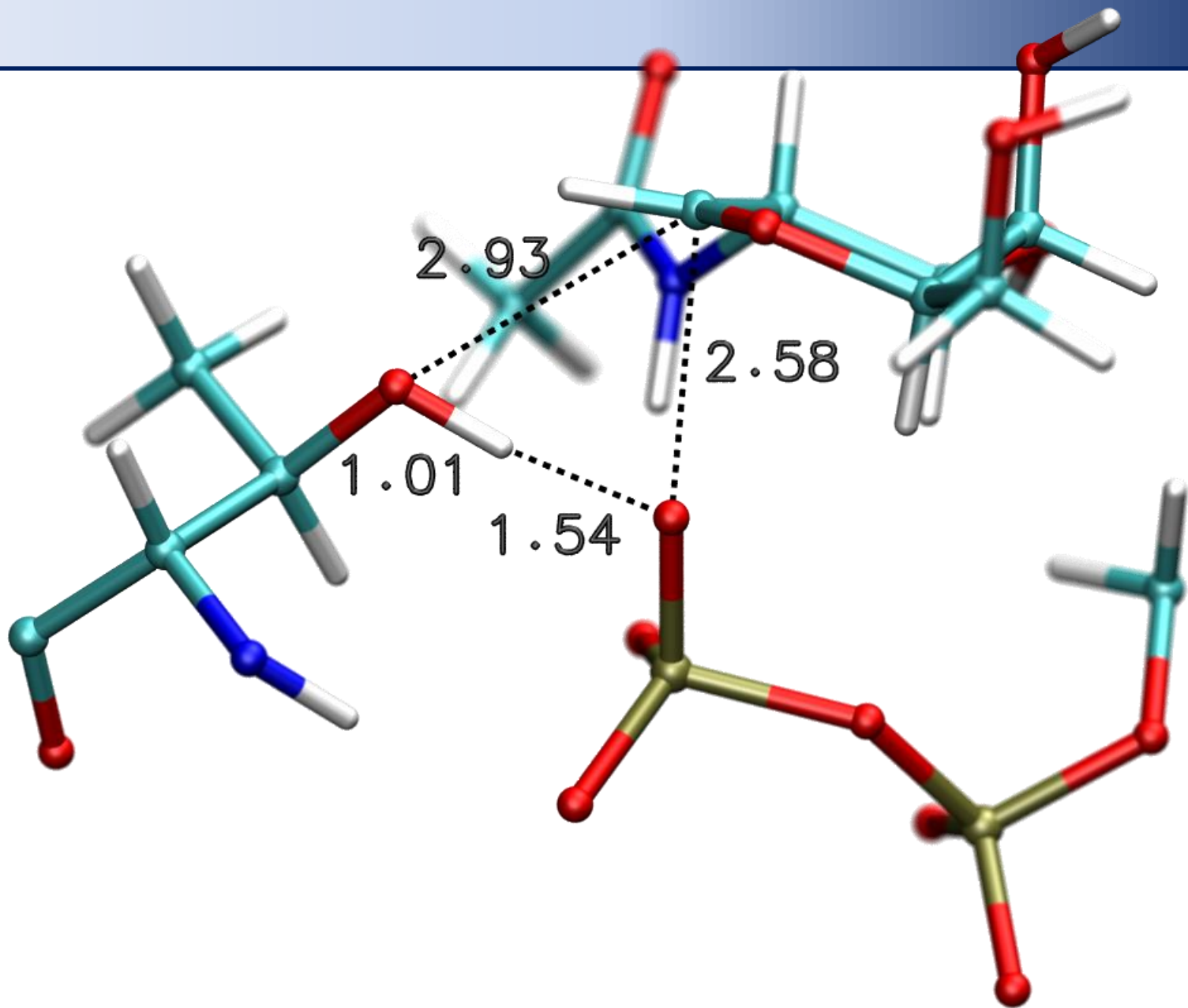


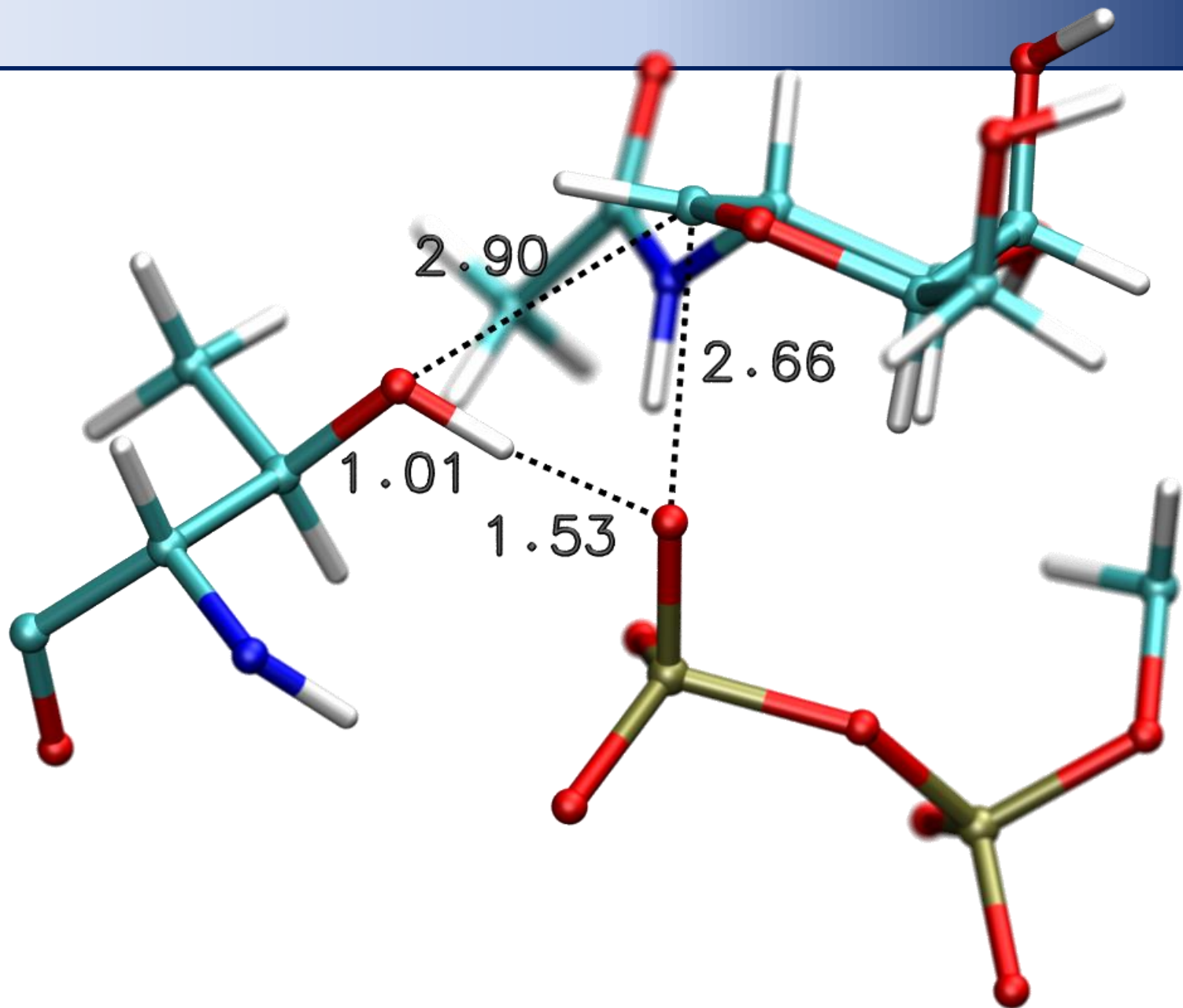


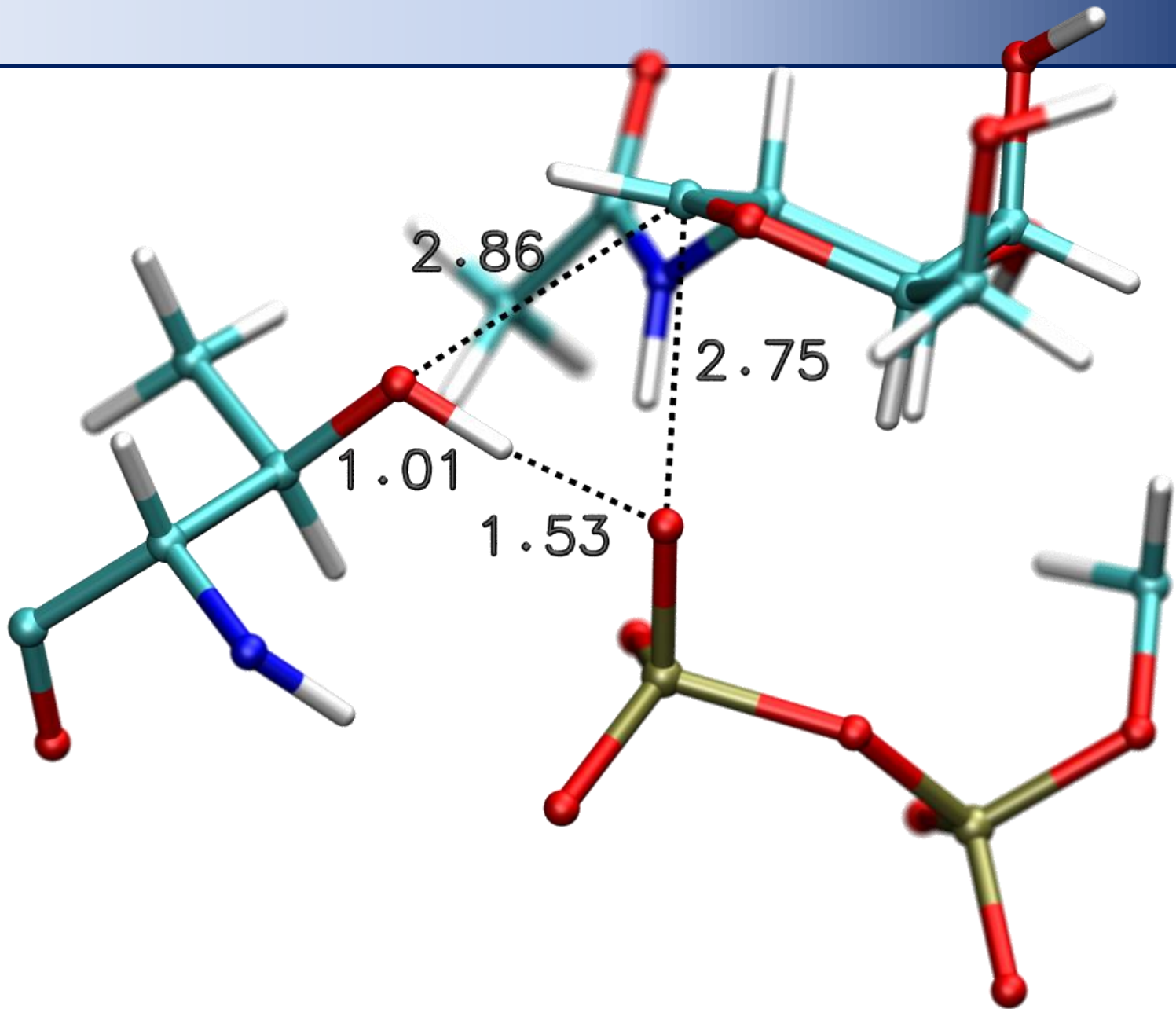


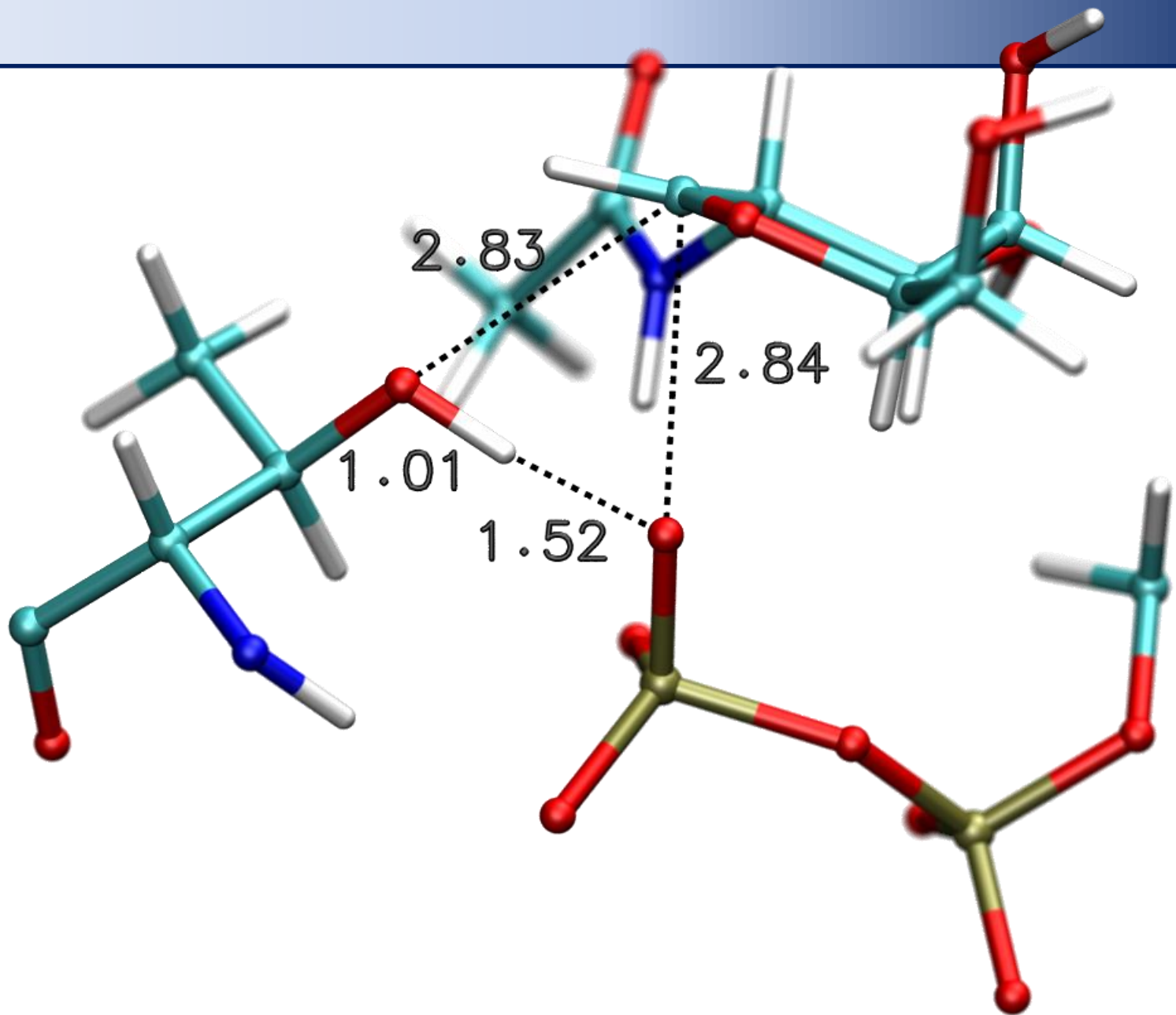


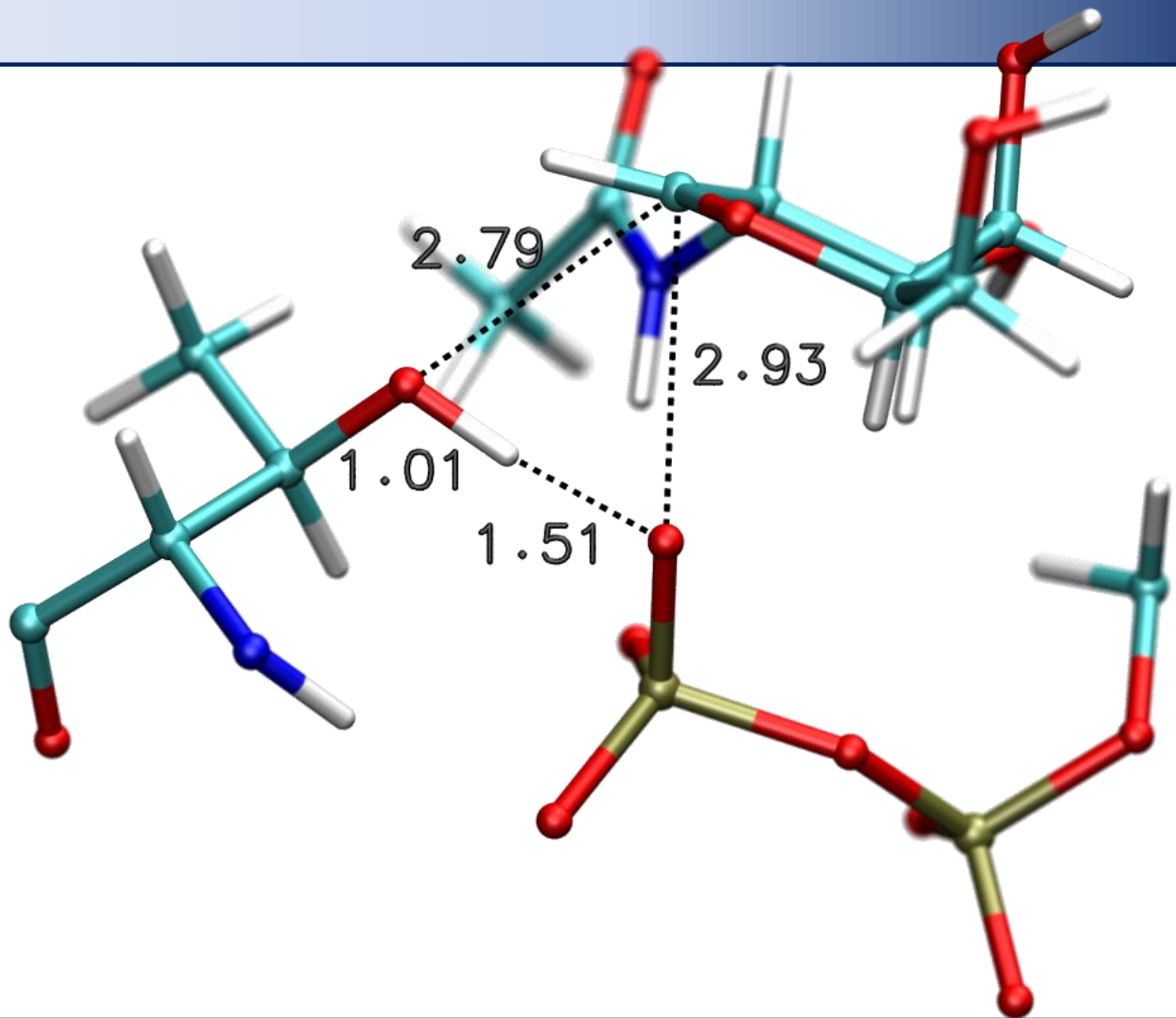


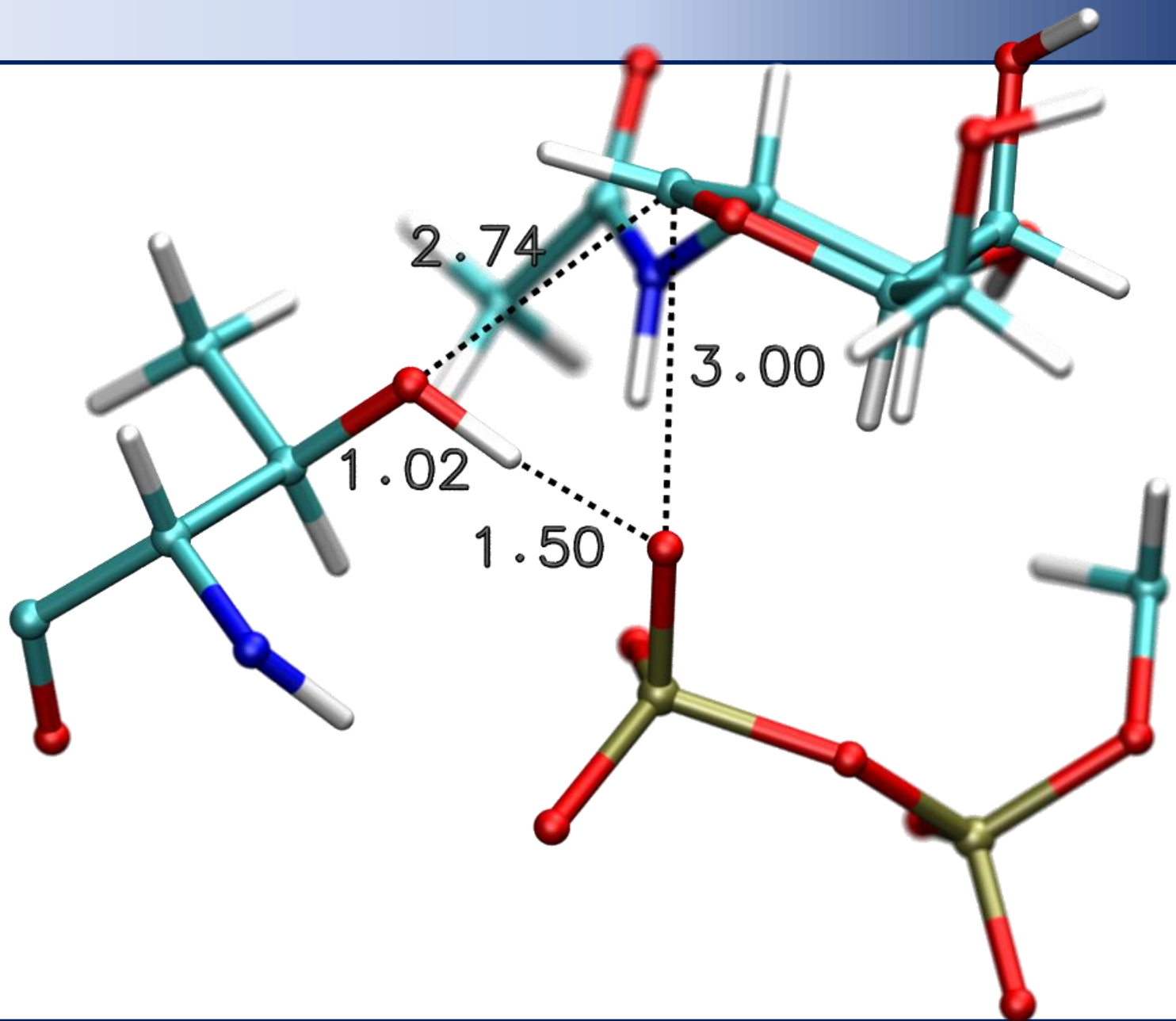


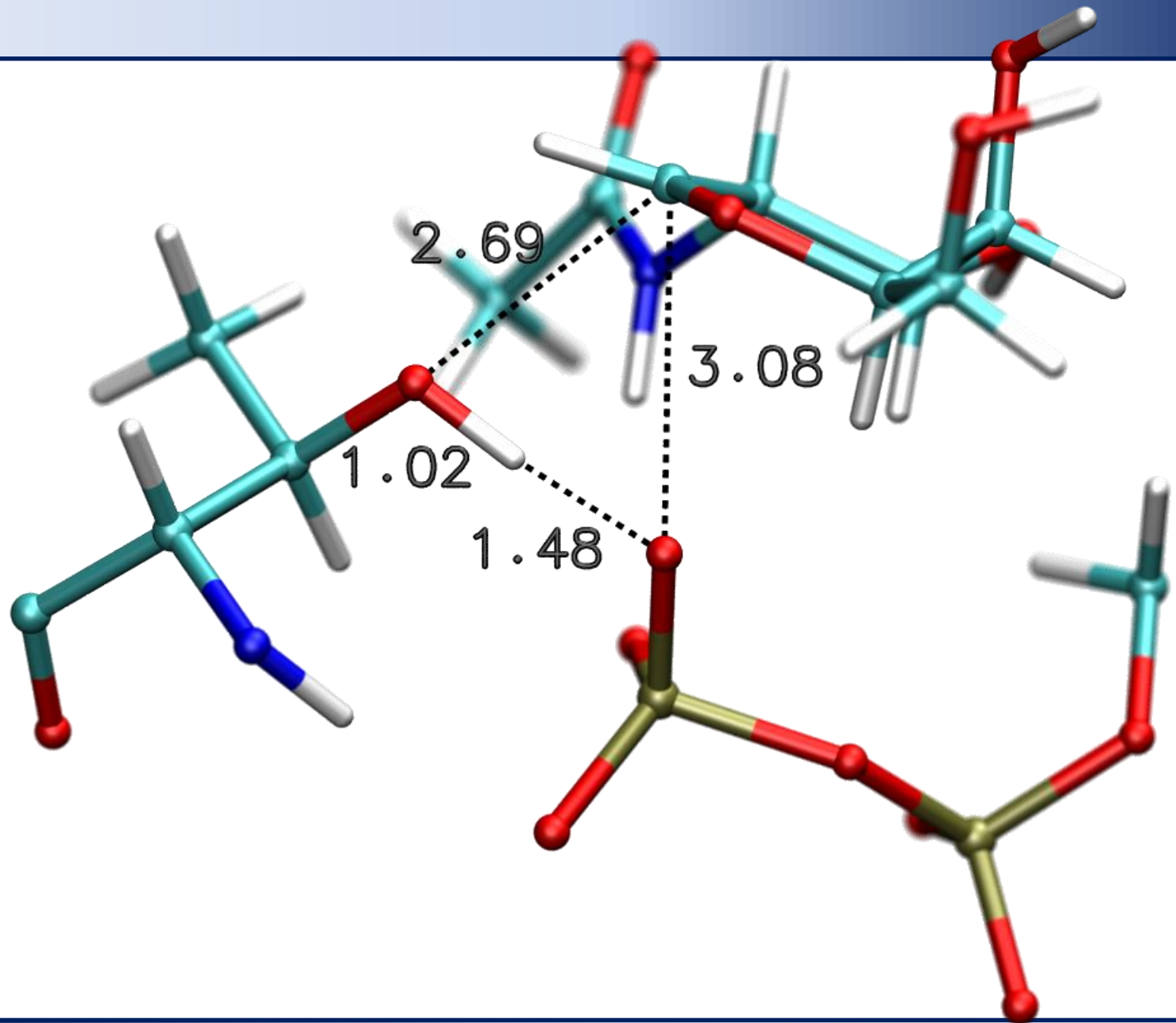


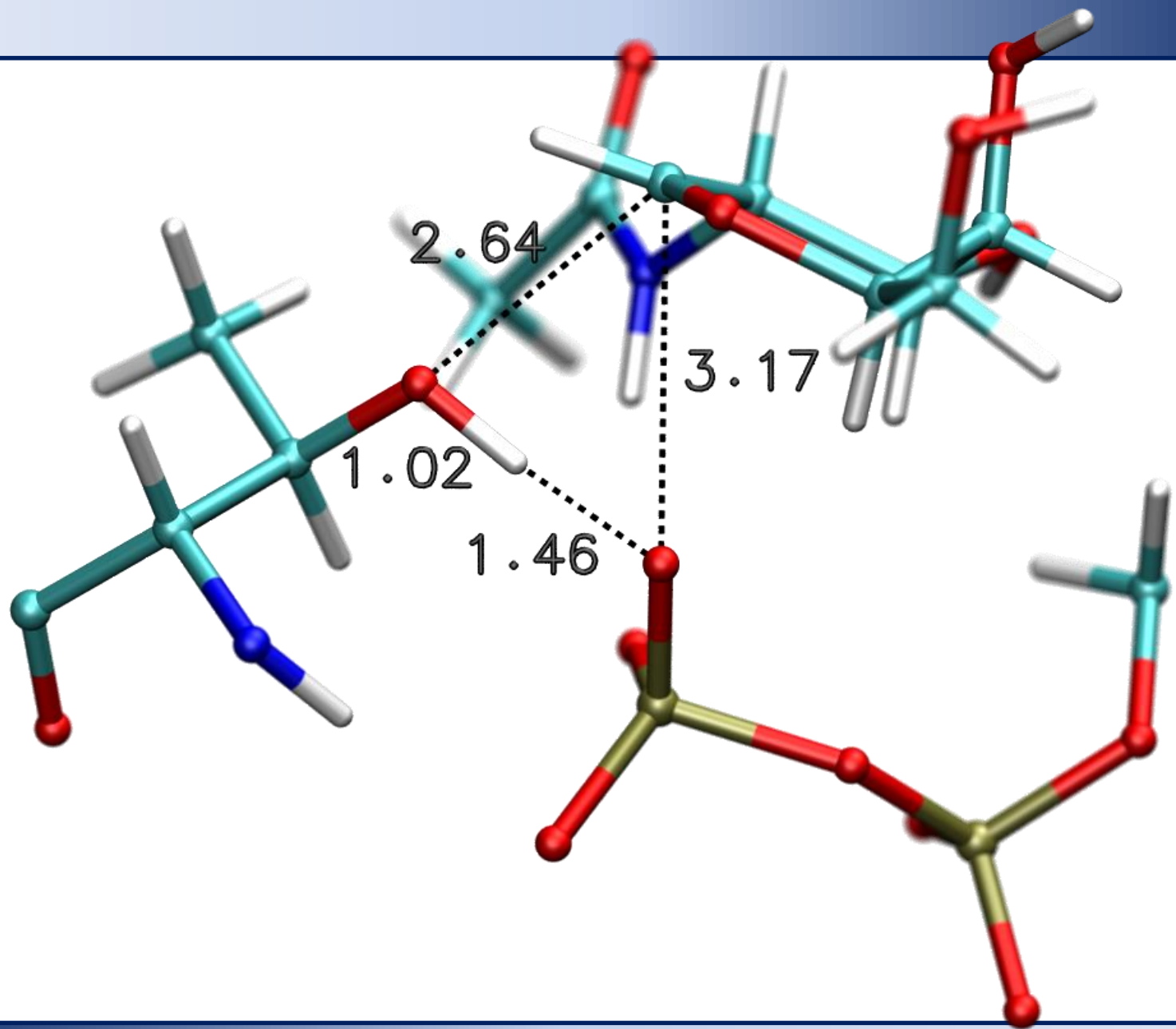


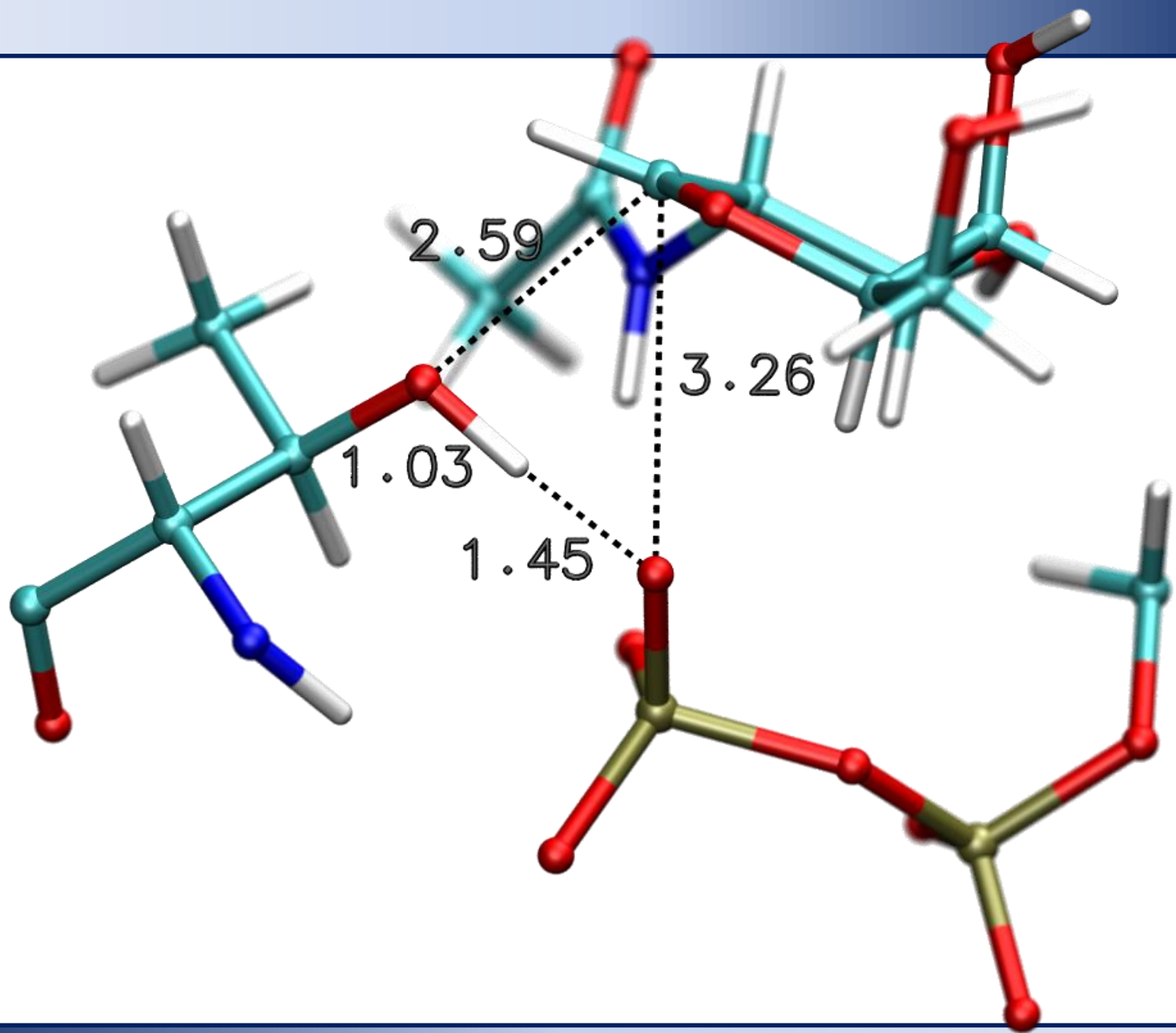


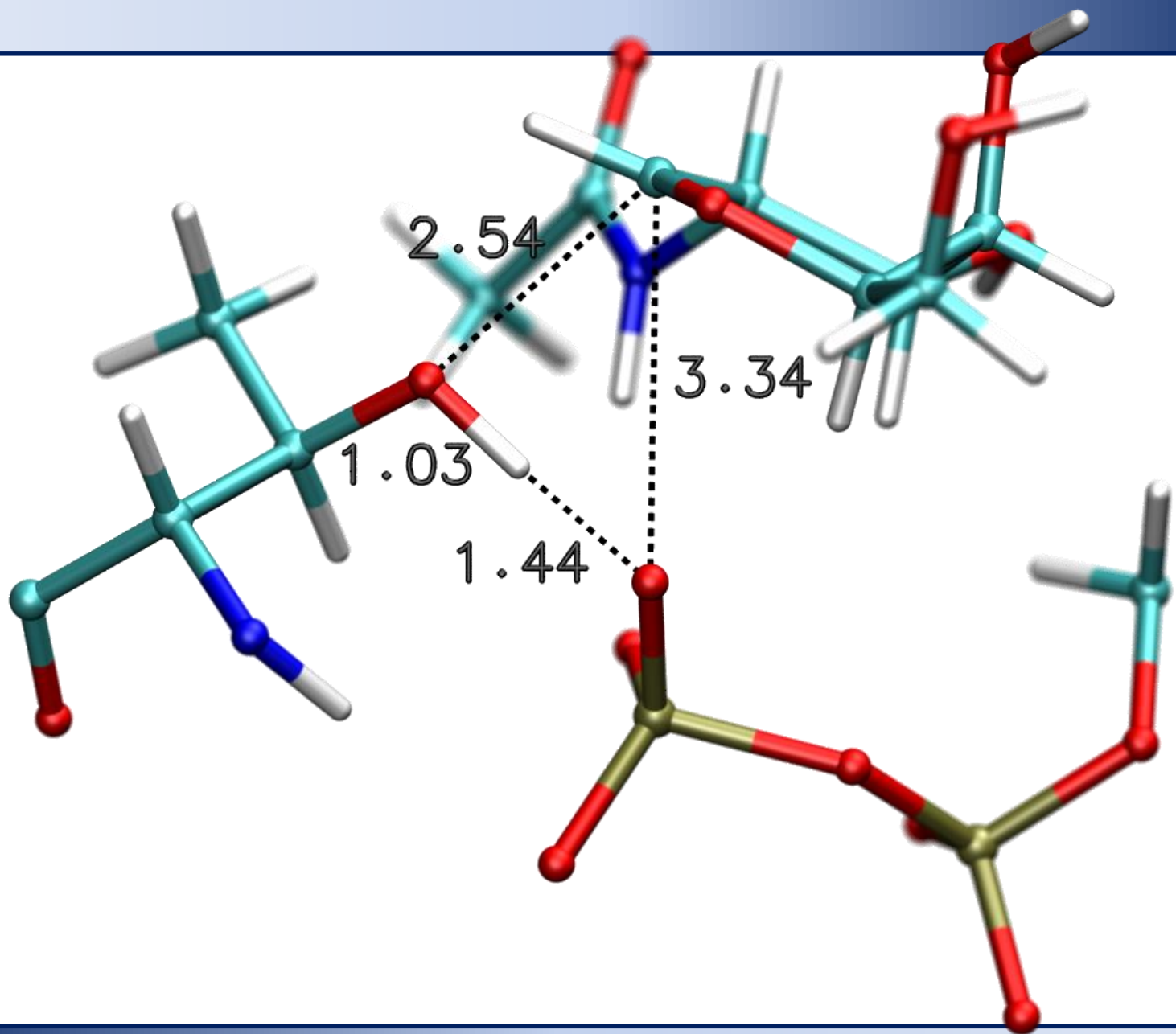


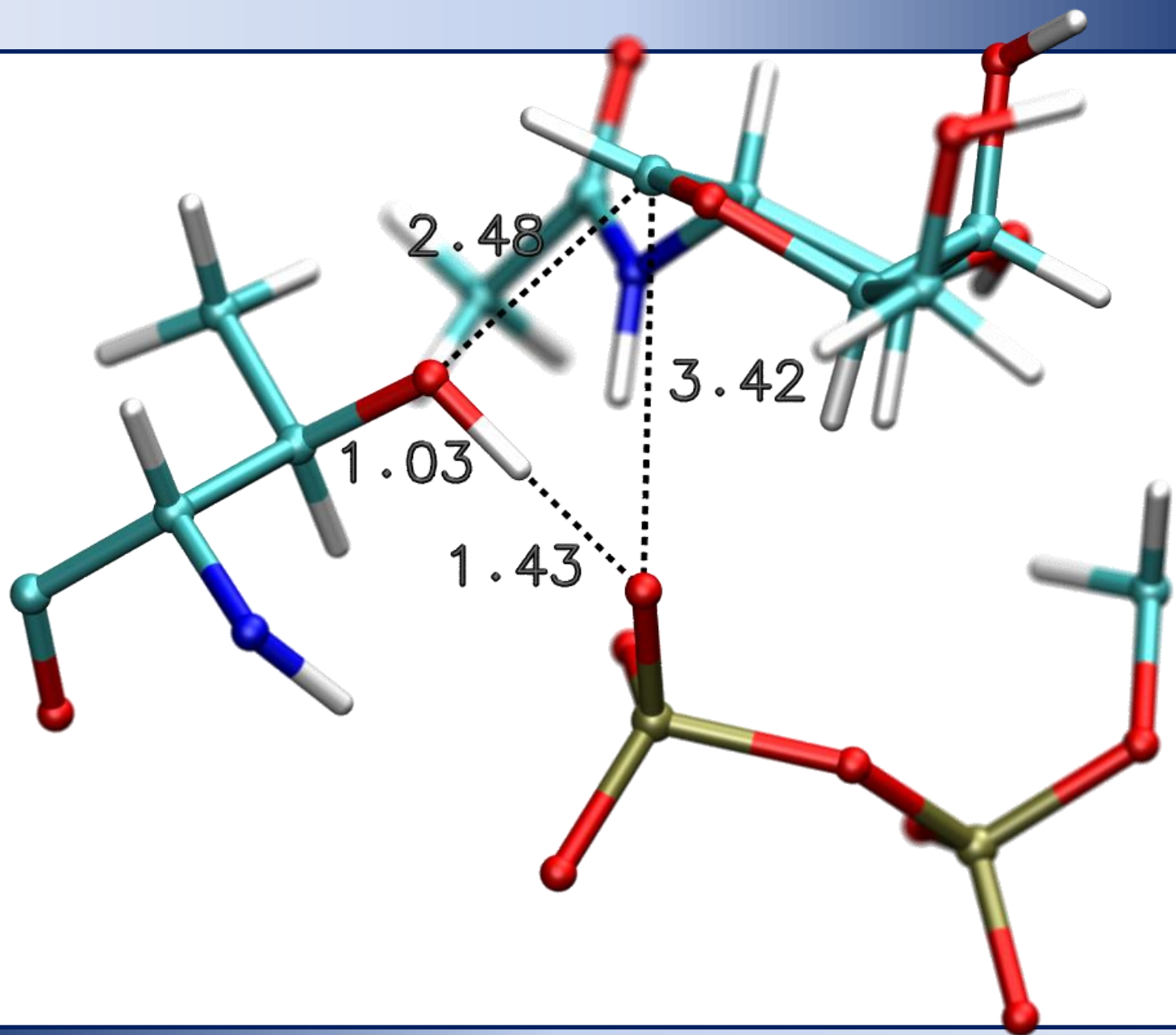


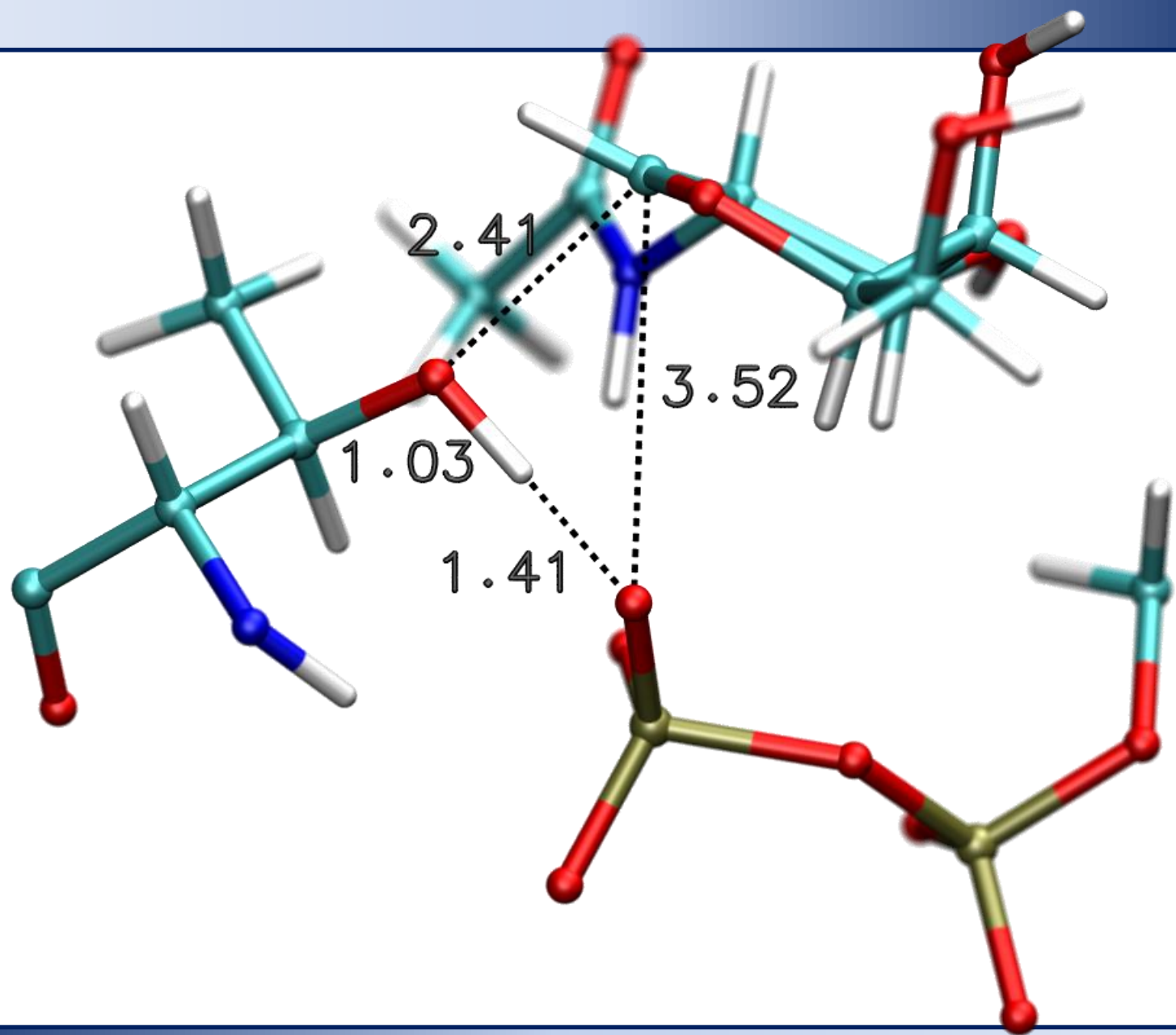


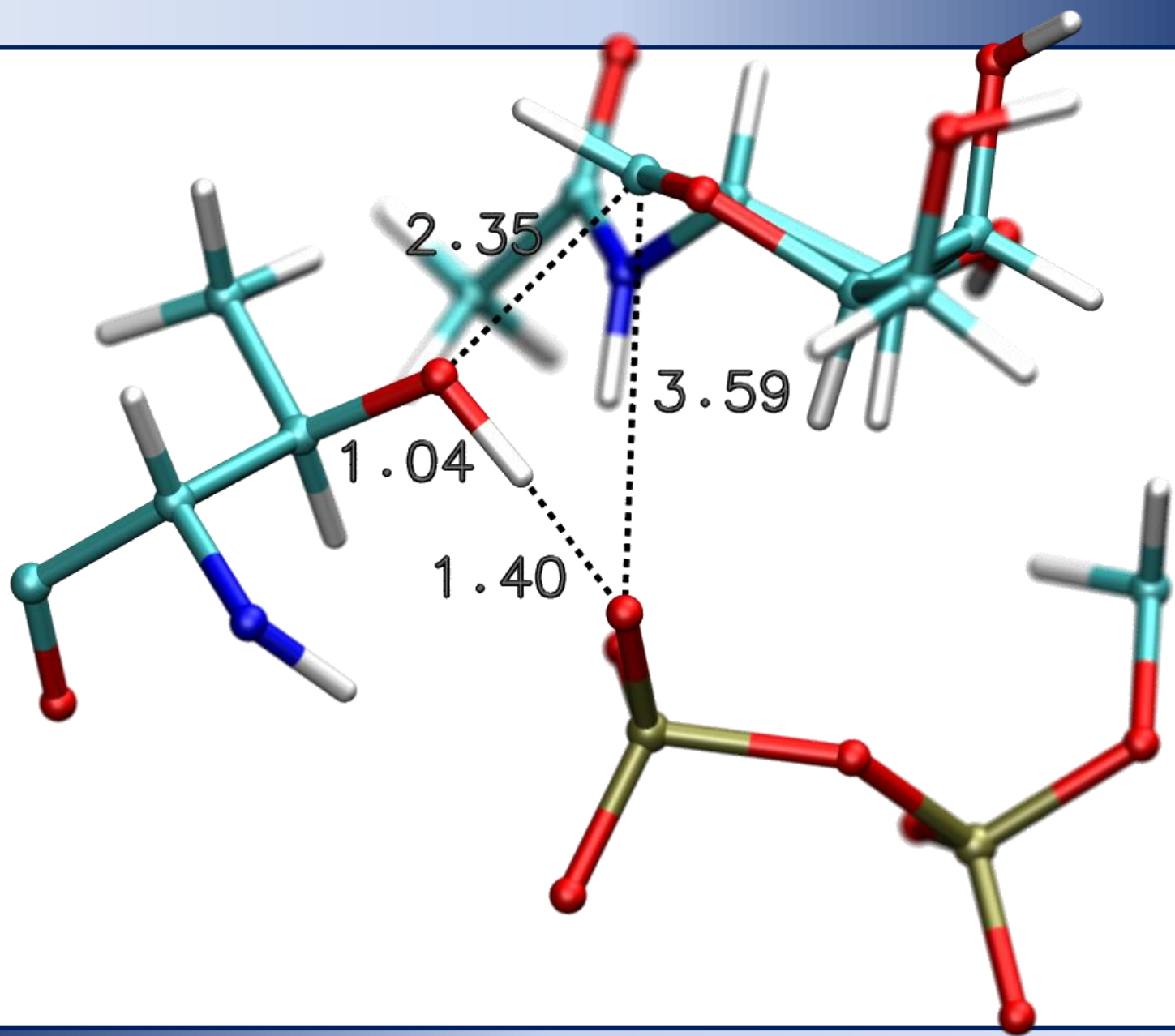


