Department of Chemical Engineering and Chemical Technology

Imperial College London

Molecular Dynamics Modelling of Skin and Hair Proteins

A thesis submitted for the degree of Doctor of Philosophy

Jan Kazimierz Marzinek

ABSTRACT

The binding free energy is one of the most important and desired thermodynamic properties in simulations of biological systems. The propensity of small molecules binding to macromolecules of human bio-substrates regulates their sub-cellular disposition. This subject is fundamental in transdermal permeation and hair absorption of cosmetic actives. Biomechanical and biophysical properties of hair and skin are related to keratin as their major constituent. A key challenge lies in predicting molecular and thermodynamic basis as the result of small molecules interacting with alpha helical keratin at the molecular level. In addition, elastic properties of human skin which are directly related to the interactions of keratin intermediate filaments remain a challenging subject.

Molecular dynamics (MD) simulations provide a possibility of observing biological processes within atomistic resolution providing more detailed insight into experimental results. However, MD simulations are limited in terms of the achievable time scales. Hence, in this thesis MD simulations were employed in order to provide better understanding of the experimental results conducted in parallel and to overcome the main limiting factor of MD – the simulation time. For this purpose, thermodynamic and detailed structural basis have been delivered for small molecules interacting with keratin explaining and validating experimental data. On the top of this the fast free energy prediction tool has been built within all-atom force field by a use of steered molecular dynamics alone. Within the coarse grain approach, the force field was developed for the application of elastic properties of human skin enabling orders of magnitude faster than all-atom force fields simulations. The application of the coarser representation enabled assessing the influence of the natural moisturizing factor composed of small molecules on the elastic properties of the outermost human skin layer. In this work, MD results reached excellent agreement with the experimental data.

ACKNOWLEDGEMENTS

First of all, I would like express my sincere thanks to my supervisors, Professor Guoping Lian, Professor Athanasios Mantalaris, Professor Efstratios Pistikopoulos and Dr Massimo Noro for their support, guidance and for providing wonderful environment which led me to this stage. In particular I would like to thank Dr Yanyan Zhao for the experimental part of my work and fruitful collaboration over the whole period of my PhD. I place on record, my sincere gratitude to Dr Peter Bond and his group (University of Cambridge), Dr Alfonso De Simone (Imperial College London) and Professor Daan Frenkel (University of Cambridge) for stimulating discussions and productive collaboration. I am also very grateful to Dr Robert Farr, Dr Anna Akinshina and Dr Janhavi Raut for scientific support of my work. I would like to acknowledge MULTIMOT ITN, EC's Seventh Framework Programme [FP7/2007-2013] under Grant Agreement No 238013 for the sponsorship as well as Unilever R&D (Colworth Science Park, Sharnbrook) the hosting organization. I would also like to thank to High Performance Computing Service at Imperial College London and in particular to Matt Harvey and Terrance Crombie for technical support. Finally, I would like to express my deep thanks to my parents who were continuously providing me with the support, guidance and encouragement in my work.

DECALARATION OF ORIGINALITY

The work described within this thesis contains genuine and original work conducted by the author except where referenced. Results presented herein have not been used or submitted for the award of any degree at any other University.

COPYRIGHT DECALARATION

'The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work'

TABLE OF CONTENTS

Abstract 2

Acknowledgments 3

Declaration of Originality 4

Copyright Declaration 5

Table of contents 6

Nomenclature 9

List of Figures 15

List of Tables 23

List of Publications 25

CHAPTER I **Introduction** 27

1.1. Aims and Objectives 29

CHAPTER II **Literature Review** 33

2.1. Binding of Small Molecules 33

2.2. The Outermost Skin Layer - Stratum Corneum 36

2.3. Hair Structure 40

2.4. Keratins (Type I & II Intermediate Filaments) 44

CHAPTER III **Methodology** 49

3.1. Introduction 49

3.2. Molecular Dynamics 49

3.2.1. Steered Molecular Dynamics 55

3.2.2. Free Energy Calculations 56

3.2.3. Coarse Graining 61

3.2.3.1. Boltzmann Inversion 63

3.2.3.2. Iterative Boltzmann Inversion 64

CHAPTER IV	**Molecular and Thermodynamic Basis fot EGCG-Keratin Interaction**	67
	4.1. Introduction	67
	4.2. Objectives	68
	4.3. Methodology - Molecular Dynamics	69
	4.3.1. Computational Details	69
	4.3.2. Self-Assembly Setup	70
	4.3.3. Biased Simulation Setup	71
	4.4. Results - Molecular Dynamics	72
	4.4.1. Monomeric Self-Assembly Simulations	72
	4.4.2. Multilayer Self-Assembly Simulations	75
	4.4.3. Assessment of Biased MD Simulation Protocol	77
	4.4.4. Steered MD and Umbrella Sampling	81
	4.5. Conclusion	84
CHAPTER V	**Fe^{2+} - α-Helix Wool Keratin Complex Studies via Molecular Dynamics**	86
	5.1. Introduction	86
	5.2. Objecives	88
	5.3. Methodology - Molecular Dynamics	89
	5.3.1. Computational Details	89
	5.3.2. Simulation Setup	90
	5.4. Results - Molecular Dynamics	92
	5.4.1. Assessment of Binding Sites	92
	5.4.2. Free Energy Calculations	95
	5.5. Concluison	100
CHAPTER VI	**Ligand Binding Keratin α-Helix – Free energy via MD and SMD Simulations**	102
	6.1. Introduction	102
	6.2. Objecives	103
	6.3. Methodology - Molecular Dynamics	105
	6.3.1. Computational Details	105
	6.3.2. Simulation Setup - α-Helix - Ligand Complexes	107
	6.3.3. Simulation Setup - Steered MD amd Umbrella Sampling	108
	6.4. Results - Molecular Dynamics	109
	6.4.1. Keratin - Ligand Interactions	109
	6.4.2. The most stable protein-ligand coordinates	112

6.4.3. Steered Molecular Dynamics amd Umbrella Sampling 113
6.5. Conclusion 120

CHAPTER VII **Coarse-Grained Force Field Development and Application for Skin Keratins** 122
7.1. Introduction 122
7.2. Objecives 123
7.3. Methodology 124
7.3.1. The choice of the sequence 124
7.3.2. System Setup - All Atom Simulation 126
7.3.3. System Setup - Coarse Grain Approach 128
7.4. Results 130
7.4.1. All Atom Simulation 130
7.4.2. Coarse-Grained Force Field 135
7.4.2.1. Representation of Intermediate Filaments 140
7.4.2.2. Grafting Terminals on the Intermediate Filament Surface 142
7.4.2.3. The application of the Force Field 144
7.4.2.4. The Role of Natural Moisturizing Factor 148
7.5. Conclusion 151

CHAPTER VIII **Conclusions and Future Perspectives** 153
8.1. Overal Conclusions 153
8.2. Future Perspectives 156

Bibliography 158

Appendix I **Molecular and Thermodynamic Basis for EGCG-Keratin Interaction – Experimental Studies** 176

Appendix II **Fe^{2+} - Wool Keratin α-Helix Complex Studies via Isothermal Titration Calorimetry** 184

Appendix III **Coarse-Grained Force Field Development and Application for Stratum Corneum Keratins – Experimental Validation** 189

NOMENCLATURE

k_{φ}	Dihedral force constant
X_t	Overall concentration of ferrous gluconate in the cell
Q	Partition function
C_{unit}	Coulomb force constant
$H(\vec{r_i}, \vec{p_i})$	Hamiltonian – total potential energy
$u_i(\lambda)$	Harmonic potential - additional energy term
$V^{n=1}$	Tabulated potential at the 1st step 1
$\vec{a_i}$	Acceleration
k_{θ}	Angle force constant
k_B	Boltzmann constant
k_b	Bond force constant
V_0	Cell volume
q_i	Charge of atom i
q_j	Charge of atom j
φ	Dihedral angle
k_{φ}	Dihedral force constant
r_{ij}	Distance between atoms i and j
θ^0	Equilibrium angle
b^0	Equilibrium bond distance
φ^0	Equilibrium dihedral angle
η^0	Equilibrium distance in Urey-Bradley potential
ξ^0	Equilibrium Improper Angle
σ	Finite distance when interatomic potential is zero
$\vec{F_i}$	Force exerted on atom i
q	Independent degree of freedom
l_0	Initial coordinate of the ligand

m_i	Mass of atom i
$\vec{p}_i$	Momentum of atom i
n_{jk}	Multiplicity
N_i	Number of steps sampled in simulation
$\vec{r}_i$	Position of particle i in phase space
V^n	Potential for the current step
λ	Reaction coordinate
β	Reciprocal of the thermodynamic temperature
κ	Scaling factor
ρ_w	Specific density of water
n	Step number
n_S	Stoichiometry of binding
T	Temperature
t	Time
R	Universal gas constant
θ	Valence angle
V	Volume
g_i	Weight applied to each sampling window
ξ_{ijkl}	Angle between planes
r^{-6}	Attractive Lennard-Jones potential term
$P_i^B(\lambda)$	Biased distribution
$F_i^B(\lambda)$	Biased free energy
$H^B(\vec{r}_i, \vec{p}_i)$	Biased Hamiltonian
$V^B(\vec{r}_i)$	Biased potential
K_a	Binding constant
ΔH	Change in enthalpy
ΔS	Change in entropy
ΔG	Change in Gibbs free energy
$V(q)$	Coarse-grained potential
$V_\theta(\theta)$	Coarse-grained potential of angles

$V_b(b)$	Coarse-grained potential for bonds
$V_\varphi(\varphi)$	Coarse-grained potential for dihedral angles
η_{ijk}	Distance between atoms in Urey-Bradley potential
$P^{n=1}$	Distribution from the first step
F_i^B	Free energy associated with adding harmonic potential
$U(\vec{x})$	Harmonic potential
$H(q)$	Histogram - distribution
dv_i	Injection volume
E_{kin}	Kinetic energy
N_E	Number of the effective degrees of freedom
M_t	Overall concentration of keratin in the cell
ΔV^n	Potential difference (between the current and the reference)
ϕ_{angle}	Potential energy of angles
V_{bon}	Potential energy of bonded interactions
ϕ_{bond}	Potential energy of bonds
$\phi_{torsions}$	Potential energy of dihedral angles
$V_{Coulomb}$	Potential energy of electrostatic interactions
V_{VdW}	Potential energy of van der Waals interactions
V^{n+1}	Potential from the next step
$\phi_{impropers}$	Potential of improper dihedrals
$P(q)$	Probability distribution
V^{ref}	Reference atomistic potential
P_{ref}	Reference full atomistic distribution
r^{-12}	Repulsion Lennard-Jones potential term
K_{SMD}	Spring constant of the harmonic potential in steered molecular dynamics
ΔQ	Total heat change
$V(\vec{r_i})$	Total potential energy
$P_i^U(\lambda)$	Unbiased distribution
$A_i(\lambda)$	Unbiased free energy

$H^U(\vec{r_i}, \overrightarrow{p_i})$	Unbiased Hamiltonian
$\phi_{Urey-Bradley}$	Urey-Bradley potential
$P(q)$	Volume normalized distribution
$P_b(b)$	Volume normalized distribution of bonds
$P_\varphi(\varphi)$	Volume normalized distribution of dihedral angles
$P_\theta(\theta)$	Volume normalized distribution or angles
ΔG	Free energy difference
µg	Microgram
µm	Micrometer
ACEMD	Accelerated bio-molecular dynamics software
AFM	Atomic force microscope
AMBER	Assisted model building with energy refinement force field
BET	Brunauer–Emmett–Teller isotherm
BI	Boltzmann inversion
C	Celsius
CD	Circular dichroism
C_{egcg}	Concentration of EGCG in the syringe
C_{eq}	Free ligand concentration
CHARMM	Chemistry at Harvard Macromolecular Mechanics (force field)
CN	Coordination number
COM	Centre of mass
Cα	Central carbon atom
EM	Electron microscopy
FEP	Free energy perturbation
g	Gram
GAB	Guggenheim, Anderson and de Boer isotherm
GROMACS	Groningen Machine for Chemical Simulations (software)
HFE	Hydration free energy
HPLC	High performance liquid chromatography
IBI	Iterative Boltzmann inversion
IF	Intermediate filament
IOD	Ion-oxygen distance
ITC	Isothermal titration calorimetry
K	Kelvin
K_b	Binding coefficient
kcal	Kilocalory

kDa	Kilodalton
K_L	Multilayer binding constant
K_s	Monolayer binding constant
LIE	Linear interaction energy
LINCS	Linear constraints solver for molecular simulations
logD	Octanol-water distribution coefficient
logP	Octanol-water partition coefficient
M	Mol
MAE	Mean absolute error
mg	Milligram
min	Minute
mL	Millilitre
mm	Millimetre
mM	Millimol
MM/PBSA	Molecular mechanics Poisson-Boltzmann surface area
MW	Molecular weight
N	Number of atoms
nm	Nanometre
NMF	Natural moisturizing factor
NMR	Nuclear magnetic resonance
NPT	Constant number of atoms, pressure and temperature
ns	Nanosecond
NVT	Constant number of atoms, volume and temperature
P	Pressure
PME	Particle Mesh Ewald
PMF	Potential of mean force
ps	Picosecond
q	Amount of ligand bound per mole of protein
q_m	Amount of monolayer binding of ligand per mole of protein
R^2	Squared correlation coefficient
RDF	Radial distribution function
REMD	Replica exchange molecular dynamics
RMSF	Root mean square fluctuations
s	Second
SASA	Solvent accessible surface area
SMD	Steered molecular dynamics
US	Umbrella sampling
VOTCA	Versatile Object-oriented Toolkit for Coarse-graining Applications

WCA	Weeks Chandler Andersen
WHAM	Weighted histogram analysis method
μs	Microsecond

LIST OF FIGURES

CHAPTER II

Figure 2.1. The structure of main green tea catechins. 34

Figure 2.2. The structure of the epidermis starting from the outermost layer: Stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, stratum basale. 36

Figure 2.3. The schematic representation of the stratum corneum as the "brick and mortar model". Bricks represent corneocytes and mortar the lipid matrix. Keratin bundles are expressed inside the corneocyte. 37

Figure 2.4. Water concentration versus the skin depth investigated by Raman spectroscopy. 38

Figure 2.5. Mechanism of the elastic properties for human outermost skin layer keratin intermediate filaments hypothesised by Jokura et al.[22] based on the experimental studies. Tubes represent keratin intermediate filaments to which unstructured "heads" (N- terminals) and "tails" (-C terminals) are attached. Circles denote natural moisturizing factor (NMF). It is proposed that the lack of the NMF causes a stronger interaction between keratin bundles and their collapse – low elastic properties. With the presence of NMF the required elasticity is reached lowering the attraction between filaments. 40

Figure 2.6. The representation of keratin in the structure of the hair fiber. 41

Figure 2.7. The structure of hair follicle with epithelial/hair keratins displayed in different parts. 42

Figure 2.8. Schematic representation of keratin tripartite domain: non-helical head and tail and helical rod. 44

Figure 2.9. The structure of intermediate filament (IF). 45

Figure 2.10. The structure of α-helix (O – oxygen, N – nitrogen, C – carbon, R – amino acid side chain, dotted line represents the hydrogen bond stabilizing the secondary structure). 47

Figure 2.11. Coiled-coil conformation: A) Parallel arrangement of two helical chains, B) Axial projection. 47

CHAPTER III

Figure 3.1. The example of force versus time profile obtained from steered molecular dynamics with applied constant velocity of 0.01 nm ns^{-1} to the centre of mass of ligand (EGCG) being within protein surface (keratin). 56

Figure 3.2. Schematic representation of umbrella sampling method: red circle- protein, blue circle – ligand. Along the reaction coordinate (protein – ligand distance) independent simulations are run with additional harmonic potential (virtual spring) where ligand is restrained by its centre of mass. Each window provides with the probability of finding (histogram) according to the harmonic potential. Overlapping windows provide the input to weighted histogram analysis method which provides the potential of mean force (free energy along the distance) of protein-ligand interactions. 61

Figure 3.3. The standard "bottom up" approach for the coarse graining based on the all atom simulation. First, the full atomistic reference trajectory needs to be known. Then, by reproducing bonded and non-bonded distributions from full atomistic trajectory as well as structural and/or thermodynamic properties, the coarse grained force field is built often accompanying by the experimental input. 63

Figure 3.4. The work flow of the iterative Boltzmann inversion (IBI) method. 65

CHAPTER IV

Figure 4.1. Segments and the corresponding sequence of keratin 1 and 10 used for MD simulations: A) Schematic representation of chosen segments 1A (blue), 1B (purple), and 2A (red); B) Helical segment 2A of keratin 1; C) Coiled-coil 1B segment of keratin 1/ keratin 10; D) Helical segment 1A of keratin 1. 69

Figure 4.2. Monomeric self-assembly simulation of EGCG molecule and keratin at 298K: A) Buried hydrophobic surface area between EGCG and keratin during the simulation; B) Number of hydrogen bonds between keratin and EGCG versus simulation time. 72

Figure 4.3. Final configurations of MD simulations: A) Segment 1A (red) and its two aromatic residues (blue) stacked via aromatic attraction to two EGCG (green) rings; B) Segment 1B coiled coil (purple) and EGCG (yellow) occupying three different residues via hydrogen bonds (red dotted lines) at 308K; C) Segment 1A (blue) and EGCG (yellow) occupying three different residues via hydrogen bonds at 298K. 74

Figure 4.4. The buried surface area between keratin and EGCG at 318 K. After 47 ns EGCG molecule detached suggesting weaker interactions at higher temperatures. 75

Figure 4.5. Conformations throughout the simulation time of keratin coiled coil segment (purple) and 20 EGCG molecules (yellow) at 298 K. The formation of clusters is observed by further spreading EGCG molecules over the keratin surface as multilayer. 76

Figure 4.6. Binding of 20 EGCG molecules to keratin segment 1B at 298K: A) The number of hydrogen bonds between 20 EGCG molecules; B) The number of hydrogen bonds between keratin and 20 EGCG molecules at 298K. At the constant number of hydrogen bonds between EGCG molecules the increase in EGCG-keratin hydrogen bonds is observed. This proves the spreading of EGCG molecules over the keratin surface as a multilayer assembly. 76

Figure 4.7. Hydrophobic surface area of 20 EGCG molecules during the simulation of binding to keratin (segment 1B). EGCG molecules cluster and elongate on the keratin surface which results in decreasing their overall hydrophobic area. 77

Figure 4.8. An example of steered molecular dynamics simulation of pulling EGCG away from the keratin surface: A) Pulling force versus simulation time - Two breaking maxima correspond to hydrogen bond breakage at 0.105 ns and 0.171 ns; B) Snapshots from steered molecular dynamics simulation corresponding to the first and second maximum pulling force respectively. 78

Figure 4.9. Force versus time of pulling EGCG 4 nm away from keratin 1 (α-helical segment 1A) at 298K using: A) velocity of 1 nm ns^{-1} and spring constant of 360 kcal $mol^{-1}nm^{-2}$; B) velocity of 10 nm ns^{-1} and spring constant of 120 kcal $mol^{-1}nm^{-2}$. 79

Figure 4.10. Potential of mean force profiles obtained from windows generated by different pulling simulations of EGCG (Figure 4.9) from the same initial coordinates of keratin-EGCG complex. Nearly identical values of the free energy difference were achieved. 79

Figure 4.11. A) The convergence of the potential of mean force curves. By increasing the simulation time in each window a reasonable convergence is achieved. Hence, in all systems the umbrella sampling windows were carried out for 50 ns; B) Umbrella sampling histograms along the reaction coordinate (distance). 20 umbrella potentials with spacing of 0.2 nm and 50 ns of sampling show excellent overlap. 80

Figure 4.12. Force versus time for pulling EGCG 4 nm away from keratin coiled-coil at 308K using pulling rate of 10 nm ns^{-1} and spring constant of 120 kcal $mol^{-1}nm^{-2}$. Three very similar profiles were obtained from the same initial coordinates providing approximately the same breaking point (maximum) confirming simulation reproducibility. 80

Figure 4.13. Force versus time profiles of pulling EGCG away from different keratin segments (1B – black and green; 1A – red and orange, 2A – blue) and corresponding potential of mean force curves at: A) 298 K; B) 308 K; C) 318K. 81

Figure 4.14. The average number of hydrogen bonds between EGCG and keratin versus the binding free energy difference (at 298 K: R^2 = 0.81, at 308 K: R^2 = 0.96, at 318 K: R^2 = 0.94) as well as EGCG/keratin buried surface area (determined from the NPT equilibration prior pulling simulations) with respect to the binding free energy difference (R^2 = 0.88). 82

Figure 4.15. The example of the number of hydrogen bonds between EGCG and water molecules decreasing over the simulation time at 298 K. EGCG during the binding process expelled water molecules from the keratin surface. 83

Figure 4.16. The linear correlations between the maximum pulling force (each value obtained from 0.4 ns of steered molecular dynamics simulations) and the free energy obtained by umbrella sampling (each point corresponds to 1μs of sampling) calculations (R^2 = 0.99 in all cases). 84

CHAPTER V

Figure 5.1. Three possible types of binding site hypothesised by Jurinovich et al. 87

Figure 5.2. Examples of initial configurations used in MD simulations for the binding site assessment: Type I - Fe^{2+} interaction with backbone carbonyl oxygen (ALA67) and amide nitrogen (SER68); Type II - Fe^{2+} interacting with side chain hydroxyl oxygen (SER68) and backbone carbonyl oxygen (ASP64); Type III - Fe^{2+}coordinated by GLU71 side chain carboxylate oxygens. Cartoon representation of helix – blue. Atoms are shown as red oxygens, cyan carbons, grey hydrogens, dark blue nitrogens, and yellow Fe^{2+}. 91

Figure 5.3. The most stable MD simulation of Fe^{2+} binding to α-helix at three suggested sites type I, II and III. Type III binding site was shown to be the most stable, with the ion remaining bound to the acidic side chain carboxylate oxygen atoms for the whole 100 ns simulation. 92

Figure 5.4. Final conformations of ferrous gluconate with α-helix: A) Fe^{2+} stacked in between two glutamic acid (GLU79 and GLU82) side chain carboxylate groups, coordinated by water molecules; B) Fe^{2+} and water molecules bound to a GLU71 side chain carboxylate oxygen. Iron mediates ionic interactions with a gluconic moiety which also forms a hydrogen bond (red dotted line) with glutamic acid; Cartoon representation of helix – blue; Fe^{2+} – yellow sphere, gluconate – green; protein atoms: oxygens (red), carbons (cyan), hydrogen (grey), nitrogen (blue). 93

Figure 5.5. Fe^{2+} binding to α-helix at type II binding site. After 7 ns, Fe^{2+} moved away from type II binding site, initially located between GLU57 and THR58, to type III binding site occupying only carboxylate oxygens of GLU57, and remained there until the end of the 100 ns simulation. The distance with the carboxylate oxygen atoms remained stable throughout. Atoms are shown as red oxygens, cyan carbons, grey hydrogens, dark blue nitrogens and yellow Fe^{2+}. 94

Figure 5.6. Distances of Fe^{2+} from polar atoms of binding site of type I, II and III as a function of the simulation time. Type I and II did not show any stability in contrast to type III where in all cases iron ion remained within GLU or ASP side chain oxygens. OE1, OE2 – side chain oxygens, O – backbone oxygen, N – backbone nitrogen. 94

Figure 5.7. The assessment of different binding constants (48 – 960 kcal mol^{-1} nm^{-2}) in steered molecular dynamics simulations of pulling Fe^{2+} away from the acidic pocket at the constant pulling rate of 10 nm ns^{-1}. The force versus time profiles provided breaking points (maxima) with spring constants of 240, 480 and 960 kcal mol^{-1} nm^{-2}. Lower values of the spring strength did not allow extracting iron from keratin surface. 96

Figure 5.8. Force versus time profiles from steered molecular dynamics simulations of pulling ferrous ion away from the acidic keratin pocket with a force constant of 480 kcal mol^{-1} nm^{-2} at 1, 5 and 10 nm ns^{-1}. 96

Figure 5.9. A) Potential of mean force (PMF) curves for Fe^{2+} – α-helix polypeptide segment of keratin interaction at 298, 308 and 318 K, together with B) an example of umbrella sampling histograms at 298 K. 97

Figure 5.10. The average number of hydrogen bonds between acidic side chain oxygens and water molecules at 298, 308 and 318 K as a function of the distance of Fe^{2+} and keratin α-helix acidic binding site. Hydrogen bonds were calculated from umbrella sampling windows based on a 0.35 nm cut-off distance between donor and acceptor atoms as well as a hydrogen-donor-acceptor angle of 30°, and then averaged over the 100 ns simulation time of each window. 99

CHAPTER VI

Figure 6.1. Molecules studied herein: A) Citric acid, B) Dihydrogen citrate, C) Hydrogen citrate, D) Citrate, E) Catechin, F) Oleic acid as well as studied in Chapter IV and V G) Epigallocatechin-3-gallate and H) Iron (II) gluconate respectively. 105

Figure 6.2. Interactions of A) oleic acids (orange) with keratin (blue) at 200 ns, and B) catechin (cyan) with keratin (blue) at 200 ns. The mass concentration of ligand in both cases is 5%. 109

Figure 6.3. A) The number of hydrogen bonds per molecule between: oleic acid – oleic acid, oleic acid – keratin, catechin – catechin, catechin – keratin over the simulation time. B) Hydrophobic solvent accessible surface area (SASA) of all catechin or oleic acid molecules over the simulation time. 110

Figure 6.4. Citric acid binding to keratin at different pH values: A) pH < 3.29 – Citric acid binds to keratin via hydrogen bonds only; B) pH = 4.5 – Hydrogen- and dihydrogen citrate formed small salt aggregates via sodium ions and stacked on the keratin surface via hydrogen bonds and electrostatic interactions, C) pH > 6.39 – Citrate formed salt cluster of 26 molecules stabilized by sodium ions mediating the interactions between citrate molecules as well as citrate – keratin charged residues; D) Number of hydrogen bonds over the simulation time (averaged over 200 data points - every 1 ns). Red – oxygen, blue – nitrogen, cyan – carbon white – hydrogen), citric acid (purple), dihydrogen citrate (black), hydrogen citrate (yellow), sodium ions (green); Cartoon representation of protein – blue. 111

Figure 6.5. The protocol employed for the choice of the most stable α-helix-ligand coordinates based on the catechin as an example: A) 200 ns snapshot of the molecules which are directly bound to the keratin (green surface representation for catechins with the highest number of hydrogen bonds and red surface representation for catechins remaining catechins; white surface representation, blue cartoon and space filling: red – oxygens, blue - nitrogens, cyan - carbons, white - hydrogens for helical keratin); B) The average number of hydrogen bonds calculated over last 50 ns for all molecules being within keratin surface; C) The keratin-catechin buried surface area calculated for two molecules which exhibited the highest number of hydrogen bonds. The buried surface (green and blue) was averaged over 100 data points (black and red respectively). 112

114

Figure 6.6. The example pulling simulation of the most stable catechin (green surface representation). The pulling direction according to our previous study[16] is perpendicular to the helical backbone axis. During the SMD simulation at approximately 0.1 ns the force reaches the maximum when the last hydrogen bond is broken. The distance between the keratin helical segment and the catechin starts increasing at this point reaching 2.4 nm at the end of the simulation. Red surface representation for remaining catechins obtained by MD self-assembly simulations; white surface representation, blue cartoon and space filling: red – oxygens, blue - nitrogens, cyan - carbons, white - hydrogens for helical keratin.

Figure 6.7. The rupture force obtained from steered molecular dynamics (SMD) simulations for 7 molecules (oleic acid, catechin, citric acid, dihydrogen citrate, hydrogen citrate, citrate and ferrous ion) versus the used spring constant. Each point corresponds to the average value of 10 identical pulling simulations. The best fitted curve was obtained in each case with a power function (R^2=0.96-0.98). 116

Figure 6.8. Potential of mean force obtained from umbrella sampling method for different molecules binding to keratin helical segment. 117

Figure 6.9. Histograms of umbrella sampling windows of: A) Catechin; B) Citric acid; C) Citrate; D) Dihydrogen citrate; E) Hydrogen citrate; F) Oleic Acid. 117

Figure 6.10. A) Binding free energy (ΔG) of small neutral molecules to helical motif as a function of *logP* ; B) Binding free energy of citric acid at different protonation states; C) Protein-ligand buried area (A)as a function of the binding free energy for neutral molecules. The equation is best described by: $A = 1.3321 - 0.431\Delta G$. 115

Figure 6.10. Binding free energies (ΔG) obtained from US of small molecules binding the keratin helical motif versus the maximum pulling force averaged over 10 pulling simulations at the constant loading rate (velocity of 10 nm ns^{-1} and a spring constant of 120 kcal mol^{-1} nm^{-2}) according to the keratin-EGCG study. 118

Figure 6.11. Binding free energies (ΔG) obtained from US of small molecules binding the keratin helical motif versus the maximum pulling force averaged over 10 pulling simulations at the constant loading rate (velocity of 10 nm ns^{-1} and a spring constant of 120 kcal mol^{-1} nm^{-2}) according to the keratin-EGCG study.[134] The linear dependency is built based on EGCG, catechin, citric acid and dihydrogen citrate with R^2=0.98 ($\Delta G = -0.1224F_{Max} + 3.9346$). Ferrous ion, citrate, hydrogen citrate and oleic acid did not follow the linear correlation due to their extreme properties. 119

CHAPTER VII

Figure 7.1. The choice of the terminal surrogate: full sequence of keratin 10: N terminal (146 residues) and approximated surrogate of 60 residues with conserved basic and acidic ending as well as glycine rich motif (residues marked in: blue – basic, red – acidic, green – glycine, yellow – remaining part of surrogate, black – those omitted in the surrogate sequence). 126

Figure 7.2. The temperature annealing cycle of 10 ns – driving the system from 300 to 500K sufficient number of times enables to explore different conformations and hence the whole phase space in terms of positions and momenta. 127

Figure 7.3. Weeks Chandler Andersen (WCA) potential used for all beads in the coarse-grain model to reproduce bonded parameters within the simulation of 10 μs. 129

Figure 7.4. Initial helical structure of the 60 residual surrogate together with the obtained random coil after simulation of 80 ns in the NPT ensemble at the temperature of 800K. 130

Figure 7.5. Mean distances map of the surrogate terminal at 300K. The trajectory was divided into two parts: I – 300-1200 ns and II – 1200-2100 ns. Both parts represent similar map of mean distances as well as the whole trajectory of 300-2100 ns. First 300 ns were omitted for the system equilibration. 131

Figure 7.6. Root mean square fluctuations (RMSF) with respect to the initial random coil structure of given residue index (central carbon atom) averaged over the simulation time of part I and II at 300K. 132

Figure 7.7. Central carbon atoms end to end distance (first and last residues) distributions of part I and II full atomistic simulations at 300K. 132

Figure 7.8. Distributions of the polypeptide hydrophobic solvent accessible surface area of the part I and II at 300 K. 132

Figure 7.9. Distributions of the polypeptide gyration radius of part I and II of all atom simulation at 300 K. 133

Figure 7.10. Radial distribution functions (RDFs) between different types of beads obtained from fully atomistic simulation at 300 K. 134

Figure 7.11. Bonded distributions extracted from fully atomistic simulation at 300 K of: A) 58 angles between 3 consecutive central carbon atoms; B) 57 dihedral angles between 4 consecutive central carbon atoms. 135

Figure 7.13. The example of refining tabulated potential based on the one of the angles between 3 consecutive central carbon atoms. The raw data obtained by inverting the distribution from fully atomistic simulation is refined through the: A) interpolation and B) extrapolation. 136

Figure 7.14. Coarse-grained distributions after 10 μs of NVT simulation at 300K using Weeks Chandler Andersen potential versus distributions obtained from fully atomistic simulation of: A) 58 angles and B) 57 dihedral angles. 136

Figure 7.15. Radial distribution functions (RDFs) between 5 types of beads (polar, nonpolar, basic, acidic, glycine) obtained from: fully atomistic simulation (black) at 300 K and the coarse-grained simulation of 10 μs after 10 000 of iterations (red) using iterative Boltzmann inversion method. 138

Figure 7.16. Final tabulated potentials with the 2 nm cut-off between 5 types of beads. These potentials provided well converged radial distribution functions shown in Figure 7.15. 139

Figure 7.17. Distributions at 300K of A) 58 angles and B) 57 dihedral angles of the coarse-grained model with the 15 potentials (5 types of beads: polar, nonpolar, basic, acidic, glycine) after 10 μs of the simulation versus distributions from fully atomistic simulation. 139

Figure 7.18. Distributions of the surrogate polypeptide gyration radius: coarse-grain model (red) versus fully atomistic simulation (black). 140

Figure 7.19. The representation of the intermediate filament (12.5 nm in length) with 5 types of beads: Acidic – red, basic – blue, glycine – green, polar – grey, nonpolar – orange. 141

Figure 7.20. Representation of the intermediate filament in two axial projections along its: A) diameter and B) length. Acidic – red, basic – blue, glycine – green, polar – grey, nonpolar – orange. 144

Figure 7.21. The initial and the final conformation of the intermediate filament (blue) with grafted chains (red). After 5 ns the rapid adsorption of terminals was observed with formed loops. 145

Figure 7.22. Radial distribution functions (number of atoms away from the intermediate filament (IF) surface in two dimensions – XY components) obtained from NVT simulation at 300 K of one IF with 16 grafted terminals residues. RDFs correspond to distribution of terminals residues: A) Basic of N and C terminal; B) N1, C1, N10, C10 terminals; C) N1+N10 terminal and C1+C10 terminal; D) N+C terminals. 146

Figure 7.23. Results by Akinshina et al: Volume fractions of residues away from the intermediate filament surface: A) Basic of N and C terminal; B) N1, C1, N10, C10 terminals; C) N1+N10 terminal and C1+C10 terminal; D) N+C terminals. Results from the coarse-grain model can be compared to all atom model (AA) in this figure. 146

Figure 7.24. Initial configuration for the simulation of studying interactions between keratin filaments (glycine – green, polar – grey, nonpolar – orange, acidic – red, basic – blue). 147

Figure 7.25. Radial distribution functions (number of atoms away from the intermediate filament surface in two dimensions – XY components) obtained from NVT simulation at 300 K of terminals residues in the system of 7 intermediate filaments (each with 16 terminals). RDFs correspond to distribution of terminal residues of the central IF and its grafted chains: A) Basic of N and C terminal; B) N1+N10 terminal and C1+C10 terminal. 148

Figure 7.26. Radial distribution functions of (number of atoms away from the intermediate filament surface in two dimensions – XY components) obtained from NVT simulation at 300 K of all terminals residues in the system of 1 intermediate filament with 16 terminals attached: black – without NMF, red – with addition of acidic and basic beads of NMF, green – with addition of extra NMF polar beads, blue – with addition of extra NMF nonpolar beads, yellow – with full composition of NMF. 142

LIST OF TABLES

CHAPTER II

Table 2.1. The moisture content of hair at different relative humidity. 43

Table 2.2. The nomenclature of known mammalian keratins. 46

CHAPTER IV

Table 4.1. Details of self-assembly simulations for keratin-EGCG interactions. 71

Table 4.2. The contribution of the average number of hydrogen bonds between EGCG and acidic residues as well as occupied amino acids. 74

CHAPTER V

Table 5.1. Thermodynamic parameters for ferrous gluconate binding to keratin at different temperatures of 298, 308 and 318 K obtained by isothermal titration calorimetry which is described in Appendix II. Data obtained by fitting the isothermal titration calorimetry data to equation (A.II.1) and (A.II.2). Results from molecular dynamics simulations (ΔG_{MD}) are also provided. Mean absolute error (MAE) of the free energy between experiment and molecular simulations is below 3.1%. 98

CHAPTER VI

Table 6.1. A) Small molecules evaluated in this study with varying octanol/water ($logP$) partition coefficient.
B) Charged species evaluated in this study: negatively charged protonation states (depending on the pH) of citric acid and ferrous ion. Citric acid pKa values (pK_{a1}=3.29, pK_{a2}=4.77, pK_{a3}=6.39). 106

Table 6.2. Details of MD simulations for keratin segment 1A and small molecules with their 5% mass concentration. The pH is set according to pK_a values provided in the reference. 107

Table 6.3. Keratin α-helix-ligand complex: Binding free energies from umbrella sampling (ΔG_{MD}); binding free energies validated experimentally (for EGCG and Fe^{2+}) and experimentally obtained with other proteins (ΔG_{Exp}); buried surface area of uncharged molecules; rupture force obtained as the average value over 10 identical pulling simulations at the constant loading rate (velocity of 10 nm ns^{-1} and spring constant of 120 kcal mol^{-1} nm^{-2}) according to the keratin-EGCG study. 120

CHAPTER VII

Table 7.1.	The amino acid composition of the most abundant stratum corneum keratin pair: KRT1/KRT10 in terms of charged amino acids and the content of glycine.	125
Table 7.2.	Amino acid composition of the real intermediate filament (50 nm in length) made of KRT1/KRT10 pair and its approximated model (1/4 in length) used in simulations.	141
Table 7.3.	The composition of natural moisturizing factor mixture in terms of acidic, basic, polar, nonpolar, glycine and water mass concentration.	149

LIST OF PUBLICATIONS

Journal Publications

1. Marzinek JK, De Simone A, Frenkel D, Bond PJ, Harvey M, Noro MG, Lian G, Mantalaris A, Pistikopoulos EN. Coarse-Grained Force Field Development and Application for Interactions between Keratin Intermediate Filaments. To be submitted.

2. Marzinek JK, Bond PJ, Lian G, Zhao Y, Noro MG, Mantalaris A, Pistikopoulos EN, Han L. Free Energy Predictions of Ligand Binding to an α-Helix using Steered Molecular Dynamics and Umbrella Sampling Simulations. Journal of Chemical Information and Modeling, 2014, doi: 10.1021/ci500164q.

3. Zhao Y[+], Marzinek JK[+], Bond PJ, Chen L, Qing Li, Mantalaris A, Pistikopoulos EN, Noro MG, Han L, Lian G. A Study on Ferrous Ion Binding to Keratin Using Isothermal Titration Calorimetry and Molecular Dynamics Simulation. Journal of Pharmaceutical Sciences, 2014, 103: 1224-1232, doi: 10.1002/jps.23895. [+]Equal contributors.

4. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG, Han L, Chen L. Molecular and Thermodynamic Basis for EGCG-Keratin Interaction - Part I: Molecular Dynamics Simulations. AIChE Journal, 2013, 59: 4816–4823. doi: 10.1002/aic.14220. Manuscript published as "Best Paper Presentation".

5. Zhao Y, Marzinek JK, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro M, Han L. Molecular and Thermodynamic Basis for EGCG - Keratin Interaction – Part II: Experimental Investigations. AIChE Journal, 2013. 59: 4824–4827. doi: 10.1002/aic.14221. Manuscript published as "Best Paper Presentation".

Selected Conference Presentations and Posters:

1. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN. Molecular Dynamics Modelling of Skin and Hair Proteins – Review. Manuscript and presentation at the MULTIMOD ITN workshop, Lyon, France, November 2013.

2. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG. Ligand Binding Free Energy Predictions Based on single Steered Molecular Dynamics for Hydrophobic and Hydrogen Bonding governed Interactions with α-Helix. Presentation and abstract at the American Institute of Chemical Engineers (AICHE), 2013 Annual Meeting in San Francisco, USA, November 2013.

3. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG. Virtual Microscope – Small Molecules onto Keratin Fibres. Presentation at the "Strictly Science" meeting – Unilever R&D, Colworth Science Park, UK, September 2013.

4. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN. The Assessment of Iron (II) Binding Sites onto Wool Keratin via Molecular Dynamics. Presentation and manuscript. MULTIMOD ITN; Khalkidhiki, Greece, May 2013.

5. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG. Small Molecules onto Keratin Fibres - Molecular Dynamics and Experimental Validation. Invited presentation the Internal UCC Symposium - Unilever Lecture Theatre, Department of Chemistry, University of Cambridge, Cambridge, UK, March 2013.

6. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG, Han L. Molecular Dynamics and Isothermal Titration Calorimetry Study on Catechin Binding to Keratin. Abstract and presentation at the 2nd Annual CCP-BioSim Conference: Frontiers of Biomolecular Simulation, University of Nottingham, Nottingham, UK, March 2013.

7. Lian G, Marzinek JK, Raut J, Farr R, Zhao Y, Wang L, Yang L. Integrative Multiscale Modelling of Materials and Processes. Poster presentation, Science Fair, Unilever R&D - Colworth Science Park, Sharnbrook, UK, December 2012.

8. Marzinek JK, Lian G, Mantalaris A, Pistikopoulos EN. The Investigation of Green Tea Catechin Binding to Keratins by Molecular Dynamics Experimental Validation- Abstract and presentation at the annual meeting of American Institute of Chemical Engineers 2012, Pittsburgh, USA, November 2012.

9. Marzinek JK, Lian G, Mantalaris A, Pistikopoulos EN. Binding Free Energy Calculations of EGCG with Stratum Corneum Keratins by Molecular Dynamics Simulations. Presentation and manuscript. MULTIMOD ITN: 2nd Workshop Proceedings, Italy, Milan, October 2012.

10. Lian G, Marzinek JK, Zhao Y. Molecular and Thermodynamic Study on Plant Active Binding to Proteins. Unilever Science Fair - Poster and oral presentation, Colworth Science Park, Sharnbrook, UK, April 2011.

Awards

American Institute of Chemical Engineers (AICHE), 2012 Annual Meeting, Pittsburgh, USA. “Best Presentation” award in the session: “Industrial applications of Molecular Simulations” and invitation to the AICHE Journal as the “Best Paper”. The abstract and presentation: Marzinek JK, Lian G, Mantalaris A, Pistikopoulos E.N. The Investigation of Green Tea Catechin Binding to Keratins by Molecular Dynamics and Experimental Validation.

CHAPTER I

Introduction

Nowadays, computational chemistry plays a key role in predicting thermodynamic and structural properties in petroleum, chemical and pharmaceutical industry. This originates from the increasing costs of experimental work as well as often extreme and unreachable conditions at the laboratorial level. In addition, experimental methods cannot deliver detailed insight into atomistic behaviour provided by molecular simulations. Recently, molecular simulations have been widely used in all industrial branches. The significance of molecular simulations is thus rapidly expanding due to the increasing industrial needs as well as fast development of computational resources. The latter one though is still a limiting factor for molecular simulations in terms of reachable time and length scales. However, the 2013 Nobel Prize in Chemistry was awarded "for the development of multiscale models for complex chemical systems" reaching excellent agreement with experimental results.

Molecular dynamics (MD) is a computational method which enables to observe the movement of atoms with atomistic details. It is an excellent tool for studying e.g. new pharmaceuticals targeting macromolecules in the drug discovery branch.[1] The detailed binding site assessment, conformation changes as well as strength of the interactions can be assessed within this method. However, MD has its limitations which are the achievable time scales and thus the size of the studied systems. The propensity of small molecules (ligands) binding macromolecules (e.g. proteins) is quantified by the binding free energy difference directly related to the binding constant. The binding constant determines the equilibrium concentration of free ligand, protein and protein-ligand complex. The possible pharmacological and/or toxic effect depends on the free energy of binding. Recently, within MD simulation framework there has been an increasing number of studies focused on the binding free energy predictions of small molecules to proteins often by accompanying experimental validation.[2-4] Although a great agreement between a simulation and experiments was gained in many cases,

free energy methods within MD simulations using explicit solvent are still computationally expensive. The challenge still lies in obtaining binding free energy in reasonable time scales.

Apart from detailed observations and lower costs in comparison to experiments the application of MD arises from the limitations of laboratorial methods. For instance isothermal titration calorimetry (ITC), the most commonly used method in protein-lignand systems can not be applied for very low or high affinity interactions - the equilibrium association constant can be accurately measured only at the range of $10^2 – 10^9$ M-1.[135] ITC requires high concentration of reactants which can lead to the incorrect capturing of the heat capacity in low affinity interactions and to e.g. aggregation or non-availability of the compounds in high affinity interactions.[174] MD simulations by overcoming these problems serve a promising tool in drug discovery branch.

The binding affinity of small molecules to macromolecule is fundamental in transdermal permeation as well as absorption to human hair of small active molecules. For instance, the binding free energy of ligands to the outer skin layer protein regulates their concentration in lower layers where most of the vital profits are delivered. In particular, there is an increasing interest in understanding the binding of small ligands used in personal care products. Natural green tea extract with a high content of polyphenols were previously shown to have many benefits onto human skin and hair: anti-oxidant[6], anti-tumor[7], hairloss prevention and hair growth stimulation[8] as well as photoprotective[9] properties. Iron ion used in hair care products has been attributed to act a mordant which has capability to stabilize protein-ligand interactions[10-12] and in particular natural dyestuff with the main protein of human hair providing a permanent colour.[13,14] Other molecules with experimentally observed benefits in personal care products involve citric acid[15] and oleic acid.[16] Stratum corneum (SC), the outermost skin layer of the epidermis constitutes the main barrier for skin permeation. The main protein of SC is known to be keratin. Keratin accounts of over 80% [17] of the outermost skin layer dry mass and 65 to 95% [18] of hair. Keratin is the main structural protein responsible for the appearance and physicochemical properties of skin and hair. It is mainly composed of alpha helical structure which is the most common motif in globular proteins.[19] The strength of the interactions as well as structural basis (binding sites and conformational changes) between small actives and the most prevalent helical motif part of keratin is a key subject in understanding their role and remain unknown.

Elastic properties of human skin are directly related to the interactions of keratin intermediate filaments (IFs). Keratin IFs are composed by pairing acidic and basics keratin chains providing a 50 nm in length and 8 nm

in diameter bundle.[20,21] Natural moisturizing factor (NMF) composed of free amino acids and ions was previously shown to have beneficial influence on the skin elasticity and its moisturizing properties.[22]

Due to the system size, MD simulations involving interactions between IFs with application of NMF solution (small molecules) using atomistic force fields with explicit solvent have unreachable time scales. A coarse-graining approach has capability to overcome this issue and can be applied in order to develop a specific case force field. In coarse-graining, specified group of atoms is represented as a one sphere which significantly improves the simulation performance. Hence, the challenge lies in building and applying a coarse-grained force field with the above implementation.

1.1. Aims and Objectives

The aim of this thesis is to employ MD simulations in order to investigate the thermodynamic and conformational changes of small active molecules with varying properties binding to the main biosubstrate protein of human skin and hair (keratin). In this work it is aimed to explain in more details and to validate experimental work conducted at the same time by Dr Yanyan Zhao (China Agricultural University) who studied keratin-small molecules interactions. In addition, the objective was to assess the observed fast free energy protocol observed for keratin-EGCG system for small molecules with varying properties binding the most prevalent secondary structure motif (α-helix) of keratin. On the top of this the aim was to develop and apply the coarse-grain force field for the outermost skin layer (SC) which would enable orders of magnitude faster than all-atom force fields simulations and predictions of skin elastic properties. The experimental work conducted by Dr Alfonso De Simone (Imperial College London) was used to confirm the observed (within coarse-grain and atomistic MD) secondary structure of the studied surrogate polypeptide.

In particular, the objectives of this thesis within MD have been defined:

1. **Investigate molecular and thermodynamic basis for keratin-EGCG interactions validating experimental data (Chapter IV)**

   The objective of this chapter is to study interactions of the main active green tea polyphenol EGCG (epigallocatechin-3-gallate) with skin keratin using MD simulation framework and to provide detailed insight into experimental results. Dr Yanyan Zhao (China Agricultural University) within ultrafiltration method observed the multilayer assembly of EGCG molecules on the keratin surface. Using isothermal titration calorimetry (ITC) the binding free energy at 298K was provided. All experimental results are presented in Appendix I. Hence, the main challenge of this chapter was to validate multilayer binding of EGCG as well as the free energy using MD. With this information the more detailed insight in comparison to experiment was aimed: the assessment of the type of interactions involved as well as residues with the highest binding affinity. Although in the free energy calculations the only one reaction coordinate (distance perpendicular to the helical axis) was studied a good agreement between experiment and simulations was gained. MD model involves fast free energy predictions for keratin-EGCG system using pulling simulations alone – the linear correlation of the rupture force obtained from steered molecular dynamics (SMD) versus the binding free energy.

2. **Investigate Fe^{2+} – wool keratin complex validating experimental ITC study (Chapter V)**

   The main challenge of this part is to provide more detailed atomistic and thermodynamic insight into the experimental data for the ferrous ion interacting with wool keratin. The experimental results obtained by Dr Yanyan Zhao using ITC (presented in Appendix II) were best fitted by the "one set of identical site model". In addition, the binding free energies at three different temperatures as well as significant entropic contribution were obtained. Thus the aim of this chapter was to assess and define the binding sites involved in this interaction, validate free energies as well as explain the entropic contribution. In MD simulations the nonbonded model with Lennard-Jones potential for iron (II) ion was tested.

It appeared that the simplest nonbonded model provided a surprising great agreement between experimental free energies and those obtained by simulations. In addition, MD simulations explained the "one set of binding" site model used to interpret the experimental data as well as the significant entropic contribution in terms of displacement of water molecules from the binding pocket.

3. **Build a fast free energy method for small molecule – α-helix interactions (Chapter VI)**

The aim of this chapter was to test the linear correlation of the rupture force versus the free energy observed for keratin-EGCG system (Chapter IV) for other small molecules with significantly varying properties (octanol/water partition coefficient and protonation states), binding the helical motif of keratin. In addition, it was aimed to assess the type of interactions involved, provide structural basis and compare obtained free energies to existing experimental data. Using steered molecular dynamics (SMD) with conserved pulling conditions (loading rates, reaction coordinate) according to keratin-EGCG system, it was observed that the rupture force versus the free energy correlation works for 4 out of 8 molecules studied herein. It was shown that molecules which did not follow the correlation have properties which significantly vary from EGCG i.e. have overall charge greater than ±1 (citrate, hydrogen citrate, Fe^{2+}) or they octanol/water partition coefficient is very high. Although the model does not involve many molecules, this protocol provides a promising methodology for fast free energy predictions with future implementation for other systems.

4. **Build and apply a coarse grain force field for elastic properties of human skin (Chapter VII)**

The objective of this chapter was to build and apply the coarse-grain force field for interactions between keratin intermediate filaments which are directly related to the elastic properties of human skin. Within the coarse grained representation it was aimed to reproduce explicit solvent atomistic simulation by the use of "water free" force field with tabulated non-bonded and bonded potentials which will significantly save the computational time.

In addition the objective was to assess the influence of the natural moisturizing factor (NMF) on the interactions between keratin intermediate filaments (IFs). The coarse-grain model involving water within tabulated potentials based on the all atom simulation in explicit solvent was developed. The force field was applied validating existing experimental data. The secondary structure of the keratin terminal surrogate was experimentally validated by Dr Alfonso De Simone using circular dichroism (CD). Experimental results are presented in Appendix III. In addition the influence of NMF on the elastic properties of skin was assessed.

Literature Review

The relevant literature of this thesis is reviewed within this chapter. First, the importance of the binding affinity followed by the benefits of small molecules interacting with proteins is presented. Subsequently, the anatomy of human skin and hair as well as the role of their main protein – keratin is extensively depicted.

2.1. Binding of Small Molecules

The propensity of small molecules (ligands) binding macromolecules (e.g. proteins) is quantified by the binding free energy difference ΔG directly related to the binding constant K_a:

$$\Delta G = -RT \ln K_a \tag{2.1}$$

Where, R is the universal gas constant and T denotes the temperature. The binding constant determines the equilibrium concentration of free ligand $[L]$, protein $[P]$ and protein-ligand $[PL]$ complex:

$$K_a = \frac{[PL]}{[P][L]} \tag{2.2}$$

With known equilibrium concentration of all species it is possible to predict the required concentration of the applied compound. Hence, the free energy plays a key role in pharmaceutical industry and drug discovery branch. It is also fundamental in the intestinal absorption[23], transdermal permeation[17,24-28] as well as skin and hair care applications.[29,30]

Natural ingredients (plant actives) have recently gained a wide popularity due to their nearly non-toxic properties as well as large number of attributed benefits, such as anti-oxidant, anti-tumor and anti-

inflammatory.[30-41] For instance, some of those molecules involve polyphenols[42], iron compounds[10-12], citric acid[15] and oleic acid.[16]

Polyphenols are found in food, plants and beverages. Tea is one of the most widely consumed beverages in the world which is abundant in catechins (polyphenols). Green tea catechins have been reported to have antioxidant[6,43-46] and antimutagenic[47-52] properties. Studies have also detected an association between applications of polyphenols and decreased cancer risk.[53-56] The anti-tumor effect of green tea has also been shown.[7,57,58] Catechins are often used in skin personal-care products acting as powerful antioxidants which are able to protect skin from ultraviolet damage.[59,60] Catechins are believed to be able to penetrate skin quickly and give hydration benefit without a greasy feel.[42] Green tea polyphenols have also been attributed to skin anti-ageing properties[61] which are associated with eliminating fine lines, scars and pigmentation and the appearance of pores.[62,63] Polyphenols can reduce cellulite and are excellent as anti-bacteria agents for treating pimples and acne.[64] Furthermore, they are able to prevent foot and body from odour, due to their anti-fungal property.[65] Tea extracts are also used in hair care products. They are able to treat dyed, damaged and chemically treated hair providing them with a shine, softness and smoothness.[8] It was previously reported that tea catechins have hairloss prevention property and are able to stimulate the hair growth.[8, 66] Tea polyphenols consists of four main catechins: EC, ECG, EGC and EGCG which are presented in Figure 2.1.