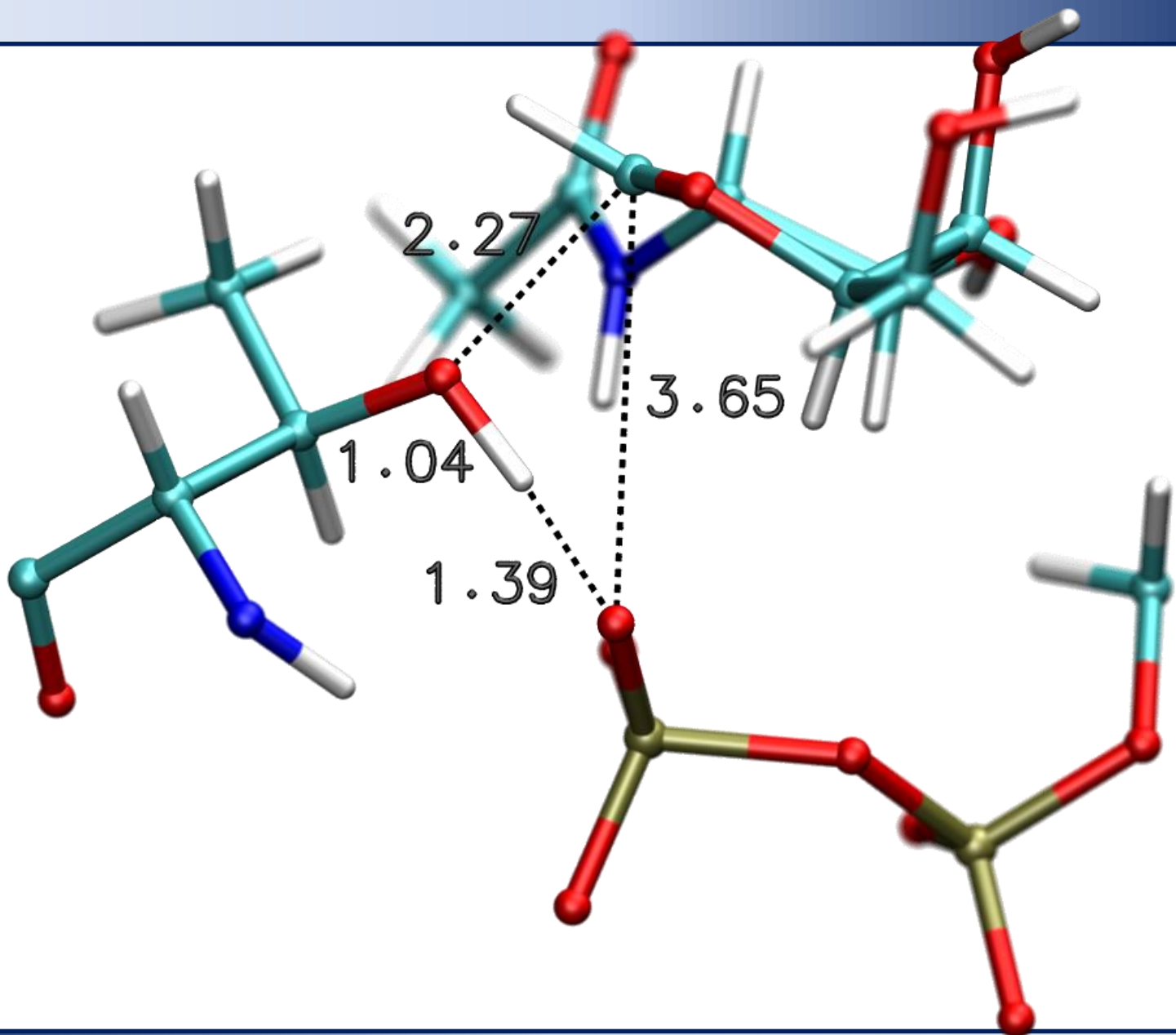


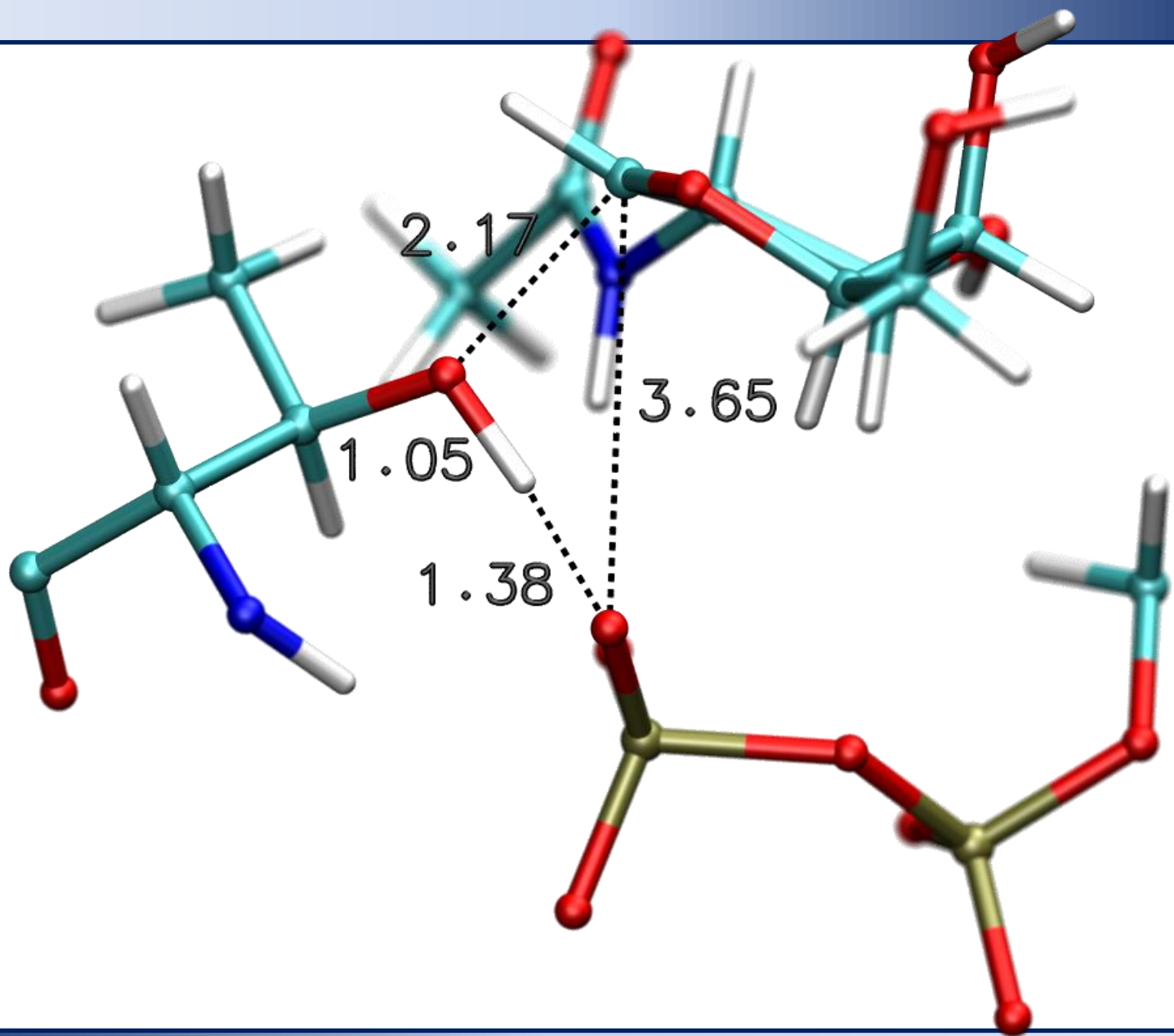


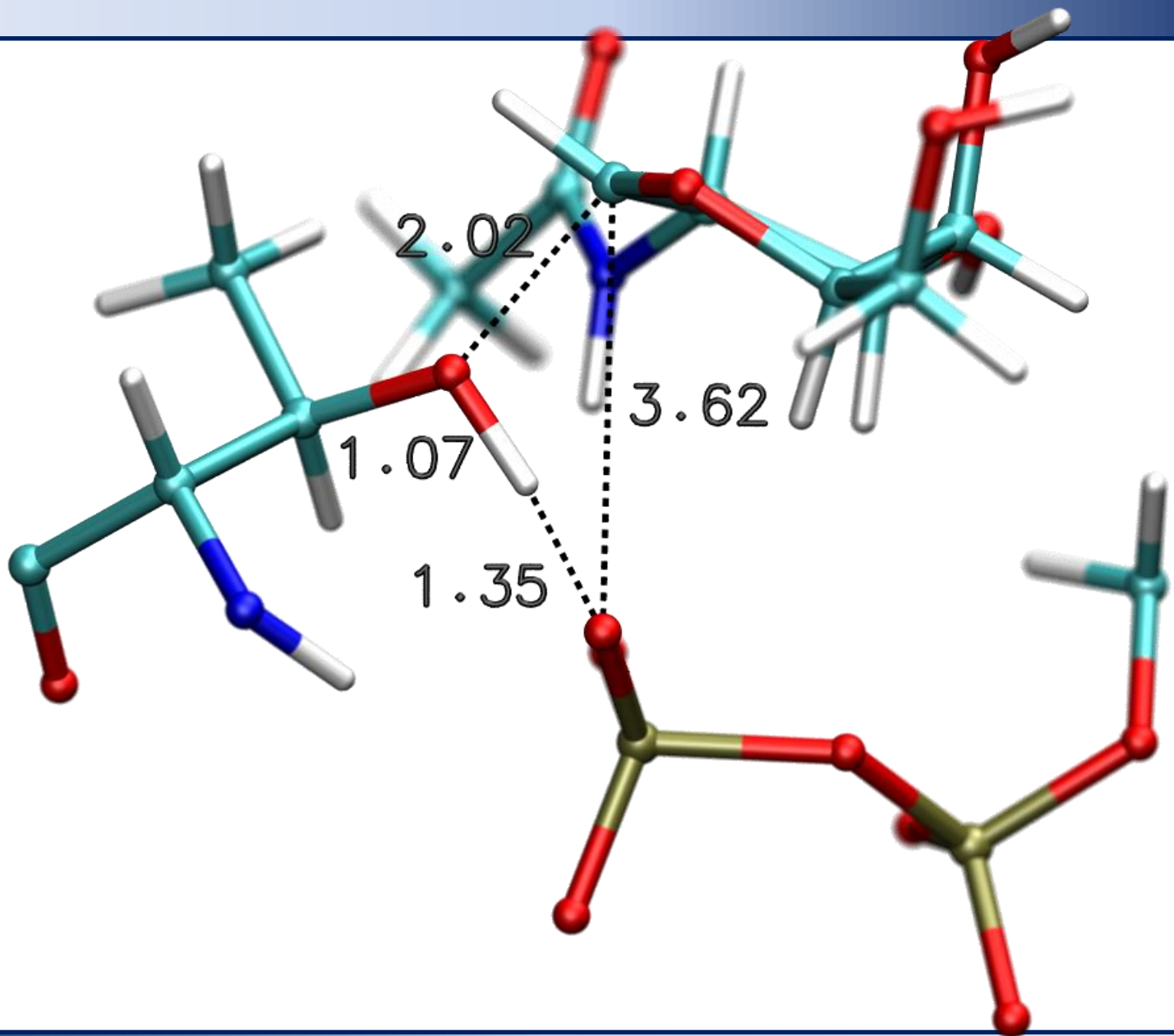


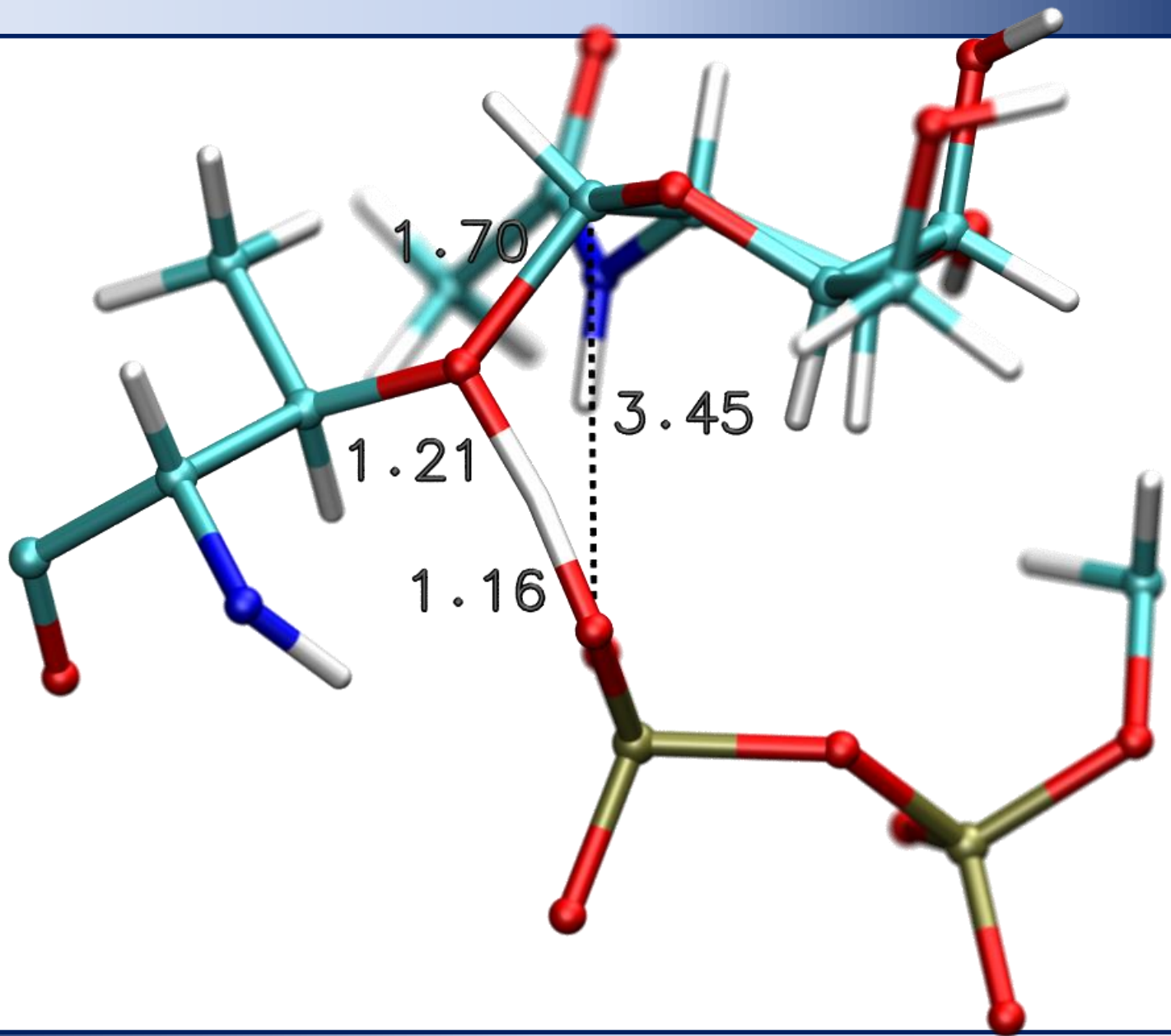


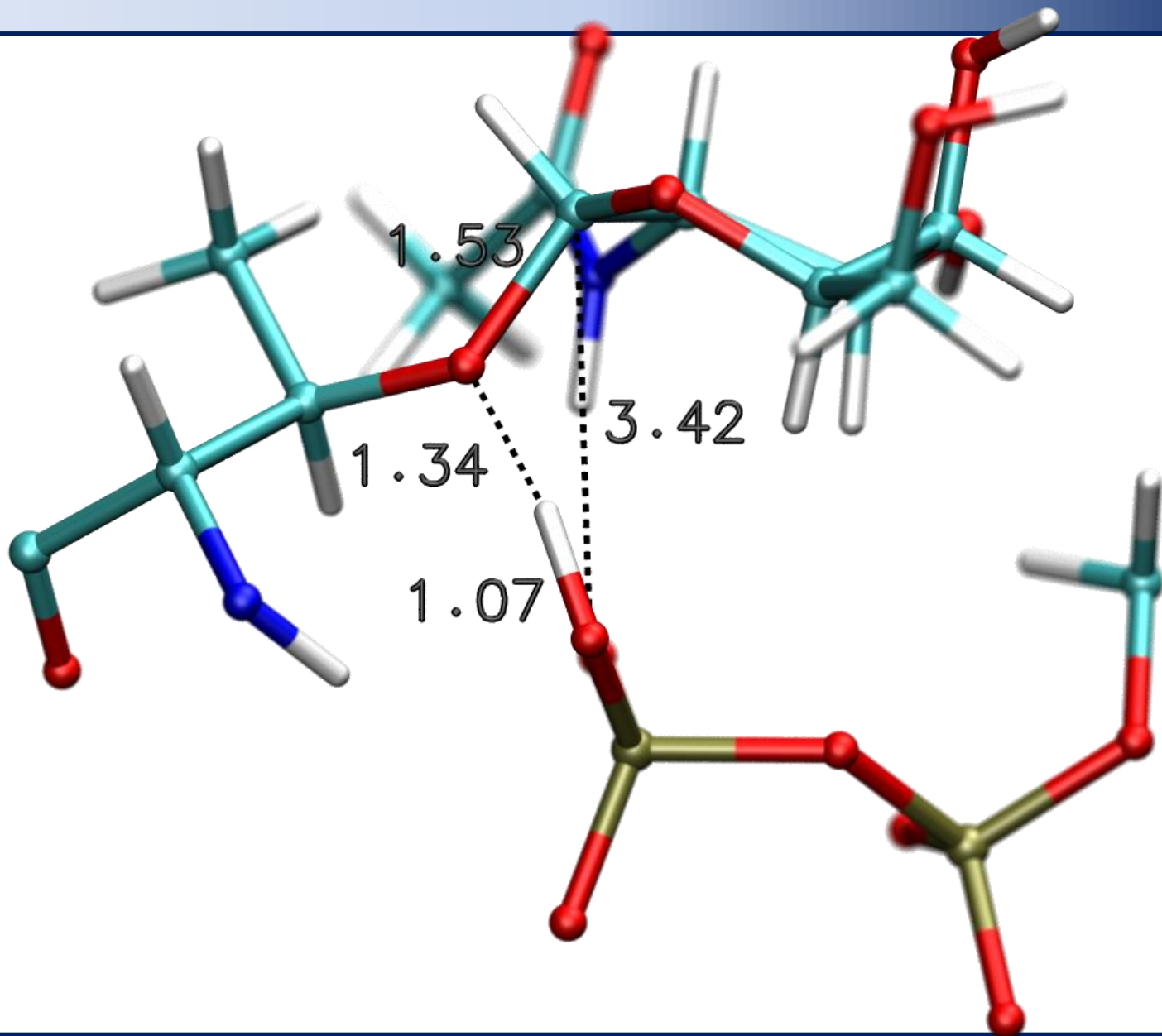


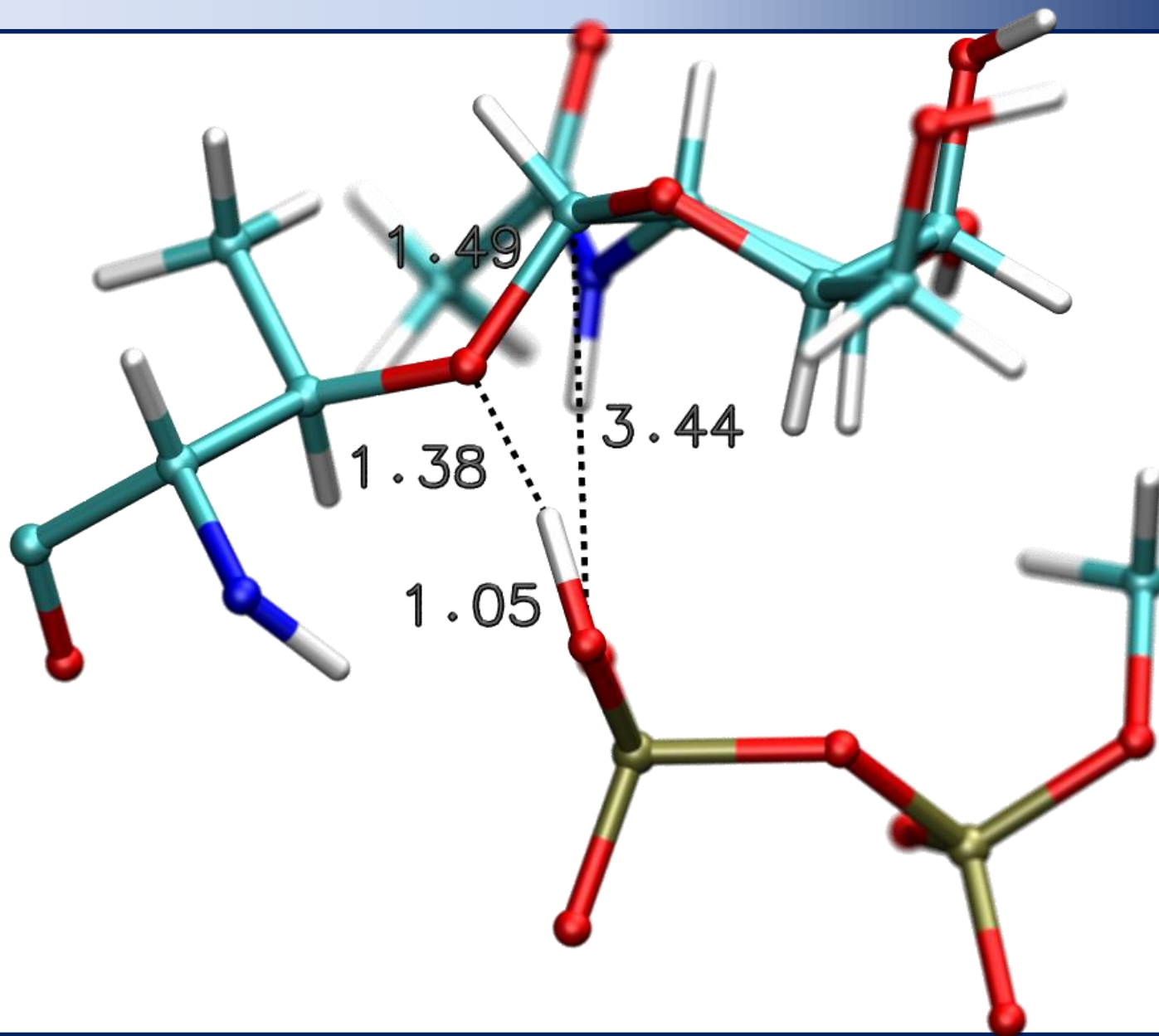


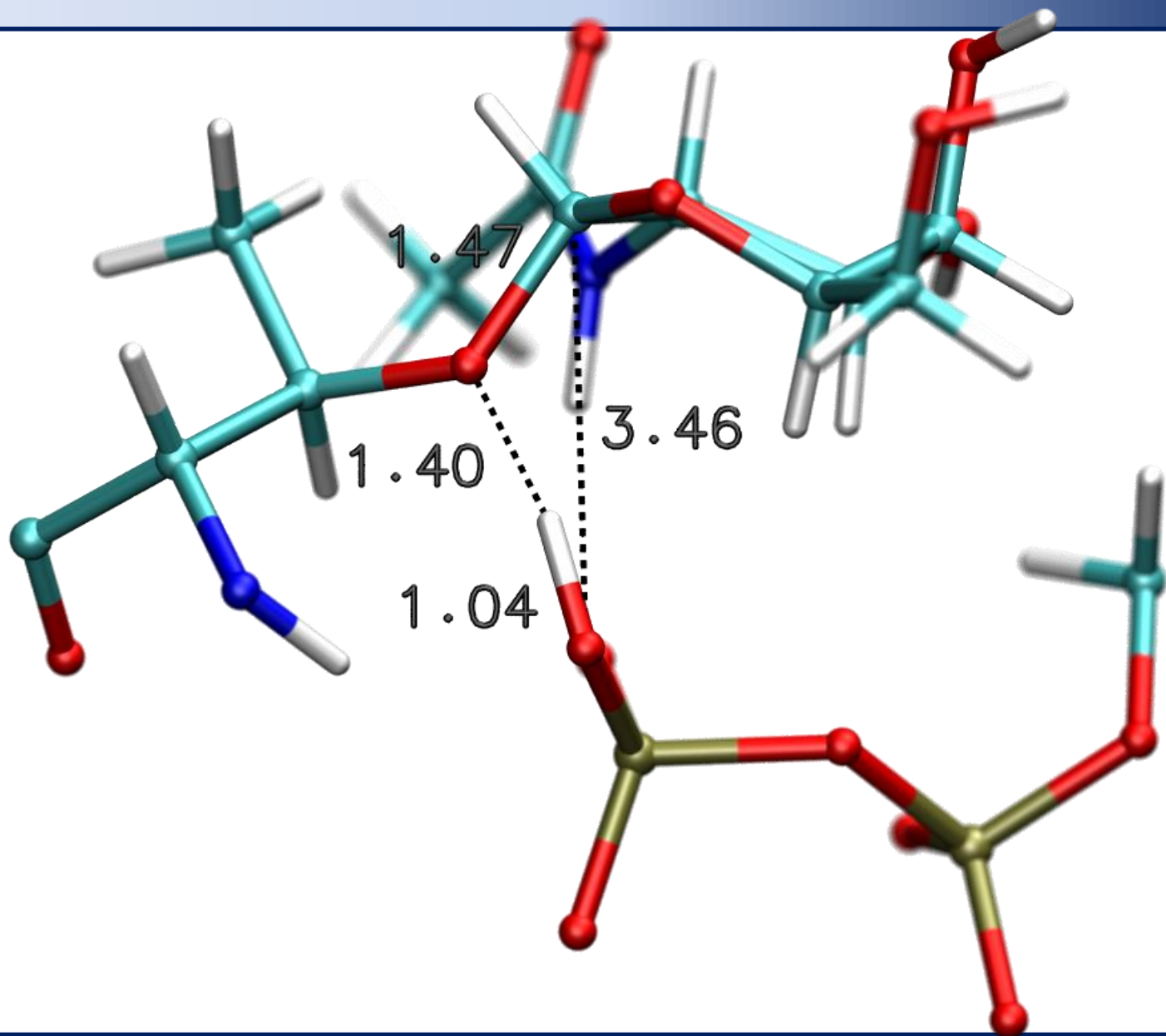


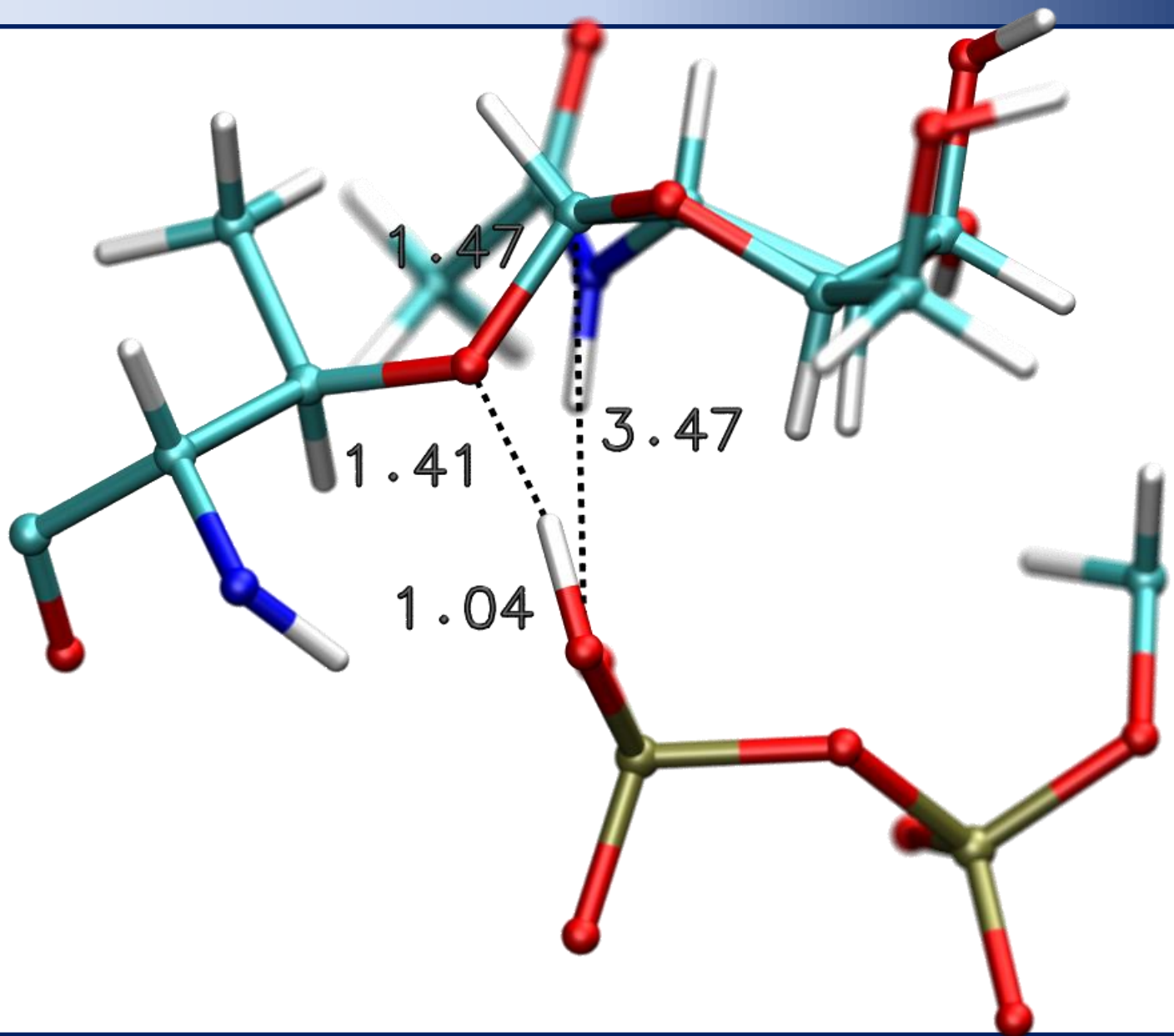


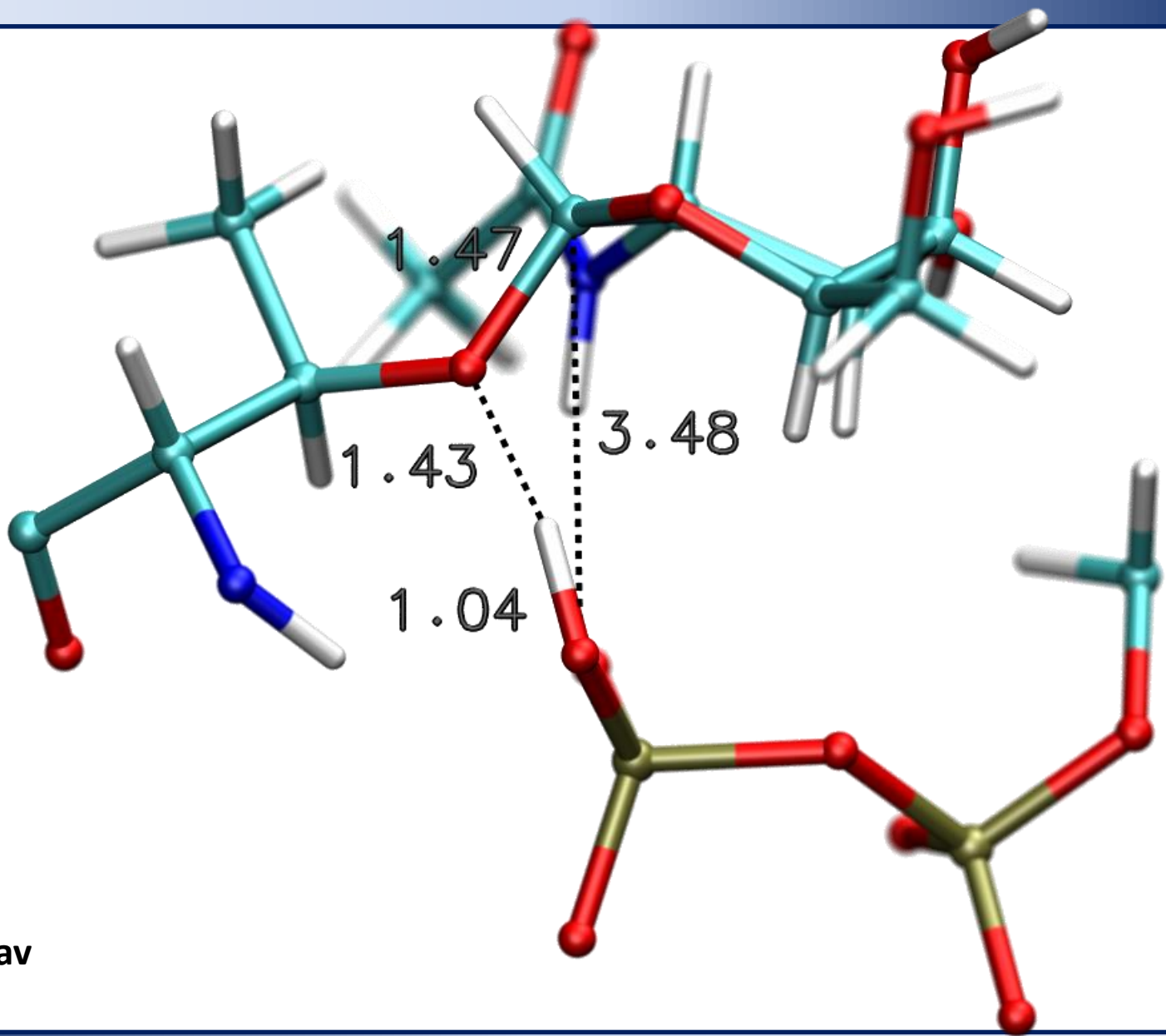






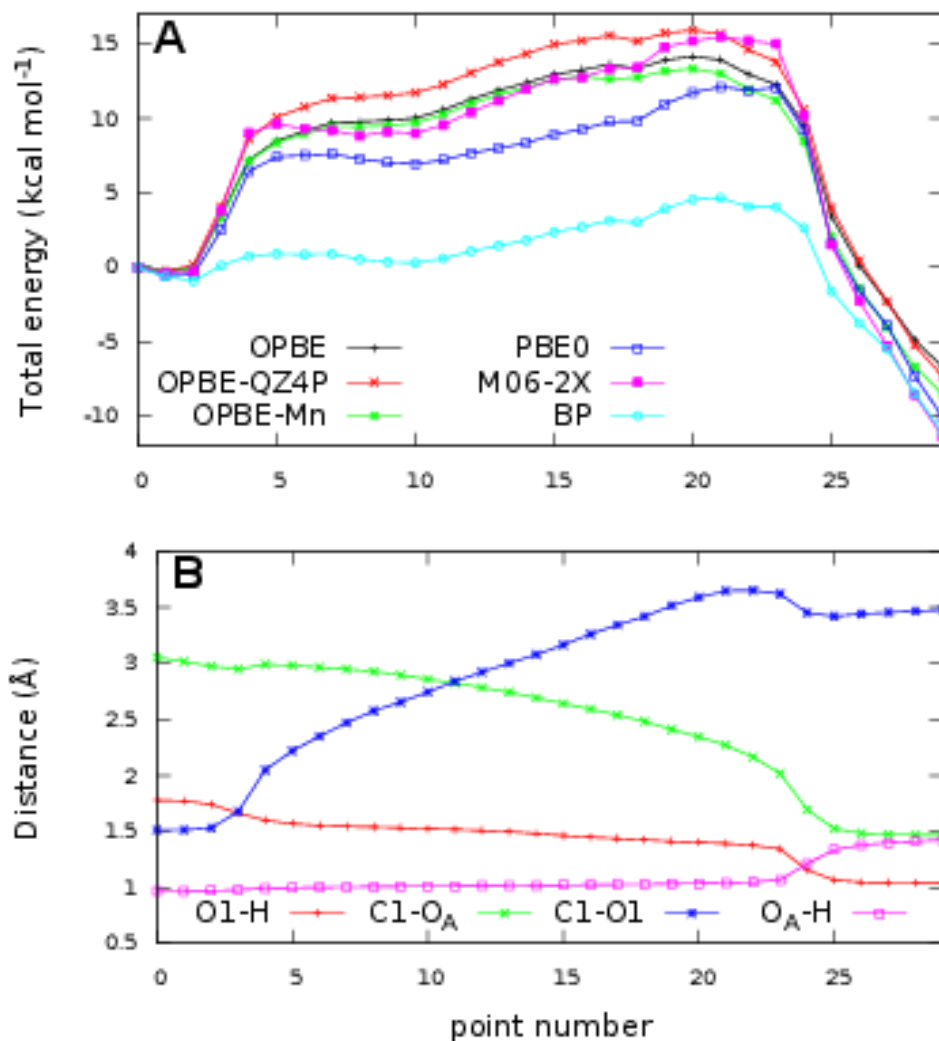






koncový stav

Výsledek



Specifika metod

Kvantově mechanické metody:

- výpočetní náročnost narůstá s požadovanou přesností výpočtu a velikostí studovaného modelu
- jedná se o výpočetně (CPU) tak i datově (RAM) náročné výpočty
- urychlení pomocí paralelního spouštění je možné, ale většinou špatně škáluje (škálování není u velmi přesných metod lineární)
- **paralelní spouštění** je vhodné spíše na SMP uzlech, při spouštění na klastrech vyžaduje rychlé datové propojení výpočetních uzlů

Molekulárně dynamické simulace (za použití molekulové mechaniky):

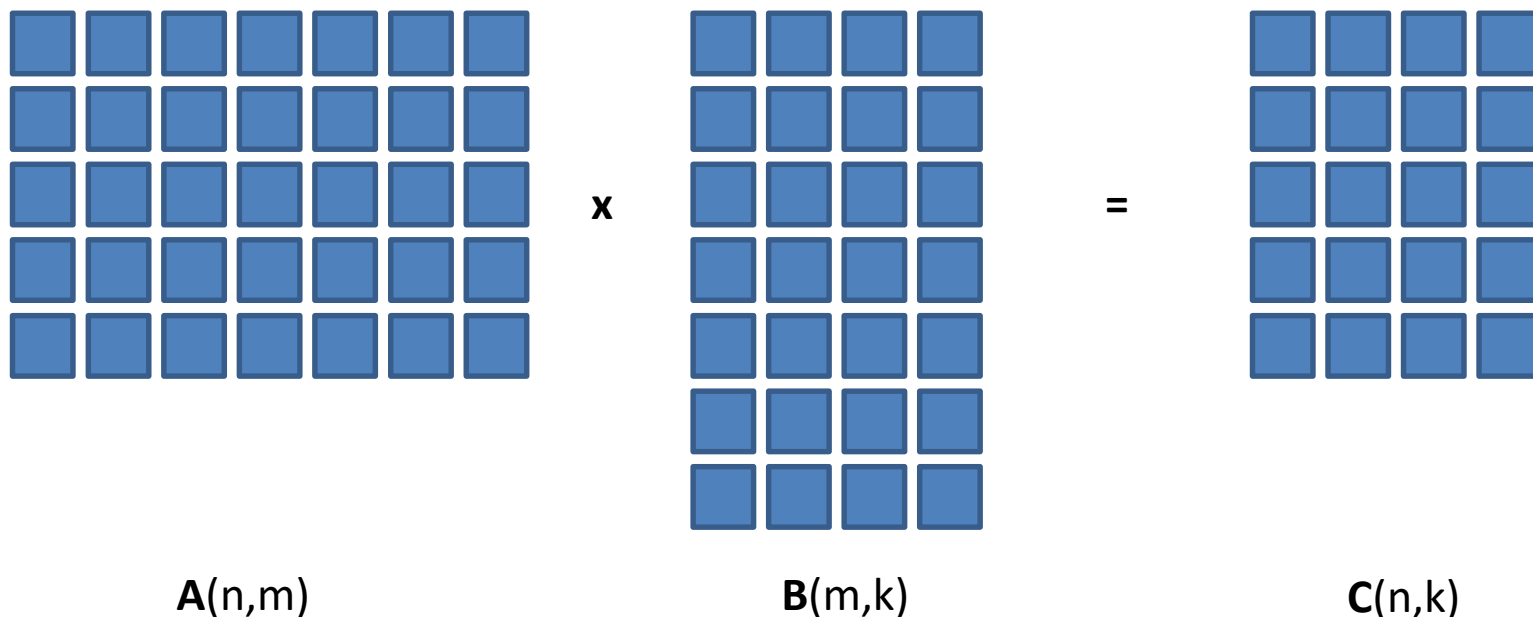
- výpočetní náročnost narůstá s velikostí modelu a požadovanou délkou vzorkování