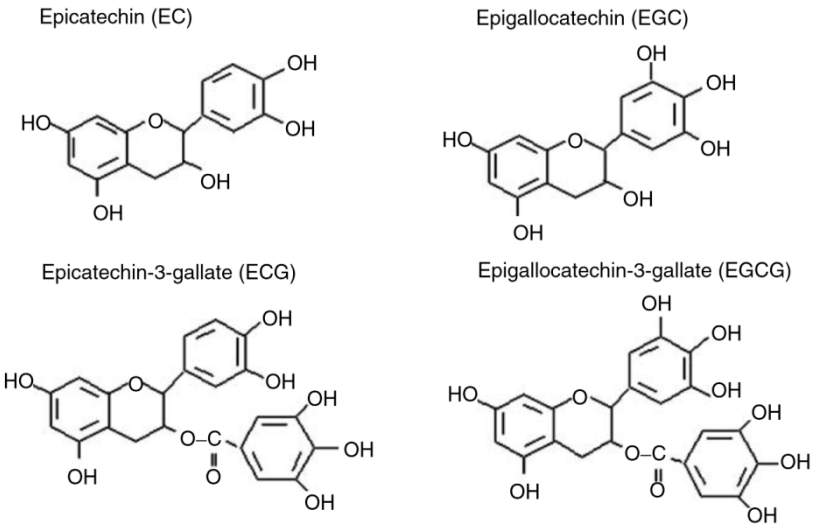


Figure 2.1. The structure of main green tea catechins.[67]

The major constituent (50-80% of all tea catechins) and the most active polyphenol of green tea is epigallocatechin-3-gallate (EGCG).[68] The eight phenolic groups of EGCG can serve hydrogen bond to biomolecules. The octanol-water partition-coefficient *logP* = 2.08[69] of EGCG may suggest its hydrophobic interactions with macromolecules. However, EGCG is a hydrophilic molecule being poorly soluble in lipid systems.[175] Previously, EGCG has been shown to bind to salivary proline-rich proteins[70,71], fibronectin and fibrinogen and histidine-rich glycoproteins[72], and more recently to protein such as the 67 kDa laminin receptor.[73] Results have also shown that EGCG has a high binding affinity to type III Intermediate Filament – Vimentin.[74] It has been previously investigated that the hydrophobic interaction and hydrogen bonds are crucial in binding mechanism of EGCG to Amyloid β-Peptide.[75] Recently, Zhao et al.[23] showed experimentally that EGCG binds to gastric mucin by a combination of hydrophobic interactions and hydrogen bonding. In their study, multilayer binding of polyphenol was observed by ultrafiltration. Their isothermal titration calorimetry (ITC) data provided the binding affinity decrease with the temperature increase.

Iron ion plays a fundamental role in many biological processes: dioxygen transport, electron transfer, oxidation or in stabilization of pharmaceuticals with protein complexes.[10-12] In hair and textile dyeing industry iron plays a significant role as a mordant providing hair with a more permanent colour.[13,14] Many proteins have capability to uptake, transport, storage and export the iron.[76-79] High water soluble iron compounds e.g. ferrous sulphate or ferrous gluconate, were shown to have higher bioavailability than poorly water soluble ones and hence they are often used instead.[76] Ferrous ion was experimentally reported to be involved in binding to *Escherichia coli* ferritin[80] and yeast frataxin.[81] Previous ITC study on Fe^{2+} binding to recombinant human H-Chain ferritin[82] demonstrated large positive entropy changes. The free energy obtained by their ITC measurements was fitted by one set of identical binding site model obtaining the value of -7.09 kcal mol^{-1} at 25°C and pH 6.5. Jurinovich et al.[83] using quantum mechanical – molecular mechanical (QM/MM) reported the possibility of three different interactions between ferrous ion and an -helical wool keratin peptide and side chain functional groups were reported to be the most stable conformation. However, this QM/MM simulation was based upon an ideal protein environment and the solvent effects were neglected. Both the binding thermodynamic properties and structural binding site assessment between keratin a-helix and ferrous ion remained unknown.

2.2. The Outermost Skin Layer - Stratum Corneum

Human skin acts as a major protection for our body. It is the biggest organ of the integumentary system and has the area of approximately 1.6 – 1.8 m^2.[84] The main function of the skin is to retain water inside the body and protect it from external substances. The epidermis is the outer layer of the skin which builds up the cutis together with the dermis. The structure of epidermis is presented in Figure 2.2. It is made of 5 main layers starting from the outermost: Stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, stratum basale.

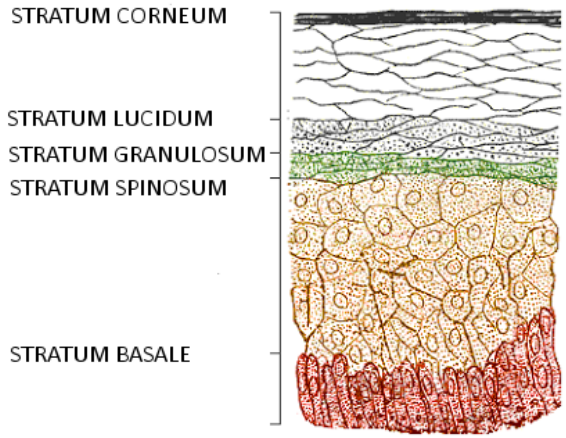


Figure 2.2. The structure of the epidermis starting from the outermost layer: Stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, stratum basale.[85]

Benefits which were attributed to small molecules are related to the lower layers of human skin. Hence, the transdermal delivery of drugs and skincare applications depend on the biology and structure of the outermost skin layer – stratum corneum (SC). The strength of the interactions between topically applied substances and the SC thus regulates their concentration in deeper parts of the skin.

The principal role of SC is controlling the barrier function against invasion of foreign substances and the ability of the skin to retain the water.[86] SC is composed of dead, flattened layers of corneocytes and intracellular lamellae. Corneocytes constitute the final product of epidermal keratinocytes which is continuously being renewed. Throughout this lifecycle cells vary in their shape and composition until the migration to the outer layers of the epidermis where they fall away. Approximately twenty days are needed for the migration and maturation of keratinocytes from the basal layer to form the corneocytes in the stratum corneum.[87] Corneocytes are about 1 µm thick with 50 µm diameter.[25] The results showed that healthy

corneocytes are anucleated and form a series between 10 to 20 layers (depending on the site in the body). The major constituent of the SC is protein - keratin. Keratinisation or cornification is the process of converting the material of epidermal cell into keratin. Keratin intermediate filaments (IFs) constitute approximately 85-90% of the dry mass of the stratum corneum.[26,27] The organization of the keratin intermediate filaments plays the most important role as a barrier property of skin which include holding of water, skin appearance and elasticity.[88] The results show that keratins fill the corneocyte cytoplasm forming keratin bundles aggregating in a parallel pattern.

In Figure 2.3 stratum corneum is shown as a bricks-and-mortar model representation where corneocytes are arranged like bricks in a lipid matrix which forms the mortar phase.[27,28] The most common keratin pair expressed in SC is keratin 1/keratin10 (KRT1/KRT10).[89,90]

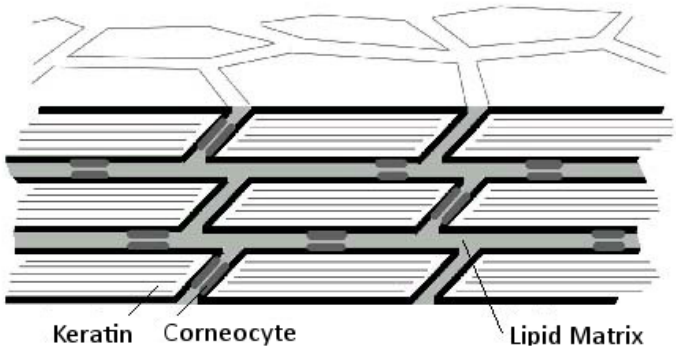


Figure 2.3. The schematic representation of the stratum corneum as the "brick and mortar model".[91] Bricks represent corneocytes and mortar the lipid matrix. Keratin bundles are expressed inside the corneocyte.

Corneocytes are protected by surrounding cornified envelope which is the thin insoluble layer consisting of many proteins (e.g. loricrin, involucrin and small proline-rich proteins) which are cross-linked to each other.[88] The cornified envelope acts as a barrier against diffusion of external substances. Therefore, the binding affinity of substances to the keratin intermediate filaments which are inside the corneocyte can be decreased. The lipid channel (lipid matrix) which fills the extra-cellular spaces between corneocytes consists of the three main components: ceramide sphingolipids, cholesterol and free fatty acid.[86] The lipid matrix constitutes approximately 10-15 % dry mass of the SC.[26,27]

The pH of normal skin is acidic and varies from 4.2 to 5.6.[92] The acidic property of the SC results from the enzymes hydrolysing the phospholipids of membranes which generate free fatty acids.[93] The acidic mantle

(sebum and perspiration which are on the skin surface) protects the skins from damage, sun, wind, the colonization of bacteria and constitutes moisture barrier via the absorption of water by amino acids and salt.[92] The main component of the skin is water. Water concentration in the stratum corneum plays a significant role in the barrier property of skin as well as the skin appearance. This hydration state constitutes one of the crucial factors for SC as a barrier function. Water as a natural component of SC affects its plasticity and acts as a penetration enhancer.[94] The information on the gradient of water content in the outermost layer of the skin is a crucial factor in the transdermal delivery of small molecules. "Apart from lipids and protein, water is another major substance in stratum corneum and is secondary to keratinized corneocytes in mass".[27] Water content along the SC changes from approximately 30% at the skin surface to approximately 70% at the innermost layer.[95] The average water concentration in vivo in SC (normal hydration level) is approximately 30%. However, in vitro it may vary from 45 to 70%.

The results of Raman spectroscopy showed the sigmoidal shape of the water content versus the skin depth (Figure 2.4). [95,96]

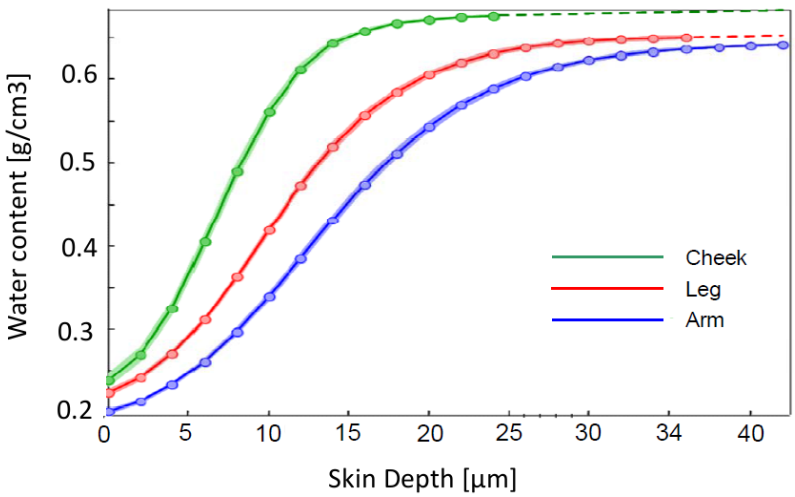


Figure 2.4. Water concentration versus the skin depth investigated by Raman spectroscopy.[96]

Keratin intermediate filaments present within corneocytes are embedded in a "broth" which is a mixture of water, free amino acids, ions and other non-ionic compounds. This composition is often called "Natural Moisturizing Factor" (NMF) which plays a crucial role in skin properties in terms of the transdermal permeation, elasticity and moisture content.[97-100] The decrease in the NMF concentration results in the dry and

itchy skin which may cause further disorders and diseases.[101] For instance, the applications of water based moisturizers causes the soak of SC layers obtaining illusion of moisturized skin.

That causes NMF to be diluted and evaporation of water becomes much higher. Water without the presence of NMF has no capacity for providing elastic properties for human skin. The effect of the NMF in the SC has been previously studied by Jokura et al. [22] using nuclear magnetic resonance (NMR) and electron microscopy. It has been shown that in the absence of NMF keratin IFs associate stronger and the interactions between keratin bundles are increased. In addition, the mobility of IF was reduced and the increase of water content did not result in better elasticity without the presence of NMF. The greater mobility, the better SC elastic properties as well as decreased attractive forces between keratin bundles were achieved with application of the solution composed of free amino acids. The effect of different type of amino acids was assessed: neutral to basic (glycine and lysine) provided with excellent SC elasticity improvement and acidic residues were shown to have much lower capacity for the SC elasticity restoration. Those results showed that the role of NMF is to prevent the attractive forces between keratin IFs providing the SC layer with elastic properties as well moisturized healthy skin. The mechanism hypothesised by those experimental studies is presented in Figure 2.5. It is observed that without the presence of water extractable materials (NMF) there is a great association between keratin IFs.

Akinshina et al.[102] developed a model based on the self consistent field theory (SCF) for the interactions between keratin IFs in the SC (using the most abundant keratin pair: keratin 1 and 10). By representing IFs as charged surfaces with grafted unstructured keratin heads and tails it was claimed that terminal chains have capacity of mediating weak attraction between IFs by electrostatics and bridging. The addition of NMF (free amino acids) caused the lower attraction between keratin bundles. Results validated experimental studies by Jokura et al. and were represented as volume fraction of given types of amino acids (polar, nonpolar, acidic, basic and glycine) away from the keratin IF surface.

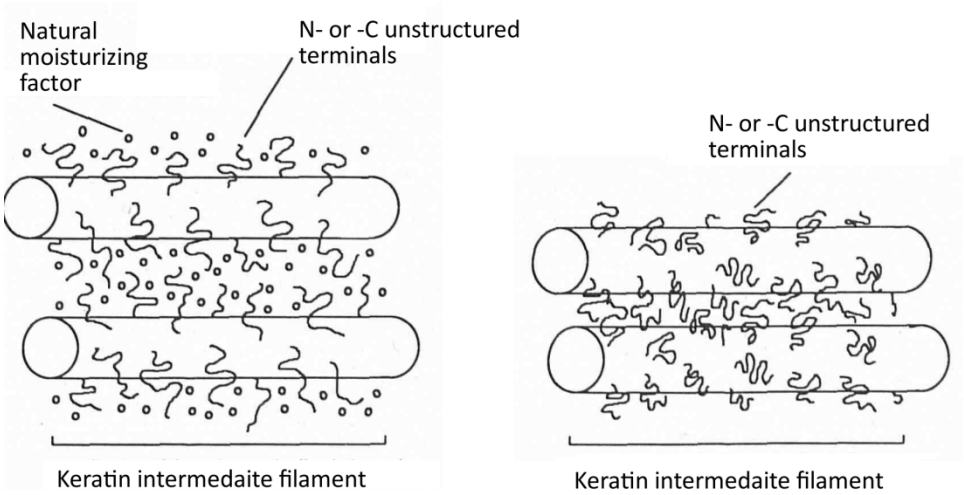


Figure 2.5. Mechanism of the elastic properties for human outermost skin layer keratin intermediate filaments hypothesised by Jokura et al.[22] based on the experimental studies. Tubes represent keratin intermediate filaments to which unstructured "heads" (N- terminals) and "tails" (-C terminals) are attached. Circles denote natural moisturizing factor (NMF). It is proposed that the lack of the NMF causes a stronger interaction between keratin bundles and their collapse – low elastic properties. With the presence of NMF the required elasticity is reached lowering the attraction between filaments.

2.3. Hair Structure

The major constituent of hair is protein - keratin. Hair is a highly hydrophilic structure with capability to absorb up to 30% of its own weight of water. Hair fiber consists of 65-95 % keratin proteins of its total weight.[18] Figure 2.6 represents the structure of hair fiber with keratin arrangement. On the image (1) the cross section of a hair fiber is presented. The cortex cells (cortical cells) are the main component of hair follicile.[103,104] These long cells are connected together in a cement environment which is rich in proteins and lipids. Small granules of melanin being responsible for the colour of hair can be also distiguished.[18] Cortex cells are made of keratin macrofibrils (in parallel orientation) of which the cross section is shown on image (2). Further magnification is shown in the image (3) of microfibrils (intermediate filaments) which form the macrofibrils. Seven to ten tetramer units form the intermediate filaments which are made of four coiled monomeric keratin chains. Keratins consist of a high concentration of cysteine which has a sulphur atom. Two sulphur atoms from different cysteines form a covalent disulfide bonds. These bonds are responsible for the stability of the tertiary structure of protein as well as the overall hair appearance (curly, straight, damaged etc).[105]

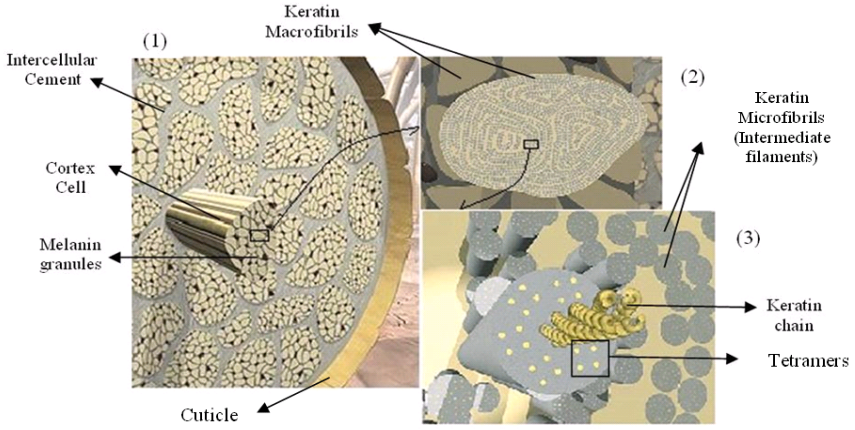


Figure 2.6. The representation of keratin in the structure of the hair fiber.[106]

The base of the hair follicle embedded in dermis constitutes a dermal papilla where the growth takes place.[107] The bloodstream provides the papilla with nourishment in order to produce hair.

A hair follicle is composed of:

- The lower part of hair follicle - the onion-shaped hair bulb,
- The inner root sheath
- The hair shaft: medulla, cortex, cuticle
- The outer root sheath

In Figure 2.7 the structure of the hair follicle is presented. Different pairs of keratins type I and type II are expressed within different parts of the hair follicle. This detail information was achieved by the use of indirect immunofluorescence microscopy.[108] In this method the sample of the hair follicle is visualized by the fluorescent microscope together with appropriate antibodies conjugated to the fluorescent labels. The primary antibody by recognizing the target biomolecule binds to it. The secondary antibody being conjugated to the fluorescent dye binds to the primary antibody localizing the target molecule. The microscope detects the distribution of the target molecule across the hair follicle sample and provides its visualization.[266]

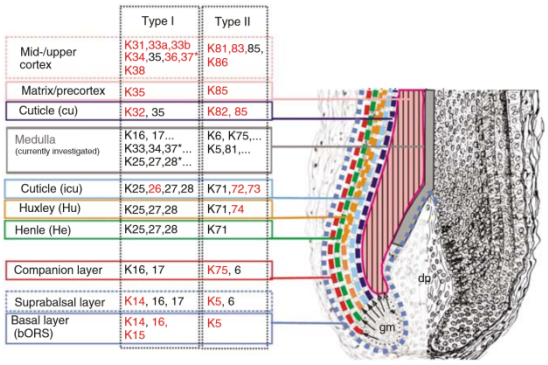


Figure 2.7. The structure of hair follicle with epithelial/hair keratins displayed in different parts.[108]

New hair is formed inside the hair bulb. Each hair follicle possesses a cavity where the dermal papilla is embedded. The lowest part of the bulb is responsible for production of new cells. During their growth and development they continuously force the cells produced previously to move upwards. In the hair bulb the special cells produce melanin pigment which provides hair with a color.[18]

The bulb consists of the hair matrix, the inner root sheath and the hair shaft which is made of: medulla, cortex and cuticle. The medulla constitutes the innermost part. The cortex contains fibres which are responsible for the strength and elasticity of hair. The cuticle is the thin outermost and providing with the protection to the cortex. The outer root sheath surrounds the follicle and protects the hair shaft within the follicle.[108]

As previously mentioned proteins account for 65-95% of the dry mass of human hair. Hair is mainly made of hard keratins due to the high amount of sulphur (cysteine).[109] The disulfide covalent bonds are strong and resistant to breaking apart. Their key role is the hair elasticity and resistance to damage under the external stresses. Although disulfide bonds are known to strengthen the hair being resistant to applications of acid, it is possible to break them by application of alkali solutions. This feature is used in straightening processes of hair where new shape is achieved which then can be stabilized by the application of acid. Hence, following the neutralization of the alkali the disulfide bonds reforms. The rebuilt disulfide bonds enable the hair to keep its new shape.[106]

Water constitutes the integral part of the hair keratin structure. The water content of hair depends directly on atmospheric humidity. Relative humidity is a term which is commonly used for describing the ratio of the water vapour amount in the air at a specific temperature to the maximum amount that air can retain at this

temperature. This is expressed as a percentage value. In Table 2.1 the water content of hair at different relative humidity at approximately 23˚C is presented.[109] It is observed that the increase of relative humidity results in the greater water sorption and hence the water content.

Table 2.1. The moisture content of hair at different relative humidities.[109]

Relative Humidity	*Approximate water content [%]*
29.2	6.0
40.3	7.6
50.0	9.8
65.0	12.8
70.3	13.6

The amount of moisture plays a crucial role in the physicochemical and cosmetic properties of hair fibres. Hair keratins possess several hydrophilic side chains and peptide bonds which are responsible for water attraction. It has been shown that at the low humidity the amino and guanidino groups are mainly responsible for the water sorption by keratin fibres. However, other results provide the information that peptide bonds play the key role in water sorption process.[109] The results from NMR of human hair show that the water protons are hydrogen bonded to keratins presenting a lower mobility than in the bulk liquid. Below 25% of relative humidity the hydrophilic residues are occupied by water molecules via the hydrogen bonds. Above this humidity, further attraction of water occurs resulting in a lower binding energy of water which has already been absorbed by the keratin. At 80% relative humidity and higher values the water on water phenomenon becomes the key mechanism.[109]

2.4. Keratins (Type I & II Intermediate Filaments)

Keratins are fibrous proteins which are long chains of amino acids linked by covalent peptide bonds connecting carbon atom of one amino acid and nitrogen of the consecutive one. Twenty amino acids are involved in forming their structure. Keratins are the major component of human skin and hair. They are also found in horns, claws, nails and feathers which suggest their strong structure.[88,110] Keratins are a significant factor in the mechanical stability and integrity of epithelial cells and tissues. They protect epithelial cells from mechanical and non-mechanical stresses which are able to cause the death of the cell.

Furthermore, some keratins also have regulatory functions where they are involved in intracellular signalling pathways, (e.g. protection from stress or wound healing). The keratins are the typical epithelia intermediate filament proteins which show a very high degree of molecular diversity. It is well established that in humans 54 functional keratin genes exist.[111] These heteropolymeric filaments are built by pairing type I and II molecules forming a dynamic network of 10-12 nm filaments.[112] Keratins of type I are acidic with a low molecular weight ranging from 40 to 64 kDa. In contrast, type II keratins consist of natural-to-basic high molecular weight proteins (from 52 to 67 kDa). The structure of monomeric keratin chain is shown in Figure 2.8. All the keratins possess the tripartite domain structure which is characteristic for intermediate filament (IF) proteins.

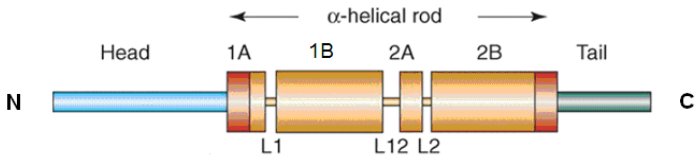


Figure 2.8. Schematic representation of keratin tripartite domain: non-helical head and tail and helical rod.[113]

The central α-helical domain ("rod") constitutes a major determinant of self-assembly. At the N- and C-termini the non-helical "head" and "tail" domains are present. The nitrogen (N) and carbon (C) terminals vary in structure and size. The α-helical rod domain (length of approximately 310 amino acids) is interrupted at three locations by non-helical linkers L1, L12 and L2. Monomeric keratin chain consists of α-helical segments 1A, 2B, 2A, 2B. Two monomeric units form a parallel left-handed coiled-coil dimer, two of which aggregate and

create a tetramer (Figure 2.9). Further assembly of tetramers involve formation of protofilaments. Four right handed, rope-like structures of protofilament dimers are the final conformations of intermediate filaments.

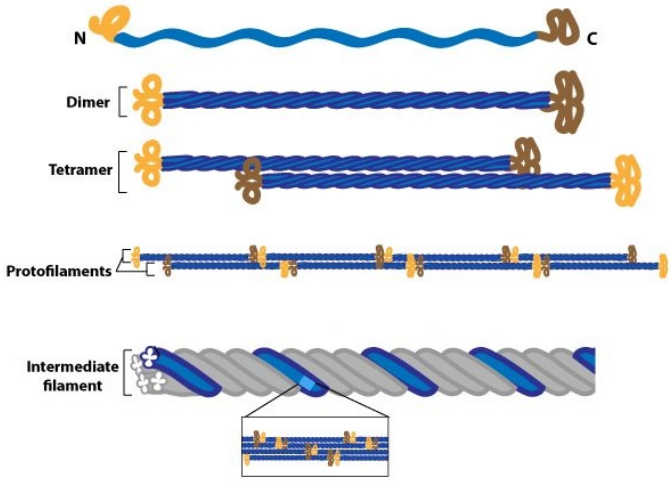


Figure 2.9. The structure of intermediate filament (IF).[114]

Keratins are divided into two types: type I (“acidic”) and II (“basic and neutral”). Within last twenty years a huge number of hair follicle-specific epithelial keratins have been discovered. It is possible to distinguish subfamily of keratin proteins. They can be divided them into "soft" keratins, which can be found in skin and "hard" keratins which hair fibers and nails are composed of. This classification comes from the high to low cysteine content in the keratin sequences where cysteines form disulfide covalent bonds making the conformation more resistant to degradation. Contrary, “soft” keratins (low cysteine content) can be broken down much easier being stabilized by hydrophobic forces and electrostatic interactions only (non-covalent).

In 2006 the “Keratin Nomenclature Committee” established a novel consensus nomenclature for mammalian keratin genes and proteins (Table 2.2).[115] There are 28 type I keratin genes (17 epithelial keratins and 11 hair keratins) and 26 type II keratin genes (20 epithelial keratins and 6 hair keratins). In total, 54 human keratin genes exist.

Table 2.2. The nomenclature of known mammalian keratins.[115]

	Type I		Type II	
Keratin types	New name	Former name	New name	Former name
Epithelial keratins	K9	K9	K1	K1
	K10	K10	K2	K2
	K12	K12	K3	K3
	K12	K12	K4	K4
	K13	K13	K5	K5
	K14	K14	K6a	K6a
	K15	K15	K6b	K6b
	K16	K16	K6c	K6e/h
	K17	K17	K7	K7
	K18	K18	K8	K8
	K19	K19	K76	K2p
	K23	K23	K77	K1b
	K24	K24	K78	K5b
			K79	K6l
			K80	Kb20
Hair follicle-specific epithelial keratins (root sheath)	K25	K25irs1	K71	K6irs1
	K26	K25irs2	K72	K6irs2
	K27	K25irs3	K73	K6irs3
	K28	K25irs4	K74	K6irs4
			K75	K6hf
Hair keratins	K31	Ha1	K81	Hb1
	K32	Ha2	K82	Hb2
	K33a	Ha3-I	K83	Hb3
	K33b	Ha3-II	K84	Hb4
	K34	Ha4	K85	Hb5
	K35	Ha5	K86	Hb6
	K36	Ha6		
	K37	Ha7		
	K38	Ha8		
	K39	Ka35		
	K40	Ka36		

As described previously, keratins possess the alpha helical rod which is responsible for the keratin arrangement. The alpha helix is the most common motif in the secondary structure of proteins and accounts for about 30% on average in globular proteins.[116] Helical conformation constitutes a right-handed spiral structure in which every amide group of amino acid (N-H) forms a hydrogen bond with oxygen of carbonyl (C=O) group of amino acid four residues earlier. The hydrogen bond is approximately 0.19 nm long and plays a key role in stabilization of the secondary structure. The α-helix consists of a 3.6 residues per turn with length of 0.54 nm (Figure 2.10).

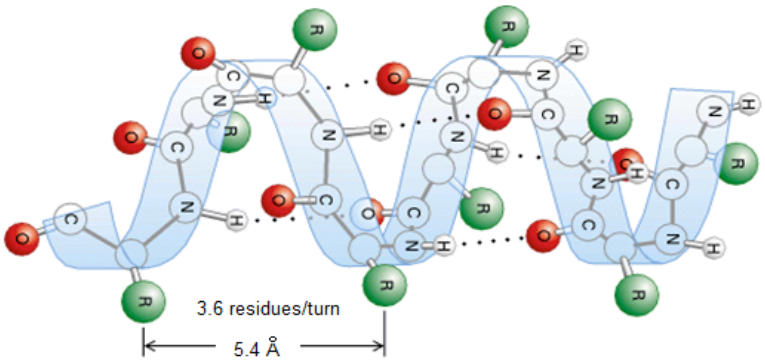


Figure 2.10. The structure of α-helix (O – oxygen, N – nitrogen, C – carbon, R – amino acid side chain, dotted line represents the hydrogen bond stabilizing the secondary structure).[117]

The α-helical conformation is the most uniform, common and predictable structure based on the sequence. It was proposed by Linus Pauling and Robert Corey in the early 1950s.[118] The α-helix is also known as the Pauling-Corey helix.

As previously mentioned, keratins have a unique property by pairing in parallel into heterodimers –type I and II (Figure 2.11). They associate within helical rod domains into a coiled-coil conformation.[118] These are stable forms in which two α-helix wrap around resulting in left handed coiled coil structure which is 48 – 50 nm in length. It is possible to represent each α-helical chain as a characteristic sequence of heptad repeat "abcdefg" (Figure 2.11).[118] This motif is repeated itself every seven amino acids along the chain. The residues "a" and "d" (the first and the fourth) are known to be occupied by nonpolar amino acids building up a hydrophobic core between α-helices. Residues "e" and "g" (the fifth and the seventh) with opposing charges attract each other by electrostatic interactions forming a "salt bridge". Residues "b", "c" and "f" are hydrophilic exposed outside the coiled-coil conformation.

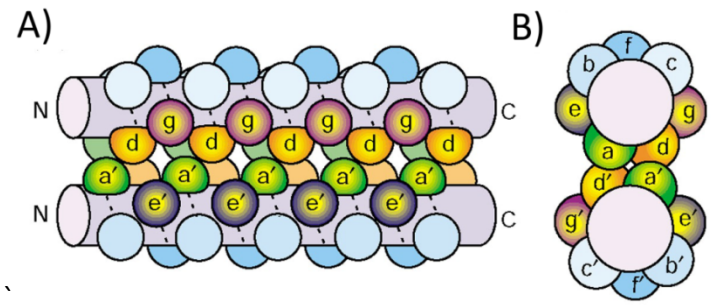


Figure 2.11. Coiled-coil conformation: A) Parallel arrangement of two helical chains, B) Axial projection.[119]

The further level of aggregation of heterodimers involves formation of tetramers. They are formed by antiparallel alignment of two coiled-coils in one of the modes: dimers overlapping 1B segments, dimers overlapping 2B segments, dimers overlapping 1B and 2B segments.[120] Keratin chains in each tetramer are linked by covalent disulphide bonds and/or ionic as well as hydrogen bond. Although the strength of the hydrogen bond is significantly weaker than the disulfide bond, the elasticity of keratins comes mainly from the flexibility of hydrogen bonds which are much larger in number. Hence, hydrogen bonds responsible for the keratin filaments stabilization are a key factor in maintaining the tertiary structure of these fibrous proteins.

CHAPTER III

Methodology

3.1. Introduction

Nowadays, computational chemistry plays a key role in predicting thermodynamic and structural properties in most of industry branches. Molecular simulations constitute perfect tools which give the opportunity to test a theory. These computational methods enable scientists to follow the atomistic scale resolution which cannot be observed using experimental techniques. The use of molecular simulations also results from the increasing costs of experimental work as well as often extreme and unreachable conditions at the laboratorial level. For instance, in predicting protein-ligand binding free energy, isothermal titration calorimetry (ITC) cannot be applied for very low or very high affinities – equilibrium association constant can be obtained at the range of $10^2 – 10^9$ M^{-1}.[135] That arises from the ITC sensitivity and thus a need of high concentrations of reactants where: for low affinity interactions the heat capacity may not be correctly captured (low sensitivity) while for the high affinity interactions may result in aggregation and hence non-availability of the studied compounds.[174] MD simulations provide an excellent opportunity to explore in details experimental results providing high resolution of the studied system.

3.2. Molecular Dynamics

Molecular Dynamics (MD) is a computer simulation method which allows us to follow the motion (dynamics) of all atoms and molecules in a given system. The positions, orientations and velocities which change over time can be observed within MD simulations. This technique enables us to follow and to understand structural (conformational) changes and dynamics of the molecules in a given system with a high degree of precision. The ergodic hypothesis states that in the system which is sampled over the long period of time and its total energy

is constant, all energy regions will be equally visited. The time average property of the system is equal to the average over all energy states - the whole phase space in terms of positions and momenta is sampled provided that the simulation is run over a sufficient time. Hence, in biological systems the ergodic hypothesis allows researchers to compare their time averaged results of e.g. binding affinity, heat of absorption, surface tension, diffusivity etc. obtained from computer simulations with the experimental (ensemble averaged) data providing better insight and understanding of the nature of the studied system. However, it should be noted that computer simulations have limitations in terms of the size of the system and reachable time scales. Thus, many methods were developed in order to improve sampling of the phase space by modifications of the Hamiltonian e.g. metadynamics[265], umbrella sampling (US)[144-145] or replica exchange molecular dynamics (REMD).[261]

MD is often called a "virtual microscope" and is widely used in the study of biological molecules and in materials science. MD numerically integrates classical Newton equations of motion:

$$\vec{F}_i = m_i \vec{a}_i \tag{3.1}$$

of atom i in a system formed N of atoms. In the above equation m_i stands for mass, the acceleration is $\vec{a}_i = \frac{d^2 \vec{r}_i}{dt^2}$ (second derivative of the position $\vec{r}_i$ with respect to time t, describing how velocity changes over time) and $\vec{F}_i$ is the force which results from the interactions with other atoms. In Hamiltonian form it can be written:

$$\dot{\vec{r}}_i = \frac{\vec{p}_i}{m_i} \tag{3.2}$$

$$\dot{\vec{p}}_i = \vec{F}_i \tag{3.3}$$

Equation (3.2) describes how the position changes in time (dot stands for time derivative) where $\vec{p}_i$ denotes momentum. The force acting upon atom i is equal to the change of momentum with respect to time as can be observed in equation (3.3). The trajectory of the system is calculated as a 6N – dimensional phase space,

where 3N stands for positions and 3N for momenta. Forces on atoms come from the total potential energy $V_i(\vec{r_i})$ (force field) of the system as presented in equation (3.4). Force is equal to changes of the total potential energy with respect to the positions:

$$\vec{F}_i = -\frac{dV(\vec{r_i})}{d\vec{r_i}} \tag{3.4}$$

In MD at each point of the calculations, 6 quantities are known for every atom: 3 positions (x, y, z) and respective forces. The Hamiltonian formulation of Newtonian dynamics is represented by the total energy as the sum of the total potential energy in terms of positions and total kinetic energy in terms of momenta:

$$H(\vec{r}_i, \vec{p}_i) = V(\vec{r}_i) + \sum_{i=1}^{3N} \frac{\vec{p}_i}{2m_i} \tag{3.5}$$

The total energy given by Hamiltonian does not change with respect to in time because there are no interactions with any external source of energy. Thus, the total energy of the system is conserved and remains constant:

$$\frac{dH}{dt} = 0 \tag{3.6}$$

The total potential energy $V(\vec{r_i})$ in MD simulations constitutes an analytical expression which is a function of all atomic positions. In order to calculate forces between atoms the interatomic potential is required $(\vec{F_i} = -\nabla_{\vec{r_i}} V(\vec{r_i}))$. It can be written as the sum of different types of interactions between atoms:

$$V(\vec{r_i}) = V_{bon} + V_{Coulomb} + V_{vdW} \tag{3.7}$$

The term V_{bon} describes bonded interaction, $V_{Coulomb}$ electrostatic interactions and V_{vdW} stands for van der Waals potential.

The first term V_{bon} of the equation (3.7) in the Charm22[121] force field which was implemented within this work describes atoms covalently bonded:

$$V_{bon} = \phi_{bond} + \phi_{angle} + \phi_{torsions} + \phi_{improers} + \phi_{Urey-Bradley} \tag{3.8}$$

The first term of the potential describes the energy corresponding to bond stretch. It is expressed as a quadratic function, harmonic potential:

$$\phi_{bond} = k_b(b - b^0)^2 \tag{3.9}$$

, where b stands for the bond length, k_b for bond force constant and b^0 represents equilibrium bond distance. The potential energy which comes from angle bending between consecutive covalently bonded atoms is described by the equation:

$$\phi_{angle} = k_\theta(\theta - \theta^0)^2 \tag{3.10}$$

In the equation 3.10, k_θ represents angle force constant, θ is the angle and θ^0 equilibrium angle. In the case that of four bonded atoms two planes can be defined: between atoms i, j and k and between atoms j, k and l. The angle φ between those two planes constitutes a dihedral angle. The dihedral angle potential is described as:

$$\phi_{torsions} = k_\varphi[1 + \cos(n_{jk}\varphi - \varphi^0)] \tag{3.11}$$

k_φ is a dihedral force constant, n_{jk} stands for multiplicity and φ^0 represents the phase shift.

Improper dihedrals are established in order to keep plane groups planar (e.g. aromatic rings):

$$\phi_{impropers} = k_{\xi}(\xi_{ijkl} - \xi^{0})^{2} \tag{3.12}$$

The term ξ_{ijkl} denotes the angle between two planes, k_{ξ} stands for the improper force constant and ξ^{0} represents equilibrium improper angle. The Urey-Bradley[122,123] term of bond-angle vibration (angle bending) using 1,3 nonbonded interactions between triplet of atoms i, j, k is described by a harmonic function:

$$\phi_{Urey-Bradley} = k_{\eta}(\eta_{ijk} - \eta^{0})^{2} \tag{3.13}$$

, where k_{η} is the force constant, η_{ijk} is the distance between atoms i and k (1, 3 atoms) in the harmonic potential and η^{0} stands for their equilibrium distance.

The non-bonded potentials are described by electrostatics $V_{Coulomb}$ and van der Waals V_{vdW}. Van der Waals interactions are described by Lennard-Jones potential which is expressed:

$$V_{vdW} = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{3.14}$$

, where ε stands for the depth of the potential well, σ is a finite distance when the inter-atomic potential is zero and r_{ij} is the distance between atoms. The term r_{ij}^{-12} describes a repulsion while r_{ij}^{-6} is the attractive long-range term. That gives an approximation of repulsive and attractive forces between non-covalently bonded atoms. Van der Waals interactions cannot be calculated for two bonded atoms as the repulsive energy would be very high. It is also not possible to implement those calculations between atoms separated by two covalent bonds due to the short distance between them hence those interactions are excluded within Van der Waals term. Atoms separated by three covalent bonds (1-4 interactions) depending on the force field are either multiplied by a factor (<1) or possess specific Lennard-Jones parameters.

Non-bond electrostatic interactions come from the different electronegativity of atoms. Different atoms can share or take electrons from each other due to their partial charges which results in a constant exchange of electrons between them.

By the assumption of point charges of atoms the electrostatic energy can be described by the Coulomb law:

$$V_{Coulomb} = C_{unit} \sum_{i<j} \frac{q_i q_j}{r_{ij}} \tag{3.15}$$

The terms q_i and q_j are the charges of atoms i and j, r_{ij} is a distance between them and C_{unit} is a Coulomb force constant. The exclusions for the electrostatic interactions involve bonded atoms and those separated by two or three covalent bonds depending on the force field.

By summing of all of the equations above the total potential energy (force field) is obtained:

$$\begin{aligned} V(\vec{r_i}) = & \sum_{bonds} k_b (b - b^0)^2 + \sum_{angles} k_\theta (\theta - \theta^0)^2 + \sum_{torsions} k_\varphi [1 + \cos(n_{jk}\varphi - \varphi^0)] + \\ & + \sum_{impropers} \frac{1}{2} k_\xi (\xi_{ijkl} - \xi^0)^2 + \phi_{Urey-Bradley}(\eta) = k_\eta (\eta_{ijk} - \eta^0)^2 + \\ & \sum_{electr} C_{unit} \frac{q_i q_j}{r_{ij}} + \sum_{vdW} 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{3.16}$$

Different force fields depending on the nature of the system are available (e.g. proteins, lipids, alkanes etc). Hence, the choice of the force field is one of the most important issues in MD simulations. For new molecules or group of atoms the force field parametrization is necessary often by accompanying experiment.

The temperature of the system in MD is directly related to the average kinetic energy. The time average total kinetic energy E_{kin} is defined as:

$$E_{kin} = \frac{N_E}{2} k_B T = \frac{1}{2} \sum_{i=1}^{N} m_i \vec{v}_i^2 \tag{3.17}$$

, where k_B is the Boltzmann constant, N_E stands for the number of effective degrees of freedom ($N_E = 3N - 3$ - in most cases the centre of mass is zero at time zero). Hence, each degree of freedom contributes $\frac{1}{2}k_BT$ which origins from the equipartition theorem.

3.2.1. Steered Molecular Dynamics

The stability of biomolecules in terms of inter- and intramolecular forces is a crucial challenge in understanding their conformational changes as well as their mechanical properties. The strength of interactions of protein-protein, protein-ligand complexes[124-126] is a key in predicting their binding propensities. Steered molecular dynamics (SMD) is a simulation method which allows scientists to observe with the atomistic resolution the unbinding process of biological complexes. SMD has been previously attempted and implemented in many systems and results can be often compared to the experimental data from atomic force microscope (AFM).[127-130] SMD has been previously shown as a very predictive method in drug discovery branch.[131] For instance, using Jarzynski equality[132] SMD is able to estimate the equilibrium thermodynamic property (binding free energy) from hundreds of non-equilibrium simulations. In SMD, a constant velocity or force is applied to group of atoms (e.g. ligand bound to a protein surface) which attached to a "virtual spring" (harmonic potential e.g. in one dimension $U(\vec{x}) = K_{SMD}(l - l_0)^2$ where, K_{SMD} is the spring constant, l is the reaction coordinate (distance), l_0 is the initial position of the ligand) is pulled away. Using SMD with constant velocity (pulling rate) applied to the ligand, the external force $\vec{F}$ exerted[133]:

$$\vec{F} = K_{SMD}(l_0 + \vec{v}t - l) \tag{3.18}$$

, where $\vec{v}$ is the pulling velocity and t is the time of the SMD simulation. During the pulling time the force builds up to the breaking points where intermolecular forces (i.e. hydrogen bonds, salt bridges and/or hydrophobic attraction) are no longer present and two molecules (protein and ligand) are separated. The example of the force versus time profile for hydrogen bonding and hydrophobic governed interactions

between ligand (EGCG) and protein (keratin) is shown in Figure 3.1.[134] The force builds up over the simulation time to the breaking point (maximum pulling force) where the hydrogen bond was broken and protein no longer interact with the ligand. Hence, SMD simulations can provide the information about stability of the interactions.

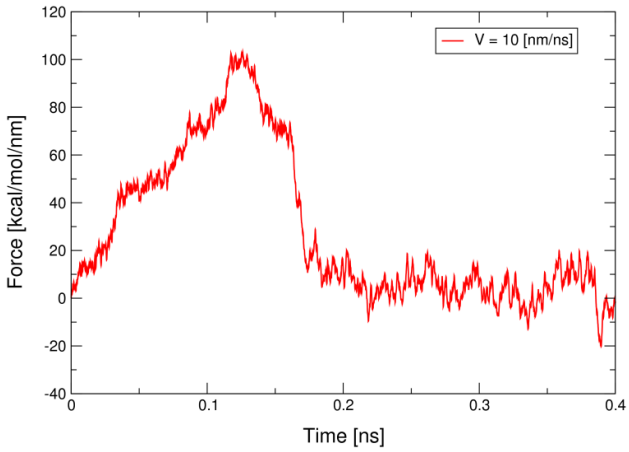


Figure 3.1. The example of force versus time profile obtained from steered molecular dynamics with applied constant velocity of 0.01 nm ns^{-1} to the centre of mass of ligand (EGCG) being within protein surface (keratin).[134]

Previously, Mai, B K et al[136,137] proposed a linear correlations of the maximum pulling force (rupture force) obtained from SMD simulations versus the binding free energy obtained from experiment and free energy calculations within MD simulation for different ligand in the treatment of A/H1N1 influenza virus. The advantage of this linear dependency is that the ligand binding affinity can be obtained from short SMD simulations (hundred of picoseconds to nanoseconds) instead of expensive and time consuming experiments or free energy calculations within MD methods (hundred of nanoseconds to microseconds).

3.2.2. Free Energy Calculations

Free energy is one of the most desired thermodynamic property in simulations of biological systems.[138,139] The difference between the final and initial state (the free energy difference) is a driving force of every process e.g. the protein-ligand binding. It involves the entropic and enthalpic contribution and is directly related to the binding constant, the latter which provides the equilibrium concentration of: the unbound free ligand, the

unbound protein and the protein-ligand complex. Recently, there is increasing number of studies reporting binding free energy calculations of small molecules to proteins via molecular dynamics (MD) and further experimental validation. The progress in computational resources and gained promising agreement between computer simulations and experiment reveals MD as a very predictive tool in the drug discovery branch. Methods for the free energy calculations involve molecular mechanics/Poisson-Boltzmann (MM/PBSA)[140,141], free energy perturbation (FEP)[142], linear interaction energy (LIE)[143] or umbrella sampling (US)[144,145]. Although these methods involving explicit solvent provide very accurate binding affinities, they are computationally very expensive.

As described previously, MD simulations provides with positions, velocities and forces of atoms over a simulation time providing with structural changes. However, those values cannot be directly compared to the experimental data. Statistical mechanics with the ergodic hypothesis allows scientists to extract desired thermodynamic property validating experimental data. Taking this into account and the fact that the entropy being a measure of the available space consisted in the free energy, the extensive sampling over a long simulation time is necessary.

In statistical mechanics the canonical partition function (in the system with constant number of atoms, constant volume and temperature) describing the statistical properties of the system in thermodynamic equilibrium is an integral of the all possible energy states (over the whole phase space in terms of positions and momenta)[146, 147]:

$$Q = \int \exp[-\beta H(\vec{r}_i, \vec{p}_i)] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i \tag{3.19}$$

, where $\beta = 1/(k_B T)$.

The Helmholtz free energy is related to the partition function via:

$$A = -1/\beta \ln Q \tag{3.20}$$

In the isothermal-isobaric ensemble (constant number of atoms N, constant pressure P and temperature T) the Gibbs free energy can be calculated.[148] However, this cannot be calculated from computer simulation. When the system is ergodic, so that the whole phase space is visited and the simulation is run over sufficient time, the ensemble average is equal to the time average.

In most cases the interesting thermodynamic property is the free energy difference between the final and the initial state (free energy is a state function). In protein-ligand free energy systems these correspond to the bound and unbound states which differ in geometry. The reaction coordinate provides the distinction between initial and final state and can be defined e.g. as a distance, angle or root mean square deviations. Whether in the protein-ligand system the distance is the reaction coordinate, the free energy along this pathway is called potential of mean force (PMF). The free energy along the reaction coordinate $A(\lambda) = -1/\beta \ln P^U(\lambda)$ can be obtained directly from MD simulations by monitoring the unbiased distributions $P^U(\lambda)$ of the system along the reaction coordinate.

Umbrella sampling (US) method for PMF calculations introduced by Torrie and Valleau[144,145] which based on the previous work[149,150] ensures that extensive sampling is reached with additional biased potential - energy term. This is achieved by multiple overlapping windows (simulations) along the reaction coordinate, each with e.g. harmonic potential (virtual spring)[148]. This additional energy $u_i(\lambda)$ term is a function of the reaction coordinate λ only. Hence, the biased total energy $H^B(\vec{r_i})$ (bias – "B" superscript) is a sum of the unbiased (unbias – "U" superscript) $H^U(\vec{r_i})$ total energy and the new bias potential $u_i(\lambda)$ where i denotes given window:

$$H^B(\vec{r_i}) = H^U(\vec{r_i}) + u_i(\lambda) \tag{3.21}$$

The goal is to obtain unbias PMF – free energy along the reaction coordinate, hence free energy of given window $A_i(\lambda)$ is required. For this purpose the unbiased distribution is needed:

$$P_i^U(\lambda) = \frac{\int \exp[(-\beta H(\vec{r}_i, \vec{p}_i))\delta[\lambda(\vec{r}_i) - \lambda] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]}{\int \exp[(-\beta H(\vec{r}_i, \vec{p}_i)) d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]} \tag{3.22}$$

However, MD simulations with additional energy terms provide the biased distribution:

$$P_i^B(\lambda) = \frac{\int \exp\{-\beta[H(\vec{r}_i, \vec{p}_i) + u_i(\lambda(\vec{r}_i))]\}\delta[\lambda(\vec{r}_i) - \lambda] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]}{\int \exp\{-\beta[H(\vec{r}_i, \vec{p}_i) + u_i(\lambda(\vec{r}_i))]\} d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]} \tag{3.23}$$

As the additional bias term depends on the reaction coordinate only and integration is done over all degrees of freedom:

$$P_i^B(\lambda) = \exp[-\beta u_i(\lambda)] \frac{\int \exp\{-\beta H(\vec{r}_i, \vec{p}_i)]\}\delta[\lambda(\vec{r}_i) - \lambda] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]}{\int \exp\{-\beta[H(\vec{r}_i, \vec{p}_i) + u_i(\lambda(\vec{r}_i))]\} d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]} \tag{3.24}$$

Thus, the unbiased distribution using equation (3.22) can be written as:

$$\begin{aligned} P_i^U(\lambda) &= P_i^B(\lambda)\exp[\beta u_i(\lambda)] \frac{\int \exp\{-\beta[H(\vec{r}_i, \vec{p}_i) + u_i(\lambda(\vec{r}_i))]\} d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]}{\int \exp[-\beta H(\vec{r}_i, \vec{p}_i)] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]} = \\ &= P_i^B(\lambda)\exp[\beta u_i(\lambda)] \frac{\int \exp[-\beta H(\vec{r}_i, \vec{p}_i)] + \exp\{-\beta u_i[\lambda(\vec{r}_i)]\} d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]}{\int \exp[-\beta H(\vec{r}_i, \vec{p}_i)] d^{N_E}\vec{r}_i d^{N_E}\vec{p}_i]} = \\ &= P_i^B(\lambda)\exp[\beta u_i(\lambda)]\langle \exp[-\beta u_i(\lambda)]\rangle \end{aligned} \tag{3.25}$$

Now, free energy of the given window can be calculated according to $A_i(\lambda) = -(1/\beta)\ln P_i^U(\lambda)$, biased distribution $P_i^B(\lambda)$ is given by MD simulations, additional energy term $u_i(\lambda)$ is taken analytically. The free energy associated with additional energy term (biased potential) F_i^B is correlated:

$$\left\langle \exp[-F_i^B(\lambda)] \right\rangle = \exp(-\beta u_i(\lambda)) = \int P_i^U(\lambda) \exp(-\beta u_i(\lambda)) d\lambda =$$
$$= \int \exp\{-\beta[A(\lambda) + u_i(\lambda)]\} d\lambda \quad (3.26)$$

Hence, the unbiased free energy of the given window can be expressed as:

$$A_i(\lambda) = -(1/\beta) \ln P_i^B(\lambda) - u_i(\lambda) + F_i^B \quad (3.27)$$

The only assumption of the equation above is the sufficient time of sampling. Taking into account that the desired PMF combines more than one window, the F_i^B of each simulation has to be calculated. However, F_i^B cannot be directly obtained from simulation and numerous methods have been introduced.[151] The most common technique is called weighted histogram analysis method (WHAM)[152] which enables to combine all the windows into the PMF curve providing with the free energy along the reaction coordinate $A(\lambda)$. In Figure 3.2 the schematic representation of US method (red circle – protein, blue circle - ligand) with the distance as a reaction coordinate is presented. Eleven simulations along the distance are run with the additional harmonic potential (spring) providing the histograms - probabilities of finding. Histograms constitute the input for WHAM method in which each sampling window obtains its weight g_i so the unbiased distribution:

$$P^U(\lambda) = \sum_i^{windows} g_i(\lambda) P_i^U(\lambda) \quad (3.28)$$

, where g_i are adjusted to minimize to statistical error of the distribution $P^U(\lambda)$ and $\sum g_i = 1$ [152] which leads to:

$$g_i = \frac{a_i}{\sum_j a_j}, a_i(\lambda) = N_i \exp[-\beta u_i(\lambda) + \beta F_i^B] \quad (3.29)$$

N_i denotes all the steps sampled in simulation i. Then, the F_i^B can be taken from equation (3.26):

$$\exp[-F_i^B(\lambda)] = \int P_i^U(\lambda)\exp[-\beta u_i(\lambda)]d\lambda \tag{3.30}$$

These need to be iterated until the convergence is reached.

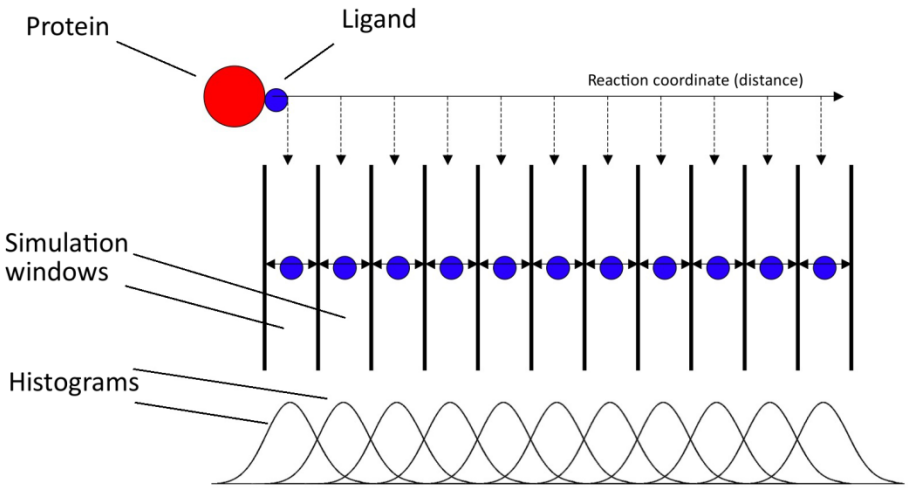


Figure 3.2. Schematic representation of umbrella sampling method: red circle- protein, blue circle – ligand. Along the reaction coordinate (protein – ligand distance) independent simulations are run with additional harmonic potential (virtual spring) where ligand is restrained by its centre of mass. Each window provides with the probability of finding (histogram) according to the harmonic potential. Overlapping windows provide the input to weighted histogram analysis method which provides the potential of mean force (free energy along the distance) of protein-ligand interactions.[153]

3.2.3. Coarse Graining

One of the most challenging issues in computational chemistry is the time-scale and the length-scale achievable in comparison to realistic systems. This comes from the computational resources available these days which is still a limiting factor in molecular simulations. To overcome this problem in order to reduce the number of degrees of freedom, one can represent group of all atom reference system by grouping them into beads. Thus, with decreased number of atoms in the system much longer time scales and larger systems can be simulated. Although, the "coarser" (more atoms represented by a bead) the model the faster the simulation, the greater decline in the number of thermodynamic reproducible properties is reached. Hence,

for given system the appropriate representation of group of atoms need to be chosen in order to reproduce the key thermodynamic and/or structural properties of interest. Coarse graining was successfully implemented in many biological systems reaching excellent agreement with the reference atomistic trajectory and experimental data.[154-165] The most common and widely used coarse grained force field for biomolecular systems is Martini[166-168], where four heavy atoms represents one bead and parametrization is based on the free energies between polar and apolar phases of the known experimental data of many compounds. However, Martini force field cannot be implemented for any system and much better precision is obtained by creating for the specific case its own force field. That ensures about the reliable reproduction of the fully atomistic simulation with well established parameters.

The typical "bottom up" approach for the coarse graining is presented in Figure 3.3. First, the reference full atomistic simulation with known bonded and non-bonded parameters for all types of atoms (according to the equations 3.7-3.16) need to be run. It is worth noting that nowadays, fully atomistic force fields which are continuously developed based on the experimental data reproduce thermodynamic and structural properties of biological system with extreme precision. Full atomistic simulation requires reliable time scales in order to sample all possible energy states, which can be the most time consuming step in coarse-graining. Secondly, the number of atoms per bead in the coarse grain representation needs to be chosen. Subsequently, bonded and non-bonded interaction potentials for the coarser representation need to be extracted. Often, the two reference full atomistic trajectories are run in parallel: one with applied exclusions of bonded from non-bonded parameters and the second with both bonded and non-bonded parameters. That ensures about the precision of the extracting bonded and non-bonded distributions which inverted provide the potentials. Subsequently, the distributions of bonds, angles and dihedral angles (bonded potentials) as well as radial distribution functions (RDF)[169] - probabilities of finding of between given types of bead as a function of their distance (non-bonded potentials) need to be extracted from fully atomistic trajectory. In addition to reproduction of full atomistic bonded and non-bonded distributions within coarse grained representation, the thermodynamic and/or structural properties of interest need to be compared. This validation is often accompanying by the experiment. Hence, the validated force filed can be implemented reaching order of magnitudes faster time and length scales.

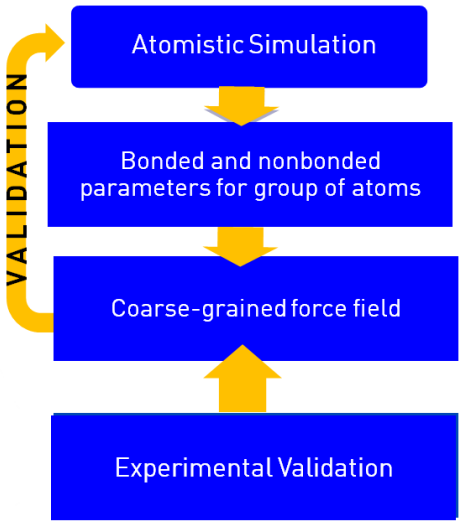


Figure 3.3. The standard "bottom up" approach for the coarse graining based on the all atom simulation. First, the full atomistic reference trajectory needs to be known. Then, by reproducing bonded and non-bonded distributions from full atomistic trajectory as well as structural and/or thermodynamic properties, the coarse grained force field is built often accompanying by the experimental input.

3.2.3.1. Boltzmann Inversion

Boltzmann inversion (BI) method, a structure-based technique is typically applied for bonded potentials [170-172] – bonds, angles and torsions (dihedral angles). In the canonical ensemble NVT (constant number of atoms N, constant volume V, constant temperature T) the probability distribution (Boltzmann distribution) is described as:

$$P(q) = Q^{-1} \exp[-\beta V(q)] \tag{3.31}$$

,where q are independent degrees of freedom (angle, torsion, bond length), $Q = \int \exp[-\beta V(q)] dq$ is the partition function. The coarse grained potential $V(q)$ (potential of mean force) can be obtained by the inverted Boltzmann probability $P(q)$ of a given variable q:

$$V(q) = -k_B T \ln P(q) \tag{3.32}$$

The normalization factor Q can be omitted as it would enter the coarse-grained potential as an additive irrelevant constant. $P(q)$ is a volume normalized distribution of bonds $P_b(b)$, angles $P_\theta(\theta)$ or dihedrals $P_\varphi(\varphi)$ based on the histograms $H(q)$:

$$P_b(b) = \frac{H_b(b)}{4\pi r^2}, P_\theta(\theta) = \frac{H_\theta(\theta)}{\sin\theta}, P_\varphi(\varphi) = H(\varphi) \tag{3.33}$$

b denotes the bond length, θ is the angle between three consecutive atoms, and φ states for the torsion angle between two planes formed by four consecutive atoms.

Hence, the bonded coarse grain potential can be written as a sum of bonds, angles and torsions:

$$V(q) = V(b,\theta,\varphi) = V_b(b) + V_\theta(\theta) + V_\varphi(\varphi) \tag{3.34}$$

$$V_q(q) = -k_B T \ln P_q(q)$$

Inverted distributions (potentials) need be smoothed for the continuity, interpolated in poorly sampled regions as well as extrapolated. This can be achieved within tabulated potentials instead of parameters used in all-atom force fields described by e.g. harmonic potential for bonds and angles.

3.2.3.2. Iterative Boltzmann Inversion

Iterative Boltzmann Inversion (IBI) method is mainly applied with non-bonded parameters in order to reproduce the reference P_{ref} radial distribution functions (RDFs) of two given types of beads - probabilities of finding at a certain distances. However, in cases of "coarser" representations this should be implemented for both bonded and non-bonded parameters in the one work flow. The IBI method block scheme is presented in Figure 3.4.

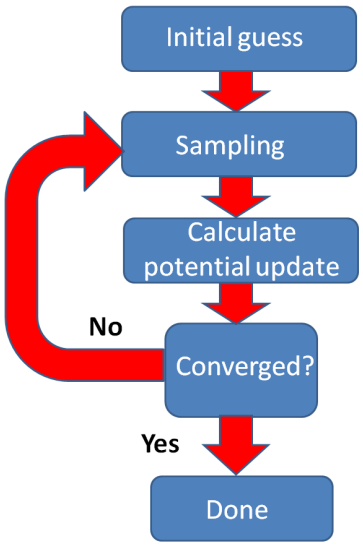


Figure 3.4. The work flow of the iterative Boltzmann inversion (IBI) method.[170]

First, the initial guess for tabulated potentials (step 1) needs to be known. This can be calculated according to the equation:

$$V^{n=1} = -k_B T \ln P_{ref} \tag{3.35}$$

, where n is the step number. Subsequently the simulation is run (sampled) for the sufficient time to extract the distributions from the first step $P^{n=1}$. The P_{ref} is compared with the $P^{n=1}$ and once the convergence is reached the workflow finishes – the reference distribution function was reproduced. Otherwise the rescaled potentials are applied and the consecutive simulation is run. The potential of the next step V^{n+1}:

$$V^{n+1} = V^n + \kappa \Delta V^n \tag{3.36}$$

, where κ is the scaling factor $\in (0,1)$ and V^n is the potential for the current step, the ΔV^n is the difference between the potential form the current step and the reference potential:

$$\Delta V^n = V^{ref} - V^n = k_B T \ln \frac{P^n}{P_{ref}} \tag{3.37}$$

Due to the fact that given potential between two types of beads influences the other potentials, it is recommended that at each step the only one potential is rescaled. As previously mentioned the convergence is reached once all the reference potentials are reproduced within coarse grained representation in terms of distributions $P^n = P_{ref}$.

CHAPTER IV

Molecular and Thermodynamic Basis for EGCG-Keratin Interaction

4.1. Introduction

As previously mentioned, binding of small molecules to macromolecules of human bio-substrates regulates their sub-cellular disposition. Transdermal permeation is an important route for delivering benefits of these small active molecules to the lower skin layers. There is an increasing interest in understanding the binding of plant extracts, such as green tea with a high content of polyphenols due to many health benefits attributed to them (section 2.1). Stratum corneum (SC), the outermost skin layer constitutes the first barrier for skin permeation in which keratin constitutes the major component (80% of its dry mass).[17] Hence, the strength of the SC keratin-EGCG interaction regulates the delivery of vital profits. The most abundant pair in the upper epidermis and SC forming IFs is keratin type 1 and 10, thus constituted the subject of this chapter. EGCG, the most active green tea polyphenol was selected as ligand.

Previously, Zhao et al.[23] showed that EGCG binds to gastric mucin by a combination of hydrophobic interactions and hydrogen bonding and multilayer binding of polyphenol was suggested. In addition, a decrease in binding affinity with increasing temperature was observed. However, thermodynamic equilibrium properties and structural basis for interactions between EGCG and keratin remain unknown and have never been studied before.

4.2. Objectives

The objective of this chapter is to study interactions of the main active green tea polyphenol (EGCG) with SC keratin using MD simulation framework. At the same time Dr Yanyan Zhao et al.[173] provided the experimental insight in the keratin-EGCG interactions. Experimental studies of the thermodynamic properties of keratin-EGCG interactions have been quantitatively determined using ultrafiltration, high-performance liquid chromatography (HPLC) and isothermal titration calorimetry (ITC). At 298 K the experimental binding free energy ΔG_{Exp} = -6.37 kcal mol^{-1} as well as the multilayer isotherm describing the assembly of EGCG on the keratin surface were obtained. All experimental results are presented in Appendix I. Hence, the main challenge of this chapter was to validate experimental results and provide more detailed insight to keratin-EGCG system using MD simulations. Thus, the main objectives within MD in studying keratin-EGCG interactions were as following:

- Validate the multilayer assembly of EGCG molecules,
- Validate the experimental binding free energy at 298K,
- Assess the influence of the temperature on the binding affinity,
- Assess the type of interactions which dominate the binding process,
- Evaluate residues with the highest binding affinity.

For this purpose, self-assembly simulations totalling ~6 μs were first used to generate possible keratin-EGCG bound states at a range of concentrations and temperatures. In addition, steered constant-velocity molecular dynamics (SMD) simulations were used to pull EGCG molecules away from the keratin surface. Subsequently, snapshots from the SMD simulations provided starting points for the umbrella sampling (US) method. Although in this chapter the only one reaction coordinate was studied, free energy calculations of ~12 μs using different equilibrium configurations at varying temperatures with explicit solvent reached an excellent agreement between simulations and experimental results.[134, 173]

4.3. Methodology – Molecular Dynamics

4.3.1. Computational Details

The structure of epigallocatechin-3-gallate (EGCG) was built using Chimera 1.5.3[176]. To create the EGCG topology the CHARMM General Force Field (CGenFF)[177-180] for organic molecules (program version 0.9.1 beta) was employed and translated into GROMACS[181] topology format using in-house code. A coiled coil keratin fragment was modelled based on the cortexillin I domain (Protein Data Bank entry: 1D7M). The cortexillin and keratin 1/keratin 10 residues heptad repeats ("abcdefg") were predicted using Matcher ("Algorithm for matching coiled-coil proteins").[182] Each helical fragment of the cortexillin coiled coil consists of 101 residues, of which 71 residues corresponding to segment 1B of keratin type 1 and 10 were chosen. Predicted cortexillin repeats were mutated into the keratin sequences. Alpha helical segment 1A (35 residues) and 2A (19 residues) of keratin 1 were built using Chimera 1.5.3 software. The amino acid sequences of all keratin segments were taken from the Human Intermediate Filament Database.[183] Initial structures were minimized via the steepest descent algorithm. The chosen segments, constructed models and corresponding sequences are presented in Figure 4.1.

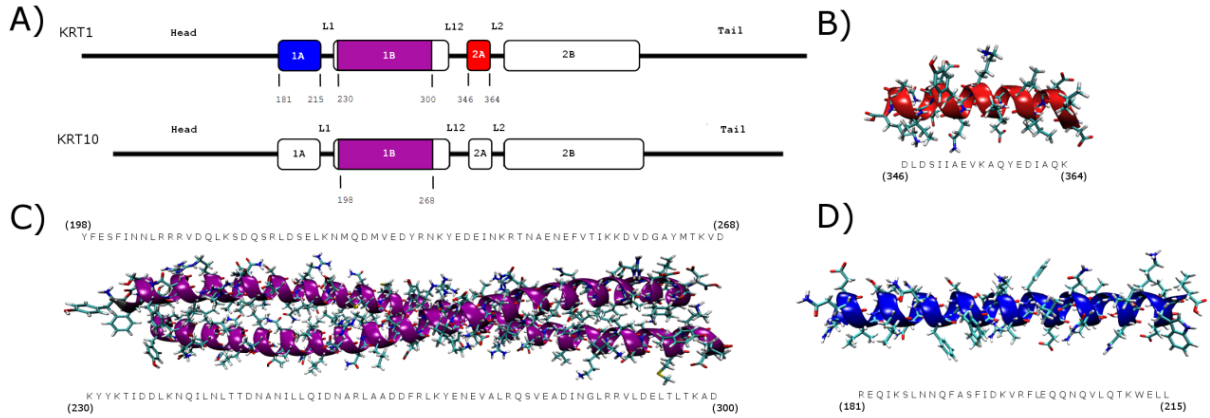


Figure 4.1. Segments and the corresponding sequence of keratin 1 and 10 used for MD simulations: A) Schematic representation of chosen segments 1A (blue), 1B (purple), and 2A (red); B) Helical segment 2A of keratin 1; C) Coiled-coil 1B segment of keratin 1/ keratin 10; D) Helical segment 1A of keratin 1.

All MD and SMD simulations were performed using the GROMACS 4.5.4[181] package with the Charmm/CMAP[121] force field and the TIP3P[184] water model. Classical equations of motion were integrated with a Verlet leap-frog algorithm using a 2 fs time step. The LINCS[185] algorithm was employed to constrain all bond lengths. A 1.4 nm cut-off distance was used for the short-range neighbor list, updated every five steps (10 fs), and van der Waals interactions were cutoff at 1.4 nm. The Particle Mesh Ewald (PME) method[186,187] was used for the electrostatics with a 1.2 nm real space cut-off. The velocity rescale thermostat and Parinello-Rahman[188] barostat were used to maintain the temperature and pressure (at 1 bar) respectively. Protein or protein and ligand were coupled to one bath, while the remaining water and ions were coupled to the other. The initial velocities were set to follow a Maxwell distribution. Periodic boundary conditions were used in all directions. All simulations were performed on two Linux clusters at Imperial College London: Department of Chemical Engineering Computing Service, URL: imperial.ac.uk/chemicalengineering and High Performance Computing Service, URL: imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing.

4.3.2. Self-Assembly Setup

Different keratin segments were placed at the centre of a cubic unit cell: Segment 1B with the length of app. 10 nm into a 12x12x12 nm box, segment 1A with the length of app. 5nm into the 8x8x8 nm box and segment 2A with the length of app. 3 nm (6x6x6 nm). To mimic the fact that both the helix and coiled coil segments are part of the Intermediate Filament, terminal residues were treated in their uncharged state. Initially, one EGCG molecule was placed into the box in a random position near to the given keratin segment. Each system involved one keratin segment (1A, 2A or 1B) and one EGCG molecule. The simulation box was filled with water using the TIP3P[184] water model, to which 150 mM NaCl was added on top of the neutralizing ions (6 Na^+ to segment 1B, 3 Na^+ to segment 2A and 1 Cl^- to segment 1A). Following energy minimization using the steepest descent algorithm, 0.2 ns of sequential NVT (constant number of atoms, constant volume and temperature) and NPT (constant number of atoms, constant pressure and temperature) equilibration simulations were conducted, with position restraints applied to protein and ligand heavy atoms. Production simulations were then performed in the NPT ensemble at 298, 308 and 318 K. In total, 51 EGCG-keratin MD simulations of 100 ns were run with one EGCG molecule placed at different initial positions around given keratin segment. This involved, in each case for three different temperatures: 8 simulations with coiled-coil segment 1B, 5

simulations with helical segment 1A and 4 simulations with helical segment 2A. The number of simulations corresponds to the size of the given protein segment and thus different possibilities to place EGCG randomly. Following the same procedure in order to assess higher concentration of EGCG, 3 additional simulations of 200 ns at 298 K were performed with 20 EGCG molecules placed a minimum of 6 nm away from the coiled coil fragment starting from different initial coordinates of EGCG molecules. Each system and corresponding self assembly simulation is listed in Table 4.1.

Table 4.1. Details of self-assembly simulations for keratin-EGCG interactions.

System No.	Segment	Number of EGCG molecules	Simulation time [ns]	Temperature [K]	Number of simulations
1.	1A	1	100	298	5
2.	1A	1	100	308	5
3.	1A	1	100	318	5
4.	2A	1	100	298	4
5.	2A	1	100	308	4
6.	2A	1	100	318	4
7.	1B	1	100	298	8
8.	1B	1	100	308	8
9.	1B	1	100	318	8
10.	1B	20	200	298	3

4.3.3. Biased Simulation Setup

The final coordinates of the keratin-EGCG complex from the self-assembly MD simulations were set as the initial configurations for SMD simulations. For each complex, the COM of an EGCG molecule was increased at a constant rate (velocity), being pulled up to 4 nm away. The ligand pulling direction provides different initial configurations for biased umbrella simulation and hence can significantly influence the potential of mean force. However, in this study it was aimed to explore many PMF profiles starting from different protein-ligand

complexes with the conserved reaction coordinate (distance). Thus, the ligand was pulled in a direction perpendicular to the protein backbone axis in all cases. As the steered MD simulation proceeded the pulling force was observed to increase, up to the point where the first keratin-EGCG hydrogen bond was broken. The breaking point of each hydrogen bond corresponds to a local maximum in the pulling force.

Steered MD was used to extract the initial coordinates for umbrella sampling windows (binding free energy calculations). A weighted histogram analysis method (WHAM)[152,189] was then applied to combine all windows into the potential of mean force (PMF) curve to estimate the binding free energy difference along the reaction coordinate. The statistical error estimate of each PMF curve was performed by dividing the whole trajectory into parts of 10 ns. The average number of hydrogen bonds between protein and EGCG was counted based on the cutoff distance between the donor and the acceptor atoms below 0.35 nm and a hydrogen-donor-acceptor maximum angle of 30˚. The buried area between the protein and ligand was computed by subtracting solvent accessible surface area (SASA)[190] of the protein and EGCG from SASA of keratin-EGCG complex. Both hydrogen bonds and buried area were extracted from equilibrium simulations, prior to SMD simulations and US.

4.4. Results – Molecular Dynamics

4.4.1. Monomeric Self-Assembly Simulations

Simulated tea catechin (EGCG) molecule became attached to multiple (between 1 and 5) residues at the end of each simulation replica in the system for a single EGCG and one keratin segment (Table 4.1. – Simulation no. 1 - 9). An example of the Keratin-EGCG hydrophobic buried area as well as the number of hydrogen bonds versus simulation time is presented in Figure 4.2.

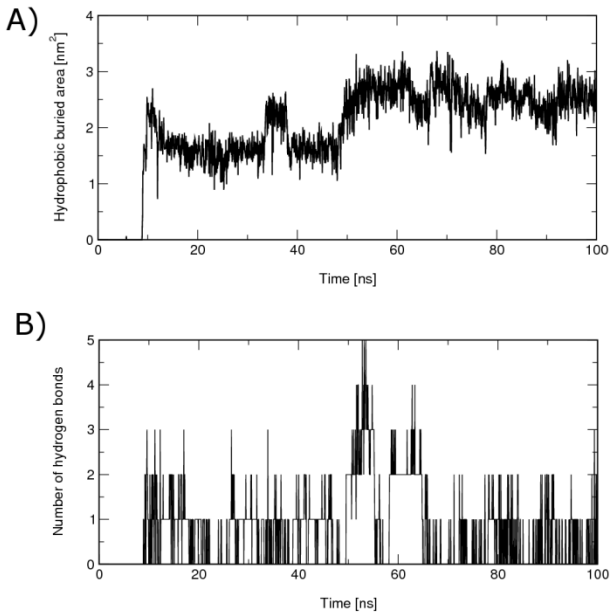


Figure 4.2. Monomeric self-assembly simulation of EGCG molecule and keratin at 298K: A) Buried hydrophobic surface area between EGCG and keratin during the simulation; B) Number of hydrogen bonds between keratin and EGCG versus the simulation time.

Hydrophobic interactions played a major role in the initial stages of protein-ligand assembly, followed by a slow optimization via formation of hydrogen bonds. In many cases, stacking interactions were observed in which the aromatic rings of EGCG became packed against those of phenylalanine or tyrosine (Figure 4.3 A). Side chains of acidic residues were predominantly occupied by catechin, supported by the fact that EGCG contains 8 hydrogen donors and 11 hydrogen acceptors (Figure 4.3. B, C). This confirms previous investigations in which EGCG was shown to interact specifically with charged amino acids.[191] In this study, once an acidic side chain oxygen or nitrogen had created a hydrogen bond with the EGCG hydroxyl group, it was subsequently persistent for the remaining simulation time (Figure 4.3. B and C).

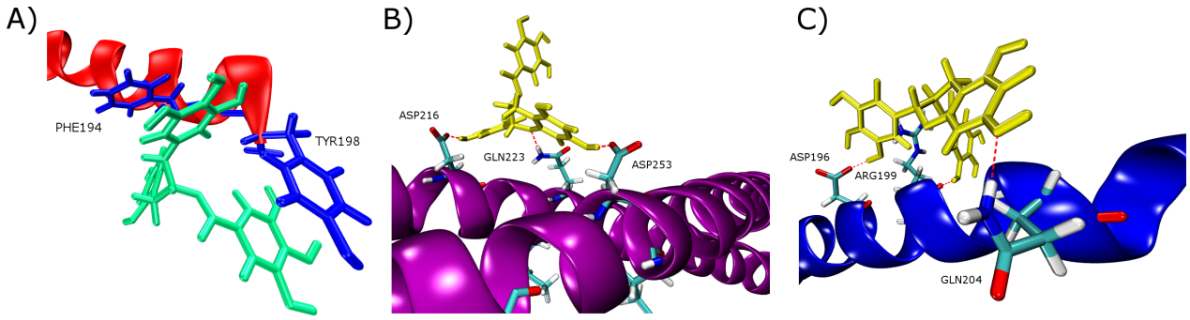


Figure 4.3. Final configurations of MD simulations: A) Segment 1A (red) and its two aromatic residues (blue) stacked via aromatic attraction to two EGCG (green) rings; B) Segment 1B coiled coil (purple) and EGCG (yellow) occupying three different residues via hydrogen bonds (red dotted lines) at 308K; C) Segment 1A (blue) and EGCG (yellow) occupying three different residues via hydrogen bonds at 298K.

In Table 4.2, the occupied keratin residues, as well as the average number of hydrogen bonds formed between EGCG and acidic residues of the coiled coil segment 1B, during 100 ns of simulation at 298 K are presented. The binding contribution of acidic amino acids varies widely from ~30 to 98 % due to the exposed hydrogen bond acceptors of their side chains, interacting with the multiple EGCG hydroxyl groups.

Table 4.2. The contribution of the average number of hydrogen bonds between EGCG and acidic residues as well as occupied amino acids.

Initial position of EGCG around the Keratin segment 1B	Average no of h-bonds (EGCG-protein)	Average no of h-bonds (EGCG-charged residues)	Contribution of h-bonds with charged residues [%]	Occupied Residues
1.	1.034	0.485	47	Asp258, Glu264
2.	0.799	0.249	31	Glu239, Ser247
3.	1.355	1.034	76	Asp225, Asn226, Asp233
4.	1.169	1.144	98	Asp251, Asp260, Glu261
5.	2.484	2.429	98	Glu261, Asp292
6.	1.002	0.604	60	Tyr230, Phe231, Glu232
7.	0.746	0.541	73	Asp272, Glu249, Asn275
8.	1.560	1.365	88	Asp265, Ala231, Tyr266

At 318 K, in 3 out of 17 self-assembly monomeric EGCG-keratin simulations (each of 100 ns), the EGCG molecule stacked onto the keratin surface after ~2-5 ns, and remained attached for ~40-60 ns before detaching, suggesting weaker interactions at higher temperatures. In Figure 4.4 keratin-EGCG total buried area versus the simulation time at 318 K is presented. After 47 ns EGCG at 318K detached from the keratin surface. In contrast, at 298 and 308 K, EGCG always remained bound to the keratin surface.

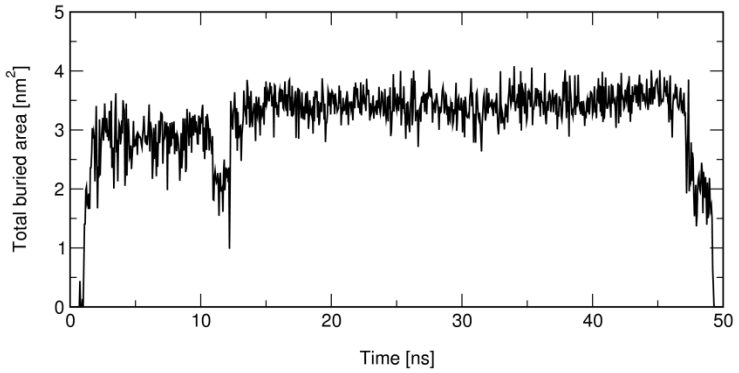


Figure 4.4. The buried surface area between keratin and EGCG at 318 K. After 47 ns EGCG molecule detached suggesting weaker interactions at higher temperatures.

4.4.2. Multilayer Self-Assembly Simulations

In the simulation of 20 EGCG molecules and keratin coiled fragment (Table 4.1. – Simulation no. 10), EGCG molecules tended to bind to keratin as clusters. Snapshots from this simulation over the 200 ns are presented in Figure 4.5. Initially, individual EGCG molecules were placed 6 nm away from coiled coil part and away from each other. Within 10 ns, EGCG aggregated into small clusters (2-3 EGCG), and further clustering was observed by 40 ns (~4-8 molecules) driven primarily by hydrophobic attraction. Within the next 30 ns, three clusters of 6-7 EGCG were formed and one adsorbed onto the protein surface. After 120 ns, a cluster of all 20 molecules were attached to the coiled coil, and gradually spread over the protein surface. The snapshot at 200 ns (Figure 4.5) shows an elongated EGCG cluster adsorbed onto the protein as multilayer.

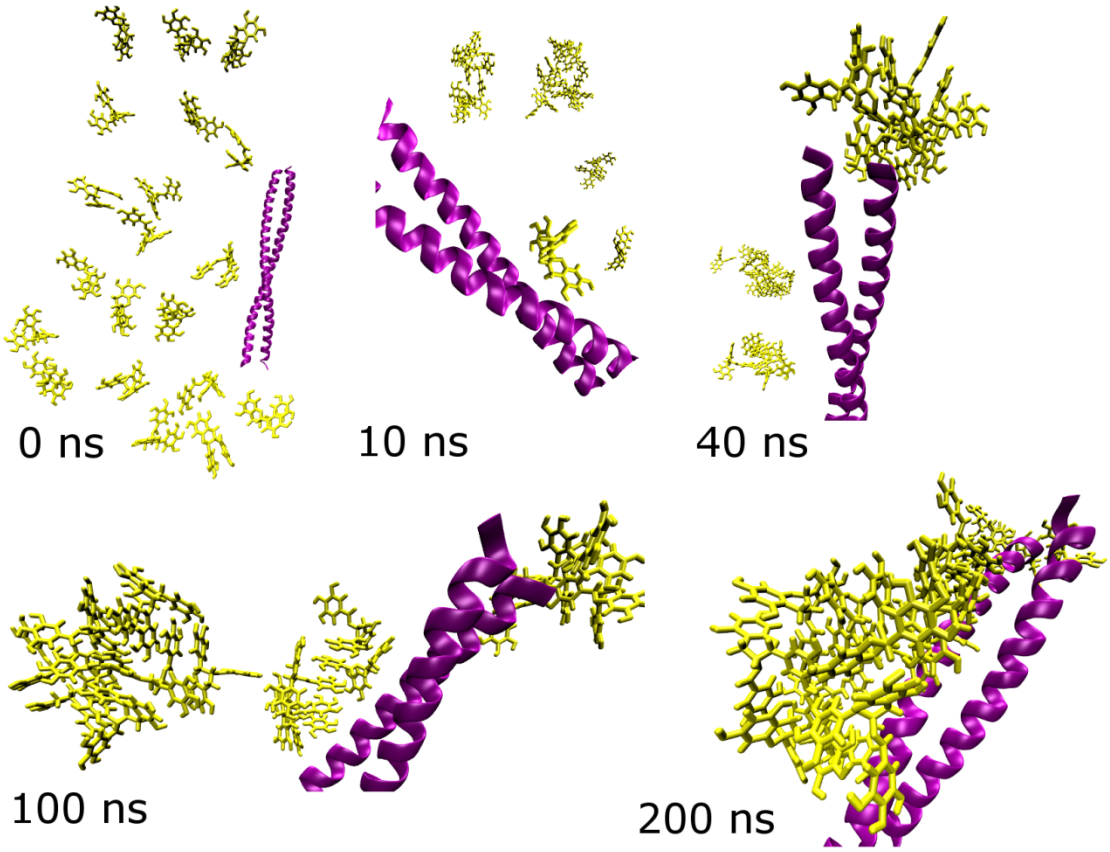


Figure 4.5. Conformations throughout the simulation time of keratin coiled coil segment (purple) and 20 EGCG molecules (yellow) at 298 K. The formation of clusters is observed by further spreading EGCG molecules over the keratin surface as a multilayer assembly.

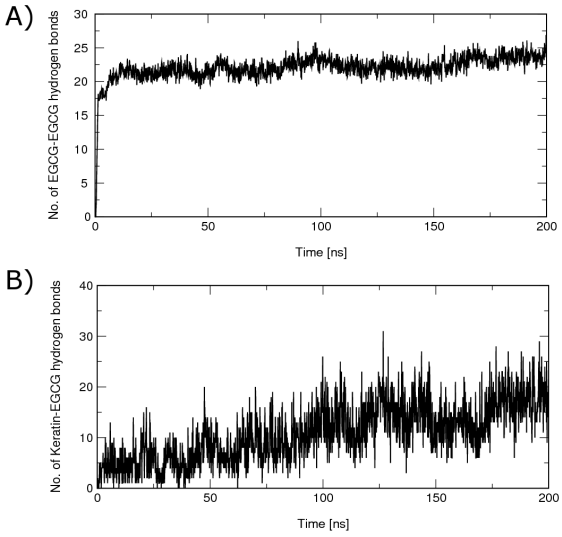


Figure 4.6. Binding of 20 EGCG molecules to keratin segment 1B at 298K: A) The number of hydrogen bonds between 20 EGCG molecules; B) The number of hydrogen bonds between keratin and 20 EGCG molecules at 298K. At the constant number of hydrogen bonds between EGCG molecules the increase in EGCG-keratin hydrogen bonds is observed. This proves the spreading of EGCG molecules over the keratin surface as a multilayer assembly.

In Figure 4.6 the number of hydrogen bonds between all 20 EGCG molecules and keratin-EGCG over the simulation time is presented. It may be observed that EGCG molecules bind to each other within first 2 ns via hydrogen bonds forming small clusters by further formation of big aggregates. The number of hydrogen bonds between EGCG molecules and the coiled coil part of keratin show an increasing number due to the spreading over the keratin surface. That is confirmed by covering the EGCG hydrophobic area (Figure 4.7). As nearly 50% of SC keratin residues are nonpolar, the hydrophobic collapse followed by hydrogen bonding optimization represented the primary mechanism for EGCG aggregation and binding to keratin. It can be therefore confirmed that the formation of EGCG multilayer is in agreement with previous investigations of ultrafiltration and HPLC experiments in mucin-EGCG system[23] and simultaneous keratin-EGCG experimental approach presented in Appendix I.

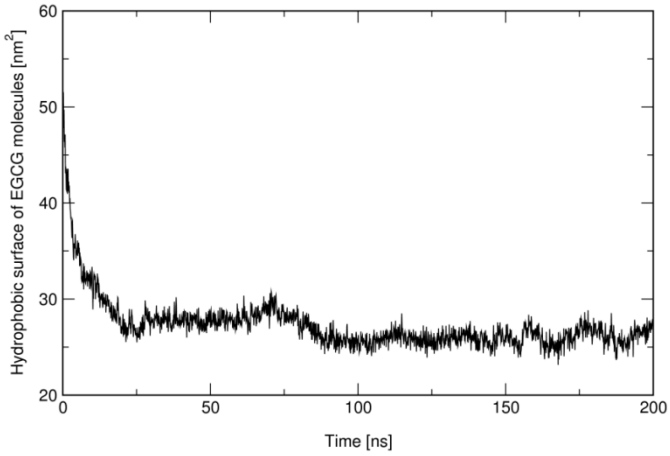


Figure 4.7. Hydrophobic surface area of 20 EGCG molecules during the simulation of binding to keratin (segment 1B). EGCG molecules cluster and elongate on the keratin surface which results in decreasing their overall hydrophobic area.

4.4.3. Assessment of Biased MD Simulation Protocol

As previously mentioned the ligand pulling direction was conserved in all protein-ligand conformations and corresponded to the axis perpendicular to the protein helical axis. An example of SMD simulation is shown in Figure 4.8. The first force maximum was reached at 0.105 ns corresponding to the breakage of the first hydrogen bond with a value of 77 kcal mol^{-1} nm^{-1}. The second maximum of the pulling force was at 0.171 ns

with a value of 75 kcal mol^{-1} nm^{-1}. For such SMD trajectories, the breaking point of the complex is defined as the highest value of the reached force.

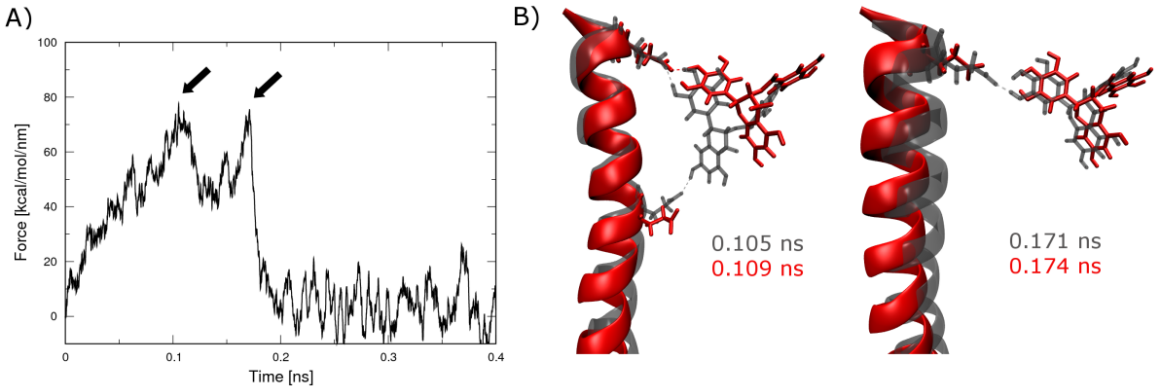


Figure 4.8. An example of steered molecular dynamics simulation of pulling EGCG from keratin surface: A) Pulling force versus time - Two breaking maxima correspond to hydrogen bond breakage at 0.105 ns and 0.171 ns; B) Snapshots from steered molecular dynamics simulation corresponding to the first and second maximum pulling force respectively.

As the initial coordinates for US windows were obtained from SMD simulations, the effect of pulling rate and virtual spring constant in SMD on the binding free energy was assessed. EGCG was pulled away from segment 1A at 298 K for 0.4 ns at 10 nm ns^{-1} and for 4 ns at 1 nm ns^{-1} using harmonic spring constants of 120 kcal mol^{-1} nm^{-2} and 360 kcal mol^{-1} nm^{-2} respectively. By using the spacing of approximately 0.2 nm along the reaction coordinate (distance) for both SMD simulations, the initial configurations for US windows were extracted. Subsequently, 20 US window simulations of 50 ns each were calculated using a harmonic biased potential for both cases, with a spring constant of 120 kcal mol^{-1} nm^{-2}. The WHAM method was then applied to obtain the final PMF curve neglecting the first 3 ns of each window for the system equilibration. With both different pulling rates and spring constants nearly the same values (0.15 kcal/mol difference) of binding free energy difference were achieved. The corresponding plots of force versus time are presented in Figure 4.9 A and B as well PMF profiles in Figure 4.10.

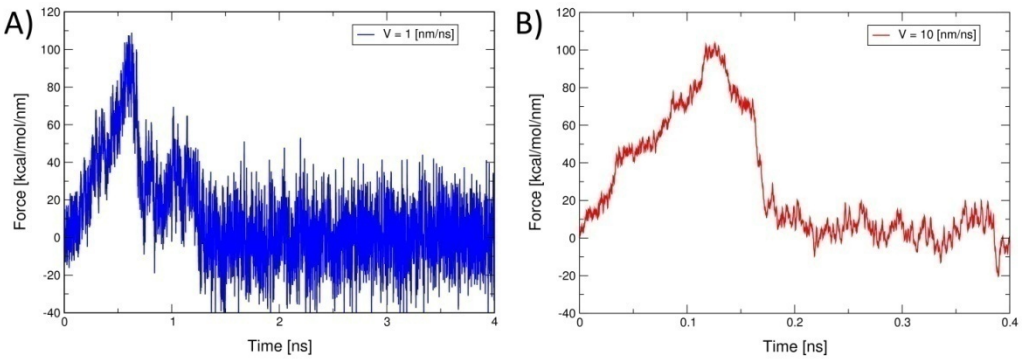


Figure 4.9. Force versus time of pulling EGCG 4 nm away from keratin 1 (α-helical segment 1A) at 298K using: A) velocity of 1 nm ns^{-1} and spring constant of 360 kcal $mol^{-1}nm^{-2}$; B) velocity of 10 nm ns^{-1} and spring constant of 120 kcal $mol^{-1}nm^{-2}$.

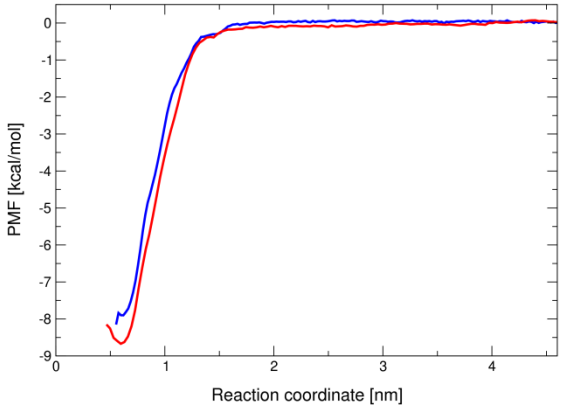


Figure 4.10. Potential of mean force profiles obtained from windows generated by different pulling simulations of EGCG (Figure 4.9) from the same initial coordinates of keratin-EGCG complex. Nearly identical values of the free energy difference were achieved.

Hence, the fast pulling rate of 10 nm ns^{-1} was used in all subsequent keratin-EGCG SMD simulation systems, with a spring constant of 120 kcal mol^{-1} nm^{-2}. The corresponding methodology for calculating the binding free energy difference (0.2 nm spaced 50 ns simulations with harmonic spring constant of 120 kcal mol^{-1} nm^{-2}) was applied. The examples of the convergence of the PMF as well as umbrella sampling histograms showing excellent overlap is presented in Fig 4.11 A and B respectively.

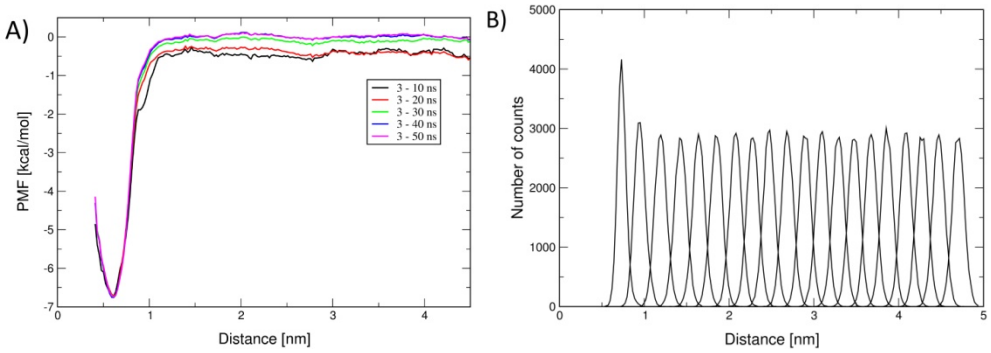


Figure 4.11. A) The convergence of the potential of mean force curves. By increasing the simulation time in each window a reasonable convergence is achieved. Hence, in all systems the umbrella sampling windows were carried out for 50 ns; B) Umbrella sampling histograms along the reaction coordinate (distance). 20 umbrella potentials with spacing of 0.2 nm and 50 ns of sampling show excellent overlap.

To confirm the reproducibility of fast pulling for each system, SMD was performed three times with the same pulling conditions. An example of SMD simulation (force versus time) for removing EGCG from the keratin coiled coil fragment at 308 K from the same initial coordinates is presented in Figure 4.12. Nearly identical trajectories were produced with approximately the same rupture force (maximum pulling force).

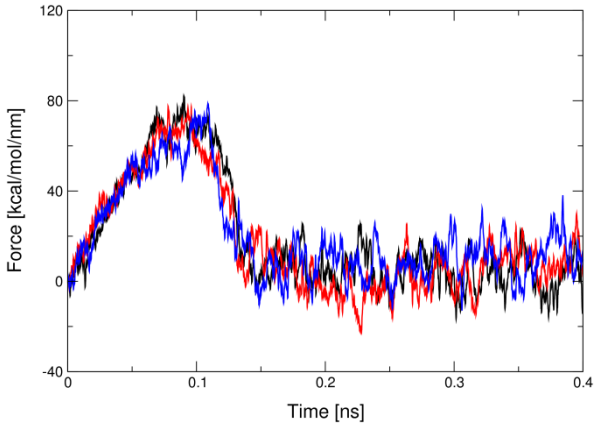


Figure 4.12. Force versus time for pulling EGCG 4 nm away from keratin coiled-coil at 308K using pulling rate of 10 nm ns^{-1} and spring constant of 120 kcal $mol^{-1}nm^{-2}$. Three very similar profiles were obtained from the same initial coordinates providing approximately the same breaking point (maximum) confirming simulation reproducibility.

4.4.4. Steered MD and Umbrella Sampling

The free energy calculations were performed based on the results from monomeric (one EGCG molecule and given keratin segment) self-assembly simulations. From all 17 self-assembly simulations at 298K and 308K we extracted the final coordinates of the EGCG-keratin complex and ran 17 SMD simulations at given temperature to pull the EGCG away from the keratin surface. As at 318K in 3 cases EGCG spontaneously detached, the same procedure with 14 self-assembly results was followed. For a given temperature, all maximum pulling forces (breaking points) were compared and the ones with the highest, lowest and median values of the maximum force were chosen. Those SMD trajectories were used to extract the initial coordinates for free energy calculations. 5 SMD trajectories for 298K, 3 for 308 K and 3 for 318 K were used. In Figure 4.13 results from SMD simulations and US are presented. The force versus time profiles and corresponding PMF curves (the same colours) can be observed for a given temperature. The energy minima of PMF curves correspond to the initial coordinates at which EGCG was adsorbed to the keratin surface. As the distance between those two groups increased resulting in the PMF increase, eventually reaching a plateau between 1-2 nm in all cases.

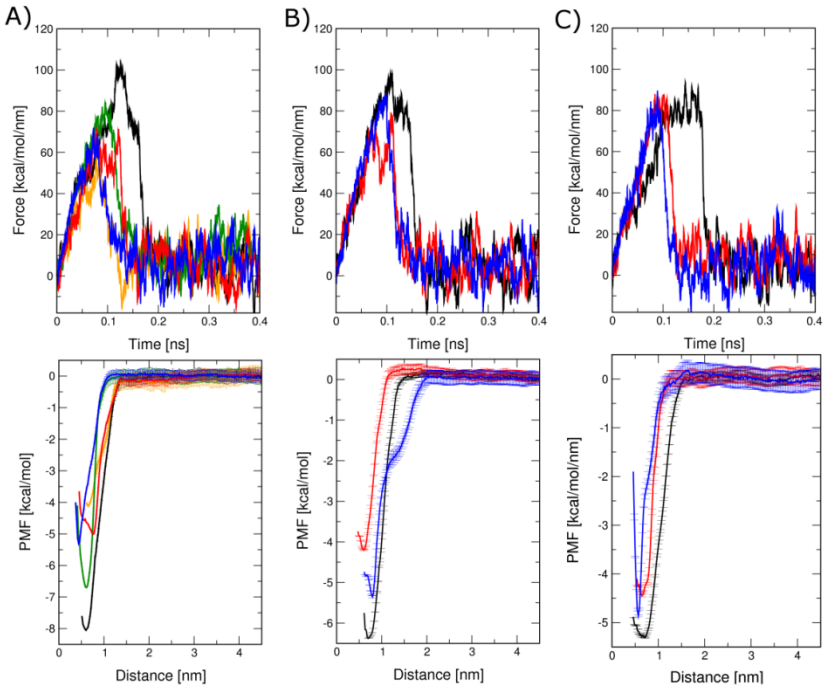


Figure 4.13. Force versus time profiles of pulling EGCG away from different keratin segments (1B – black and green; 1A – red and orange, 2A – blue) and corresponding potential of mean force curves at: A) 298 K; B) 308 K; C) 318K.

The correlation between the average number of hydrogen bonds and buried EGCG/keratin surface area versus binding free energy is presented in Figure 4.14. At 308 and 318K the correlation provided the R^2=0.94-0.96 suggesting a predictive model of the free energy based on the average number of hydrogen bonds. However, at 298K the point corresponding to the free energy (ΔG) value of -5.3 kcal mol^{-1} obtained from one simulation provided only 0.87 of the average number of hydrogen bonds significantly weakening the correlation (R^2=0.81). This data point corresponds to the case where aromatic stacking of EGCG with tyrosine/phenylalanine residues was dominant (conformation shown in Figure 4.3 A). Aromatic residues which constitute approximately 10% of the SC keratin type 1 and 10 amino acids are thus very important in the initial stabilization/assembly of EGCG molecules on the keratin surface. Hence it is not possible to predict the binding free energy based on the average number of hydrogen bonds due to the presence of the aromatic interactions.

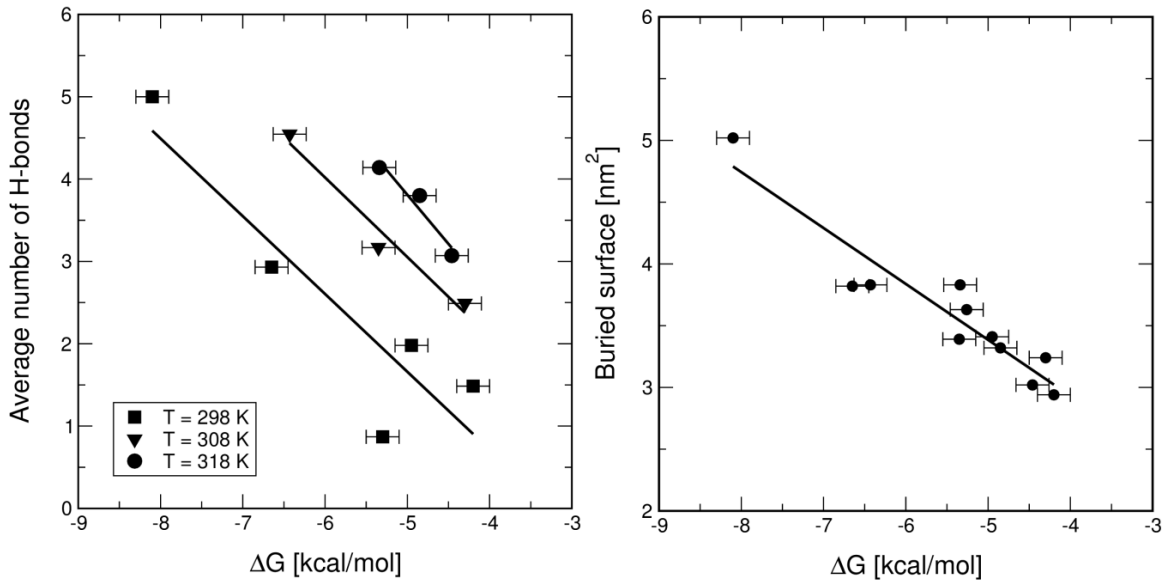


Figure 4.14. The average number of hydrogen bonds between EGCG and keratin versus the binding free energy difference (at 298 K: R^2 = 0.81, at 308 K: R^2 = 0.96, at 318 K: R^2 = 0.94) as well as EGCG/keratin buried surface area (determined from the NPT equilibration prior pulling simulations) with respect to the binding free energy difference (R^2 = 0.88).

Figure 4.14 B also shows that the higher the value of the buried surface area between EGCG and keratin, the greater the magnitude of ΔG, at all three temperatures. In all cases, EGCG expelled water molecules from the keratin surface and became directly bound to the keratin during adsorption, suggesting the potential for a favourable entropic contribution to the binding process. An example of decreasing number of hydrogen bonds between EGCG and water over the binding process to keratin is presented in (Figure 4.15).