- z důvodu nízké algoritmické složitosti je možné výpočty provádět pomocí **GPGPU**
- vytváří velké množství dat (trajektorie)
- urychlení výpočtu pomocí paralelního spouštění je snadné
- paralelizaci lze provést na několika úrovních (výpočet sil, více chodců nebo replik), pro poslední dva případy je možné dosáhnout lineárního škálování

Cvičení 1

1. Co určuje časová náročnost výpočtu $O(N)$?
2. Kolikrát se prodlouží výpočet potenciální energie molekuly benzenu kvantově chemickou metodou CCSD(T), pokud změníme použitou bázi z aug-cc-pVDZ na aug-cc-pVTZ? Počet bázových funkcí je 192 pro aug-cc-pVDZ a 414 pro aug-cc-pVTZ.
3. Pokud bude doba výpočtu potenciální energie za použití metody CCSD(T)/aug-cc-pVDZ trvat 5 hodin, jak dlouho bude trvat výpočet za použití metody CCSD(T)/aug-cc-pVTZ?
4. Reakce prvního řádu katalyzovaná enzymem má jeden rychlost určující krok s aktivační Gibbsovou energií 18 kcal/mol. Jaký je poločas reakce při 300 K?
5. Jakou délku by musela mít molekulárně dynamická simulace jednoho komplexu enzymu se substrátem z