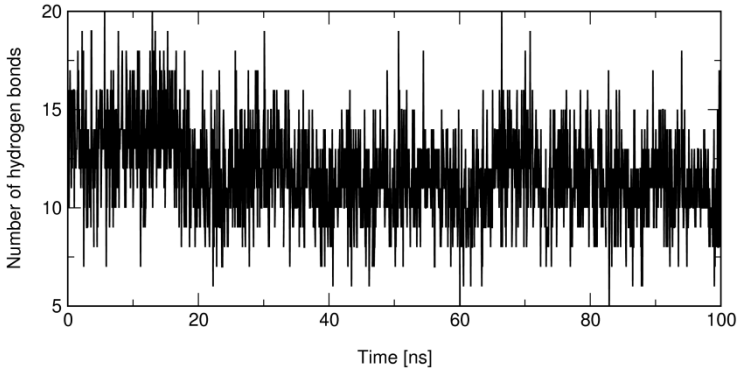


Figure 4.15. The example of the number of hydrogen bonds between EGCG and water molecules decreasing over the simulation time at 298 K. EGCG during the binding process expelled water molecules from the keratin surface.

As described above, from each SMD simulation, the maximum pulling force is obtained. A linear correlation of the rupture force (obtained in all cases at the same pulling rate) versus the free energy ΔG of binding is observed in Figure 4.16. With a single pulling reproducible (similar rupture forces obtained – Figure 4.12) SMD simulation for 0.4 ns it is possible to estimate the binding affinity in this system instead of time consuming US calculations (20 windows of 50 ns, 1 μs per PMF curve). The implication is that based on the strong linear correlation it may be possible to estimate the free energy for other small molecules binding the most common α-helical motif via hydrogen bonds and/or hydrophobic collapse using single SMD simulation.

The binding free energy at a given temperature was estimated on the basis of the maximum and minimum observed ΔG values, corresponding to ΔG_{298} = -6.2 kcal mol^{-1}, ΔG_{308} = -5.4 kcal mol^{-1} and ΔG_{318} = -4.9 kcal mol^{-1} respectively. The binding affinity decreases as temperature increases due to the increase in the kinetic energy (mobility) of molecules, hence leading to weaker interactions. The experimental validation has been conducted by Yanyan Zhao (China Agricultural University) and is presented in Appendix I. The value obtained by ITC experiment at 298 K resulted in ΔG_{Exp}= -6.37 kcal mol^{-1} and provides excellent agreement to the result

obtained via MD simulations. A multilayer assembly of EGCG molecules has been observed within ultrafiltration which confirms computational findings.

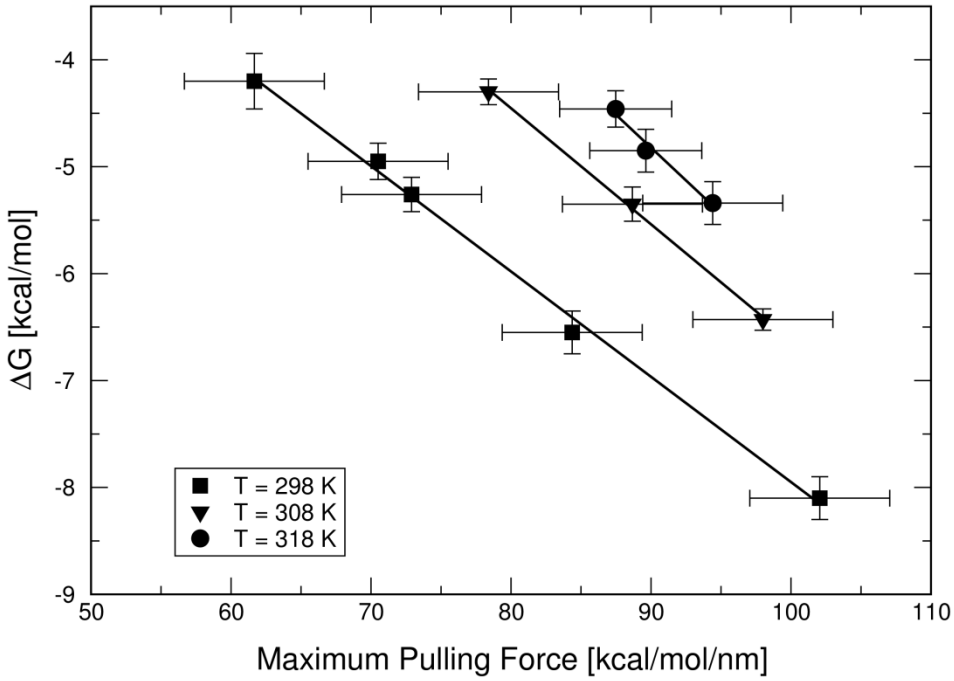


Figure 4.16. The linear correlations between the maximum pulling force (each value obtained from 0.4 ns of steered molecular dynamics simulations) and the free energy obtained by umbrella sampling (each point corresponds to 20 windows of 50 ns which corresponds to 1μs of sampling) calculations (R^2 = 0.98 in all cases).

4.5. Conclusion

The biological activity of EGCG mainly depends upon its equilibrium concentration, the binding free energy. SC constitutes the first barrier for transdermal permeation of active molecules. Due to the high content of keratin in SC (80% of its dry mass) the strength of the keratin-EGCG interaction regulates its concentration in the deeper skin layers. An understanding of those molecular interactions of the most active green tea polyphenol (EGCG) with skin protein requires characterization of the thermodynamic properties of the system as well as atomic-scale structure and dynamics. In Appendix I the experimental results are presented for EGCG binding keratin. The multilayer isotherm based on ultrafiltration and HPLC has been reported. In addition, the binding free energy of ITC experimental data at 298 K (ΔG_{Exp} = -6.37 kcal mol^{-1}) was obtained. Hence, the aim of this work was to validate and provide more detailed understanding of keratin-EGCG interactions using MD

simulations i.e. type of interactions governing the binding process, residues with the highest binding affinity and the influence on temperature on the binding free energy.

Previously, EGCG was shown to bind vimentin (type III intermediate filaments) by affinity chromatography, two-dimensional electrophoresis, and mass spectrometry.[74] For the first time in this work, it has been shown that EGCG binds to the helical and coiled coil part of type I and II human keratin intermediate filament (IF) expressed in the stratum corneum. MD simulations of various keratin-EGCG interaction systems have been performed for a total of 18 μs starting from different initial configurations at three different temperatures. Computer simulations revealed that hydrophobic assembly and in many cases aromatic interactions followed by hydrogen bonding are the key mechanisms governing the binding of EGCG to keratin. This result is with agreement with experimental studies where EGCG bound to serum albumin by hydrophobic collapse and hydrogen bonding.[192] In this chapter, using MD the significance of the interactions with acidic amino acid side chains with EGCG has been also shown. The previously experimentally suggested[23] multilayer binding of EGCG to mucin via both hydrogen bonds and hydrophobic collapse is confirmed to be responsible for binding in the EGCG-keratin system. Similarly, MD simulations also confirm the multilayer assembly of EGCG on the keratin surface revealed by ultrafiltration experiment (Appendix I). In those MD simulations, EGCG molecules formed clusters which bind to keratin and spread over its surface.

Although in the free energy calculations the initial coordinates for US windows are pulling pathway dependent, with the conserved reaction coordinate in all protein-ligand complexes (distance perpendicular to the helical axis) a good agreement between experiment and simulations was achieved. With the conserved reaction coordinate the linear correlation between breaking points (rupture forces of different EGCG-keratin complexes) from short EGCG pulling simulations and the binding free energy difference from time consuming US simulations was observed. This indicates that it may be possible to make good estimates of the binding free energy for other small molecules to the alpha helical keratin surface from SMD simulations alone. MD simulations also indicated that the increase in temperature results in a decrease in the EGCG binding affinity towards keratin and the average number of hydrogen bonds between them which is in agreement with the EGCG-mucin system.[23] In this work, all MD results providing better understanding of keratin-EGCG interactions confirmed experimental part reaching excellent agreement.

CHAPTER V

Fe^{2+} - α-Helix Wool Keratin Complex Studies via Molecular Dynamics

5.1. Introduction

Ferrous ion is plays a key role in a number of biological processes such as dioxygen transport, electron transfer, oxidation as well as stabilizing drug-protein complexes.[10,12,193] Alpha helix is the most common protein secondary structure and accounts for approximately 30% of globular proteins.[194] Keratin, the main structural protein of wool, skin, hair, nails is mainly composed of α-helices hence used in these investigations. The atomic level study of structural behaviour i.e. binding sites and the role of solvent effect as well as the equilibrium thermodynamic properties of iron cations binding to alpha helical motif are essential to understand the equilibrium concentration of the free iron. Further investigations are necessitated by a detailed description of the thermodynamics i.e. equilibrium binding properties (free energy, enthalpic and entropic contribution) thus the maximum reachable concentration as well as detailed atomic conformations i.e. binding sites and the role of water in the binding process. Due to the advantage of yielding atomically-detailed data regarding the assessment of binding sites and associated conformational changes within proteins, MD simulations of ligand-protein interaction have recently provided complementary information.[195-197] A study[83] using quantum mechanical – molecular mechanical (QM/MM) approaches reported the possibility of three different interactions between ferrous ion and an α-helical peptide (Figure 5.1). However, this simulation was based upon an ideal protein environment and neglected solvent effects. Both the binding thermodynamic properties and structural binding site assessment between α-helix and ferrous ion remain unknown and never been studied within MD simulations. The bioavailability of iron compounds is also determined by their solubility. High water soluble iron compounds, such as ferrous sulphate and ferrous gluconate, were proved to have higher bioavailability than poorly water soluble ones.[198] Therefore, ferrous

gluconate was chosen here to study the binding affinity with alpha helical rich keratin of wool according to experimental results presented in the Appendix II.

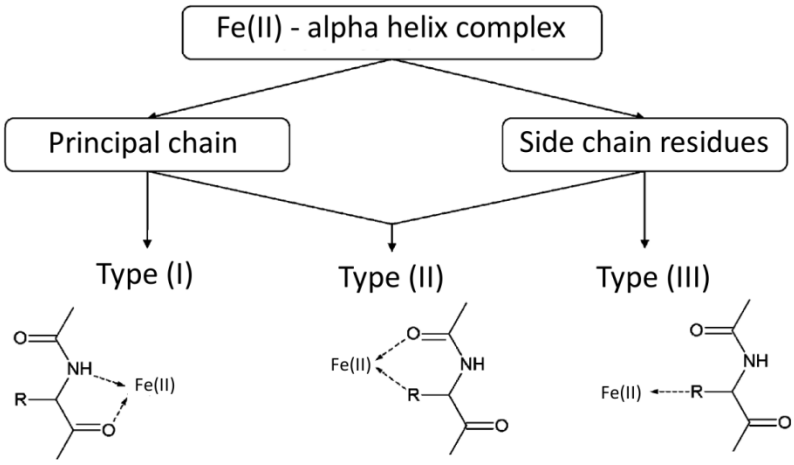


Figure 5.1. Three possible types of binding site hypothesised by Jurinovich et al.[83]

The interactions between iron (II) ion and keratin have never been studied within MD simulations hence the challenge lies in choosing the optimal way of treating ferrous ion at the atomistic level.

In computer simulations various models representing ions exist. This involves several polarized force fields[236-238] which takes into account the polarization effects and charge transfer as well as: bonded[264], cationic dummy[248] and nonbonded models.[263, 251] In the bonded model the ion is treated as a covalent part of the ligating residue which involves bond, angle, dihedral angle, point charge and van der Waals (vdW) parameters. However, with covalent bond the binding free energy within MD simulations cannot be accurately estimated. In the cationic dummy model the ion with its vdW parameters does not create any covalent bonds but the charges are placed between ion and ligands to mimic the directionality of valence bonds.[248] The nonbonded model involve its overall charge and vdW parameters being flexible with its coordinates. This simplified model does not take into account polarization, charge transfer, charge distribution or covalent bonds also possible with surrounding ligands. Although many sophisticated models exist, the nonbonded model due to its simplicity is the most widely used by MD users hence applied in this study. The appropriate choice of vdW (Lennard-Jones potential within MD) parameters is thus the crucial step in realistic representation of ions in the surrounding environment. For instance, in 1990 by reproducing hydration free energies (HFE) within free

energy perturbation, Lennard-Jones parameters for metal cations interacting with water were created and adopted in early AMBER force filed.[219] The Zn^{2+} Lennard-Jones parameters produced by QM calculations fitting the Zn^{2+}–H_2O potential energy surface together with experimental results as well as HFEs are currently widely applied in MD.[218] The weak point of previously developed nonbonded models is their application with different water models and treatment of the long ranged electrostatics. In the recent study (2013)[200] by the application of the thermodynamic integration (TI), Lennard-Jones parameters for 24 metal (II) ions with incorporated the most widely used PME[186, 187] method for long range electrostatics were built. Parameters were created for the most commonly used 4 water models: TIP3P[184], SPC/E[211], TIP4P[210], $TIP4P_{EW}$.[196] In these calculations experimental HFEs, ion-oxygen distance (IOD) in the first salvation shell as well as coordination numbers (CNs) were aimed to be reproduced. However, it appeared that it is not possible to incorporate all three parameters within one nonbonded potential. Hence, the three sets of Lennard-Jones parameters for all 4 water models were created separately: one reproducing HFEs with ±1 kcal mol^{-1} accuracy without keeping CNs, one reproducing experimental IOD with good agreement and low accuracy of CNs and the third one (named CM) reproducing HFEs with ±2 kcal mol^{-1} and conserving the experimental CNs. In this chapter, aiming at the simple and computationally less expensive nonbonded model for iron (II) ion with the combination of TIP3P solvent and Charmm22/CMAP[121] force field, the CM model was chosen.

5.2. Objectives

Experimental studies of the thermodynamic properties of wool keratin interacting with ferrous ion have been quantitatively determined using ITC experiment by Dr Yanyan Zhao (China Agricultural University). These results are presented in Appendix II. Within experiment, at three different temperatures (298, 308, 318 K) free energies together with significant entropic contribution were obtained by fitting the data to the "one set of binding sites" model. Hence, the main challenge of this study is to provide more detailed molecular insight into the experimental results of ferrous ion binding helical wool keratin using MD simulations. In particular the following objectives have been defined:

- Evaluate the keratin binding site for iron ion validating ITC data,
- Validate experimental binding free energies at three different temperatures,
- Explain the significant entropic contribution obtained by ITC.

For this purpose, herein using explicit solvent within MD simulation framework the interaction between ferrous gluconate and keratin were investigated. Lennard-Jones parameters for PME electrostatics for nonbonded model with Fe^{2+} predicted within the recent study[200] – CM model (described in the introduction of this chapter) were employed. The choice arises from the incorporated reproduced experimental HFEs with conserved CNs within this simple and widely used nonbonded model. MD simulations presented in this chapter by employing simple nonbonded model for iron ion together with TIP3P solvent agreed extremely and surprisingly well with the ITC results providing better understanding of experimental data. Both experimental and simulation insights resulted in the published paper.[199]

5.3. Methodology – Molecular Dynamics

5.3.1. Computational Details

The structure of ferrous gluconate was built using Chimera 1.5.3.[176] In order to create the topology, the CHARMM General Force Field (CGenFF)[177-180] for organic molecules (program version 0.9.6 beta) was employed. The topology was then processed and translated into the GROMACS[181] format using in-house code. As previously mentioned the CM set of Lennard-Jones parameters for Fe^{2+} in the TIP3P[184] water model were taken from the recent study by Li et al. [200]

The keratin sample used for the ITC study was a mixture of proteins of 46 kDa and 65 kDa (type I and II) extracted from sheep wool. A survey of known keratin sequences from sheep wool was conducted. Therefore type II keratin microfibrillar component 5 (http://www.uniprot.org/uniprot/P25691) was considered with MW of 55 kDa being close to the average of the mixture of proteins used in experiment. Interactions of ferrous ion with proteins are mainly governed by electrostatic interactions. In wool, acidic and basic residues accounts for approximately 13% and 21% respectively.[201] Hence, the helical segment 1A with a similar content of charged amino acids (11.4% of acidic and 20% of basic) was chosen. The choice of this helical segment arises from the fact that the helical rod of the whole protein sequence accounts for over 61% (307 out of 502 residues). In addition head and tail domains have a very low content of acidic residues (7 out of 502 residues, 1.4%). The exposure of all residues in coiled coil conformation of two helices (sample used in experiment) is approximately equal the potential binding site accessibility in single helix. Therefore the binding free energy in units per mol can be compared between MD an ITC experiment.

Segment 1A of keratin was built using the PyMol software. The structure was then processed through the H++ online tool (http://biophysics.cs.vt.edu/H++) [202-204] to predict likely pKa values at pH=4.5, in accordance with our experiments. Ionisable keratin side chains with respect to the pKa were assigned their default, charged state. The energy of the initial structure was then minimized in vacuum via the steepest descent (SD) algorithm.

All MD simulations were performed using the GROMACS 4.5.5 [181] package with the CHARMM22/CMAP [121] force field and the TIP3P[184] water model. Classical equations of motion were integrated with a Verlet leap-frog algorithm using a 2 fs time step. The LINCS[185] algorithm was employed to constrain bond lengths. A 1.4 nm cut-off distance was used for the short-range neighbor list, updated every five steps (10 fs), and van der Waals interactions were cut off at 1.4 nm. The PME method[186,187] was used for the long range electrostatics with a 1.2 nm real space cut-off. The velocity rescale thermostat with additional stochastic term and Parinello-Rahman barostat[188] were used to maintain the temperature and pressure (at 1 bar). The initial velocities were set to follow a Maxwell distribution. Periodic boundary conditions were used in all directions. All simulations were performed on two Linux clusters at Imperial College London: Department of Chemical Engineering Computing Service, URL: imperial.ac.uk/chemicalengineering and High Performance Computing Service, URL: imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing.

5.3.2. Simulation Setup

Initially, a keratin segment of α-helix was placed in the centre of a cubic box of 512 nm^3. The termini were treated in their uncharged state being a part of the 55kDa keratin. A previous study hypothesized three possibilities for Fe^{2+} interaction modes with α-helical polypeptide fragments.[83] This involved: type I, in which ferrous ion interacted with amide nitrogen and carbonyl oxygen atoms of the polypeptide backbone; type II, in which ferrous ion interacted with backbone carbonyl oxygens together with the side chain atoms of uncharged polar residues (including serine glutamine, asparginine or threonine); and type III, in which only charged side chain functional groups (of aspartate and glutamate residues) were responsible for forming interactions. Therefore, to assess each of these possible α-helix binding site modes, Fe^{2+} were manually placed within 0.2 nm of polar atoms of each of the three mentioned possible interaction sites. In particular, type II sites involved either serine, asparagine or glutamine side chain oxygens in combination with neighbouring residue backbone

carbonyl oxygens. In the case of type III, the most likely binding sites were negatively charged acidic residues, and hence the Fe^{2+} was placed in proximity of side chain functional oxygens of either aspartic or glutamic acid carboxylate groups. In the case of the type I model, Fe^{2+} was placed in proximity of backbone polar atoms of different residues (polar, nonpolar, basic and acidic). Examples of initial configurations (type I, II and III) for ferrous-keratin complexes are shown in Figure 5.2. The scope of the study involved the assessment of the stability of each binding site in order to identify the potential binding pocket responsible for mediating the interaction of keratin with other small molecules.

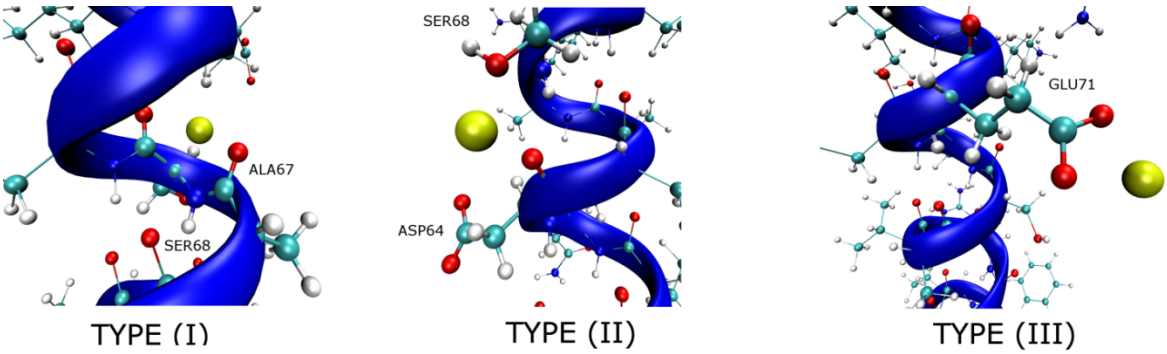


Figure 5.2. Examples of initial configurations used in MD simulations for the binding site assessment: Type I - Fe^{2+} interaction with backbone carbonyl oxygen (ALA67) and amide nitrogen (SER68); Type II - Fe^{2+} interacting with side chain hydroxyl oxygen (SER68) and backbone carbonyl oxygen (ASP64); Type III - Fe^{2+} coordinated by GLU71 side chain carboxylate oxygens. Cartoon representation of helix – blue. Atoms are shown as: red oxygens, cyan carbons, grey hydrogens, dark blue nitrogens, and yellow Fe^{2+}.

Due to the fact that ferrous gluconate is fully dissociated in aqueous solution, gluconic acid moieties and Fe^{2+} were placed separately and then TIP3P[184] water molecules filled the remainder of the box. The initial mass concentration of ferrous gluconate was set to 3% which resulted in 20 molecules. 12 ferrous ions were placed within three different hypothesized binding sites and the remaining ions and gluconic acid moieties were placed randomly around the keratin α-helical segment. To neutralize the excess system charge resulting from the presence of acidic keratin α-helix residues, one Na^{+} ion was added. Using the SD algorithm, energy minimization was then conducted to relax any steric overlap between solute and solvent. The system was then equilibrated in the canonical ensemble (NVT) for 0.2 ns and then in the isothermal-isobaric ensemble (NPT – constant number of atoms, pressure and temperature) for another 0.2 ns, allowing only water and neutralizing ions to move freely. Production simulations were then run for 100 ns at 3 different temperatures: 298, 308 and

318 K. Free energy calculations were performed using the umbrella sampling (US) method. Initial coordinates for US windows were obtained by steered molecular dynamics (SMD) simulations, in which Fe^{2+} was gradually pulled away from the α-helical segment of keratin surface. A constant velocity was applied to Fe^{2+} with a harmonic potential ("virtual spring") providing initial configurations for free energy calculations along the distance between the two species. A systematic assessment of the influence of different binding constants, as well as pulling velocities on the force versus time profiles in the SMD simulations were conducted, and are presented within the results section.

5.4. Results – Molecular Dynamics

5.4.1. Assessment of Binding Sites

The three types likely binding sites (I, II and III) for Fe^{2+} binding to protein reported in the literature [83] were assessed based on 100 ns simulation trajectories. In Figure 5.3, the most stable examples of Fe^{2+} binding to α-helical segment of keratin at type I, type II and type III sites as a distance over the simulation time are presented. This distance corresponds to the one between the centre of mass of polar atom(s) of a given binding site and ferrous ion. It can be noted that only the distance of iron from type III binding site remained constant. The most stable interaction of molecular coordination structure is shown in Figure 5.4 A which corresponds to the type III in Figure 5.3.

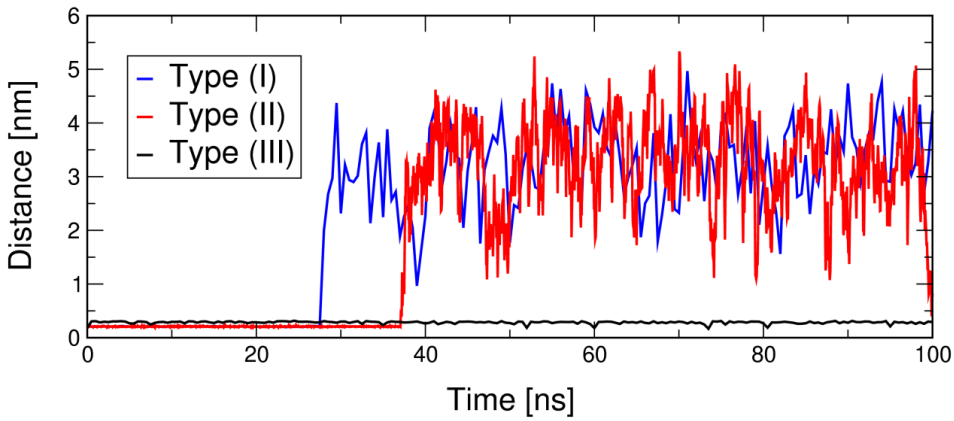


Figure 5.3. The most stable MD simulation of Fe^{2+} binding to α-helix at three suggested sites type I, II and III. Type III binding site was shown to be the most stable, with the ion remaining bound to the acidic side chain carboxylate oxygen atoms for the whole 100 ns simulation.

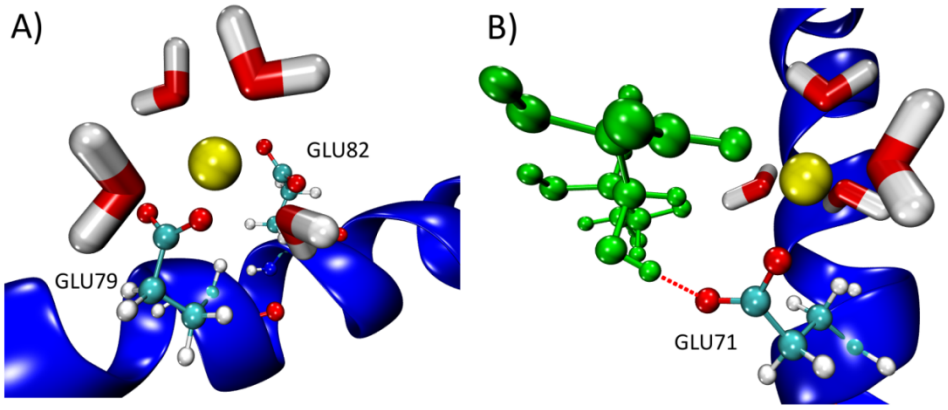


Figure 5.4. Final conformations of ferrous gluconate with α-helix: A) Fe^{2+} stacked in between two glutamic acid (GLU79 and GLU82) side chain carboxylate groups, coordinated by water molecules; B) Fe^{2+} and water molecules bound to a GLU71 side chain carboxylate oxygen. Iron mediates ionic interactions with a gluconic moiety which also forms a hydrogen bond (red dotted line) with glutamic acid; Cartoon representation of helix – blue; Fe^{2+} – yellow sphere, gluconate – green; protein atoms: oxygens (red), carbons (cyan), hydrogen (grey), nitrogen (blue).

In one case, Fe^{2+} was placed within 0.2 nm of both the side chain carboxylate oxygens of GLU57 and the neighbouring side chain hydroxyl oxygen of THR58 which corresponded to type II bind site. The corresponding snapshots together with the distance of Fe^{2+} from the centre of mass of both GLU57 and THR58 side chain oxygens are presented in Figure 5.5. More examples of type I, II and III stability, in terms of the distance from the interacting site versus time, are presented in Figure 5.6.

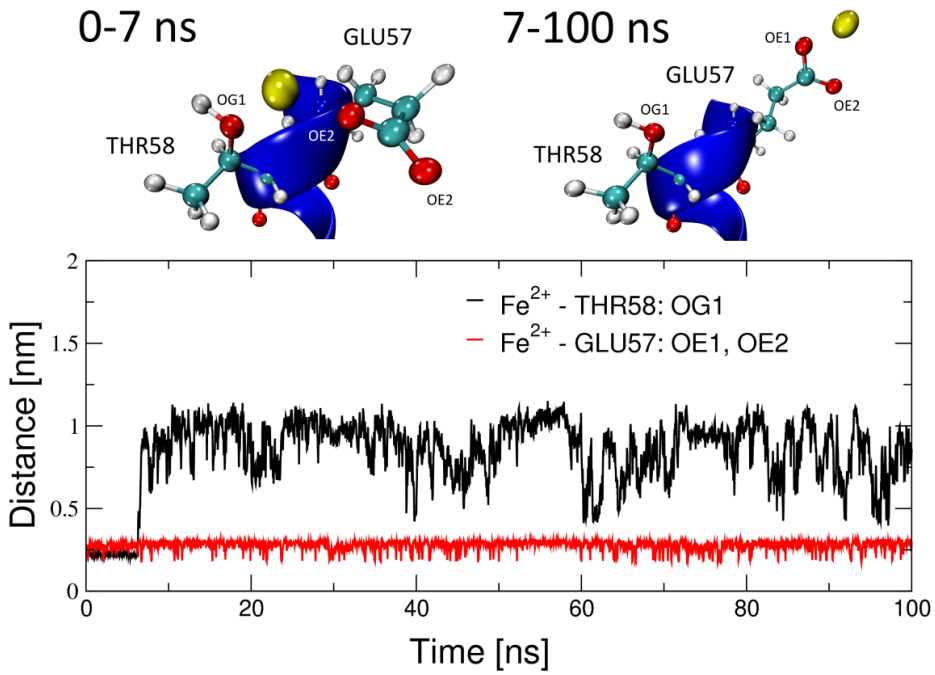


Figure 5.5. Fe^{2+} binding to α-helix at type II binding site. After 7 ns, Fe^{2+} moved away from type II binding site, initially located between GLU57 and THR58, to type III binding site occupying only carboxylate oxygens of GLU57, and remained there until the end of the 100 ns simulation. The distance with the carboxylate oxygen atoms remained stable throughout. Atoms are shown as red oxygens, cyan carbons, grey hydrogens, dark blue nitrogens and yellow Fe^{2+}.

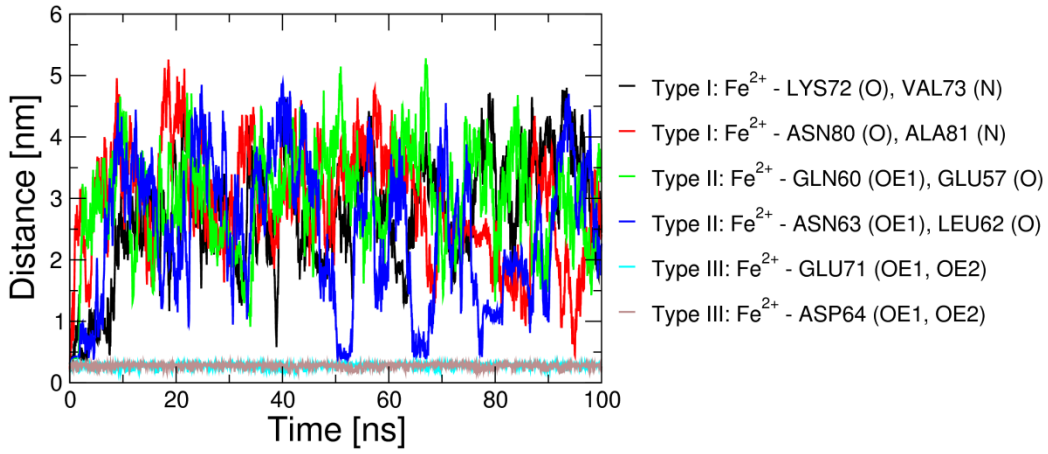


Figure 5.6. Distances of Fe^{2+} from polar atoms of binding site of type I, II and III as a function of the simulation time. Type I and II did not show any stability in contrast to type III where in all cases iron ion remained within GLU or ASP side chain oxygens. OE1, OE2 – side chain oxygens, O – backbone oxygen, N – backbone nitrogen.

A recent study [83] suggested three possible binding types between Fe^{2+} and protein amino acid residues of ferritin (as illustrated in Figure 5.1 and 5.2). The results from the molecular simulation confirmed that only ferrous ions at type III side chain carboxylate oxygens of both glutamic and aspartic acid remained stable for the whole simulation time (Figure 5.3, Figure 5.6). Ferrous ions at type I and II locations drifted away, suggesting weak interactions. This is further confirmed in Figure 5.5: Fe^{2+} remained only 7 ns in the type II configuration and then drifted towards the side chain oxygens of GLU57, remaining exclusively coordinated by the carboxylate group until the end of the 100 ns simulation. In this case, the distance of Fe^{2+} from the GLU57 carboxylate oxygen atoms remained constant for the whole simulation time.

The gluconic acid moieties of ferrous gluconate appear to have very small affinity towards wool keratin. Of 20 initially placed gluconic acid anions, only 1-2 (at 298 K) or 3-5 (at 308 K) became stacked onto the α-helix surface via ionic interactions and hydrogen bonds (Figure 5.3.B). The mediating atom of ionic interactions between α-helix and the gluconate tended to be the coordinated ferrous ion. However, those weak interactions did not remain for significant proportions of the simulation time, and hence are not further discussed here.

5.4.2. Free Energy Calculations

From the MD simulation, the final conformations of the Fe^{2+} - α-helix complex at type III binding mode (ferrous ion bound to glutamic acid side chain functional groups – Figure 5.4A) were extracted. Initial coordinates for the US windows were obtained using SMD with a constant pulling rate. First, an assessment of the effect of different force constants at the same velocity was performed. For this purpose, Fe^{2+} was pulled 2 nm away from keratin α-helical segment, with a velocity of 10 nm ns^{-1}, with harmonic spring constants of: 960, 480, 240, 120 and 48 kcal mol nm^{-2}. The force versus time profiles of these SMD simulations are presented in Figure 5.7. During the assessment of force constants at the constant pulling rates, the weaker force constants (120 and 48 kcal mol^{-1} nm^{-2}) were insufficient to extract Fe^{2+} from the keratin type III acidic pocket, and a force breaking point was not reached. With the stronger force constants (240, 480 and 960 kcal mol^{-1} nm^{-2}) a maximum pulling force corresponding to the breaking point was obtained. This rupture force corresponds to the point when the ionic bond between the ion and the acidic carboxylate group was broken. It is worth noting that with an increase of the strength of the spring, the breaking points were reached within shorter times, but

the results were otherwise comparable. Thus, an intermediate spring constant of 480 kcal mol^{-1} nm^{-2} was chosen for subsequent pulling simulations.

Similarly, to assess the influence of different pulling rates, velocities of 1, 5 and 10 nm ns^{-1} were applied, with a spring constant of 480 kcal mol^{-1} nm^{-2} (Figure 5.8). Application of different pulling velocities produced identical force versus time profiles. Hence, the fastest pulling rate of 10 nm ns^{-1} (with a spring constant of 480 kcal mol^{-1} nm^{-2}) was chosen to produce the initial coordinates for US windows.

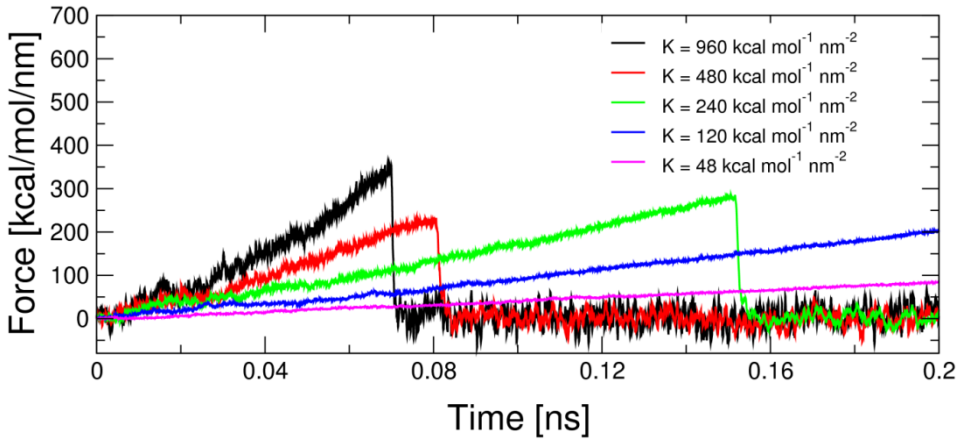


Figure 5.7. The assessment of different binding constants (48 – 960 kcal mol^{-1} nm^{-2}) in steered molecular dynamics simulations of pulling Fe^{2+} away from the acidic pocket at the constant pulling rate of 10 nm ns^{-1}. The force versus time profiles provided breaking points (maxima) with spring constants of 240, 480 and 960 kcal mol^{-1} nm^{-2}. Lower values of the spring strength did not allow extracting iron from keratin surface.

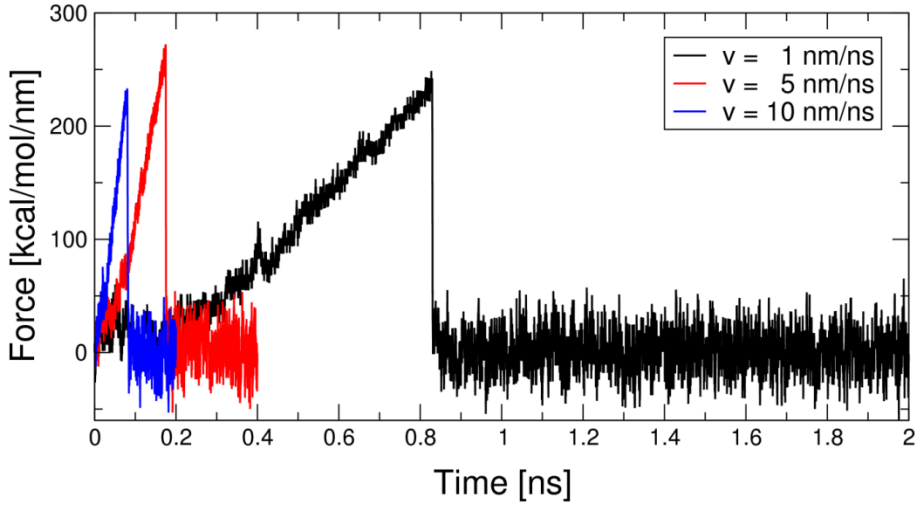


Figure 5.8. Force versus time profiles from steered molecular dynamics simulations of pulling ferrous ion away from the acidic keratin pocket with a force constant of 480 kcal mol^{-1} nm^{-2} at 1, 5 and 10 nm ns^{-1}.

The SMD trajectory was used to extract initial coordinates for free energy calculations by US method. Since the system is largely governed by ionic interactions, extensive sampling was required in order to obtain a reasonable histogram overlap. Thus, a mean spacing of 0.1 nm with 100 ns sampling per US window was used along this coordinate. As a result, 22 US windows of 100 ns each were calculated at each of three temperatures. The harmonic spring constant in the US windows ranged between 120 and 480 kcal mol^{-1} nm^{-2}. The weighted histogram analysis method WHAM[152,189] was then used to combine all windows and yield the PMF curve (free energy along the reaction coordinate). The first 10 ns of each window were omitted for system equilibration. In Figure 5.9 the PMF curves are shown at three different temperatures for Fe^{2+} binding to α-helix together with an example of US histograms, showing excellent overlap.

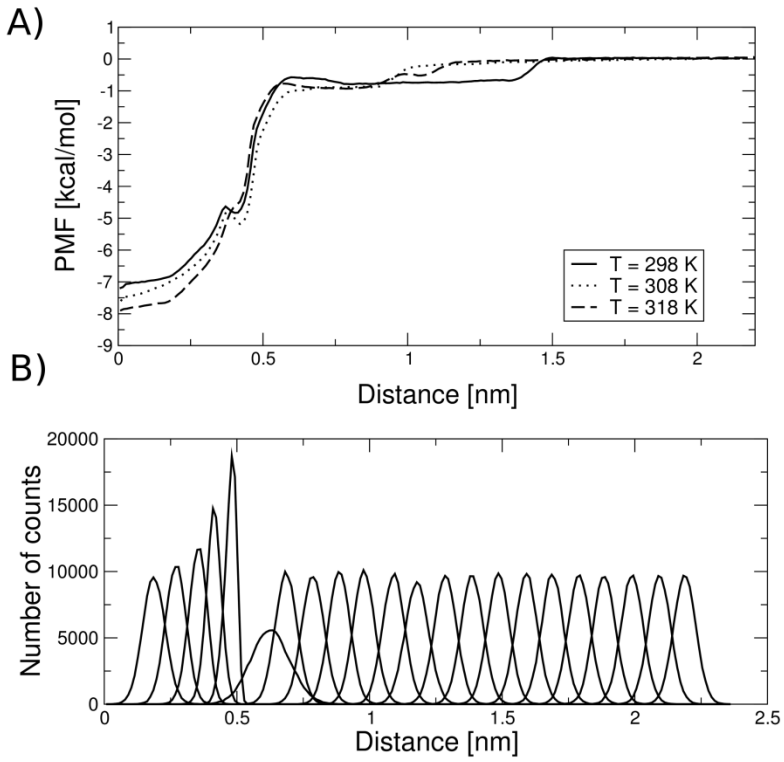


Figure 5.9. A) Potential of mean force (PMF) curves for Fe^{2+} – α-helix polypeptide segment of keratin interaction at 298, 308 and 318 K, together with B) an example of umbrella sampling histograms at 298 K.

The first PMF minimum (Figure 5.9) corresponds to the configurations in which Fe^{2+} strongly interacted with acidic carboxylate oxygens of α-helix. The second local minimum at approximately 0.4 nm represents the first coordination shell when ferrous ion was cross-linked to the acidic side chain via a water molecule, as confirmed by the step-wise increase in hydrogen-bonds between α-helix and water at this distance (Figure 5.10). The PMF then increases with distance, and reaches a plateau at around ~1-1.5 nm.

The free energy is the difference between the initial and the final state at plateau. The obtained free energies were ΔG_{298} = -7.2 kcal mol^{-1}, ΔG_{308} = -7.6 kcal mol^{-1} and ΔG_{318} = -7.9 kcal mol^{-1}, yielding a striking agreement with the experimental ITC data (Table 4.1 - ΔG_{EXP}) described in details in the Appendix II. The increasing temperature resulted in an increase of the binding energy in both experiments and MD simulations (Table 5.1). The error was estimated to be ± 0.2 kcal mol^{-1} and was assessed by dividing the whole sampling trajectory into five blocks of 20 ns and calculating the individual PMF curves. Although the excellent agreement between binding affinities obtained by MD simulations and experiment are surprising and fortunate it is shown that with the simple nonbonded model for iron ion and TIP3P water model a great accuracy can be reached.

Table 5.1. Thermodynamic parameters for ferrous gluconate binding to keratin at different temperatures of 298, 308 and 318 K obtained by isothermal titration calorimetry which is described in Appendix II. Data obtained by fitting the isothermal titration calorimetry data to equation (A.II.1) and (A.II.2). Results from molecular dynamics simulations (ΔG_{MD}) are also provided. Mean absolute error (MAE) of the free energy between experiment and molecular simulations is below 3.1%.

Temperature (K)	ΔH_{EXP} [kcal/mol]	$T\Delta S_{EXP}$ [kcal/mol]	ΔG_{EXP} [kcal/mol]	ΔG_{MD} [kcal/mol]	MAE*
298	-0.86±0.05	6.57±0.14	-7.43±0.19	-7.20±0.20	3.09
308	-1.05±0.04	6.64±0.13	-7.69±0.17	-7.60±0.20	1.17
318	-1.12±0.05	6.88±0.18	-8.00±0.23	-7.90±0.20	1.25

$^{*}\mathrm{MAE} = \left|\frac{\Delta G_{EXP} - \Delta G_{MD}}{\Delta G_{EXP}}\right| \times 100\%$

In order to show the reduced entropic cost for Fe^{2+} binding due to water displacement from the acidic type III binding site at increasing temperatures, the average number of hydrogen bonds between acidic side chains and water molecules as a function of US distance (Figure 5.9) was calculated. The hydrogen bonds were collected based on a 0.35 nm cut-off distance between donor and acceptor atoms as well as a hydrogen-donor-acceptor angle of 30˚, and then averaged over the 100 ns simulation time of each window.

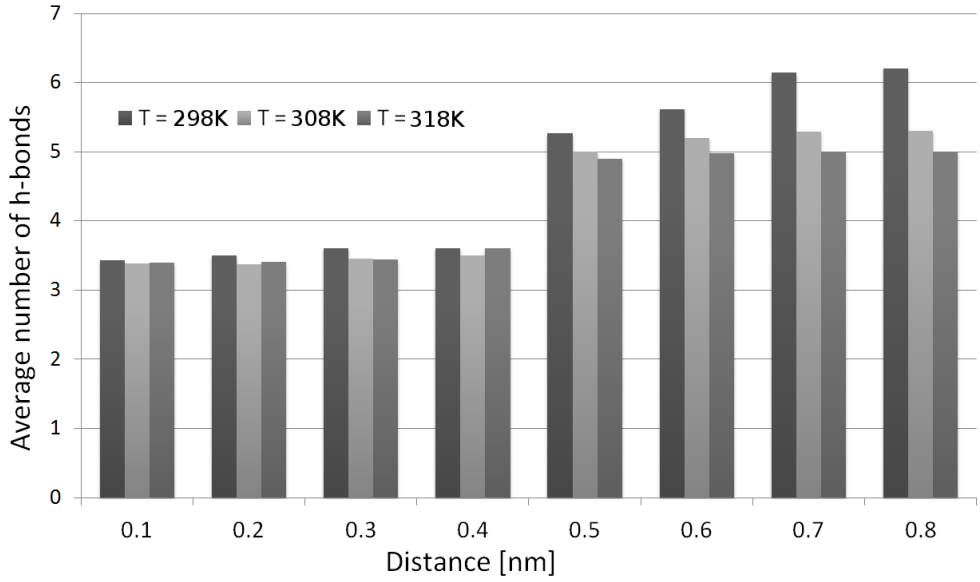


Figure 5.10. The average number of hydrogen bonds between acidic side chain oxygens and water molecules at 298, 308 and 318 K as function of the distance of Fe^{2+} and keratin α-helix acidic binding site. Hydrogen bonds were calculated from umbrella sampling windows based on a 0.35 nm cut-off distance between donor and acceptor atoms as well as a hydrogen-donor-acceptor angle of 30˚, and then averaged over the 100 ns simulation time of each window.

A previous ITC study on Fe^{2+} binding to Human H-Chain Ferritin also demonstrated large positive entropy changes.[82] The free energy obtained by their ITC measurments was nearly identical (-7.09 kcal mol^{-1}, 298 K, pH 6.5) in comparison to MD and ITC results in this work (ΔG_{MD} = -7.2 kcal mol^{-1} , ΔG_{Exp} = -7.43 kcal $mol^{-}1$ at 298 K). The large entropy change was hypothesized to be due to the displacement of water molecules bound to protein acidic residues and was further confirmed by the MD simulation. With the increase of the kinetic energy (temperature), there was a greater propensity to release the water molecules from the α-helix surface, leading to the higher entropic contribution and thus the greater magnitude in the binding free energies. The average number of hydrogen bonds between acidic pocket and water molecules within US windows in Figure 5.10 confirms the entropic findings. It is observed that within 0.4 nm, the acidic side chains were occupied by ferrous ion, thus blocking access to water. The release of ferrous ion from the α-helix surface at distances greater than 0.4 nm was associated with an increasing number of hydrogen bonds between the acidic carboxylate oxygens and water molecules. Once the distance between ferrous ion and a-helix surface became greater than 0.7 nm, the average number of hydrogen bonds remained constant. As higher temperatures resulted in a greater mobility of solvent molecules, the lower stability of associated hydrogen bonds with

protein was observed. This explains the greater entropic contribution associated with the displacement (release) of water during the binding of ferrous ion at higher temperatures.

5.5. Conclusion

Ferrous ion is plays a key role in a number of biological processes.[10,12,205] Alpha helix is the most common protein secondary structure and accounts for approximately 30% of globular proteins.[206] The Fe^{2+} - α-helix interactions have never been studied in the water environment. Previous study[83] using quantum mechanical – molecular mechanical (QM/MM) approaches reported the possibility of three different types of interactions between ferrous ion and an α-helical peptide (Figure 5.1 and 5.2) neglecting solvent effects. Both the binding thermodynamic properties and structural binding site assessment between α-helix and ferrous ion remained unknown.

For the first time in this work, the Fe^{2+} - α-helix rich keratin complex were studied by MD simulations in explicit water environment by the use of simple nonbonded model for ion. The purpose of this work was to validate and provide better understanding of experimental results conducted at the same time by Dr Yanyan Zhao (China Agricultural University) which are presented in Appendix II. ITC experiments were best fitted by "one set of identical binding sites". Herein, MD results explained that ferrous gluconate interaction with α-helix rich keratin is dominated by the binding of ferrous ion to acidic side chain carboxylate group oxygens (Type III). In other cases (Type I and II) ferrous ion drifted away suggesting weak interactions. It was observed by both experiment and MD simulations that higher temperatures facilitated the binding process. There was a great agreement between ITC and MD simulation in terms of obtained binding free energies at three different temperatures. Although the excellent agreement between binding affinities obtained by MD simulations and experiment are surprising and fortunate it is shown that with the simple nonbonded model for iron ion and TIP3P water model a great accuracy can be reached. It also appeared that the studied within MD simulation short helical polypeptide (35 residues) was representative enough in terms of the number of acidic residues in the whole protein used in experiment. The helical part in the whole protein accounts for 61%. The remaining unstructured terminals had negligible number of acidic residues thus the free energy in kcal mol^{-1} could be compare to experiment where helix only is responsible for binding. ITC data suggested the entropically driven process and that with the increase of temperature the greater the entropic contribution. This is was explained

by MD that due to the increased mobility of solvent molecules and hence lower stability of hydrogen bonds of ferrous ion with the acidic pocket. A greater displacement of water from the α-helix surface upon Fe^{2+} binding was observed at lower temperatures, confirming the importance of the entropic contribution.

CHAPTER VI

Ligand Binding Keratin α-Helix – Free energy via MD and SMD Simulations

6.1. Introduction

The binding free energy is one of the most important and desired thermodynamic properties in simulations of biological systems.[138,207] Recently, there has been an increasing number of studies focused on predicting the binding free energy of small molecules to proteins via molecular dynamics (MD) simulations in combination with experimental measurements.[2-4] The progress made suggests that MD may soon be a practical predictive tool in drug discovery. Although MD methods which involve explicit solvent often provide very accurate prediction of binding affinities[134, 173, 195, 197, 199, 208-209], they are computationally expensive. For instance, to yield well converged properties from US calculations, sampling over timescales ranging from hundreds of nanoseconds to microseconds is often required.

In drug design, steered molecular dynamics (SMD) simulations are able to provide information about mechanical properties as well as the structural changes within proteins, or in protein-protein and protein-ligand complexes. [212-215] SMD has been widely used and the results are often able to reproduce results from experimental data obtained via atomic force microscope (AFM). In SMD, a constant force or velocity is applied to a specified group of atoms attached to a "virtual spring". SMD has been shown to be a very predictive tool in drug design [216] providing binding affinity estimates using the Jarzynski equality.[217] However, this method requires hundreds to thousands pulling simulations. In chapter IV (Figure 4.16), a linear dependency between the maximum pulling force obtained from SMD simulations and free energies obtained via US (experimentally validated) for helical part of keratin interacting with epigallocatechin-3-gallate (EGCG) was observed.[134, 173] Previously, only Mai et al [136,137] proposed a similar linear correlation at constant loading rate (pulling velocity times the spring constant) between the averaged over 4 SMD trajectories rupture force (maximum pulling

force) and the experimental or MM/PBSA binding free energies. In their study different ligands binding to swine influenza A/H1N1 virus followed the linear model. Hence, being tempted by those finding, in this chapter the linear dependency for keratin-EGCG system at 298K is tested for other small molecules with varying properties. The implication of the model is that it may be possible to predict the binding free energy within couple of SMD simulations (hundreds picoseconds to nanoseconds) instead of time consuming free energy calculations (hundreds of nanoseconds to microseconds). Previous theoretical studies using the Bell model[220] have shown that the loading rate exhibits an exponential relationship with the rupture force.[221-224] Herein, the SMD simulations have been also carried out to assess this dependency for ligands with significantly ranging properties.

6.2. Objectives

Due to the fact that the a-helix is the most prevalent protein secondary structure type accounting of 30% of the average globular protein it was aimed at test our a-helix-EGCG linear dependency at 298 K (Figure 4.16: rupture force versus the free energy) for several small molecules which ranging properties i.e. overall charge (protonation state) as well as octanol/water partition coefficient which defined as the logarithm of the concentrations ratio of un-ionized compound in octanol and water solutions:

$$logP = \left[\frac{(solute)_{ocatnol}}{(solute)_{water}^{un-ionized}}\right] \quad (6.1.)$$

For this purpose, 7 ligands binding to helical keratin segment were tested, consisting of: a) oleic acid, catechin and citric acid, which vary significantly in their octanol/water partition coefficient $logP$ values; and b) citric acid in several different protonation states (hydrogen citrate, dihydrogen citrate, citrate), as well as the studied in chapter V complex with ferrous ion.[199] In addition, the aim was to establish the dependency of the applied spring constant versus the rupture force within SMD simulations.

To reach the above aims, the following objectives for all molecules within MD simulation framework have been defined:

- Assess the type of interactions for binding the keratin helical segment,
- Apply SMD protocol with ranging force constants and define the correlation of the spring constant versus the rupture force,
- Calculate the free energy for each molecule by US and compare with available experimental data,
- Check the validity of the keratin-EGCG linear correlation: rupture force versus the free energy.