předchozího úkolu tak, abyste pozorovali přeměnu substrátu s 50 % pravděpodobností?
6. Určete počet integračních kroků, které bude nutné provést v simulaci z úkolu 5 za předpokladu, že bude integrační krok 0,125 fs (QM/MM dynamika za použití CPMD).
7. Určete strojový čas, který by bylo nutné na simulaci vynaložit, za předpokladu, že se jeden integrační krok počítá 5 sekund. Hodnotu diskutujte.
8. Určete strojový čas, který je zapotřebí vynaložit na molekulárně dynamickou simulaci fragmentu celulosy o délce 1 μ s ve vodním boxu o celkovém počtu 408609 atomů na jedné grafické kartě typu GTX 1080 za NPT podmínek? Pro řešení použijte data poskytnutá zde: <http://ambermd.org/gpus/benchmarks.htm#Benchmarks>

Modelové problémy a systémy

Násobení matic



Využití:

- hledání vlastních čísel a vektorů čtvercových matic (kvantová chemie)
- řešení soustavy lineárních rovnic (QSAR, QSPR)
- transformace (posunutí, rotace, škálování - zobrazení a grafika)

Opakování/samostudium:

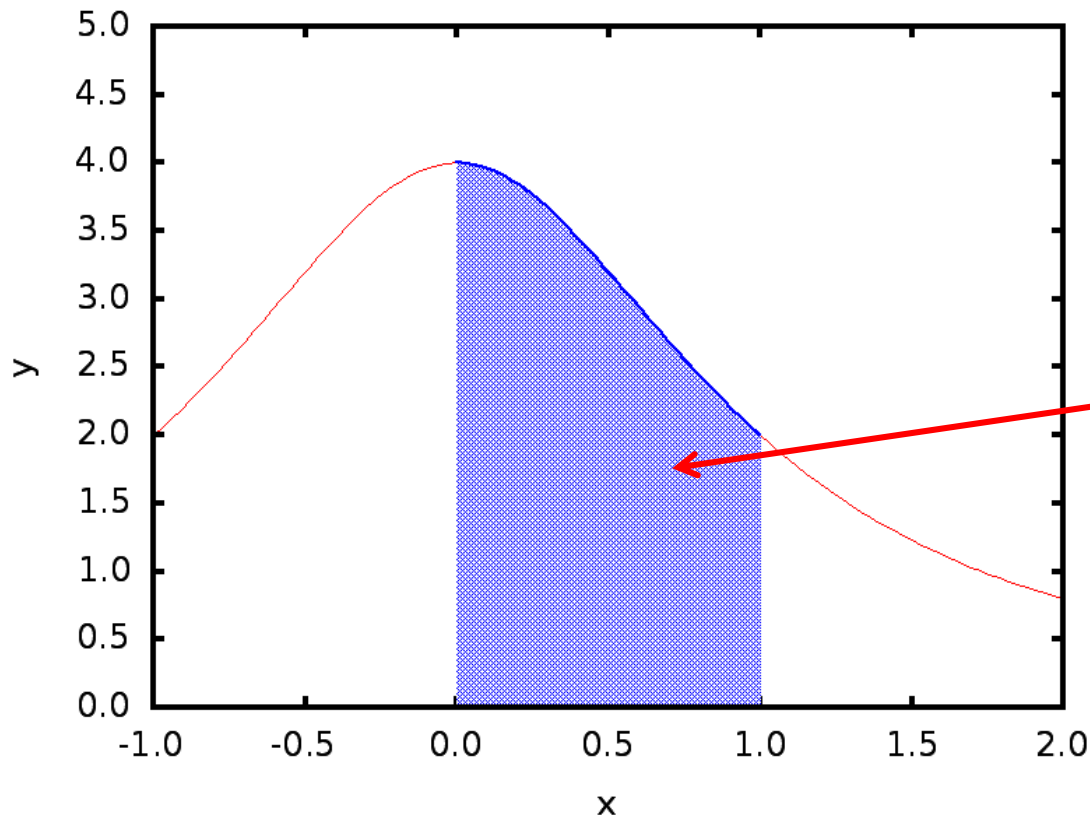
- Jak se násobení matic provádí?
- Kolik operací je nutné provést?