For the first time, in this work it is approached to test the linear correlation of the free energy versus the rupture force (obtained from couple of SMD simulations) of small molecules with significantly varying properties, binding the most common helical motif. Initially, the type of interactions involved in binding process is assessed. Subsequently, for each molecule the correlation of the applied spring constant within SMD simulations versus the rupture force is obtained. Umbrella sampling was then used to calculate the binding free energy and is compared to existing experimental data reaching very good agreement. Finally, the linear correlation of the binding free energy versus the rupture force is obtained for molecules which do not exhibit significantly varying properties i.e. with $-1.721 < logP < 2.08$ or the maximum overall charge of ± 1. Based on this correlation within MD, it is possible to predict the binding affinity of small molecules interacting with α-helix based on couple of SMD simulations (hundreds of picoseconds to nanoseconds) instead of time consuming free energy calculations (hundreds of nanoseconds to microseconds).

6.3. Methodology – Molecular Dynamics

6.3.1. Computational Details

The ligand molecules studied here are shown in Figure 6.1. These molecules have varying octanol/water partition coefficient ($logP$) and different protonation states as presented in Table 6.1.

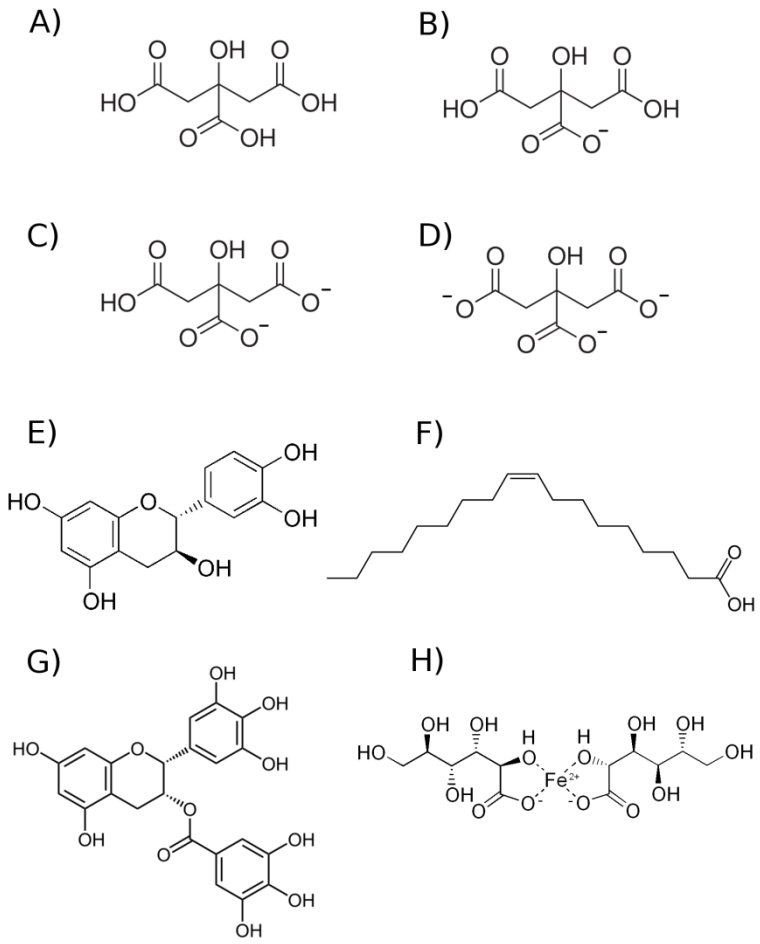


Figure 6.1. Molecules studied herein: A) Citric acid, B) Dihydrogen citrate, C) Hydrogen citrate, D) Citrate, E) Catechin, F) Oleic acid as well as studied in Chapter IV and V G) Epigallocatechin-3-gallate and H) Iron (II) gluconate respectively.

Table 6.1.

A) Small molecules evaluated in this study with varying octanol/water (*logP*) partition coefficient.

Molecule Name	CAS number	*logP*	Molecular Weight [Da]
Oleic Acid	112-80-1	7.64 [225]	282
Epigallocatechin-3-gallate	989-51-5	2.08 [226]	458
Catechin	154-23-4	0.51[227]	290
Citric Acid	72-92-9	-1.721 [228]	192

B) Charged species evaluated in this study: negatively charged protonation states (depending on the pH) of citric acid and ferrous ion. Citric acid pKa values (pK_{a1}=3.29, pK_{a2}=4.77, pK_{a3}=6.39).[229]

Molecule Name	Overall charge	pH range	Molecular Weight [Da]
Dihydrogen citrate	-1	3.29 – 4.77	191
Hydrogen citrate	-2	4.77 – 6.39	190
Citrate	-3	> 6.39	189
Fe^{2+}	+2	N/A	58

The structure of small molecules was built using Chimera 1.5.3 [176] The parameters for each molecule used the CHARMM General Force Field (CGenFF) [179,180,230] for organic molecules (program version 0.9.6 beta), translated into the GROMACS[181] format. Alpha helical segment 1A (35 residues) of keratin 83 with high contribution of charged residues (20% of basic, 11.43 % of acidic) was built using the PyMol software. [231] The amino acid sequence was provided with the Human Intermediate Filament Database.[183] Keratin segment was minimized in vacuum using the steepest descent algorithm.

In order to be consistent with previous study on keratin-EGCG, the CHARMM22/CMAP [121] force field with TIP3P[184] water model was used in all MD and SMD simulations. All simulations were performed with the GROMACS 4.6[181] package compatible with graphics processing units (GPUs). With a time step of 2 fs, equations of motion were integrated through Verlet leap-frog algorithm. Bond lengths were constraint with the LINCS [185] algorithm. The cut-off distance was 1.4 nm for the short-range neighbor list and van der Waals interactions. The PME method[186,187] was applied for the long range electrostatic interaction with a 1.4 nm cut-off. The velocity rescale thermostat with additional stochastic term[232] and Parinello-Rahman barostat[233] were used to

maintain the temperature at 298 K and pressure at 1 bar. The initial velocities were set according to the Maxwell distribution. Periodic boundaries were used in all directions. GPUs were used for self-assembly simulations and central processing units (CPUs) for US and SMD simulations.

6.3.2. Simulation Setup: α-Helix – Ligand Complexes

Initially, keratin segment 1A was placed at the centre of a cubic unit cell of 512 nm^3 (8x8x8 nm). Terminal residues remained in their uncharged state. Small molecules were placed randomly around the keratin fragment to reach 5% mass concentration. The simulation box was filled with approximately 16,000 TIP3P[184] water molecules and Na+ and Cl- ions to neutralize the charge of the system. Energy minimization was then performed via the steepest descent algorithm. Two step equilibration (0.2 ns each) in the canonical ensemble (NVT) and isothermal-isobaric ensemble (NPT) was conducted, respectively. During the equilibration, positions of protein and ligand heavy atoms were restrained allowing only water and ions to move to soak the system. The NPT ensemble at 298 K was used for the production simulation run of 200 ns. In total, 5 keratin – small molecule systems were run and are listed in Table 6.2. In system number 4 (Table 6.2) the pH was fixed to approximately 4.5 which corresponded to the contribution of dihydrogen citrate in 65% and hydrogen citrate of 35% which resulted in 52 and 28 molecules respectively, reaching the overall mass concentration of 5%.

Table 6.2. Details of MD simulations for keratin segment 1A and small molecules with their 5% mass concentration. The pH is set according to pK_a values provided in each respective reference.

System No.	Ligand	pH	Number of ligands	Simulation time [ns]	Temperature [K]
1	Citric Acid	<3.29[229]	80	200 ns	298
2	Oleic Acid	< 8.5[234]	55	200 ns	298
3	Catechin	<8.7[235]	53	200 ns	298
4	Dihydrogen Citrate/Hydrogen Citrate	4.5	52/28	200 ns	298
5	Citrate	> 6.39[229]	82	200 ns	298

In order to assess the type of interactions governing protein-ligand binding, the number of hydrogen bonds as well as the overall hydrophobic surface area of small molecules was calculated over the simulation time. Hydrogen bonds were based on the cutoff distance between the donor and the acceptor atoms below 0.35 nm and a hydrogen-donor-acceptor maximum angle of 30˚. The overall hydrophobic area of small molecules was computed by numerical-integration of their hydrophobic solvent accessible surface area (SASA)[190]. The buried area between protein and ligand was obtained by calculating the individual SASA of ligand, protein and protein-ligand complex. The most stable protein-ligand complex for each molecule was chosen based on the highest number of hydrogen bonds and/or the buried area. The protocol for those calculations is presented within the results.

6.3.3. Simulations Setup: Steered MD and Umbrella Sampling

For each of the molecule studied herein the only one (the most stable) keratin-ligand complex was chosen for further SMD and free energy simulations. All the ligands which were cluster layered as well as water and ion molecules were removed from the system leaving only protein covered with the first layer of small molecules. That arises from consistency with our keratin-EGCG study[16] where no molecules were present on the pulling pathway of EGCG. Subsequently, protein with the first layer of small molecules was placed in the edge of the cubic box which was filled with TIP3P water model and ions to neutralize the system charge. After energy minimization by SD algorithm each keratin-small molecules system was equilibrated for 1 ns in NVT and then the NPT ensemble. From NPT equilibration the buried SASA between protein and the most stable ligand were extracted. Subsequently, a constant velocity was applied to its COM and pulled 2-2.5 nm away, effectively into bulk solvent. In order to be consistent with keratin-EGCG[134] system, the pulling distance, direction (perpendicular to helical rod) and velocity of 10 nm ns^{-1} were conserved. The influence of the varying spring constants (from 1 to 2000 kcal mol^{-1} nm^{-2}) on the rupture force was applied. 10 SMD simulations with a given spring constant of each protein-ligand complex were performed. The rupture force corresponded to the arithmetic average of 10 obtained values. SMD simulations were used to extract the initial coordinates for binding free energy calculations (US). In order to generate sufficient sampling, 0.1 nm spacing along the reaction coordinate (distance) was used. Each US window ran for 50 ns with the spring constant of 480 kcal mol^{-1} nm^{-2}. The weighted histogram analysis method (WHAM)[152] was used to combine all windows into the

PMF curve to estimate the binding free energy difference. Estimation of statistical error was performed for each PMF curve by dividing the whole trajectory of 50 ns into blocks of 10 ns.

6.4. Results – Molecular Dynamics

6.4.1. Keratin-Ligand Interactions

The number of hydrogen bonds between ligands the keratin segment was calculated for each system over the 200 ns of the simulation. In addition, for molecules with *logP* >0 (oleic acid, catechin – system number 2 and 3 in Table 6.2) the number of hydrogen bonds between ligands was also collected. The number of hydrogen bonds was divided by the total number of molecules and averaged over 1 ns windows for clarity (200 points). In order to observe possible hydrophobicity-driven interaction, the hydrophobic SASA of all catechin or oleic acid molecules was calculated over the 200 ns of the simulation.

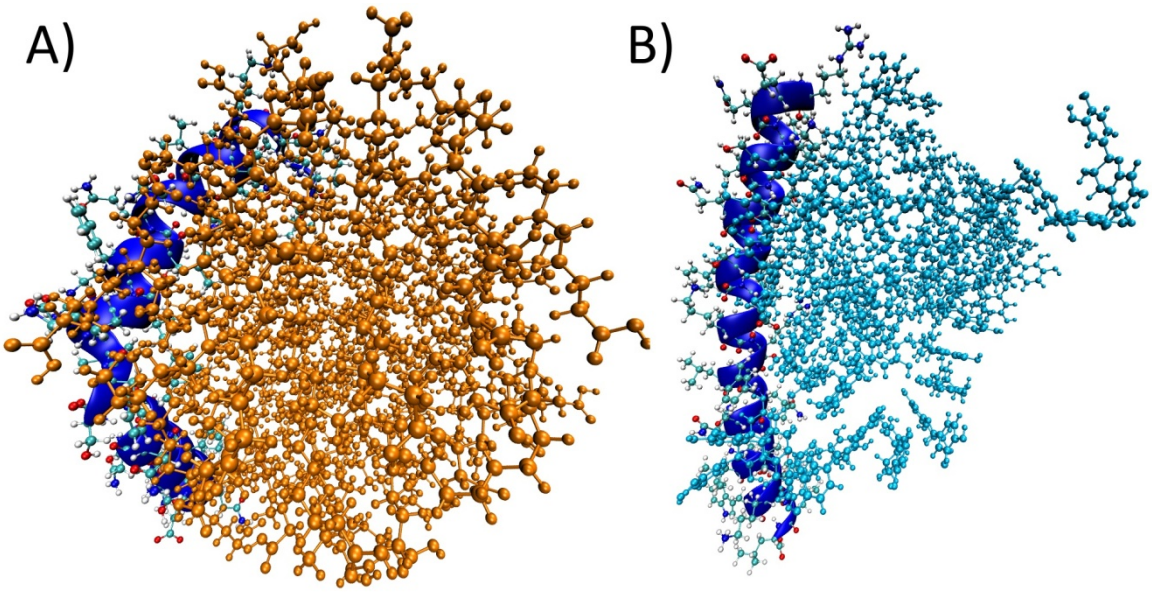


Figure 6.2. Interactions of A) oleic acids (orange) with keratin (blue) at 200 ns, and B) catechin (cyan) with keratin (blue) at 200 ns. The mass concentration of ligand in both cases is 5%.

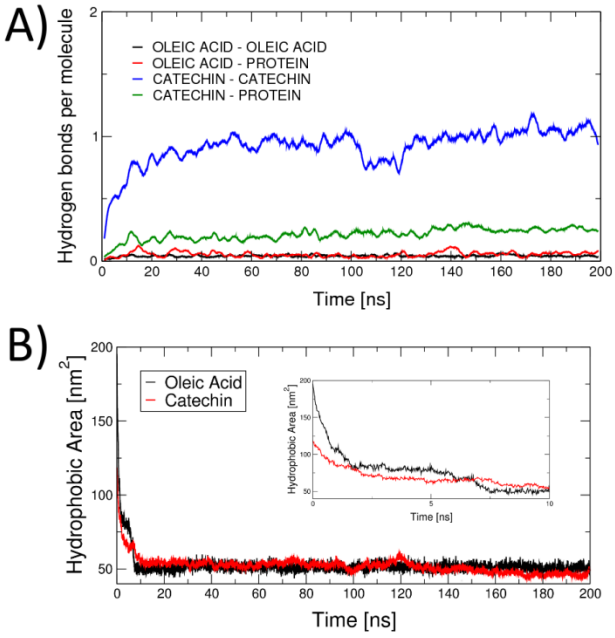


Figure 6.3. A) The number of hydrogen bonds per molecule between: oleic acid – oleic acid, oleic acid – keratin, catechin – catechin, catechin – keratin over the simulation time. B) Hydrophobic solvent accessible surface area (SASA) of all catechin or oleic acid molecules over the simulation time.

All oleic acid and catechin molecules clustered on the keratin surface forming a multilayer assembly (Figure 6.2 A and B) as previously observed for keratin-EGCG.[134, 173] In Figure 6.2 B the overall hydrophobic SASA of oleic acid and catechin is shown to decrease rapidly in the first 10 ns due to cluster formation and binding to keratin. The hydrophobic SASA of oleic acid dropped from 200 to 50 nm^2 whereas that of catechin from 120 to 48 nm^2. During the self-clustering and keratin binding process the number of hydrogen bonds per molecule of catechin-catechin molecules increased to ~1 within 10 ns (Figure 6.3 A) and a multilayer assembly was also formed (Figure 6.2 B). However, the number of hydrogen bonds per molecule within the catechin cluster and keratin varied around 0.25 (Figure 6.3 A). This shows that in addition to hydrophobic interaction, hydrogen bonding is also involved in the process of binding to keratin, which is consistent with the presence of 6 hydrogen bond acceptors, and 5 hydrogen bond donors in catechin, as well its *logP* value of 0.51. For oleic acid clusters, the number of hydrogen bonds per molecule remained below 0.1. Similarly, the number of hydrogen bonds between oleic acid assembly and keratin was low, suggesting that hydrophobic interactions dominate. The *logP* value of 7.64 for oleic acid as well as the presence of only 2 hydrogen bond acceptors and 1 donor supports our observations of hydrophobic assembly as the main driving force.

For citric acid, hydrogen citrate, dihydrogen citrate and citrate (Table 6.2 - system number 1, 4-5), the number of hydrogen bonds per molecule between ligands and keratin was calculated over the simulation time. Citric acid had the smallest number of hydrogen bonds per molecule (Figure 6.4 D). The increase of the overall charge (dihydrogen citrate of -1, hydrogen citrate of -2, citrate of -3) resulted in an increase in the number of hydrogen bonds. At pH 4.5 (Table 6.2 - System number 4) hydrogen- and dihydrogen citrate formed small salt clusters (2-5) via ionic bonds mediated by sodium ions (Figure 6.4 B). This led to stronger interactions in comparison to citric acid. For citrate (Table 6.2 - System number 5), electrostatic interactions were dominant. It was observed (Figure 6.4 C) that 26 citrate molecules formed a salt aggregate stabilized by sodium ions.

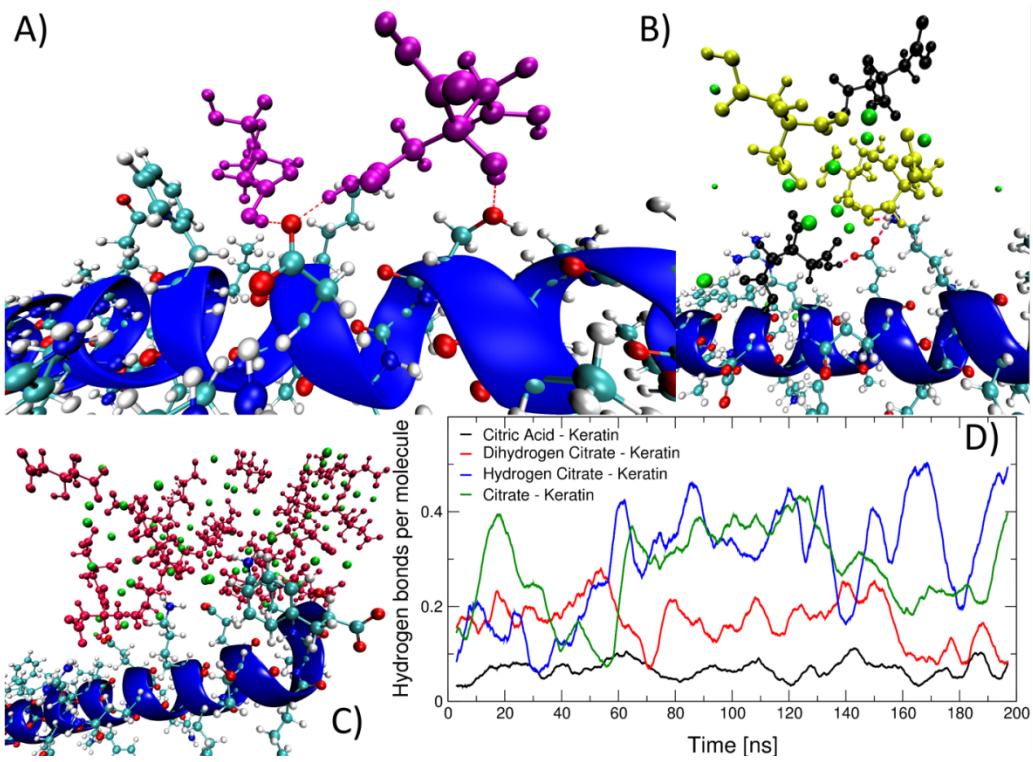
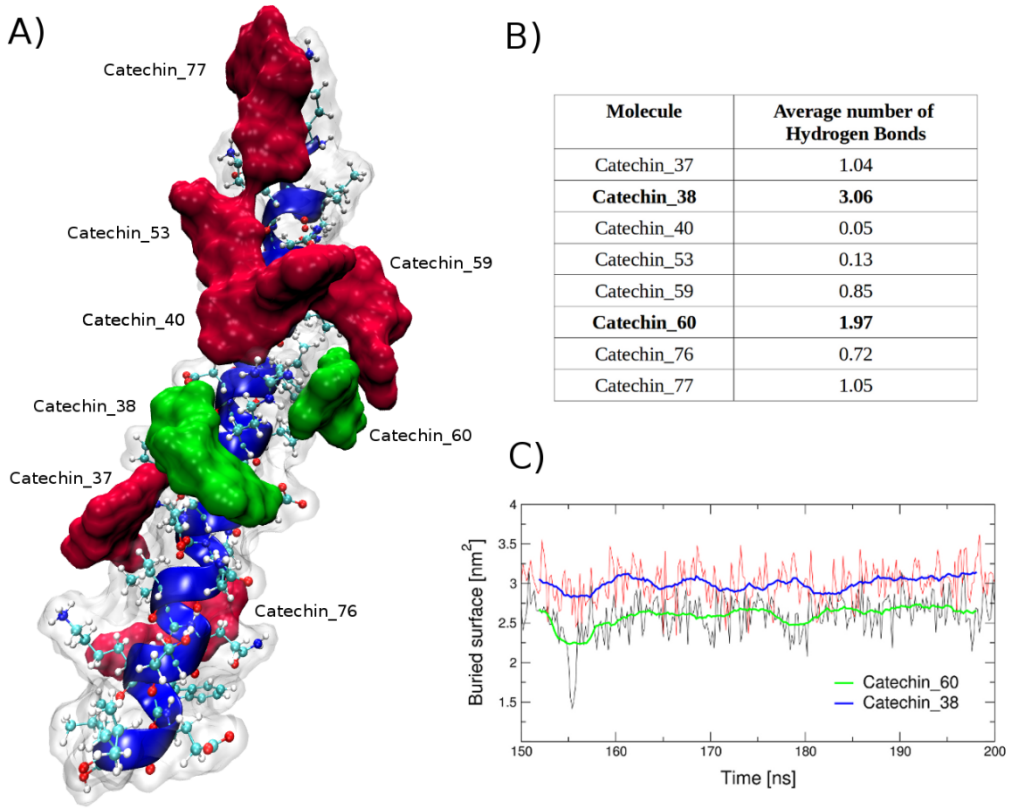


Figure 6.4. Citric acid binding to keratin at different pH values: A) pH < 3.29 – Citric acid binds to keratin via hydrogen bonds only; B) pH = 4.5 – Hydrogen- and dihydrogen citrate formed small salt aggregates via sodium ions and stacked on the keratin surface via hydrogen bonds and electrostatic interactions, C) pH > 6.39 – Citrate formed salt cluster of 26 molecules stabilized by sodium ions mediating the interactions between citrate molecules as well as citrate – keratin charged residues; D) Number of hydrogen bonds over the simulation time (averaged over 200 data points - every 1 ns). Red – oxygen, blue – nitrogen, cyan – carbon white – hydrogen), citric acid (purple), dihydrogen citrate (black), hydrogen citrate (yellow), sodium ions (green); Cartoon representation of protein – blue.

6.4.2. The most stable protein-ligand coordinates

From each MD simulation all the molecules being directly bound to the keratin surface were assessed in terms of the stability. All the cluster layered small molecules which did not directly interact with keratin surface were in those calculations omitted. The stability assessment was based on the protein-ligand average number of hydrogen bonds and/or the buried area. The protocol for the choice of the most stable configurations is presented in Figure 6.5 based on the catechin as an example.



Molecule	Average number of Hydrogen Bonds
Catechin_37	1.04
Catechin_38	**3.06**
Catechin_40	0.05
Catechin_53	0.13
Catechin_59	0.85
Catechin_60	**1.97**
Catechin_76	0.72
Catechin_77	1.05

Figure 6.5. The protocol employed for the choice of the most stable α-helix-ligand coordinates based on the catechin as an example: A) 200 ns snapshot of the molecules which are directly bound to the keratin (green surface representation for catechins with the highest number of hydrogen bonds and red surface representation for catechins remaining catechins; white surface representation, blue cartoon and space filling: red – oxygens, blue - nitrogens, cyan - carbons, white - hydrogens for helical keratin); B) The average number of hydrogen bonds calculated over last 50 ns for all molecules being within keratin surface; C) The keratin-catechin buried surface area calculated for two molecules which exhibited the highest number of hydrogen bonds. The buried surface (green and blue) was averaged over 100 data points (black and red respectively).

From the last frame at 200 ns all the catechins being directly attached to keratin are chosen (Figure 6.5 A, red and green surface representation for catechins). Subsequently, the average number of hydrogen bonds between each those molecules and keratin are calculated over last 50 ns of the simulation time and averaged over all frames (Figure 6.5 B). Two molecules (Catechin_38 and Catechin_60 - green surface representation in Figure 6.5 A) were shown to have the highest average number of hydrogen bonds and are further analyzed in terms of the keratin-ligand buried area (Figure 6.5 C). It appeared that Catechin_38 exhibited the greater contribution of the buried area (averages taken over 100 data points) and hence is chosen as the most stable configuration and used for SMD and free energy simulations (US). This stability-protocol assessment was employed for oleic acid, citric acid, hydrogen citrate, dihydrogen citrate and citrate. The keratin-ligand coordinates of EGCG and Fe^{2+} for SMD protocol were taken from Chapter IV and V of this thesis.

6.4.3. Steered Molecular Dynamics and Umbrella Sampling

Each system consisted of the keratin and the first layer of small molecules which were directly bound to keratin. The remaining clustered molecules were removed in order to be consistent with our previous study where no other molecules were present on the pulling pathway. Prior to the pulling simulations, from a 1 ns system equilibration simulation (in the NPT ensemble), mean values of protein-ligand buried surface area for uncharged molecules (catechin, citric acid, oleic acid) were obtained and are listed in Table 6.3. The pulling simulation of the assessed previously most stable ligand bound to keratin coordinates was performed at ranging spring constants. The example of the pulling simulation of the most stable catechin is shown in Figure 6.6. The pulling direction according to chapter IV and V is perpendicular to the helical backbone axis. During the SMD simulation the force reached the maximum (rupture force) at approximately 0.1 nm where the last hydrogen bond was broken. The distance between the catechin and keratin segment COM started increasing at the same time reaching the 2.4 nm at the end of the simulation.

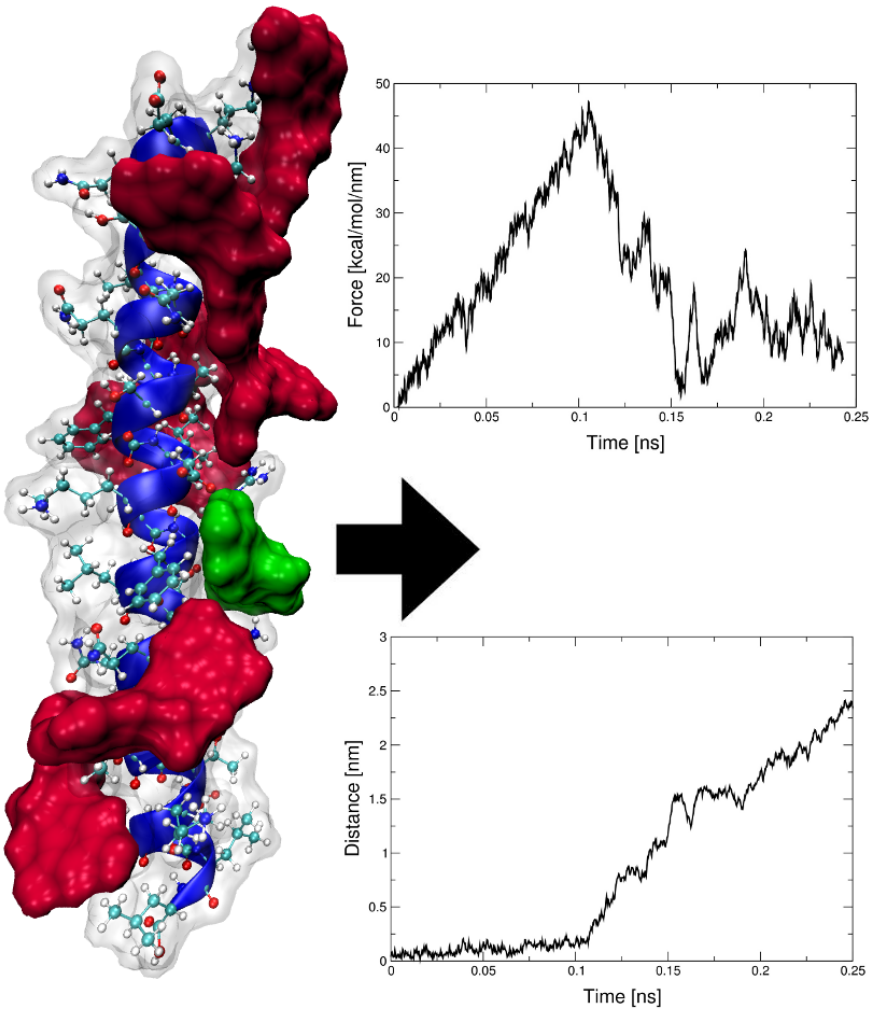


Figure 6.6. The example pulling simulation of the most stable catechin (green surface representation). The pulling direction according to our previous study[16] is perpendicular to the helical backbone axis. During the SMD simulation at approximately 0.1 ns the force reaches the maximum when the last hydrogen bond is broken. The distance between the keratin helical segment and the catechin starts increasing at this point reaching 2.4 nm at the end of the simulation. Red surface representation for remaining catechins obtained by MD self-assembly simulations; white surface representation, blue cartoon and space filling: red – oxygens, blue - nitrogens, cyan - carbons, white - hydrogens for helical keratin.

Results from the SMD protocol with ranging spring constants are presented in Figure 6.7. Those results correspond in each protein-ligand for one, the most stable coordinates assessed in the previous step. The plots present the maximum pulling force (rupture force) averaged over 10 repeated identical pulling simulations versus the used harmonic spring constant. The error was estimated as the maximum and minimum of the obtained 10 rupture forces with given spring constant. It is observed that the increase in the stiffness of the spring resulted in an increase of the maximum pulling force.

This relationship is theoretically described by an exponential function[239-242] derived from the Bell model[220] and our results correlated with such dependency with the R^2 = 0.90-0.94. However, significantly better results were obtained with the power function (R^2 =0.96-0.98). This arises from the very fast loading rate used in simulations in comparison to AFM experimental studies where exponential dependency is reached.[243-247]

SMD trajectories provided the initial configurations for US windows which were equally spaced in 0.1 nm windows yielding a total of 21-27 windows of 50 ns each. The free energy of binding (PMF) is the difference between the minima where the small molecule is adsorbed to the protein and the plateau value when interactions are effectively absent (Figure 6.8). The umbrella sampling histograms of each molecule are shown in Figure 6.9 presenting excellent overlap. The free energies of molecular binding to the keratin alpha helical segment obtained via US are listed in Table 6.3 (ΔG_{MD}) and are compared with available experimentally-determined (ITC) results (ΔG_{Exp}) binding other proteins or directly validated (EGCG, ferrous ion). The binding free energies from US exhibit a good agreement with experimental results. The highest magnitude (ΔG=-9.56 kcal mol^{-1}) of binding free energy was for oleic acid, which has the highest *logP* value of 7.64.

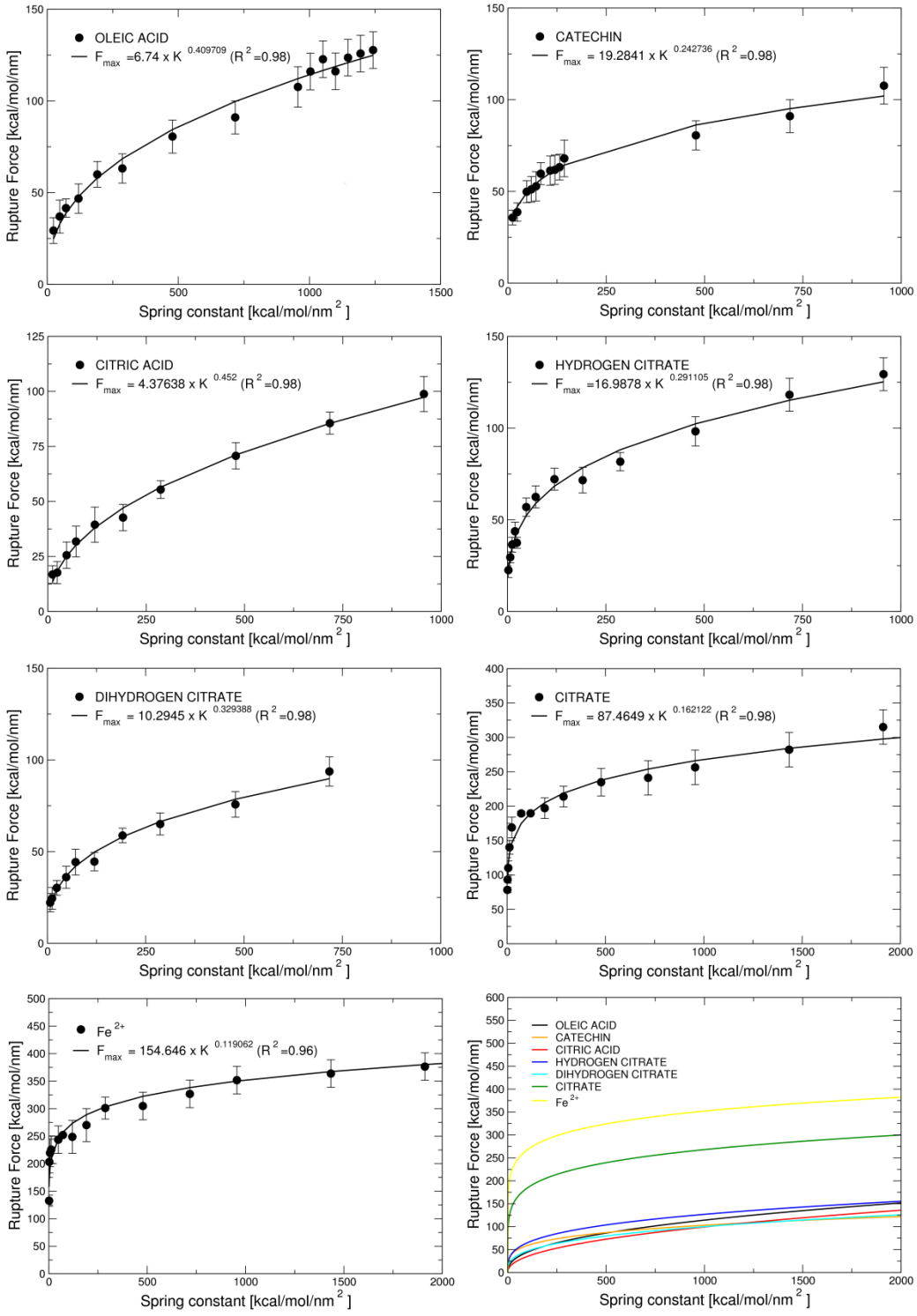


Figure 6.7. The rupture force obtained from steered molecular dynamics (SMD) simulations for 7 molecules (oleic acid, catechin, citric acid, dihydrogen citrate, hydrogen citrate, citrate and ferrous ion) versus the used spring constant. Each point corresponds to the average value of 10 identical pulling simulations. The best fitted curve was obtained in each case with a power function (R^2=0.96-0.98).

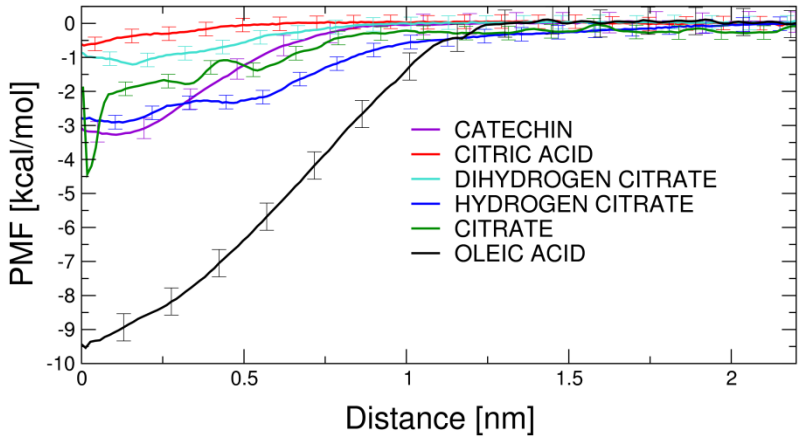


Figure 6.8. Potential of mean force obtained from umbrella sampling method for different molecules binding to keratin helical segment.

Figure 6.10 A shows the binding free energy difference from US versus the *logP* values of uncharged molecules studied herein. The higher the *logP* the greater the magnitude of the binding free energy was observed. It may be noted that an increase in pH resulted in an increase of free energy for citric acid. The increase of the pH above the respective pKa values resulted in a decrease of the overall charge and hence the increase in electrostatic interactions (Figure 6.10 B). The correlation of the buried area for uncharged molecules (EGCG, oleic acid, catechin and citric acid) as a function of the free energy is presented in Figure 6.10 C. It can be observed that the higher the buried area between the alpha helix and ligand the stronger the binding affinity.

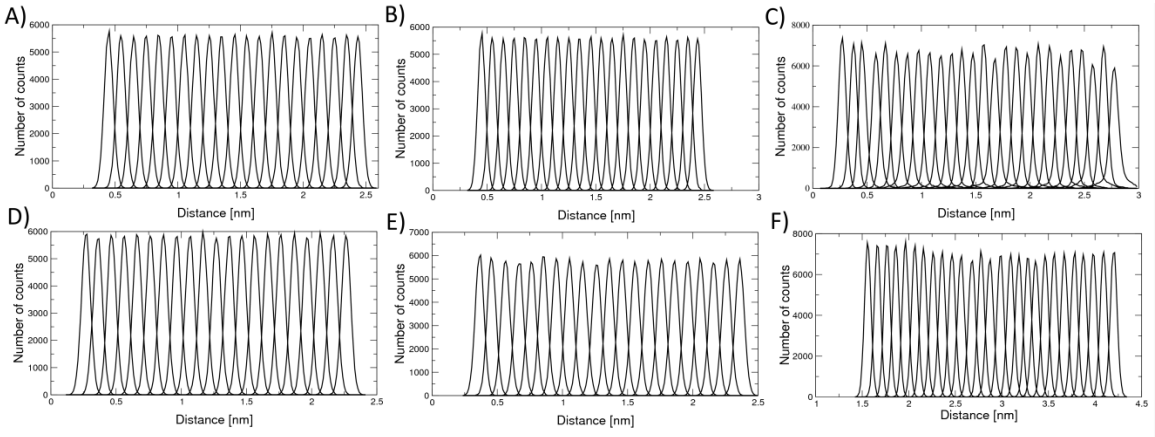


Figure 6.9. Histograms of umbrella sampling windows of: A) Catechin; B) Citric acid; C) Citrate; D) Dihydrogen citrate; E) Hydrogen citrate; F) Oleic Acid.

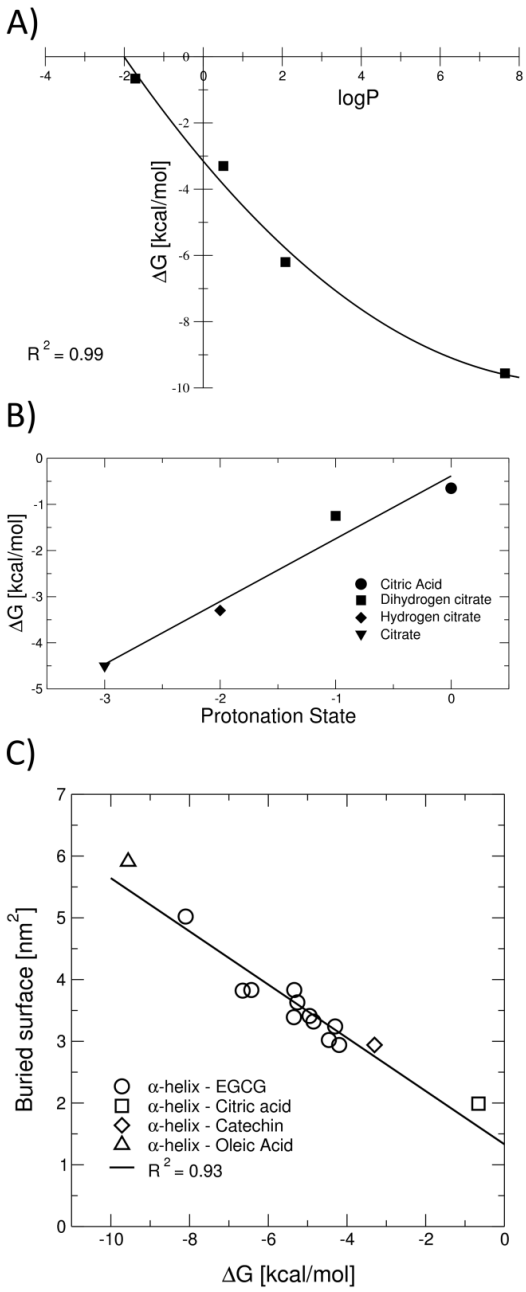


Figure 6.10. A) Binding free energy (ΔG) of small neutral molecules to helical motif as a function of *logP* ; B) Binding free energy of citric acid at different protonation states; C) Protein-ligand buried area (A)as a function of the binding free energy for neutral molecules. The equation is best described by: $A = 1.3321 - 0.431\Delta G$.

The main goal of this research was to assess the validity of the linear dependency from keratin-EGCG system (Figure 4.16): rupture force versus the binding free energy for 7 ligands studied herein. For this purpose, the rupture force for each molecule is extracted as the average value over 10 SMD trajectories from SMD protocol with the spring constant of 120 kcal mol^{-1} nm^{-2} (according to the keratin-EGCG study). All the rupture forces for

each molecule are listed in Table 6.3. In Figure 6.11 the rupture forces versus the binding free energies obtained from US for all molecules is presented. The linear correlation with R^2=0.98 correspond to previously studied EGCG in chapter IV, citric acid, catechin and dihydrogen citrate. It can be observed that molecules which did not follow the correlation (oleic acid, hydrogen citrate, ferrous ion and citrate) possess extreme properties i.e. significantly differ from EGCG properties (*logP* >2.08 or the maximum overall charge exceeds ± 1).

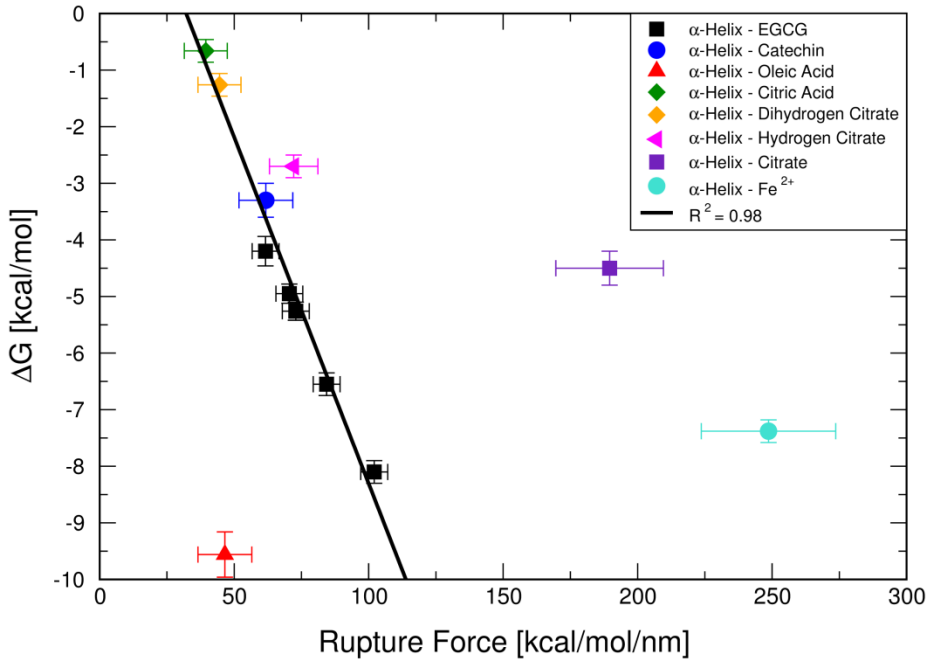


Figure 6.11. Binding free energies (ΔG) obtained from US of small molecules binding the keratin helical motif versus the maximum pulling force averaged over 10 pulling simulations at the constant loading rate (velocity of 10 nm ns^{-1} and a spring constant of 120 kcal mol^{-1} nm^{-2}) according to the keratin-EGCG study.[134] The linear dependency is built based on EGCG, catechin, citric acid and dihydrogen citrate with R^2=0.98 ($\Delta G = -0.1224F_{Max} + 3.9346$). Ferrous ion, citrate, hydrogen citrate and oleic acid did not follow the linear correlation due to their extreme properties.

Table 6.3. Keratin α-helix-ligand complex: Binding free energies from umbrella sampling (ΔG_{MD}); binding free energies validated experimentally (for EGCG and Fe^{2+}) and experimentally obtained with other proteins (ΔG_{Exp}); buried surface area of uncharged molecules; rupture force obtained as the average value over 10 identical pulling simulations at the constant loading rate (velocity of 10 nm ns^{-1} and spring constant of 120 kcal mol^{-1} nm^{-2}) according to the keratin-EGCG study.

Molecule name	ΔG_{MD} [kcal/mol]	ΔG_{Exp} [kcal/mol]	Protein-Ligand buried surface [nm^2]	Rupture Force [kcal/mol/nm^{-1}]
Oleic Acid	-9.56	-9.45[249]/-10.16[250]	5.91	46.54
EGCG	-6.20	-6.37[173]	-	-
Catechin	-3.30	-2.87[252]	2.94	61.78
Citric Acid	-0.66	-	1.99	39.44
Dihydrogen Citrate	-1.26	-	-	44.54
Hydrogen Citrate	-2.70	-	-	72.12
Citrate	-4.50	-5.30[253]	-	189.59
Fe^{2+}	-7.20	-7.38[199]	-	248.67

6.5. Conclusion

Binding affinity of small molecules to macromolecules is one of the most important and thermodynamic property. Recent advances in MD simulations enable predictions of binding free energy with high accuracy. However, methods within MD simulation framework involving explicit solvent are computationally very extensive and demanding. In previous work (chapter IV) on α-helix-EGCG system a linear dependency between the binding free energy (obtained from extensive US) and the rupture force (obtained from short SMD simulations) was presented. Previous work of Mai et al.[136,137] on swine influenza A/H1N1 virus suggested a similar linear correlation between the binding free energy and the maximum pulling force for different ligands. Hence, our keratin-EGCG linear dependency was encouraged to be tested for several additional ligands with significantly varying properties binding the most common alpha helical motif.

Initially, the type of interactions involved in the binding process was assessed. This involved hydrophobic collapse for oleic acid, electrostatic interactions for citrate, hydrophobic interactions and hydrogen bonding for catechin, as well as combination of hydrogen bonds and electrostatics for hydrogen citrate and dihydrogen

citrate. Subsequently, for the most stable molecule in each system the SMD protocol with ranging force constants was applied. It appeared that the power functions of the rupture force versus the applied spring constant described the data well. The binding affinity of each molecule towards keratin helix was then calculated by US and compared with available experimental data. A good agreement was reached. Finally, the validity of the keratin-EGCG linear correlation: rupture force versus the free energy was assessed. It was shown that it is possible to calculate the free energy based on short SMD simulations at constant loading rate for molecules which do not vary significantly in their properties (-1.721< *logP* <2.08 or the maximum overall charge of ±1).

Hence, the protocol for fast free energy predictions using CHARMM22/CMAP force field and TIP3P model for those ligands binding the most common alpha helical motif is provided:

1. Apply SMD protocol of pulling the ligand away from the helix perpendicular to its rod 10 times by its centre of mass with the spring constant of 120 kcal mol^{-1} nm^{-2} and velocity of 10 nm ns^{-1},
2. The arithmetic average (F_{max}) of the maximum pulling forces obtained in step 1 is correlated with the binding free energy: $\Delta G = -0.1224F_{Max} + 3.9346$ where ΔG in units of kcal mol^{-1} and F_{max} in units of kcal mol^{-1} nm^{-1}.

Although the above model is based on low number of molecules and further investigations are necessary in order to build a predictive tool, for the first time in this study molecules with significantly varying properties were taken into account for binding free energy predictions based on couple of SMD simulations. In this chapter it is shown that this approach could be a starting point in the new methodology fast for free energy predictions based on short SMD simulations (hundreds of picoseconds to nanoseconds) instead of time consuming free energy calculations (hundreds of nanosecond to microseconds) and may provide a new pathway in drug design binding affinity establishment. It is shown that SMD is a very predictive tool for free energy calculations.

CHAPTER VII

Coarse-Grained Force Field Development and Application for Skin Keratins

7.1. Introduction

Keratin intermediate filaments (IFs) which are present in the outermost skin layer (SC) within dead, flattened and anucleated cells (corneocytes) are embedded in a "broth" which is a mixture of water, free amino acids, ions as well as other non-ionic compounds. As previously mentioned this mixture composition is called "Natural Moisturizing Factor" (NMF) which plays a significant role in properties of human skin such as: transdermal permeation of small actives, elasticity and the water content.[97,98,100,254] The decrease in the NMF concentration results in the dry and itchy skin which may cause further disorders and diseases.[101] For instance, the applications of water based moisturizers causes the soak of SC layers obtaining illusion of moisturized skin. That causes NMF to be diluted and evaporation of water becomes much higher. Water itself without the presence of NMF has no capacity for providing elastic properties for human skin.[22]

Jokura et al[22] using experimental methods investigated that the absence of NMF in SC causes much stronger association of keratin IFs. Additionally, the mobility of IF was reduced and the increase of the moisture content did not result in better SC elasticity without NMF. The greater the mobility of IFs the better elastic properties of the outermost skin layer reached. Jokura et al. observed that the application of NMF composed of free amino acids resulted in decreased attractive forces between keratin bundles. Different types of amino acids were assessed: neutral to basic (glycine and lysine) provided SC with great elastic properties while acidic residues were shown to have decreased capacity for the SC elasticity restoration. Those results were summarised with the statement that "the role of NMF is to prevent the attractive forces between keratin IFs" providing the SC layer with elastic properties as well moisturized healthy skin.

Akinshina et al.[102] developed a model based on the self consistent field theory (SCF) for the interactions between keratin IFs in the SC (keratin pair: keratin 1 and 10). By representing IFs as charged surfaces with grafted unstructured keratin heads and tails it was claimed that terminal chains have capacity of mediating weak attraction between IFs by electrostatics and bridging. The addition of NMF (free amino acids) caused the lower attraction between keratin bundles. Results validated experimental studies by Jokura et al. and were represented as volume fraction of given types of amino acids (polar, nonpolar, acidic, basic and glycine) away from the keratin IF surface.

The IF diameter of 7.8 nm as well as the distance between keratin IFs of 8.2 nm in a healthy skin was provided by Norlen et al. [110] in the cryo-electron microscopy study. It is also known that the length of the IF corresponds to 50 nm.[88] The elastic properties of the outermost skin layer and thus interactions between keratin IFs within MD have never been studied before. That arises from the large size of this system which all atom force fields have no capacity in reaching reasonable time scales.

7.2. Objectives

Due to the above limits of all atom force fields, the aim of this chapter was to develop and apply the coarse-grained (CG) force field which will enable studying interactions between keratin IFs and the effect of NMF. For this purpose, the following objectives were defined:

- Run all atom reference simulation in explicit water,
- Generate a water free CG force field which will reproduce full atomistic representation with 5 types of beads: polar, nonpolar, acidic, basic, glycine - each representing one amino acid,
- Build a representation of keratin IFs,
- Apply the force field for interactions between keratin IFs and assess the influence of NMF,
- Validate the secondary structure of the surrogate within circular dichroism (CD).

For this purpose the 60 residues surrogate representing non-helical keratin terminal was chosen. Temperature annealing was employed for the all atom simulation in explicit water in order to explore all possible conformations (energy states) of the surrogate. Subsequently, by representing each amino acid as one bead with its mass corresponding to its central carbon atom in the coarse grain representation, bonded and non-bonded distributions were extracted and reproduced within Boltzmann inversion and iterative Boltzmann inversion method respectively. A great agreement was reached between fully atomistic and coarse grain model in terms of bonded and non-bonded distributions as well as the compactness of the polypeptide (radius of gyration). The coarse-grain force field was applied with IFs represented by cylindrically shaped conformations. The influence of NMF was assessed confirming Jokura et al. experimental studies. Finally, circular dichroism (CD) experiment was performed by Dr Alfonso De Simone (Imperial College London) on the synthesised surrogate polypeptide. The obtained experimental result provided the typical spectra of a random coil confirming the computational findings. The experimental part is presented in Appendix III. In this work, for the first time a coarse-grain force field including water within tabulated potentials was built and applied for the interactions between intermediate filaments.

7.3. Methodology

7.3.1. The choice of the sequence

As previously mentioned keratin IFs are formed by pairing acidic (type I) and basic (type II) keratin. The most abundant keratin pair in the SC is keratin 1 and keratin 10 (KRT1/KRT10).[89,90] The sequences of KRT1/KRT10 were provided by keratin IF database (www.interfil.org).[183] It can be noted that two “heads” and two “tails” (non-helical terminals) of the most common keratin pair in SC possess glycine rich motifs. In Table 7.1 the amino acid composition of SC keratin pair in terms of charged residues and glycine content is presented. The analysis is based on the acidic pH of the skin where histidine residues are in their protonated state.

Table 7.1. The amino acid composition of the most abundant stratum corneum keratin pair: KRT1/KRT10 in terms of charged amino acids and the content of glycine.

	Overall charge	Glycine content [%]	Number of amino acids
KRT1 – N terminal	+10	40	180
KRT1 – C terminal	+9	46	151
KRT10 – N terminal	+6	48	146
KRT10 – C terminal	+7	56	124
KRT1/KRT10 – helical rod	-13/-21	1.6/3.2	313/314

It can be noted that the overall charge of helical part of the coiled-coil domain composed of KRT1/KRT10 corresponds to -34 while all sticking out non-helical terminals ("heads" and "tails") possess the charge of +32. Hence, the crucial role is played by electrostatic interactions between keratin IFs and "heads" and "tails". The glycine content of all terminals in SC keratin pair is above 40% while in the helical part below 5%. It can be noted that acidic residues in all terminals are mainly situated close to the keratin IF surface while basic amino acids near the terminal endings. All SC keratin terminals have amino acid sequence longer than 100 residues which is not reachable by experimental synthesis. Due to this fact and the computational resources needed for long polypeptides with explicit water, the sequence of the terminal was approximated to 60 residues surrogate based on the KRT10 - terminal N. The terminal endings (basic and acidic of 16 and 10 residues respectively) were conserved as well as 34 residues of glycine rich part neighbouring with the basic ending (Figure 7.1).

Keratin 10: N terminal - 146 residues:

MSVRYSSSKHYSSSRS
GGGGGGGGCGGGGGVSSLRISSSKGSLGGGFSSGGFSGGSFSRGSSGGGCFGGSSGGYGGLGGF
GGGSFRGSYGSSSFGGSYGGSF
GGGSFGGGSFGGGSFGGGGFGGGGFGGGFGGGFGGDGGLLSGNE

Surrogate terminal - 60 residues:

MSVRYSSSKHYSSSRSGGGSFGGGSFGGGSFGGGGFGGGGFGGGFGGGFGGDGGLLSGNE

Figure 7.1. The choice of the terminal surrogate: full sequence of keratin 10: N terminal (146 residues) and approximated surrogate of 60 residues with conserved basic and acidic ending as well as glycine rich motif (residues marked in: blue – basic, red – acidic, green – glycine, yellow – remaining part of surrogate, black – those omitted in the surrogate sequence).

7.3.2. System setup – All Atom Simulation

The reference full atomistic simulation was run using ACEMD[255,256] software under graphic process unit (GPU) with AMBER99SB*-ILDN-Q[257] force field. The surrogate sequence has been taken as an input to the initial structure for the full atomistic reference simulation. Alpha helical segment consisting of the surrogate sequence has been built using PyMol software and was placed in the cubic unit cell of 857.375 nm^3. Due to the acidic skin pH, histidine was treated in its protonated state. The box was filled with the 27000 TIP3P[184] water molecules and two CL- ions were added to reach the neutral charge of the system. After the energy minimization using conjugate-gradient[258] algorithm with 1fs time step, the NVT of 1ns and NPT of 1 ns with position restraints dynamics of backbone atoms were subsequently performed using 2 fs time step for the system equilibration. The langevin[259] thermostatic control and Berendsen barostat[260] with the 1 atm pressure were applied using the velocity-Verlet integration. The PME[186-187, 255] was used for long-range electrostatics with the cut-off distance of 0.9 nm for the direct-space interactions. In order to obtain a random coil, the production run of 80 ns simulation in NPT ensemble was performed at temperature of 800K. Obtained random coil was an initial input for the temperature annealing simulation. By driving the system from 300K to 500K in the NPT ensemble, different energy states can be explored. This technique provides the possibility of avoiding the energy minima trap. Contrary, using a constant temperature the final structure may be dependent upon the initial structure due to the local energy minima. Replica exchange molecular dynamics (REMD)[261] is another method to enhance sampling by exploring different energy states with replicas simulated at different

temperatures. REMD is computationally very expensive and not applicable under the graphic process unit (GPU) within ACEMD software. However, using temperature annealing different conformations can be visited and the system has capability to explore the whole phase space in terms of the momenta and positions. For this purpose a series of cycles (each of 10 ns) was applied (Figure 7.2):

1. Heating up the system from 300K to 500K of 2 ns,
2. Simulation of 500K for 3ns,
3. Cooling down the system from 500K to 300K for 3 ns,
4. Equilibrium simulation at 300K for 2ns (the collection of data).

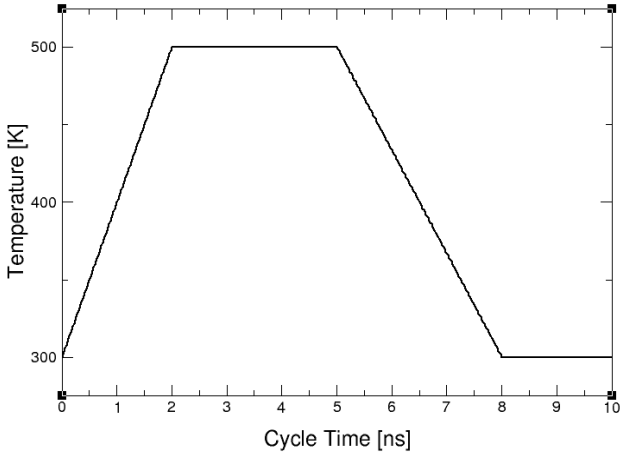


Figure 7.2. The temperature annealing cycle of 10 ns – driving the system from 300 to 500K sufficient number of times enables to explore different conformations and hence the whole phase space in terms of positions and momenta.

The temperature annealing was run until the trajectory at 300 K (last 2 ns of each cycle) divided into two parts represented similar structural properties. That ensured that all energy states i.e. conformations were explored at this temperature. Subsequently, this trajectory (at 300 K) was used for calculating distributions for bonded and non-bonded interactions. All distributions were calculated between central carbon atoms of each residue. In order to reproduce bonded parameters, distributions of angles (3 consecutive central carbon atoms) as well as dihedral angles (4 consecutive central carbon atoms) over the trajectory were calculated. The distributions of bonds were omitted due to the simplification assumed in the coarse grain representation by the use of distance constraints (LINCS)[185] between Cα atoms corresponding to 0.38 nm.

Non-bonded interactions were extracted as radial distribution functions (RDFs) between different types of beads. RDF is a probability of finding of the particle density (type B) at a distance r around particles of type A. RDFs calculations were performed with excluding 3 possible neighbours. The RDFs were calculated between 5 types of beads: polar (serine, tyrosine, and asparagine), nonpolar (valine, phenylalanine, and leucine), basic (methionine with NH3+ group, arginine, lysine and protonated at skin pH histidine), acidic (aspartic acid, glutamic acid) and glycine. The choice of the type beads corresponds to the intramolecular electrostatic interactions (acidic, basic) within keratin, hydrophobic interactions (polar, nonpolar) and the crucial glycine rich repeating motifs.

7.3.3. System Setup – Coarse Grained Approach

As previously mentioned the aim of this work was to build the coarse-grained force field with 5 types of beads. Each of 60 beads represented one amino acid of given type with its mass. The distances between "bonded" coarse-grain beads were constraint to 0.38 nm (using LINCS algorithm) according to full atomistic simulation distances between central carbon atoms. In order to reproduce bonded (angles and dihedrals) and non-bonded (RDFs) distributions obtained from all atom simulation, Boltzmann inversion method was used for this purpose using VOTCA software.[170] For bonded parameters the calculated histograms of angles and dihedral angles were normalized according to the equation 3.33 and inverted (at T=300K) into tabulated potentials (equation 3.34). For non-bonded parameters RDFs (P_{ref}) were similarly inverted according to the equation 3.35. The post-processing methodology of the obtained potentials i.e. inverted bonded and non-bonded distributions was as following:

- Clipping poorly sampled regions – deleting values with irregular shape,
- Resampling (interpolating) the potential – changing the grid,
- Extrapolation at the borders,
- Exporting the potential to the table supported by GROMACS software.

The exported tables are the input for the coarse-grain force field and simulation can be performed. In total for 60 residues polypeptide distributions of 58 angles, 57 dihedral angles as well as 15 RDFs (5 types of beads) were aimed to be reproduced.

The coarse-graining methodology was as following:

1. Reproduce within CG representation all atom bonded distributions (angles, dihedrals). For this purpose a simulation of 10 μs with the post-processed inverted bonded distributions (potentials) using non-bonded Weeks Chandler Andersen (WCA)[262] potential for all beads. WCA is a fully repulsive potential presented in Figure 7.3 which enables exploring all possible angles and dihedral angles thus used in this chapter.

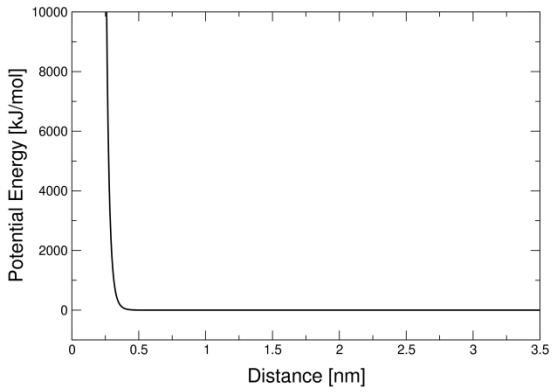


Figure 7.3. Weeks Chandler Andersen (WCA) potential used for all beads in the coarse-grain model to reproduce bonded parameters within the simulation of 10 μs.

2. Reproduce all-atom 15 RDFs by CG model with bonded potentials included from the step 1. The input for the iterative Boltzmann inversion method are reference RDFs from full atomistic simulation as well as initial guess of tabulated potentials (inverted RDFs).

3. Assess the influence of non-bonded tabulated potentials from point 2 on the distributions of bonded potentials obtained in step 1 within 10 μs simulation.

All coarse-grain simulations were run using NVT ensemble. The cut off distance for non-bonded potentials were set to 2 nm. The equations of motion were integrated using Verlet leap-frog algorithm. The velocity rescale thermostat with additional stochastic term was used to maintain the temperature at 300 K. The initial velocities were set according to the Maxwell distribution. Periodic boundary conditions were used in all directions.

7.4. Results

7.4.1. All Atom Simulation

As previously mentioned the first full atomistic NPT simulation involved 60 residues α-helix at 800 K in order to obtain a random coil. The obtained unstructured surrogate after 80 ns simulation at 800 K is presented in Figure 7.4.

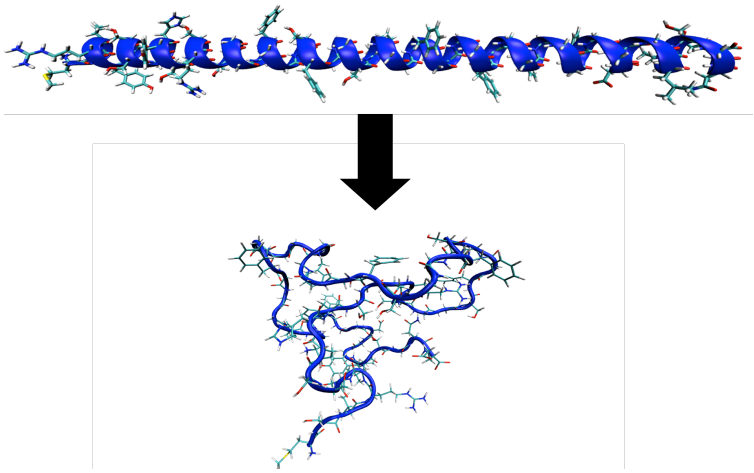


Figure 7.4. Initial helical structure of the 60 residues surrogate together with the obtained random coil after simulation of 80 ns in the NPT ensemble at the temperature of 800K.

Subsequently, 210 cycles of 10 ns (2.1 μs in total) have been conducted using NPT ensemble. This simulation time was sufficient to explore the whole phase space which has been assessed by dividing the whole trajectory at 300 K (combined last 2ns of each cycle into one trajectory) into: part I) 300-1200 ns and part II) 1200-2100 ns) which represented similar structural properties. That ensured that all energy states were visited. In the Figure 7.5 the mean distances maps of the whole simulation and both part I and II between given residue index are presented. The mean distance for each residues pair was calculated as the distance between their centres

of masses and averaged over the whole simulation time. First 300 ns (30 cycles) for the system equilibration were omitted. It is observed that nearly identical mean distances maps were obtained in both parts of the simulation suggesting that all conformations were explored at 300K.

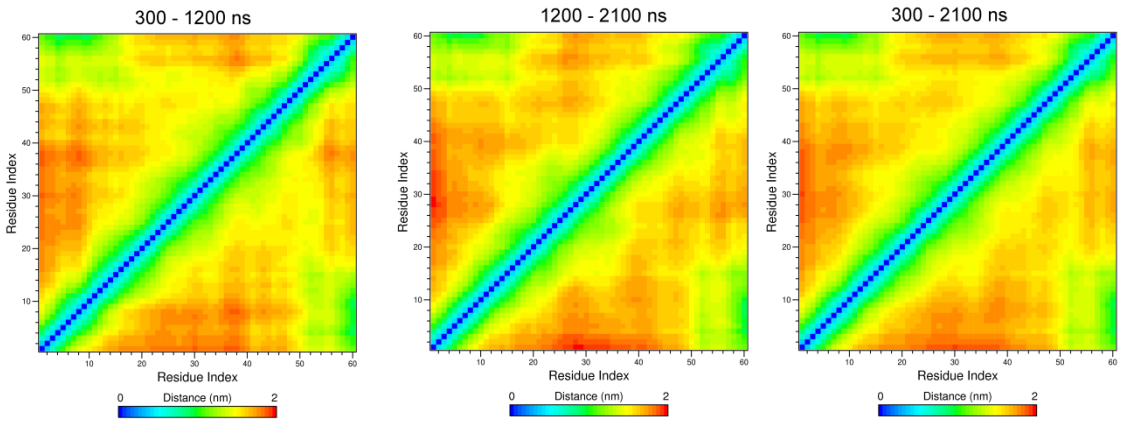


Figure 7.5. Mean distances map of the surrogate terminal at 300K. The trajectory was divided into two parts: I – 300-1200 ns and II – 1200-2100 ns. Both parts represent similar maps of mean distances as well as the whole trajectory of 300-2100 ns. First 300 ns were omitted for the system equilibration.

Root mean square fluctuation (RMSF i.e. standard deviations averaged over the simulation time) of central carbon atoms with respect to the initial random coil was calculated for both parts of the full atomistic simulation. In Figure 7.6 RMSF is shown for both part I and II of the simulation providing similar deviations of each central carbon atom. The end to end distance between central carbon atoms of terminal residues (1st and 60th) was also calculated for both parts of the simulation. In Figure 7.7 nearly identical distributions are shown. Similarly the distribution of the hydrophobic solvent accessible surface area (SASA)[190] shows similar profiles (Figure 7.8) together with the radius of gyration (Figure 7.9) calculated for the backbone atoms.

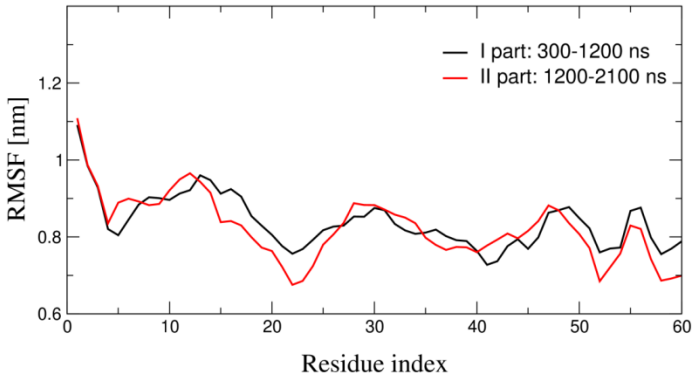


Figure 7.6. Root mean square fluctuations (RMSF) with respect to the initial random coil structure of given residue index (central carbon atom) averaged over the simulation time of part I and II at 300K.

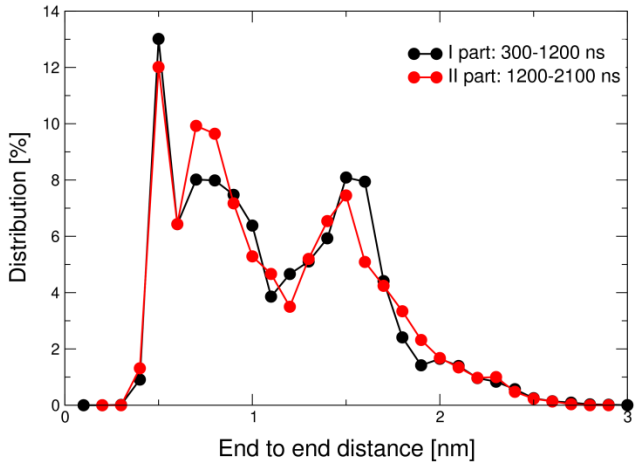


Figure 7.7. Central carbon atoms end to end distance (first and last residues) distributions of part I and II full atomistic simulations at 300K.

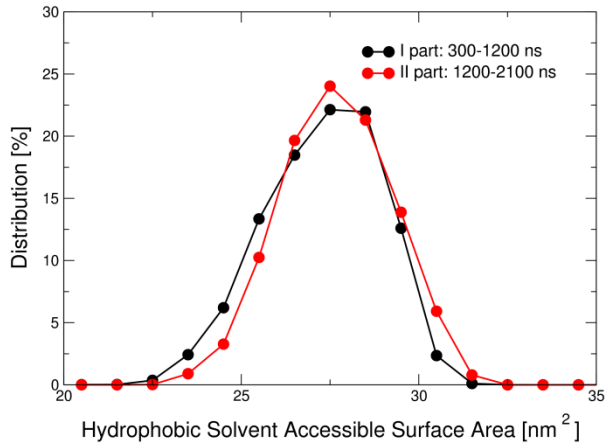


Figure 7.8. Distributions of the polypeptide hydrophobic solvent accessible surface area of the part I and II at 300 K.

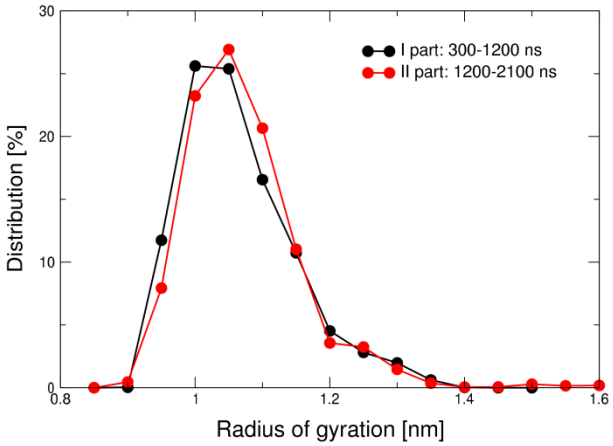


Figure 7.9. Distributions of the polypeptide gyration radius of part I and II of the all atom simulation at 300 K.

Once all conformations (energy states) were explored within the full atomistic simulation at 300 K, distributions of angles and dihedral angles together with 15 RDFs between different types of beads can be extracted. In Figure 7.10 RDFs are presented for given pair of beads. RDFs were normalized so that area below the curve is equal to 1 and then cut at the 2 nm distance where negligible interactions are present. Similarly the distributions of angles and dihedrals were obtained. In Figure 7.11 distributions of 58 angles and 57 dihedrals at 300 K from fully atomistic simulation are shown.

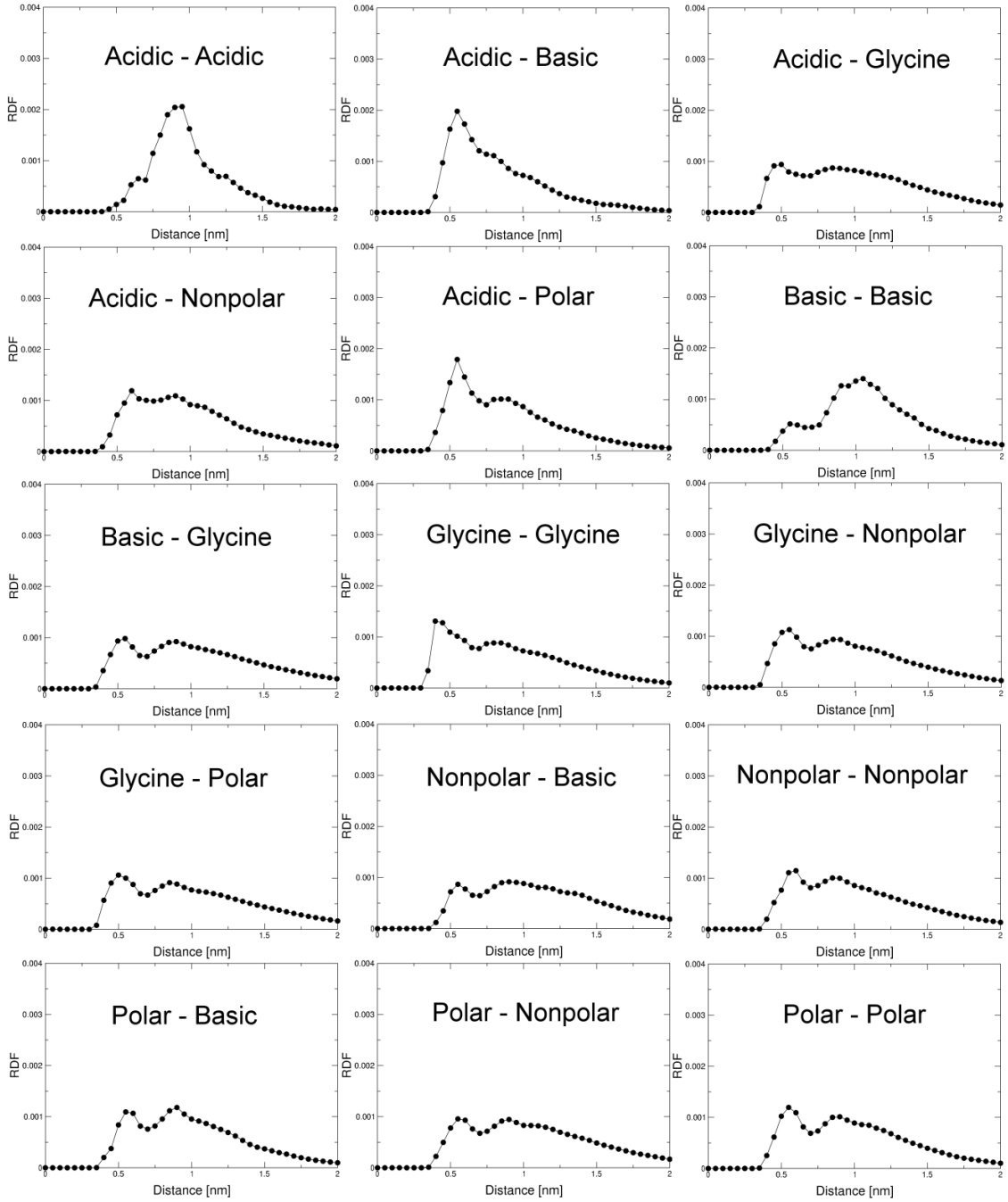


Figure 7.10. Radial distribution functions (RDFs) between different types of beads obtained from fully atomistic simulation at 300 K.

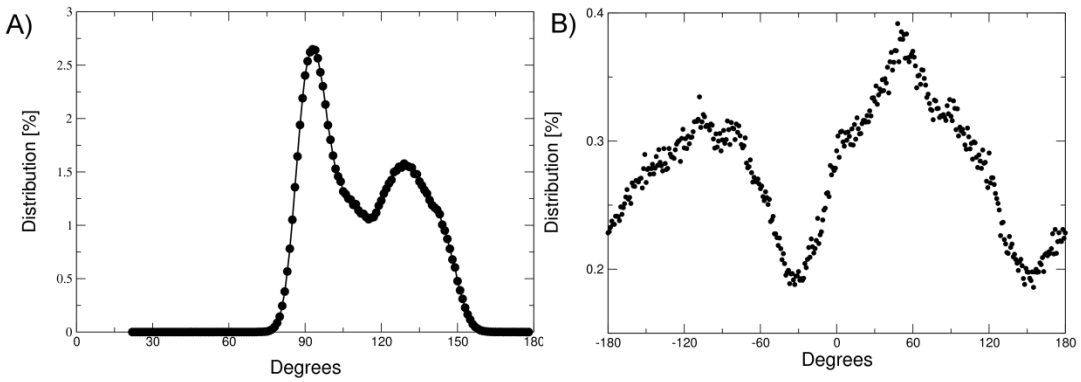


Figure 7.11. Bonded distributions extracted from fully atomistic simulation at 300 K: A) 58 angles between 3 consecutive central carbon atoms; B) 57 dihedral angles between 4 consecutive central carbon atoms.

7.4.2. Coarse-Grained Force Field

The aim of the coarse-grained model is to reproduce fully atomistic simulation. As mentioned before the electrostatics and water is assumed to be included in the coarse grain simulation. In Figure 7.12 fully atomistic simulation box with explicit solvent model is shown as well as the aimed coarse grained representation with 60 beads, each representing one amino acid.

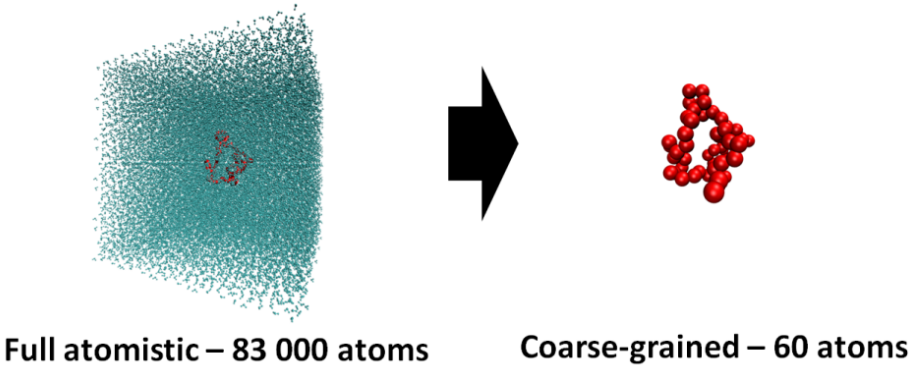


Figure 7.12. The concept of coarse graining from fully atomistic simulation (app. 83 000 atoms) to coarse-grained representation (60 beads). Each amino acid in the coarse-grained model is represents by one bead with the mass of the residue. Electrostatics and explicit water model from fully atomistic simulation were aimed to be included within tabulated non-bonded potentials.

According to the methodology in section 7.3.3, initially distributions of 58 angles and 57 dihedral angles (Figure 7.11) were aimed to be reproduced using BI method within a simulation of applied WCA potential for all beads. The post-processing methodology of each angle and dihedral angle was employed and the example of one of the angles is shown in Figure 7.13.

First, poorly sampled regions of the potential are removed. Secondly the data is resampled (interpolated) within cubic function using minimum and maximum values of the area as well as the grid spacing (Figure 7.13 A). Finally, the data is extrapolated to the borders using the cubic function and translated into the GROMACS software as tabulated potential - the input for the coarse-grain simulation (Figure 7.13 B).

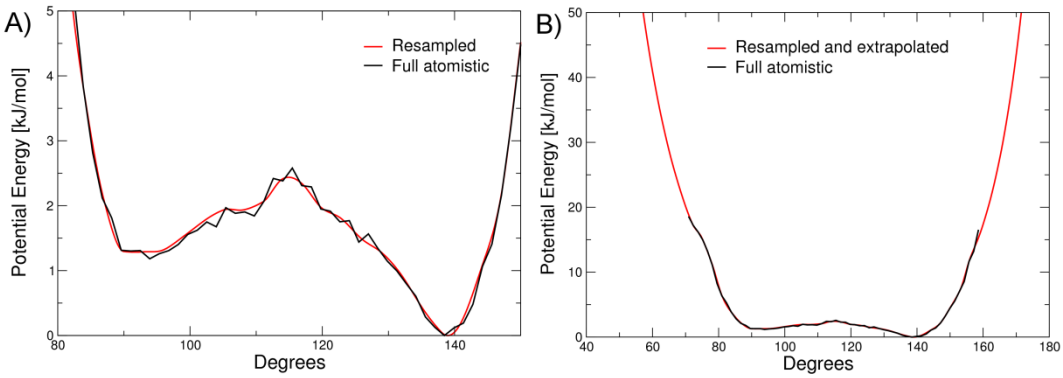


Figure 7.13. The example of refining tabulated potential based on the one of the angles between 3 consecutive central carbon atoms. The raw data obtained by inverting the distribution from fully atomistic simulation is refined through the: A) interpolation (resampling) and B) extrapolation.

The post-processed tabulated bonded potentials (58 angles and 57 dihedral angles) were used as an input for the simulation using Weeks Chandler Anderson (WCA) non-bonded potential for all beads. The production CG run was set to 10 μs in the NVT ensemble. In the Figure 7.14 the results of the coarse-grain distributions of all angles and dihedral angles in comparison to the fully atomistic simulation are presented. Excellent agreement was reached between fully atomistic simulation and the coarse-grain model.

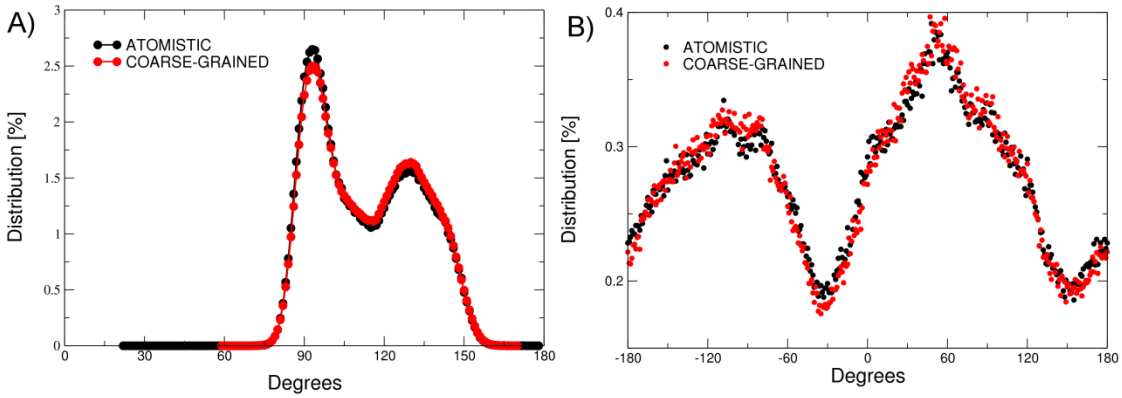


Figure 7.14. Coarse-grained distributions after 10 μs of NVT simulation at 300K using Weeks Chandler Andersen potential versus distributions obtained from fully atomistic simulation of: A) 58 angles and B) 57 dihedral angles.

The reproduced bonded parameters were used as the input for the IBI to iteratively obtain the reference RDFs within CG representation. The input for the IBI method in VOTCA software is: fully atomistic reference RDFs together with the tabulated potentials (inverted RDFs) for the first production run. The IBI method employed herein is described in the methodology section (equations 3.35 – 3.37). Initially, the first simulation with the mentioned input is run. Secondly, the obtained coarse-grain RDFs are compared with the fully atomistic RDFs (P_{ref}). The potential for the next step ($n+1$) is equal to the potential for the n step plus the scaled by κ the difference between step n and the reference atomistic simulation (equation 3.36 - 3.37). Initially κ was set to one and each simulation (iteration) time in the NVT ensemble was set to 1 ns. At each step one of 15 potentials was refined. This comes from the fact that the change applied to the one potential affect all the remaining. RDFs of coarse-grain model versus the RDFs from full atomistic simulation were assessed every 45 steps so that each potential was refined 3 times. After 8000 iterations of 1 ns where coarse-grain RDFs were close enough to the atomistic RDFs the iteration step was increased to 5 ns to obtain longer period of sampling and κ set to 0.5 due to the under- or overestimation of neighbouring steps. Subsequently, 1000 iterations of 5 ns each were run with the κ set to 0.1. The next 500 iterations were run with κ equal to 0.1 with the each run of 100 ns. The final step involved 45 iterations of 500 ns each with κ of 0.1. In total, approximately 10 000 iterations were run.

The tabulated non-bonded potentials from the last iteration step were the input for the production run together with previously extracted bonded tabulated potentials to assess all non-bonded distributions (RDFs) in comparison to the fully atomistic simulations. In the Figure 7.15 all coarse-grain RDFs obtained from 10 μs simulation with tabulated bonded and non-bonded potentials after approximately 10000 iterations using IBI method versus the RDFs from fully atomistic simulation are shown. A great convergence was reached between RDFs obtained from fully atomistic simulation and the coarse-grained model using tabulated potentials. Due to the electrostatics included within non-bonded tabulated potentials, the distributions with basic or acidic beads represented the lowest convergence. All potentials which provided distributions shown in Figure 7.15 are presented in Figure 7.16. It can be observed that only acidic-nonpolar, acidic-basic, glycine-nonpolar and basic-polar interactions show attractive interactions. All remaining provided with the repulsive potentials.

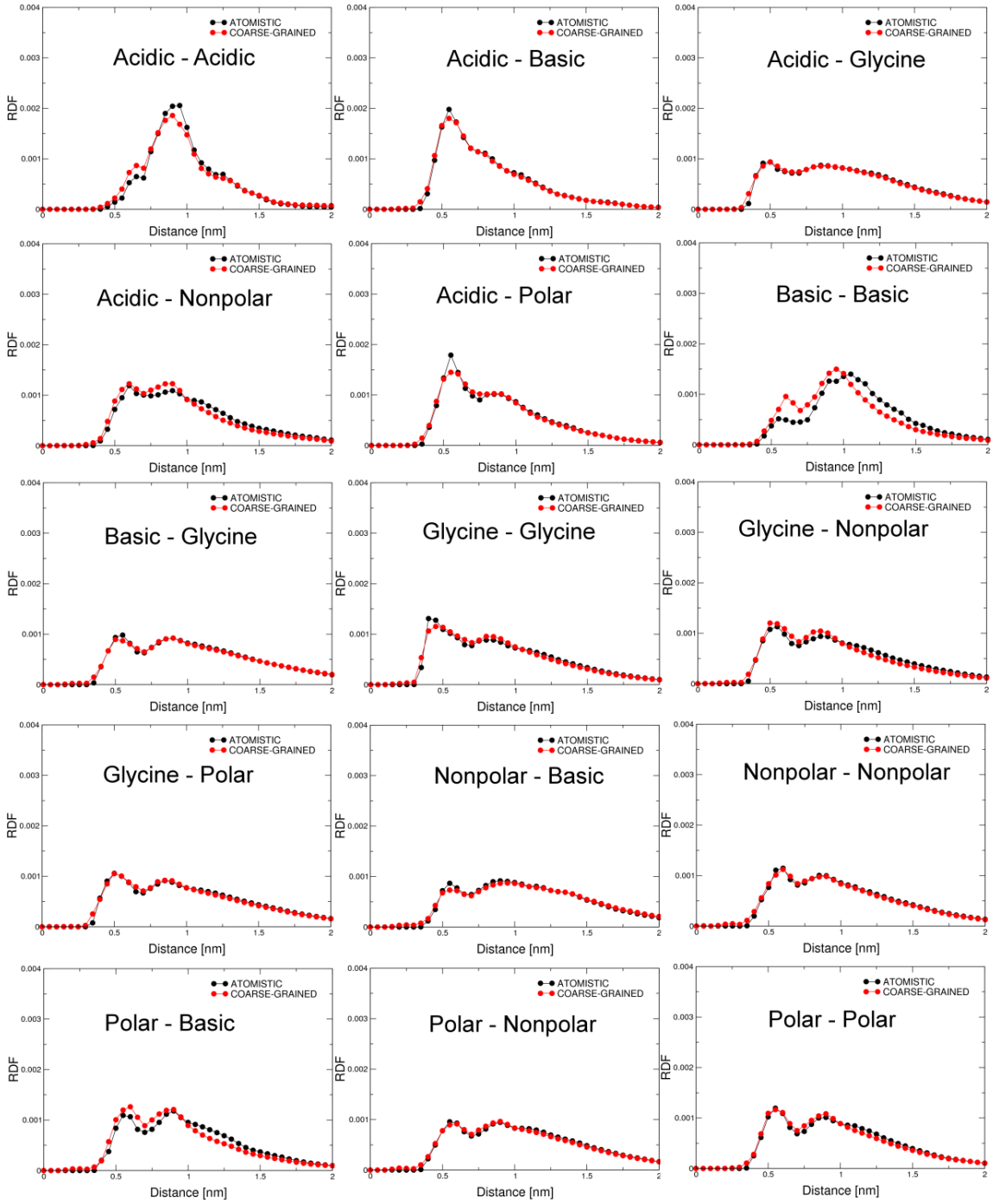


Figure 7.15. Radial distribution functions (RDFs) between 5 types of beads (polar, nonpolar, basic, acidic, glycine) obtained from: fully atomistic simulation (black) at 300 K and the coarse-grained simulation of 10 μs after 10 000 of iterations (red) using iterative Boltzmann inversion method.

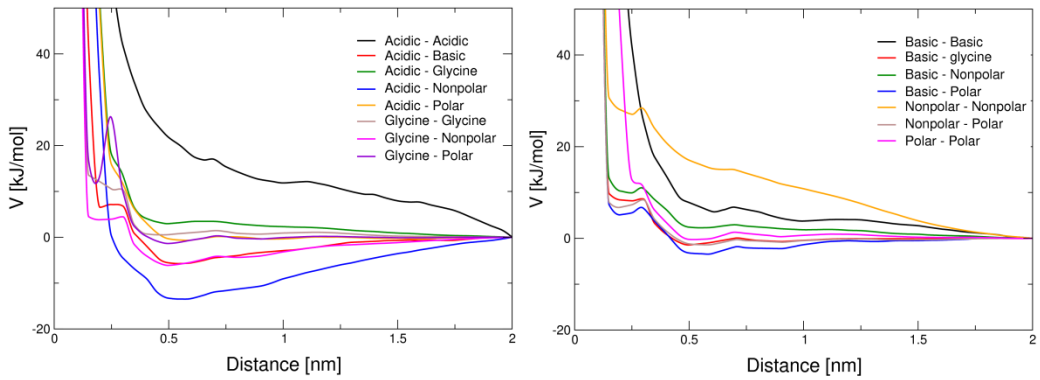


Figure 7.16. Final tabulated potentials with the 2 nm cut-off between 5 types of beads. These potentials provided well converged radial distribution functions shown in Figure 7.15.

Due to the fact that angles and dihedral distributions for the coarse-grain model were calculated using WCA potential for non-bonded interactions the bonded distributions were also calculated in order to ensure about their validity with new non-bonded interaction potentials. In Figure 7.17 distributions of all angles and dihedrals are shown together with distributions from fully atomistic simulation. It can be observed that bonded distributions with new non-bonded potentials show similar agreement with those used with WCA potential in the Figure 7.14.

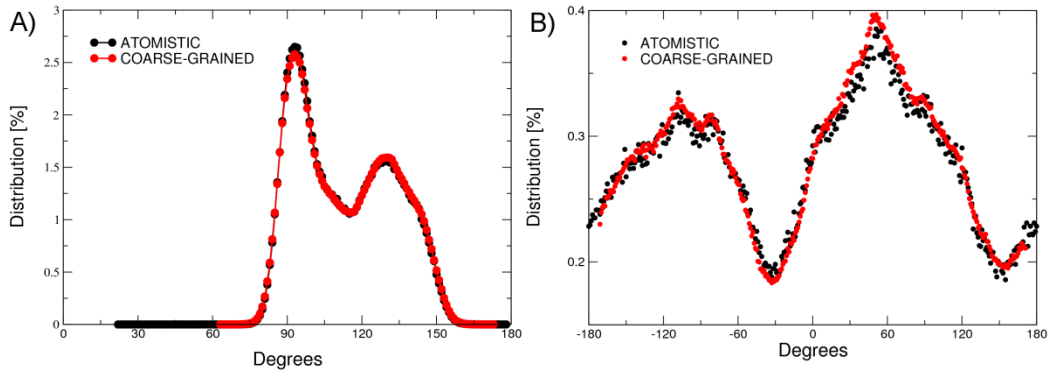


Figure 7.17. Distributions at 300K of A) 58 angles and B) 57 dihedral angles of the coarse-grained model with the 15 potentials (5 types of beads: polar, nonpolar, basic, acidic, glycine) after 10 μs of the simulation versus distributions from fully atomistic simulation.

Although there is a great agreement between the coarse-grain force field and fully atomistic simulation for the both bonded and non-bonded distributions the main challenge is reproduction of the structural properties of the whole polypeptide. The radius of gyration (compactness of the system – mean distances of central carbon atoms of all residues with respect to the centre of mass of the whole polypeptide averaged over the

simulation time) was also extracted from the coarse-grain model and compare to the fully atomistic simulation (Figure 7.18). Very good agreement was gained between coarse grain model and atomistic simulation. The developed force field will be used for studying the interactions between keratin IF.

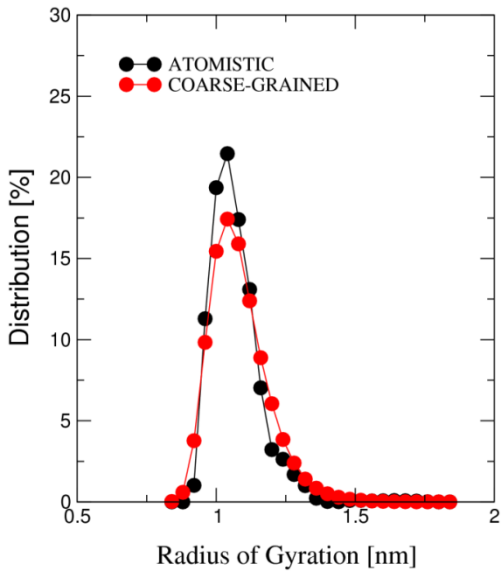


Figure 7.18. Distributions of the surrogate polypeptide gyration radius: coarse-grain model (red) versus fully atomistic simulation (black).

7.4.2.1. Representation of Intermediate Filaments

The goal of this work is to apply the developed force field with 5 types of beads: polar, nonpolar, basic, acidic and glycine together with the representation of intermediate filaments (IFs). The IF diameter, length and the distance between IFs according to the cryo-electron microscopy by Norlen et al.[110] is 7.8nm, 50 nm and 8.2 nm respectively. In order to build a realistic representation of the IF, the full sequence analysis of the coiled-coil part of keratin pair: KRT1 and KRT10 were conducted. All helix forming amino acids were divided into the CG types of beads: polar – serine, threonine, tyrosine, asparginine, glutamine; nonpolar – alanine, valine, isoleucine, leucine, methionine, phenylalanine, proline, cysteine; basic – histidine, arginine, lysine; acidic – aspartic acid, glutamic; and glycine. The IF consists of 16 coiled coils (32 chains in total) represented by the KRT1 (313 residues) and KRT10 (314 residues) pairs. The sequence analysis of the IF helixes made of 16 helical rods of KRT1 and 16 KRT10 is presented in the Table 7.2 together with the approximated model which represents ¼ in length (12.5 nm) of the real (50 nm) IF. For the productions runs to save the computational

time the ¼ IF representation will be copied across periodic boundary conditions to mimic infinite in length IF. The diameter of the IF model was conserved according to experimental data being set to 7.8 nm. With known dimensions of the IF and its surface area as well as its overall charge of -544 the surface charge density is equal to: -0.443e/nm^2. In the IF ¼ model representation conserving similar contribution of different types of beads the charge density is nearly identical and equal to -0.444e/nm^2.

Table 7.2. Amino acid composition of the real intermediate filament (50 nm in length) made of KRT1/KRT10 pair and its approximated model (1/4 in length) which will be used in simulations.

Amino acid type	Intermediate Filament (50 nm in length) – 10 032 residues	Intermediate Filament ¼ model (12.5 nm in length) – 2516 residues
Polar	1344	338
Nonpolar	1888	471
Basic	3184	799
Acidic	3376	848
Glycine	240	60

In order to obtain the closest number of atoms in the ¼ IF representation (2508 = 0.25 of the whole IF of 10 032) of the real IF, it is represented by the tube of 68 line chains with 37 residues. That provides 2516 residues in total. The coiled-coil sequence is known to possess heptad repeats “abcdefg”. Hence, conserving the heptad repeats in each of 68 strands of 37 residues and placing uniformly remaining residues the approximation of the IF is built. In Figure 7.19 the ¼ of the IF model is shown. The cuff off distance (2 nm) in the non-bonded tabulated potentials enabled leaving the inside of the IF model empty.

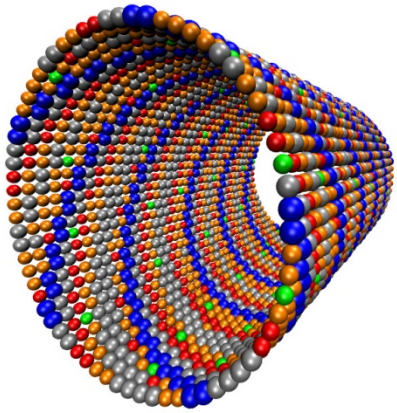


Figure 7.19. The representation of the intermediate filament (12.5 nm in length) with 5 types of beads: acidic – red, basic – blue, glycine – green, polar – grey, nonpolar – orange.

7.4.2.2. Grafting Terminals on the Intermediate Filament Surface

With known potentials for angles (3 consecutive beads), dihedrals (4 consecutive beads) in 60 residues polypeptide the real sequence of keratin "tails" and "heads" which will be grafted onto the keratin IF surface need to approximated. Below the approximation of all terminals is presented together with the real sequence of each of terminals with 5 types of beads: A-Acidic, B-Basic, G-glycine, P-Polar, N-Nonpolar. The main challenge was to conserve the total number amino acids the number of given amino acid type, glycine repeats in the middle of each chain as well as basic and acidic residues on terminal endings.

1. **K10 – N terminal (146 residues):**

- **Real sequence:**

BPNBPPPPBBPPPPBPGGGGGGGGNGGGGGNPPNBNPPPBGPNGGGNPPGGNPGGPNPBGPPGGGNNGGPPGGPGGN
GGNGGGPNBGPPGPPPNGGPPGGPNGGGPNGGGPNGGGPNGGGGNGGGGNGGGNGGGNGGAGGNNPGPA

Number of amino acids: Acidic: 2, Basic: 9, Polar: 40, Glycine: 70, Nonpolar: 25

- **Approximated sequence with known angles/dihedral angles:**

BPNBPPPPBBPPPPBPGGGPNGGGPNGGGGGGPBPPPPPPPBPGGGPNNGGGPNGGGGPBPGGGPNGGGPNNGGGPN
NGGGPBPGGGPNGGGPNNGGGGPNNGGGPNGGGPNGGGPNGGGGNGGGGNGGGNGGGNGGAGGNNPGPA

Number of amino acids: Acidic: 2, Basic: 9, Polar: 38, Glycine: 72, Nonpolar: 25

2. **K1 – N terminal (180 residues):**

- **Real sequence:**

BPBPNPPBPGPBPGGGNPPGPNGNNPPPBBPPPPPPBBPGGGGGBNPPNGGGGGPNGNGGGNGPBPNNPNGGP.BPNP
NPNNBG.GGBGPGNGGGPGGGGNGGGGNGGGGNGGGGNGGGGNGGGNGPGGGGNGGGGNGGGGPGGGPGNNNNN
GGNPANPNPPPNNPN.NPNANANANPBNBPBA

Number of amino acids: Acidic: 5, Basic: 16, Polar: 44, Glycine: 72, Nonpolar: 43

- **Approximated sequence with known angles/dihedral angles:**

BPNBPPPBPPPBPGGGPNNGGGPBPPPBBPPPBPGGGPBPGGGGNNPGPNNGGGGPBPPPBPGGGPNNGGGPNNGGG
GPBPPP)BPGGGNNPGPNNGGGNNPGPNNGGGNNPGPNNGGGNNPGPNNGGGNNPGPNNGGGPNNGGAGGNGGA
GGNGGAGGNGGGPBPGGGPBPGGGPNNGGA

Number of amino acids: Acidic: 4, Basic: 15, Polar: 49, Glycine: 74, Nonpolar: 38

3. **K10 – C terminal (124 residues):**

- **Real sequence:**

GGGBGGGPNGGGPGGGPPGGGPPGGGBGGGBGGPPGGGPGGGPPGGGPPGGGPGGGPPPGGBGGPPPGGPGGGPPG
GGGGGPGGGPPGGGPPPGGGPGGGPPPGGBBPPP PGPNGAPPPBGNBB

Number of amino acids: Acidic: 1, Basic: 9, Polar: 41, Glycine: 70, Nonpolar: 25

- **Approximated sequence with known angles/dihedral angles:**

GGGPBPGGGPNGGGPNGGGPNGGGPBPGGGPBPGGGPNGGGPNGGGPNGGGPNGGGGPBPGGGGGGGPNGGGPN
GGGGGGPNGGGPNGGGPNGGGPBPPPBPGGGNGGAGGNNPGGGPBPPPBB

Number of amino acids: Acidic: 1, Basic: 9, Polar: 30, Glycine: 69, Nonpolar: 15

4. **K1 – C terminal (151 residues):**

- **Real sequence:**

PGANNNPNPNPNPPPBPPNPGGGPBGGGGGGPGPGGPPPGPGGGPPGPGGGGGGGBGPPGPGGPPPGPGGGPPGPGG
GGGGBGPPGPGPPPGGPBGGPGGGGGGPPGGBGPGGGPPGGPNGGBGPPPGGNBPPGGPPPNBNNPPPPPGNPB

Number of amino acids: Acidic: 1, Basic: 10, Polar: 58, Glycine: 69, Nonpolar: 13

- **Approximated sequence with known angles/dihedral angles:**

APGPNGGGGNNPGPNNGGGPBPPPPPBPGGGPNGGGGGGGGPNGGGPNGGGGGGGGGPBPGGGGGPNGGGPNGGG
PBPPPPPPPPPPBPGGGPNGGGPNGGGPBPGGGPNGGGPBPPPPPBPGGGGGNNPGGGPNGGGPBPPPPPPPPPB

Number of amino acids: Acidic: 1, Basic: 10, Polar: 55, Glycine: 69, Nonpolar: 16

One keratin IF of 50 nm consists of 32 chains which gives 64 terminals sticking out from the surface. With ¼ representation of keratin I, that provides 8 grafted “tails” and 8 “heads” (16 terminals). Coiled coil is built up of two parallel KRT1/KRT10 monomers while tetramers (2 coiled coils) by the anti-parallel arrangement. In that case, two N terminals of KRT1/KR10 are grafted next to each other. That approach is also taken for two C terminals which stick on the same chain of the IF. All 16 terminals with the initial line chain conformation were uniformly grafted on the keratin IF surface which is presented in Figure 7.20. The box length in the Z axis corresponds to the length of the keratin IF (12.5 nm) being copied across periodic boundary conditions.

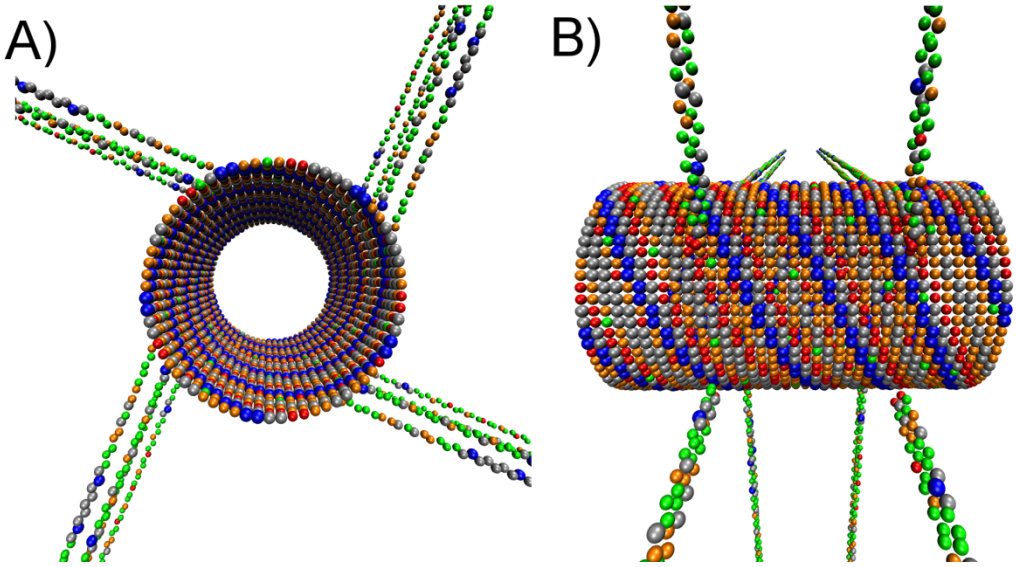


Figure 7.20. Representation of the intermediate filament with grafted terminals in two axial projections along its: A) diameter and B) length. Acidic – red, basic – blue, glycine – green, polar – grey, nonpolar – orange.