Numerická integrace

Výpočet určitých integrálů je možné provádět numerickými metodami, které se používají pokud:

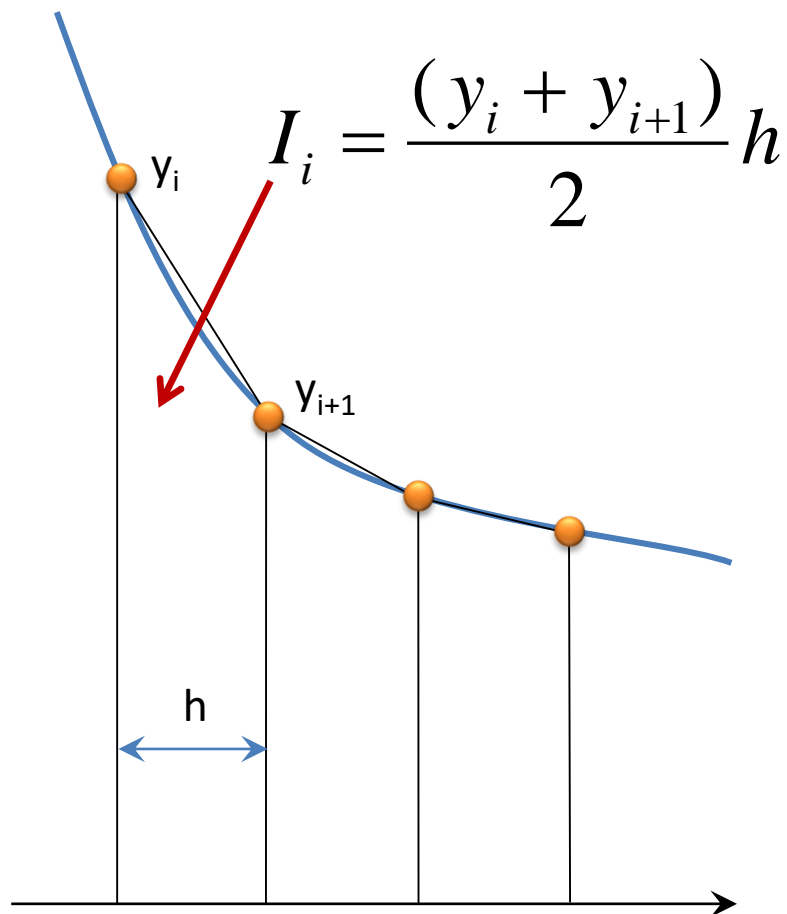
- funkci není možné integrovat analyticky
- analytická integrace je prakticky nerealizovatelná (přesnost vs výpočetní náročnost)



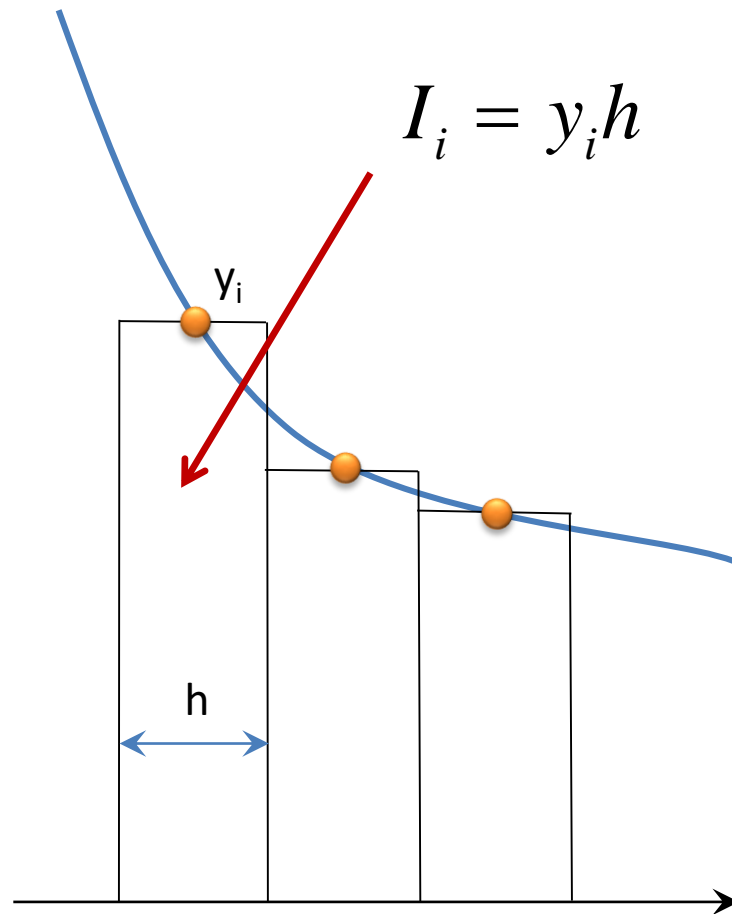
$$I = \int_0^1 \frac{4}{1+x^2} dx$$

určitý integrál je plocha pod křivkou
v rozsahu integračních mezí

Metody numerická integrace



lichoběžníková metoda



obdélníková metoda



Fulleren C_{60}

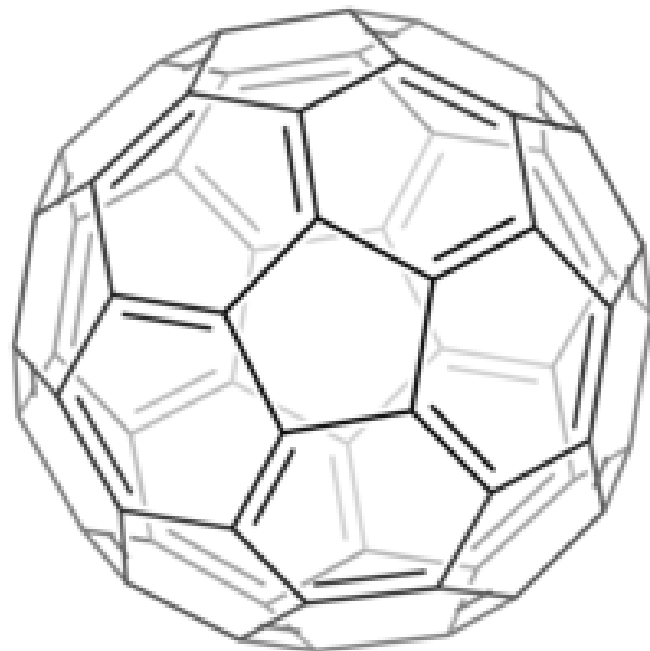
<https://en.wikipedia.org/wiki/Buckminsterfullerene>

Úkoly:

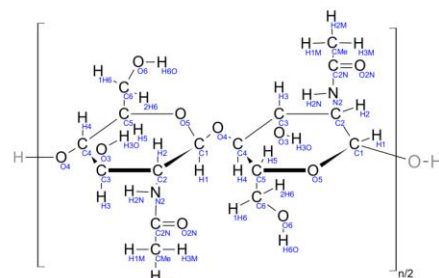
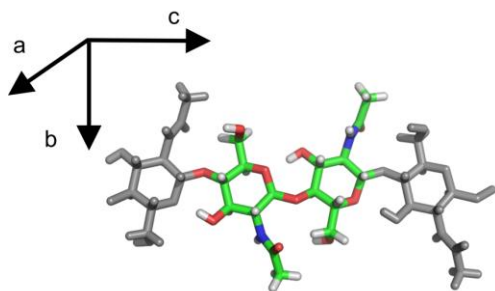
- vytvoření modelu molekuly C_{60}
- optimalizace geometrie
- výpočet molekulárních vibrací

Metody:

- semiempirická kvantově-chemická metoda PM6

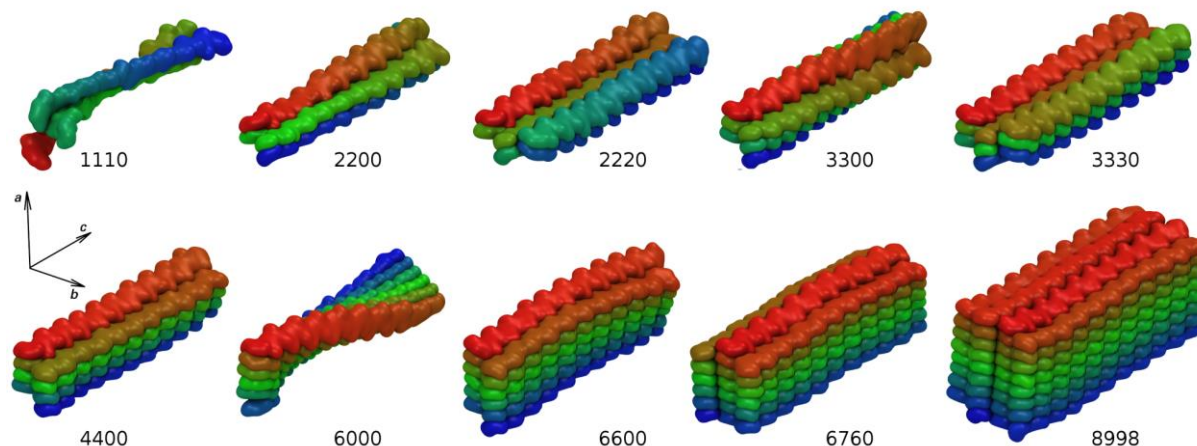


Chitinové vlákna



stavební jednotka

mechanické vlastnosti
chitinových nanovláken



Úkoly:

- MD simulace vlákna 6000

Strelcova, Z.; Kulhanek, P.; Friak, M.; Fabritius, H.-O.; Petrov, M.; Neugebauer, J.; Koca, J. The structure and dynamics of chitin nanofibrils in an aqueous environment revealed by molecular dynamics simulations. *RSC Adv.* **2016**, *6* (36), 30710–30721
DOI: [10.1039/c6ra00107f](https://doi.org/10.1039/c6ra00107f)

Souvislost s kurzem C2115

Násobení matic:

- limitující faktory související s architekturou počítačů (paměťová propustnost)
- optimalizované knihovny pro numerické výpočty (BLAS, LAPACK, Intel MKL, AMD MCL)

Numerická integrace:

- limitující faktory související s architekturou počítačů (zaokrouhlovací chyby a jejich dopad na výsledek integrace)
- paralelizace výpočtu (OpenMP versus MPI)

Fulleren C_{60} :

- spouštění výpočtů v programu Gaussian
 - v MetaCentru (PBSPro)
 - na klastru WOLF (PBSPro a Infinity)

Chitinové vlákno:

- molekulárně dynamické simulace v programu pmemd
 - škálování CPU paralelní implementace
 - srovnání běhu na CPU a GPU

Cvičení 2

Fulleren C_{60} :

1. Postavte 3D model molekuly fullerenu C_{60} a proveďte jeho optimalizaci pomocí silového pole MMFF94. Ke stavbě 3D modelu použijte strukturu ve formátu SMILES (wikipedie pro C_{60}). Výsledný model uložte ve formátu xyz. Ke stavbě použijte buď program avogadro nebo nemesis.

Chitinové vlákno:

Ekvilibrovaný model chitinového vlákna je možné nalézt v adresáři:

`/home/kulhanek/Documents/C2115/Lesson02/chitin`

topologie systému je 6000.parm7

souřadnice, rychlosti a velikost boxu je v 6000.rst7

2. Zobrazte model v programu VMD.
3. Kolik atomů model obsahuje?
4. Kolik vláken chitinu model obsahuje?
5. Jaký tvar má simulační box?

Samostudium

1. Jak se provádí násobení matic?
2. Kolik operací je zapotřebí při násobení matic provést?
3. Jaká je výpočetní komplexita násobení matic?
4. Která metoda numerické integrace je přesnější, obdélníková nebo lichoběžníková?
5. Nalezněte jiné metody numerické integrace.
6. Je možné numerickou integrací vypočítat neurčitý integrál?