7.4.2.3. The Application of the Force Field

Starting from the conformation presented in Figure 7.20 the simulation was run using position restraint dynamics for the keratin IF beads. The position of each terminal bead being the closest to the IF surface was also restraint. The low force constant of 120 kcal mol^{-1} nm^{-2} was used in the harmonic position restraint dynamic to reach realistic flexibility of the system. Simulation was run using 8 fs time step within NVT ensemble for 4.2 μs at 300K.

The immediate adsorption of the all terminals was observed after 5 ns. In Figure 7.21 the snapshots from the simulation are shown: initial and final configuration. The blue colour corresponds to the IF and red colour to the grafted terminals. Keratin "heads" and "tails" remained mainly within the IF surface forming loops due to the interactions between basic residues at the end of each terminals and acidic IF surface area.

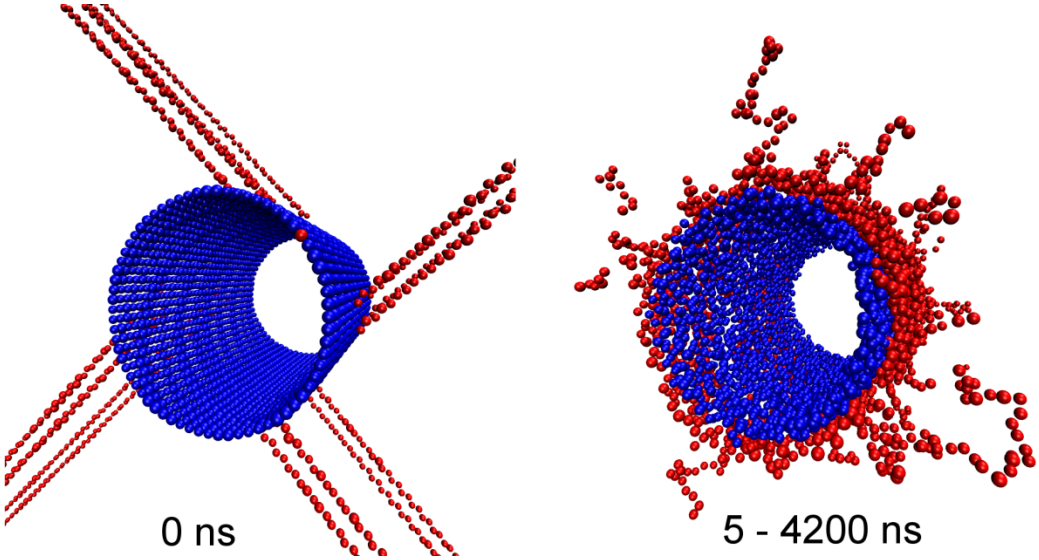


Figure 7.21. The Initial and the final conformation of the intermediate filament (blue) with grafted chains (red). After 5 ns the rapid adsorption of terminals was observed with formed loops.

Radial distribution functions (RDFs) of terminals were calculated over last 500 ns - the number of atoms at given distance away from the IF surface in two dimensions – XY axis (Z axis correspond to the IF length). In Figure 7.22 the RDFs are presented for: N and C; N+C; N1, C1, N10, C10 as well as each terminal basic residues of N and C which are present at the end of each terminal close to the bulk. All RDF results are compared to Akinshina et al[102] SCF theory model for the same system (AA – all atoms) where the amino acids volume fraction versus the distance was calculated (Figure 7.23). Distributions in both cases (coarse-grain model and Akinshina et al.) show maxima close to the surface.

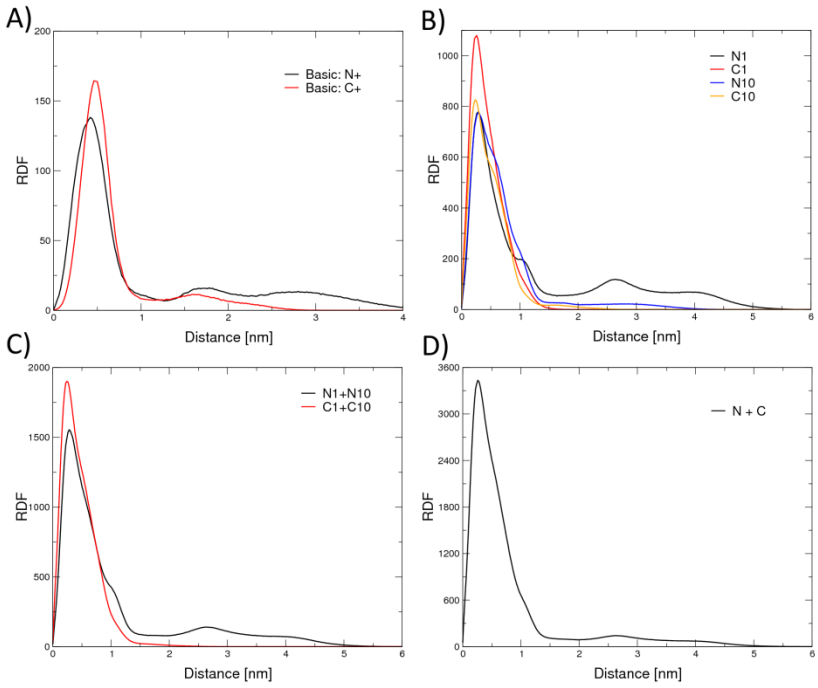


Figure 7.22. Radial distribution functions (RDFs) as the "heads" and "tails" number of atoms away from the intermediate filament surface in two dimensions – XY components averaged over NVT simulation at 300 K. RDFs correspond to distribution of terminals residues: A) Basic of N and C terminal; B) N1, C1, N10, C10 terminals; C) N1+N10 terminal and C1+C10 terminal; D) N+C terminals.

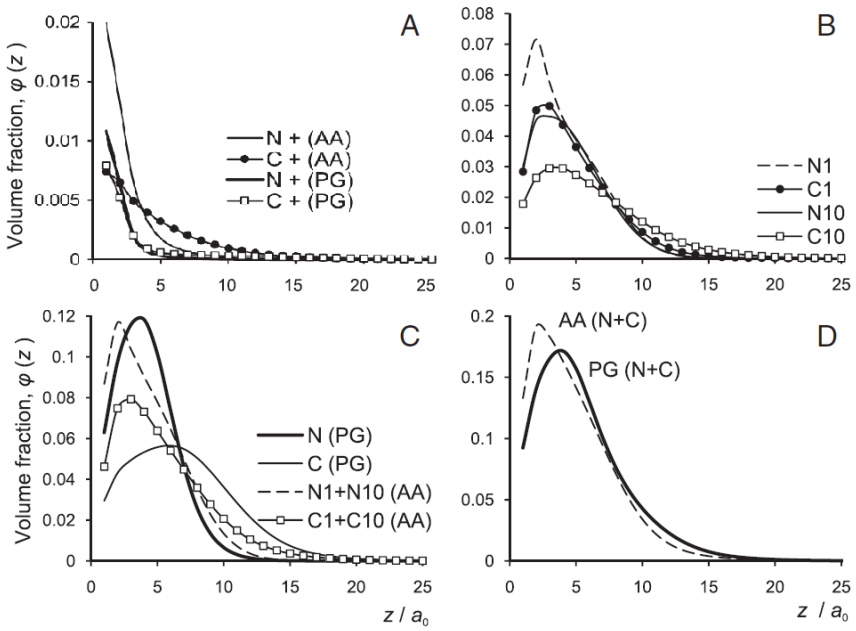


Figure 7.23. Results by Akinshina et al[102]: Volume fractions of residues away from the intermediate filament surface: A) Basic of N and C terminal; B) N1, C1, N10, C10 terminals; C) N1+N10 terminal and C1+C10 terminal; D) N+C terminals. Results from the coarse-grain model can be compared to all atom model (AA) in this figure.

RDFs of the coarse-grain model show that the highest number of atoms belonging to the N and C terminals remained close to the intermediate filament surface. Basic residues which are placed at the end of each terminal (in the bulk) had the highest distribution close to the oppositely charged surface. Hence, terminals form loops (Figure 7.21 B) and are able to bridge interactions with possible another surface of the IF. Those results are with agreement to the SCF model (Figure 7.23). It was also observed that N and C terminals have distributions which are similarly narrowed, although the N terminals have slightly higher propensity (in comparison to C terminal) for being further away from the surface.

The final configuration at 10 μs of the IF with adsorbed "heads" and "tails" was used and copied as an input for studying interaction between keratin IFs. In the Figure 7.24 the initial configuration for the simulation involving 7 IFs with grafted, adsorbed chains is presented. With the same simulation settings as for one IF the simulation was run for 10 μs.

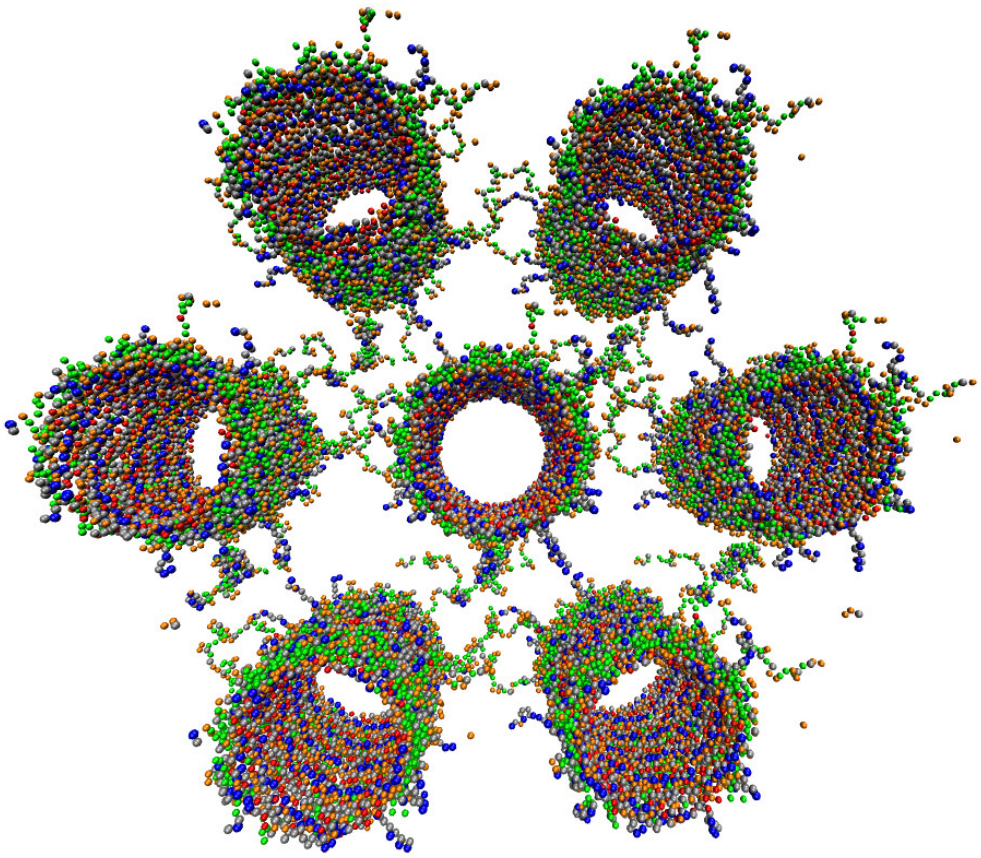


Figure 7.24. Initial configuration for the simulation of studying interactions between keratin filaments (glycine – green, polar – grey, nonpolar – orange, acidic – red, basic – blue).

It was observed over the simulation time that terminals remained mainly within each keratin IF surface to which they were attached. To assess the distributions of keratin N and C terminals grafted to IF surface, RDFs in the identical way over last 500 ns for the central filament and its grafted chains were calculated. In Figure 7.25 RDFs of basic residues as well as two N and two C terminals away from the keratin IF surface are shown. It is observed that the presence of other IFs provided slightly wider distributions (in both basic residues and all C and N terminals residues) in comparison to the single IF. That means that terminals of one IF can interact with terminals of another IF.

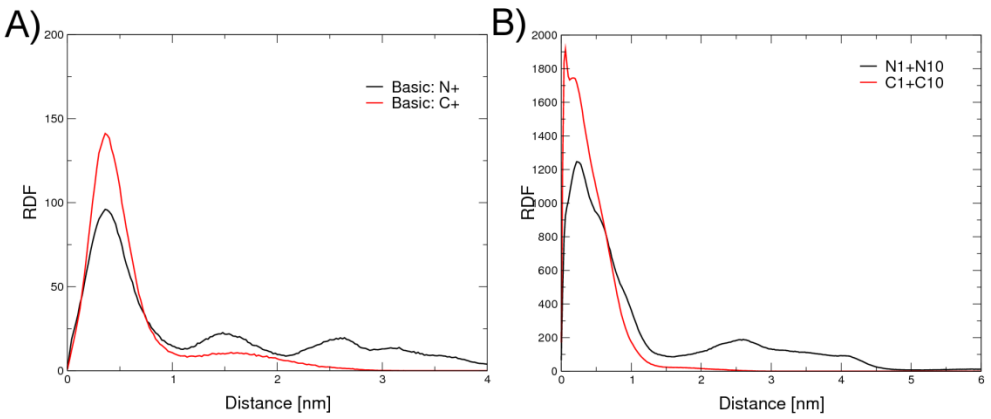


Figure 7.25. Radial distribution functions (number of atoms away from the intermediate filament surface in two dimensions – XY components) obtained from NVT simulation at 300 K of terminals residues in the system of 7 intermediate filaments (each with 16 terminals). RDFs correspond to distribution of terminal residues of the central IF and its grafted chains: A) Basic of N and C terminal; B) N1+N10 terminal and C1+C10 terminal.

7.4.2.4. The Role of Natural Moisturizing Factor

Akinshina et al. [102] provided the analysis of the NMF composition on terms of: polar, nonpolar, basic, acidic and glycine residues. In Table 7.3 the composition of NMF solution of the corneocytes in which keratin is embedded is presented.

Table 7.3. The composition of natural moisturizing factor mixture in terms of acidic, basic, polar, nonpolar, glycine and water mass concentrations.

Molecule Type	Mass Concentration [%]
Acidic	11.2
Basic	11.2
Polar	13.2
Nonpolar	25.5
Glycine	10.5
Water	27.9

With this information the assessment of the given type of bead belonging to the NMF solution was gradually performed. Similarly the results are presented as RDFs of terminals away from the IF surface. The system involved one IF model with grafted 16 terminals, which was placed in the centre of the rectangular box: 25x25x12.5 nm. The first system corresponded to (in addition to IF with grafted chains) uniformly and freely placed 11.2% of acidic and 11.2% of basic mass concentration of beads. The second system in addition to the first one corresponded to additionally placed 13.2% mass concentration of polar beads. In the third system 25.5% of the nonpolar beads were added and the fourth system corresponded to the full NMF composition (additional 10.5% of glycine). All four systems were run for 10 μs. The influence of adding given type of NMF solution bead can be assessed as RDFs of terminals away from the IF surface in XY components (Figure 7.26).

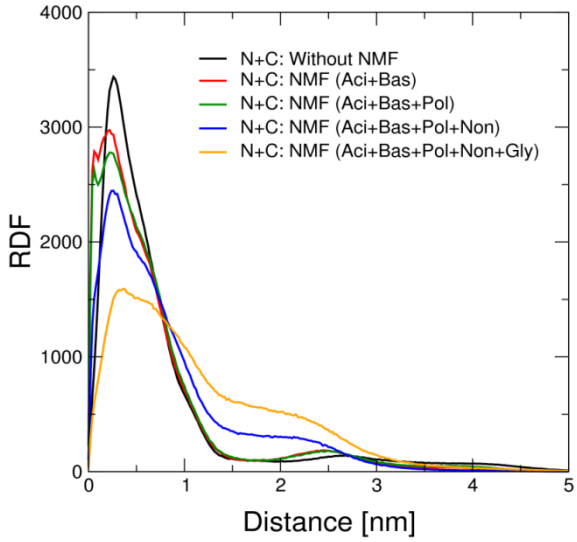


Figure 7.26. Radial distribution functions of (number of atoms away from the intermediate filament surface in two dimensions – XY components) obtained from NVT simulation at 300 K of all terminals residues in the system of 1 intermediate filament with 16 terminals attached: black – without NMF, red – with addition of acidic and basic beads of NMF, green – with addition of extra NMF polar beads, blue – with addition of extra NMF nonpolar beads, yellow – with full composition of NMF.

It can be observed that the addition of acidic, basic and polar beads of the NMF mixture slightly influenced the distribution of grafted chains which were mainly placed within keratin IF surface. However, the addition of nonpolar and finally glycine beads (full NMF mixture) provided much wider distributions of terminals away from the surface. It is concluded that the NMF role is to prevent the attractive forces between keratin IFs due to the exposure of terminals to the water solution. In that case terminals of one IF can interact with opposite IFs and its grafted chains mediating interactions and preventing the attraction between keratin bundles. In comparison, without NMF terminals which remain within IF surface cause their collapse and thus provide low elastic properties of the human outermost skin layer. The addition of glycine in the NMF solution provided the greatest recovery of terminals. That confirms Jokura et al.[22] experimental findings where glycine was observed to have the highest influence on the skin elasticity.

7.5. Conclusion

For the first time in this work, the coarse-grained force field development and its application for the most abundant stratum corneum keratin pair: KRT1/KR10 has been conducted. For this purpose, initially the sequence of the representative terminal was chosen and a long full atomistic simulation with explicit solvent of 2.1 μs was performed using temperature annealing. That ensured that all possible energy states are explored. Trajectory was divided into two parts until both represented similar structural properties. The distribution of the protein: gyration radius, hydrophobic solvent accessible surface area, root mean square fluctuation of central carbon atoms, end to end distance of terminal residues as well as mean distances map of each residue with respect to all residues were generated for both parts reaching excellent convergence. The coarse-grain force field was aimed to include explicit water model and electrostatics in tabulated potentials as well as 5 relevant types of beads (residues): polar, nonpolar, basic, acidic and glycine. Bonds were constraints to the Cα - Cα distance (0.38 nm). Angles and dihedral angles between 3 and 4 consecutive atoms respectively were reproduced by Boltzmann inversion method and with applied fully repulsive non-bonded potential (WCA) by 10 μs of the simulation. The next step involved reproduction of non-bonded potentials and the great agreement between atomistic RDFs and coarse-grain simulation was reached after 10 000 iterations using iterative Boltzmann inversion method. The radius of gyration within coarser representation reproduced all atom simulation very well.

First the developed force field was applied to the system lacking natural moisturizing factor (NMF). This involved two systems: with one IF and with seven IFs (all with attached 16 terminals). It was observed in both cases that terminals mainly remained within keratin IF surfaces without possibility of interaction in the opposite keratin bundle. That system lacks in elastic properties between keratin bundles which presumably would collapse. It was also observed that grafted terminals form loops due to presence of basic residues at the end of each chain which is in the agreement of Akinshina et al.[102] SCF model.

Secondly, the NMF solution was gradually added to the system of one IF with grafted 16 terminals. It was observed that addition of acidic, basic and polar residues did not provide the recovery for distributions of terminals which were mainly within the keratin IF surface. The addition of nonpolar and glycine beads of the NMF solution provided much wider distributions of terminals. That shows their possible interactions with opposite IF surface and its grafted chains which provides the recovery of the skin elasticity. This result is in

agreement with Jokura et al.[22] experimental study where glycine was responsible for lowering the attractive forces between keratin IFs.

Finally, the experimental validation by circular dichroism (CD) was provided by Dr Alfonso De Simone (Appendix III) on the synthesized 60 residues surrogate. It was observed that the surrogate does not exhibit any secondary structure which is in agreement with all atom and CG simulation.

CHAPER VIII

Conclusions and Future Perspectives

8.1. Overall conclusions

Today, computational chemistry plays a crucial role in predicting thermodynamic and structural properties in most industrial branches. The importance of molecular simulations is significantly increasing due to the gained agreement with experimental data. For instance, the application of molecular dynamics (MD) in the development of personal care products as well new medicines in pharmaceutical industry provide much faster technique of lower costs with much detailed insight in comparison to experimental methods. MD is a computational method which enables observing the movement of atoms with atomistic details. It is an excellent tool for studying e.g. small active molecules with beneficial influence which are used in skin and hair care products interacting with human biosubstrate (keratin). The binding sites with the highest affinity as well as structural changes of protein-ligand interactions can be assessed within atomistic details using MD. The binding affinity of small molecules to proteins is determined by the binding free energy difference. The latter one provides the binding constant which enables extracting the equilibrium concentration of free ligand, protein and protein-ligand complex. That is the crucial information for industry which further enables developing new products. MD enables extracting equilibrium thermodynamic properties of the studied system e.g. binding free energy in a less expensive way in comparison to experiment. However, MD has also its limitations - the achievable time scales and thus the size of the studied systems. That arises from the available computational resources.

In this thesis MD simulations were employed in order to provide better understanding of the experimental results conducted in parallel (Dr Yanyan Zhao, China Agricultural University). In addition the aim was to overcome the main limiting factor of MD – the simulation time. For this purpose by the use of MD simulation framewrork, thermodynamic and detailed structural basis have been delivered for small molecules with significantly varying properties interacting with keratin. MD results explained and validated experimental data extremely well. On the top of this, a fast free energy prediction tool has been built within all atom force field by a use of steered molecular dynamics (SMD) alone. Within the coarse grain approach, the force field was developed and applied for studying elastic properties of human skin. The force field enables orders of magnitude faster than all atom force fields simulations.

In particular the following achievements have been gained within MD:

1. Detailed Explanation and Validation of Experimental Results

a) Keratin –EGCG Interactions

Dr Yanyan Zhao by the use ultrafiltration experiment observed the multilayer isotherm which described the interactions of EGCG molecules with keratin. The experimental free energy obtained within isothermal titration calorimetry (ITC) at 298K corresponded to ΔG_{ITC}= -6.37 kcal mol^{-1}.

Within MD the hydrophobic collapse followed by hydrogen bonding and often aromatic stacking were observed to be the governing interactions. EGCG molecules were shown to possess the highest binding affinity towards acidic residues. The multilayer assembly of EGCG molecules on the keratin surface was confirmed within MD. In addition, the binding affinity was observed to decrease with the temperature increase. Although in the free energy calculations the only one reaction coordinate (distance perpendicular to the helical axis) was studied a good agreement between experiment and simulations was gained. The binding free energy at 298K using MD was ΔG_{MD}= -6.20 kcal mol^{-1}. The linear correlation between the binding free energy and the maximum pulling force obtained from short pulling simulations was observed. It was concluded that based on this correlation it might be possible to use SMD alone for predicting the binding free energy for other small molecules binding helical motifs in orders of magnitude faster time than free energy methods within MD.

b) Keratin – Fe^{2+} Complex

Dr Yanyan Zhao using ITC method obtained the binding free energy at three different temperatures (ΔG_{ITC298} =-7.43 kcal mol^{-1}, ΔG_{ITC308} =-7.69 kcal mol^{-1}, ΔG_{ITC318} =-8.00 kcal mol^{-1}) for ferrous ion interacting with keratin. The data was best fitted by the "one set of identical binding site" model and a significant entropic contribution was observed.

In MD simulations the nonbonded model with Lennard-Jones potential for iron (II) ion was tested. It appeared that the simplest nonbonded model provided a surprising great agreement between experimental free energies and those obtained by simulations. MD simulations explained the entropic contribution in terms of the displacement of water molecules on the protein surface. The only one stable conformation was shown to be the acidic pocket providing the insight into the one binding site used in ITC data. Binding free energies at three temperatures agreed within the 3% error to the experimental data (ΔG_{MD298} =-7.20 kcal mol^{-1}, ΔG_{MD308} =-7.60 kcal mol^{-1}, ΔG_{MD318} =-7.90 kcal mol^{-1}).

2. Overcoming the Time Scales Limitation of MD

a) Ligand – α-Helix Fast Free Energy Predictions

The fast free energy method based on couple of SMD simulations: linear correlation of the rupture force versus the free energy, developed for the keratin-EGCG system was tested for other small molecules with significantly varying properties. It was shown that this correlation work very well for molecules with -1.721 < *logP* < 2.08 and those with overall charge of ±1 binding the most common α-helical motif. Although the method is based on small number of molecules it provides a great starting point for future fast free energy predictions. In this manner the free energy can be obtained by couple of short SMD simulations (hundreds of picoseconds to nanoseconds) instead of time consuming free energy methods (hundred of nanoseconds to microseconds). On the top of this, the rupture force versus the applied spring constant correlations were shown for each molecule. The type interactions involved in α-helix – ligand complex together with thermodynamic equilibrium has been also investigated and agreed well with existing experimental data.

b) Elastic Properties of Skin - The Coarse-Grain Force Field

For the first time the coarse grain force field was developed for the interactions between keratin bundles which are directly related to elastic properties of the outermost skin layer (stratum corneum). The coarser representation involves the presence of the explicit solvent model within tabulated potentials which significantly reduced the computational time. Based on the long full atomistic simulation of 83000 atoms the coarse representation corresponded to 60 beads and reproduced the polypeptide structural properties very well. Hence, much bigger systems with longer time scales can be studied in the assessment of interactions between keratin filaments. What is more, the secondary structure of the 60 residues synthesized surrogate was validated by Dr Alfonso De Simone (Imperial College London) using circular dichroism. The agreement of a random coil was gained between experiment and simulations. The coarse-grain model was applied and assessed the influence of the natural moisturizing factor components. Glycine was shown to possess the greatest recovery on the grafted terminals which enables interactions with opposite filaments (elasticity). That is with agreement to previous experimental studies by Jokura et al.[22]

8.2. Future perspectives

The use of molecular dynamics simulations presented in this thesis enabled extracting detailed structural behaviour as well as thermodynamic properties between the main protein of human skin and hair (keratin) and small molecules used as additives in the personal care products. The industrial implementation of MD described herein provides a faster, more detailed and cheaper insight into free energy predictions than experimental methods. The developed coarse-grain force field has application in predicting the concentration of given molecules in skin care products.

The future perspectives based on results gained within this thesis are as following:

1. The developed free energy model based on pulling simulations via steered MD only may provide a new methodology for precise and fast free energy calculations. However, the model requires testing for larger number of molecules followed by experimental methods. This new method may be also implemented in other systems i.e. for other biosubstrate interacting with ligands and e.g. applied in toxicology measurements, proteins interacting with pharmaceuticals and protein-protein interactions. By gaining orders of magnitudes faster free energy predictions, this method may have a wide implementation for the current achievable time scales by molecular dynamics and possibly replace the experimental measurements.

2. The built coarse-grain force field with water included within tabulated potentials provides an implementation for personal care products development. This work by combination of experimental measurement (CD) and MD simulations leads to much faster calculations which are the main limiting factor in all-atom force fields. However, in order to further validate the force field the nuclear magnetic resonance (NMR) method will be performed to measure the polypeptide diffusivity. That property enables calculating the hydrodynamic radius, thus the radius of gyration which can be directly compared to all atom and coarse grain model. The applied force field for interactions between keratin bundles has capability of assessing the influence of NMF and other small molecules used in skin care products which will be conducted in the near future. Different concentrations of NMF were assessed for the best elasticity between IFs. However, those calculations were performed with the position restraints dynamics for the keratin IF. Hence, the free energy calculations will be performed to assess the equilibrium distance between keratin bundles in order to match the experimental value of 8.2 nm. These calculations will enable personal care industry to develop products with the optimum concentration of given type of molecules. In addition, the methodology of including the water within tabulated potentials in coarse grained force fields can be applied in many biological systems enabling studying big systems with much faster performance.

BIBLIOGRAPHY

1. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. Journal of Computer-Aided Molecular Design. **2012,** 26, 15-26.

2. Anisimov VM, Ziemys A, Kizhake S, Yuan Z, Natarajan A, Cavasotto CN. Computational and experimental studies of the interaction between phospho-peptides and the C-terminal domain of BRCA1. J Comput Aided Mol Des. **2011,** 25, 1071-1084.

3. Dubey KD, Tiwari RK, Ojha RP. Recent Advances in Protein-Ligand Interactions: MD Simulations and Binding free Energy. Curr Comput Aided Drug Des. **2013**.

4. Talhout R, Villa A, Mark AE, Engberts JB. Understanding binding affinity: a combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. Journal of the American Chemical Society. **2003,** 125, 10570-10579.

5. Kastris PL, Bonvin AM. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. Journal of the Royal Society Interface. **2013,** 10. 10.1098/rsif.2012.0835

6. Yokozawa T, Nakagawa T, Kitani K. Antioxidative activity of green tea polyphenol in cholesterol-fed rats. Journal of Agricultural and Food Chemistry. **2002,** 50, 3549-3552.

7. Yang H, Sun DK, Chen D, Cui QC, Gu YY, Jiang T, Chen W, Wan SB, Dou QP. Antitumor activity of novel fluoro-substituted (-)-epigallocatechin-3-gallate analogs. Cancer Letters. **2010,** 292, 48-53.

8. Kwon OS, Han JH, Yoo HG, Chung JH, Cho KH, Eun HC, Kim KH. Human hair growth enhancement in vitro by green tea epigallocatechin-3-gallate (EGCG). Phytomedicine. **2007,** 14, 551-555.

9. Afaq F, Katiyar SK. Polyphenols: Skin Photoprotection and Inhibition of Photocarcinogenesis. Mini-Reviews in Medicinal Chemistry. **2011,** 11, 1200-1215.

10. La MD, Magri A, Campagna T, Campitiello MA, Raiola L, Isernia C, Hansson O, Bonomo RP, Rizzarelli E. A doppel alpha-helix peptide fragment mimics the copper(II) interactions with the whole protein. Chemistry. **2010,** 16, 6212-6223.

11. Smith SJ, Du K, Radford RJ, Tezcan FA. Functional, metal-based crosslinkers for alpha-helix induction in short peptides. Chem Sci. **2013,** 4, 3740-3747.

12. Xu J. An insight into interaction of Fe2+ with glycylglycine: A DFT study. Journal of Molecular Structure: THEOCHEM. **2005,** 757, 171-174.

13. Bechtolda T, Turcanua A, Ganglbergerb E, Geissler E. Natural dyes in modern textile dyehouses — how to combine experiences of two centuries to meet the demands of the future? Journal of Cleaner Production. **2003,** 11, 499-509.

14. Ferreira ES, Hulme AN, McNab H, Quye A. The natural constituents of historical textile dyes. Chemical Society Reviews. **2004,** 33, 329-336.

15. Berenstein EF, Underbill CB, Lakkakorpi J, Ditre CM, Uitto J, Yu RJ, Scotte V. Citric Acid Increases Viable Epidermal Thickness and Glycosaminoglycan Content of Sundamaged Skin. Dermatologic Surgery. **1997,** 23, 689-694.

16. Naik A, Pechtold LARM, Potts RO, Guy RH. Mechanism of oleic acid-induced skin penetration enhancement in vivo in humans. Journal of Controlled Release. **1995,** 37, 299-306.

17. Williams AC, Barry BW. Skin Absorption Enhancers. Critical Reviews in Therapeutic Drug Carrier Systems. **1992,** 9, 305-353.

18. Bharat Bhushan. Biophysics of Human Hair: Structural, Nanomechanical, and Nanotribological Studies. Chapter: Introduction-Human Hair, Skin, and Hair Care Products. 2010. 978-3-642-15900-8 (Print) 978-3-642-15901-5 (Online).

19. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal. **1998,** 75, 422-427.

20. Heard CM, Monk BV, Modley AJ. Binding of primaquine to epidermal membranes and keratin. International Journal of Pharmaceutics. **2003,** 257, 237-244.

21. Gu LH, Coulombe PA. Keratin finction in skin epithelia: a broadening palette with surprising shades. Current Opinion in Cell Biology. **2007,** 19, 13-23.

22. Jokura Y, Ishikawa S, Tokuda H, Imokawa G. Molecular Analysis of Elastic Properties of the Stratum-Corneum by Solid-State C-13-Nuclear Magnetic-Resonance Spectroscopy. Journal of Investigative Dermatology. **1995,** 104, 806-812.

23. Zhao Y, Chen L, Yakubov G, Aminiafshar T, Han L, Lian G. Experimental and Theoretical Studies on the Binding of Epigallocatechin Gallate to Purified Porcine Gastric Mucin. The Journal of Physical Chemistry B. **2012**. doi: 10.1021/jp212059x

24. Wang LM, Chen LJ, Lian GP, Han LJ. Determination of partition and binding properties of solutes to stratum corneum. International Journal of Pharmaceutics. **2010,** 398, 114-122.

25. Zewert TE, Pliquett UF, Vanbever R, Langer R, Weaver JC. Creation of transdermal pathways for macromolecule transport by skin electroporation and a low toxicity, pathway-enlarging molecule. Bioelectrochemistry and Bioenergetics. **1999,** 49, 11-20.

26. Hansen S, Naegel A, Heisig M, Wittum G, Neumann D, Kostka KH, Meiers P, Lehr CM, Schaefer UF. The Role of Corneocytes in Skin Transport Revised-A Combined Computational and Experimental Approach. Pharmaceutical Research. **2009,** 26, 1379-1397.

27. Chen LJ, Lian GP, Han LJ. Use of "bricks and mortar" model to predict transdermal permeation: Model development and initial validation. Industrial & Engineering Chemistry Research. **2008,** 47, 6465-6472.

28. Cork MJ. The importance of skin barrier function. Journal of Dermatological Treatment. **1997,** 8, S7-S13.

29. Wang LM, Chen LJ, Han LJ, Lian GP. Kinetics and Equilibrium of Solute Diffusion into Human Hair. Annals of Biomedical Engineering. **2012,** 40, 2719-2726.

30. Aburjai T, Natsheh FM. Plants Used in Cosmetics. Phytotherapy Research. **2003,** 17, 987-1000.

31. Siebert KJ, Troukhanova NV, Lynn PY. Nature of polyphenol-protein interactions. Journal of Agricultural and Food Chemistry. **1996,** 44, 80-85.

32. Mantena SK, Meeran SM, Elmets CA, Katiyar SK. Orally administered green tea polyphenols prevent ultraviolet radiation-induced skin cancer in mice through activation of cytotoxic T cells and inhibition of angiogenesis in tumors. Journal of Nutrition. **2005,** 135, 2871-2877.

33. Record IR, Dreosti IE. Protection by black tea and green tea against UVB and UVA+B induced skin cancer in hairless mice. Research/Fundamental and Molecular Mechanisms of Mutagenesis. **1998,** 422, 191-199.

34. Singh BN, Shankar S, Srivastava RK. Green tea catechin, epigallocatechin-3-gallate (EGCG): Mechanisms, perspectives and clinical applications. Biochemical Pharmacology. **2011,** 82, 1807-1821.

35. Katiyar SK. Green tea prevents non-melanoma skin cancer by enhancing DNA repair. Archives of Biochemistry and Biophysics. **2011,** 508, 152-158.

36. Siddiqui IA, Tarapore RS, Mukhtar H. Prevention of skin cancer by green tea Past, present and future. Cancer Biology & Therapy. **2009,** 8, 1288-1291.

37. Katiyar S, Elmets CA, Katiyar SK. Green tea and skin cancer: photoimmunology, angiogenesis and DNA repair. The Journal of Nutritional Biochemistry. **2007,** 18, 287-296.

38. Yao LH, Jiang YM, Shi J, Tomas-Barberan FA, Datta N, Singanusong R, Chen SS. Flavonoids in food and their health benefits. Plant Foods for Human Nutrition. **2004,** 59, 113-122.

39. Bowe WP. Cosmetic benefits of natural ingredients: mushrooms, feverfew, tea, and wheat complex. J Drugs Dermatol. **2013,** 12, s133-s136.

40. Baumann L, Woolery-Lloyd H, Friedman A. "Natural" ingredients in cosmetic dermatology. J Drugs Dermatol. **2009,** 8, s5-s9.

41. Dweck AC. Natural ingredients for colouring and styling. Int J Cosmet Sci. **2002,** 24, 287-302.

42. dal Belo SE, Gaspar LR, Maia Campos PM, Marty JP. Skin penetration of epigallocatechin-3-gallate and quercetin from green tea and Ginkgo biloba extracts vehiculated in cosmetic formulations. Skin Pharmacol Physiol. **2009,** 22, 299-304.

43. Lotito SB, Fraga CG. (+)-Catechin as antioxidant: mechanisms preventing human plasma oxidation and activity in red wines. Biofactors. **1999,** 10, 125-130.

44. Molan AL, De S, Meagher L. Antioxidant activity and polyphenol content of green tea flavan-3-ols and oligomeric proanthocyanidins. International Journal of Food Sciences and Nutrition. **2009,** 60, 497-506.

45. Silva AR, Menezes PF, Martinello T, Novakovich GF, Praes CE, Feferman IH. Antioxidant kinetics of plant-derived substances and extracts. Int J Cosmet Sci. **2010,** 32, 73-80.

46. Fukushima Y, Ohie T, Yonekawa Y, Yonemoto K, Aizawa H, Mori Y, Watanabe M, Takeuchi M, Hasegawa M, Taguchi C, Kondo K. Coffee and green tea as a large source of antioxidant polyphenols in the Japanese population. Journal of Agricultural and Food Chemistry. **2009,** 57, 1253-1259.

47. Sharma V, Rao LJ. A thought on the biological activities of black tea. Crit Rev Food Sci Nutr. **2009,** 49, 379-404.

48. Ferguson LR, Philpott M. Nutrition and mutagenesis. Annual Review of Nutrition. **2008,** 28, 313-329.

49. Krizkova L, Chovanova Z, Durackova Z, Krajcovic J. Antimutagenic in vitro activity of plant polyphenols: Pycnogenol and Ginkgo biloba extract (EGb 761). Phytotherapy Research. **2008,** 22, 384-388.

50. Cabrera C, Artacho R, Gimenez R. Beneficial effects of green tea--a review. Journal of the American College of Nutrition. **2006,** 25, 79-99.

51. Winn RN, Kling H, Norris MB. Antimutagenicity of green tea polyphenols in the liver of transgenic medaka. Environmental and Molecular Mutagenesis. **2005,** 46, 88-95.

52. Ioannides C, Yoxall V. Antimutagenic activity of tea: role of polyphenols. Curr Opin Clin Nutr Metab Care. **2003,** 6, 649-656.

53. Khan HY, Zubair H, Faisal M, Ullah MF, Farhan M, Sarkar FH, Ahmad A, Hadi SM. Plant polyphenol induced cell death in human cancer cells involves mobilization of intracellular copper ions and reactive oxygen species generation: A mechanism for cancer chemopreventive action. Mol Nutr Food Res. **2013**.

54. Yiannakopoulou EC. Green Tea Catechins: Proposed Mechanisms of Action in Breast Cancer Focusing on The Interplay Between Survival and Apoptosis. Anticancer Agents Med Chem. **2013**.

55. Landis-Piwowar K, Chen D, Foldes R, Chan TH, Dou QP. Novel epigallocatechin gallate analogs as potential anticancer agents: a patent review (2. Expert Opin Ther Pat. **2013,** 23, 189-202.

56. Bhattacharya U, Mukhopadhyay S, Giri AK. Comparative antimutagenic and anticancer activity of three fractions of black tea polyphenols thearubigins. Nutrition and Cancer. **2011,** 63, 1122-1132.

57. Wahl O, Oswald M, Tretzel L, Herres E, Arend J, Efferth T. Inhibition of tumor angiogenesis by antibodies, synthetic small molecules and natural products. Current Medicinal Chemistry. **2011,** 18, 3136-3155.

58. Ohga N, Hida K, Hida Y, Muraki C, Tsuchiya K, Matsuda K, Ohiro Y, Totsuka Y, Shindoh M. Inhibitory effects of epigallocatechin-3 gallate, a polyphenol in green tea, on tumor-associated endothelial cells and endothelial progenitor cells. Cancer Sci. **2009,** 100, 1963-1970.

59. Yoshino S, Mitoma T, Tsuruta K, Todo H, Sugibayashi K. Effect of emulsification on the skin permeation and UV protection of catechin. Pharmaceutical Development and Technology. **2013**.

60. Duenas M, Salas E, Cheynier V, Dangles O, Fulcrand H. UV-visible spectroscopic investigation of the 8,8-methylmethine catechin-malvidin 3-glucoside pigments in aqueous solution: structural transformations and molecular complexation with chlorogenic acid. Journal of Agricultural and Food Chemistry. **2006,** 54, 189-196.

61. Zillich OV, Schweiggert-Weisz U, Hasenkopf K, Eisner P, Kerscher M. Release and in vitro skin permeation of polyphenols from cosmetic emulsions. Int J Cosmet Sci. **2013,** 35, 491-501.

62. Gianeti MD, Mercurio DG, Campos PM. The use of green tea extract in cosmetic formulations: not only an antioxidant active ingredient. Dermatol Ther. **2013,** 26, 267-271.

63. Yoon JY, Kwon HH, Min SU, Thiboutot DM, Suh DH. Epigallocatechin-3-gallate improves acne in humans by modulating intracellular molecular targets and inhibiting P. acnes. J Invest Dermatol. **2013,** 133, 429-440.

64. Li-hua S, Shu-peng S, Tian-yan Z, Xin-qiang T CLGHTYNL. Experimental study on tea polyphenols cream in animal acne model. Journal of Clinical Dermatology. **2007**.

65. Ikeda S, Kanoya Y, Nagata S. Effects of a foot bath containing green tea polyphenols on interdigital tinea pedis. Foot (Edinb). **2013**.

66. Esfandiari A, Kelley P. The effects of tea polyphenolic compounds on hair loss among rodents. Journal of the National Medical Association. **2005,** 97, 816-818.

67. Bigelow RL, Cardelli JA. The green tea catechins, (-)-Epigallocatechin-3-gallate (EGCG) and (-)-Epicatechin-3-gallate (ECG), inhibit HGF/Met signaling in immortalized and tumorigenic breast epithelial cells. Oncogene. **2006,** 25, 1922-1930.

68. Singh T, Katiyar SK. Green tea polyphenol, (-)-epigallocatechin-3-gallate, induces toxicity in human skin cancer cells by targeting beta-catenin signaling. Toxicology and Applied Pharmacology. **2013,** 273, 418-424.

69. (-)-Epigallocatechin gallate - The Chemical Database. The Department of Chemistry,University of Akron.

70. Katiyar SK, Elmets CA. Green tea polyphenolic antioxidants and skin photoprotection (Review). International Journal of Oncology. **2001,** 18, 1307-1313.

71. Charlton AJ, Haslam E, Williamson MP. Multiple conformations of the proline-rich protein/epigallocatechin gallate complex determined by time-averaged nuclear Overhauser effects. Journal of the American Chemical Society. **2002,** 124, 9899-9905.

72. Sazuka M, Itoi T, Suzuki Y, Odani S, Koide T, Isemura M. Evidence for the interaction between (-)-epigallocatechin gallate and human plasma proteins fibronectin, fibrinogen, and histidine-rich glycoprotein. Biosci Biotechnol Biochem. **1996,** 60, 1317-1319.

73. Byun EB, Choi HG, Sung NY, Byun EH. Green tea polyphenol epigallocatechin-3-gallate inhibits TLR4 signaling through the 67-kDa laminin receptor on lipopolysaccharide-stimulated dendritic cells. Biochemical and Biophysical Research Communications. **2012,** 426, 480-485.

74. Ermakova S, Choi BY, Choi HS, Kang BS, Bode AM, Dong Z. The intermediate filament protein vimentin is a new target for epigallocatechin gallate. Journal of Biological Chemistry. **2005,** 280, 16882-16890.

75. Wang SH, Liu FF, Dong XY, Sun Y. Thermodynamic analysis of the molecular interactions between amyloid beta-peptide 42 and (-)-epigallocatechin-3-gallate. J Phys Chem B. **2010,** 114, 11576-11583.

76. Dean E.Wilcox. Isothermal titration calorimetry of metal ions binding to proteins: An overview of recent studies. Inorganica Chimica Acta. **2008,** 361, 857-867.

77. Perron NR, Brumaghim JL. A review of the antioxidant mechanisms of polyphenol compounds related to iron binding. Cell Biochemistry and Biophysics. **2009,** 53, 75-100.

78. Sacco C, Skowronsky RA, Gade S, Kenney JM, Spuches AM. Calorimetric investigation of copper(II) binding to Abeta peptides: thermodynamics of coordination plasticity. J Biol Inorg Chem. **2012,** 17, 531-541.

79. Grossoehme NE, Akilesh S, Guerinot ML, Wilcox DE. Metal-binding thermodynamics of the histidine-rich sequence from the metal-transport protein IRT1 of Arabidopsis thaliana. Inorganic Chemistry. **2006,** 45, 8500-8508.

80. Bou-Abdallah F, Woodhall MR, Velazquez-Campoy A, Andrews SC, Chasteen ND. Thermodynamic analysis of ferrous ion binding to Escherichia coli ferritin EcFtnA. Biochemistry. **2005,** 44, 13837-13846.

81. Cook JD, Bencze KZ, Jankovic AD, Crater AK, Busch CN, Bradley PB, Stemmler AJ, Spaller MR, Stemmler TL. Monomeric yeast frataxin is an iron-binding protein. Biochemistry. **2006,** 45, 7767-7777.

82. Bou-Abdallah F, Arosio P, Santambrogio P, Yang X, Janus-Chandler C, Chasteen ND. Ferrous ion binding to recombinant human H-chain ferritin. An isothermal titration calorimetry study. Biochemistry. **2002,** 41, 11184-11191.

83. Jurinovich S, Degano I, Mennucci B. A Strategy for the Study of the Interactions between Metal-Dyes and Proteins with QM/MM Approaches: the Case of Iron-Gall Dye. Journal of Physical Chemistry B. **2012,** 116, 13344-13352.

84. Bender AE&DAB. Body Surface Area. A Dictionary of Food and Nutrition.New York: Oxford University Press . 1995.
Ref Type: Catalog

85. Epidermis Skin Anatomy. http://wellnessadvocate.com/?rtc=3162.

86. Das C, Noro MG, Olmsted PD. Simulation Studies of Stratum Corneum Lipid Mixtures. Biophysical Journal. **2009,** 97, 1941-1951.

87. Kang-Sickel JCC, Fox DD, Nam TG, Jayaraj K, Ball LM, French JE, Klapper DG, Gold A, Nylander-French LA. S-arylcysteine-keratin adducts as biomarkers of human dermal exposure to aromatic hydrocarbons. Chemical Research in Toxicology. **2008,** 21, 852-858.

88. Norlen L. Stratum corneum keratin structure, function and formation - a comprehensive review. International Journal of Cosmetic Science. **2006,** 28, 397-425.

89. Steinert PM. Structure, Function, and Dynamics of Keratin Intermediate Filaments. Journal of Investigative Dermatology. **1993,** 100, 729-734.

90. Paramio JM, Jorcano JL. Assembly Dynamics of Epidermal Keratins K1 and K10 in Transfected Cells. Experimental Cell Research. **1994,** 215, 319-331.

91. Adhesion and Debonding of Biological Soft Tissues – Stratum Corneum. http://dauskardt.stanford.edu/ken_wu/kenwu_research.html.

92. Korting HC, BraunFalco O. The effect of detergents on skin pH and its consequences. Clinics in Dermatology. **1996,** 14, 23-27.

93. Nishizuka Y. Membrane Phospholipid Degradation and Protein-Kinase-C for Cell Signaling. Neuroscience Research. **1992,** 15, 3-5.

94. Nair AB, Vaka SRK, Gupta S, Repka MA, Murthy SN. In vitro and in vivo evaluation of a hydrogel-based prototype transdermal patch system of alfuzosin hydrochloride. Pharmaceutical Development and Technology. **2012,** 17, 158-163.

95. Stephan Bielfeldt VSUEAVDPJDSK-PW. Assessment of Human Stratum Corneum Thickness and its Barrier Properties by In-Vivo Confocal Raman Spectroscopy. International Journal of Cosmetic Science. **2009,** 12, 479-480.

96. Hancewicz TM, Xiao Chunhong, Weissman J, Foy V, Zhang S, Misra M. A Consensus Modeling Approach for the Determination of Stratum Corneum Thickness Using *In-Vivo* Confocal Raman Spectroscopy. Journal of Cosmetics, Dermatological Sciences and Applications. **2012,** 2, 241-251.

97. Nakagawa N, Sakai S, Matsumoto M, Yamada K, Nagano M, Yuki T, Sumida Y, Uchiwa H. Relationship between NMF (lactate and potassium) content and the physical properties of the stratum corneum in healthy subjects. Journal of Investigative Dermatology. **2004,** 122, 755-763.

98. Visscher MO, Tolia GT, Wickett RR, Hoath SB. Effect of soaking and natural moisturizing factor on stratum corneum water-handling properties. Journal of Cosmetic Science. **2003,** 54, 289-300.

99. Nakagawa N, Sakai S, Matsumoto M, Yamada K, Nagano M, Yuki T, Sumida Y, Uchiwa H. Relationship between NMF (lactate and potassium) content and the physical properties of the stratum corneum in healthy subjects. Journal of Investigative Dermatology. **2004,** 122, 755-763.

100. Caspers PJ, Lucassen GW, Wolthuis R, Bruining HA, Puppels GJ. In vitro and in vivo Raman spectroscopy of human skin. Biospectroscopy. **1998,** 4, S31-S39.

101. Ehrhardt Proksch J-ML. The management of dry skin with topical emollients – recent perspectives. Journal der Deutschen Dermatologischen Gesellschaft. **2005,** 3, 768-774. DOI: 10.1111/j.1610-0387.2005.05068.x

102. Akinshina A, Jambon-Puillet E, Warren PB, Noro MG. Self-consistent field theory for the interactions between keratin intermediate filaments. BMC Biophys. **2013,** 6, 12.

103. Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. Current Biology. **2009,** 19, R132-R142.

104. Sperling LC. Hair anatomy for the clinician. Journal of the American Academy of Dermatology. **1991,** 25, 1-17.

105. GARN SM. Hair structure related to hair form. American Journal of Physical Anthropology. **1946,** 4, 252.

106. L'Oréal International 41, Rue Martre 92217 Clichy Cedex France. www.hair-science.com.

107. Hair structure. http://www.revalid.com/all_about_hair/hair_structure.html.

108. Langbein L, Rogers MA, Praetzel-Wunder S, Helmke B, Schirmacher P, Schweizer J. K25 (K25irs1), K26 (K25irs2), K27 (K25irs3), and K28 (K25irs4) represent the type I inner root sheath keratins of the human hair follicle. Journal of Investigative Dermatology. **2006,** 126, 2377-2386.

109. Robbins CR. Chemical and Physical Behavior of Human Hair. 5th ed. 2012, XXIII, 724p. ed. 2012

110. Norlen L, Al-Amoudi A. Stratum corneum keratin structure, function, and formation: The cubic rod-packing and membrane templating model. Journal of Investigative Dermatology. **2004,** 123, 715-732.

111. Mclean WHI, Moore CBT. Keratin disorders: from gene to therapy. Human Molecular Genetics. **2011,** 20, R189-R197.

112. Gu LH, Coulombe PA. Keratin function in skin epithelia: a broadening palette with surprising shades. Current Opinion in Cell Biology. Current Opinion in Cell Biology. **2007,** 19, 13-23.

113. Gu LH, Coulombe PA. Keratin finction in skin epithelia: a broadening palette with surprising shades. Current Opinion in Cell Biology. **2007,** 19, 13-23.

114. Glossary Term: Intermediate Filaments (IFs). http://www.mechanobio.info/.

115. Schweizer J BPCPLLLEMTMLOMPDRMWM. New consensus nomenclature for mammalian keratins. Journal of Cell Biology. **2006,** 174, 169-174.

116. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal. **1998,** 75, 422-427.

117. Peptide Bonds and Protein Structure. www.nslc.wustl.edu/.

118. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. Chembiochem. **2004,** 5, 170-176.

119. Fong JH, Keating AE, Singh M. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology. **2004,** 5.

120. Parry DAD, Steinert PM. Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism. Quarterly Reviews of Biophysics. **1999,** 32, 99-187.

121. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. Journal of Computational Chemistry. **2010,** 6, 459-466.

122. Vivoni A, Birke R, Lombardi J. Optimization of force constants with an Urey-Bradley force field avoiding normal mode crossings. Spectrochim Acta A Mol Biomol Spectrosc. **2001,** 57, 535-544.

123. Derreumaux P, Vergoten G. Effect of Urey-Bradley-Shimanouchi force field on the harmonic dynamics of proteins. Proteins. **1991,** 11, 120-132.

124. Vashisth H, Abrams CF. Ligand Escape Pathways and (Un)Binding Free Energy Calculations for the Hexameric Insulin-Phenol Complex. Biophysical Journal. **2008,** 95, 4193-4204.

125. Izrailev S, Stepaniants S, Schulten K. Applications of steered molecular dynamics to protein-ligand/membrane binding. Biophysical Journal. **1998,** 74, A177.

126. Li WH, Fu J, Cheng FX, Zheng MY, Zhang J, Liu GX, Tang Y. Unbinding Pathways of GW4064 from Human Farnesoid X Receptor As Revealed by Molecular Dynamics Simulations. Journal of Chemical Information and Modeling. **2012,** 52, 3043-3052.

127. Song Y, Peng W, Liu K, Yang P, Zhang WK, Zhang X. Exploring the Folding Pattern of a Polymer Chain in a Single Crystal by Combining Single-Molecule Force Spectroscopy and Steered Molecular Dynamics Simulations. Langmuir. **2013,** 29, 3853-3857.

128. Strzelecki J, Mikulska K, Lekka M, Kulik A, Balter A, Nowak W. AFM Force Spectroscopy and Steered Molecular Dynamics Simulation of Protein Contactin 4. Acta Physica Polonica A. **2009,** 116, S156-S159.

129. Lee W, Yang M, Schofield CJ, Marszalek PE. Nanomechanics of Ankyrin-R Repeats Probed by AFM and SMD Simulations. Biophysical Journal. **2010,** 98, 447A.

130. Guzman DL, Roland JT, Keer H, Kong YP, Ritz T, Yee A, Guan ZB. Using steered molecular dynamics simulations and single-molecule force spectroscopy to guide the rational design of biomimetic modular polymeric materials. Polymer. **2008,** 49, 3892-3901.

131. Li MS, Mai BK. Steered Molecular Dynamics-A Promising Tool for Drug Design. Current Bioinformatics. **2012,** 7, 342-351.

132. Jarzynski C. Nonequilibrium equality for free energy differences. Physical Review Letters. **1997,** 78, 2690-2693.

133. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wriggers W, Schulten K. Steered Molecular Dynamics. Lecture Notes in Computational Science and Engineering. **1999,** 4, 39-65.

134. Marzinek JK, Zhao Y, Lian G, Mantalaris A, Pistikopoulos EN, Bond PJ, Noro MG, Han L, Chen L. Molecular and thermodynamic basis for EGCG-Keratin interaction-part I: Molecular dynamics simulations. AIChE Journal. **2013,** 59, 4816-4823.

135. Kastris PL, Bonvin AM. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. Journal of the Royal Society Interface. **2013**, 10. 10.1098/rsif.2012.0835.

136. Mai BK, Viet MH, Li MS. Top Leads for Swine Influenza A/H1N1 Virus Revealed by Steered Molecular Dynamics Approach. J Chem Inf Model. **2010,** 50, 2236-2247.

137. Mai BK, Li MS. Neuraminidase inhibitor R-125489-A promising drug for treating influenza virus: Steered molecular dynamics approach. Biochemical and Biophysical Research Communications. **2011,** 410, 688-691.

138. Beveridge DL, Dicapua FM. Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems. Annual Review of Biophysics and Biophysical Chemistry. **1989,** 18, 431-492.

139. Deng YQ, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. Journal of Physical Chemistry B. **2009,** 113, 2234-2246.

140. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Accounts of Chemical Research. **2000,** 33, 889-897.

141. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M. Validation and use of the MM-PBSA approach for drug discovery. Journal of Medicinal Chemistry. **2005,** 48, 4040-4048.

142. Beveridge DL, Dicapua FM. Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems. Annual Review of Biophysics and Biophysical Chemistry. **1989,** 18, 431-492.

143. Hansson T, Marelius J, Aqvist J. Ligand binding affinity prediction by linear interaction energy methods. Journal of Computer-Aided Molecular Design. **1998,** 12, 27-35.

144. Torrie GM, Valleau JP. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. Chemical Physics Letters. **1974,** 28, 578-581.

145. Torrie GM, Valleau JP. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. Journal of Computational Physics. **1977,** 23, 187-199.

146. Huang K. Statistical Mechanics. John Wiley & Sons, New York ed. 1967

147. Isihara A. Statistical Physics. Academic Press, New York ed. 1971

148. Kastner J. Umbrella sampling. Wiley Interdisciplinary Reviews: Computational Molecular Science. **2011,** 1, 932-942.

149. McDonald IR SK. Machine Calculation of Thermodynamic Properties of a Simple Fluid at Supercritical Temperatures. Journal of Chemical Physics. **1967,** 47, 4766-4772.

150. McDonald IR SK. Examination of the adequacy of the 12–6 potential for liquid argon by means of Monte Carlo calculations. Journal of Chemical Physics. **1969,** 50, 2308-2315.

151. Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. Physical Review Letters. **1988,** 61, 2635-2638.

152. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. Journal of Computational Chemistry. **1992,** 13, 1011-1021.

153. GROMACS Tutorial: Umbrella Sampling. Department of Biochemistry, Virginia Tech.Site design copyright 2008-2012.

154. De Simone A, Kitchen C, Kwan AH, Sunde M, Dobson CM, Frenkel D. Intrinsic disorder modulates protein self-assembly and aggregation. Biophysics and Computational Biology. **2012,** 109, 6951-6956.

155. Horn JN, Kao TC, Grossfield A. Coarse-grained molecular dynamics provides insight into the interactions of lipids and cholesterol with rhodopsin. Advances in Experimental Medicine and Biology. **2014,** 796, 75-94.

156. Siuda I, Thogersen L. Conformational flexibility of the leucine binding protein examined by protein domain coarse-grained molecular dynamics. J Mol Model. **2013,** 19, 4931-4945.

157. Noid WG. Perspective: Coarse-grained models for biomolecular systems. Journal of Chemical Physics. **2013,** 139, 090901.

158. Xia Z, Bell DR, Shi Y, Ren P. RNA 3D structure prediction by using a coarse-grained model and experimental data. J Phys Chem B. **2013,** 117, 3135-3144.

159. Liu L, Chen SJ. Coarse-grained prediction of RNA loop structures. Plos One. **2012,** 7, e48460.

160. Liu C, Faller R. Conformational, dynamical. and tensional study of tethered bilayer lipid membranes in coarse-grained molecular simulations. Langmuir. **2012,** 28, 15907-15915.

161. Hadley KR, McCabe C. Coarse-Grained Molecular Models of Water: A Review. Mol Simul. **2012,** 38, 671-681.

162. Cellmer T, Fawzi NL. Coarse-grained simulations of protein aggregation. Methods in Molecular Biology. **2012,** 899, 453-470.

163. Su ZY, Wang YT. Coarse-grained molecular dynamics simulations of cobra cytotoxin A3 interactions with a lipid bilayer: penetration of loops into membranes. J Phys Chem B. **2011,** 115, 796-802.

164. Spijker P, van HB, Debertrand M, Markvoort AJ, Vaidehi N, Hilbers PA. Coarse grained molecular dynamics simulations of transmembrane protein-lipid systems. Int J Mol Sci. **2010,** 11, 2393-2420.

165. Hicks SD, Henley CL. Coarse-grained protein-protein stiffnesses and dynamics from all-atom simulations. Phys Rev E Stat Nonlin Soft Matter Phys. **2010,** 81, 030903.

166. Periole X, Marrink SJ. The Martini coarse-grained force field. Methods in Molecular Biology. **2013,** 924, 533-565.

167. Marrink SJ, Tieleman DP. Perspective on the Martini model. Chemical Society Reviews. **2013,** 42, 6801-6822.

168. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B. **2007,** 111, 7812-7824.

169. Fernandez M, Caballero J, Fernandez L, Abreu JI, Garriga M. Protein radial distribution function (P-RDF) and Bayesian-Regularized Genetic Neural Networks for modeling protein conformational stability: chymotrypsin inhibitor 2 mutants. J Mol Graph Model. **2007,** 26, 748-759.

170. V.Rühle CJALKKDA. Versatile Object-oriented Toolkit for Coarse-graining Applications. Journal of Chemical Theory and Computation. **2009,** 5, 3211-3223.

171. Dirk Reith MPFM-P. Deriving effective mesoscale potentials from atomistic simulations. Journal of Computational Chemistry. **2003,** 24, 1624-1636. DOI: 10.1002/jcc.10307

172. W Tschöp KKJBTBOH. Simulation of polymer melts. i. coarse-graining procedure for polycarbonates. Acta Polymerica. **1998,** 61-74.

173. Zhao Y, Marzinek JK, Chen L, Mantalaris A, Pistikopoulos EN, Han L, Lian G, Bond PJ, Nor MG. Molecular and Thermodynamic Basis for EGCG-Keratin Interaction - Part II: Experimental Investigation. AIChE Journal. **2013,** 59, 4824-4827. DOI: 10.1002/aic.14221.

174. Tame JRH, O'Brien R, Ladbury JE. Isothermal titration calorimetry of biomolecules. Biocalorimetry:Applications of Calorimetry to the Biological Sciences. **1998**, 27-38. John Wiley & Sons Ltd.: Chichester, Sussex.

175. Zhong Y, Chiou YS, Pan MH, Shahidi F. Anti-inflammatory activity of lipophilic epigallocatechin gallate (EGCG) derivatives in LPS-stimulated murine macrophages. **2012**, Food Chemistry, 134, 742-748, ISSN 0308-8146.

176. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry. **2004,** 25, 1605-1612.

177. Vanommeslaeghe K, Ghosh J, Polani NK, Sheetz M, Pamidighantam SV, Connolly JWD, MacKerell AD. Automation of the Charmm General Force Field for Drug-Like Molecules. Biophysical Journal. **2011,** 100, 611.

178. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. J Chem Inf Model. **2012,** 52, 3155-3168.

179. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. Journal of Computational Chemistry. **2010,** 31, 671-690.

180. ParamChem Interface. https://www.paramchem.org.

181. Hess B. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Abstr Pap Am Chem S. **2009,** 237.

182. MATCHER - Algorithm for Matching Coiled-Coil Proteins: http://cis.poly.edu/~jps/.

183. Szeverenyi I, Cassidy AJ, Chung CW, Lee BT, Common JEA, Ogg SC, Chen H, Sim SY, Goh WLR, Ng KW, Simpson JA, Chee LL, Eng GH, Li B, Lunny DP, Chuon D, Venkatesh A, Khoo KH, Mclean WHI, Lim YP, Lane EB. The human intermediate filament database: Comprehensive information on a gene family involved in many human diseases. Human Mutation. **2008,** 29, 351-360.

184. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics. 1983, 79:926–935.

185. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. **1997,** 18, 1463-1472.

186. Darden T, York D, Pedersen L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics. **1993,** 98, 10089-10092.

187. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. Journal of Chemical Physics. **1995,** 103, 8577-8593.

188. Parrinello M, Rahman A. Physical Review Letters. **1980,** 45, 1196-1199.

189. Hub JS, de Groot BL, van der Spoel D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. Journal of Chemical Theory and Computation. **2010,** 6, 3713-3720.

190. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The Double Cubic Lattice Method - Efficient Approaches to Numerical-Integration of Surface-Area and Volume and to Dot Surface Contouring of Molecular Assemblies. Journal of Computational Chemistry. **1995,** 16, 273-284.

191. Popovych N, Brender JR, Soong R, Vivekanandan S, Hartman K, Basrur V, Macdonald PM, Ramamoorthy A. Site Specific Interaction of the Polyphenol EGCG with the SEVI Amyloid Precursor Peptide PAP(248-286). The Journal of Physical Chemistry B. **2012,** 116, 3650-3658. doi: 10.1021/jp2121577

192. Maiti TK GKDS. Interaction of (-)-epigallocatechin-3-gallate with human serum albumin: fluorescence, fourier transform infrared, circular dichroism, and docking studies. Proteins. **2006,** 64, 355-362.

193. Smith SJ, Du K, Radford RJ, Tezcan FA. Functional, metal-based crosslinkers for alpha-helix induction in short peptides. Chem Sci. **2013,** 4, 3740-3747.

194. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal. **1998,** 75, 422-427.

195. Feenstra P, Brunsteiner M, Khinast J. Prediction of drug-packaging interactions via molecular dynamics (MD) simulations. International Journal of Pharmaceutics. **2012,** 431, 26-32.

196. Horn HW, Swope WC, Pitera JW et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. **2004**, Journal of Chemical Physics, 120, 9665-9678.

197. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. Proceedings of the National Academy of Sciences of the United States of America. **2005,** 102, 6825-6830.

198. Wilcox DE. Isothermal titration calorimetry of metal ions binding to proteins: An overview of recent studies. Inorganica Chimica Acta. **2008,** 361, 857-867.

199. Zhao Y, Marzinek JK, Bond PJ, Chen L, Li Q, Mantalaris A, Pistikopoulos EN, Noro MG, Han L, Lian G. A study on Fe^{2+}- alpha helical rich keratin complex formation using isothermal titration calorimetry and molecular dynamics simulation. Journal of Pharmaceutical Sciences. **2014**. 10.1002/jps.23895.

200. Li P, Roberts BP, Chakravorty DK, Merz KM. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. Journal of Chemical Theory and Computation. **2013,** 9, 2733-2748.

201. Corfield MC RA. The amino acid composition of wool. The Biochemical Journal. **1955,** 59, 62-68.

202. Anandakrishnan R, Aguilar B, Onufriev AV. H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Research. **2012,** 40, W537-W541.

203. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. Nucleic Acids Research. **2005,** 33, W368-W371.

204. Myers J, Grothaus G, Narayanan S, Onufriev A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. Proteins-Structure Function and Bioinformatics. **2006,** 63, 928-938.

205. Smith SJ, Du K, Radford RJ, Tezcan FA. Functional, metal-based crosslinkers for alpha-helix induction in short peptides. Chem Sci. **2013,** 4, 3740-3747.

206. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal. **1998,** 75, 422-427.

207. Deng YQ, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. Journal of Physical Chemistry B. **2009,** 113, 2234-2246.

208. Buch I, Sadiq SK, De Fabritiis G. Optimized Potential of Mean Force Calculations for Standard Binding Free Energies. J Chem Theory Comput. **2011,** 7, 1765-1772.

209. Chen PC, Kuyucak S. Accurate Determination of the Binding Free Energy for KcsA-Charybdotoxin Complex from the Potential of Mean Force Calculations with Restraints. Biophysical Journal. **2011,** 100, 2466-2474.

210. Jorgensen W L, Chandrasekhar J, Madura J et al. Comparison of simple potential functions for simulating liquid water, **1983**, Journal of Chemical Physics, 79, 926−935.

211. Berendsen HJC, Grigera JR, Straatsma TP.The missing term in effective pair potentials. **1987**, Journal of Physical Chemistry, 91, 6269-6271.

212. Izrailev S, Stepaniants S, Schulten K. Applications of steered molecular dynamics to protein-ligand/membrane binding. Biophysical Journal. **1998,** 74, A177.

213. Yu T, Lee OS, Schatz GC. Steered molecular dynamics studies of the potential of mean force for Peptide amphiphile self-assembly into cylindrical nanofibers. Journal of Physical Chemistry. **2013,** 117, 7543-7460.

214. He J, Xu L, Zhang S, Guan J, Shen M, Li H. Steered molecular dynamics simulation of the binding of the beta2 and beta3 regions in domain-swapped human cystatin C dimer. Journal of Molecular Modelling. **2013,** 19, 825-832.

215. Skovstrup S, David L, Taboureau O, Jorgensen FS. A steered molecular dynamics study of binding and translocation processes in the GABA transporter. Plos One. **2013,** 7, e39360.

216. Li MS, Mai BK. Steered Molecular Dynamics-A Promising Tool for Drug Design. Current Bioinformatics. **2012,** 7, 342-351.

217. Jarzynski C. Nonequilibrium equality for free energy differences. Physical Review Letters. **1997,** 78, 2690-2693.

218. Stote RH, Karplus M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. **1995**, Proteins, 23, 12-31.2

219. Aaqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. **1990**, Journal of Physical Chemistry, 94, 8021-8024.

220. Bell GI. Theoretical-Models for the Specific Adhesion of Cells to Cells Or to Surfaces. Advances in Applied Probability. **1980,** 12, 566-567.

221. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophysical Journal. **1997,** 72, 1568-1581.

222. Sekatskii SK, Benedetti F, Dietler G. Dependence of the most probable and average bond rupture force on the force loading rate: First order correction to the Bell-Evans model. Journal of Applied Physics. **2013,** 114.

223. Husson J, Pincet F. Analyzing single-bond experiments: Influence of the shape of the energy landscape and universal law between the width, depth, and force spectrum of the bond. Physical Review e. **2008,** 77.

224. Dudko OK, Hummer G, Szabo A. Intrinsic rates and activation free energies from single-molecule pulling experiments. Physical Review Letters. **2006,** 96.

225. Sangster J. Octanol-Water Partition-Coefficients of Simple Organic-Compounds. Journal of Physical and Chemical Reference Data. **1989,** 18, 1111-1229.

226. (-)-Epigallocatechin gallate. The Chemical Database The Department of Chemistry, University of Akron. **2012**.

227. Perrissoud D, Testa B. Inhibiting Or Potentiating Effects of Flavonoids on Carbon Tetrachloride-Induced Toxicity in Isolated Rat Hepatocytes. Arzneimittel-Forschung/Drug Research. **1986,** 36-2, 1249-1253.

228. ChemSpider - The free chemical database. ID 305. 2013.

229. Data for Biochemical Research. ZirChrom Separations, Inc.Retrieved January 11, 2012. ZirChrom Separations, Inc.Retrieved January 11, 2012.

230. Vanommeslaeghe K.et al. Atom typing and assignment of parameters and charges by analogy. In preparation . 22-2-2013.
Ref Type: Unpublished Work

231. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. 2012.
Ref Type: Computer Program

232. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. Journal of Chemical Physics. **2007,** 126.

233. Parrinello M, Rahman A. Phys Rev Lett. **1980,** 45, 1196-1199.

234. Cistola DP, Hamilton JA, Jackson D, Small DM. Ionization and Phase-Behavior of Fatty-Acids in Water - Application of the Gibbs Phase Rule. Biochemistry. **1988,** 27, 1881-1888.

235. Kim H, Choi Y, Kim JI, Jeong K, Jung S. Molecular Modeling Studies on the Chiral Separation of (+/-)-Catechins by Mono-succinyl-beta-cyclodextrin. Bulletin of the Korean Chemical Society. **2009,** 30, 1373-1375.

236. Ren P, Wu P, Ponder JW. Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. **2011**, Journal of Chemical Theory and Computation, 7, 3143-3161.

237. Grossfield A, Ren P, Ponder JW. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. **2004**, Journal of the American Chemical Society, 125, 15671-15682.

238. Jiao D, King C, Grossfield A et al. Simulation of Ca^{2+} and Mg^{2+} Solvation Using Polarizable Atomic Multipole Potential. **2006**, Journal of Physical Chemistry B, 110, 18553-18559.

239. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophysical Journal. **1997,** 72, 1568-1581.

240. Sekatskii SK, Benedetti F, Dietler G. Dependence of the most probable and average bond rupture force on the force loading rate: First order correction to the Bell-Evans model. Journal of Applied Physics. **2013,** 114.

241. Husson J, Pincet F. Analyzing single-bond experiments: Influence of the shape of the energy landscape and universal law between the width, depth, and force spectrum of the bond. Physical Review e. **2008,** 77.

242. Dudko OK, Hummer G, Szabo A. Intrinsic rates and activation free energies from single-molecule pulling experiments. Physical Review Letters. **2006,** 96.

243. FT Hane, SJ Attwood, Z Leonenko. Comparison of three competing dynamic force spectroscopy models to study binding forces of amyloid-â (1–42). Soft Matter. **2014**. 10.1039/C3SM52257A

244. L Donlon, D Nordin, D Frankel. Complete unfolding of fibronectin reveals surface interactions. Soft Matter. **2012,** 8, 9933-9940.

245. CK Lee, YM Wang, LS Huang, S Lin. Atomic force microscopy: Determination of unbinding force, off rate and energy barrier for protein–ligand interaction. Micron. **2007,** 38, 446-461.

246. TA Sulchek, RW Friddle, K Langry, EY Lau, H Albrecht, TV Ratto, SJ DeNardo, ME Colvin, A Noy. Dynamic force spectroscopy of parallel individual Mucin1–antibody bonds. Proceedings of the National Academy of Sciences of the United States of America. **2005,** 102, 16638-16643.

247. F Schwesinger, R Ros, T Strunz, D Anselmetti, HJ Güntherodt, A Honegger, L Jermutus, L Tiefenauer, Plückthun. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. Proceedings of the National Academy of Sciences of the United States of America. **2000,** 97, 9972-9977.

248. Pang Y-P. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. **2001**, Proteins: Structure, Function and Bioinformatics, 45, 183-189.

249. Barbana C, Perez MD, Sanchez L, Dalgalarrondo M, Chobert JM, Haertle T. Interaction of bovine alpha-lactalbumin with fatty acids as determined by partition equilibrium and fluorescence spectroscopy. Int Dairy J. **2006,** 16, 18-25.

250. Balendiran GK, Schnutgen F, Scapin G, Borchers T, Xhong N, Lim K, Godbout R, Spener F, Sacchettini JC. Crystal structure and thermodynamic analysis of human brain fatty acid-binding protein. Journal of Biological Chemistry. **2000,** 275, 27045-27054.

251. Babu CS, Lim C. Empirical Force Fields for Biologically Active Divalent Metal Cations in Water. **2006**, Journal of Physical Chemistry A, 110, 691-699.

252. Frazier RA, Papadopoulou A, Green RJ. Isothermal titration calorimetry study of epicatechin binding to serum albumin. Journal of Pharmaceutical and Biomedical Analysis. **2006,** 41, 1602-1605.

253. Lacal J, Alfonso C, Liu XX, Parales RE, Morel B, Conejero-Lara F, Rivas G, Duque E, Ramos JL, Krell T. Identification of a Chemoreceptor for Tricarboxylic Acid Cycle Intermediates DIFFERENTIAL CHEMOTACTIC RESPONSE TOWARDS RECEPTOR LIGANDS. Journal of Biological Chemistry. **2010,** 285, 23124-23134.

254. Nakagawa N, Sakai S, Matsumoto M, Yamada K, Nagano M, Yuki T, Sumida Y, Uchiwa H. Relationship between NMF (lactate and potassium) content and the physical properties of the stratum corneum in healthy subjects. Journal of Investigative Dermatology. **2004,** 122, 755-763.

255. Harvey MJ, De Fabritiis G. An implementation of the smooth particle-mesh Ewald (PME) method on GPU hardware. Journsl of Chemical Theory and Computation. **2009,** 5, 2371-2377.

256. Harvey M, Giupponi G, De Fabritiis G. ACEMD: Accelerated molecular dynamics simulations in the microseconds timescale. Journal of Chemical Theory and Computation. **2009,** 5, 1632.

257. Best RB, De SD, Mittal J. Residue-specific alpha-helix propensities from molecular simulation. Biophysical Journal. **2012,** 102, 1462-1467.

258. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Review of Modern Physics. **1992,** 64, 1045-1097. 10.1103/RevModPhys.64.1045

259. Schlick T. Molecular Modeling and Simulation. Springer. **2002,** 435-438.

260. Berendsen HJC, Postma JPM., van Gunsteren WF, DiNola A., Haak JR. Molecular-Dynamics with Coupling to an External Bath. Journal of Chemical Physics. **1984,** 81, 3684-3690.

261. Chebaro Y, Dong X, Laghaei R, Derreumaux P, Mousseau N. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. J Phys Chem B. **2009,** 113, 267-274.

262. Weeks JD, Chandler D, Andersen HC. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. The Journal of Chemical Physics. **1971,** 54, 5237.

263. Joung IS, Cheatham TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. **2008**, Journal of Physical Chemistry B, 112, 9020-9041.

264. Peters MB, Yang Y, Wang B et al. Structural Survey of Zinc Containing Proteins and the Development of the Zinc Amber Force Field (ZAFF). **2010**, Journal of Chemical Theory and Computation, 6, 2935-2947.

265. Laio A, Parrinello M. Escaping free-energy minima. Proceedings of the National Academy of Sciences of the United States of America. **2002**, 99,12562–12566.

266. Robinson JP, Sturgis J, Kumar GL. IHC STAINING METHODS - Immunofluorescence; Chapter 10, 5th Edition. 61-65.

APPENDIX I

Molecular and Thermodynamic Basis for EGCG-Keratin Interaction – Experimental Studies

Experiments conducted by Dr Yanyan Zhao (China Agricultural University, Beijing 100083, P. R. China). Those results constitute part of the publication: Zhao Y, Marzinek JK, Chen L, Han L, Mantalaris, A Pistikopoulos EN, Lian G, Bond, PJ, Noro MG. (2013), *Molecular and thermodynamic basis for EGCG-Keratin interaction-part II: Experimental investigation*. AIChE J., 59: 4824–4827. doi: 10.1002/aic.1422.

AI.1. Introduction

The following experimental results constitute the main subject of validation by molecular dynamics study described in Chapter IV. In this chapter, ultrafiltration, high performance liquid chromatography (HPLC) and isothermal titration calorimetry (ITC) methods are used to investigate the binding of EGCG to keratin extracted from sheep wool. The ultrafiltration method has been applied to quantify the thermodynamic equilibrium of EGCG binding to keratin, and the associated thermodynamic properties have been investigated by ITC method. The binding free energy determined from the ITC measurement agrees favourably with the predicted value from MD simulations reported in Chapter IV.

AI.2. Methodology

AI.2.1. Material Preparation

EGCG was provided by J&K Scientific with a purity of 98%. A keratin sample was obtained from Professor Jian Lu's group at Manchester University. The provided keratin solution (K1S) is a mixture of two main components of which the average molecular weight is 50kDa.

The samples were lyophilized and stored at -18°C in sealed bottles until use. The keratin solution (0.0040mM) was prepared by dissolving 5mg lyophilized keratin powder into 25mL 10mM pH 6.0 PBS buffer. Different concentrations of EGCG solution were prepared by dissolving EGCG powder into 10mM pH 6.0 PBS buffer and diluting to lower concentrations.

AI.2.2. Ultrafiltration

Aliquots (2mL) of the keratin solution (0.0040mM) were mixed with EGCG solution (2mL) at different concentrations at 298 K and incubated for 20min. After incubation, each mixture (4mL) was placed into an Amicon Ultra-4 centrifugal filter unit, of 3kDa cut-off (Millipore, Beijing, China.), and centrifuged at 4,000g. Control experiments were carried out using Aliquots (2mL) of Ultra-pure water mixed with EGCG solution (2mL) at different concentrations. The amount of bound EGCG was calculated from the concentration difference between the control samples and the filtrate of the test samples using the HPLC method.

Chromatographic analysis was performed on a Hitachi HPLC system equipped with a UV/vis Detector (at 280nm wavelength). The separation was performed on a Shim-pack VP-ODS (250×4.6mm ID, 5 μm) using an aqueous Glacial acetic acid: (1% in water; solvent A): acetonitrile (solvent B) =20:80 at a flow rate of 0.8 mL min^{-1}.

AI.2.3. Isothermal Titration Calorimetry

For ITC measurements, keratin solution (0.0040 mM) was prepared by dissolving keratin in 10mM PBS buffer at pH 6.0. EGCG solution (0.3027mM) was then dissolved in the same buffer used for keratin solution. An ITC200 titration microcalorimeter (MicroCal, Inc., Beijing, China) was used at 298 K. The duration of each injection was 5s, and the time interval between injections was 180s. The solution in the cell was stirred at 500 rpm by the syringe to ensure thorough mixing. Control experiments were also carried out for the titration of EGCG solution into buffer and buffer into the keratin solution. The heat of the control experiments was subtracted from the titration experiment of EGCG and keratin titration, to remove the effects of association and dilution.

AI.3. Results

AI.3.1. HPLC for EGCG

A typical chromatogram obtained in analysis of the EGCG by HPLC is shown in Figure AI.1. Figure AI.2 shows the standard curve used in determining the amount of EGCG by HPLC.

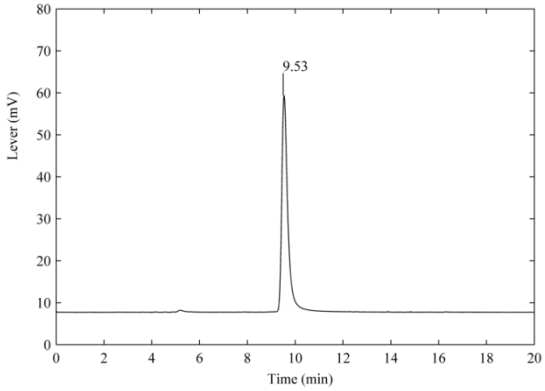


Figure AI.1. HPLC chromatogram obtained for EGCG (30mg/mL).

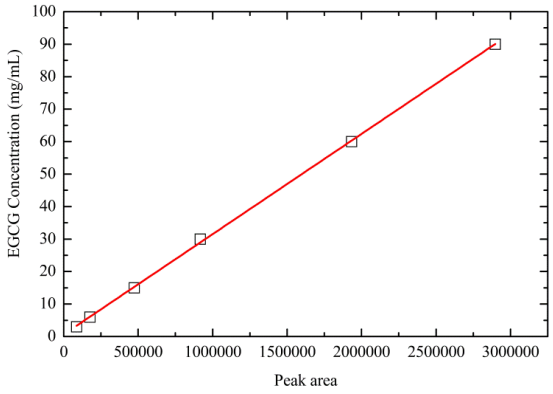


Figure AI.2. Standard curve used to determine the concentration of EGCG by HPLC.

AI.3.2. EGCG-keratin binding assay

As shown in Figure AI.3, the shape of the absorption equilibrium curve of EGCG and keratin binding implies multilayer binding equilibrium. The BET isotherm model developed by Brunauer et al. [Brauner et al. *J. Am. Chem Soc*. 1938; 60:309-319] is usually used to describe multilayer adsorption. However, the BET equation was developed to describe gas adsorption as a function of partial pressure.

For molecular adsorption from a liquid system, the GAB equation, or the modified BET equation is used to fit the data [Ebadi et al. *Adsorption*. 2009;15:65-73]:

$$q = q_m \frac{K_S C_{eq}}{(1 - K_L C_{eq})(1 - K_L C_{eq} + K_s C_{eq})} \quad \text{(AI.1)}$$

, where q is the amount of ligand bound per mole of protein (mol mol^{-1}); C_{eq} is the free ligand concentration (M); q_m is the amount of monolayer binding of ligand per mole of protein (mol mol^{-1}); K_s is the monolayer binding constant (M^{-1}); and K_L is the multilayer binding constant (M^{-1}).

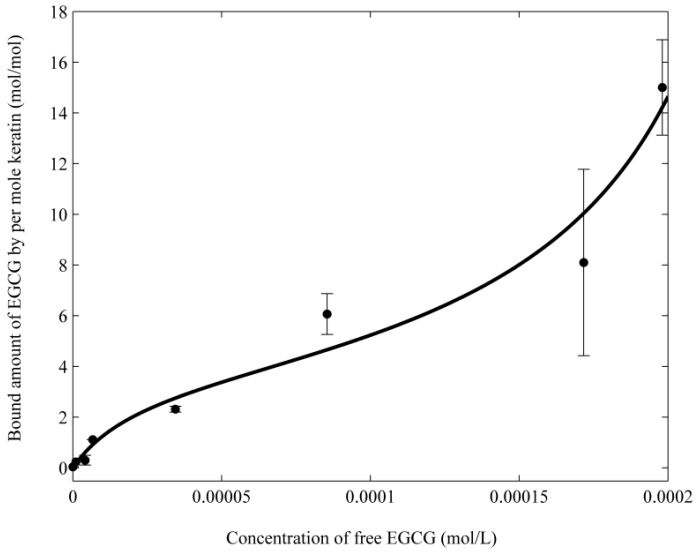


Figure AI.3. EGCG-keratin isotherm measured by ultrafiltration. Data is best fitted by GAB equation.

Table AI.1. Fitted isotherm parameters of EGCG-keratin binding by ultrafiltration.

Parameters	GAB fitting	95% confidence bounds
K_s (M^{-1})	4.62E+04	(3.46e+04, 1.27e+05)
K_L (M^{-1})	3770	(3548, 3991)
q_m (mol/mol)	3.70	(1.11, 4.88)
R^2	0.9920	-

The best fitted values of the GAB absorption parameters are listed in Table AI.1. When the GAB equation is used to describe the binding of two molecules in a liquid system, e.g. liquid phase adsorption [Ebadi et al. *Adsorption*. 2009;15:65-73] or protein binding [Poncet-Legrand et al. *J. Agric. Food Chem*. 2007;55:9235-9240], the binding constant is often quoted as in units of M^{-1}. However, for solid phase adsorption where the solid phase is fixed, a dimensionless binding coefficient (or distribution coefficient) is often quoted [Yadav et al. *Chem. Eng. Sc*. 2009;64:1480-1487].

From the binding constant of Table AI.1, the binding coefficient of EGCG to wool keratin can be also estimated as:

$$K_b = \left.\frac{q}{MW_{Keratin} C_{eq} / \rho_w}\right|_{C_{eq} \to 0} = \frac{q_m K_S \rho_w}{MW_{\mathrm{ker}\,atin}} \qquad \text{(AI.3)}$$

Where $MW_{Keratin}$ is the molecular weight of wool keratin and ρ_w is the specific density of water.

By considering the molecular weight of keratin as 50 kDa, the derived binding coefficient of EGCG to keratin is 3.53.

AI.3.3. ITC measurement

ITC data of EGCG-keratin binding are plotted as a graph of heat flux against time. The peak-by-peak values are integrated and normalized to obtain the enthalpy change per mole of injectant (ΔH, kcal mol^{-1}) against the molar ratio (EGCG/keratin).

Figure AI.4 shows the results of titration of EGCG (0.3027mM, at pH 6.0, 298 K) into keratin (0.0040mM, at pH 6.0, 298 K).

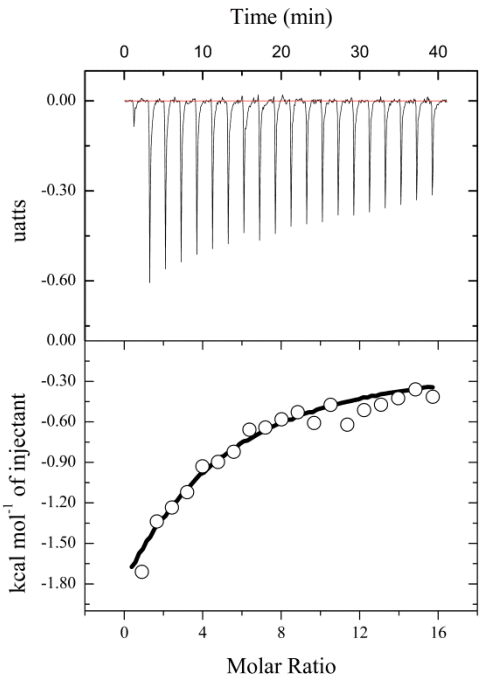


Figure AI.4. Titration of EGCG (0.3027mM) to keratin (0.0040mM) observed by ITC.

With multi-layer binding, the measured change of enthalpy should be resolved into two components, one associated with the mono-layer binding and the other associated with consecutive multi-layer binding. Assuming GAB equilibrium of binding, it can be derived that the data presented in Figure AI.4 can be fitted according to the following equation for the total heat for EGCG and keratin binding:

$$Q = M_t V_0 \left\{ q_m \frac{K_S C_{eq}}{(1 + K_S C_{eq})} \Delta H_1 + \left[q_m \frac{K_S C_{eq}}{(1 - K_L C_{eq})(1 - K_L C_{eq} + K_S C_{eq})} - q_m \frac{K_S C_{eq}}{(1 + K_S C_{eq})} \right] \Delta H_2 \right\} \quad \text{(AI.4)}$$

Where M_t is the concentration of keratin in the cell; ΔH_1 and ΔH_2 are the molar heat of EGCG-keratin interaction of the first and consecutive layer respectively; V_0 is the cell volume (200 µL); and C_{eq} is the concentration of free (unbound) EGCG, which is calculated from the total EGCG concentration and the bound EGCG concentration in the cell as follows:

$$X_t = C_{eq} + M_t q_m \frac{K_S C_{eq}}{(1 - K_L C_{eq})(1 - K_L C_{eq} + K_S C_{eq})} \quad \text{(AI.5)}$$

The measured enthalpy change of each injectant is:

$$\Delta H = (Q(i) - Q(i-1) + \frac{dv_i}{V_0}(\frac{Q(i) - Q(i-1)}{2}))/(C_{egcg} \times dv_i) \quad \text{(AI.6)}$$

Where C_{egcg} is the concentration of EGCG in the syringe, 0.3027mM; and dv_i is the volume of the i^{th} injection. In equations AI.4 – AI.6, q_m, K_s, K_L, ΔH_1 and ΔH_2 are unknown parameters. The values of q_m, Ks and K_L have been obtained from the ultrafiltration experiments. To prevent-over fitting, 95% confidence bounds of q_m, K_S and K_L are used as boundaries. Hence, we attained the values of q_m, K_s, K_L, ΔH_1 and ΔH_2, as listed in Table AI.2. The free energy of the first layer ΔG is related to the equilibrium constant by the following relationship:

$$\Delta G = -RT \ln K_S \quad \text{(AI.7)}$$

Where R is the universal gas constant (8.314 J mol^{-1} K^{-1}); and T is the temperature (K). Using equation (AI.7), the free energy corresponding to the equilibrium constant of the first layer binding is determined to be ΔG = -6.37 kcal mol^{-1}, which validates the value predicted by molecular dynamics simulations reported in Chapter IV (ΔG = -6.20 kcal mol^{-1}). The binding constant of the consecutive layer is an order of magnitude lower than that of the first layer and the corresponding binding free energy ΔG= -4.90 kcal mol^{-1}. For both the first and consecutive multi-layer, the ITC data showed ΔH<0 and ΔS>0. This suggests EGCG binding to keratin is governed by both hydrophobic interactions and hydrogen bonding which is consistent with the results from MD simulations. The enthalpy contributed to the majority of ΔG, indicating the binding process is enthalpy driven.

Table AI.2. Thermodynamic parameters of EGCG binding to keratin at 298 K. Data obtained by fitting to ITC data using GAB binding isotherm.

Temperature (K)	q_m (mol/mol)	K_S (M^{-1})	K_L (M^{-1})	ΔH_1 (kcal/mol)	ΔH_2 (kcal/mol)	ΔG_1 (kcal/mol)	ΔG_2 (kcal/mol)	ΔS_1 (kcal/mol/K)	ΔS_2 (kcal/mol/K)
298	**3.056**	**46759**	**3917**	**-4.81**	**-3.38**	**-6.37**	**-4.90**	**0.0052**	**0.0051**

AI.4. Conclusion

The equilibrium and thermodynamic properties of EGCG binding to keratin has been measured by ultrafiltration and ITC. The equilibrium data is best fitted by the GAB isotherm, suggesting multilayer EGCG binding to keratin rather than simple monolayer binding. The thermodynamic properties of EGCG binding to keratin have been characterized by ITC. The free energies calculated from both ultrafiltration and ITC are in good agreement with the result of molecular dynamic simulations reported in Chapter IV.

Fe^{2+} - Wool Keratin α-Helix Complex Studies via Isothermal Titration Calorimetry

Experiments conducted by Dr Yanyan Zhao (China Agricultural University, Beijing 100083, P. R. China). The work constitutes the experimental part of the publication:

Zhao Y, Marzinek JK, Bond PJ, Chen L, Qing Li, Mantalaris A, Pistikopoulos EN, Noro MG, Han L, Lian G. *A Study on Ferrous Ion Binding to Keratin Using Isothermal Titration Calorimetry and Molecular Dynamics Simulation*. Journal of Pharmaceutical Sciences, 2014. *DOI 10.1002/jps.23895.*

AII.1. Introduction

Isothermal titration calorimetry (ITC) has been applied in this study to investigate thermodynamic properties of the wool keratin with ferrous ion. Results presented in this appendix show that the interaction between Fe^{2+} and wool keratin are entropically driven and the increase of the temperature facilitate the binding process by the increase of the entropic contribution. The ITC data was best fitted by the one set of identical binding sites stating that only one type of the binding site is present.

AII.2. Methodology

AII.2.1. Materials

The bioavailability of compounds involving iron is determined by their solubility. Ferrous gluconate is known to have increased solubility hence used in this study. Ferrous gluconate dihydrate was obtained from J&K science, Beijing, China. The purity was 98%. Keratin samples were separated from sheep wool fibres and

kindly provided by Professor Jian Lu's group at Manchester University, UK.

The keratin in water solution was a mixture of two main components with molecular weights of about 46 kDa and 65 kDa. For the analysis of ITC data, a molecular weight (MW) of 50 kDa was suggested by the provider. Water used was tap water filtered with a Milli-Q system (Millipore, Beijing, China).

AII.2.2. Sample Preparation

Keratin samples were dissolved in water and the pH was measured to be at 6.08. The concentration of keratin solution was 0.5 mg/mL. By taking a MW of the keratin sample at 50 kDa, the concentration of keratin was 1.0×10^{-5} M. Ferrous gluconate solution (6.0×10^{-4} M) was prepared by dissolving ferrous gluconate dihydrate powder into pure water and the pH was measured to be 4.5. Ferrous gluconate was regarded as completely dissociated at pH 4.5.

AII.2.3. ITC Measurements

A nanoisothermal titration calorimeter (Nano-ITC low volume TA) was used. The solution of keratin was placed in a 190 µL sample cell and ferrous gluconate solution was loaded into the injection syringe. Both keratin and ferrous gluconate were degassed. The titration schedule included addition of 2 µL per injection of titrant with 25 injections. Each injection time was 5 s with 300 s intervals. The solution in the cell was mixed at 300 rpm by the syringe. Control experiments were also performed for the titration of ferrous gluconate solution into water. In order to remove the dilution and hydrolysis effect, the heat of the control experiments was subtracted from the titration experiment. Titrations of ferrous gluconate to keratin have been carried out at 25, 35 and 45 °C. Data analysis of ITC mesurements was conducted using Nano Analyze software provided by TA Instruments, including baseline-adjusted experimental titration data, peak-integrated background-subtracted and proper binding model fitting.

AII.3. Results

AII.3.1. ITC Studies on Fe^{2+} Binding

The ITC results of ferrous gluconate titration at 25°C are shown in Figure AII.1 as a plot of heat flux against time on the top, and on the bottom as peak-integrated, background-subtracted, concentration-normalized molar heat flow per aliquot versus injection numbers.

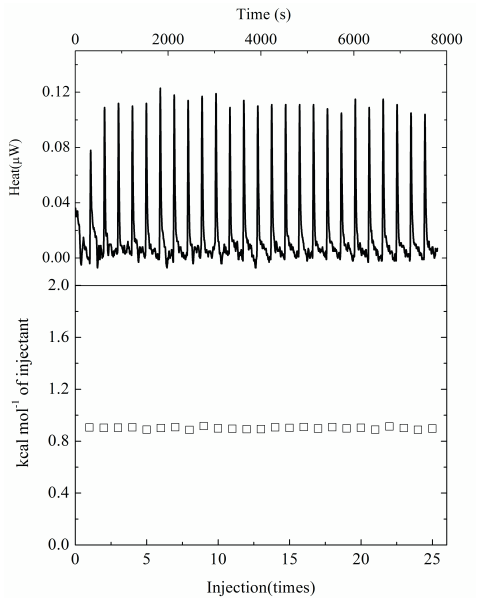


Figure AII.1. ITC measurement of ferrous gluconate ($6.0x10^{-4}$M, in pure water solution, 25°C) titrated into pure water.

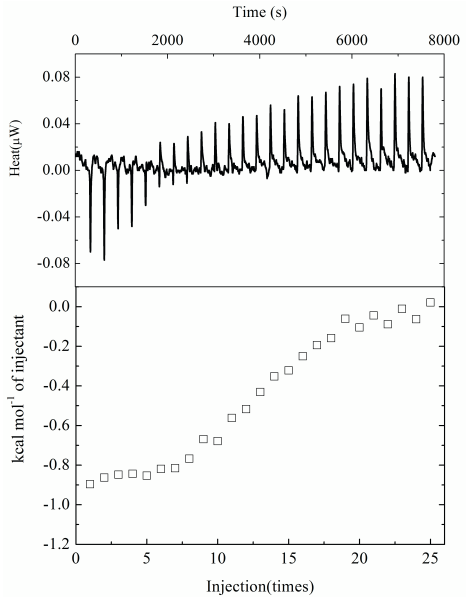


Figure AII.2. ITC measurement of ferrous gluconate ($6.0x10^{-4}$M, in pure water solution, 25°C) into keratin solution ($1.0x10^{-5}$M, in pure water solution, 25°C). Squared points corresponds to the observed enthalpy

changes of ferrous gluconate and keratin after subtracting the control experiment.

The control experiment for injection of ligands into buffer consisted of a series of equal heats of dilution. Ferrous gluconate dissociated after titration into water solution, and was an endothermic process. Figure AII.2 shows the results of titration of ferrous gluconate (6.0×10^{-4} M, dissolved in pure water solution at 25°C) into keratin solution (1.0×10^{-5} M, in pure water solution, 25°C). After subtracting the blank experiment of ferrous gluconate titration into water, the interaction of ferrous gluconate with keratin resulted as an exothermic process.

The one set of identical binding sites model is used to fit the ITC experimental data:

$$Q = \frac{n_S M_t \Delta H V_0}{2}[1 + \frac{X_t}{n_S M_t} + \frac{1}{n_S K_a M_t} - \sqrt{(1 + \frac{X_t}{n_S M_t} + \frac{1}{n_S K_a M_t})^2 - \frac{4X_t}{n_S M_t}}] \quad \text{(AII.1)}$$

$$\Delta Q = (Q(i) - Q(i-1) + \frac{dv_i}{V}(\frac{Q(i) - Q(i-1)}{2})) \quad \text{(AII.2)}$$

Here M_t is the overall concentration of keratin in the cell; X_t is the overall concentration of ferrous gluconate in the cell; Q is the total heat change of each injection; V_0 is the cell volume (190 μL); dv_i is the injection volume for the i^{th} injection. The fitting parameters for the interaction between ferrous gluconate and keratin, including enthalpy change (*ΔH*), number of binding ferrous ion per mol of keratin [mol/mol]

(n_S), and the binding constant (K_a). Figure AII.3 shows this model fitted the data well. It also shows that increasing temperature resulted in an increase of the binding free energy. All the fitted parameters (*n*, K_a and *ΔH*) and calculated free energy change (*ΔG*) and entropy change (*ΔS*) by euqation (AII.3) and (AII.4) are shown in Table AII.1.

$$\Delta G = -RT \ln K_a \quad \text{(AII.3)}$$

$$\Delta S = \frac{\Delta H - \Delta G}{T} \quad \text{(AII.4)}$$

Standard errors of these parameters were estimated by the data fitting to equation (AII.1) and (AII.2).

Table AII.1 reveals that 1 mole keratin can bind approximately 9 (8.59 ±0.25) moles of ferrous gluconate. The large positive entropy values obtained indicate that Fe^{2+} binding to keratin is entropically driven.

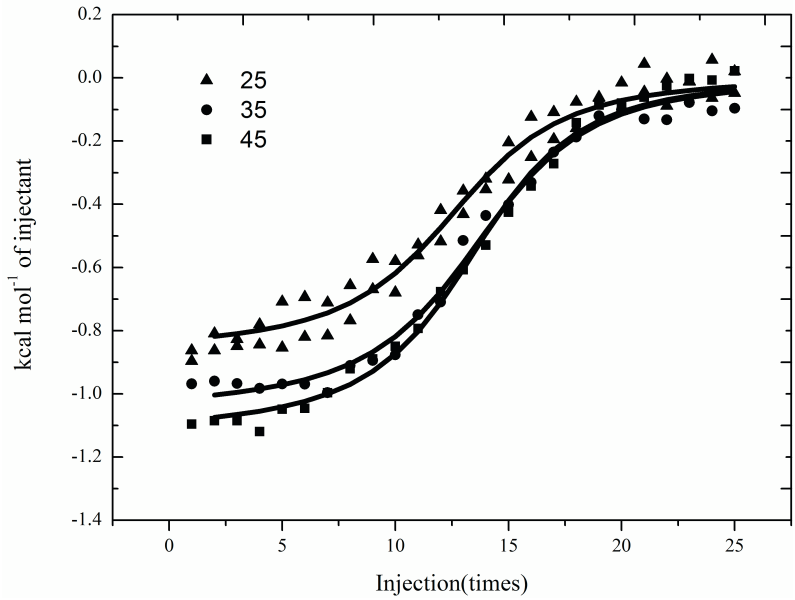


Figure AII.3. ITC measurement of ferrous gluconate ($6.0x10^{-4}$M, in pure water solution) into keratin solution ($1.0x10^{-5}$M, in pure water solution) after subtracting the control experiment. The solid line is the fitted enthalpy changes using a "one set of binding sites" model at 25, 35, and 45°C.

Table AII.1. Thermodynamic parameters for ferrous gluconate binding to keratin at different temperatures of 25, 35 and 45 °C. Data obtained by fitting the isothermal titration calorimetry data to equation (AII.1) and (AII.2). Results from molecular dynamics simulations (ΔG_{MD}) are also provided for comparison. Mean absolute error (MAE) of the free energy between experiment and molecular simulations is below 3.1%.

Temperature(°C)	n_S	ΔH[kcal/mol]	$T\Delta S$[kcal/mol]	K_a	ΔG[kcal/mol]
25	8.59±0.34	-0.86±0.05	6.57±0.14	$2.81\pm\mathbf{1.04}\times10^5$	-7.43±0.19
35	9.42±0.24	-1.05±0.04	6.64±0.13	$2.85\pm\mathbf{0.79}\times10^5$	-7.69±0.17
45	9.22±0.27	-1.12±0.05	6.88±0.18	$3.17\pm0.99\times10^5$	-8.00±0.23

AII.4. Conclusion

In summary, the Fe^{2+} and α-helix rich keratin binding in water environment was studied by ITC experiment. ITC experiments revealed that the interaction is exothermic and 1 mole of keratin can bind 9 moles of ferrous gluconate. Higher temperatures facilitated the binding process which is entropy driven. Experimental results were validated by molecular dynamics simulations in Chapter V and excellent agreement was reached.

Coarse-Grained Force Field Development and Application for Stratum Corneum Keratins – Experimental Validation

Experiment conducted by Dr Alfonso De Simone (Faculty of Natural Sciences, Department of Life Sciences, South Kensington Campus, SW1 2AZ, London, United Kingdom) at the University of Cambridge (Chemistry Department, Lensfield Road, CB 1EW, Cambridge, United Kingdom).

AIII.1. Introduction

The purpose of the experimental validation is to assess the secondary structure of the approximated surrogate terminal (60 residues). For this purpose circular dichroism (CD) was employed. In this chapter it is approached to confirm and validate full atomistic as well coarse-grain model from Chapter VI which clearly showed the unstructured terminal.

AIII.2. Materials and Methods

The keratin sample was synthesized by Life Technologies company with 96% purity measured by high performance liquid chromatography (HPLC). The amount provided corresponded to 10 mg. The solution was prepared by dissolving 2.2 mg in a 50 mM sodium acetate buffer. The pH of the solution was measured to be 4.5 (adjusted to the skin pH of 4.3-5.5). CD machine (Jasco J-810) was used for measurements which were conducted at 300K according to the simulations.

AIII.3. Results - Circular Dichroism

The result from CD is a multiple acquisition of 10 spectra and is presented in Figure AIII.1. At this wavelength (200-250 nm) the secondary structure can be assessed. The spectra shows typical graph for the random coil.

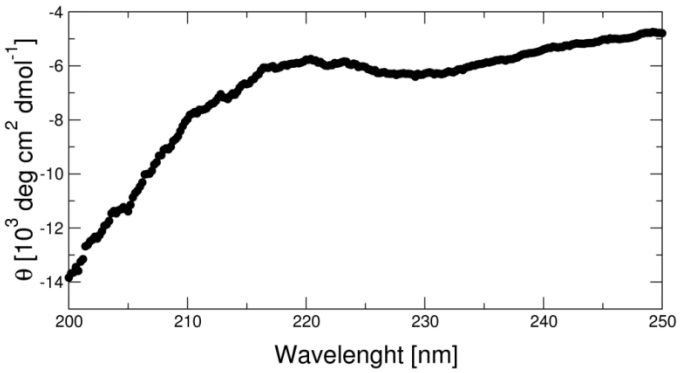


Figure AIII.1. Circular dichroism spectra obtained using 10 acquisitions. The shape of the curve is typical for a random coil state and excludes the presence of secondary structure.